Technical University of Crete

**Department of Electronics &
Computer Engineering**

Master Thesis

Subject:
**Statistical Evaluation of Image Segmentation Algorithms' Responses Employed to
Detect Corrosion Damage on Stonework**

Implemented by:
Kapsalas Petros (June 2006)

Supervisory Committee:
Prof. Zervakis Michalis   (Thesis Advisor)
Prof. Kalaitzakis Konstantinos (Co-Examiner)
Prof. Stavrakakis Georgios   (Co-Examiner)

<u>Acknowledgements</u>

I wish to express my sincere appreciation and thanks to Professor Michalis Zervakis for his guidance and his support during the implementation of this thesis as well as for his assistance on the preparation of this manuscript.

I am also grateful to Dr. Noni Maravelaki-Kalaitzaki for her invaluable guidance and her support throughout the duration of this work as well as for her important observations and suggestions on the improvement of the current thesis.

Last but not least, I wish to thank my family and especially my parents for their support and encouragement as well as my friends for sharing my thoughts worries and expectations during all these years.

Abstract

Corrosion damage of industrial materials and artwork objects form an aspect of high importance nowadays and gathers the interest of many researchers from different scientific fields. The main aim of these research efforts is to extract reliable information on the extent and the types of degradation and thus to propose techniques for effective reconstruction. A challenging issue in the field of corrosion damage estimation is the development of non-destructive to the material evaluation methodologies.

The current work introduces a novel approach of deterioration damage analysis based on computer vision techniques for non-destructive quantitative and qualitative evaluation of degradation effects on stonework. Thus, we have developed various segmentation approaches each of which handling in a different way the background in-homogeneities. The detection schemes, implemented in this work, aim at approaching accurately the topology of corrosion patterns while preserving their shape and size features. Thus, methods of adaptive thresholding, based on features of the local background, are initially employed while other techniques that involve Region Growing segmentation or fusion of detection results are also tested. The corrosion damage effects derived by the segmentation procedure are subsequently quantified by the means of several statistical metrics. In this thesis we are also focused towards the performance evaluation and the potential of segmentation processes in correctly detecting and localizing decay effects. A semi-automated framework for validating the algorithms' performance is thus developed. The framework implementation includes image dataset depicting representative decay effects, ground truth overlays, and source code for extracting ground truth matrixes and performance curves. This framework guarantees reliable and objective estimation of segmentation algorithms' performance while it allows informed experimental feedback for the design of improved segmentation schemes. Further to exploiting the robust points of each segmentation approach, this work also studies the corrosion mechanisms by investigating the way that degradation state is reflected onto the size of the segmented decay areas and their relative intensities over the background. At the final stage of this work we perform shape analysis on the segmented decay patterns. The analysis scheme is mainly based on boundary information and aims at investigating the way that cleaning state/and or exposure conditions are reflected on the segments' shape features. Furthermore, through studying the decay patterns' shape and in particular the existence of holes/and or nested regions within the body of the segmented areas we can track the occurrence of specific degradation mechanisms. Shape features considered in combination with size and intensity characteristics of degraded areas may aid the classification of corrosion damage.

Our detection methodologies and performance analysis framework is tested on a variety of images capturing from micro- to macro-scale characteristics of corrosion damage. Thus, the current work involves an examination of the limitations and the potential of various monitoring modalities to determine corrosion damage. The experts inspect the entire detection procedure and performance evaluation and the derived results proved to be in accordance with their own judgments and with previous chemical studies on the same surfaces.

Problem Specification

1.1. Introduction

The stone deterioration encountered on stonework has concentrated increasing concern of researchers during the last decades. The factors inducing stone decay can be discriminated into two main categories, according to their origin. Thus, they are associated either to weathering agents or to anthropogenic activities. With the development of industrial activities and the corresponding increase in air pollution, scientific studies on its effects became more frequent. Already in the 1930s several papers on this scientific topic were published. Later studies aimed at investigating the mechanisms of stone-pollutant reactions and at evaluating the deterioration rate of stones. Through the following subsections of 1st chapter we are focused towards examining the effects induced by the basic pollutants and their physico-chemical interactions.

1.2. Sulphation Processes

In modern urban atmosphere, sulphur dioxide (SO_2) attacks calcite $CaCO_3$ in calcareous stone producing gypsum ($CaSO_42H_2O$), which forms black crusts at rain sheltered surfaces and accelerates erosion at areas exposed to rain. The airborne particles accumulated on stone surfaces have always been considered to enhance the gypsum crust formation and thus it is believed that they should be removed from the surface to decrease the effects of SO_2. Black crusts presence on stone surfaces not only causes the aesthetical alteration of the stonework but also accelerate further decay due to their catalyzing action. Cammufo in [1] described black crusts as gypsum crystals with a radial structure normal to the underlying rock and a uniform distribution of black carbonaceous particles. These are the result of dry deposition of sulphur dioxide upon calcareous stone surfaces protected from rain. The black crust consists of carbonaceous particles from petroleum derivatives and clay minerals that are characteristic components of soil dust, sea spray of natural origin, and heavy metals [2-4]. The heavy metal compounds observed are mainly Fe, Ti, Pb, Mn, Cu, V, Cr, Zn, and Ni compounds. These various impurities on stone surfaces have long been considered to affect the gypsum formation in one way or another although their role has not been extensively studied. In some studies, the impurities are considered to accelerate the formation of gypsum. Cheng et al. [5] have observed that carbonaceous particles collected from combustion sources accelerated the sulphation reaction of marble. Sabbioni et al. [6] have studied this on marble, travertine and different mortar samples. However, Hutchinson et al. [7] found that fly ash emitted by the combustion of fuel oil or coal fire had no considerable effect on sulphation of limestone. Ausset et al. [8] have found that depending on the lithotype, fly ash had no or only a slightly accelerating effect on gypsum formation. Hutchinson et al. [7] have studied the sulphation in the presence of metal oxides on pure calcium carbonate powder and limestone. No significant influences on limestone sulphation by metal oxides were observed, whereas slight increases were found on pure calcium carbonate powder.

Further analysis on the structure of black crusts, revealed that their thickness varies from 100μm up to several mm according to the litho type and the atmospheric pollution. FTIR and XRD analyses on the black crusts revealed that we could determine sub-layers with different chemical composition and structural characteristics. The inner crust layers usually present empty cavities and residual particles of calcite as a result of the sulphation processes and the consequent migration of calcium ions (Ca^{+2}) outward from such sites.

The weathering factors (rain, temperature, winds, humidity etc) play an activate role in the development and the evolution of black crusts. More specifically, the black crusts absorb a great amount of humidity resulting in the development of lamellar texture and flaking of the stone. It is apparent that humidity's action in combination with the absorption of various metal oxides can lead to the abruption of the crust and extensively to loss of the stone material and aesthetic damage.

In the subsequent figures we can observe black crusts (screened under Secondary Electronic Microscopes (SEM)) as well as calcite crystals, metals and other decay patterns embedded into the crust's matrix.
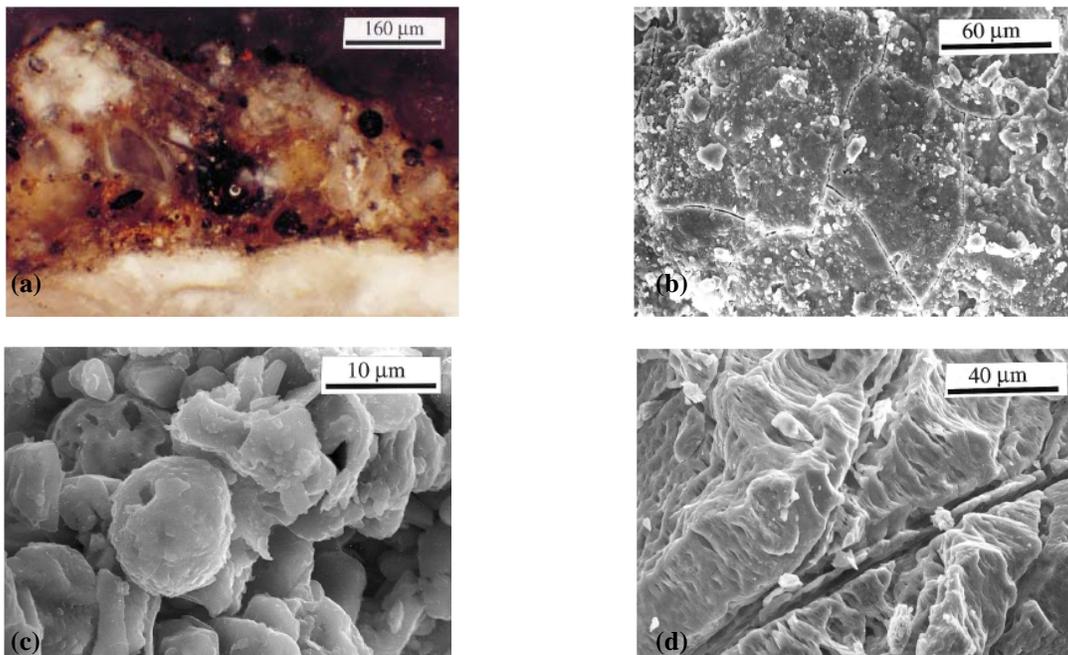
Figure 1: (a) Calcite crystals in the crust (polished cross-section), **(b)** Gypsum plaques on black crust with micro-cracks (SEM), **(c)** Spongy particles on the internal surface of the black crust (SEM), **(d)** Corroded calcite crystals on washed marble surface (SEM)

1.3.   Carbonaceous particles and the effects that they arise on stone monuments

The presence of fly ash and carbonaceous particles in the atmosphere has been reported by various authors [9], who established that they play an important role in the overall deterioration of materials [10]. The main environmental damage elect on the carbonate structural components and binders used in ancient masonry is the transformation of calcium carbonate into gypsum due to wet and dry deposition of SO_2. Grey-to-black crust formation is produced by gypsum crystals and atmospheric deposition, including carbonaceous particles which, because of their high specific surface and heavy metal content, act as catalytic support to the

heterogeneous oxidation of SO₂ [11]. The interaction between carbonaceous particles and stones was confirmed by studies involving simulation experiments in laboratory exposure systems [12]. The carbon compounds present in the alteration patinas on building materials may have four different origins:

1) Calcium carbonate, deriving almost exclusively from the underlying materials;

2) Deposition of atmospheric particles containing elemental and organic, primary and secondary compounds;

3) Biological weathering due to the action of microorganisms such as fungi and lichens, which produce oxalic acid that reacts with the underlying materials, leading to the formation of calcium oxalate;

4) Surface treatments (oils, waxes, etc.) frequently used in the past to protect historic monuments.

The deterioration phenomenology aroused by each of the above factors varies significantly. The deposition though of airborne particles, derived by the incomplete fuel combustion, plays the most important role. This is further evidenced by the high concentration of heavy metals and fly ash in crust specimens. Several investigations were carried out in the recent years to comprehend the action of carbonaceous particles and in general airborne particles carrying strong primary acidity. According to [13] the presence of fly ash in association with medium humidity levels lead to re-crystallization of the superficial gypsum layer. Optical inspection methods revealed that the appearance of coloration ranges from green to reddish-brown depending upon the duration of exposure and the chemical composition of the substrate. The catalyzing action of carbonaceous materials is also considered to be critical for the evolution of decay especially in urban and industrial areas. Many researchers claim that when a particle of ash is stuck on the stone surface it creates a hole around it by the acidity that it releases and it was observed that gypsum was crystallized in the margins of the crater.

The weathering phenomena such as wind's velocities and humidity levels enforce the action of the carbonaceous particles. Apparently, in areas where winds of high velocities occur, dust, smoke, and particles originating from oil and coal combustion are transferred in larger distances and with a brunt that enables the entrapment of black particles in the crust's matrix. This is mainly responsible for black color of crusts occurring in sheltered regions. Through figure 2(a)-(c) we can observe the occurrence of gypsum and other organic deposits in the crust's matrix.
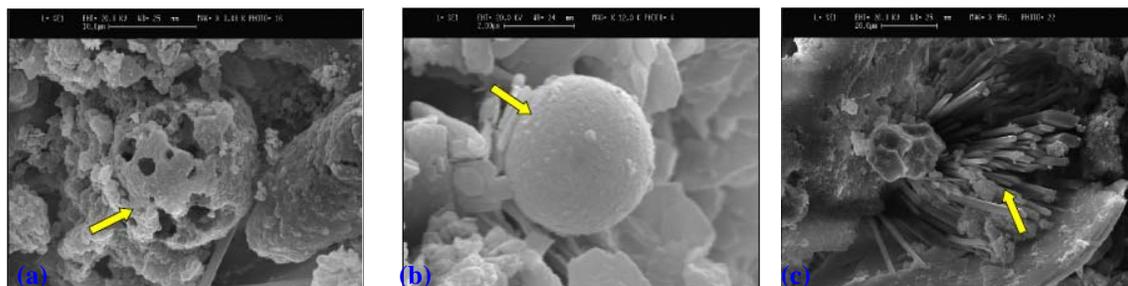
Figure 2: SEM micrographs of black crusts showing (a) fly ash particles derived from combustion of oil-based fuels and (b) alumino-silicate particles characteristic of coal combustion emission;(c) fibrous ettringite crystals at the interface between concrete and black crust.

Carbonaceous particles also affect areas submitted to rain's action by causing dissolution of carbonate minerals and intensifying the development of carbonate salts. Thus, we can observe zones of color alteration on the surface material while large pore sizes can also be observed due to the disintegration of the stonework. Unsheltered surfaces located in urban areas usually exhibit a reddish-brown to yellow layer perfectly adhering the stone surface, which is commonly referred to as skin of the marble.

1.4.   Biodeterioration of Stone

Biodeterioration processes on stones are one of several damage functions influencing the disintegration and destruction of historical monuments and sculptures. The impact of physical and chemical factors determined by exposure conditions such as moisture, atmospheric and anthropogenic influences, and the nature of the stone itself needs to be considered. Early investigations on stone biodeterioration considered that bio-corrosion follow the initial deteriorating effects of inorganic agents. These agents were thought to condition stone surfaces for microbial contamination due to structural changes and the enrichment of inorganic and organic nutrient substrates. However, recent investigations on the biodeterioration of stones, have found that bio-deteriorating effects can be clearly detected in the early stages of stone exposure.

The microbial colonization of stones commonly starts with phototrophic organisms, which build up a visible protective bio-film enriched with inorganic and organic biomass on the nutrient-depleted stone surface. Phototrophic microorganisms may grow on the stone surface or may penetrate some millimeters into the rock pore system. They do not seem to grow under thin stone. However, recent investigations revealed the presence of phototrophs even under rock scales a few millimeters thick providing shelter against desiccation and intense UV-radiation from sunlight [14]. Early investigations on the action of phototrophic microorganisms claimed that they do not have any direct effect on the deterioration of stone except for the aesthetically detrimental effect due to their pigments and that under certain climatic conditions they provided a protective film on the stone surface regulating humidity and temperature. However, more recent investigations have stressed the importance of phototrophs in the physical and chemical deterioration of stones, especially when fed by anthropogenic derivatives under moderate climates. These biodeterioration processes, are characterized by the excretion of organic acids (especially on marble and limestone), the uptake and accumulation of sulphur and calcium into their cells, the alteration of stone-forming minerals and the enlargement of pores due to the penetration of hyphae and roots, thus loosening stone particles from the parent rock material mainly on granitic rocks. This attack is intensified by the growth of biofilms weakening the mineral lattice by repeated wetting and drying cycles.

Algae and cyano-bacteria are usually the first colonizers of historical monuments due to their photosynthetic nature. These microorganisms can deteriorate stone either chemically or mechanically and their presence is generally detected through the formation of patina or crusts. Lichens, which are highly

resistant to desiccation and extreme temperatures, are also frequently associated with biodeterioration of stone, as are mosses and vascular plants.

The establishment of heterotrophic micro-flora on rocks is possible even without the participation of other bacteria. In this case the microorganisms use organic substrates derived by the rock material. The deposition of various particulate and organic compounds on stone surfaces from air-pollution as well as organic biomass contributes to the nutrient supply. Heterotrophic bacteria penetrate into the rock by hyphal growth and by bio-corrosive activity. Their deteriorating activity also includes discoloration of stone surfaces and mechanical stress to the stone structure.

Table 1 provides a summarization of the bio-deteriorative factors and the phenomenology that they arise.

Bio-Deteriorative Factor	Alterations Induced
Autotrophic Bacteria	Black Crust, black-brown layers, exfoliation, powdering
Heterotrophic Bacteria	Black Crust, black layers, exfoliation, color change
Actinomycetes	Whitish-gray powder, color alteration, white efflorescence
Cyano-Bacteria	Patinas and sheets of various colors and consistency
Fungi	Colored stains and patches, exfoliation, pitting
Algae	Patinas and sheets of various colors and consistency
Lichens	Crusts, Patches and pitting
Mosses and liverworts	Discoloration, green-gray patches
Higher Plants	Cracks, collapse, detachment of materials

1.5. Weathering Effects on the Deterioration of Stone Surfaces

Weathering conditions play an important role in the onset and the evolution of degradation phenomena. Several investigations were conducted with the aim to assess the corrosion mechanisms and the interaction between the climate conditions and pollutant-induced deterioration. Temperature and humidity levels were revealed to affect the structural integrity and color alteration of stone monuments. More specifically, the presence of temperatures not extremely high but of medium levels is capable of generating cracks on the stone's surface. This is mainly observed on marbles with a high content of dolomite and aluminum. As a result, these types of rocks usually develop a spongy texture. The atmosphere's moisture reinforces the gradual evolution of the desquamation of the stone. Furthermore, high temperatures are considered to be responsible for the color alteration (weakness) as well as for the formation of alveolate texture. The combined action of rain and winds leads to the dissolution of minerals, embedded into the stone matrix, and thus to the development of strains. In particular, the rain's action in combination with high humidity levels leads to disintegration of the stone material and more extensively to increase of the porous size and material loss (always depending upon the lithotype).

Variations on the degree of corrosion are observed amongst the sheltered and the unsheltered areas. Thus, it should be noted that in sheltered areas, where dry decay phenomena prevail, black crusts of more anomalous relief and friability are encountered. In the vast majority of the cases, water leaks and rain-wash cause discoloration of the stone and formation of orange-brown stains owing to the dissolution of carbonate

minerals. Frost also plays an important role as it penetrates the external rock layers and consequently it increases the friability of the rock.

1.6.   Non-Destructive Evaluation of Corrosion Damage

As it was discussed previously, the occurrence of black crusts on stone surfaces not only causes aesthetical damage but also leads to further degradation due to the catalyzing actions of its constituents. Furthermore, the cavities formed in the body of black crusts; function as locations of deposits accumulation. Thus, further dry deposition phenomena are intensified. Moreover, it should be stated that when black crusts absorb humidity it might lead to disintegration of the crust's structure and extensively to loss of the stone material. Thus, it is obvious that the development of chemical cleaning methods is essential not only for the restoration of degraded areas but also for preventing further corrosion phenomena. The cleaning methods applied on stone surfaces should aim at removing pollutant deposits while maintaining the color and the structural integrity of the stonework. Several investigations were carried out with the objective to develop accurate methods of removing decay effects. The key step for selecting the most appropriate restoration method is the determination of the decay effects origin and chemical composition. In other words, the diagnosis methods employed should approach accurately the basic features of degradation. The diagnosis procedures used so far are based on assessing the degradation state through chemical or electrochemical methods, which is always accompanied by obtaining stone specimen. Thus, it is considered to be destructive as it leaves the studied stonework more exposed to corrosion agents.

The implementation of non-destructive diagnostic approaches capable to be applied in-situ is currently an aspect of great concern. The diagnosis methods introduced in this work satisfy the above criteria. It involves obtaining images of the surfaces in concern through various Monitoring Modalities that provide screening of the specimen at various scales and thus studying different options of degradation. More specifically, the Monitoring Systems employed in this work is a (Fiber Optics Microscope, a digital camera system and a reflectography system operating at the visible (vis), infrared (ir) and near infrared (nir) spectral bands). The determination of the type and the severity of decay are based on accurate segmentation. Several statistical measures are introduced in this work to quantify corrosion effects. Further to assessing the severity of degradation we also approached the type and chemical composition of decay patterns. More specifically, the decay patterns are discriminated into two main categories. "Black Particles" which are associated to the presence of carbonaceous particles, dust and other organic/inorganic deposits and "White Particles" that represent gypsum crystals or re-crystallized $CaCO_3$. The type of degradation is approached through studying the size distribution of decay patterns, their shape features and spatial arrangement.

1.7.   Related Work

Several works have been done with the objective of extracting information regarding corroded areas on artworks. The most of the work, though, is focused towards assessing the occurrence of degradation effects on old paintings and less on detecting corrosion damage on stonework. This is mainly observed due to the

large variations of corrosion effects (flaws, material loss, discoloration, black crusts etc.) and the lithotype diversity. Moropoulou et. al. in [15] introduced a non-destructive technique of assessing the occurrence of corroded areas on stonework. Through this approach, treated and untreated stone surfaces are monitored via a thermography screening system. Subsequently the histograms of the obtained images are extracted with the intention of estimating their temperature distributions and identifying the problematic areas. An automated approach of quantifying corrosion damage through measuring color alteration was also implemented by Lebrun [16]. Color alteration is quantified by computing the Euclidean distance in a (pseudo)-L*a*b* colour space. Many investigations for accurate geometric analysis of the material surfaces affected by degradation have also taken place [17]. At these cases, topographical acquisition of data is often used. Gelli et. al in [18] employ an automated approach known as "Shape from Shading Method" in order to perform reconstruction of degraded stone surfaces. Furthermore, methods of characterizing the stone structure and detecting regions of material loss were developed in the study of Moltedo et al. [19] while Boukouvalas et al. in [20] introduce computer vision techniques for the detection and classification of mineral veins encountered on ceramic tiles surfaces. Further to assessing the corrosion state through measuring colour alteration several investigations were also conducted to assess the disintegration of stonework. Mahadevan et. Al. in [21] introduces an automated system of locating cracks on stonework. The procedure employs Gabor filters to locate the topology of flaws and subsequently uses skeletonization and geometric correction to reduce noise instances and to eliminate splitting of cracks. Numerous investigations have been carried out thus far with the objective to evaluate corrosion damage on old paintings. Pappas and Pitas in [22], employed image processing (IP) techniques to diagnose corrosion defects and perform reconstruction of the digital image on the locations of degraded areas. Kokla et. Al in [23] also approached the aspect of old paintings decay by relating the distribution of watercolor/ink intensity, under visible and infrared radiation. The way that corrosion damage affects the structural integrity of aerospace materials has also concentrated great concern through the last years [24]. An early attempt to segment degraded areas on metals was performed in [25]. Through this approach, eddy currents and infrared thermography inspect the decay effects and the information gathered is fused with the employment of several statistical and probabilistic algorithms. A further approach aiming at recognizing corroded areas on aerospace materials and classifying them according to their type was introduced in [26]. Through this approach the image was transformed to HIS color space to calculate color damage while for evaluating the texture features of degradation the method of the co-occurrence matrix was employed. Five types of corrosion damage were classified through a probabilistic method of decision-making. A similar study is also reported in [27], focused towards recognizing the various defects encountered on a cold mill strip with the aid of binary decision.

1.8. Objectives of the Current Work

The initial objective of the current work was to develop image segmentation algorithms mainly based on texture and intensity characteristics in order to accurately detect the topology and extent of decay areas

while preserving their shape. More specifically, we implemented 8 image segmentation schemes that perform decay areas' detection by handling in a different way the background in-homogeneities. Thus we have implemented local processing algorithms that either apply dynamically selected thresholds or determine the thresholds according to hypotheses made on the local intensities distributions. Size and shape preservation are also considered as issues of high importance. Further to detecting decay patterns at their real topology and extent we also assigned labels to the segmented areas to quantify decay's presence. Such metrics are the number of segments, their size distribution, the relative intensity of decay areas over the background as well as their spatial arrangement.

At a subsequent step we have developed a semi-automated approach of extracting the Ground Truth Matrix and have evaluating each algorithm's response both at determining the exact topology of corrosion patterns and at approaching their extent. Through the algorithms' performance evaluation the role of experts is critical. The experts' role is mainly absorbed in inspecting the Ground Truth extraction processes while they also pose the criteria for determining . These criteria essentially determine the features of an appropriate segmentation approach. Usually the expert's opinion is absorbed in the process of Ground Truth Matrix estimation. In this work we consider the potential of a detector to segment all susceptible areas (even those that do not correspond to decay effects) as indicative of its efficiency. Such a response, is considered preferable by the experts. Besides the comparison of several algorithmic approaches, in this paper we investigate how exposure or even cleaning conditions are reflected in the size and the relative intensities of the corroded areas (over the background). This aspect is approached by using statistical tests to assess the significance of differences observed in the decay characteristics of the examined structures. These tests mainly contribute in exemplifying the mechanisms and the efficiency of chemical cleaning as well as in understanding the procedure of crusts' development.

The shape features of decay patterns detected on surfaces of different structural and cleaning state are also examined. Through the shape feature analysis we investigate how the structural and cleaning effects are reflected on the shape features of the segmented decay areas. Furthermore, through studying the decay patterns' shape we can assess the occurrence of specific phenomena that may have taken place in the evolution of corrosion. Shape features considered in combination with size and intensity characteristics of degraded areas may aid the classification of corrosion damage.

Structuring this thesis, we follow the directions of the current work. Thus, through chapter 2A & 2B we discuss the implemented segmentation algorithms by providing information on basic concepts of the theoretical background and details of the implemented segmentation procedures. Subsequently, in sub-chapter 2C we summarize the results derived by the application of the segmentation algorithms on the images test set. Chapter 3A discusses issues concerning the algorithms' performance evaluation (by providing information associated both to aspects of the theoretical background and related work). Moreover, through chapter 3A we also analyze the theory of tests of statistical significance and study the shape descriptors and

their usefulness in machine vision applications. Sub-chapter 3B provides information concerning the implementation of the Ground truth extraction and of the process that measures the algorithmic responses. The last sub-section of 3B discusses the process of shape features extraction. Finally, through chapter 4 we present the results derived by the algorithms' performance study and by the tests of statistical significance and shape analysis.

2. Segmentation Approaches

2.A. Theoretical Background of the Segmentation Algorithms

Image Filtering

When an image is acquired by a camera or other imaging modality, often the vision system for which it is intended is unable to use it directly. The image may be corrupted by random variations in intensity, variations in illumination, or poor contrast that must be dealt with in the early stages of vision processing.

At first we provide a brief review of discrete linear systems and frequency analysis and then we discuss some filtering techniques. The Gaussian smoothing filter is covered in depth.

2.A.1. Linear Systems

Many image processing operations can be modeled as a linear system:
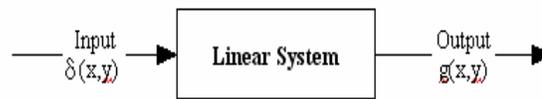
For a linear system, when the input to the system is an impulse $\delta(x,y)$ centered at the origin, the output $g(x,y)$, is the system's impulse response. Furthermore, a system whose response remains the same irrespective of the position of the input pulse is called a space invariant system:

A linear space invariant (LSI) system can be completely described by an impulse response $g(x,y)$ as follows:
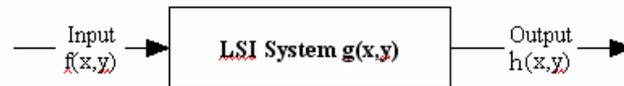
Where $f(x, y)$ and $h(x,y)$ are the input and output images, respectively. The above system must satisfy the following relationship:

$$af_1(x,y) + bf_2(x,y) \Rightarrow ah_1(x,y) + bh_2(x,y)$$

Where $f_1(x,y)$ and $f_2(x,y)$ are the input images, $h_1(x,y)$ and $h_2(x,y)$ are the output images corresponding to f_1 and f_2, and a and b are constant scaling factors.

For such a system, the output $h(x,y)$ is the convolution of $f(x,y)$ with the impulse response $g(x,y)$ and is

defined as: $h(x,y) = f(x,y) * g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')g(x-x',y-y')dx'dy'$.

For discrete functions, this becomes:

$$h[i,j] = f[i,j] * g[i,j] = \sum_{k=1}^{n}\sum_{l=1}^{m} f[k,l]g[i-k,j-l]$$

A	B	C
D	E	F
G	H	I

h[i,j]

$$\mathbf{h[i, j]= Ap_1+Bp_2+Cp_3+Dp_4+Ep_5+Fp_6+Gp_7+Hp_8+Ip_9}$$
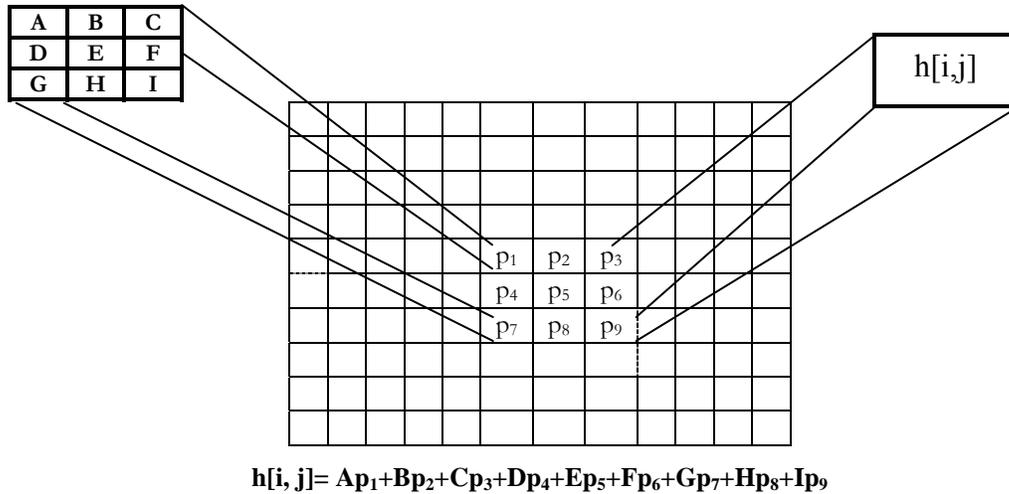
Figure 1: An example of a 3x3 convolution mask. The origin of the convolution mask corresponds to location E and the weights A,B,…,I are the values of g[-k,-l], k,l=-1,0,+1.

If f and h are images, convolution becomes the computation of weighted sums of the image pixels. The impulse response, g[i, j], is referred to as a convolution mask. For each pixel [i, j] in the image, the value h[i,j] is calculated by translating the convolution mask to pixel [i,j] in the image, and then taking the weighted sum of the pixels in the neighborhood about [i, j] where the individual weights are the corresponding values in the convolution mask. This process is illustrated in figure 1 using a 3x3 mask. Convolution is a linear operation, since

$$h[i,j] = g[i,j] * \{a_1 h_1[i,j] + a_2 h_2[i,j]\} = a_1 \{g[i,j] * h_1[i,j]\} + a_2 \{g[i,j] * h_2[i,j]\}$$

for any constants a1 and a2. In other words, the convolution of a sum is the sum of the convolutions, and the convolution of a scaled image is the scaled convolution. Convolution is a spatially invariant operation, since the same filter weights are used throughout the image. However, a spatially varying filter requires different weights in different parts of the image.

2.A.2. Linear Filters

As it is known, images are often corrupted by random variations in intensity values, called noise. Some common types of noise are *salt and pepper noise*, *Gaussian noise*, and *impulse noise*.

Linear smoothing filters are good filters for removing Gaussian noise and in most cases, the other types of noise as well. A linear filter is implemented using the weighted sum of the pixels in successive windows. Typically, the same pattern of weights is used in each window, which means that the linear filter is spatially invariant and can be implemented using a convolution mask. If different filter weights are used for different parts of the image, but the filter is still implemented as a weighted sum, then the linear filter is spatially varying. Any filter that is not a weighted sum of pixels is a nonlinear filter. Non-linear filters can be spatially invariant; meaning that the same calculation is performed regardless if the position in the image is spatially varying. The median filter, presented in 2.A.3, is a spatially invariant non-linear filter.

2.A.2.1. Mean Filter

One of the simplest linear filters is implemented by a local averaging operation where the value of each pixel is replaced by the average of all the values in the local neighborhood:

$$h[i,j] = \frac{1}{M} \sum_{(k,l) \in N} f[k,l] \tag{1}$$

where M is the total number of pixels in the neighborhood N. For example, taking 3x3 neighborhood about [i, j] yields:

$$h[i,j] = \frac{1}{9} \sum_{k=i-1}^{i+1} \sum_{k=j-1}^{j+1} f[k,l] \tag{2}$$

We can observe that the convolution operation of subsection 2.A.1 is reduced to simple local averaging operations. This shows that a mean filter can be implemented as a convolution mask.

2.A.3. Median Filter

The main problem with local averaging operations is that they tend to blur sharp discontinuities in intensity values in an image. An alternative approach is to replace each pixel value with the median of the gray values in the local neighborhood. Filters using this technique are called *median filters*.

Median filters are very effective in removing salt and pepper and impulse noise while retaining image details because they do not depend on values, which are significantly different from typical values in the neighborhood. Median filters work in successive image windows in a fashion similar to linear filters. However, the process is no longer a weighted sum. For example, take a 3x3 window and compute the median of the pixels in each window centered around [i, j]:

1) Sort the pixels into ascending order by gray level.
2) Select the value of the middle pixel as the new value for pixel [i, j].

The results of median filtering using a 3x3 mask to eliminate salt & pepper noise is illustrated in fig. 2.

Figure 2: The results of a median filter on an image corrupted by salt and pepper noise. *Left:* Noisy image. *Right:* Smoothed image.

2.A.4. Gaussian Smoothing

Gaussian filters are a class of linear smoothing filters with the weights chosen according to the shape of a Gaussian function. The Gaussian smoothing filter is a very good filter for removing noise drawn from a normal distribution. The zero-mean Gaussian function in one dimension is:

$$g(x) = \exp\left(-\frac{\pi}{2\sigma^2}\right)$$ (3)

where the Gaussian spread parameter σ determines the width of the Gaussian. For image processing, the two-dimensional zero-mean discrete Gaussian function is described in (2).

$$g[i,j] = \exp\left(-\frac{\left(i^2 + j^2\right)}{2\sigma^2}\right)$$ (2)

Gaussian functions have five properties that make them particularly useful in early vision processing. These properties indicate that the Gaussian smoothing filters are effective low-pass filters from the perspective of both the spatial and frequency domains, are efficient to implement, and can be used for practical vision applications. The five properties are summarized below.

1. In two dimensions, Gaussian functions are rotationally symmetric. This means that the amount of smoothing performed by the filter will be the same in all directions. In general, the edges in an image will not be oriented in some particular direction that is known in advance; consequently, there is no reason a priori to smooth more in one direction than in another. The property of rotational symmetry implies that a Gaussian smoothing filter will not bias subsequent edge detection in any particular direction.

2. The Gaussian function has a single lobe. This means that a Gaussian filter smooths by replacing each image pixel with a weighted average of the neighboring pixels such that the weight given to a neighbor decreases monotonically with distance from the central pixel. This property is important since an edge is a local feature in an image, and a smoothing operation that gives more significance to pixels farther away will distort the features.

3. The Fourier transform of a Gaussian image has a single lobe in the frequency spectrum. This property is straightforward corollary of the fact that the Fourier transform of a Gaussian itself is itself a Gaussian as will be shown below. Images are often corrupted by undesirable high-frequency signals (noise and fine texture). The desirable image features, such as edges, will have components at both high and low frequencies. The single lobe in the Fourier transform of a Gaussian means that the smoothed image will not be corrupted by contributions from unwanted high-frequency signals, while most of the desirable signals will be retained.

4. The width, and hence the degree of smoothing, of Gaussian filter is parameterized by σ, and the relationship between σ and the degree of smoothing is very simple. A larger σ implies a wider Gaussian filter and greater smoothing. Engineers can adjust the degree of smoothing to achieve a compromise between excessive blur of the desired image features (too much smoothing) and excessive undesired variation in the smoothed image due to noise and fine texture (too little smoothing).

5. Large Gaussian filters can be implemented very efficiently because Gaussian functions are separable. Two-dimensional Gaussian convolution can be performed by convolving the image with a one-dimensional filter oriented orthogonal to the Gaussian used in the first stage. Thus, the amount of

computation required for a 2-D Gaussian filter grows linearly in the width of the filter mask instead of growing quadratically.

2.A.5. Mathematical Morphology

Mathematical morphology gets its name from the study of shape. This approach exploits the fact that in many machine vision applications, it is natural and easy to think in terms of shapes when designing algorithms. A morphological approach facilitates shape-based, or iconic, thinking. The fundamental unit of pictorial information in the morphological approach is the binary image.

The intersection of any two binary images A and B, written $A \cap B$ is the binary image that has 1 at all pixels p that are 1 in both A and B. Thus,

$$A \cap B = \{p \mid p \in A \text{ and } p \in B\}.$$

The union of A and B, written as $A \cup B$, is the binary image that which is 1 at all pixels p which are 1 in A or 1 in B (or in both). Symbolically,

$$A \cup B = \{p \mid p \in A \text{ or } p \in B\}.$$

Let Ω be a universal binary (all 1) and A in a binary image. The complement of A is the binary image, which interchanges the 1s and 0s in A. Thus,

$$\overline{A} = \{p \mid p \in \Omega \text{ and } p \notin A\}.$$

The vector sum of two pixels in p and q with indices [i, j] and [k, l] is the pixel p+q with indices [i+k, j+l]. The vector difference p-q is the pixel with indices [i-k, j-l]. If A is a binary image and p is a pixel, then the translation of A by p is an image given by

$$A_p = \{a + p \mid a \in A\}.$$

2.A.5.1. Dilation

Translation of a binary image A by pixel p shifts the origin of A to p. If $A_{b1}, A_{b2}, \cdots, A_{bn}$ are translations of the binary image A by the pixels of the binary image $B = \{b_1, b_2, \cdots, b_n\}$, then the union of the translations of A by the 1 pixels of B is called the dilation of A by B and is given by

$$A \oplus B = \bigcup_{b_i \in B} A_{b_i} \tag{3}$$

Dilation has both associative and commutative properties. Thus, in a sequence of dilation steps the order of performing operations is not important. This fact allows breaking a complex shape into several simpler shapes, which can be recombined as a sequence of dilations.

2.A.5.2. Erosion

The opposite of dilation is erosion. The erosion of a binary image A by a binary image B is 1 at a pixel p if and only if every 1 pixel in the translation of B to p is also 1 in A. Erosion is given by

$$A \ominus B = \{p \mid B_p \subseteq A\} \tag{4}$$

Often the binary image B is a regular shape, which is used as a probe on image A and is referred to as a *structuring element*. Erosion plays a very important role in many applications. Erosion of an image by a structuring element results in an image that gives all locations where the structuring element is contained in the image. During a dilation operation every pixel in the structuring element will be present in the final dilated image, including the pixel not contained in the original object. But during erosion operation the pixel at the origin of the structuring element will be removed because the entire structuring element is not within the object. Conversely, in the case where the entire structuring element does fit within the original object, there will be no change to the final dilated or eroded image (i.e., no pixels will be added or deleted at that point).

Dilation and erosion exhibit a dual nature that is geometric rather than logical and involves a geometric complement as well as a logical complement. The geometric complement of a binary image is called its *reflection*. The reflection of a binary image B is that binary image B' which is symmetric with B about the origin, that is

$$B' = \left\{ -p \middle| p \in B \right\} \tag{5}$$

The geometric duality of dilation and erosion is expressed by the relationships

$$\overline{A \oplus B} = \overline{A} \ominus B' \tag{6}$$

and

$$\overline{A \ominus B} = \overline{A} \oplus B' \tag{7}$$

Geometric duality contrasts with logical duality:

$$\overline{A \bigcup B} = \overline{A} \bigcap \overline{B} \tag{8}$$

and

$$\overline{A \bigcap B} = \overline{A} \bigcup \overline{B} \tag{9}$$

also called deMorgan's law.

Erosion and dilation are often used in filtering images. If the nature of noise is known, then a suitable structuring element can be used and a sequence of erosion and dilation operations can be applied for removing the noise. Such filters affect the shape of the objects in the image.

The basic operations of mathematical morphology can be combined into complex sequences. For example, erosion followed by dilation with the same structuring element (probe) will remove all of the pixels in regions, which are too small to contain the probe, and will leave the rest. This sequence is called opening. As an example, if a disk-shaped probe image is used, then all the convex or isolated regions of pixels smaller than the disk will be eliminated. This forms a filter that suppresses positive spatial details. The remaining pixels show those regions, which were too small for the probe, and these could be the features of interest, depending on the application.

The opposite sequence, a dilation followed by erosion, will fill-in holes and concavities smaller than the probe. This is referred to as closing. Again, what is removed may be just as important as what it remains. Such filters can be used to suppress spatial features or discriminate against objects based upon their size. The

structuring element used does not have to be compact or regular, and can be any pattern of pixels. In this way features made up of distributed pixels can be detected.

2.A.6.   Region Segmentation Methods

The segmentation problem is first defined in [27] and is repeated here for reference, Given a set of pixels I and a homogeneity predicate *P(.)*, find a partition S of the image I into a set of n regions R_i,

$$\bigcup_{i=1}^{n} R_i = I \cdot$$

The homogeneity predicate and partitioning of the image have the properties that any region satisfies the predicate

$$P(R_i) = \text{True}$$

for all I, and any two adjacent regions cannot be merged into a single region that satisfies the predicate

$$P(R_i \cup R_j) = \text{False}.$$

The homogeneity predicate P(.) defines the conformity of all points in the region R_i, to the region model. The process of converting a gray value image into a binary image is a simple form of segmentation where the image is partitioned into two sets. The algorithms for thresholding to obtain binary images can be generalized to more than two levels. The thresholds in the algorithms are chosen by the developers. To make segmentation robust to variations in the scene, the algorithm should be able to select an appropriate threshold automatically using the samples of image intensity present in the image. The knowledge about the gray values of objects should not be hard-wired into an algorithm; the algorithm should use knowledge about the relative characteristics of gray values to select the appropriate threshold. This simple idea is useful in many computer vision algorithms.

2.A.6.1.   Automatic Thresholding

To make segmentation more robust, the threshold should be automatically selected by the system. Knowledge about the objects in the scene, the application, and the environment should be used in the segmentation algorithm in a form more general than the fixed threshold value. Such knowledge may include:

- Intensity characteristics of objects
- Sizes of the objects
- Fractions of the image occupied by the objects.
- Number of different types of objects appearing in an image.

A thresholding scheme that uses such knowledge and selects a proper thresholding value for each image without human intervention is called an automatic thresholding scheme. Automatic thresholding analyzes the gray value distribution in an image, usually by using a histogram of the gray values, and uses the knowledge about the application to select the most appropriate threshold. Since the knowledge employed in these schemes is more general, the domain of applicability of the algorithm is increased.

Suppose that an image contains n objects $O_1, O_2, \cdots, O_n$, including the background, and gray values from different populations $\pi_1, \cdots, \pi_n$ with probability distributions $p_1(z), p_2(z), \cdots, p_n(z)$. In many applications, the probabilities $P_1, P_2, \cdots P_n$ of the objects appearing in an image may also be known. Using this knowledge, it is possible to rigorously formulate the threshold selection problem. Since the illumination geometry of scene controls the probability distribution of intensity values $p_i(z)$ in an image, one cannot usually pre-compute the threshold values. As we will see, most methods for automatic threshold selection use the size and probability of occurrence and estimate the intensity distributions by computing histograms of the image intensities.

Many automatic thresholding schemes have been used in different applications. These algorithms can be employed to handle cases such as light objects against dark background or dark objects in a bright background. Some algorithms can be generalized to handle object gray values from an arbitrary set of pixel values.

Limitations of the Histogram Methods

The histogram-based segmentation approaches are useful in those applications where objects have constant gray values. If the illumination is different in different parts of a scene, then a single threshold may not be sufficient to segment the image, even if the image contains only one object. In such cases one must use techniques that effectively partition an image, arbitrarily, and select thresholds for each sub-image independently. If the images are complex, these approaches will also perform poorly.

The most basic limitation on the histogram-based approaches is due to the fact that a histogram throws away spatial information about the intensity values in an image. The histogram describes the global intensity distribution. Several images with very different spatial distributions may have similar histograms. The global nature of a histogram limits its applicability to complex scenes. It does not exploit the important fact that points from the same object are usually spatially close due to surface coherence.

2.A.6.2.   Region Representation

Regions are used in many contexts and can be represented in many alternative forms. Different representations are suitable in different applications. Some applications require computations only for a single region, while others require relationships among different regions of an image. In this section, we will discuss a few commonly used representations of regions and study their features. It must be mentioned here that regions can also be represented as closed contours. Most region representations can be classified into one of the following three classes:

1) Array representations
2) Hierarchical representations
   - Pyramids
   - Quad Trees

---

3) Symbolic representations

2.A.6.3. Split and Merge Methods

A simple intensity-based segmentation usually results in too many regions. Even in images where most humans see very clear regions with constant gray value, the output of a thresholding algorithm may contain many extra regions. The main reasons for this problem are high-frequency noise and a gradual transition between gray values in different regions.

After the initial intensity-based region segmentation, the regions may need to be refined or reformed. Several approaches have been proposed for post-processing such regions obtained from a simple segmentation approach. Some of these approaches use domain-dependent knowledge, while other applications use knowledge about the imaging process. The refinement may be done iteratively by a person or automatically by a computer. In an automatic system, the segmentation will have to be refined based on object characteristics and general knowledge about the images.

Automatic refinement is done using a combination of split and merge operations. Split and merge operations eliminate false boundaries and spurious regions by merging adjacent regions that belong to the same object, and they add missing boundaries by splitting regions that contain parts of different objects. Some possible approaches for refinement include:

o Merge adjacent regions with similar characteristics.
o Remove questionable edges.
o Use topological properties of the regions.
o Use shape information about objects in the scene.
o Use semantic information about the scene.

The first three approaches use only information about image intensity combined with other domain-independent characteristics of regions.

2.A.6.4. Region Growing

In many images, the gray values of individual regions are not nearly constant and more sophisticated techniques must be used for segmentation. The best techniques are those based on the assumption that the image can be partitioned into regions that can be modeled by simple functions. This idea can be applied naturally for region segmentation.

The segmentation problem leads to an algorithm that starts with seed regions and then grows the regions to form larger regions satisfying these constraints. The homogeneity predicate can be based on any characteristic of the regions in the image such as average intensity, variance, texture or color.

The region growing approach begins by partitioning the image into *nxn* regions where n is typically between 5 and 9. Regions are merged if simple planar or bi-quadratic function can be fit to the gray values in both regions. The planar and bi-quadratic models are a linear combination of basis functions. The basis functions span the variable-order bivariate polynomials so the model is

$$f(x,y,a,m) = \sum_{i+j \leq M} a_{ij} x^i y^j,$$

where the order m of the model is restricted to $0 \leq m \leq 2$. This means that the region models are restricted to planar and bi-quadratic functions. The homogeneity predicate is based on the distance of points in a region from the function that models the region:

$$\chi^2(R,a,m) = \sum_{(x,y) \in R} d^2(x,y,a,m)$$

where the distance is ordinary Euclidean distance:

$$d^2(x,y,a,m) = [g(x,y) - f(x,y,a,m)]^2.$$

The gray value g(x, y) at point (x,y) in the image plane is the gray value of the pixel at that image location. Given a set of points R, the problem is to find the order m of the model and the model parameters that maximize the error function $\chi^2(R,a,m)$. This is a least squares problem that can be solved using singular value decomposition.

2.B. Approaches Employed for Segmentation

Principles and Requirements for Segmentation

The development of algorithmic approaches that detect accurately the locations where decayed areas occur, aids the extraction of reliable assessments on the extent of decay phenomena. The occurrence of noise effects as well as the in-homogeneity of the stone surface leads to the induction of false positive and false negative spots. The presence of these spots alters the estimation of the decay state, thus their elimination arises to be of high importance.

In order to design a detector that performs accurate determination of the decay areas we should at first identify the peculiarities of the problem.

- The objects of interest are very small. They are visible as dark or white particles in the FOM image.

- These small objects are in an inhomogeneous background reflecting the structure of the marble surface. The background structure may be brighter in some parts of the image than white particles on other parts. The same observation is also valid for the case of black particles. For this reason, a simple threshold method cannot be used for segmentation. The employed detector should take under consideration the local characteristics of the neighborhood of the image.

- Another problem is the typically low contrast of the very small objects to the background, which is close to the noise caused by the inhomogeneous stone structure. Due to the growth of the black and white particles there is no absolute lower bound to their contrast. Obviously the aim must be to be as sensitive as possible to the systematic variations caused by the deterioration patterns while suppressing those random variations caused by noise. This means that the segmentation has to take into account dynamically the local gray value variation.

In order to overcome the difficulties arisen by the limitations of the detection procedure, an efficient spot detector should take under consideration the following specifications.

- It should be insensitive to large-scale intensity variations. These are characterized by low spatial frequencies and are usually associated with the presence of mineral veins or other features of the stone.

- As the size of the spots is approximately known but may vary, the detector should be adapted to an expected size but should not be too specific. The prior assumption about the shape of the spots is that they are round resulting to an angular isotropic operator. Thus the segmented regions that are line or dot-shaped are not considered corresponding to deterioration patterns and thus they are eliminated by the employment of the appropriate morphological operator.

- Spots of high contrast should be detected even in an area of high noise level whereas in areas of low noise level spots of low contrast are expected to occur.

- Sub-areas that depict a non-uniformity of the underlying texture are more susceptible to be decayed.

To provide robust segmentation results based on the above specifications, we implemented and tested several algorithmic schemes that can be classified into different categories depending on the way they handle background in-homogeneities. The first step towards the implementation of an efficient spot detector is to decouple the detection of useful information from the background activity. This is achieved by the first algorithmic approach, which employs a broadband high-pass filter to enhance the decay areas and remove the general structure of the background. The segmentation process in this first approach is conducted through a simple thresholding technique that sets a global threshold from the statistical analysis of the entire image. The disability of such methods to eliminate the induction of false positive and false negative spots leads to the employment of the next category that uses adaptive thresholding schemes. Thus, we tested algorithmic approaches that perform thresholding based on characteristics of the local background structure using also some knowledge of the extent and spatial arrangement of decay patterns. All the above methods, however, use information from the histogram of the sub-regions in order to select an appropriate threshold. A fundamental limitation of such approaches is that they completely ignore information regarding the spatial relations of intensity values. In order to overcome this limitation, we also tested a local region growing segmentation approach. The basic goal here is to select local thresholds dynamically, based on an iterative evaluation of the labeling quality achieved by each threshold value. At each iteration, the initially selected area is grown according to a thresholding similarity predicate aiming at producing compact areas, while avoiding the merging of different regions. In an effort to further reduce the segmentation errors introduced due to the local background variations, we also implemented a more elaborate growing scheme that uses prior knowledge of the expected size of spots and the inter-spot distance. This procedure is quite reliable in detecting spot locations even in low contrast between the spot and its background. However, the detected shape is distorted and the boundary of the individual spots is smoothed. In order to address the effective shape detection of decay spots, we tested a category of local morphological operators. This approach

preserves the original spot shape, at the price of more false positive spots and merged spots that should be separated. In order to exploit the strength of both concepts (accurate topology detection and shape preservation) a morphological fusion algorithm was implemented that expands the areas detected by the local region growing approach up to the size derived by the morphological operators. These algorithms are briefly presented in the next section. In the subsequent diagram we present a classification of the detection processes developed in this thesis.
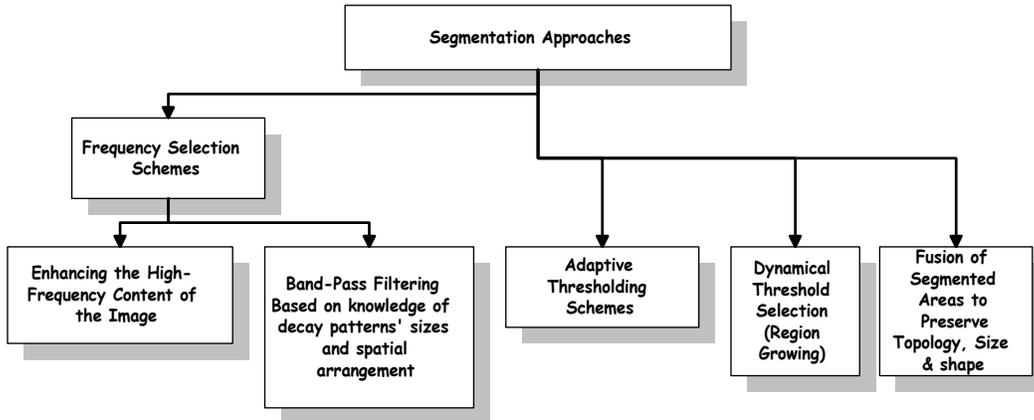
Diagram 1: Classification of the segmentation approaches.

Segmentation Approaches

2.B.1. Detection based on Frequency selection and Thresholding

In this section we discuss segmentation approaches that involve frequency-selective filtering followed by thresholding. Most algorithmic schemes employ high-pass filtering to enhance the discernibility of discontinuities of the stone structure. The high-pass filtering process removes the low frequency content of the image that mostly reflects the background activity. Following the filtering processes, they extract the histogram of the detail[1] image as to determine appropriate thresholds. Other algorithmic schemes induce band-pass filtering accompanied by a dual thresholding scheme. Through this process, they aim at maintaining on patterns with specific frequency content while suppressing background in-homogeneities associated to noise artifacts.

2.B.1.1. High-Pass Filtering and Thresholding

In order to perform image acquisition, a high-pass filtering scheme is employed. Considering the small size of the deterioration particles, a low pass filter with a wide kernel would be able to remove them while conserving the background of the image. Conversely, a high pass filter can be used to detect the decayed areas. A high pass filtered image can be derived as the difference between the original and a low-pass filtered version of the image. This can be written as follows.

$$f'(x, y) = f(x, y) - G_\sigma[f(x, y)]$$

---

[1] *Detail image is the image obtained through the frequency selection process.*

where f(x, y) is the original image and $G_\sigma[f(x, y)]$ and f'(x, y) represent the low-pass and the high-pass versions of the image respectively. As a low pass filter, a Gaussian filter with a wide kernel is employed. The parameters of the Gaussian filter are chosen to be suitable for the detection of the objects of interest. The size σ is selected to be larger than the expected size of the majority of the decay patterns. Here σ=2.75 was chosen, which means that the spatial variations at a scale larger than this is attenuated. The region of support of the Gaussian filter is 21x21 pixels. This size was selected in order not to eliminate the extent of the deterioration patterns. Subsequently, the high-pass filtered image is being thresholded in order to determine the areas where black or white particles occur. The procedure of evaluating the thresholds on the high pass filtered image starts by extracting the histogram. The form of the histogram is quite similar to the Gaussian function representation. The threshold values were determined to be located at the edges of the histogram by a rather empirical process. If Th_1 is located at the left of the histogram and Th_2 is located at the right, then all pixels with values greater than Th_2 are considered to comprise the white particles while the pixels with values lower than Th_1 are defined as black spots. An attempt to modify the threshold values would lead to the occurrence of more false positive or false negative spots. Further experiments on threshold adjustment revealed that the threshold values are located closely to the outliers determined by robust-fit approach. If these thresholds detect actual crust formation, the remaining pattern will be homogeneous. Other patterns detected are most likely due to noise and are expected to result in spurious isolated small random formation. Thus it arises to be essential the use of morphological filtering that would eliminate all the noise effects. The currently discussed detection process detects the decayed areas without considering the local gray value variations. This reduces the accuracy of the method as the false positives and the false negatives are increased.

2.B.1.2.   Weighted Difference of Gaussians (DoG) Detector
Principles for the design of the Method

The Difference of Gaussians Detector (DoG) employs a frequency selection process that performs band-pass filtering of the original image as to enhance discontinuities related to the presence of decay []. The detector makes use of the knowledge of the approximate size of decay spots. It also requires an idea of the inter-spots' distance. The precise knowledge of both sizes is not crucial however. In order to be independent of the local noise level a method is used that is adaptive to the local variations of the gray values. The idea is to give to two Gaussian convolution kernels different weights. We will initially analyze the detection of white spots and the detection of black spots requires the controversial procedure.

The positive kernel is assigned a weight *w* smaller than 1:

$$I' = \left(wG_{\sigma_+} - G_{\sigma_-}\right)*I$$

The decision criterion for a spot is:

$$I'(x, y) > 0.$$

This means that for a spot to be detected, the local average defined by the kernel of size σ_+ has to be larger by a factor $\frac{1}{w}$ than the local average defined by the kernel of size σ_-. The important point of this

---

where f(x, y) is the original image and $G_\sigma[f(x, y)]$ and f'(x, y) represent the low-pass and the high-pass versions of the image respectively. As a low pass filter, a Gaussian filter with a wide kernel is employed. The parameters of the Gaussian filter are chosen to be suitable for the detection of the objects of interest. The size σ is selected to be larger than the expected size of the majority of the decay patterns. Here σ=2.75 was chosen, which means that the spatial variations at a scale larger than this is attenuated. The region of support of the Gaussian filter is 21x21 pixels. This size was selected in order not to eliminate the extent of the deterioration patterns. Subsequently, the high-pass filtered image is being thresholded in order to determine the areas where black or white particles occur. The procedure of evaluating the thresholds on the high pass filtered image starts by extracting the histogram. The form of the histogram is quite similar to the Gaussian function representation. The threshold values were determined to be located at the edges of the histogram by a rather empirical process. If Th_1 is located at the left of the histogram and Th_2 is located at the right, then all pixels with values greater than Th_2 are considered to comprise the white particles while the pixels with values lower than Th_1 are defined as black spots. An attempt to modify the threshold values would lead to the occurrence of more false positive or false negative spots. Further experiments on threshold adjustment revealed that the threshold values are located closely to the outliers determined by robust-fit approach. If these thresholds detect actual crust formation, the remaining pattern will be homogeneous. Other patterns detected are most likely due to noise and are expected to result in spurious isolated small random formation. Thus it arises to be essential the use of morphological filtering that would eliminate all the noise effects. The currently discussed detection process detects the decayed areas without considering the local gray value variations. This reduces the accuracy of the method as the false positives and the false negatives are increased.

2.B.1.2.   Weighted Difference of Gaussians (DoG) Detector
Principles for the design of the Method

The Difference of Gaussians Detector (DoG) employs a frequency selection process that performs band-pass filtering of the original image as to enhance discontinuities related to the presence of decay []. The detector makes use of the knowledge of the approximate size of decay spots. It also requires an idea of the inter-spots' distance. The precise knowledge of both sizes is not crucial however. In order to be independent of the local noise level a method is used that is adaptive to the local variations of the gray values. The idea is to give to two Gaussian convolution kernels different weights. We will initially analyze the detection of white spots and the detection of black spots requires the controversial procedure.

The positive kernel is assigned a weight *w* smaller than 1:

$$I' = \left(wG_{\sigma_+} - G_{\sigma_-}\right)*I$$

The decision criterion for a spot is:

$$I'(x, y) > 0.$$

This means that for a spot to be detected, the local average defined by the kernel of size σ_+ has to be larger by a factor $\frac{1}{w}$ than the local average defined by the kernel of size σ_-. The important point of this

---

criterion is that it is invariant with respect to the scale of *I*. This means that a low contrast spot in an inhomogeneous background is detected equally well as a high contrast spot in an area of high noise level. In other words the threshold depends dynamically on the contrast ratio between the center part of the detector and the peripheral part.

Details of the implementation

The Gaussian detector consists of several steps. At first the original image f(x, y) is low pass filtered using a Gaussian kernel with standard deviation σ equal to 4 pixels.

$$f_1(x,y) = f(x,y) - G_4[f(x,y)] \tag{17}$$

The difference of Gaussian filtering consists of the subtraction of one smoothed version of the image from another having a different degree of smoothing. Two Gaussian kernels with different standard deviations are used to smooth the image. The standard deviations of the Gaussian kernels are chosen to reflect the dimension of the black and white particles and the inter-particle distances. The weighted version of Gaussian method is employed by assigning a weight of 0.8 to the kernel of larger width for the detection of black spots while the inverse procedure is followed for the detection of the white spots. The conditions used for the detection of black and white particles respectively are addressed by the following equations.

For the detection of black spots:

$$f_2(x,y) = 0.8G_6[f_1(x,y)] - G_{0.25}[f_1(x,y)] \tag{18}$$

For the detection of the white spots:

$$f'_2(x,y) = 0.8G_{0.25}[f_1(x,y)] - G_6[f_1(x,y)] \tag{19}$$

The images resulting from the Gaussian filters f_2 and f'_2 were segmented using the following procedure. The standard deviation of the filtered image was calculated and a first threshold equal to k_1 times this standard deviation was applied. Afterwards the standard deviation was recalculated by using only the pixels beyond the initial threshold. The final threshold was set as k_2 times the recalculated standard deviation. According to previous studies on detecting clusters of calcifications on digital mammograms [28], k_1 and k_2 should belong in the range [1, 3] if the standard deviation of the histogram of $f_2(x, y)$ is greater than 1. In our application these constants are selected experimentally and for the detection of white spots they are set to $k_1 = k_2 = 1.5$, while for the detection and of black spots k_1 and k_2 are set to 2 and 3, respectively. Following to the detection of decayed areas morphological filtering is performed in order to eliminate spots, lines and various other abrupt changes in the background that do not correspond to objects of interest. Since the Gaussian detector does not preserve the shape of the spots, this scheme provides reliable information about the location of a decayed area but not for its shape.

The previous global thresholding methods do not take under consideration specific features of the local background, thus inducing many false positive and false negative areas. The implementation of neighbor-based segmentation procedures that employ thresholds based on intensities of the neighboring

pixels is more efficient at suppressing instances of over-segmentation. The latter mainly occur due to the dynamically varying stone structure. In the subsequent sections we introduce three approaches of neighborhood-based segmentation that rely on stochastic hypotheses of the local intensity distributions.

2.B.2.   Segmentation Approaches Based on Local Thresholds

Neighborhood-based threshold selection aims at exploiting local characteristics to reduce the false positive and false negative segmentation areas, induced via the High Pass filtering. Initially, a detection scheme identical to this analyzed in 2.B.1.1 takes place. The detected areas are labeled in compact regions using an 8-neighborhood criterion. For each label the centroid's co-ordinates are calculated and stored. Subsequently, a window centered at the specific co-ordinates is applied on the detail image. The extent of the window is selected as not to eliminate the size of deterioration patterns while preserving a homogeneous background. In the sub-areas determined by the applied windows, the histogram is extracted and some statistical metrics indicative of the distribution of gray levels are evaluated. The derived values are then used to calculate thresholds. We test three methods of threshold selection that reflect different hypotheses regarding the local intensity distributions

- Initially, we assume a normal distribution of the local intensities. Thus the mean and the standard deviation are considered as representative measures of the intensities' distribution. In this case, the threshold is determined via the mean and the standard deviation values.

- A further hypothesis on the intensities values assumes non-parametric distribution. Thus the threshold depends upon the median and the quartiles values.

- Finally in the Robust Fit Thresholding approach we hypothesize that the local intensities obey to the normal distribution and a curve fitting approach is recruited to extract the outliers. The outliers in this case correspond to components of the histogram that depart from a normal distribution with the same mean and standard deviation as those measured from the sub-region's histogram.

2.B.2.1.   Thresholding by Using the Mean-Variance Criterion

Following to the application of the square window the histogram of the sub-area is extracted and the mean intensity value along with the standard deviation are evaluated. All the image locations that satisfy the equation $p(i,j) \leq (\text{Mean} - 1.5*\text{s}\tan\text{dard\_deviation})$ are considered to comprise black spots. The threshold applied in order to detect white particles is $\text{Th} = (\text{Mean} + 1.5*\text{s}\tan\text{dard\_deviation})$. Thus the pixels that satisfy the condition I>Th are treated as components of the white particles. The sequence of images illustrated below demonstrates the result image extracted at each stage of the algorithm. More specifically, figure 3(a) depicts the results derived after the application of the High Pass Filtering process while 3(b) and 3(c) illustrate the centers of gravity of the segmented areas and the decay patterns detected by the current method respectively.
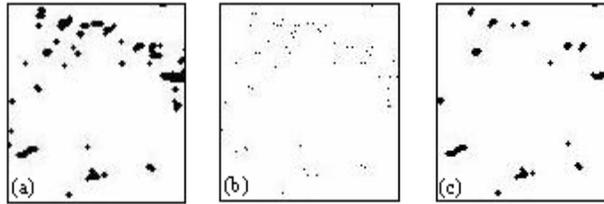
Figure 3: (a) Black particles detected by the High Pass Filtering Process, (b) Centers of gravity of the segmented black particles, (c) Black particles detected by the application of the Mean Variance Threshold.

By observing figure 3(b), it can be seen that a particle, which is seen as an entity by the human eye, is considered as a set of 2 or more adjacent labels by the Labeling Algorithm. Furthermore by observing the results illustrated above, it is obvious that the Mean-Variance Thresholding Method segmented deterioration particles, which are smaller in size. This assessment is justified by the fact that it performs more focused localization by using the search window and this way we avoid merge of adjacent areas and reduce incorrect segmentation.

2.B.2.2.   Thresholding by using the Box plot Outlier Criterion

The histogram-extracted statistical values in this case, are the median gray value, the lower quartile and the upper quartile. The threshold applied for the detection of black spots is Th_1= Upper_Quartile - 1.5*Inter_Quartile while the threshold value for the detection of white spots is Th_2= Upper_Quartile - 1.5*Inter_Quartile.

2.B.2.3.   Thresholding by using the Robust Fit Method

Finally, in the Robust Fit Thresholding approach we assume that the local intensities obey to the normal distribution and a curve fitting approach is recruited to extract outliers that depart from normal the distribution. At first we evaluate the distribution of gray levels in sub-region defined by the square window. Subsequently, a normal distribution is fit through robust-fitting as to avoid the effects of outliers. The robust fit function uses iteratively reweighed least squares algorithm and weights at each iteration are calculated by applying the bi-square function to the residuals from the previous iteration. This algorithm assigns lower weights to points that do not fit well the histogram. Subsequently, the weights derived from the above procedure are stored in a vector. While traversing the vector from the head to the end element, the position of the first nonzero element corresponds to the threshold denoted by the symbol Th_1. All pixels with gray values that are lower than Th_1 are detected as pixels that constitute black particles. The procedure followed to segment white particles is quite similar. More specifically, while traversing the same vector the position of the last nonzero element is defined to be equal to the upper threshold Th_2. All pixels with gray values beyond Th_2 are detected as points of white particles.

2.B.3.   Sub-Region Decomposition Algorithm

The sub-region decomposition approach initially employs frequency selection to enhance instances of stone structure abnormalities. Subsequently, a tool of selecting susceptible regions is developed. The

---

determination of these regions is based on texture features reflected on the sub-regions' histograms. More specifically, the histogram's Sub-Region Decomposition as representative measures of the intensities' distribution symmetry are employed. Regions with high Sub-Region Decomposition values usually represent areas of decay patterns occurrence and thus they are marked as susceptible. At the thresholding stage of the method, only areas with high Sub-Region Decomposition levels are checked via an adaptive thresholding scheme. The goal of the method is that it attains to reduce the computational time while avoiding the induction of false positive and false negative instances of segmentation (due to local processing).

Sub-Region Decomposition Based Tests

Initially, the image is high-pass filtered, as reported in 2.B.1.1. The derived image is identified under the term detail image. Since black and white particles are small isolated regions, they produce heavy outliers in the detail image. The problem is thus reduced to that of detecting outliers. The detail image is first divided into square non-overlapping regions of extent 41x41 pixels. The window's size was selected properly to consider the limitations of the detection procedure. More specifically, the window's extent should be quite larger than the objects of interest (black/white particles) in order avoid splitting the decayed areas. On the other hand, the window's size should not be too extensive because in such case specific characteristics of the local background are not preserved. In each of the decomposed sub-region the histogram is extracted and the Sub-Region Decomposition metrics as measures of the asymmetry and impulsiveness of the distribution are validated.

Sub-Region Decomposition are higher order statistics. In the following equations the definitions of these statistical parameters are reported. For a random variable X the skewness is given by the following equation.

$$\gamma_3 = \frac{E\left[(X - E[X])^3\right]}{\left(E\left[(X - E[X])^2\right]\right)^{\frac{3}{2}}} \tag{10}$$

And is a measure of the asymmetry of the distribution. An estimate of the skewness is given in equation (11).

$$\hat{\gamma}_3 = \frac{\sum_{i=1}^{N}(X_i - m)^3}{(N-1)\sigma^3} \tag{11}$$

By m and σ we indicate the estimates of mean and standard deviation over N observations of X_i (i=1…N). Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative the histogram components are spread more to the left than to the right. If skewness is positive, the data are spread more to the right. The skewness for the normal distribution is zero.

Similarly, for a random variable kurtosis is defined as:

$$\gamma_4 = \frac{E\left[(X - E[X])^4\right]}{\left(E\left[(X - E[X])^2\right]\right)^2} \tag{12}$$

Kurtosis is a measure of the tails of the distribution. An estimate of kurtosis is given by:

$$\hat{\gamma}_4 = \frac{\sum\limits_{i=1}^{N}(Xi-m)^4}{(N-1)\sigma^4} - 3 \qquad (13)$$

Where m and σ are defined as in (11). Kurtosis is a measure of how outlier prone a distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier prone than the normal distribution have kurtosis less than 3. If a region contains black/white particles then due to their impulsive nature the symmetry of the detail image histogram is destroyed. It is also evident that the tails of the distribution are heavier and hence the kurtosis assumes a high value. Therefore a statistical test based on Sub-Region Decomposition is effective in finding regions with asymmetrical and heavier tailed distributions. The detection problem is posed as a hypothesis problem where the null hypothesis, H_0 corresponds to the case that no problematic regions occur in the sub-region against the alternative H_1. This hypothesis-testing problem is reduced to the following decision rule based on the Sub-Region Decomposition.

$$\Gamma(x) = \begin{cases} 0 & \gamma_3 \leq T_1 \quad \text{or} \quad \gamma_4 \leq T_2 \\ 1 & \gamma_3 > T_1 \quad \text{or} \quad \gamma_4 > T_2 \end{cases} \qquad (14)$$

T_1 and T_2 are experimentally determined thresholds. Once the regions containing the deteriorative patterns are determined by the above test, the thresholding procedure estimates the locations where decay areas prevail.

Determination of Thresholds

After the decomposition of the image and the evaluation of Sub-Region Decomposition values, the image is reconstructed and the susceptible regions are examined in order to detect the presence of black or white spots. According to this method, in each sub-region to which the detail image is decomposed the histogram is extracted. Some values concerning the distribution of gray-levels are calculated from the histogram. More specifically, the computed values are: lower quartile (under the identification Q_1), upper quartile (denoted by Q_3), inter-quartile range (denoted by R_f) and median m. The Boxplot Outlier method of thresholding determines the outliers to be the part of data, which is outside the region (Q_1-kR_f, Q_3+kR_f). The parameter k is usually taken belong in the range [1.5, 3].

The algorithmic schemes discussed thus far are based on either global or adaptive thresholding schemes. However, the algorithm developer determines some of the threshold parameters. The implementation of segmentation approaches that employ dynamically varying thresholds may provide a more efficient discrimination between decay areas and noise artifacts. Through the Region Growing algorithm discussed below, the thresholds are dynamically selected and are based on iterative evaluation of labeling quality for each threshold value. The algorithm developer does not apply any external threshold as it is determined to be the one that induce the least change between two consecutive steps of the iteration.

2.B.4. Region Growing Algorithm

The Region Growing Algorithm starts by applying a high pass filtering process to the examined image. The high pass filtering is identical to the process reported in section 2.B.1.1. Subsequently, all pixels with intensity values under the median level are selected as seed pixels. A region is grown around a seed pixel by appending its 4 connected neighbours that satisfy the following condition.

$$p(i,j) \leq (1-t)\frac{F_{min} + F_{max}}{2} \tag{15}$$

Where $p(i, j)$ is the pixel being checked, F_{max} and F_{min} are current maximum and minimum values of the region being grown and t is the region growing tolerance parameter. The value of t is not selected by the user; the best t value is automatically derived for each segmented structure by repeating the growth with multiple values of t in the interval [0.01, 0.4]. The above repetition induces high computational complexity as the tolerance is varied from 0.01 to 0.4 with a step equal to the inverse of the seed pixel value. The value that introduces the least change to the feature vector from one step to the following is chosen as the optimal tolerance value. The features studied are the centre of gravity of the segmented regions and their size. The algorithm determines the value of t that results in the minimal change in the vector of two features with respect to the previous t value in the sequence by computing a normalized distance between consecutive vectors. The vector with the minimal distance indicates the best choice of t. The process followed for the determination of regions considered as white particles is ideal. Except for the fact that the condition used to check for inclusion is

$$p(i,j) \geq \frac{(1+t)*(F_{max} + F_{min})}{2} \tag{16}$$

Following to the detection procedure, a morphological filtering process is applied to eliminate noise artefacts as well as other abrupt changes of the background that do not correspond to deterioration patterns. The development of the Region Growing Algorithm was considered to be a reliable tool towards the diminishment of the false positive and false negative rate induced by the High-Pass filtering scheme. Furthermore it guarantees high detection accuracy as it maintains to local characteristics of the stone background by taking under consideration the grey value variations on the neighbourhood of seed pixels.

2.B.5. Detection via Morphological Operators

As it was previously mentioned, the Weighted Difference of Gaussians Detector accurately approaches the topology of decay areas. However, the shape of the spots is distorted: the boundary of the individual spots is smoothed. For further analysis, however, the original shape should be preserved, in particular when the spots have bizarre boundary. This is done with a morphological filter operation, the theoretical background of which is provided below.

The two basic morphological operations are erosion and dilation. These two operations are defined as follows:

$$B\ominus\Theta = \min\{B(x+m_1, y+m_2) - M(m_1, m_2) \,|\, (x+m_1),(y+m_2) \in D_B; (m_1, m_2) \in D_M\} \tag{20}$$

$$B\oplus M = \max\{B(x-m_1, y-m_2) + M(m_1, m_2) \,|\, (x-m_1),(y-m_2) \in D_B; (m_1, m_2) \in D_M\} \tag{21}$$

Equation (20) defines the erosion operation while equation (21) defines dilation. Furthermore, B(x, y) denotes an image, while M is the structuring element. According to the literature, dilation operation tends to eliminate or reduce the dark details encountered on an image, depending on the size and shape of the structuring element. Thus dilation is frequently used in order to erode the holes and extend the acres. In contrast, morphological erosion is employed for the removal of the bright details from the image; the latter correspond to areas that are smaller in extent from the structuring element. Dilation and erosion are complementary operations in the sense that dilation in the background is the same as erosion on the object. $M \ominus B = (M \oplus B^c)^c$. By combining erosion and dilation, the important morphological operations of opening and closing are computed. The definitions of opening and closing are given as:

$$M \circ B := M \oplus (M \ominus B) \tag{22}$$

$$M \bullet B := M \ominus (M \oplus B) \tag{23}$$

As it is obvious, opening and closing are dual operations and therefore, black and white particles are treated differently. Based on these operations Top-Hat and Bot-Hat transform of an image are defined by the equations (14) and (15) respectively.

$$R := B - (M \circ B) \tag{24}$$
$$R := B - (M \bullet B) \tag{25}$$

The top hat and bot-hat transforms enhance objects of interest that have size smaller than the structuring element. In the current approach a disk of 13 pixels diameter was employed as a structuring element. The top-hat transform is used for the detection of white particles while the bot-hat transform is used in the case of black spots. The thresholding scheme followed for the detection of deterioration patterns is ideal to this reported in the section 2.B.1.2. Figure 4 illustrates the result of detection through Morphological operations and DoG detector respectively.
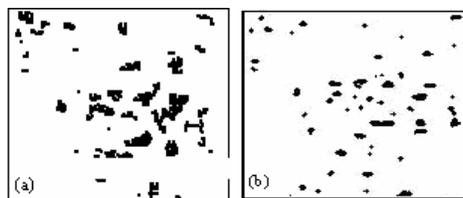
Figure 4: (a) Black particles detected on a stone surface by the Morphological operators. (b) The same surface after the application of the DoG detector.

A brief observation of the above images indicates that Morphological operators tend to merge adjacent decayed areas while preserving the features of their shape.

Finally, a process of fusing the results segmented by both DoG and Morphological detectors was developed in order to accurately approach the topology and the extent of degraded regions.

2.B.6. Reconstruction by Conditional Thickening

The shape of the spots determined by the morphological method is better preserved than the detection by the Gaussian detector, as far as false positive locations of spots and merged regions are concerned. The Gaussian detector determined the spots and their topology. In order to reconstruct the shape of the spots optimally, both methods are combined. The idea was to exploit the strength of both concepts, by detecting the spots with the Gaussian detector and expanding them, but not allowing their merge or grow beyond the size given by the result of the morphological filter operation. For this purpose, a morphological conditional thickening is applied. The operator $\otimes$ of X relative to Y with the pair of structuring elements (M_1, M_2) is defined as follows:

$$(M1, M2) \otimes XY = Y \cap \left(X \cup \left((M1 \ominus X) \cap (M2 \ominus X^c) \right) \right) \tag{26}$$

For the segmentation of the black and white particles, the structuring element M has the following structures:

$$M_{1,1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad M_{1,2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad M_{2,1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad M_{2,2} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

And every rotation of these matrices around 90^0.

The result of the conditional thickening of X onto Y with the structuring element M is the conjunction of the partial results of the equation above reported for every pair (M_{i1}, M_{i2}).

$$E = \bigcup_{i=1\cdots 8} M_i \otimes XY \tag{27}$$

The equation is applied for E=X until E does not change any more. The special property of M is that either X increases until the boundaries of Y are reached, or two subsets of X are separated through a line of one pixel in width. Consequently, E is always a subset of Y and contains as many objects as the intersection of X and Y. A conjunction of two objects of X is prevented by the structuring element M, because it extends X only with pixels of Y not disturbing the topology of X. This means that spots detected by the Gaussian detector are both extended by topologically unimportant pixels and the results are always intersected with the corresponding ones of the morphological method. In (27), X represents the result of the detection process and Y represents the result of the reconstruction one giving the shape of the spots. The intersection in each step assumes that after the conditional thickening, the remaining spots are present in both X and Y. Therefore, the intersection step determines the place and the number of the black\white particles, Y determines the shape, while the structuring element M prevents the confluence of several spots.

2.C. Segmentation Results

2.C.1. Experimental Setup

The studied images represent degraded stone regions monitored via a Fiber Optics Microscope (FOM) system and a digital camera. The studied FOM images depict sheltered and unsheltered areas obtained from the columns of the National Archaeological Museum (Athens). The images represent sheltered and unsheltered surfaces and are subdivided to those illustrating reedings and flutings. This discrimination is performed due to the different degradation and structural effects encountered on the corresponding surfaces. Thus, reedings represent areas more exposed to the rain and winds' action and consequently black crusts encountered on these areas tend to be thinner in thickness than the crusts encountered on the corresponding flutings surfaces. However, reedings present flaws and more granular texture due to the removal of stone grains. On the other hand, regarding the unsheltered areas, these tend to develop more lamellar texture and crusts thinner in thickness. The latter observation can be explained by considering that the water fluency results in removing the deposited materials. The discoloration of the unsheltered surfaces and the formation of reddish-brown or brown-black strains should be considered as an effect of the dissolution of the substrate due to water's action. This work further to assessing the structural effects of corrosion damage also aims at extracting the effects of cleaning interventions.

The applied cleaning treatments on the investigated marble surfaces (FOM) were an ion-exchange resin paste with deionized water (DS), a biological paste (BP) of 1000 ml deionized water, 50 gr $(NH_2)_2CO$, 20 ml $(CH_2OH)_2CHOH$ and approximately 800 gr sepiolite, and a wet micro-blasting method (WMB) springing spherical particles of calcium carbonate, with diameter lower than 80 μm, with a maximum function pressure of 0.5 bar. The proportion of water and spherical particles of calcium carbonate in the device's commixture barrel was 3:1. In order to assess the cleaning performance, chemical investigations with the aid of destructive techniques were also performed on the cleaned surfaces. The results of the chemical analysis are subsequently used to estimate the effectiveness of the cleaning methods in removing decay. In this work we assess the severity of degradation in terms of the size of the detected decay areas and the alteration of the relative (to the background) intensities on areas of corrosion damage prevalence.

The digital camera system is also recruited to investigate the modality's potential in accurately segmenting decay areas. The image screened via this system corresponds to a stone surface where adjacent strips of cleaned and uncleaned areas occur. The cleaning process was conducted by a Nd:YAG laser system used to partially remove the crust. The energy fluency of the Nd:YAG laser was fixed at 6.3 J/cm². Throughout the cleaning process, some parameters such as the laser pulses are modified resulting in the removal of crust layers differing in thickness. Each cleaned strip was obtained by increasing the number of laser pulses per spot from one up to six; a 40% area overlap was recorded between adjacent spots.

The evaluation of the derived results is performed through visually inspecting the segmented corroded areas and validating the statistical metrics. As it was discussed previously, the evaluation process is focused

---

towards estimating the effects of cleaning methodologies while identifying some features of degradation detected on various surfaces. The whole evaluation procedure is conducted by the experts.

2.C.2. Qualitative Evaluation

At first, we will present images depicting the segmentation of corroded areas. Due to reasons associated to space saving the images presented below, show decay areas detected only by the Conditional Thickening Algorithm. Figures 1 and 2 illustrate instances of marble surfaces where different corrosion damage prevails. At first, fig. 1 depicts an untreated surface located on a sheltered fluting while fig. 2 presents a stone surface located at the same areas after it has been treated by WMB method.
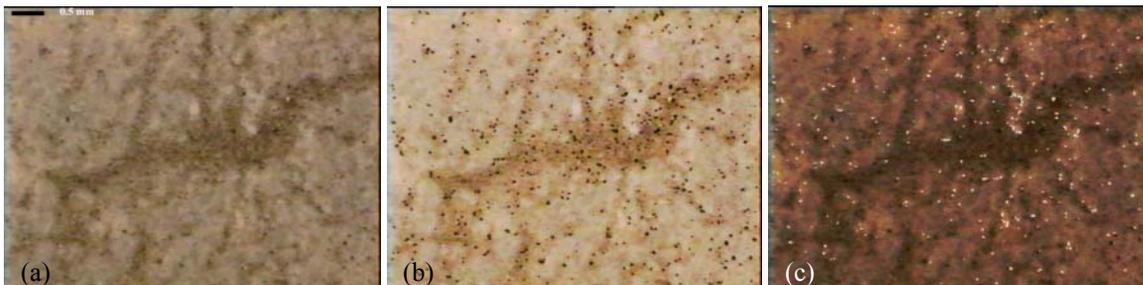
Figure 1: (a) untreated marble surface located at a sheltered columns fluting. (b) Segmented black particles overlaid on the original image. (c) Segmented white particles overlaid on the original image.
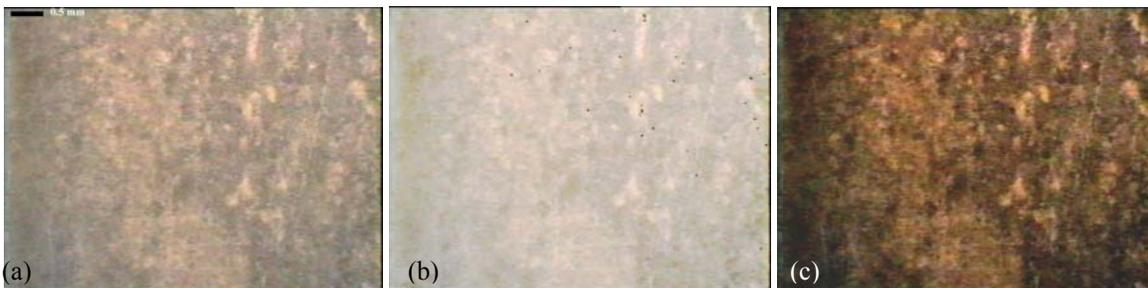
Figure 2: (a) Sheltered columns fluting treated by the WMB method (b) Segmented black particles overlaid on the original image. (c) Segmented white particles overlaid on the original image (No white particles were detected at this case).

By inspecting the above reported figures it can be observed that a significant diminishment of corrosion damage effects takes place after the cleaning process. The elimination of corrosion damage is illustrated by the lower extent and the smaller number of degradation patterns. Similar analysis is also performed to images derived by chemical cleaning with the aid of the other recruited methodologies (namely BP and DS). The assessment drawn concerning the efficiency of each of the cleaning methods verifies that all of them attain to remove the degradation patterns quite effectively. However, variations in the results still occur. Such variations are associated to the degree of corrosion that they remove regarding both the extent and the thickness of crusts. The peculiarities and the efficiency of the cleaning methods will be discussed more extensively in chapter 4.

Further to estimating the efficiency of the cleaning approaches, this thesis also aims at studying the types of corrosion encountered on surfaces of different exposure. Figure 3(a) and 4(a) illustrate decay on a marble surface located on an unsheltered fluting and a sheltered reeding respectively, while (b) and (c) depict black and white particles as they were segmented by the Conditional Thickening Algorithm.
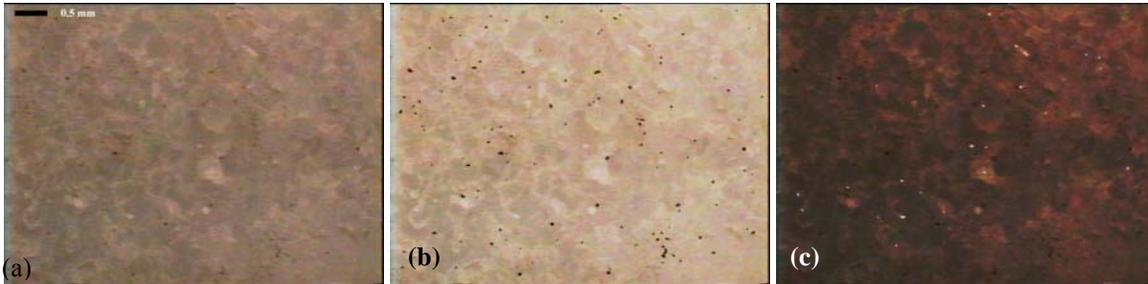
Figure 3: (a) untreated marble surface located at an unsheltered columns fluting. (b) Segmented black particles overlaid on the original image. (c) Segmented white particles overlaid on the original image.
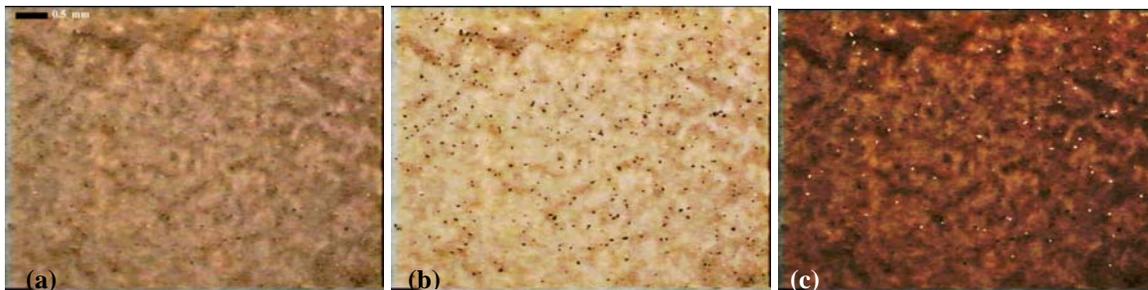

Figure 4: (a) untreated marble surface located at a sheltered columns reeding. (b) Segmented black particles overlaid on the original image. (c) Segmented white particles overlaid on the original image.

Visual evaluation of the segmented decay areas, leads to the assessment that corrosion effects seem to be more extended in sheltered untreated flutings while they appear to be diminished on images depicting sheltered untreated reedings. In particular, sheltered areas and column flutings accumulate the atmospheric deposition, while unsheltered areas and column reedings, being more exposed to rain's and wind's action, show lower amount of decay produce.

Through the subsequent section, the decay determined via the aid of the recruited algorithms is quantified. As it was mentioned above, the quantification process is conducted by introducing some statistical measures such as the number of the segmented decay patterns, the fraction of the studied surface covered by such spots of deterioration as well as the distribution of their size, thus exemplifying not only the extent but also the severity of degradation.

2.C.3. Quantitative Evaluation

The decay effects detected on the studied images are quantified by measuring the number of spots, the percentage of area covered by such spots and their average size and spatial distribution. Table 1 depicts the mean number of black and white particles detected on surfaces with different exposure to rain action, before and after cleaning (total 25 images).

Table 2-1: Number of black particles segmented on each of the studied surfaces

	High Pass Filtering	Region Growing Algorithm	Sub-Region Decomposition Algorithm	Mean Variance (Labeling Method)	Box Plot (Labeling Method)	Robust Fit (Labeling Method)	DoG Detector	Conditional Thickening Algorithm
Sheltered Untreated Flutings (Black)	856	1103	1026	843	700	800	924	744
Sheltered Untreated Flutings (White)	777	1303	1219	1244	1165	1065	923	705
Sheltered Treated Flutings (DS) (Black)	10	127	4	54	14	135	36	62
Sheltered Treated Flutings (DS) (White)	8	72	88	91	35	14	16	13
Sheltered Treated Flutings (BP) (Black)	26	40	38	24	17	23	5	23
Sheltered Treated Fluting (BP) (White)	18	56	19	68	42	26	44	48
Sheltered Treated Flutings (WMB) (Black)	42	23	10	54	20	48	23	3
Sheltered Treated Flutings (WMB) (White)	24	27	43	11	9	4	6	9
Sheltered Untreated Reedings (Black)	571	620	637	532	574	503	798	498
Sheltered Untreated Reedings (White)	322	237	261	234	256	223	291	232
Sheltered Treated Reedings (Black)(Ds)	20	51	43	38	29	26	58	10
Sheltered Treated Reedings (White)(Ds)	26	17	51	32	33	31	26	11
Sheltered Treated Reedings (Black)(BP)	20	33	35	34	34	26	27	19
Sheltered Treated Reedings (White) (BP)	15	18	15	20	16	20	40	30
Unsheltered Untreated Flutings (Black)	294	280	264	256	263	223	334	225
Unsheltered Untreated Flutings (White)	85	34	48	51	61	49	72	56
Unsheltered Treated Flutings (Black) (Ds)	38	79	60	46	22	39	35	14
Unsheltered Treated Flutings (White) (Ds)	15	8	10	19	18	20	14	11

Several conclusions can be drawn from Table 1. The mean number of either black or white particles expressing decay is significantly reduced after chemical cleaning. Sheltered surfaces and flutings show more decay patterns than unsheltered surfaces and reedings. These results can reasonably well be interpreted by the fact that sheltered areas and column flutings accumulate the atmospheric deposition, while unsheltered areas and column reedings, being more exposed to rain's and wind's action, show lower amount of decay produce. An attempt to compare the implemented algorithms in terms of the number of the segmented decay areas reveals the existence of differences. These mainly occur due to the discernibility provided by each of them. For instance, it was observed that the Sub-Region Decomposition as well as the DoG algorithms tends to split decay areas thus providing a greater number of regions as degraded. In particular, the above response of the Sub-Region Decomposition algorithm might be owed to the size of the window used. Through some experiments it was revealed that an automated adaptive selection of window's size (according to the texture variation) attains better determination of the deterioration patterns' topology and extent.

The percentage of surfaces covered by corrosion damage is also considered as a measure of the severity of decay. The results derived by measuring this metric are presented in table 2.

Table 2-2: Percentage of the studied stone surfaces covered by either black or white particles.

	High Pass Filtering	Region Growing Algorithm	Skew ness & Kurtosis Algorithm	Mean Variance (Labeling Method)	Box Plot (Labeling Method)	Robust Fit (Labeling Method)	DoG Detector	Conditional Thickening Algorithm
Percentage of Surface Covered by Black Particles								
Flutings Located Sheltered Areas								
Untreated Areas	2.82%	3.6%	2.27%	2.42%	2.57%	2.38%	2.5%	3.85%
Areas Treated by WMB	0.08%	0.07%	0.12%	0.09%	0.05%	0.11%	0.03%	0.02%
Areas Treated by BP	0.08%	0.12%	0.1%	0.07%	0.06%	0.11%	0.09%	0.08%
Areas Treated by DS	0.19%	0.24%	0.14%	0.09%	0.08%	0.4%	0.37%	0.28%
Reedings Located Sheltered Areas								
Untreated Areas	1.61%	1.57%	1.52%	1.34%	1.37%	1.25%	1.62%	1.83%
Areas Treated by BP	0.09%	0.05%	0.08%	0.09%	0.09%	0.06%	0.05%	0.07%
Areas Treated by DS	0.12%	0.15%	0.09%	0.1%	0.09%	0.07%	0.08%	0.07%
Flutings Located Unsheltered Areas								
Untreated Areas	0.63%	0.62%	0.51%	0.58%	0.49%	0.53%	0.47%	0.66%
Areas Treated by DS	0.08%	0.12%	0.09%	0.13%	0.07%	0.09%	0.05%	0.04%
Percentage of Surface Covered by White Particles								
Flutings Located Sheltered Areas								
Untreated Areas	2.95%	3.82%	3.61%	3.86%	3.32%	3.27%	1.87%	2.84%
Areas Treated by WMB	0.04%	0.05%	0.05%	0.07%	0.04%	0.06%	0.02%	0.03%
Areas Treated by BP	0.09%	0.09%	0.05%	0.14%	0.10%	0.06%	0.12%	0.11%
Areas Treated by DS	0.12%	0.11%	0.08%	0.22%	0.08%	0.05%	0.10%	0.07%
Reedings Located Sheltered Areas								
Untreated Areas	0.71%	0.78%	0.59%	0.73%	0.57%	0.64%	0.67%	1.02%
Areas Treated by BP	0.04%	0.03%	0.03%	0.03%	0.02%	0.05%	0.07%	0.04%
Areas Treated by DS	0.07%	0.04%	0.08%	0.08%	0.09%	0.09%	0.05%	0.04%
Flutings Located Unsheltered Areas								
Untreated Areas	0.14%	0.11%	0.12%	0.13%	0.11%	0.14%	0.12%	0.16%
Areas Treated by DS	0.04%	0.02%	0.02%	0.05%	0.05%	0.05%	0.03%	0.04%

It can be observed that the data presented in table 2 exhibit significant similarities with those presented in table 1. The effect of cleaning is reflected, by the recruited algorithms, through the drastic reduction of the surface's percentage covered by decay areas. The area measurements indicate that the DS method is less able of diminishing the percentage of the surface covered by black crusts. In addition, the surface measurements provide valuable information about the extent of decay encountered on areas with different exposure to climate conditions. It is revealed that sheltered flutings are covered by decay particles in a greater extent comparing with the other studied surfaces. In contrast, decay effects in a lower degree cover unsheltered areas.

Another interesting quantitative measure concerns the size distribution of decay particles. In general it can be observed that the severity of degradation is reflected on larger size of the segmented decay areas. Thus, black particles of larger extent were encountered on flutings than on reedings. This observation holds true for the surfaces located both on sheltered and unsheltered areas. The above assessment is quite

reasonable, since flutings function as areas of pollutants' accumulation, while reedings represent more washed-out areas. Thus, the results obtained are in accordance with the evaluation from chemical analysis.

2.C.3.1.   Study of the Decay Patterns Size Distribution

Through this subsection we investigate whether the structural and cleaning state is reflected onto the size of decay areas. Thus, after the detection procedure has been accomplished labels are assigned to the segmented areas and their extent and spatial distribution are calculated. The distribution of sizes is approached here in terms of the lower-quartile, median and upper-quartile decay area sizes. In table 3 we present the median value of the decay patterns extent as an indicative measure of their size distribution.

Table 2-3: Median of the decay patterns extent

	High Pass Filtering	Region Growing Algorithm	Skewness & Kurtosis Algorithm	Mean Variance (Labeling Method)	Box Plot (Labeling Method)	Robust Fit (Labeling Method)	DoG Detector	Conditional Thickening Algorithm
Flutings Located Sheltered Areas								
Untreated Areas	11	12	10	9	10	8	8	16
Areas Treated by WMB	5	8	8	5	5	5	5	10
Areas Treated by BP	8	10	8	8	8	10	8	12
Areas Treated by DS	8	8	8	8	8	8	8	8
Reedings Located Sheltered Areas								
Untreated Areas	10	8	8	10	10	8	8	13
Areas Treated by BP	5	8	5	8	8	5	4	9
Areas Treated by DS	8	8	8	8	10	8	3	10
Flutings Located Unsheltered Areas								
Untreated Areas	8	10	8	8	8	8	8	12
Areas Treated by DS	5	8	5	8	8	8	5	8

It can be seen that decay patterns of greater extent are encountered on sheltered flutings. This observation supports the experts' initial judgment that degradation state is more severe on these areas. As it was previously stated, these areas correspond to cavities of pollutants accumulation and thus black crusts of greater thickness is encountered. Another important assessment that can be drawn regards the decay patterns sizes after chemical cleaning. From table 3 we can conclude that decay patterns' extent is eliminated after chemical treatment. However, to establish conclusions based on these observations we should previously check the statistical significance of extent's alterations. Such approaches are introduced in chapter 4. Finally, it is revealed that the Conditional Thickening tends to detect areas larger in extent than the other algorithms. This is also an aspect that is more thoroughly investigated through chapter 4 and reflects the algorithm's ability to approach decay areas while preserving their size and shape characteristics

2.C.3.2.   Approach of the Crusts' Thickness

Up to this point, we investigate only area measurements on the detected decay patterns. However, intensity information from the original image on these patterns is also of great concern, since it relates with

the depth of the crust accumulated on the surface. In this study, the aspect of crust thickness is approached in a rather qualitative point of view in that darker formation implies more light absorption and, thus, thicker formation of black crust. Since the intensity distribution does not reflect heavy tails and resembles better the normal distribution, the reported measures reflect the mean and the upper and lower quartiles of the distribution over the particles of all images of the same type considered.

The results presented in Table 4 approach issues that can be studied through intensity distributions. First, it indicates the change in intensity levels before and after chemical cleaning with various methods. Then it reveals the effects of exposure as reflected to intensity mean. Finally, it summarizes the effect of cleaning in intensity distributions from the same piece of material (partially cleaned). Indeed, it is verified that the algorithmic approaches derive intensity distributions shifted to lower values when applied to surfaces with black crusts of higher thickness (sheltered flutings). Moreover, after chemical cleaning the intensity mean of the detected particles is increased, since the crust's thickness is diminished on these areas and thus they appear brighter and less disturbing in a macroscopic point of view. This result also indicates that even though chemical cleaning does not completely eliminate all decay formations, it manages to reduce the thickness of remaining crust patterns relative to their original state.

Table 2-4: Mean intensity of the studied surfaces as the recruited algorithms derived it.

	High Pass Filtering	Region Growing Algorithm	Skew ness & Kurtosis Algorithm	Mean Variance (Labeling Method)	Box Plot (Labeling Method)	Robust Fit (Labeling Method)	DoG Detector	Conditional Thickening Algorithm
Flutings Located at Sheltered Areas								
Mean Intensity Of black spots (Diagnosis)	55,560	48,603	63,793	55,601	53,235	53,255	58,063	57,89
Mean Intensity Of black spots (Biol Paste)	77,166	79,157	83,612	83,811	83,413	83,641	85,846	68,2
Mean Intensity Of black spots (WMB)	93,916	66,495	93,91	91,911	91,010	90,023	87,266	73,211
Mean Intensity Of black spots (Ds)	60,705	68	69,355	69,195	65,212	72,527	68,272	75,74
Reedings Located at Sheltered Areas								
Mean Intensity Of black spots (Diagnosis)	90,67	88,226	104,44	92,583	93,289	86,643	100,71	95,978
Mean Intensity Of black spots (BP)	103,9	103,57	110,36	107,73	110,19	110,9	112,41	101,12
Mean Intensity Of black spots (Ds)	96,63	95,71	108,35	101,37	98,716	97,182	105,11	105,148
Flutings Located at Unsheltered Areas								
Mean Intensity Of black spots (Diagnosis)	95,72	95,593	102,81	97,098	93,924	96,885	103,39	96,31
Mean Intens. Of black spots(Ds)	99,90	93,029	91,23	96,53	93,31	95,722	99,431	98,326

2.C.4. Corrosion Damage Estimation on Macroscopical Images

Further to segmenting deterioration effects encountered on FOM images, we are also interested in investigating the algorithms' efficiency when applied to images depicting macroscopical effects of degradation. Thus, we examine corrosion effects on images obtained by a digital camera system and a reflectography system operating at the *visible, infrared* and *near-infrared* spectrums of the band. Subsequently we illustrate a stone specimen depicting adjacent treated and untreated strips. The treated strips were obtained by

cleaning the stone material via the aid of Nd:Yag laser. Figure 5 illustrates the stone specimen discussed here. In the bottom of the images we can observe the successive co-occurrence of treated and untreated zones.
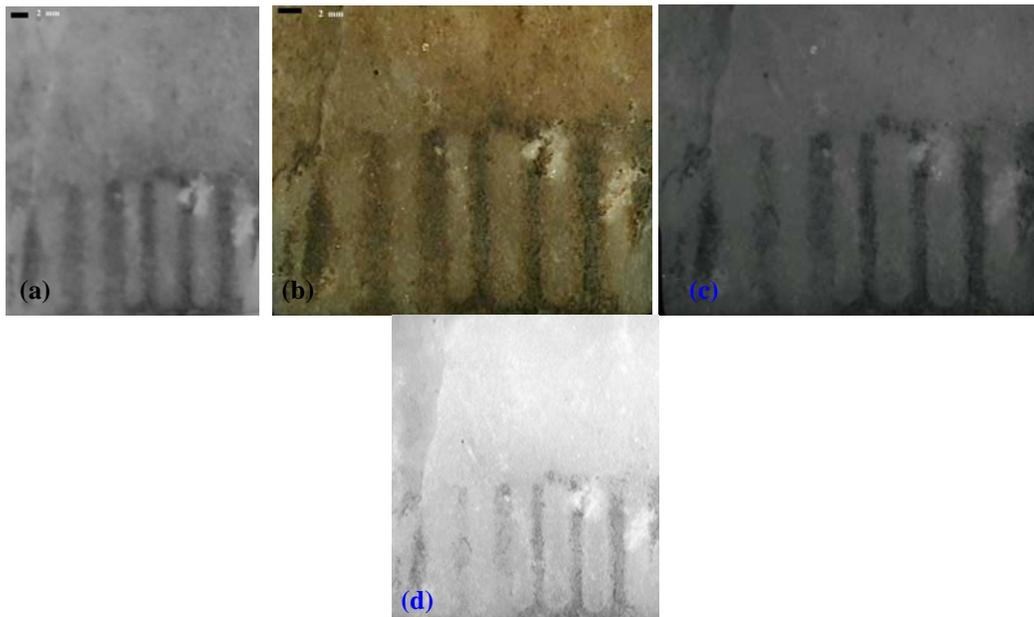

Figure 5: Images depicting the stone specimen as is it was monitored via the aid of (a) digital camera, (b) reflectography in the visible (c) infrared, (d) and near infrared spectral bands.

The studied images were inspected to detect the presence of black and white particles. Regarding the segmented areas found on the digital camera and the reflectography (vis) images we could state that the algorithms cannot effectively determine the white particles' presence. However, white particles are detected considerably better on the stone material monitored under (ir) and (nir) options of illumination. This can be interpreted by considering the small inter-particle distance between black spots and gypsum crystals in the crust matrix in association to the resolution of the digital camera. As it regards to the detection of black particles, we should notice that the responses of the algorithms tend to converge and this is expected if we consider the low resolution of the digital camera system that does not provide much information for texture features. To judge the observations stated above we present the decay patterns detected on 5(a) by the Conditional Thickening, the Region Growing and the DoG respectively.
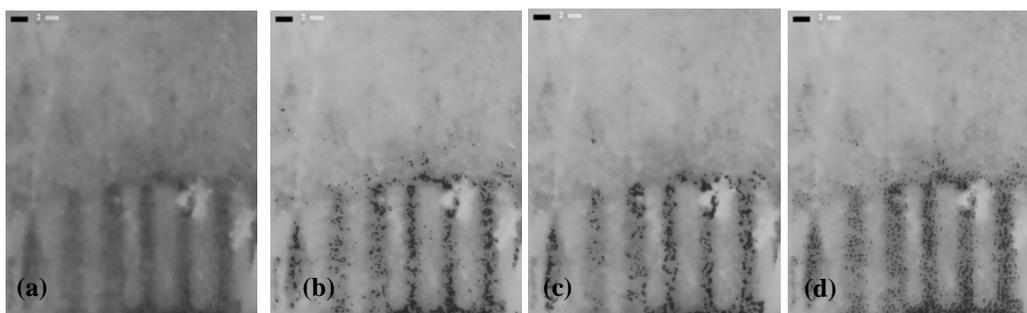
Figure 6: Detected decay areas overlaid in the original image (a) original digital camera image, (b) areas segmented by the Conditional Thickening overlaid on the original image, (c) Decay patterns detected by the Region Growing overlaid on the original image and (d) degraded regions detected through the DoG also overlaid on the original image.

It can be seen that the degraded areas determined by all 3 algorithms are arranged to cloze locations. This means that the algorithms accurately determine the topology of decay areas but segment areas of different extent. At this point, we should make clear that the objective of our detection processes is not to segment areas of intensity alteration induced by corrosion damage, but rather to determine the individual decay patterns appearing within any background structure (corroded or cleaned), which lead to the formation of black crusts beyond the color alteration effects. This applies especially for fig. 6 where segmentation does not aim to distinguish cleaned from corroded areas but to detect decay patterns on each of these areas.

3. Performance Evaluation Schemes

3.A. Theoretical Background and Related Work

3.A.1. Introduction

Performance evaluation of computer vision algorithms has received increasing attention in recent years [29-34]. This thesis aims at developing an automated framework for objective performance evaluation of region segmentation algorithms. Earlier work in performance evaluation of image segmentation algorithms compared algorithmic schemes that segment images via the aid of some parameters. In those comparisons, the training to select parameter values for the algorithms was done manually by the developers. Manual training produces results that are dependent on the knowledge, skill and effort of the algorithms' developer. Thus, an automated training and subsequent performance evaluation has proved [30] to provide performance at least as good as that produced with careful manual tuning by the humans. Through this chapter we discuss several aspects on the methodologies employed for evaluation of algorithms' potential to segment areas at their exact location and extent. The use of multiple test sets of images provides the basis for a test of the significance of performance differences between algorithms.

The Receiver Operating Characteristic (ROC) analysis comes from statistical decision theory and was originally used during World War II for the analysis of radar images. The first applications of this theory within the medical area occurred during the late 1960s. Today the ROC analysis is a widespread method in the medical field and many textbooks and articles have descriptions of it. From the computer science point of view, ROC analysis has been increasingly used as a tool to evaluate and discriminate effects among different methods. The ROC curve relies heavily on notations as sensitivity and specificity and these values depend on the specific data set. Even though the values for sensitivity and specificity in theory lie in the interval [0; 1], in practice the borders are decided by the data set.

Further to evaluating the performance of region segmentation algorithms and studying their differences in the segmentation procedure this chapter also introduces issues related to statistical analysis and statistical tests of significance. Statistical analysis is mainly concerned with determining in which extent the properties of a population **A** is related to the properties of another population **B**. Such analyses are more useful when observed differences are small compared to experimental imprecision and variability of the studied data. Statistical tests are formulated to assess whether the two populations differ in a corresponding property by a considerable amount ε. Such tests generally test hypotheses such as "The 2 populations do not differ significantly" (the latter statement is usually referred as null hypothesis). This hypothesis is tested by ascertaining the probability that the statement is true. If the probability value is small, the experimenter concludes that the difference is quite unlikely to be caused by random sampling. (S)He concludes instead that the populations are different.

Shape feature study is usually employed as a further tool to estimate the differences between the algorithms' responses. Thus, through this chapter we also discuss several issues associated to shape description and representation. Shape description is a very important issue in digital image processing and in pictorial pattern recognition. It provides descriptions of objects according to their shape, which can be used for object recognition and classification. Therefore one of the most important applications of shape description and representation is in robotic vision and medical imaging. Shape description can also provide techniques suitable for image coding that permit image transmission at low bit rates. A further application of shape analysis concerns the aspect of content-based image retrieval in digital libraries. The latter aspect is thoroughly investigated through the recent years with the aim of introducing shape features capable of providing sufficient retrieval accuracy. The shape descriptors should also be able to find noise affected shapes and defective shapes, which are tolerated by human inspection. A multitude of techniques have been developed in the recent years. A categorization of these techniques reveals that they can be classified into internal and external shape descriptors. The former ones are based on the description of the shape contour (e.g., Fourier descriptors, B-spline description). The latter ones are mainly area descriptor algorithms (e.g., quad-trees, skeletons, shape decomposition algorithms).

3.A.2.   ROC Analysis

Receiver Operating Characteristic (ROC) curves has long been used to evaluate algorithms' performance in many fields (e.g. signal processing and machine vision). The ROC curve provides information on the tradeoff between the hit rate (true positives) and the false alarm rates (false positives). In order to draw the ROC curve, both positive and negative instances are considered. The ROC curves are widely used for diagnosing as well as for judging the detection ability of different segmentation algorithms. Although theories about ROC curves have been established and computation methods and computer software are available for cross-sectional design, limited research for estimating ROC curves and their summary statistics has been done for repeated measure designs, which are useful in many applications, such as biological, medical and health services research.

3.A.2.1.   Notations

ROC analysis is commonly used to evaluate diagnostic tool and, in order to get an understanding of its use; this section describes some of the notations surrounding a diagnostic tool.

It is important to distinguish between disorder and diagnosis. A patient either has or has not a specific disorder during the period of testing. However, most medical tests are compared to the diagnosis of the disorder and the measure is how well the result of the test corresponds to the diagnosis made. The diagnosis is not necessarily the correct answer but it is what we have to work with during the testing. Usually, diagnose is made by using one or more tests. If only one test is used this test is called the gold standard. A new test

that is found to work better than the current gold standard might replace it and become the new gold standard in the future.

The first step when performing a test is to create a sample group. Different ways to create sample groups are described in the subsequent section. In the following, it is assumed that a sample group has been chosen, diagnosed and tested. Let p_i be the probability that patient i will get a positive diagnosis and q_i be patient i's probability of a positive test. The prevalence, P, of the positive diagnosis in the population is theoretically P = mean (p_i). The level of the test, Q, is Q =mean (q_i). We also define P' = 1- P and Q' = 1 -Q.

Table 1: Relations between the measurement probabilities of the outcome, prevalence, and level of a test defined in the text.

Diagnosis	Test Results		
	Positive	Negative	
Positive	TP	FN	P
Negative	FP	TN	P′
	Q	Q′	1

In general, four possible decisions and two types of errors are made when comparing a test result with a diagnosis as shown in Table 1. If both diagnosis and test are positive, it is called a true positive. The probability of a (TP) to occur is estimated by counting the true positives in the sample and divide by the sample size. If the diagnosis is positive and the test is negative it is called a false negative (FN). False positive (FP) and true negative (TN) are defined similarly. The values described are used to calculate different measurements of the quality of the test. The first one is sensitivity, SE, which is the probability of having a positive test among the patients who have a positive diagnosis.

$$SE = TP/(TP + FN) = TP/P \tag{1}$$

Specificity, SP, is the probability of having a negative test among the patients who have a negative diagnosis.

$$SP = TN/(FP + TN) = TN/P^{'} \tag{2}$$

Efficiency is defined as $EFF = TP + TN$. All three measurements will be used frequently in this report. Two other measurements that can be used are the predictive value of a positive test, $PVP = TP/(TP + FP) = TP/Q$ and the predictive value of a negative test, $PVN = TN/(TN + FN) = TN/Q^{'}$.

3.A.2.2. Sampling

To be able to perform a diagnostic test, a sample of the population must be collected to perform the test on. There are essentially three different ways to create a sample:

Naturalistic sampling

> The sampling is done by gathering a number N_0 of people that are representative of the population in interest. Each patient in the sample are both diagnosed and tested. This is a natural way to do a sampling but difficult to perform in practice.

Retrospective sampling

If the test is costly and/or risky, one might not want to test every person in the sample. In retrospective sampling, N_0 people are drawn from the population and all are diagnosed. This sample is called the screening sample. From this sample, a random sample of N_1 people with positive diagnosis and a random sample of N_2 people with negative diagnosis are drawn. Then only these N_1+N_2 are tested.

Prospective sampling

Prospective sampling is done when a number of patients N_0 is drawn from the population and tested. Then a random sample of N_1 patients with a positive test is selected together with a random sample of N_2 patients with a negative test. Only these N_1+N_2 patients are diagnosed.

From a computer science point of view, a data file is received containing numerical values representing symptoms for a specific disease together with the diagnose. Tests are performed on the data, i.e. apply an algorithm on the data to receive an output that symbolizes the diagnosis. Mostly, all cases are tested and therefore naturalistic sampling is used (if the clinicians have created the data set in a representative way). It can also be viewed as retrospective sampling is used. However, in that case the value of N_0 is unknown.

3.A.2.2.1. Estimates and standard errors for naturalistic sampling

The estimations of the prevalence, $\hat{P}$, the level, $\hat{Q}$ together with the estimates for TP, FN, FP, TN, and EFF are all unbiased and they all follow the same pattern:

$$\mathrm{SERR}\left(\hat{X}\right)=\sqrt{\left(\hat{X}\left(1-\hat{X}\right)\right)/N_o} \tag{3}$$

Where X is one of the measures mentioned earlier. This standard error can be estimated by replacing X with $\hat{X}$ in the formula.

Sensitivity and specificity becomes a bit harder. The estimates are

$$SE = T\hat{P}/\hat{P} \quad \text{And} \quad SP = T\hat{N}/\hat{P}' \tag{4}$$

But these estimates are not unbiased. The bias depends on the size of the sample N_0 as well as on P. This bias gets smaller as N_0 gets larger and if it is large enough compared to P the problem can be avoided. In the case of a low-risk population, N_0 must be larger than in a higher-risk population. If N_0 is large enough, the standard errors can be estimated in the following way:

$$
\begin{aligned}
SERR\left(S\hat{E}\right) &= \sqrt{SE(1-SE)/(N_0 P)} \\
SERR\left(S\hat{P}\right) &= \sqrt{SP(1-SP)/(N_0(1-P))}
\end{aligned}
\tag{5}
$$

3.A.2.3. The ROC curve

A decision support system gives output in the interval [0; 1] where 0 denotes a negative and 1 denotes a positive diagnosis. By introducing a cut-off somewhere in the interval the output is binarised and compared to the true diagnosis. Each cut-off corresponds to a point on a ROC curve. The ROC curve has the sensitivity plotted vertically and the reversed scale of the specificity on the horizontal axis. The scale of the

horizontal axis is also called the false positive rate[2]. The sensitivity and specificity, and therefore the performance of the system, vary with the cut-off. If several tests are performed on the same sample, ROC curves can be used to compare their performance. Another way to use ROC curves is to see the performance of a decision support system. The correspondence between cut-off and performance can be shown by an example.
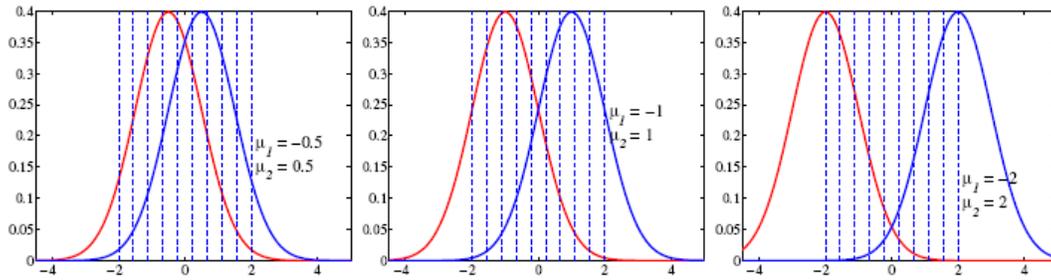

Figure 3.A-1: Three pairs of distributions with large, medium and small overlap respectively.

Assume that the result of the decision support system forms two normal distributions, one for the healthy patients and one for the ill patients. The threshold is placed at different positions to divide the distributions. The sensitivity and specificity is calculated for each position and the resulting points are plotted as a ROC curve. The curve gives a picture of the performance of the system. In Figure 1, three examples with ten different decision thresholds are shown. The overlap between the distributions is largest in the first example and smallest in the third. When plotting the sensitivity versus (1-specificity) values for these cut-offs we get the curves shown in Figure 2.
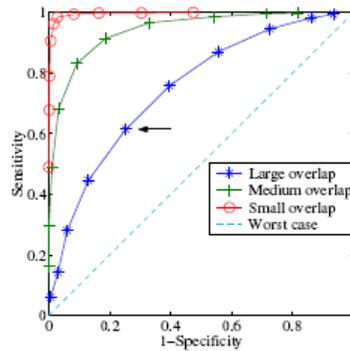
Figure 3.A-2: The corresponding ROC-curves for three pairs of distributions with large, medium, and small overlap respectively.

3.A.2.3.1. Points on the ROC curve

Each point on the curve corresponds to a specific pair of sensitivity and specificity and the complete curve gives an overview of the overall performance of a test. When comparing ROC-curves of different tests, good curves lie closer to the top left corner and the worst case is a diagonal line (shown as a dashed line in

---

[2] *It is called the false positive rate since* $1\text{-}SP = 1 - \dfrac{TN}{P'} = \dfrac{FP + TN\_TN}{FP + TN} = \dfrac{FP}{FP + TN}$

Figure 2). There are methods for estimating confidence intervals for ROC curves as well. When concentrating on a specific point on a ROC curve it might be interesting to find the confidence region for the sensitivity and specificity. The standard error is calculated by Equation (5). However, as said earlier, it is important to have large enough sample. The confidence region, CR, is then given by $CR = X \pm t_{a(2),n-1} SERR(X)$ where X is the estimated sensitivity (or specificity). If the number of objects, P or P' is small it is better to use a binomial distribution instead. Sometimes, there are two curves present and then it is interesting to compare sensitivity (or specificity) at a given point on the ROC curves. If the data used to construct the two curves are the same we have paired samples and only the samples that disagree with each other are used. The test statistic is given by

$$\chi^2 = \frac{(|t - f| - 1)^2}{t + f},$$

Where t is the number of samples correctly classified by the first decision support system but incorrectly by the second. The sample correctly classified by the second system but not by the first is denoted f. The statistic is corrected for small sample sizes and should be compared to the χ^2-table with one degree of freedom.

3.A.2.3.2.   The Area under the ROC curve

The total area under the ROC-curve is a measure of the performance of the diagnostic test since it reflects the test performance at all possible cut-off levels. The area lies in the interval [0.5, 1] and the larger area, the better performance. Assume that a high value from the method indicates that diagnosis is positive and a low value indicates that diagnosis is negative. The area is then a measurement of the probability that the distribution of the positive diagnosis is statistically larger than the distribution of the negative diagnosis. Many articles discuss the area and how to calculate and interpret it. In experiments, there is usually only a finite set of points on the ROC-curve. Therefore it is only possible to find an approximation of the area under the curve. There are several ways to calculate the area under a ROC curve. First, the trapezoidal rule can be used but gives an underestimation of the area. Second, it is possible to get a better approximation of the curve by fitting the data to a bi-normal model with maximum-likelihood estimates. After that it is possible to get a better estimate of the area. A third way to calculate the area is to use the Mann-Whitney U statistic (also known as the non-parametric Wilcoxon statistic).

In comparing two areas there is often a need to compare different methods applied on the same data set and compare the ROC curves in order to determine which method is the best. In such cases it is important to take the correlation between the areas that is induced by the data into account. If this is done, the standard error is reduced and the power of the comparison increases. In other words, it is easier to detect differences in areas if the correlation is counted for.
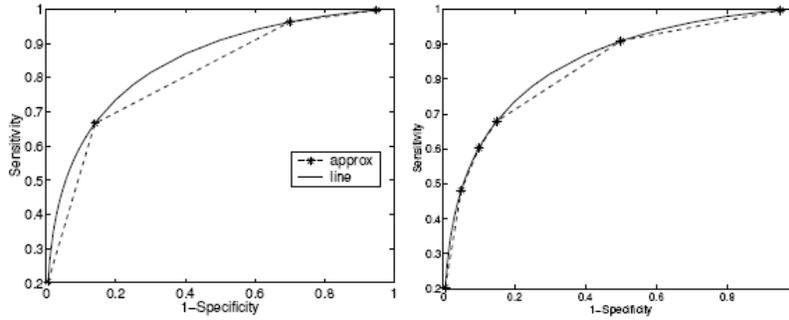
Figure 3.A-3: An Example of how few points makes an underestimate of the area under the ROC-curve.

When comparing two areas the critical ratio is defined by (6)

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}} \tag{6}$$

A_1 and A_2 are the two areas and SE_1 and SE_2 the corresponding standard errors and r is the quantity representing the correlation between the two areas due to working on the same set of data. It is important to point out that a non-significant difference between areas for two methods does not imply equivalence between the methods. In order to say that two methods are equivalent, a definition of equivalence must be made. The meaning of a non-significant difference is that there is insufficient evidence to show a difference smaller then a specific amount.

Some papers discuss the fact that when someone is comparing the full areas, equal weight is given to all false positive error rates. Often the clinician is interested only in differences with in a specified interval, e.g., the area where the specificity lies in the interval [0.8 1]. In these cases comparisons of partial areas should be done instead.

3.A.2.4.   Precision Recall Curves

Though an information theoretic approach such as mutual information can produce a useful method for ranking algorithms relative to one another, it does not produce an intuitive performance measure. Thus, attention is focused towards a richer evaluation tool: precision-recall curves. A standard evaluation technique in the information retrieval community, it has been found that the precision-recall framework is both effective and intuitive. Furthermore, in some cases it is more appropriate than the related receiver operating characteristic (ROC) curve or the Bayes risk, which is a summary statistic of the ROC curve.

A precision-recall curve is a parameterized curve that captures the trade-off between accuracy and noise as a detector's threshold varies. Precision is the fraction of detections that are true positives rather than false positives, while recall is the fraction of true positives that are detected rather than missed. In probabilistic terms, precision is the probability that the detector's signal is valid, and recall is the probability that the ground truth data was detected. Equations (7) and (8) provide mathematical definitions of precision (p) and recall (r) for convenience.

$$p = \frac{\text{TP}}{\text{TP}+\text{FP}} \qquad (7) \qquad\qquad r = \frac{\text{TP}}{\text{TP}+\text{FN}} \qquad (8)$$

These two measures are particularly meaningful for a classifier that is a detector of a sparse signal X, so that the two classes are X and $\hat{X}$. For such a detector, we are interested in how many true X were missed (recall), and how many declared X were true (precision). A downstream application may be characterized in terms of how much true signal is required to succeed and how much noise can be tolerated. Recall gives the former and precision the latter. A graphical representation of precision-recall curves is given in the diagram below presenting the case of an ideal system along with the performance curves of a system that demonstrates a less efficient performance.
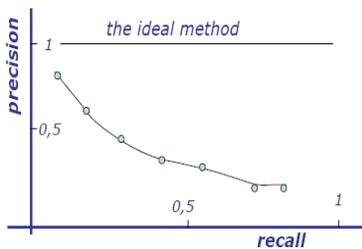

Figure 3.A-4: The precision-recall of both an ideal systems and a system demonstrating high false positive and false dismissals are presented.

A particular application can define a relative cost α between these quantities, which focuses attention at a specific point on the precision-recall curve. The F-measure, defined as:

$$F = \frac{1}{\alpha P^{-1} + (1-\alpha)R^{-1}} \qquad (9)$$

captures this trade-off as the weighted harmonic mean of P and R. The F-measure is valued between 0 and 1, where larger values are more desirable.

3.A.3.   Statistical Tests

The basic function of statistical analysis is to make judgments about the real world on the basis of incomplete information. Specifically, it is desired to determine the nature of some phenomenon based on a finite sampling of that phenomenon. The sampling procedure will produce a distribution of values, which can be characterized by various moments of that distribution. It has been proved that the distribution of a random variable is given by the binomial distribution function, which under certain limiting conditions can be represented by the normal probability density distribution function and the Poisson distribution function. In addition, certain physical phenomena will follow distribution functions that are non-normal in nature. It has been seen that the characteristics, or statistics, of the distribution functions themselves can be characterized by sampling probability density distribution functions. Generally these distribution functions are also non-normal particularly in the small sample limit.

3.A.3.1. Statistical Distribution Functions

In practice, the moments of any sampling distribution have values that depend on the sample size. If we were to repeat a finite sample having N values a large number of times, then the various moments of that sample will vary. Since sampling the same parent population generates them all, we might expect the sampling distribution of the moments to approach that of the parent population as the sample size increases. If the parent population is represented by a random variable, its moments will approach those of the normal curve and their distributions will also approach that of the normal curve. However, when the sample size N is small, the distribution functions for the mean, variance and other statistics that characterize the distribution will depart from the normal curve.

3.A.3.2. The t-Density Distribution Function

At first we consider the range of values for the mean x that we can expect from a small sampling of the parent population N. Let us define the amount that the mean x of any particular sample departs from the mean of the parent population x_p as

$$t \equiv {(x - x_p)} / {\sigma_{\bar{x}}} \tag{10}$$

Here the variable t is normalized by the best un-biased estimate of the standard deviation of the mean $\sigma_{\bar{x}}$ so as to produce a dimensionless quantity whose distribution function we can discuss without worrying about its units. Clearly the distribution function of t will depend on the sample size N. The differences from the normal curve are represented in the following figure. The function is symmetric with a mean, mode, and skewness equal to zero. However, the function is rather flatter than the normal curve so the kurtosis is greater than three, but will approach three as N increases. The specific form of the t-distribution is

$$f(t) = \frac{\Gamma\left[\tfrac{1}{2}(N+1)\right]}{\sqrt{\pi N}\, \Gamma\left(\tfrac{1}{2} N\right)} \left[1 + \frac{t^2}{N}\right]^{-(N+1)/2} \tag{11}$$

which has a variance of

$$\sigma_t^{\,2} = N/(N-2) \tag{12}$$

Generally, the differences between the t-distribution function and the normal curve are negligible for N>30, but even this difference can be reduced by using a normal curve with a variance given by equation (12) instead of unity. At the out set we should be clear about the difference between the number of samples N and the number of degrees of freedom v contained in the sample. Subsequently, we introduce the concept of "degrees of freedom" when determining the variance. The variance of both a single observation and the mean was expressed in terms of the mean itself. The determination of the mean reduced the number of independent information points represented by the data by one. Thus the factor of (N-1) represented the remaining independent pieces of information, known as the degrees of freedom, available for the statistic of

interest. The presence of the mean in the expression for the t-statistic (11) reduces the number of degrees of freedom available for t by one.
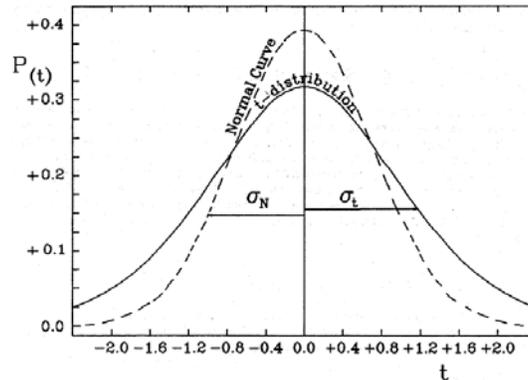

Figure 3.A-5: Shows a comparison between the normal curve and the t-distribution function for N=8. The symmetric nature of the t-distribution means that the mean, median, mode, and skewness will all be zero while the variance and kurtosis will be slightly larger than their normal counterparts. As N→∞, the t-distribution approaches the normal curve with unit variance.

3.A.3.3.   The χ^2-Density Distribution Function

In the literature, the parameter χ^2 is introduced as a measure of the mean square error of a least square fit to some data. We chose that symbol with the current use in mind. Define:

$$\chi^2 = \sum_{j=1}^{N} \left(x_j - \overline{x}_j\right)^2 \Big/ \sigma_j^2 \qquad (13)$$

Where σ_j^2 is the variance of a single observation. The quantity χ^2 is then sort of a normalized square error. Indeed, in the case where the variance of a single observation is constant for all observations we can write

$$\chi^2 = N\overline{\varepsilon^2}/\sigma^2 \qquad (14)$$

where ε^2 is the mean square error. However, the value of χ^2 will continue to grow with N so that some authors further normalize χ^2 so that

$$\chi_v^2 = \chi^2/v \qquad (15)$$
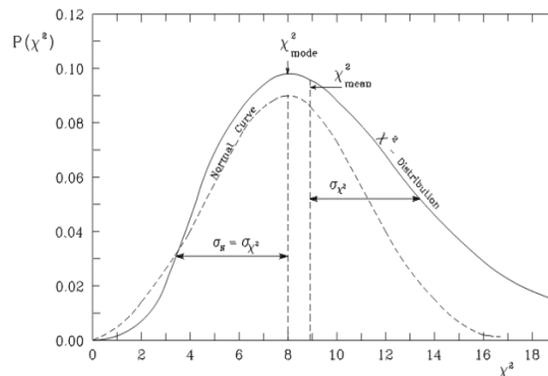
**Figure 3.A-6:** compares the χ^2- distribution with the normal curve. For N = 10 the curve is quite skewed near the origin with the mean occurring past the mode (χ^2= 8). The Normal curve has μ = 8 and σ^2 = 20. For large N, the mode of the χ^2-distribution approaches half the variance and the distribution function approaches a normal curve with the mean equal the mode.

Here the number of degrees of freedom (i.e. the sample size N reduced by the number of independent moments present in the expression) does not appear explicitly in the result. Since χ^2 is intrinsically positive, its distribution function cannot be expected to be symmetric. Figure 6 compares the probability density distribution function for χ^2, as given by

$$f(\chi^2) = \left[2^{N/2}\Gamma(1/2\,N)\right]^{-1} e^{-\chi^2/2}(\chi^2)^{1/2(N-2)} \tag{16}$$

with the normal distribution function.

The moments of the χ^2 density distribution function yield values of the variance, mode, and skewness of

$$\left.\begin{array}{l} \sigma^2_{\chi^2} = 2N \\ \chi^2_m = N-2 \\ s = \sqrt{2/N} \end{array}\right\} \tag{17}$$

As N increases, the mode increases approaching half the variance while the skewness approaches zero. Thus, this distribution function will also approach the normal curve as N becomes large.

3.A.3.4.   The F-Density Distribution Function

So far, the cases considered had their moments generated by the sampling process and are all generated from samples of the same size (i.e. the same value of N). We can ask how the sample size could affect the probability of obtaining a particular value of the variance. For example, the χ^2 distribution function describes how values of the variance will be distributed for a particular value of N. How could we expect this distribution function to change relatively if we changed N. The probability density distribution of the ratio of two variances, or more specifically F is defined to be

$$F_{12} = \left(\frac{(\chi_1^2/\nu_1)}{(\chi_2^2/\nu_2)}\right) = \left(\frac{\chi_{\nu1}^2}{\chi_{\nu2}^2}\right) \tag{18}$$

This can be shown to have the rather complicated density distribution function of the form

$$f(F) = \frac{\Gamma\left[\tfrac{1}{2}(N_1+N_2)\right]N_1^{\tfrac{1}{2}N_1}N_2^{\tfrac{1}{2}N_2}F_{12}^{\tfrac{1}{2}(N_1-1)}}{\Gamma\left(\tfrac{1}{2}N_1\right)\Gamma\left(\tfrac{1}{2}N_2\right)(N_1F+N_2)^{\tfrac{1}{2}(N_1+N_2)}} = \frac{\Gamma\left[\tfrac{1}{2}(\nu_1+\nu_2)\right]}{\Gamma\left(\tfrac{1}{2}\nu_1\right)\Gamma\left(\tfrac{1}{2}\nu_2\right)}\left[\frac{\nu_1}{\nu_2}\right]^{\nu_1/2}\frac{F_{12}^{(\nu1-1)/2}}{\left(1+F_{12}\nu_1/\nu_2\right)^{(\nu1+\nu2)/2}} \tag{19},$$

where the degrees of freedom ν_1 and ν_2 are N_1 and N_2 respectively. The shape of this density distribution function is displayed in Figure 7.

The mean, mode and variance of F-probability density distribution function are:

$$\left.\begin{array}{c} \overline{F} = N_2/(N_2 - 2) \\[4pt] F_{m0} = \dfrac{N_2(N_1 - 2)}{N_1(N_2 - 2)} \\[6pt] \sigma_F^2 = \dfrac{2(N_2 + N_1 - 2)N_2^2}{N_1(N_2 - 4)(N_2 - 2)^2} \end{array}\right\} \tag{20}$$

As it is expected, the F-statistic behaves very much like a χ^2 except that there is an additional parameter involved. However, as N_1 and N_2 both become large, the F-distribution function becomes indistinguishable from the normal curve. While N_1 and N_2 have been presented as the sample sizes for two different samplings of the parent population, they really represent the number of independent pieces of information (i.e. the number of degrees of freedom give or take some moments) entering into the determination of the variance σ_n^2 or alternately, the value of χ_n^2. The F-statistic can be used to provide the distribution of variances resulting from a change in the number of values of a_i thereby changing the number of degrees of freedom as well as a change in the sample size N.
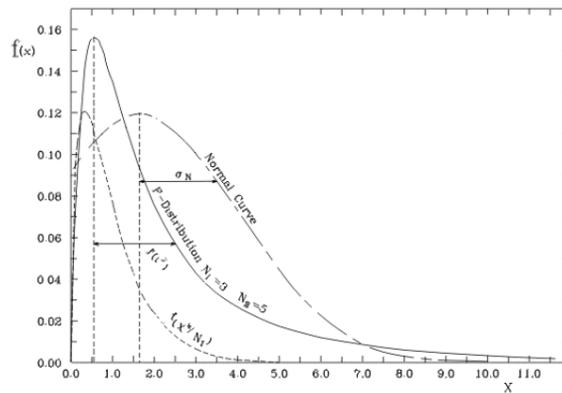

Figure 3.A-7: Shows the probability density distribution function for the F-statistic with values of $N_1 = 3$ and $N_2 = 5$ respectively. Also plotted are the limiting distribution functions $f\left(\chi^2/N_1\right)$ and $f\left(t^2\right)$. The first of these is obtained from f(F) in the limit of $N_2 \to \infty$. The second arises when $N_1 \to 1$. One can see the tail of the $f(t^2)$ distribution approaching that of f(F) as the value of the independent variable increases. Finally, the normal curve which all distributions approach for large values of N is shown with a mean equal to $\overline{F}$ and a variance equal to the variance for f(F).

Since the t, χ^2, and F density distribution functions all approach the normal distribution function as $N \to \infty$, the normal curve may be considered a special case of the three curves. What is less obvious is that the t- and χ^2 density distribution functions are special cases of the F density distribution. From the defining equations for t [equation (10)] and χ^2 [equation (13)] we see that:

$$\lim_{N \to 1} t^2 = \chi^2 \tag{21}$$

From equations (14) and (15) the limiting value of the normalized or reduced χ^2 is given by

$$\lim_{v \to \infty} \chi_v^2 = 1 \tag{22}$$

Kapsalas Petros—Master Thesis 57

So that

$$\lim_{\substack{N_1 \to N \\ N2 \leftarrow \to \infty}} F = \chi^2/N \tag{23}$$

Finally t can be related to F in the special case where

$$\lim_{\substack{N_1 \to 1 \\ N_2 \to N}} F = t^2 \tag{24}$$

Thus we see that the F probably density distribution function is the general generator for the density distribution functions for t and χ^2 and hence for the normal density distribution function itself.

3.A.3.5. Significance of the Statistical Analysis

Much of statistical analysis is concerned with determining the extent to which the properties of a sample reflect the properties of the parent population. This could be re-stated by obtaining the probability that the particular result differs from the corresponding property of the parent population by an amount ε. These probabilities may be obtained by integrating the appropriate probability density distribution function over the appropriate range. Problems formulated in this fashion constitute a statistical test. Such tests generally test hypotheses such as "this statistic does not differ from the value of the parent population". Such a hypothesis is often called null hypothesis for it postulates no difference between the sample and the value for the parent population. We test this hypothesis by ascertaining the probability that the statement is true or possibly the probability that the statement is false. Statistically, one never "proves" or "disproves" a hypothesis. One simply establishes the probability that a particular statement (usually a null hypothesis) is true or false. If a hypothesis is sustained or rejected with a certain probability p the statement is often said to be significant at a percent level corresponding to the probability multiplied by 100. That is, a particular statement could be said to be significant at the 5% level if the probability that the event described could occur by chance is .05.

3.A.3.6. The "Students" t-Test

In order to establish the extent to which a particular mean value $\overline{x}$ obtained from a sampling of N items from some parent population actually represents the mean of the parent population. To do this we must establish some tolerances that we will accept as allowing the statement that $\overline{x}$ is indeed "the same" as $\overline{x}_p$. We can do this by first deciding how often we are willing to be wrong. That is, what is the acceptable probability that the statement is false? For the sake of the argument, lets us take that value to be 5%. We can re-write equation (10) as

$$\overline{x} = \overline{x}_p \pm \sigma_x - t \tag{25}$$

And thereby establish a range δ in x given by

$$\delta = |\overline{x} - \overline{x}_p| = \sigma_x - t \tag{26}$$

Or for the 5% level as

$$\delta_{(5\%)} = \sigma_x - t_{5\%} \tag{27}$$

Now we have already established that the t-distribution depends only on the sample size N so that we may find $t_{5\%}$ by integrating that distribution function over that range of t that would allow for it to differ from the expected value with a probability of 5%.

$$0.05 = 2\int_{t5\%}^{\infty} f(t)dt = 2\left(1 - \int_{0}^{t5\%} f(t)dt\right) \tag{28}$$

The value of t will depend on N and the values of δ that result and are known as the confidence limits of the 5% level. There are numerous books that provide tables of t for different levels of confidence for various values of N (e.g. Croxton et al). For example if N is 5, then the value of t corresponding to the 5% level is 2.571.

Thus we could say that there is only a 5% chance that $\overline{x}$ differs from $\overline{x}_p$ by more than 2. 571 $\sigma_{\overline{x}}$. In the case where the number of samples increases to $\overline{x}_p$, the same confidence limits drop to 1.96 σ_x. We can obtain the latter result simply by integrating the 'tails' of the normal curve until we have enclosed 5% of the total area of the curve. Thus it is important to use the proper density distribution function when dealing with small to moderate sample sizes. There integrals set the confidence limit appropriate for the small sample sizes.

We may also use this test to examine additional hypotheses about the nature of the mean. Consider the following two hypotheses:

   a. The measured mean is greater than the mean of the parent population (i. e $\overline{x} > \overline{x}_p$),

And

   b. The measured mean is less than the mean of the parent population. (ι.e $\overline{x} < \overline{x}_p$).

While these hypotheses resemble the null hypothesis, they differ subtly. In each case the probability of meeting the hypothesis involves the frequency distribution of t on just one side of the mean. Thus the factor of two that is present in equation (28) allowing for both "tails" of the t-distribution in establishing the probability of occurrence is absent. Therefore the confidence limits at the p-percentile are set by

$$\left.\begin{array}{l} p_a = \int_{t_p}^{\infty} f(t)dt \quad = 1 - \int_{0}^{t_p} f(t)dt \\ p_b = \int_{-\infty}^{-t_p} f(t)dt = 1 - \int_{-t_p}^{0} f(t)dt \end{array}\right\} \tag{29}$$

Again one should be careful to remember that one never "proves" a hypothesis to be correct, one simply finds that it is not necessarily false. One can say that the data are consistent with the hypothesis at the p-percent level.

As the sample size becomes large and the t density distribution function approaches the normal curve, the integrals in equations (28) and (29) can be replaced with

$$p = 2\,\text{erfc}\!\left(f_p\right) = 2\!\left[1 - \text{erf}\!\left(f_p\right)\right]$$
$$p_{a,b} = \text{erfc}\!\left(\pm t_p\right) = 1 - \text{erf}\!\left(\pm t_p\right)$$

(30)

Where erf(x) is called the error function and erfc(x) is known as the complimentary error function of x respectively. The effect of sample sizes on the confidence limits, or alternately the levels of significance, when estimating the accuracy of the mean was first pointed out by W.S. Gossett who used the pseudonym "Student" when writing about it. It has been known as "Student's t-Test" ever since. There are many other uses to which the t-test may be put

3.A.3.7. Practical Application of a t-Test

To determine whether an individual measurement is a typical member of the population as a whole requires knowledge of the variation of individual measurements, i.e., the standard deviation of the population. Similarly, to determine the degree of error associated with a sample mean requires knowledge of the variability of sample means, i.e., the standard error of the mean. Hence, to determine whether there is a significant difference between the means of two samples, knowledge is required of the degree of variability of the difference between two sample means. Consider two different populations, for each sample, the mean is calculated and the difference between the two means (C*-T*) represents the degree of divergence. Imagine that the experiment is repeated many times over and that several estimates of C*-T* are obtained. The distribution of the sample means is shown in figure 8(a) while figure 8(b) demonstrates the difference (C*-T*). If the distribution of the two populations is normally distributed, then the distribution of the differences between pairs of means taken from these two populations will also be normally distributed. Hence, we can use the standard normal distribution to test whether there is a true difference between the two means in the experiment.
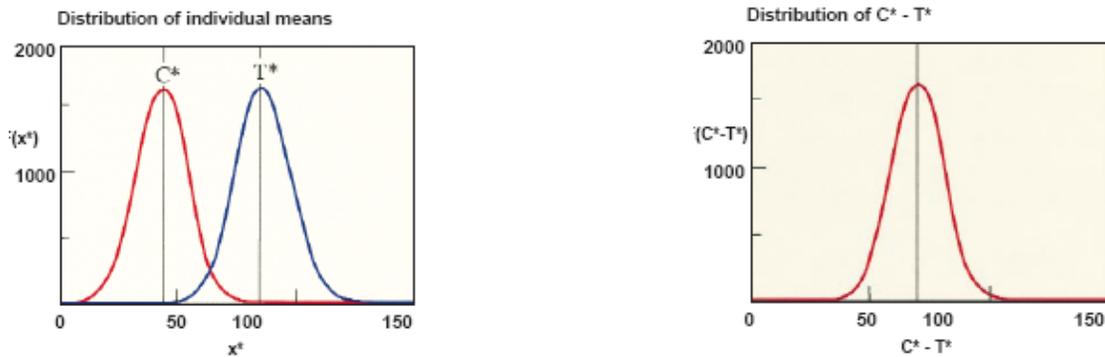
Figure 3.A-8: Distribution of the difference between two sample means.

Comparing the difference between two means "Student's t-test"

Consider an experiment designed to test the null hypothesis that no difference occurs between the studied populations. Examination of the group means reveals that they differ (C*-T*) which suggests that the two populations significantly differ. To decide between these two alternatives, we compare the treatment

---

effect (C*-T*) with the degree of variation pooled from both groups by carrying out a 't' test. The formula for the "t-test", one of the commonest procedures used in data analysis, is as follows:

$$t = C^* - T^* / s\sqrt{(1/n_1 + 1/n_2)} \tag{31}$$

Where 's' is an estimate of the standard deviation based on both samples jointly and n_1 and n_2 are the number of observations within each group. Hence, the value of 't' is the ratio of the difference between the means to the degree of variation between patients combined from both groups.

When 't' is calculated, the difference between the two means is converted so that it becomes a member of the t-distribution with the predetermined degrees of freedom. Small differences between the means, which are more likely to have arisen by chance, result in values of 't' which lie close to the mean of the distribution. Larger mean differences result in 't' values further out in the tail of the distribution. For instance, when 't' is equal to or greater than 2.78 (the value at P=0.05) for 4 degrees of freedom, the value is in the zone of the distribution which includes the 5% most extreme values. This is an unlikely value of 't' to have been obtained by chance alone and therefore, we conclude that there is a real difference between the two means. Figure 9 illustrates the t-distribution for 4 degrees of freedom as well as the locations on the distribution where t takes the values 2.78 and -2.78.
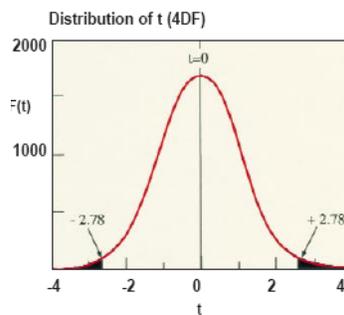
Figure 3.A-9: The 't' distribution for 4 degrees of freedom (DF).

3.A.3.8. NON-PARAMETRIC TESTS

The mode and median

An alternative approach to the analysis of non-normally distributed data is to use a distribution-free or non-parametric test (non-parametric to distinguish the tests from 'parametric' tests which are based on the normal distribution). These tests are easy to carry out and can be used regardless of the shape of the underlying distribution as long as the samples being compared can be assumed to come from a distribution of the same general shape. As discussed previously, when a distribution deviates significantly from normality, the arithmetic mean is often a poor description of its central tendency (Figure 10). Nevertheless, there are two additional statistics, which can be used to describe the central tendency of such a distribution. First, the mode is the value of the variable 'x' with the highest frequency, i.e., the maximum point of the curve. Second, the median is the middle value of 'x', i.e., if all the values of 'x' were listed in ascending or descending order, the median would be the middle value of the array. Little progress has been made in devising statistical tests,

which are based on the mode, but there are two tests which can be used to test the differences between the medians of two samples.
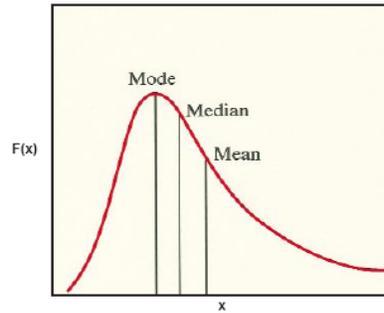

Figure 3.A-10: Mode, median and arithmetic mean of the distribution.

The test presented in this sub-section is known as the Mann-Whitney and also as the Wilcoxon test. Equivalent forms of the test are appeared in the literature under various names, probably partly because of the intuitive appeal of the test procedure. Although primarily a two-sample test, Mann-Whitney test may be applied in many different situations other than the usual two-sample situation.

An intuitive approach to the two-sample problem is to combine both samples into a single ordered sample and then assign ranks to the sample values from the smallest value to the largest, without regard to which population each value came from. Then the test-statistic might be the sum of the ranks assigned to those values from one of the populations. If the sum is too small (or too large), there is some indication that the value from that population tend to be smaller (or larger, as the case may be) than the values from the other population. Hence the null hypothesis of no differences between populations may be rejected if the ranks associated with one sample tend to be larger than those of the other sample.

Ranks may be considered preferable to the actual the actual data for several reasons. First, if the numbers assigned to the observations have no meaning by themselves but attain meaning only in an ordinal comparison with the other observations, the numbers contain no more information than the ranks contain. Such is the nature of ordinal data. Second, even if the numbers have meaning but the distribution function is not a normal distribution function; the probability theory is usually beyond our reach when the test statistic is based on the actual data. The probability theory of statistics based on ranks is relatively simple and does not depend on the distribution in many cases.

3.A.3.9. The Mann-Whitney Test

The data consists of two random samples. Let $X_1, X_2, \cdots, X_n$ denote the random sample of size n from population 1 and yet $Y_1, Y_2, \cdots, Y_m$ denote the random sample of size m from population 2. Assign the ranks 1 to m+n. Let $R(X_i)$ and $R(Y_j)$ denote the rank assigned to X_i and Y_j for all i and j. For convenience, let N=n+m.

If several sample values are exactly equal to each other (tied), assign to each the average of the ranks that would have been assigned to them had there been no ties. Prior to proceeding in the description of the Mann-Whitney Test the assumptions made are provided.

Assumptions

1. Both samples are random samples from their respective populations.

2. In addition to independence within each sample, there is mutual independence in the two samples.

3. The measurement scale is at least ordinal.

As it was discussed previously, the Mann-Whitney test comprises the non-parametric equivalent of the unpaired't-test'. Subsequently, through steps (1-5) we summarize the application of the Mann-Whitney U-test.

1) List observations in order of magnitude within each group. Assign ascending ranks 1, 2, 3, ... to the whole set of observations with repeated values, called 'ties', given the mean of the ranks within that run.

2) Sum the ranks of each row R_X, R_Y.

3) Calculate U_X and U_Y, e.g., $U_X = \{n_X(n_X+1)/2 + (n_X n_Y)\}\text{-}R_X$ where n_X and n_Y are the number of units in each group. A similar equation can be constructed for U_Y by substituting n_Y and R_Y.

4) Take whichever is the smaller of U_X and U_Y to the corresponding statistical tables. U has to be equal to, or LESS than, the tabulated value for significance, i.e., low values of U indicate significance.

3.A.4. Theoretical Background Associated to Shape descriptors

Shape is an important visual feature and it is one of the features used to describe image content. However, shape representation and description is quite difficult task. This is because when a 3-D real world object is projected onto the 2-D image plane, one dimension of object information is lost. As result, the shape extracted from the image only partially represents the projected object. To make the problem even more complex, shape is often corrupted with noise, defects, arbitrary distortion and occlusion. Shape representation generally looks for effective and perceptually important shape features based on either shape-boundary information or boundary plus interior content. Various features have been designed, including shape signature, signature histogram, shape invariants, moments, curvature, shape context, shape matrix, spectral features etc. These various shape features are often evaluated by how accurately they allow one to retrieve similar shapes from a designated database. However, it is not sufficient to evaluate a representation technique only by the effectiveness of the features employed. This is because the evaluation ignores the other important characteristics of a shape representation technique. For example in the new multimedia application content-based image retrieval (CBIR), efficiency is envisaged as equally important as effectiveness due to the online retrieval demand. In fact, MPEG-7 has set several principles to measure the shape descriptor, that is, good retrieval accuracy, compact features general applications, low computational complexity, robust retrieval performance and hierarchical coarse to fine representation [35]. Many shape representation and description

techniques have been developed in the past. A number of new techniques have been proposed in recent years.

a)   Classification of shape representation and description techniques

Shape representation and description techniques can be generally classified into two classes of methods: contour-based methods and region-based methods. The classification is based on whether shape features are extracted from the contour only or are extracted from the whole shape region. Under each class, the different methods are further divided into structural approaches and global approaches. This sub-class is based on whether the shape is represented as a whole or represented by segments/sections (primitives). These approaches can be further distinguished into space domain and transform domain, based on whether the shape features are derived from the spatial domain or from the transformed domain. The whole hierarchy of the classification is shown in Fig. 11. Subsequently, some of these techniques are discussed in details.

b)   Contour-based shape representation and description techniques

Contour shape techniques only exploit shape boundary information. There are generally two types of very different approaches for contour shape modeling: continuous approach (global) and discrete approach (structural). Continuous approaches do not divide shape into sub-parts; usually a feature vector derived from the integral boundary is used to describe the shape. The measure of shape similarity is usually a metric distance between the acquired feature vectors. Discrete approaches break the shape boundary into segments, called primitives using a particular criterion. The final representation is usually a string or a graph (or tree), the similarity measure is done by string matching or graph matching. In the following we discuss these two types of approaches.
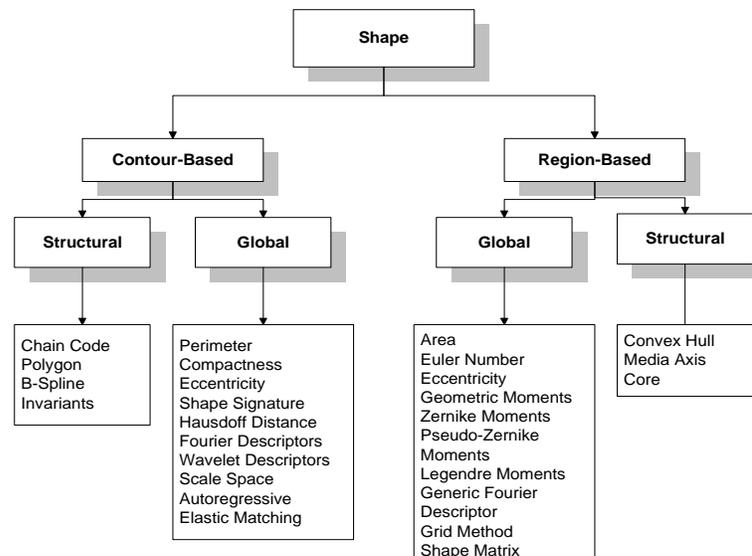

Figure 3.A-11: Classification of shape representation and description techniques.

3.A.4.1. Global methods

Global contour shape techniques take the whole shape contour as the shape representation. The matching between shapes can either be in space domain or in feature domain. For shape description, there is always a trade-off between accuracy and efficiency. On the one hand, shape should be described as accurately as possible; on the other hand, a shape description should be as compact as possible to simplify indexing and retrieval. Efficient offline feature extraction is also desirable. Simple global shape descriptors are compact; however, they are very inaccurate shape descriptors. They need to be combined with other shape descriptors to create practical shape descriptors. Correspondence based shape matching and signature based matching are not suitable for online shape matching, because they all involve the 2-D matching of two shapes. However, if partial matching is a requirement, methods based on Hausdorff distance can be a choice. Elastic matching and wavelet methods are complex to implement and match. Autoregressive (AR) methods involve matrix operations which are expensive and it is difficult to associate AR descriptors with any physical meaning. The implementation and matching of CSS is complex. However, the perceptually meaningful and compact features are appealing for shape description and online retrieval. Fourier descriptor is simple to implement, and involves less computation by either using fast Fourier transform (FFT) or using truncated Fourier transform computation. The resulting descriptor is also compact and the matching is very simple. Compared with CSS, FD is simpler to compute and more robust. Boundary moment descriptor is similar to Fourier descriptor, and is easy to acquire. However, unlike Fourier descriptor, only the few lower order moment descriptors have physical interpretation.

3.A.4.2. Structural methods

Another member in the shape analysis family is the structural shape representation. With the structural approach, shapes are broken down into boundary segments called primitives. Structural methods differ in the selection of primitives and the organization of the primitives for shape representation. Common methods of boundary decomposition are based on polygonal approximation, curvature decomposition and curve fitting. The result is encoded into a string of the general form: $S = s_1, s_2, \cdots s_n$; where s_i may be an element of a chain code, a side of a polygon, a quadratic arc, a spline, etc. s_i may contain a number of attributes like length, average curvature, maximal curvature, bending energy, orientation etc. The string can be directly used for description or can be used as input to a higher level syntactic analyzer. In the following we describe methods of shape representation and description using S.

3.A.4.3. Simple shape descriptors

Common simple global descriptors are area, circularity (perimeter^2/area), eccentricity (|length of major axis- length of major |/length of major axis), major axis orientation, and bending energy. These simple global

descriptors usually can only discriminate shapes with large differences; therefore, they are usually used as filters to eliminate false hits or combined with other shape descriptors to discriminate shapes. They are not suitable to be standalone shape descriptors. Other simple global contour shape descriptors have been proposed by [36]. These descriptors include convexity, ratio of principle axis, circular variance and elliptic variance.

Boundary moments are also used to reduce the dimensions of the boundary representation. Assuming the shape boundary has been represented as a shape feature signature $z(i)$, the rth moment m_r and central moment μ_r can be estimated as:

$$m_r = \frac{1}{N} \sum_{i=1}^{N} [z(i)]^r \tag{32}$$

$$\mu_r = \frac{1}{N} \sum_{i=1}^{N} [z(i) - m_1]^r \tag{33}$$

Where N is the number of boundary points. The normalized moments $\overline{m}_r = m_r / (\mu_2)^{3/2}$ and $\mu_r = \mu_r / (\mu_2)^{3/2}$ are invariant to shape translation, rotation and scaling. Less noise-sensitive shape descriptors can be obtained from $F_1 = (\mu_2)^{1/2} / m_1$, $F_3 = \mu_4 / (\mu_2)^2$. The method in [36] treats the amplitude of the shape signature function $z(i)$ as a random variable v and creates a histogram $p(v_i)$ from $z(i)$. Then, the rth moment is obtained by

$$\mu_r = \sum_{i=1}^{K} (v_i - m)^r p(v_i) \quad \text{and} \quad m = \sum_{i=1}^{K} v_1 p(v_i) \tag{34}$$

The advantage of boundary moments is that it is easy to implement. However, it is difficult to associate higher order moments with physical interpretation.

Spectral descriptors overcome the problem of noise sensitivity and boundary variations by analyzing shape in spectral domain. Spectral descriptors include Fourier descriptor (FD) and wavelet descriptor (WD), they are derived from spectral transforms on 1-D shape signatures. One of the most widely used shape description methods is FD [37, 38, 39]. Conventional FD methods only deal with closed curve, however, Lin et al. and Mitchell et al. used FD to describe partial shapes [40, 41]. Granlund introduced the Fourier invariants, which describe the rotational symmetry of shapes [37, 38]. Rauber proposed a UNL FD (named after Universidade Nova de Lisboa, Portugal), which is able to describe, disjointed or articulated contour shape [42]. The UNL FD is acquired by applying 2-D Fourier transform on the UNL transformed shape image. Even though a feature selection process is followed, the dimension of the feature vector acquired this way is very high. Richard and Hemami introduced a complex distance measurement, called the true distance measurement, for measuring the similarity between two set of FDs [43]. Since the true distance measurement requires two Fourier transforms for each matching, it involves 15 times more computation than a normal distance measurement. Rui et al. [44] proposed a distance measurement to classify similarity transformed characters using Fourier transformed coefficients. This distance measurement is the weighting sum of the variance of magnitude ratios and the variance of phase difference between two sets of Fourier coefficients. The Fourier coefficients are derived from Fourier reconstructed shape boundary rather than from original

boundary. This is not different from FD derived from a smoothed boundary. Eichmann et al. proposed the use of a short-time Fourier descriptor (SFD) for shape description [39], however, Zhang and Lu have found that SFD is outperformed by conventional FD methods in shape retrieval [45]. This is because SFD cannot capture global shape features although it can capture local shape features more accurately. Recently, several researchers have proposed the use of WD for shape description [46-48]. Although WD has the advantage over FD in that it is of multi-resolution in both spatial space and spectral space, the increase of spatial resolution will certainly sacrifice frequency resolution. For example, in [47], only wavelet coefficients of the few low frequencies are used to represent shape. Most importantly, the complicated matching scheme of wavelet representation makes it impractical for online shape retrieval. In [48] the similarity measurement algorithm needs $2^L \times N$ all-level shift matching, where L is the number of levels of resolution of the wavelet transform and N is the number of normalized boundary points. In [47] the number of matching for similarity measurement is not only large but also dependent on the complexity of the shape, since the similarity measurement is the all level shift matching of all the zero-crossing points of the wavelet approximation of the shape. Apart from the matching complexity, the dyadic wavelets used can rarely associate the feature segments on the shape boundary. Therefore, WD suffers the same drawback in primitive determination as that in the structural approach. FD is backed by the well-developed and well-understood Fourier theory. The advantages of FD over many other shape descriptors are (i) simple to compute; (ii) each descriptor has specific physical meaning; (iii) simple to do normalization, making shape matching a simple task; (iv) captures both global and local features. With sufficient features for selection, FD overcomes the weak discrimination ability of those simple global descriptors. FD also overcomes the noise sensitivity and difficult normalization in the shape signature representations. Most FD based works are dedicated to character recognition and object classification. The complex coordinates and the cumulative angle function are dominantly used in these works to derive FD. However, Zhang and Lu have found that for general shapes, the centroid distance function is the most desirable shape signature to derive FD. They have also found that 10 FD features are sufficient to represent shape; this is a significant reduction in dimensions of FD compared with 60 FD features usually used in shape representation. Their results show that FD outperforms CSS method in terms of retrieval performance and robustness.

3.A.4.4. Chain code representation

Chain code describes an object by a sequence of unit-size line segments with a given orientation. The method was introduced in 1961 by Freeman [49] who described a method permitting the encoding of arbitrary geometric configurations. In this approach, an arbitrary curve is represented by a sequence of small vectors of unit length and a limited set of possible directions, thus termed the unit-vector method. In the implementation, a digital boundary of an image is superimposed with a grid, the boundary points are approximated to the nearest grid point, then a sampled image is obtained. From a selected starting point, a

chain code can be generated by using 4-directional or 8-directional chain code. N-directional chain code is also possible, it is called general chain code [50]. If the chain code is used for matching it must be independent of the choice of the first boundary pixel in the sequence. One possibility for normalizing the chain code is to find the pixel in the border sequence which results in the minimum integer number if the description chain is interpreted as a base four number that pixel is then used as the starting pixel. Alternatively, the boundary can be represented by the differences in the successive directions in the chain code instead of representing the boundary by relative directions. This can be computed by subtracting each element of the chain code from the previous one and taking the result modulo n, where n is the connectivity.

After these operations, a rotationally invariant chain code is obtained by a cyclic permutation, which produces the smallest number. Such a normalized differential chain code is called the shape number. Chain code derived in this way is not scale invariant. Although it is possible to scale two similar shapes into the same size, the resulted shape numbers can have a different number of digits, making it impractical to do matching between two shapes. The chain code usually has high dimensions and is sensitive to noise. It is often used as an input to a higher-level analysis. For example, it can be used for polygon approximation and for finding boundary curvature, which is a important perceptual feature. Iivarinen and Visa derive a chain code histogram (CCH) for object recognition [51]. The CCH is computed as $p(k)=n_k/n$, where n_k is the number of chain code values k in a chain code and n is the number of links in a chain code. The CCH reflects the probabilities of different directions present in a contour. The CHH is translation and scale invariant; however, it is only invariant to a rotation of 90°. Therefore, the normalized CHH (NCHH) is proposed. It is defined as $p(k)=l_k n_k /l$, where n_k is the same as in CHH, l_k is the length of the direction k and l is the length of the contour. Although CHH reduces the dimensions of chain code representation, it does not solve the noise sensitivity problem.

Methods of analyzing the shape features in order to perform diagnosis and classification are currently in considerable use in the aspect of medical imaging as well as to numerous industrial applications. Yet, despite the apparent advantages of this approach, it is not used to analyze corrosion processes. The main reason of this lies in the problem of describing the corrosion damage by morphology, i.e color, texture and shape.

3.A.5.   Corrosion Damage Classification According to Shape and Texture Features

An attempt to classify corrosion damage effects according to their texture and shape features was conducted by Choi and Kim in [25]. The issue of classifying the corrosion defects according to their shape, texture and color typology starts by determining a vector of 18 features (5 color, 2 texture, 11 morphological shape attributes). Prior to the analysis the vector of the attributes of all the objects in the test sample is integrated and normalized. To check if the parameters meet the criterion of non-homogeneity ($\sigma^2>0$),

integrity parameters are analyzed. If this condition is satisfied, object attributes description parameters can be treated as the coordinates of some n-dimensional space. Resting on this approach, the object under analysis can be represented by a point on the n-dimensional space. The closer the distance between the attributes (point) of an object on the space are, the more is the similarity of the objects. On the same token, the farther the distances between the attributes are, the less alike they are in terms of the attributes. In accordance with this, clusters are formed on the space of attributes. The task of classification is to determine the ranges of the clusters corresponding to various objects and to set the rules being the functions of the attributes, which are used to divide the objects into classes. The clustering of corrosion patterns was performed in the space of attributes and subsequently, in order to evaluate the functional relationship between the objects co-ordinates on the classification plane and their co-ordinates (attributes) in the original multidimensional space, a linear regression model has been selected. Through the latter procedure, the lines dividing the classification plane into semi-spaces are derived. The results obtained through the clusterization approach showed that the corrosion defects can be classified according to their typology with a significant accuracy.

3.B.   Implementation
3.B.1.   Introduction

Several measures are employed for validating the algorithmic results. In this chapter, we discuss an automated framework for objective performance evaluation of region segmentation algorithms concerning images depicting decay effects, monitored by various imaging modalities. The framework includes image data sets of degraded stone surfaces (screened by FOM and Digital Camera), development of image segmentation algorithms, source code for algorithms extracting ground truths, inspection of the derived ground truth matrixes by experts, a tool for scoring of performance metrics, a tool for selecting algorithm parameters and a tool for comparing algorithms' performance. This framework guarantees reliable and objective estimation of segmentation algorithms' performance while it allows informed experimental feedback for the design of improved segmentation schemes.

As it was discussed in the previous chapter, the results derived from the application of the implemented algorithmic schemes, indicate some differences in the segmentation results. More specifically, it is observed that some of them tend to split the detected areas into adjacent small in size spots (the sub-band decomposition algorithm as well as the labeling algorithms tend to operate in this way). Others succeed in providing reliable information concerning the topology of decay patterns, while distorting their extent and shape. The objective of the performance study is to assess the potential and the limitations of the recruited algorithmic schemes in segmenting degradation patterns, while exploiting individual features associated with the robust points and the drawbacks of each approach.

Further to exploiting the robust points of each segmentation procedure and validating their performance we have also implemented automated processes of performing tests of statistical significance (T-Tests & Mann Whitney U-test). These tests study the way that the cleaning state and/or the exposure of the stone material is associated with individual features of the segmented decay areas. Finally, in the last subsection of this chapter (3.B.5) we discuss processes related to decay patterns boundary extraction and shape feature analysis.

3.B.2.   Ground Truth Matrix Extraction

The determination of a Ground Truth Matrix of decay regions is critical as it provides a test bed for measuring the algorithms' performance and comprehending the differences between them in the segmentation procedure. In this work we introduce a semi-automated approach of extracting the Ground Truth Matrix. We first extract the decay areas consistently detected by the examined algorithmic procedures. This process is considered in this section. The entire Ground Truth Matrix extraction process is supervised by the experts as explained in 4.1.

Through the Ground Truth Extraction Approach, we check in pairs the areas segmented by all algorithms. The procedure starts by labeling the segments detected by each algorithm. For a pair of

segmented and labeled images, figure 1 illustrates the processes of managing the non-overlapping and partially-overlapping labels. The various steps are presented in the following subsections
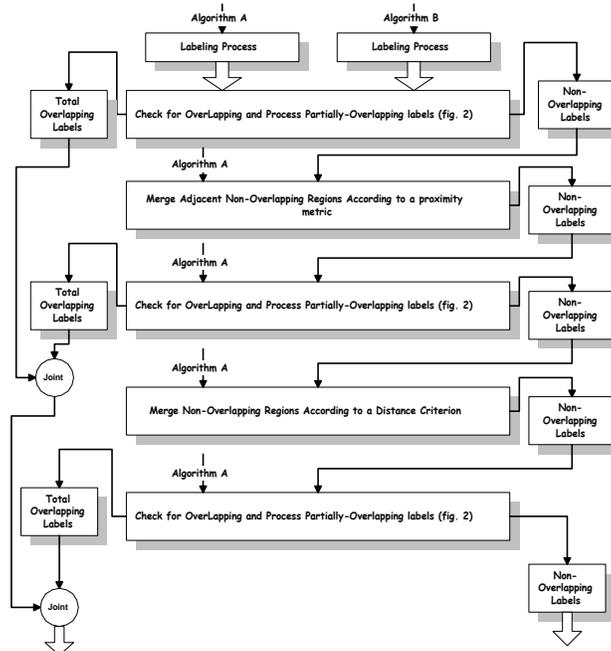

Figure 1: Flowchart illustrating the overlap extraction procedure.

This scheme is applied on the segmented images by each and every algorithm in an incremental way, as to extract the Ground Truth Matrix of decay areas. In fact, at the ith step of the process (i>1) the input images involve the result of the algorithm A_i and the image representing the non-overlapping labels obtained from the $(i-1)^{th}$ step of this process.

In subsections 3.B.2.1 and 3.B.2.2 we discuss some introductory to the Check for Overlapping Issues. More specifically, we at first (3.B.2.1) discuss some general principles related to region segmentation and label assignment processes and subsequently (&3.B.2.2) we analyze the label assignment more extensively (as it was implemented in this thesis).

3.B.2.1.   Principles Related to Region Segmentation and Labeling

Prior to the analysis of the checking for overlap procedure, a stage of labeling the segmented decay patterns is carried out. The specific definition of label assignment that we use is the same as used by Hoover et. al. [52] and is repeated here for reference. The labeling of R pixels representing decay areas, into labels r_1, r_2, …, r_n is defined by the following properties.

- Every region is spatially connected. Our implementation currently uses eight-connectedness as the definition of spatially connected.

- $\forall r_i, r_j \in R$ *if* $i \neq j$, $r_i \bigcup r_j = \emptyset$. Regions do not overlap each other.

3.B.2.2. Analysis of the Labeling Approach

Initially, it should be said that the structure, used to keep information for the determined labels is a stack whose elements represent a discrete label. The structure used to store information for each specific label is a list and in turn, each of the list's components represents a pixel's co-ordinates. The image is traversed from left to right until the first black pixel is encountered. The specific pixel is then considered to belong to the 1st label, the structure used for storing information is constructed and the pixels co-ordinates are stored. Subsequently, its 8-neighbourhood is examined. If a black pixel is met then it is considered to belong to the same label. While traversing the image, if we encounter a black pixel we check all the labels constructed so far and whether one of the pixel's 8-neighbors has been labeled, then the current pixel is considered to belong to the same label (as its neighbor). Otherwise a new label is constructed and the pixel's locations are stored. The condition stating that one pixel cannot be present in 2 distinct labels is valid in our approach. The process is continued until the whole image is checked.

3.B.2.3. Check for Overlapping Labels

The Check for Overlapping Approach aims at extracting overlapping regions detected by two different algorithms. The process initially checks whether a spot detected by algorithm A_i is also detected by algorithm A_j. The steps of the checking for overlap approach are summarized below:

1) Traverse the stack containing labels detected by Ai. For each label we examine whether A_j also detected a region at the same locations.

2) If **Yes** then check whether they **Totally** or **Partially overlap**;

   a) If the label detected by A_j totally overlaps the corresponding label segmented by A_i, assign the label to the cluster containing the **Totally Overlapping Labels**.

   b) Otherwise assign the label to the cluster containing the **Partially Overlapping Labels**.

3) If No (A_j did not segment an area at the corresponding locations), assign the label detected by A_i to the cluster containing the **Non-Overlapping Labels**

4) Repeat steps 1-3 until the end of the stack is found

Once the above procedure is terminated we have obtained three clusters of labels (***Totally Overlapping, Partially Overlapping and Non-Overlapping***).

At this point we should make clear that the aim of the Ground Truth Extraction is to determine compact areas that correspond to susceptible degraded regions. Totally overlapping labels are included in the Ground Truth. Visual inspection reveals that the partially overlapping labels often correspond to larger in extent regions that became split. In an attempt to segment the degraded regions as compact areas that represent decay patterns at their actual size we further process the partially overlapping patterns of A_i to attain total overlap to the labels of A_j.

3.B.2.4. Processing the Partially Overlapping Labels

Through this procedure we consider the partially overlapping labels of A_i that are obtained by the above process, in combination with the areas segmented by the algorithm A_j. Initially, the partially overlapping labels of algorithm A_i, are blown via a conditional thickening operator up to the point that may cover the entire corresponding label segmented by A_j. The operator of thickening label A_i towards the pair of structuring elements E_1, E_2 is defined as follows:

$$(E_1, E_2) \otimes A_i A_j = A_i \cap \left(A_j \cup \left((E_1 \ominus A_j) \cap (E_2 \ominus A_j^C) \right) \right) \tag{1}$$

The pair of structuring elements E_{i1} and E_{i2} controls the direction of expansion. To cover spatial expansion in many directions, we use 8 pairs of such elements for either black or white spots. The first two pairs are given as:

$$(E_{11}, E_{12}) = \left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \right) \text{ and } (E_{21}, E_{22}) = \left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \right)$$

The remaining pairs are obtained from these matrix combinations through rotations every 90°. Finally, the conditional thickening operator is obtained as a combination of individual results for every pair (E_{i1}, E_{i2}):

$$Area = \bigcup_{i=1}^{8} (E_{i1} E_{i2}) \otimes A_i A_j \tag{2}$$

The equation is applied for Area=A until Area=B or Area does not change any more. The special property of E is that A increases until the boundaries of B are reached. The segmented areas derived after the processing of the partially overlapping spots are labelled and assigned to the cluster containing the total overlapping labels. Figure 2 illustrates the algorithmic procedure described above.
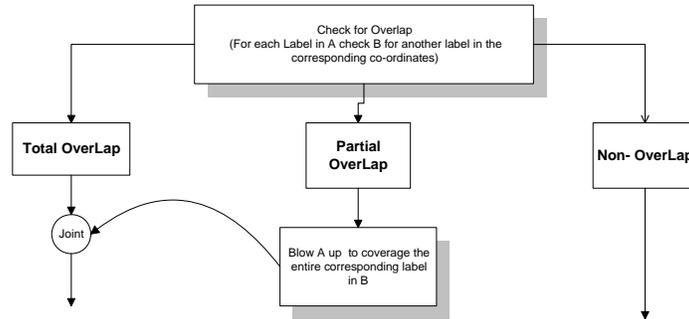
Figure 2: Check for overlap spots and processing the partially overlapping labels.

The next step of the Ground Truth extraction approach involves processing the Non-Overlapping Labels. The Ground Truth Matrix involves all labels segmented by all or just some algorithms, recognizing the potential of algorithmic failure in spot detection. Thus, even non-overlapping patterns between two algorithms mat actually be parts of the Ground Truth of the problem.

3.B.2.5.   Processing the Non-Overlapping Labels

To determine the steps followed by the Ground Truth Extraction Approach we visually and statistically inspect the non-overlapping spots. The assessments drawn by this inspection reveal that the non-overlapping labels usually correspond to small in size areas arranged in a very close distance. Furthermore, for the majority of the non-overlapping regions segmented by A_i (Regions which were segmented by A_i and not by A_j), it can be observed that A_j also detected spots at the area around, but not at exactly the same locations. The predefined spots are usually small in extent and their presence is associated to the potential of the algorithmic schemes in effectively discriminating small objects in inhomogeneous backgrounds or in backgrounds of high noise levels. According to the experts' judgment, these small patterns are associated to areas large in extent and they should be considered as an entity. In order to overcome these instances of over-segmentation; a process of merging adjacent non-overlapping spots is developed. Initially, we process adjacent non-overlapping regions.

Merging Adjacent Non-Overlapping Areas

According to the procedure, all the non-overlapping areas are labeled and the centroid of each label is calculated. Subsequently, we measure the Euclidean distance between a label's centroid to the centroids of its neighboring labels and if the distance is lower than a predefined threshold $T_{Distance}$, the adjacent areas are merged by filling the inter-spot distance. In the current implementation, the distance threshold $T_{Distance}$ was chosen to be equal to 8 pixels. This value was selected to reflect the mean diameter value of the total overlapping labels. Thus it is considered that labels with such distance could belong to the same pattern. After the merging of neighboring patterns, the derived areas are labeled again and the procedure illustrated in Fig. 2 takes place. The total-overlapping labels derived through the latter step are also assigned to the cluster containing the total overlapping labels. Figure 4 presents a diagram of the Merging Approach.
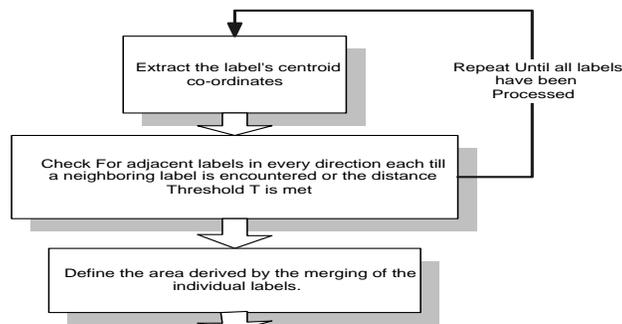
Figure 3: Flowchart depicting the process of merging adjacent labels.

The size and spatial arrangement of the derived non-overlapping labels are inspected again to determine the next steps of the approach.

Merging According to a Distance Criterion

The non-overlapping labels provided by the previous step are labeled and the centroid of each label is extracted. A window of size 31x31 is applied both at the centroid of the label in image A_i and at the

---

corresponding co-ordinates in image A$_j$. The label in the window defined in A$_i$ is submitted to morphological erosion by a structuring element (disk). The morphological erosion is iterated by increasing the radius by 1 at each iteration and terminates when either the label in window A$_i$ overlaps an existent label in window A$_j$ or the radius value reaches an upper bound. This procedure is repeated for each of the segmented areas and the radius values at which the morphological operation terminates are stored to calculate the median erosion value. Subsequently, morphological erosion is performed on all original areas of Ai with a disk-structuring element of the size of the median. Finally, the process in figure 2 is applied again, to derive the new overlapping labels.
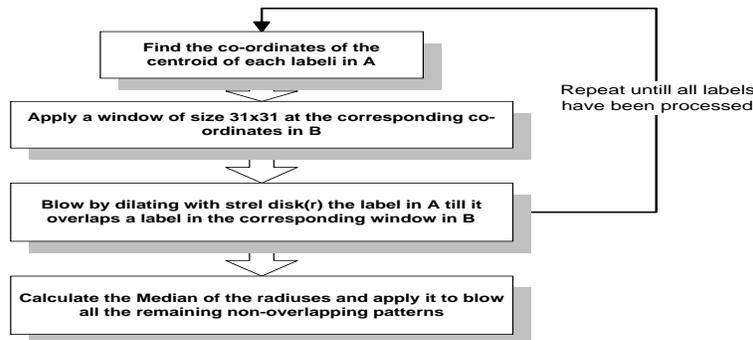
Figure 4: Flowchart of the procedure employed to calculate the trans-image distance.

3.B.2.6. Ground Truth Matrix Formation

As it was discussed previously, the extraction of the Ground Truth Matrix involves fusing of the areas segmented by all the algorithms. The fusion takes place by examining the implemented algorithms in pairs in order to extract the overlapping segments. The non-overlapping segments obtained at each stage are subsequently checked towards the results of a consecutive algorithm. The process proceeds until all methods have been examined.

Figure 5 shows that the Ground Truth stems from the union of the Non-Overlapping (obtained when the process terminates) and the total overlapping labels, derived at each step of the extraction approach. Through a brief visual inspection of the segmented degraded regions, it can be verified that the total overlapping patterns, correspond to areas larger in extent than the non-overlapping. However, the experts considered that the non-overlapping areas should also be present in the Ground Truth as these spots correspond to regions that are likely to represent decay effects.

Figure 5 illustrates the steps followed in order to define the Ground Truth Matrix. The process identified under the term "**Manage Segmented Areas**" correspond to the procedure illustrated in fig.1.
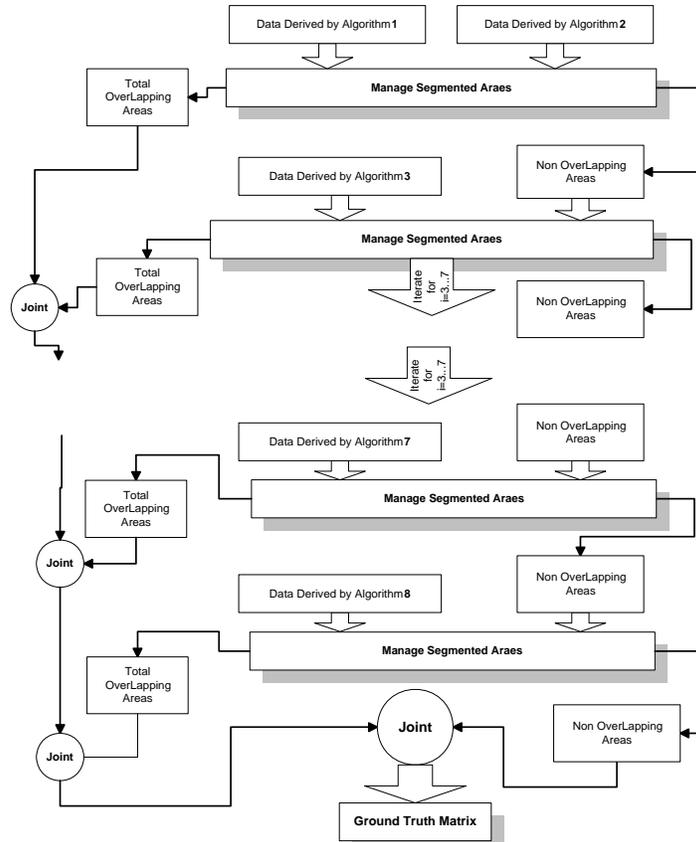
Figure 5: Flowchart of the Ground Truth Extraction Approach.

3.B.3.   Receiver Operating Curves

The segmentation of an image through an algorithmic approach (AS) is compared to the Ground Truth (GT) specification of that image to count instances of correct segmentation, under-segmentation, over-segmentation, missed regions, and noise regions. The definitions of these metrics are based on the determination of overlap. In particular in the current thesis, the overlapping is considered in terms of pixels and in terms of labels. For the first case, an instance of "correct segmentation" is recorded if and only if a pixel is detected by both an AS and the GT. At this case, we measure as true positives the number of pixels segmented by the AS that are also determined by the GT. Consequently as false positive instances of segmentation (FPs) we consider the pixels segmented as degraded by the AS not by the GT, while TNs and FNs correspond to pixels that were correctly and incorrectly segmented as cleaned regions. On the other hand, in the case of "label-based checking" true positives are determined as the number of labels segmented by the AS that overlap a corresponding area detected by the GT. The FP, TN, and FN rates are also defined accordingly.

3.B.3.1.   Implementation Analysis

In a system based on human inspection of the segmentation results, when an area is detected as degraded (by a detection methodology) the expert reviews the result image to inspect the segmented patterns.

In order to make the diagnosis, the experts have in mind a pre-determined threshold value (or cut-off point) for what a degraded and a non-degraded region look like. It is important to note that experts may have considerable variations in how they discriminate decayed areas from one another and they might use different threshold cut-off values mainly depending upon their experience.

In this thesis we adopt a semi-automated process of detection evaluation. Throughout this approach, the Receiver Operating Characteristic curves are extracted by automatically modifying the algorithmic parameters in a range of values (by adjusting values from more to less strict) and by performing decay patterns' segmentation for each parameter modification. The detected areas are subsequently compared towards the GT in order to define the pixels corresponding to TPs, TNs, FPs and FNs instances of segmentation. The sensitivity and specificity values were calculated according to these measures as

$$Sensitivity = \frac{TP}{TP + FN}$$
$$Specificity = \frac{FP}{FP + TN}$$

(3)

The pixel-based approach provides information on the effectiveness of each algorithm in determining decay effects at their real extent. However, it cannot assess their efficiency in providing robust estimation of segments' topology. Such analysis is performed through a similar procedure based on measuring the number of labels segmented by both the AS and the GT. In the latter approach, the areas segmented by the GT and AS are labeled and subsequently the structures, which keep information for the labels, are traversed. For each label derived by the AS, we check whether the GT also detected a label at the corresponding locations. The number of segmented labels defines the FPs, TPs and FNs. After the determination of these metrics, the sensitivity and specificity along with the precision[3] and recall[4] measures are calculated. Figures 6(a) and (b) illustrate the label-based and the pixel-based approaches respectively.
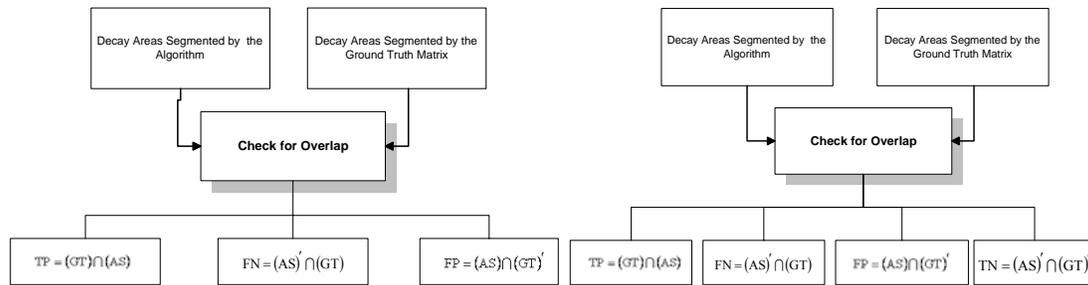
**Figure 6:** (a) Check for Overlap performed through the label based approach (b) Check for Overlap performed through the pixel based approach.

The process determined as "Check for Overlap" in fig. 6(both (a) and (b)) derives both the labels segmented by the GT and not the AS along with the labels detected by the AS and not the GT

---

[3] $precision = \frac{TP}{TP + FP}$ [2] $recall = \frac{TP}{TP + FN}$

The processes followed to determine instances of correct or incorrect segmentation demonstrate high computational complexity. This especially occurs when performing label-based convergence check as the areas segmented through the application of each threshold are labeled and subsequently the traversing of the store structures and the checking for overlap induces high computational cost. Throughout the next section, we provide an overview regarding the modification of algorithmic parameters.

3.B.3.2.   Parameter Modifications

Initially, in order to determine the range of parameter levels, we adjust some marginal values and observe each algorithm's responses. The borders of the range are subsequently defined according to these tests.

At first we study the performance of the High-Pass Filtering algorithm. Thus, we start by modifying the thresholds between marginal values. More specifically, after the acquisition of the detail image (see section 2.B.1.1), the image's histogram is extracted and the lower and upper quartiles of intensity levels are calculated. At a further step the algorithm is executed iteratively by modifying the threshold Th between the lower and the upper quartiles with a step equal to 1. The segmented areas acquired after each modification are compared towards the GT. This formula of parameter tuning is applied both for the pixel-based and for the label-based performance study.

The sub-region decomposition algorithm segments the corroded areas via the use of multiple parameters. As it is thoroughly discussed in section 2.B.3, after the detail image has been decomposed into square non-overlapping regions the histogram of each of the regions is extracted and Sub-Region Decomposition values as measures of the histogram's asymmetry are calculated. At a further stage, only the blocks that demonstrate Sub-Region Decomposition values above some experimentally determined thresholds (referred as T_1 and T_2) are considered for corrosion defects occurrence. This final stage is performed via determining the threshold through the lower and upper quartile gray levels $(\text{Th} = \text{Lower\_Q} - k_1(\text{Upper\_Q} - \text{Lower\_Q})$. To determining the trade-off between sensitivity and specificity the values of k_1 as well as the thresholds T_1 and T_2 are modified to capture all possible combinations. More specifically, T_1 and T_2 are modified by assigning values, which lay in the interval [0, 4] with a step equal to 0.2. Thus for Skewness=T_1 $\{T_1 \in [0,4] \mid \forall \; Kurtosis \in [0,4] \; and \; \forall \; k_1 \in [-3,3]\}$ the detection procedure takes place and the segmented areas are checked towards the ground truth to form the ROC and the precision recall curves. As it can be observed, the extraction of the performance curves is conducted by employing a costly (computationally) approach. This mainly occurs because the sensitivity and specificity values change rapidly and thus a low sampling rate guarantees more accurate illustration of the algorithm's response.

As it concerns to the Region Growing Algorithm, according to chapter 2.B.4 all pixels with intensity levels under the median are selected as seed pixels. A region is grown around a seed pixel by appending its 8

---

connected neighbours that satisfy a similarity predicate based on a tolerance parameter (t). In the attempt to measure the sensitivity versus the specificity in this approach a tuning parameter α is induced. The range of values assigned to α fall in the interval [-10 10] and the segmentation procedure is iterated with a step equal to 0.2 as reported in equation (4).

$$p(i,j) \leq \alpha(1-t)\frac{F_{min}+F_{max}}{2}$$ (4)

Where $p(i,j)$ is the pixel being checked, F_{max} and F_{min} are current maximum and minimum values of the region being grown. For each **α,** the optimal t value is automatically derived for each segmented structure by repeating the growth with multiple values of t in the interval [0.01, 0.4] (at each iteration, t is increased by a quota equal to the inverse of the seed pixel value). The t-value that introduces the least change to the feature vector (the elements of the feature vector is the segmented areas size and center of gravity) from one step to the following is chosen as the optimal tolerance value. The segmented areas acquired after each modification of α are checked for overlap towards the GT.

The extraction of the performance curves in the case of the adaptive thresholding schemes is conducted through modifying the threshold values Th and not the windows' extent. However some experiments showed that adaptation of the windows' size, according to the size of the texture elements, encountered in the studied image, increases the algorithm's performance. This can be explained if we take under consideration that the appropriate selection of the window's size reduces the splitting effect and FNs/FPs induction. In this process though, as it concerns to the Mean-Variance criterion the applied threshold is given by the equation $Th = Mean - k_1 std \quad k_1 \in [-4,4]$. Thus, the procedure starts by assigning k_1=-4 and is iterated by incrementing k_1 with a step equal to 0.1 until k_1=4. A quite similar process is also employed for the determination of performance curves in the case of the BoxPlot labeling algorithm. The threshold applied in this case can be expressed as $Th = Lower\_Q - k_1(Upper\_Q - Lower\_Q) \quad k_1 \in [-3,3]$.

The modification of algorithmic parameters in the case of the DoG detector is also computational costly. As it was discussed, in section 2.B.1.2 the detection scheme involves twin thresholding. At first the histogram of the detail image is extracted, the standard deviation of intensity levels is evaluated and an initial threshold of the form $Th_1 = k_1 std_1$ is applied. A second threshold $Th_2 = k_2 std_2$ is applied only to pixels above Th_1. In order to extract reliable information concerning the performance of the algorithm, for every $k_1 \in [0.2,5]$ we assign to k_2 the values belonging in the interval **0.2: 0.1: 5**. In this way all the possible combinations between k_1 and k_2 are covered.

As it was discussed in section 2.B.6, the Conditional Thickening Algorithm stems by a fusion of the segmented areas derived by the DoG and the Morphological Detectors. Thus, it is obvious that the performance curves extraction induces high computational cost, because 4 parameters should be adjusted. In particular, the parameters k_1 and k_2 employed in the DoG detector were modified in parallel with the

k_1' and k_2' employed by the Morphological Detector and the results acquired by the two algorithmic approaches are reconstructed through the conditional thickening operators.

3.B.4. Implementation of Statistical tests

3.B.4.1. T-Test

The objective of the t-test application is to assess some individual characteristics of the cleaning process. More specifically, the aim of the testing is to estimate whether the cleaning interventions introduces alterations on the crusts' thickness. In order to increase the population of the examined samples, we decompose the original image into sub-images of equal size. Thus, the testing process starts by decomposing the image obtained at the output of the segmentation procedure (image size 576x768 pixels) into 6 sub-images (of size 288x256 pixels). Subsequently, the segmented decay areas occurring in each of the sub-images are labeled and the distributions of intensities within decay areas are extracted.

The procedure of acquiring the intensities' distribution operates as described below:
A mask of the same dimensions as the image is used. At first, the sub-image depicting the detected black particles is scanned from the upper left to the right bottom corner. When a black pixel is encountered the value 1 is assigned to the element of the mask with the same co-ordinates as the studied pixel. After the whole procedure has terminated, the co-ordinates of the mask's elements with values equal to 1 correspond to problematic regions. At a further step, the mask is multiplied with the image element by element. All the non-zero results are sorted and statistical values are extracted. The statistical parameters in concern are: the mean intensity level and the standard deviation of intensities. Subsequently, the above statistical parameters, obtained for each sub-image, are used to calculate the t-statistic. The formula recruited for the computation of the t-statistic is summarized in the following equation:

$$t = \frac{M_1 - M_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{1}$$

Where s is an estimate of the standard deviation based on both samples jointly and n_1 and n_2 are the number of observations within each group and is defined in (2).

$$s = \sqrt{\frac{(n_1 - 1) \times std_1 + (n_2 - 1) \times std_2}{n_1 + n_2 - 2}} \tag{2}$$

where std_1 and std_2 denote the standard deviations measured in the populations 1 and 2, respectively, with

$$degrees\ of\ freedom\ (df) = n_1 + n_2 - 2 \tag{3}$$

Through this test we examine whether the distribution of intensity values corresponding to crusts located at cleaned areas are laid to higher values comparing to those belonging to crusts encountered on unsheltered surfaces. The key factor for selecting the t-test is that the distribution intensities obeys to normal distribution. Figure 7 presents a flowchart depicting the steps followed through the T-test and the Mann-Whitney U-Test.
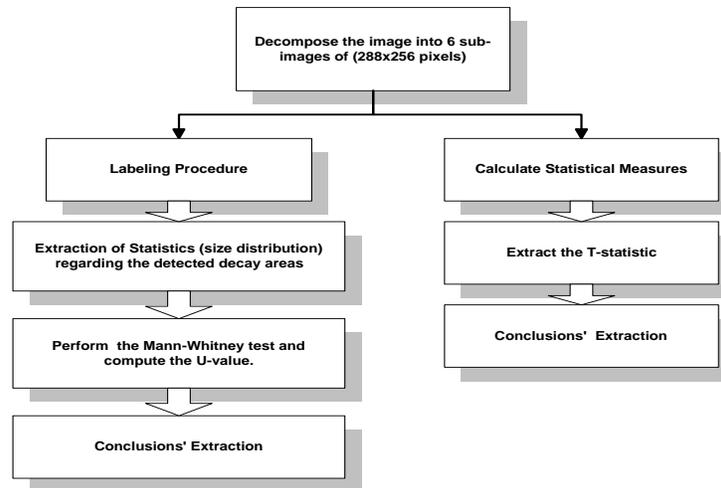

Figure 7: Block diagram illustrating the statistical tests performed.

Further to the t-test, the Mann-Whitney U-test is also used to exemplify how the recruitment of the cleaning intervention processes influences the size of the detected decay patterns. A visual inspection of the results indicates that the decay patterns encountered on cleaned areas tend to provide a distribution of pattern sizes, which is laid to smaller values. The Mann-Whitney U-test is used to determine whether the observed differences are statistically significant.

3.B.4.2. Mann-Whitney U-test

After the detection procedure has taken place, the image derived is decomposed into sub-regions (as depicted in fig. 7). Further to the decomposition step, the decay areas segmented on each sub-image are labeled and statistical measures regarding the median; mean and standard deviation of the decay area sizes are calculated. The measured statistical results are grouped according to the type of surface that they belong. Prior to selecting the appropriate statistical test, to assess the occurrence of discrepancies on label sizes, the distributions of decay area sizes (as they were determined through the various algorithms) revealed that they significantly depart from the normal distribution. The Mann-Whitney U-Test was employed to estimate the significance of difference between the studied populations. The steps followed by the Mann Whitney U-test are:

1. List observations in order of magnitude within each group. Assign ascending ranks 1, 2, 3, ... to the whole set of observations with repeated values, called 'ties', given the mean of the ranks within that run.

2. Sum the ranks of each population R_A R_B.

3. Calculate U_A and U_B, e.g., $U_A = \{n_A(n_A+1)/2 + (n_An_B)\}-R_A$ where n_A and n_B are the number of patients in each group. A similar equation can be constructed for U_B by substituting n_B and R_B.

4. Take whichever is the smaller of U_A and U_B to the statistical table. U has to be equal to, or LESS than, the tabulated value for significance, i.e., low values of U indicate significance.

The shape features of the segmented decay patterns are considered as important characteristic able to indicate the variations in the patterns' shape features, as they were determined through the various employed algorithms. Another aspect also considered in this study is the extraction of shape descriptors that reflect deviations induced by the cleaning interventions. The shape descriptors recruited in the current study are mainly associated

3.B.5.   Shape Features Extraction

One set of features used to provide additional information on the decay patterns is some features regarding their shape characteristics. Before extracting the shape features, the decay patterns should have been accurately detected. Thus the shape extraction process is performed after the detection procedure has already taken place and the detected decay areas are labeled. Shape representation is an important problem in image processing and pattern recognition. A good shape representation makes it easier for a shape to be stored, transmitted, compared against and recognized. Numerous techniques for shape representation have been developed and several criteria for a good shape representation shape been established.

- Efficiency: simplicity and compactness
- Accuracy: accurate and complete reconstruction.
- Effectiveness: suitability for shape analysis and shape recognition.

3.B.5.1.   Boundary Sequences and Cross-Sections

Following to the determination of the objects boundaries, we proceeded to storing information regarding the exact co-ordinates of the boundary pixels. Through this stage we extract boundary sequences as contour based shape representations. The boundary sequence is defined in this thesis as an ordered sequence of boundary pixel locations in clockwise order. The boundary sequence has proven to be a simple compact representation for an arbitrary shape and can describe contour of the shape accurately and completely. However, the boundary sequence is not suitable for computing shape features. Therefore a cross-section generation algorithm is proposed in sub-section 3.5.2, which determines all the cross sections in a given shape by tracing the boundary sequence of the shape once. The cross-sections are similar to run lengths in the shape so shape features can be easily computed. The boundary sequence is also useful in computing the minimum bounding rectangle (MBR), perimeter, Euler number of the shape that cannot be computed easily in the run-length encoding representation. A boundary sequence extraction method is also described in sub-section 3.5.3, which generates a boundary sequence for each shape in a binary image by scanning the image only once. The cross-section generation algorithm and the boundary sequence extraction method work well even if a shape has holes.

3.B.5.2.   Cross Section Generation Algorithm

A shape is represented as region R in a digital binary image and a clockwise boundary sequence of the region is defined as follows:

$$R = \{(x,y) | (x,y) \in \text{ shape}\}$$
$$BS(R) = \{(x_u, y_u), u = 0 \cdots, n-1\}$$

Where $|x_k - x_{k-1}| \le 1$ and $|y_k - y_{k-1}| \le 1$ for $k \ge 1$. A vertical cross-section, $C(x_i, y_l : y_u)$ can be defined as

$$C(x, y_l : y_u) = \left\{ (x,y) \left| \begin{matrix} y_l \le y \le y_u, \ (x,y) \in R \\ (x, y_l) \text{ and } (x, y_u) \in BS(R) \end{matrix} \right. \right\}$$

Where (x, y_l) and (x, y_u) represent the lower and the upper end-points of the vertical cross-section respectively.

When a boundary sequence is traced sequentially, there are three types of x-value difference (Δx) between a current point and its previous one. If $\Delta x > 0$, the trace from the current point to the previous point is called a right trace, or we say that the trace is a trace with a positive Δx. If $\Delta x < 0$ the trace is   called a left-trace. If $\Delta x = 0$, the trace with positive (resp. negative) Δy is called a down trace. (resp. up-trace). If consecutive traces consist of right- or left traces, there are only six cases as shown in fig. 10.
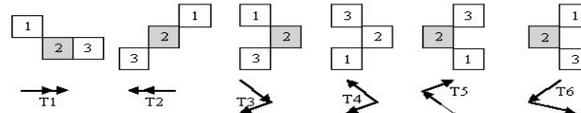
Figure 8: Six cases of consecutive traces.

The T_1 and T_2 in fig.8 represent cases of not changing the direction of traces. Then the middle point p_2 in T_1 (resp. T_2) is registered as the upper (resp. lower) end-point of a vertical cross-section as shown in fig. 10.

The T_3, T_4, T_5 and T_6 represent typical cases of changing the direction of traces. In case of changing in clockwise fashion, e.g, T_3 and T_5, the rightmost point (p_2 in T_3) or the leftmost point (p_2 in T_5) is a vertical cross-section in itself. However, the extreme points in T_4 and T_6 are determined as the middle points in a vertical cross-section, because the trace direction changes in a counter-clockwise fashion. Whether the trace direction changes in a clockwise or in a counter-clockwise fashion can be easily determined by testing difference of y-values between previous and next points of the extreme point. For example, the difference in T_3 is greater than 0, while that in T_4 is negative.

When there are down-traces or up-traces between the consecutive right- or left traces, there is a valid run (a, b in fig. 9). If the previous trace is a right trace, top points of the vertical run (a, b in fig. 9). If the previous trace is a right trace, top point of the vertical run is registered as the upper end point (a in fig. 9). If it is a left-trace, bottom point is registered as the lower end-point (b in fig. 9) When there is a vertical run in case of T_3 or T_5, its top and bottom points are registered as the upper and the lower end-points respectively.
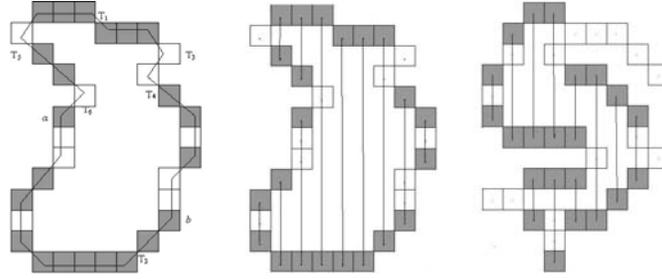
Figure 9: Examples of cross-sections

The vertical cross-section generation algorithm is summarized as follows:

// Input: A boundary sequence for a region R,

$BS(R) = \{(x_0, y_0), (x_1, y_1), \cdots, (x_{n-1}, y_{n-1})\}$, where (x_i, y_i) denotes the boundary pixels co-ordinates and it is derived by the boundary sequence extraction process described in sub-section 3.3.4.

1. Trace BS(R) until a point (x_i, y_i) whose x-difference $\Delta x_i (= x_i - x_{i-1}) \neq 0$ is found. If there is not such a point, $Insert\text{-}CS\text{-}List\left(x_0, \min\{y_i | x_i, y_i \in BS(R)\}\right)$ and $Insert\text{-}CS\text{-}List\left(x_0, \max\{y_i | x_i, y_i \in BS(R)\}\right)$ Then exit.

2. Set the direction flag $F_d = \Delta x_i$ and the starting point $s = (x_i, y_i)$. Set $Y_{pre} = y_i$ and $Y_{current} = y_i$.

3. Set $j = (i + 1) \bmod n$ and move to next point (x_j, y_j). Compute $\Delta x_j (= x_j - x_i)$.

   3.1   If $\Delta x_j = F_d$, $Insert - CS - List(x_i, Y_{current})$. Then set $Y_{pre} = Y_{currenti}$ and $Y_{current} = y_i$.

   3.2 $\Delta x_j = -F_d$, then

      3.2.1   If $F_d \times (Y_{pre} - y_j) \geq 0$, $Insert - CS - List(x_i, Y_{current})$ and $Insert - CS - List(x_i, y_i)$. Then,

         set $Y_{pre} = y_i$, $Y_{current} = y_j$, and $F_d = \Delta x_j$.

      3.2.2   Else, set $Y_{current} = y_j$ and $F_d = \Delta x_j$.

   3.3   If $\Delta x_j = 0$ and $F_d = \Delta y_j$, then $Y_{current} = y_j$.

4. Repeat step 3 until $(x_j, y_j) = s$

Step 1 in the algorithm is to find a starting point whose previous pixel is to its left or right. For a vertical shape with one-pixel width, the algorithm outputs only one vertical cross-section.

The CS-list consists of $(x_{max} - x_{min} + 1)$ buckets. A bucket [x] keeps y-values of end-points of vertical cross-sections $C(x, y_l : y_u)$'s being sorted in ascending order. The procedure $Insert - CS - List(x, y)$ inserts y into appropriate location among y-values in bucket[x] as shown in figure 9. Eventually, we can determine vertical cross-sections by pairing the y-values of each bucket[x]. In figure 9 there are three vertical cross-sections, $C(a, 3 : 7)$, $C(a, 9 : 11)$ and $C(a, 13 : 16)$.
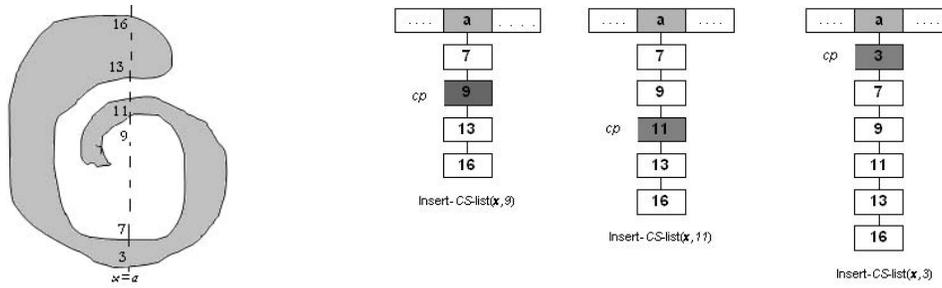
Figure 10: Examples of insert-CS-list(x, y)

If a hole H is in a region R, we can determine vertical cross-sections y merge-sorting y-values of bucket[x] I CS-list for R and those in CS-list for H. However, in CS-List for H, y_u should be increased by 1 and y_l decreased y 1, because BS(H) is the boundary sequence that is obtained y following the boundary of the hole.

3.B.5.3.   Boundary Sequence Extraction Method

From above discussions, we know that a region can be described effectively by its boundary sequence. If a region R includes holes, $H_1, H_2, \ldots\ldots, H_m$, it should be represented as a boundary descriptor, BD(R), defined as follows.

$$BD(R): BS(R) \rightarrow BS(H_1) \rightarrow \cdots\cdots \rightarrow BS(H_m)$$

The boundary sequence extraction method scans the image in a raster-scan manner as in TV. There are three different states of current scan pixel, p in a scan line as follows and the state transition diagram is represented in figure 11.

$S_0 : p \in$ background region
$S_1 : p \in$ foreground region
$S_2 : p \in$ hole region
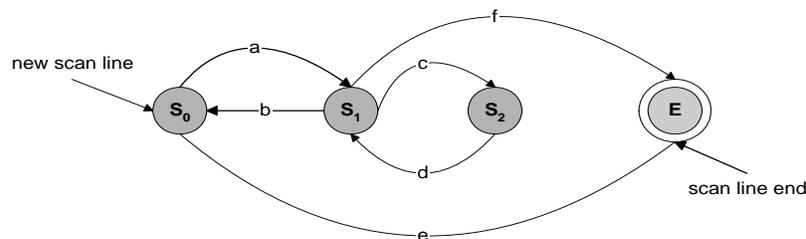
Figure 11: State transition diagram.

The state transition diagram is initialized when a new line starts and finishes at location of end-pixel in the scan line starts (state E). When a new object region or a new hole region is being extracted, pixels of extracted boundary sequence are marked with appropriate label that is assigned differently to each object region and each hole region. The boundary sequence can be extracted by the boundary following operation

---

(discussed in the subsequent section). A nested hole counter is used to determine which state is the next one when a background pixel is found in S_1. The transition conditions are summarized as follows.

a, d: When a foreground or a labeled pixel is met.

b: When a background pixel is met and the nested count is zero.

c: When a background pixel is met and the nested count is not zero.

e: When there remain only the background pixels in the current scan line.

f: When no background pixel is found.

In each state, following operations are performed.

S_0: When a foreground pixel is found, a new boundary descriptor for the new object region is generated through boundary following.

S_1: When a background pixel is found, a new hole boundary sequence for the hole region is generated through boundary following. The hole boundary sequence is attached to appropriate boundary descriptor.

S_2: When a foreground pixel is found, a new boundary descriptor for the new object region in a hole is generated.

Figure 12 shows examples of extracting boundary sequences in an image. When all pixels in a scan-line are background pixels and the transition condition e occurs. If a foreground pixel is found in S_0, then a new boundary descriptor is generated through following the boundary of corresponding object region (R_1), as shown in fig 12(a). The transition condition b occurs in fig. 12(b) when a background pixel is met in S1 and the nested-hole counter is zero, while the transition condition c occurs in fig. 12(d) because the nested-hole counter is not zero. The hole (H1) is detected in fig. 12(c) because the background pixel is found in S_1. A new boundary descriptor for object region (R3) is also extracted in fig. 12 are as follows.

$$BD(R_1): BS(R_1)$$
$$BD(R_2): BS(R_2) \rightarrow BS(H_1)$$
$$BD(R_3): BS(R_3)$$
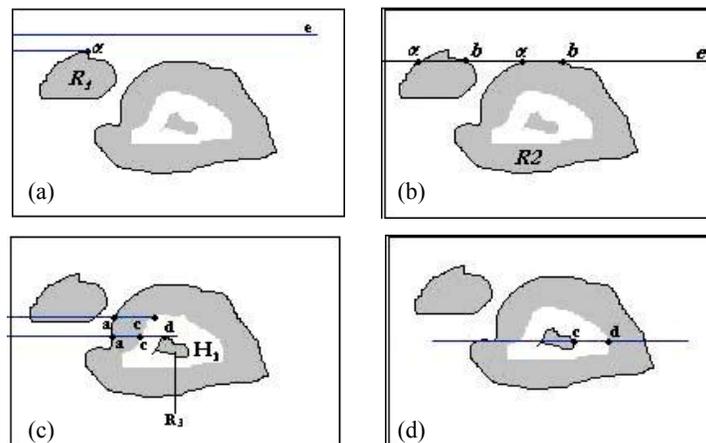
Figure 12: Examples of boundary sequences.

3.B.5.4.   Boundary Following

The boundary of a connected component S is the set of pixels of S that are adjacent at [27]. In most applications one wants to track pixels on the boundary in a particular order. One common approach is to track all pixels of a region in a clockwise sequence. In the current approach, a simple boundary following algorithm was implemented. The boundary following algorithm selects a starting pixel $s \in S$ and tracks the boundary until it comes back to the starting pixel, assuming that the boundary is not at the edge of the image. Subsequently we present the steps of the algorithm and the and in fig. 13 the boundary found by this algorithm for an 8-connected region.

Boundary Following Algorithm

1)   Find the starting pixel $s \in S$ for the region using a systematic scan from left to right and from top to bottom of the image.

2)   Let the current pixel in boundary tracking be denoted by c. Set c=s and let the 4-neighbor to the west of s be $b \in \overline{S}$.

3)   Let the 8-neighbors of c starting with b in clockwise order be $n_1, n_2, \cdots, n_8$. Find ni for the first I and that is in S.

4)   Set c=n$_i$ and b=n$_{i-1}$.

5)   Repeat steps 3 and 4 until c=s
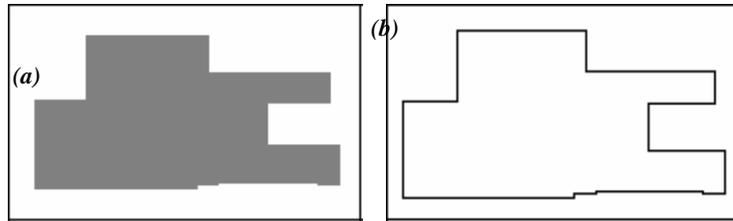
Figure 13: Results of the boundary following algorithm. (a): Original binary object. (b) Calculated Boundary

3.B.5.5.   Shape features from cross-sections

Let the set of vertical cross-sections for a region R be Ω. Then, the region can be represented as the union of the vertical cross-sections in Ω.

$$R = \bigcup_{C(x, y_1 : y_u) \in \Omega} C(x, y_1 : y_u).$$

Let define the length of a vertical cross-section $C(x, y_1 : y_u)$ as follows.

$$l(x, y_1 : y_u) = y_1 - y_u + 1$$

We can easily compute the perimeter, compactness eccentricity and boundary moments of a region R from its boundary sequence. The Euler number is defined as the number of components minus the number of holes. So, it can be also determined easily from the number of the holes N(H) in R.

$$E = 1 - N(H).$$

The compactness of an area is computed by the ratio:

$$Compactness = \frac{(Perimeter)^2}{Area}$$

The perimeter value is derived by measuring the length of the boundary sequence. More specifically, to calculate the compactness parameter, once we have extracted the boundary sequence, we subsequently look up the list of the labels to derive the corresponding labels co-ordinates. Thus , $Area = Number\ of\ pixels\ in\ the\ label$. The centroid of a region R can be determined directly as follows, while the vertical projection should be previously in run-length encoding representation.

$$\bar{x} = \frac{1}{A} \sum_{C(x,y_l:y_u) \in \Omega} (x.l(x, y_l : y_u))$$

$$\bar{y} = \frac{1}{2A} \sum_{C(x,y_l:y_u) \in \Omega} ((y_l + y_u).l(x, y_l : y_u))$$

In order to measure the eccentricity of an area we extract the major and the minor axis from each boundary sequence. As it is known from the literature, the minor axis corresponds to the axis that connects the two point of the boundary with the maximum Euclidean distance and we store this distance as the length of the major axis. Once the major axis has been extracted, we take all the projections on the major axis and extend them until they reach points of the boundary. Subsequently, the minor axis is selected to be the projection of maximum length.

$$Eccentricity = \frac{|Major\ Axis - Minor\ Axis|}{Major\ Axis}$$

Another feature which was also measured but is not included in the reports that are provided in the results chapter is the vertical projection. This parameter can be computed as follows by summing vertical cross-sections in each bucket.

$$V[i] = \sum_{C(x,y_l:y_u) \in \Omega \& x=i} l(x, y_l : y_u).$$

Orientation of an elongated shape is defined as orientation of the axis of least inertia in [Machine Vision (Jain)] and it can be computed from three parameters a, b and c as follows:

$$\alpha = \sum_{(x,y) \in R} (x - \bar{x})^2 \Rightarrow \alpha = \sum_{C(x,y_l:y_u) \in \Omega} (x - \bar{x})^2 l(x, y_l : y_u)$$

$$b = \sum_{(x,y) \in R} (x - \bar{x})(y - \bar{y}) \Rightarrow b = \sum_{C(x,y_l:y_u) \in \Omega} (x - \bar{x}) f(n_u, n_l)$$

$$c = \sum_{(x,y) \in R} (y - \bar{y})^2 \Rightarrow c = \frac{1}{6} \sum_{C(x,yl:yu) \in \Omega} (x - \bar{x}) \cdot g(n_u, n_l)$$

Where,

$$f(n_u, n_l) = |n_u|(|n_u| + 1) - |n_l|(|n_l| + 1)$$

$$g(n_u, n_l) = |n_u|(|n_u| + 1)(2|n_u| + 1) - |n_l|(|n_l| + 1)(2|n_l| + 1),$$

$$n_u = y_u - \bar{y}, \quad and \quad n_l = y_l - \bar{y}.$$

That is $\tan(2\theta) = \frac{b}{(a - c)}$. Once the orientation of each label is computed, the information is stored and we extract a histogram presenting the orientation of the objects boundaries encountered in each of the examined images. The results derived through this procedure are recruited to assess whether a significant difference occurs in the orientation values of the decay patterns detected on surfaces submitted to different weathering conditions.

Based on each label's boundary sequence and its centroid co-ordinates, we attempt to evaluate some boundary moments. By $z(i), i = 1, 2, \ldots, N$ are the Euclidean distances of the boundary pixels to the centroid (where N is the number of these pixels) the pth moment is defined as

$$m_p = \frac{1}{N} \sum_{i=1}^{N} [z(i)]^p$$

And the pth central moment is defined as

$$M_p = \frac{1}{N} \sum_{i=1}^{N} [z(i) - m_1]^p$$

In addition, low order moments are defined and extracted which are less sensitive to noise. These are named F$_1$ and F$_3$. These are defined below:

$$F_1 = \frac{\left[\frac{1}{N} \sum_{i=1}^{N} [z(i) - m_1]^2 \right]^{1/2}}{\frac{1}{N} \sum_{i=1}^{N} z(i)} \quad F_3 = \frac{\left[\frac{1}{N} \sum_{i=1}^{N} [z(i) - m_1]^4 \right]^{1/4}}{\frac{1}{N} \sum_{i=1}^{N} z(i)} \quad \text{where} \quad m_1 = \frac{1}{N} \sum_{i=1}^{N} [z(i)].$$

It has been shown [36] that the F$_3$-F$_1$ is the best moment to represent roughness of micro calcifications. As features for representing the shape of micro calcifications, we have used the median and the mean value of F$_3$-F$_1$ along the labels encountered on the same image.

An illustration of the system that we have developed for the extraction of shape features is presented in the block diagram of figure 14.
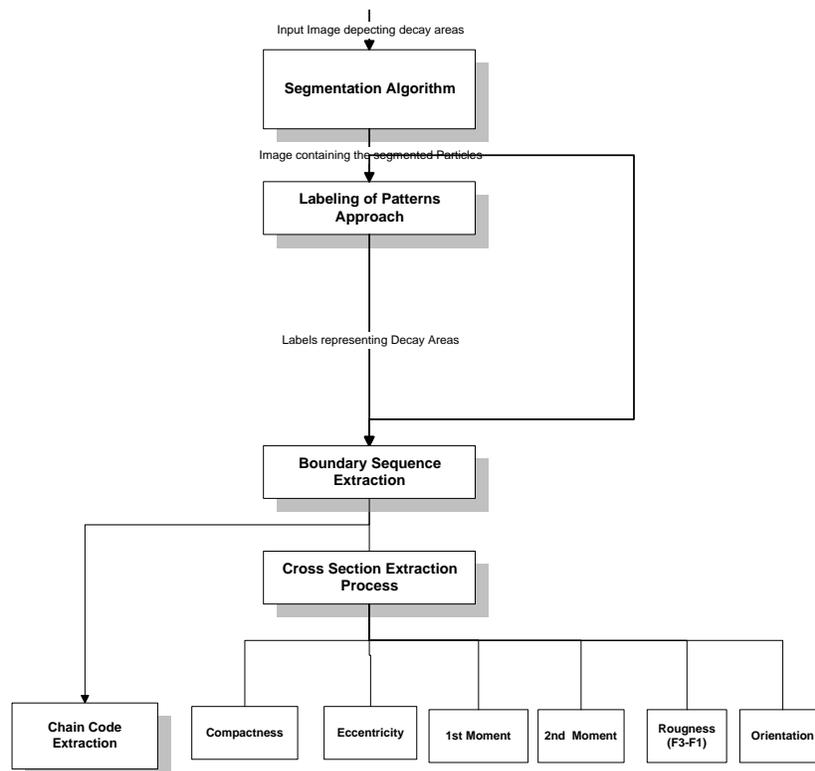

Figure 14: Block diagram illustrating the system developed for extracting shape features.

4. Results

4.1. Overview

This work initially validates the potential and the limitations of each of the recruited algorithms in effectively determining the topology and the extent of decay patterns. At a further step, we study the nature (size distribution and spatial arrangement) of decay patterns that are either segmented by all the algorithms or by each one individually. The latter study aims at extracting significant differences between the algorithms' responses. As it was extensively discussed in chapter 1, another objective of the current work is to study the size and intensities of degraded areas as representative measures of the severity of degradation. In order to estimate whether significant differences occur between decay patterns segmented on various surfaces, we perform several statistical tests. The aim of these tests is to assess the significance of gray scale intensities alterations induced by cleaning or by the different exposure conditions. Furthermore, these tests are used to evaluate changes (reduction) on decay patterns' sizes caused by the application of cleaning treatments or other structural effects. These tests contribute to assessing the mechanisms and the efficiency of chemical cleaning and to understanding the formation of crusts.

Another aspect that is also considered regards the effectiveness of several imaging modalities in providing reliable information on the extent and the severity of decay effects. The monitoring systems studied in this work are: (1) Fiber Optics Microscopy (FOM), (2) Digital Camera, (3) Reflectography in the visible (Vis), near infrared (NIR) and infrared (IR) spectral bands. The studied stone specimens correspond to marble surfaces, where adjacent regions of cleaned and un-cleaned crusts prevail. The cleaning process was conducted by a Nd:YAG laser system used to partially remove the crust. Throughout the cleaning process, some parameters such as the laser pulses are modified resulting in the removal of crust layers differing in thickness. The afore-mentioned surface is depicted via the FOM in 33 images and via the Reflectography (Vis), (IR), (NIR) and digital camera by one image respectively. The images are studied to validate the accuracy of each imaging modality and to estimate the degree to whether the monitoring systems converge in the determination of corrosion.

Finally, we examine how the cleaning and structural effects are reflected on the shape of degraded areas. This is estimated through studying the shape features and the compactness (the presence of nested regions and holes in the segments. These are associated with the occurrence of crusts' discontinuities) of the detected decay spots. The incentive of the shape analysis is to define object attributes that contribute to the classification of stone corrosion damage.

The algorithmic results are evaluated by using images demonstrating representative effects of corrosion damage. More specifically, for the algorithms' performance study, we use two FOM images presenting decay phenomena that prevail on sheltered and unsheltered areas and an image obtained by a digital camera system. The experts selected the afore-mentioned images in order to extract the Ground Truth Matrix of corrosion effects. For the rest of comparisons we use many images depicting stone material where various corrosion phenomena prevail.

Chapter 4 is structured according to the four directions of this work. Section 4.1 examines and compares the responses of the recruited algorithmic schemes and is sub-divided into three parts. Through subsections 4.2.1-4.2.2 we discuss the Ground Truth Matrix Extraction and we compare the algorithms' efficiency in determining the extent and the topology of decay areas. The third sub-part of 4.2 examines the shape features of decay patterns segmented by each individual algorithm. Section 4.3 studies the algorithms' performance curves while 4.4 explain the results of the employed statistical tests in assessing the cleaning and structural effects. The alterations in the shape features due to cleaning and structural effects are discussed extensively in 4.5. Finally 4.6 investigate the potential and the limitations of several imaging modalities in accurately monitoring corrosion damage.

4.2. Comparison of Algorithms' Performance Over the Ground Truth

At the first stage of this work, the Ground Truth Matrix of each studied image is extracted and decay areas segmented by each of the implemented algorithms are compared towards the Ground Truth. Through this approach we attempt to investigate the extent and spatial arrangement of the decay patterns detected by all the algorithms and of those segmented by each individual algorithm and not by all. In the last subsection of 4.2, the shape features segmented by each individual algorithmic scheme are studied.

4.2.1. Ground Truth Matrix and Visual Evaluation

As it was discussed in the previous chapter, the extraction of Ground Truth Matrix of decay areas is critical in the whole procedure of measuring the algorithms' performance and accuracy. The role of the expert is important in the whole procedure. This person provides the most objective way of defining whether the topology, extent and spatial distribution of the detected decay areas resemble to his/her own judgment of decay patterns prevalence on the specific stone surface. In our case, the degraded areas segmented in the Ground Truth, stem from a union of the areas segmented by all the algorithms (Commonly Detected Areas (CDAs)) and those that were detected by each of them but not by all (SDAs). As it was analyzed in the previous chapter, in order to extract the Ground Truth Matrix, each blob is expanded as to cover the max area detected by two or more areas in the same local region. The rest of patterns detected by each algorithm individually usually correspond to areas smaller in extent; their presence though in the Ground Truth was considered as significant. The images, on which we extracted the Ground Truth Matrix, were selected by the experts to be representative of the surfaces studied throughout this work. More specifically, we evaluated the Ground Truth on three surfaces (**a**) an untreated sheltered fluting (fig. 1), (**b**) an untreated unsheltered fluting (fig. 2) and (**c**) a stone surface depicting the co-existence of treated and untreated strips monitored by the Digital Camera system (fig. 3). In any case, the number of decay patterns detected in each image is quite large as to form a quite valid statistical set for algorithmic comparisons. In figure 1(a) (untreated sheltered fluting), a surface depicting rapidly varying background stone structure is illustrated.
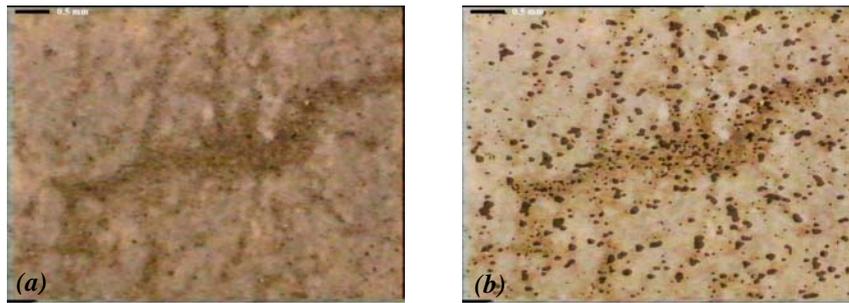
Figure 1: (a) Depicts stone specimen located on a column's fluting on sheltered surface (as monitored by the FOM (magnificationx50)), (b) the derived Ground truth Matrix overlaid on the original image.

As it can be observed, the Ground Truth Matrix segmented decay patterns large in extent, determining in this way extensive susceptible areas. Experts evaluated the Ground Truth Matrix[5] and it was assessed that all the segmented areas represent regions where corrosion damage prevails. At a further step, we studied the decay patterns that were detected by each algorithm individually (Separately Detected Areas (SDAs[6]))) and are not included in the CDAs. The SDAs tend to be arranged in isolated neighborhoods. As a measure of the spatial arrangement we employed the mean of the minimum Euclidean inter-particle distance. Thus, we measured this metric for both SDAs and CDAs. The results revealed that, for the former case this distance metric tends to be slightly larger than that for the latter. This is more observable in the cases of the Region Growing, High-Pass Filtering, and Sub-Region Decomposition Algorithms at their application to images depicting rapidly varying background structure (untreated sheltered surfaces). More specifically, experiments that were carried out revealed that the mean of the minimum Euclidean inter-particle distance concerning the CDAs was equal to 11.6 pixels while for SDAs this distance metric was equal to 14.5 pixels. However, the use of such a metric in estimating the characteristics and the arrangement of CDAs and SDAs requires further statistical processing to assess the significance of discrepancies. Figure 2 illustrates the case of the unsheltered untreated fluting and the extracted Ground Truth Matrix.
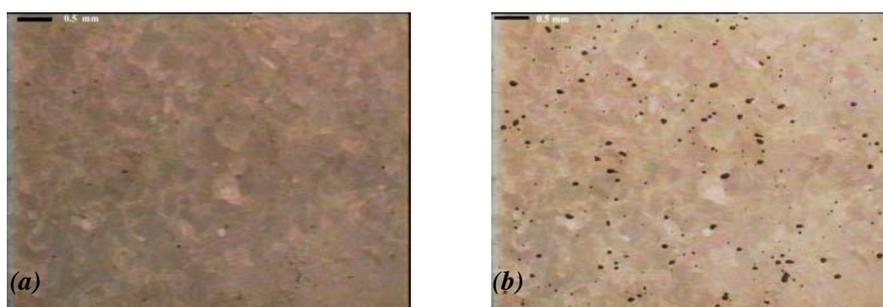
Figure 2: (a) An untreated stone specimen located on a column's fluting on unsheltered surface (as it was monitored by the FOM system (magnificationx50)), (b) the derived Ground truth Matrix overlaid on the original image.

By observing images 1(a) and 1(b) in combination with 2(a) and 2(b) it becomes evident that the Ground Truth Matrix segmented less (in number) degradation particles in the case of the unsheltered

---

[5] The mathematical expression that describes the labels in the Ground Truth is given by:
$\text{Ground Truth} = (\text{CDAs}) \bigcup_{i=1}^{8} (\text{SDAs})_i$ (index i denotes the ith algorithm).

[6] If we denote by SA_i the decay areas detected by the ith algorithm then $(\text{SDAs})_i$ can be obtained by the equation
$(\text{SDAs})_i = SA_i \bigcap (\text{CDA})^c$.

untreated fluting. This estimation is in accordance with the experts' initial judgment concerning the degradation state encountered on sheltered and unsheltered surfaces. Considering now the extent of CDAs and SDAs, it is revealed that the latter are smaller. However, investigation of their spatial arrangement in terms of the distance metric discussed previously (mean of the minimum Euclidean inter-particle distance), reveals that the metric's value is close for both CDAs and SDAs. Thus, regarding the CDAs, the distance metric is 18.12 pixels while the corresponding value for the SDAs is 19.23 pixels. Comparing the spatial arrangement of SDAs and CDAs for the cases of (fig. 1 and fig.2) we can argue that SDAs in the first case tend to be more isolated while this does not happen in the second case. This effect seems to be associated with the capability of the segmentation algorithms in detecting decay patterns on inhomogeneous surfaces. According to the authors' judgment, decay patterns related to texture irregularities, are usually arranged in neighborhoods (areas consisting of numerous degraded spots) that either correspond to "broken regions" of a single area or simply represent sporadically distributed corrosion patterns. These areas are more precisely determined on in-homogeneous images as they induce heavy outliers on the histograms of the local background. Thus, they can be detected even by using relax thresholds. Such decay patterns correspond to areas that overlap the CDAs and their small inter-particle distance further explains their neighborhood arrangement. On the other hand, decay patterns occurring on smooth background usually correspond to SDAs as their segmentation requires effective adaptation of the detection parameters since their occurrence is not obvious. The small inter-particle distance can be explained by considering that spots prevailing on smooth background represent to less porous crusts, where less deposited materials are entrapped.

Finally, in order to provide a visual inspection of the Ground Truth matrix for the case of the digital camera image we present fig 3.
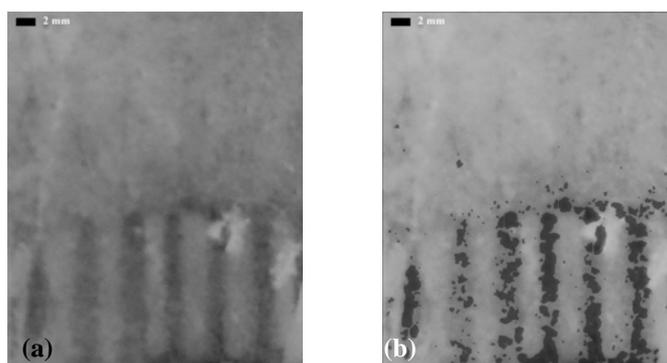
Figure 3: (**a**) A stone surface monitored by the digital camera, (**b**) The derived Ground truth Matrix overlaid on the original image.

Figure 3(a) depicts a stone surface which was partially cleaned via the aid of an Nd:Yac laser cleaning approach. On the stone material we can observe the co-existence of cleaned and uncleaned strips. According to the experts' estimation, the ground truth matrix has effectively determined the presence of degradation particles. At this point, we should make clear that the objective of the detection processes, implemented in this work, is not to segment areas of color (intensity) alteration induced by corrosion damage, but to determine the individual decay patterns, which lead to the formation of black crusts. The presence of small in extent regions is limited in the ground truth indicating that the implemented

algorithms agree in determining the topology of degradation particles. This can be explained by the fact that the digital camera provides low-resolution levels and thus the segmentation procedures are mainly based on large color alterations and do not require high sensitivity.

The validation of the algorithms' performance and the extraction of their robust points and drawbacks are conducted by inspecting the statistical results. Through the statistical data reported in he next section, the degree to which each of the algorithms matches the Ground Truth is elucidated.

4.2.2. Comparison in terms of size and number of the Detected Decay Areas

Through the statistical results presented in this section, we study the discrepancies on the responses of the implemented algorithms. More specifically, the effectiveness of the algorithmic schemes is evaluated through measuring the number or the fraction of decay areas, segmented by each algorithm, that overlap the CDAs. Some further features of the segmentation algorithms are also assessed through studying the size distribution of $SDAs_i$ and labels that overlap the CDAs.

4.2.2.1. Validation of the Algorithms Responses Regarding the Untreated Sheltered Fluting

At first, the case of the untreated sheltered fluting depicted in fig. 1 is investigated. Table 1 illustrates the number of CDAs and SDAs while Table 2 reports the percentage of overlap between CDAs and the labels segmented by each of the individual algorithms.

Table 4-1: The results derived by the check for overlapping procedure concerning the image depicted in fig.1 (untreated surface located on sheltered flutings).

	# Black Labeled Areas
CDAs	355
Checking for Overlap concerning the Region Growing	
Region Growing	1042
Labels that do not overlap to the CDAs	403
Labels Overlap to the CDAs	687
Checking for Overlap concerning the Sub-Region Decomposition	
Sub-Region Decomposition	1036
Labels that do not overlap to the CDAs	324
Labels Overlap to the CDAs	712
Checking for Overlap concerning the HighPass Filtering	
High Pass Filtering	989
Labels that do not overlap to the CDAs	273
Labels Overlap to the CDAs	716
Checking for Overlap concerning the Mean-Variance(Labeling Method)	
Mean-Variance (Labeling Method)	893
Labels that do not overlap to the CDAs	92
Labels Overlap to the CDAs	801
Checking for Overlap concerning the BoxPlot (Labeling Method)	
BoxPlot (Labeling Method)	706
Labels that do not overlap to the CDAs	82
Labels Overlap to the CDAs	624
Checking for Overlap concerning the RobustFit (Labeling Method)	
RobustFit (Labeling Method)	836
Labels that do not overlap to the CDAs	89
Labels Overlap to the CDAs	747
Checking for Overlap concerning the Difference of Gaussians Detector (DoG)	
Difference of Gaussians Detector (DoG)	998
Labels that do not overlap to the CDAs	303
Labels Overlap to the CDAs	695
Checking for Overlap concerning the Conditional Thickening	
Conditional Thickening	716

Labels that do not overlap to the CDAs	64
Labels Overlap to the CDAs	652

From the results of Table 1, it becomes obvious that the studied algorithms demonstrate discrepancies in the number of decay areas that they segment. This number varies from 1042 in the case of the **R**egion **G**rowing algorithm to 706 in the case of the **B**ox Plot Labeling Method. In any case, their number is much greater than the number of corroded areas segmented in the Ground Truth. However, this cannot provide a consistent metric of algorithms' performance, because a single corroded area may be illustrated as a neighborhood of "broken regions". An effort to detect similarities on the responses of the employed algorithms revealed that the Region Growing, Sub-Region Decomposition and the High-Pass Filtering algorithms tend to provide close segmentation results. A more accurate metric of the algorithms efficiency is the percentage of overlap between decay particles segmented by each individual algorithm and the CDAs. Table 2 presents the percentage of overlap between the CDAs and the labels detected by each individual algorithm.

Table 4-2: Percentage of the CDAs covered by the decay areas segmented by each individual algorithm

	Region Growing Algorithm	Skewness &Kurtosis Algorithm	High Pass Filtering Algorithm	MeanVariance(Labeling Algorithm)	Box Plot (Labeling Algorithm)	Robust Fit (Labeling Algorithm)	Difference of Gaussians (DoG) Detector	Conditional Thickening Algorithm
Sheltered Untreated Fluting(x50)								
Percentage of coverage between CDAs and each Algorithm	47.1%	37.5%	36.8%	31.6%	30.8%	29.2%	26.9%	47.7%
Unsheltered Untreated Fluting(x50)								
Percentage of coverage between CDAs and each Algorithm	44.9%	29.3%	34.1%	37.1%	33.8%	38.2%	38.2%	43.4%
Digital Camera Image								
Percentage of coverage between CDAs and each Algorithm	34.5%	36.3%	38.8%	39.1%	32.7%	35.4%	35.8%	41.2%

By studying Table 2 we can conclude that regarding the FOM images the Region Growing and the Conditional Thickening Algorithm present better performance. On the other hand, we can see, that the Robust Fit and the Sub-Region Decomposition Algorithms seem to be less effective. The latter conclusion can be explained by considering that these algorithms employ windows of fixed size (61x61 pixels) to perform local processing. This in turn results in splitting the segmented areas and in providing poor discrimination between the locations of decay patterns prevalence and the stone background. Experiments that were carried out by adapting the window's size according to the extent of texture elements show that the performance of the algorithms may become better. Table 3 summarizes the size distribution of SDAs$_i$ and of areas that overlap the CDAs.

Table 4-3: Results depicting the distribution of sizes of both SDAs₍ᵢ₎ and labels that overlap to the CDAs for each of the implemented algorithms (referring to fig.1)

	SDAs	Labels that **Overlap** CDAs		SDAs	Labels that **Overlap** the CDAs
Region Growing Algorithm			**Box Plot (Labeling Algorithm)**		
Lower Quartile	6	8	Lower Quartile	5	8
Median	6	13	Median	8	13
Upper Quartile	17	21	Upper Quartile	11	21
Skewness &Kurtosis Algorithm			**Robust Fit (LabelingAlgorithm)**		
Lower Quartile	5	8	Lower Quartile	5	7
Median	8	14	Median	8	12
Upper Quartile	12	25	Upper Quartile	12	20
High Pass Filtering Algorithm			**Difference of Gaussians (DoG Detector)**		
Lower Quartile	5	8	Lower Quartile	5	8
Median	8	15	Median	8	11
Upper Quartile	16	26	Upper Quartile	11	21
Mean Variance (Labeling Algorithm			**Conditional Thickening Algorithm**		
Lower Quartile	5	8	Lower Quartile	8	13
Median	8	13	Median	12	21
Upper Quartile	11	21	Upper Quartile	19	33

It can be observed that the distribution of sizes of the SDAs₍ᵢ₎ tend to be laid in lower values and this holds true for all the algorithms. A more dissect inspection of the statistical results illustrated in table 3 shows that most of the algorithms (except for the conditional thickening) studied in this work agree on the sizes of the degraded areas. In other words, they provide close size distributions of decay patterns that either overlap or do not overlap the CDAs. The Conditional Thickening Algorithm is the only one that demonstrates a distribution of sizes shifted to higher values. The latter observation was also assessed in previous stages of this work (chapter 1) and is associated to the operation of the algorithm (blowing of the labels segmented by DoG until either the borders of the corresponding labels segmented by the Morphological Algorithm are met, or two adjacent labels are merged).

4.2.2.2. Validation of the Algorithms' Responses Regarding the Untreated Unsheltered Fluting

Subsequently, we investigate similar statistical results concerning the responses of the algorithms in extracting the topology of decay areas prevalence for the other two cases (Unsheltered Untreated Fluting (x50) and Digital Camera Image). Tables 4 and 5 investigate the same aspects as those discussed in tables 1 and 3 respectively and they refer to the case of the unsheltered untreated fluting.

Table 4-4: Number of decay areas that either overlap or do not overlap the CDAs (the surface examined is an unsheltered untreated flutings).

	# Black Labeled Areas
CDAs	100
Checking for Overlap concerning the Region Growing	
Region Growing	206
Labels that do not overlap to the CDAs	50
Labels Overlap to the CDAs	150
Checking for Overlap concerning the Sub-Region Decomposition	
Sub-Region Decomposition	214
Labels that do not overlap to the CDAs	74
Labels Overlap to the CDAs	141
Checking for Overlap concerning the HighPass Filtering	
High Pass Filtering	213
Labels that do not overlap to the CDAs	24
Labels Overlap to the CDAs	181
Checking for Overlap concerning the Mean-Variance(Labeling Method)	
Mean-Variance (Labeling Method)	195
Labels that do not overlap to the CDAs	54
Labels Overlap to the CDAs	141

Checking for Overlap concerning the BoxPlot (Labeling Method)	
BoxPlot (Labeling Method)	134
Labels that do not overlap to the CDAs	29
Labels Overlap to the CDAs	105
Checking for Overlap concerning the RobustFit (Labeling Method)	
RobustFit (Labeling Method)	177
Labels that do not overlap to the CDAs	44
Labels Overlap to the CDAs	132
Checking for Overlap concerning the Difference of Gaussians Detector (DoG)	
Difference of Gaussians Detector (DoG)	180
Labels that do not overlap to the CDAs	44
Labels Overlap to the CDAs	144
Checking for Overlap concerning the Conditional Thickening	
Conditional Thickening	120
Labels that do not overlap to the CDAs	27
Labels Overlap to the CDAs	93

Table 4 reveals that some of the implemented algorithms detect close number of degraded areas and appear a similar fraction of these areas also segmented in CDAs. More specifically, it can be seen that the three initially studied algorithms namely High-Pass Filtering, Sub-Region Decomposition and Region Growing demonstrate significant similarities in the number of detected decay patterns. At this point, we should clarify that the above results are only presented as to illustrate the way that the segmented areas become split. They do not provide a consistent metric algorithmic comparison. The percentage of overlap between the CDAs and decay areas segmented by each individual algorithm provides a more reliable metric for estimating the discrepancy in the algorithmic responses.

Table 4-5: Results depicting the size distribution of SDAs and labels that overlap the CDAs, for the case of the unsheltered untreated fluting (fig. 2).

	SDAs	Labels that **Overlap** the CDAs		SDAs	Labels that **Overlap** the CDAs
Region Growing Algorithm			**Box Plot (Labeling Algorithm)**		
Lower Quartile	6	5	Lower Quartile	5	10
Median	6	8	Median	8	14
Upper Quartile	10	14	Upper Quartile	14	21
Skewness &Kurtosis Algorithm			**Robust Fit (LabelingAlgorithm)**		
Lower Quartile	6	5	Lower Quartile	6	8
Median	6	8	Median	8	14
Upper Quartile	9	14	Upper Quartile	14	25
High Pass Filtering Algorithm			**Difference of Gaussians (DoG Detector)**		
Lower Quartile	9	8	Lower Quartile	6	5
Median	10	11	Median	6	8
Upper Quartile	18	23	Upper Quartile	9	17
Mean Variance (Labeling Algorithm)			**Conditional Thickening Algorithm**		
Lower Quartile	6	8	Lower Quartile	8	11
Median	9	14	Median	10	20
Upper Quartile	14	26	Upper Quartile	16	37

Similar to table 3, an assessment that can be drawn by the above table is that the size distribution of SDAs tend be laid to cloze levels for all the algorithms (except for the Conditional Thickening). Regarding the Conditional Thickening Algorithm, it appears a distribution of decay pattern sizes laid at higher levels both for the SDAs and for the areas that overlap the CDAs. Finally, Table 5 evidences that decay patterns that overlap the CDAs, tend to be larger in extent comparing to SDAs. The latter observation supports the experts' initial judgment that SDAs correspond to areas small in size.

4.2.2.3. Validation of the Algorithms Responses on the image obtained by the Digital Camera

Through tables 6 and 7 we investigate the features of the SDAs and the labels that overlap the CDAs regarding the digital camera image. More specifically, table 6 reports the number of labels detected by each of the implemented algorithms and they overlap or do not overlap the CDAs.

Table 4-6: Number of labels segmented by each algorithm and either overlap ort do not overlap the CDAs.

	# Black Labeled Areas
CDAs	95
Checking for Overlap concerning the Region Growing	
Region Growing	237
Labels that do not overlap to the CDAs	77
Labels Overlap to the CDAs	160
Checking for Overlap concerning the Sub-Region Decomposition	
Sub-Region Decomposition	539
Labels that do not overlap to the CDAs	105
Labels Overlap to the CDAs	434
Checking for Overlap concerning the HighPass Filtering	
High Pass Filtering	225
Labels that do not overlap to the CDAs	43
Labels Overlap to the CDAs	182
Checking for Overlap concerning the Mean-Variance (Labeling Method)	
Mean-Variance (Labeling Method)	297
Labels that do not overlap to the CDAs	62
Labels Overlap to the CDAs	235
Checking for Overlap concerning the Box Plot (Labeling Method)	
Box Plot (Labeling Method)	235
Labels that do not overlap to the CDAs	73
Labels Overlap to the CDAs	162
Checking for Overlap concerning the Robust Fit (Labeling Method)	
Robust Fit (Labeling Method)	320
Labels that do not overlap to the CDAs	29
Labels Overlap to the CDAs	291
Checking for Overlap concerning the Difference of Gaussians Detector (DoG)	
Difference of Gaussians Detector (DoG)	488
Labels that do not overlap to the CDAs	140
Labels Overlap to the CDAs	348
Checking for Overlap concerning the Conditional Thickening	
Conditional Thickening	372
Labels that do not overlap to the CDAs	51
Labels Overlap to the CDAs	321

The results reported in table 6 reveal that the algorithms' performance is significantly different from in the case of the digital camera image. This fact is associated to the potential of the digital camera monitoring system to depict details of the stone structure. According to table 6, the High Pass Filtering Algorithm presents a quite better response than the others (a greater fraction of the spots that it detects overlap to the CDAs). Furthermore, it can be seen that all the algorithms agree on the topology of decay particles due to the low resolution provided by the digital camera. Thus, the detection of deterioration spots is based almost exclusively on large area intensity alterations. An effort to compare the spatial arrangement between the $SDAs_i$ and the labels that overlap the CDAs we can see that no significant discrepancies occur. In contrast, a study of the SDAs and CDAs sizes' distribution indicates that ($SDAs_i$) tend to be smaller in extent.

Table 4-7: Results depicting the distribution of sizes of both the overlapping and the non-overlapping patterns for each of the implemented algorithms (referring to fig.3)

	SDAs	Labels that **Overlap** the CDAs		SDAs	Labels that **Overlap** the CDAs
Region Growing Algorithm			**Box Plot (Labeling Algorithm)**		
Lower Quartile	5	5	Lower Quartile	4	5
Median	7	10	Median	6	11
Upper Quartile	9	18	Upper Quartile	17	22
Skewness &Kurtosis Algorithm			**Robust Fit (Labeling Algorithm)**		
Lower Quartile	4	6	Lower Quartile	4	5
Median	6	9	Median	6	10
Upper Quartile	10	15	Upper Quartile	19	18
High Pass Filtering Algorithm			**Difference of Gaussians (DoG Detector)**		
Lower Quartile	6	5	Lower Quartile	6	6
Median	10	11	Median	8	12
Upper Quartile	12	23	Upper Quartile	12	20
Mean Variance (Labeling Algorithm			**Conditional Thickening Algorithm**		
Lower Quartile	5	5	Lower Quartile	6	6
Median	7	11	Median	8	12
Upper Quartile	16	25	Upper Quartile	12	25

The previous statement is further verified in table 7. Thus, we can see that the size distribution of SDAs is laid to lower values. This observation though is not as clear as in the previous tables. Moreover, several other conclusions can be drawn concerning the segmentation performance of each individual algorithm. Thus, we can see that the Conditional Thickening approach segmented areas larger in extent while the areas segmented by the Sub-Region Decomposition algorithm are generally smaller due to splitting.

In an effort to summarize the assessments derived through the preceding discussions we focus towards three main drifts. At first, we investigate which of the implemented algorithms attains to approach better the topology of CDAs. Such results are reported in tables 1, 5 and 8. A conclusion that can be drawn is that the labels segmented by Conditional Thickening, approach better the topology of CDAs. Regarding the remaining algorithms, the fraction of the labels that overlap the CDAs is $\approx > 70\%$. Considering this fraction in association to the different types of the studied surfaces we can state that for the case of the sheltered untreated fluting the latter fraction is reduced due to the split of the segmented areas (larger in extent areas tend to be broken into sub-regions). This effect is less observable in the case of the unsheltered untreated fluting. The tests performed through tables 2, 4, & 6 though; do not provide consistent metrics for estimating the accuracy of the algorithms. However, if they are viewed in association with the Precision Recall and the ROC curves they reflect significant features of the segmentation results. In order to study the discrepancies between the algorithms in effectively determining the size of decay patterns, we present Table 2. Through studying the results reported in this table we can see that, concerning the FOM images, the Region Growing and the Conditional Thickening algorithms approach better the CDAs. The corresponding response for the other algorithms indicates that some of them tend to present better performance in the sheltered untreated fluting while others present a better performance in the unsheltered untreated fluting or the digital camera image. A noteworthy point is the fraction of CDAs covered by segments detected by the Robust Fit (Labeling Algorithm) and Sub-Region Decomposition Algorithms. At these cases, it is obvious that the percentage of coverage varies according to the surface studied. In order to optimize their performance, the

decomposition window should be automatically selected according to the extent of the texture patterns encountered on each image. The third and most important drift of the preceding study is to measure the extent and the spatial arrangement of SDAs and of areas overlapping the CDAs. The former is obtained by providing the sizes' distributions of the corresponding labels, while the latter is assessed through measuring the mean of the minimum Euclidean inter-particle distance. The results of sizes' distributions study indicate that the SDAs tend to be smaller in extent. A further conclusion that can be drawn by this process is that all algorithms (except for the Conditional Thickening) segment areas of similar extent. Finally, regarding their spatial arrangement (estimated through the mean of the minimum inter-particle distance), it can be seen that SDAs tend to be more isolated especially on surfaces depicting a rapidly varying background structure. Such results are not so obvious for the other studied surfaces.

Further to assessing the degree of discrepancy in segmenting the decay areas in their exact location and extent we also conducted shape analysis of the detected decay patterns. Through this analysis we aim at investigating the shape features derived by each of the algorithms and at assessing how these converge to the shape features determined by the Ground Truth (GT).

4.2.3.   Comparative Study on the shape features

Through this shape analysis, we investigate the differences on the shape of segments detected by various algorithms. The shape features metrics recruited in this study are thoroughly discussed in chapter 3. In particular, the differences are evaluated in terms of patterns' compactness, eccentricity and roughness as well as their 1st and 2nd central moments. Compactness represents the roughness of an object's boundaries relative to its area. The smallest value of compactness is 12.56, which is for circle. As the circle deviates towards a more complicated shape, compactness becomes larger. Eccentricity, on the other hand provides information regarding the elongation of an object while moments reflect the distance of an object's boundaries to its centroid. Roughness metric (referred as (F_3-F_1) in the literature) has proven to be a quite effective measure to reflect roughness of micro-calcifications in digital mammograms. In the following diagrams we illustrate the distribution of the features' values. At each case we investigate whether the decay patterns segmented by different algorithms present similar shape features to segments illustrated in the GT. This analysis may provide a further tool for determining the algorithm that better approaches decay patterns' shape. In order to assess this potential, we investigate corrosion patterns detected on representative types of stone surfaces. More specifically the section is organized as follows: 4.2.3.1 sub-section examines the decay patterns' shape features detected on the unsheltered untreated fluting while 4.2.3.2 & 4.2.3.3 study the cases of the sheltered untreated fluting and the stone surface monitored via the digital camera respectively.

4.2.3.1.   Study of the Shape features concerning the Unsheltered Untreated Fluting

At first, we illustrate the distributions of shape features regarding the case of the unsheltered untreated fluting. Figures 4(a) through (e) depict the distribution of the afore-mentioned shape features.
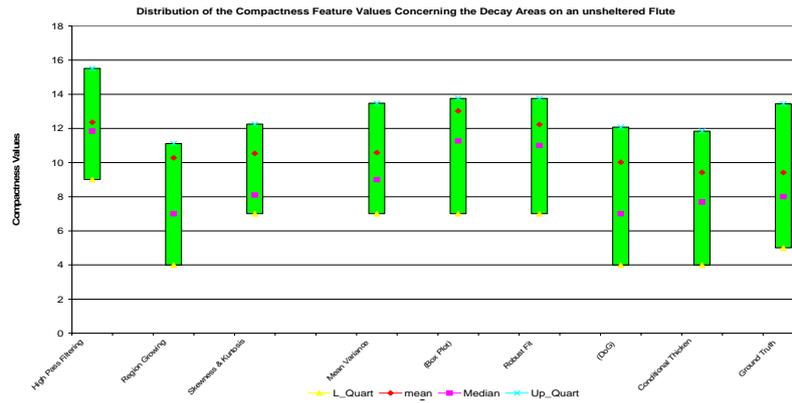
Figure 4(a): Distribution of compactness values regarding the deterioration patterns present on the unsheltered fluting.

By observing fig. 4(a), we can assess that there are some similarities between the algorithms' responses. The median and the mean values of each feature are considered as indicative measures of the feature values distribution. In particular, in fig 4(a) we can observe that the Conditional Thickening Algorithm and the GT tend to segment degraded areas of similar compactness values. According to this diagram, the DoG detector also derives a distribution of compactness values laid at close levels. Significant similarities in the distribution of compactness values are also reported between the three labeling algorithms.
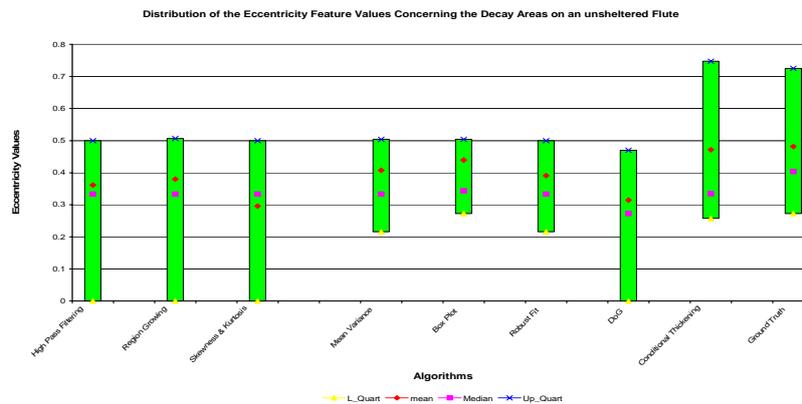
Figure 4(b): Distribution of eccentricity values regarding the deterioration patterns present on the unsheltered untreated fluting (fig. 2).

Figure 4(b) provides information on the eccentricity of the segmented black particles detected on the unsheltered untreated fluting. Here, we can assess again the similarity in the responses of the Conditional Thickening and the GT while similar agreement occurs between the three Labeling Algorithms. A noteworthy point is also the similarity in responses of the High Pass Filtering, Region Growing and Sub-Region Decomposition Algorithms.

Distribution of the First Moment Feature Values Concerning the Decay Areas on an unsheltered Flute
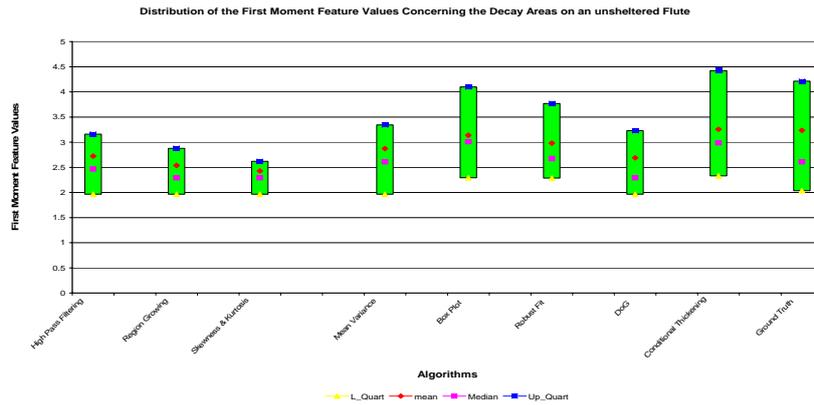
**Figure 4(c):** Distribution of 1st moment values regarding the deterioration patterns present on the unsheltered untreated fluting.

Figure 4(c) illustrates the distribution of 1st moment values of the decay areas detected by all the algorithms. This metric (1st moment) reflects the mean distance between the centroid of a label to its boundaries. As it can be observed, the distributions of the first moment values derived by the Conditional Thickening and by the GT are laid to close values. Similar assessments can be drawn regarding the 3 first presented algorithms namely (High-Pass Filtering, Region Growing Algorithm and Sub-Region Decomposition). At this point, we should make clear that the accordance between the shape features derived by the GT and those obtained by the Conditional Thickening do not necessarily mean that the latter performs better shape preservation as the GT was evaluated to provide reliable information on decay patterns topology and size and not their shape.

Figure 4(d) provides the distribution of values of the roughness metric regarding decay areas segmented on the untreated unsheltered fluting.

Distribution of the (F3-F1) Metric Values Representing roughness of the Decay Areas on an unsheltered Flute
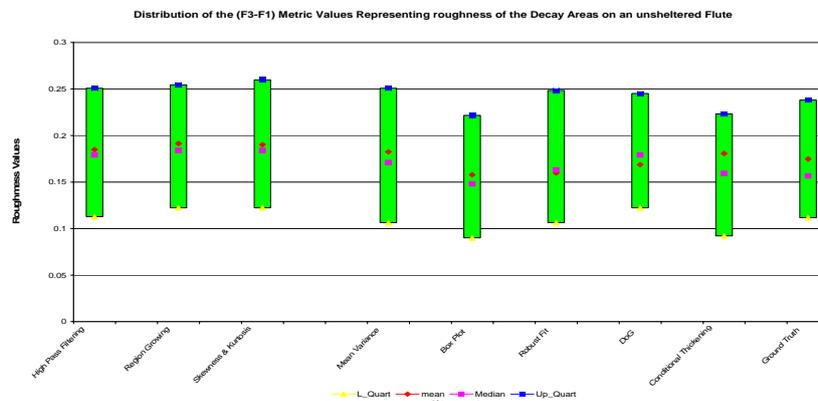
Figure 4(d): Distribution of roughness values (extracted via the F_3-F_1 metric) regarding the deterioration patterns present on the unsheltered fluting.

As it can be seen in fig. 4(d) all the tested algorithms detected decay areas of similar $(F_3$-$F_1)$ moment values. A deeper study of the presented results reveals that the corrosion patterns segmented by the Conditional Thickening Algorithm and those detected by the GT appear closer $(F_3$-$F_1)$ moment values. Similar assessment holds true for the case of the Sub-Region Decomposition, High Pass Filtering and Region Growing Algorithms. Figure 4(e) illustrates the distribution of the 2nd central moment feature values.
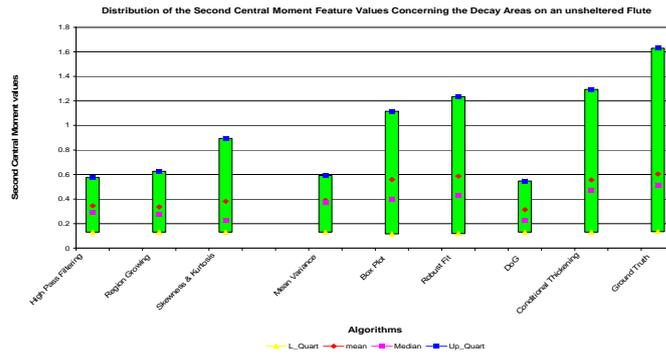
Figure 4(e): Distribution of 2nd central moment values regarding the deterioration patterns present on the unsheltered fluting.

Figure 4(e) also reveals that decay patterns segmented by different algorithms demonstrate similarities on their second central moment values. An initial observation indicates that the responses of the Conditional Thickening and the GT are in close agreement. The Sub-Region Decomposition, Region Growing and High Pass Filtering Algorithms also segment decay areas of similar 2nd central moments.

Summarizing the similarities observed between the distributions of the shape feature values, we can assess that the Conditional Thickening Algorithm seems to provide results closer to those derived by the GT. Moreover, the current study highlights the occurrence of similarities between the recruited algorithms such as among the 3 Labeling algorithms, as well as between High Pass Filtering, Region Growing and Sub-Region Decomposition. Such shape feature analysis along with the study of the topology and size of the segmented decay areas provide a test-bed for validating the potential of the algorithms in performing accurate detection. In the subsequent sections (4.1.3.2 and 4.1.3.3) similar studies are carried out and are associated to decay patterns encountered on the sheltered untreated fluting and on the surface monitored by the digital camera.

4.2.3.2. Study of the Shape features concerning the Sheltered Untreated Fluting

A study of the features values distributions concerning the case of the sheltered untreated fluting provides information associated to the algorithms' responses when handling in-homogenous background. The median and mean values are again considered to be indicative of the distributions. Thus decay patterns' shape features are summarized through figures 5(a) to 5(e). Initially, fig. 5(a) represents the distribution of compactness values.
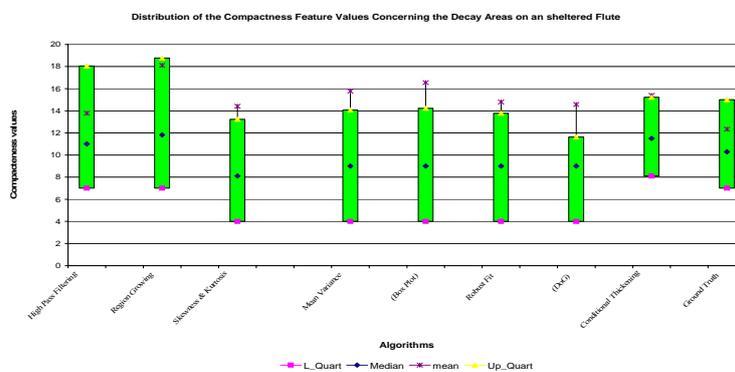

Figure 5(a): Distribution of compactness values regarding the deterioration patterns that prevail on the sheltered untreated fluting.

By observing fig. 5(a) we can see that the Conditional Thickening and the GT segment decay areas of similar compactness. Similar observations are also valid for the cases of the three labeling algorithms, the High Pass Filtering and the Region Growing. Finally, a similar agreement is drawn regarding the DoG detector and the Sub-Region Decomposition. The latter similarity between the Sub-Region Decomposition and the DoG can be associated with the results illustrated in Table 2. These verify that the Sub-Region Decomposition and the DoG tend to split decay areas into adjacent regions smaller in extent. The subsequent figures illustrate the distribution of other shapes features. Figure 5(b) investigates the eccentricity of decay patterns segmented on the stone specimen illustrated in fig. 1.
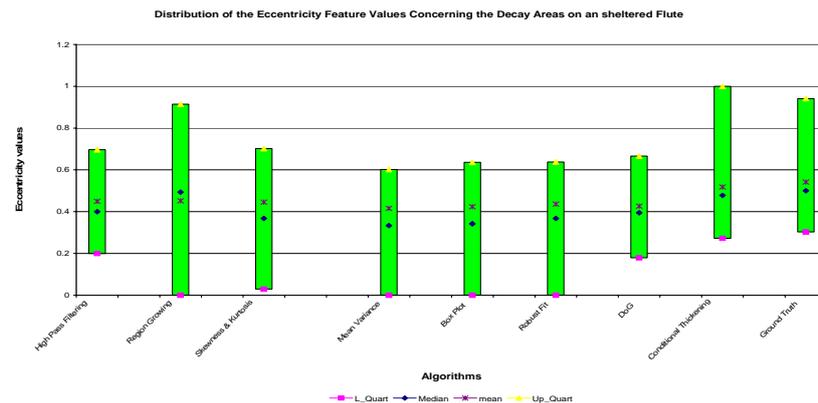

Figure 5(b): Distribution of eccentricity values regarding the deterioration patterns that prevail on the sheltered untreated fluting.

Figure 5(b) reports the distribution of the eccentricity values for the sheltered untreated fluting. Similar assessments to those presented above can be drawn by visually inspecting diagram 5(b). In particular, we can see that the GT and the Conditional Thickening seem to detect decay areas of similar eccentricity values. Moreover, the three initially presented algorithms, namely the Sub-Region Decomposition, High Pass Filtering and Region Growing also present a similar close agreement. A further comparison between the DoG detector and the Sub-Region Decomposition reveals accordance in the mean and median values while deviations are observed on the quartiles of the distributions.
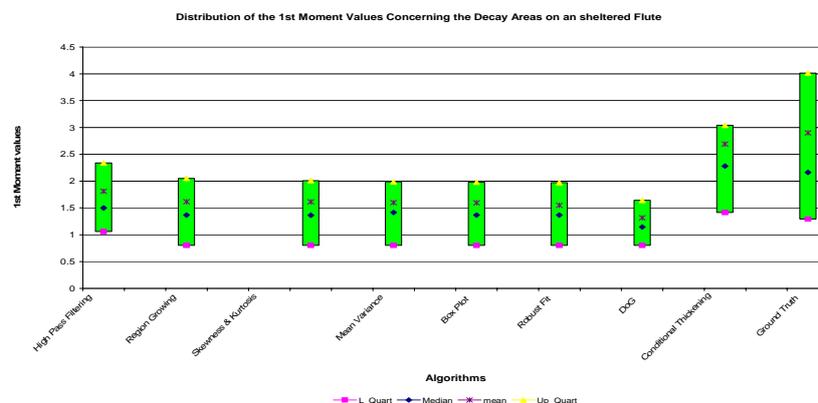
**Figure 5(c):** Distribution of 1st moment values regarding the deterioration patterns present on the sheltered fluting.

By studying the distributions of the 1st moment values depicted in 5(c), we can see that the Conditional Thickening algorithm again approaches better the GT. On the other hand, the Region Growing, Sub-Region Decomposition and the labeling Algorithms derive distributions laid to close

values. DoG deviates from the latter (it derives a distribution of 1st moment values laid to lower values) and this is expectable if we consider that this algorithm splits extended areas into sub-regions.

Figures 5(d) and (e) illustrate the distributions of the (F_3-F_1) metric and the 2nd central moments respectively.
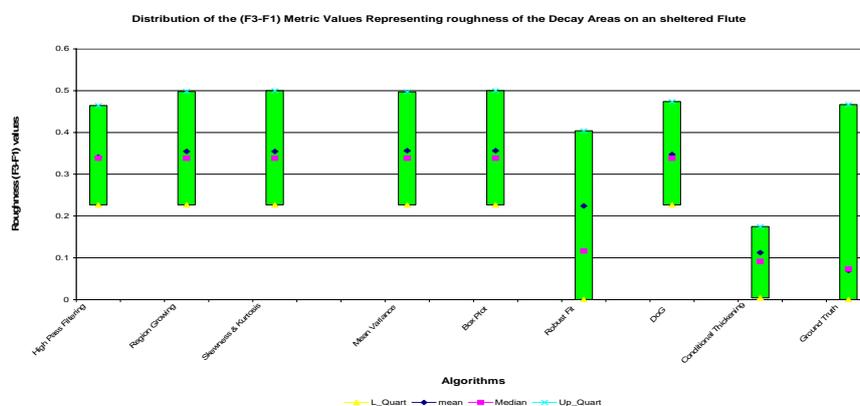

Figure 5(d): Distribution of roughness values (extracted via the (F_3-F_1) metric) regarding the deterioration patterns present on the sheltered fluting

Figure 5(d) reveals that the Conditional Thickening algorithm approximates the GT only in the median value of the distribution, while, the distribution of the GT is laid to greater values. According to this diagram, we can conclude that the Robust Fit Algorithm better matches to the GT. The other algorithms demonstrate a significant convergence in the distribution in the metric's values.
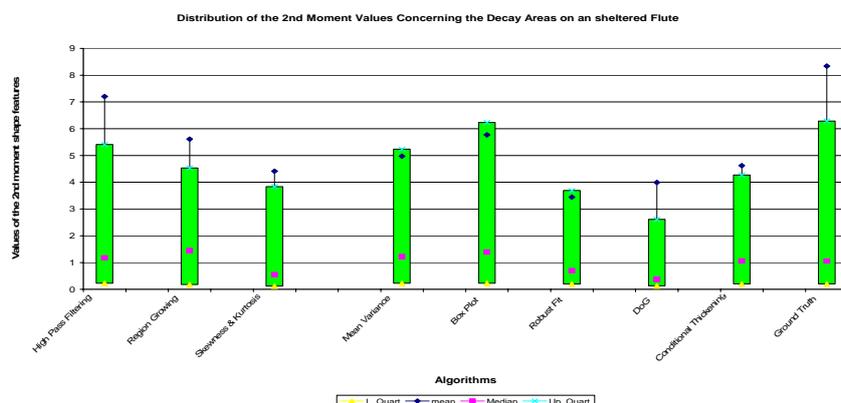
Figure 5(e): Distribution of 2nd central moment values regarding the deterioration patterns present on the unsheltered fluting.

The study of figure 5(e) also reveals some similarities in the algorithms' responses concerning the segmentation of areas with close second central moment values. Thus we can assess that the distribution of the second central moment values of the Conditional Thickening Algorithm and the GT demonstrate an agreement regarding the median level of the distribution. A brief observation of the presented results shows large deviations in the metric's mean values between the algorithms. This is mainly due to by the fact that the areas segmented by the same algorithm tend to demonstrate second central moment values with large deviations.

In an effort to summarize the performance of the implemented algorithms regarding the extraction of shape features, we could say that some algorithms such as the Conditional Thickening and the GT segment decay areas of similar shape characteristics. Such similarities were also assessed among the Labeling algorithms. A further comparison between the data illustrated fig. 4 and those depicted in figs. 5

indicates that in the case of the sheltered fluting the Sub-Region Decomposition response approximates better the DoG than in the case of the unsheltered fluting. This is mainly associated to the texture features reflected in the image and to the algorithm parameters employed through the detection process (such as the adaptation of the window's size as well as the size of the kernels of the Gaussian filters). Finally, a noteworthy observation that will be more thoroughly discussed later in this chapter concerns how decay patterns with close shape features values are arranged onto the stone surface. According to our study, large decay patterns with similar shape feature values tend to form neighborhoods in the image. However, the smaller decay patterns demonstrate more random shape features. Thus a more dissect investigation is required in order to interpret this.

The following subsection discusses the patterns' shape features measured on the decay areas segmented on the stone material screened by the digital camera.

4.1.3.3. Study of the shape features concerning the surface monitored via the Digital Camera

Figures 6(a) through (e) depict the distribution of values of the shape features regarding the decay areas segmented on the stone surface monitored by the digital camera.
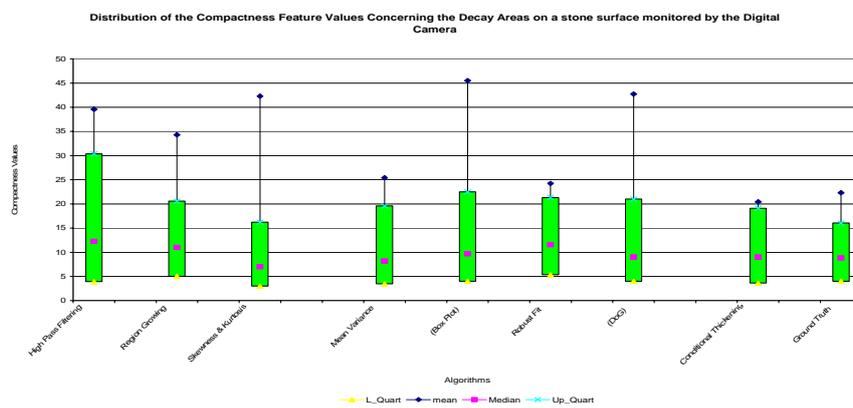
Figure 6(a): Distribution of the compactness values regarding the deterioration patterns present on the stone surface monitored by the digital camera.

By observing fig. 6(a), we can assess that similarities occur on the median levels of the distributions. The median and the mean values of each feature metric are considered as indicative measures of its distribution. In the discussed illustration we can see that the mean values are greater than the upper quartiles. This is caused due to the large standard deviation of the distributions. Robust-Fit and Conditional Thickening algorithms demonstrate the lower standard deviation values. Thus, due to the high variation observed on each distribution the extraction of similarities becomes rather dicey.

Figure 6(b) illustrates the eccentricity values of the labels detected by the implemented algorithms.
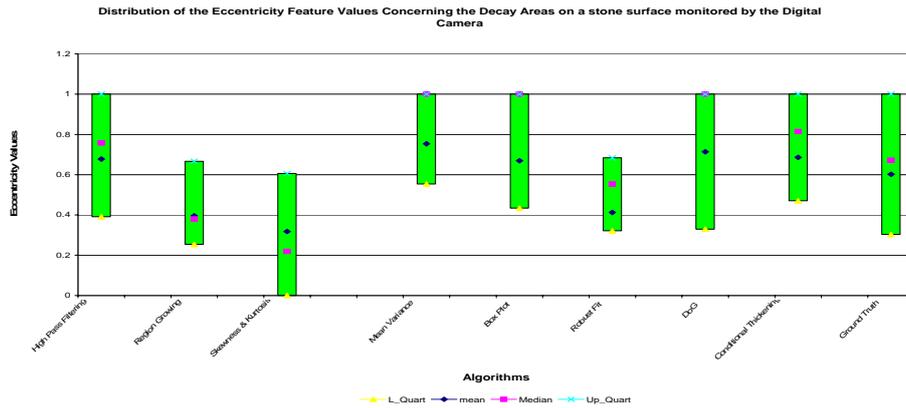
Figure 6(b): Distribution of the eccentricity values regarding the deterioration patterns present on the stone specimen monitored by the digital camera.

By observing figure 6(b), we can assess that the distributions of the eccentricity values demonstrate significant deviations. In the diagram we can see that the results derived by the Conditional Thickening deviate from those obtained by the GT. In contrast, the High Pass Filtering algorithm better approaches the GT's distribution. Such an assessment should be seen in combination with the performance curves depicted in fig. 10. According to the data reported in this diagram, the High Pass Filtering Algorithm seems to perform better. Figure 6(c), (d) and (e) represent the distribution of values regarding the 1st moment, the F_3-F_1 metric and the 2nd central moment.

Figure 6(c) represents the 1st moment values of decay patterns segmented on the digital camera image.
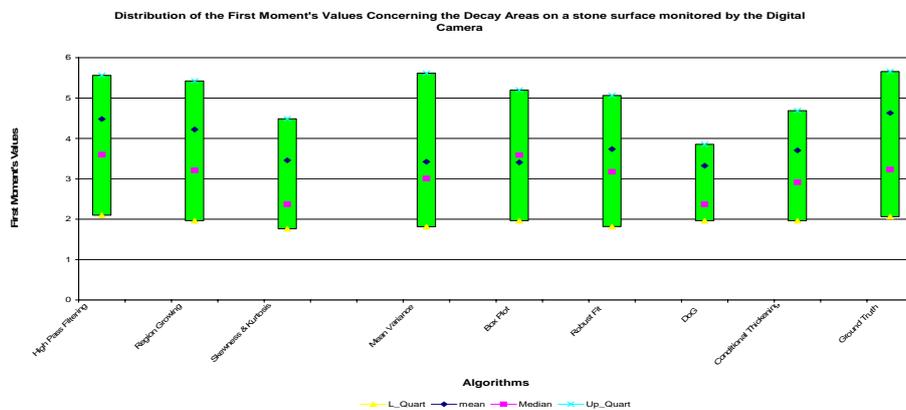

Figure 6(c): Distribution of the 1st Moment values regarding the deterioration patterns present on the stone surface monitored by the digital camera.

A brief observation of fig. 6(c) reveals that the majority of the implemented algorithms tend to derive distributions of 1st moment values that are laid to adjacent levels. This stems from the fact that according to Table 6 all the implemented algorithms (except for the Sub-Region Decomposition and the DoG) tend to segment decay areas smaller in extent than all the other algorithms. The latter observation is also reflected in fig. 5(c), where we can see that the two previously mentioned algorithms appear discrepancies from the other algorithms. Moreover, we can see that the High Pass Filtering Algorithm and the Region Growing approach better the GT. Furthermore, similarities between the Labeling

algorithms can also be seen in fig. 6(c). Through fig. 6(d) we investigate the distributions of decay patterns' 2nd central moment.
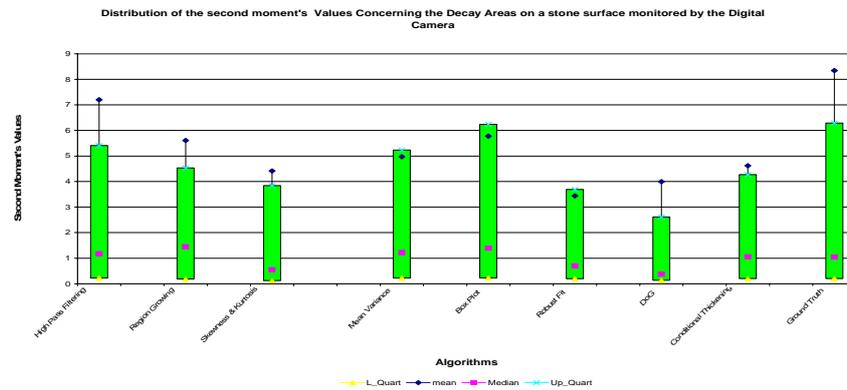

Figure 6(d): Distribution of the 2nd moment values regarding the deterioration patterns present on the stone surface monitored by the digital camera.

Figure 6(d) reveals that, the DoG, Sub-Region Decomposition and Robust Fit Algorithms do not agree with the others on the values of the 2nd segments' central moment. It can be seen that in some cases the mean values are greater than the upper quartiles, which reflects the high standard deviation of the distributions. Thus, an effort to report similarities among the studied algorithms is rather dicey.
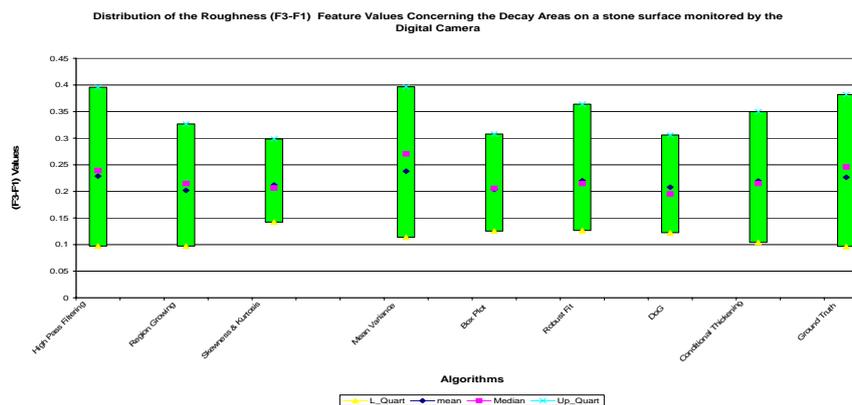
Figure 6(e): Distribution of roughness values (extracted via the F_3-F_1 metric) regarding the deterioration patterns segmented on the digital camera image.

By observing fig. 6(e) it becomes clear that the High Pass filtering algorithm better converges to the GT in the distribution of the (F_3-F_1) metric while the Conditional Thickening and the Region Growing Algorithms follow.

Further to evaluating how the recruited algorithms converge in determining the topology and the extent of the segmented areas, we also performed a more dissect study concerning algorithms' sensitivity and specificity or their precision and recall levels. The latter provide information that makes it possible to objectively and reliably compare the performance of range image segmentation algorithms while allowing informed experimental feedback for the design of more adaptive versions of the implemented segmentation algorithms.

4.3. Comparison in terms of Performance Curves

In an effort to investigate the potential of the implemented algorithmic schemes in detecting decay areas, we present diagrams depicting the performance of each algorithm versus the others. The algorithms' response is evaluated through assessing their capability at detecting the degraded areas with their correct topology and sizes. Thus, the ROC and the Precision-Recall curves are derived as robust measures of the algorithmic performances. As it was mentioned in the previous section the surfaces used to extract the Ground Truth (GT) were selected by the experts to depict representative instances of corrosion damage. Initially, the case of the fluting located on unsheltered areas is investigated.

4.3.1. Algorithms Performance Regarding the Unsheltered Untreated Surface
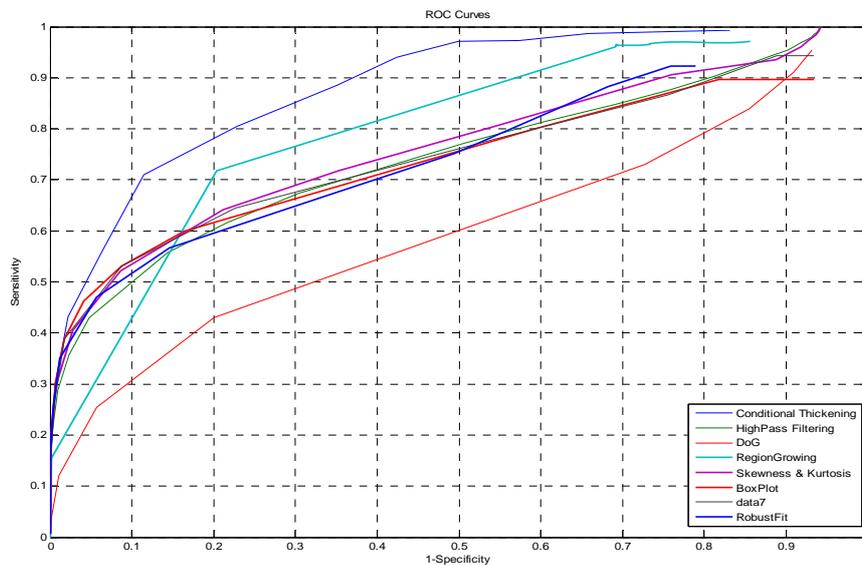
Figure 7: ROC curves depicting the performance of the implemented algorithms in the case of the unsheltered untreated fluting (illustrated in fig.2).

In fig (7) it can be seen that the Conditional Thickening Algorithm demonstrates better performance in detecting decay patterns at their real extent while the Region Growing Algorithm follows. It should also become clear that the approach of determining algorithms' specificity and sensitivity is more focused on accurately detecting their size. Thus, it is expected that algorithmic approaches that split the decayed areas such as the Labeling Algorithms, Sub-Region Decomposition and DoG will demonstrate worse performance. By observing fig. 7 we can also see that the labeling algorithms (Mean Variance, Box Plot, Robust Fit) tend to perform better than the High pass Filtering Algorithm for low values of sensitivity. This means that the Labeling Algorithms introduce less False Positive (FPs) and False Negative (FNs) areas when relaxed thresholds are applied. The preceded observation reflects the potential of local processing techniques in accurately segmenting decay spots in in-homogeneous backgrounds. Another assessment that can also be drawn, by the above figure, is that the DoG detector appears to be less efficient in preserving decay areas' actual size/and or shape. This is exemplified by considering that the latter algorithm splits the detected areas and, thus, segments many reduced in size spots. Concerning the Region Growing Algorithm, a remarkable point is that for specificity values>0.85

(1-specificity<0.15) it demonstrates poorer performance than the other algorithms while for specificity values<0.85 its performance becomes better. At this point, we should make clear that the ROC curves are based on the fraction of the pixels that have been accurately detected as degraded areas. In other words, the derived results provide information on the effectiveness of the studied algorithms in detecting the size of the degraded regions and not necessarily their topology. In conclusion, an algorithm with a ROC curve located under the others is not necessarily inefficient as it may approach accurately the locations of decay areas.

As it was discussed previously, the algorithms' performance is evaluated through the performance curves derived by studying images of various background structures. Such a study provides reliable information regarding the response of the algorithms when applied to images representing different structure discontinuities. In the following figure (fig.8), we illustrate the ROC curves depicting the performance of the implemented algorithms in the case of the sheltered untreated fluting.

### 4.3.2. Algorithms' Performance Regarding the Sheltered Untreated Surface
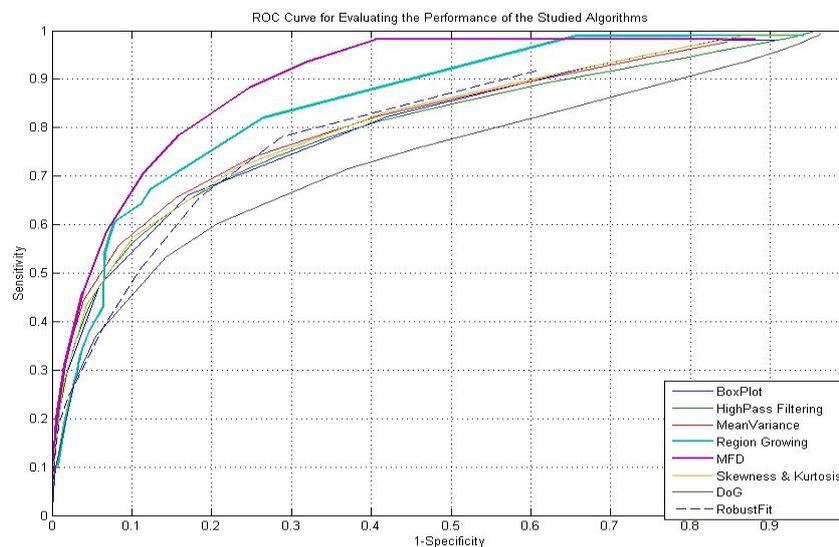
Figure 8: ROC curves depicting the performance of the implemented algorithms in the case of the sheltered untreated fluting (illustrated in fig.1).

An initial observation on the curves appearing in fig.8, reveals that the DoG algorithm demonstrates a better performance than fig. 7, which in turn may reveal a better capability of the algorithmic approach in segmenting decay areas on in-homogeneous backgrounds. Regarding the performance of the other algorithms it can be said that the Conditional Thickening and the Region Growing Algorithms perform better (than the others) for specificity levels <0.35 (1-specificity>0.65). Regarding the Region Growing Algorithm it can be seen that for high specificity levels (1-specificity<0.07) the algorithm tends to be less efficient than the others, while its performance is improved as it proceeds to lower specificity values. The labeling algorithms along with the Sub-Region Decomposition demonstrate similar performance and they seem to provide more reliable results when applied to in-homogenous background.

In an effort to assess the algorithms' potential when applied to surfaces representing various structures, we observe that the Conditional Thickening as well as the Region Growing presents a better performance in the case of the sheltered untreated fluting (fig. 8). Regarding the former algorithm, this mainly occurs because the decay patterns' topology is associated to areas where texture irregularities prevail (these areas present larger contrast to the background). Thus, they can be determined more easily through the morphological operators (we observe peaks of intensity and heavier tails on the histogram of the local background). As for the other algorithmic scheme (Region Growing) we should consider the way that this algorithm operates. More specifically, this algorithm selects seed pixels and subsequently checks for inclusion the 8-nhood of the seeds according to a condition that associates the intensity levels of the pixels included to the label, and a tolerance parameter t. While t falls in the interval [0,α], for each t we extract a vector of features concerning the label. Subsequently the optimum value of t is chosen to be the one that induce the least distance between consequent feature vectors. According to the authors' judgment, in the case of a rapidly varying texture t value is more reliable because there is usually only one t that introduces the least distance. In contrast in the case of smoother background images, there may be many t values that minimize the distance of the feature vectors and thus the first t chosen may not always correspond to the global optimum. Another important point, which should also highlighted, is the response of the labeling algorithms. We can assess that for small values of specificity, the Labeling Algorithms perform better on the sheltered untreated flutings (fig. 8) compared to the unsheltered untreated flutings (fig. 7). In other words for larger false positive (FP) rates the algorithms perform better in the surfaces presenting inhomogeneous background structures. This is expectable because texture in-homogeneities induce outliers on the histogram of the studied surfaces. Thus, even relaxed thresholds segment susceptible areas. In contrast, when operating on images of smoother background, the adjustment of low thresholds simply causes the segmentation of large compact areas that do not correspond to susceptible regions.

As opposed to the ROC curves, which study the algorithms' ability to segment decay areas at their real extent, we also examine the algorithms' potential to accurately detect corroded areas' topology (through extracting Precision Recall curves). In the previous chapter, we defined true positive (TP) labels being the labels that were detected by both the GT and the corresponding algorithm. Similarly the FNs represent areas segmented by the GT but not by the studied algorithm while FPs correspond to labels detected by the studied algorithm and not segmented by the GT. This analysis can be used as complementary to the analysis performed through ROC curves, because it investigates the capability of each algorithmic scheme in effectively determining the location of decay areas. Figure 5 shows the trade-off between an algorithm's precision and recall.
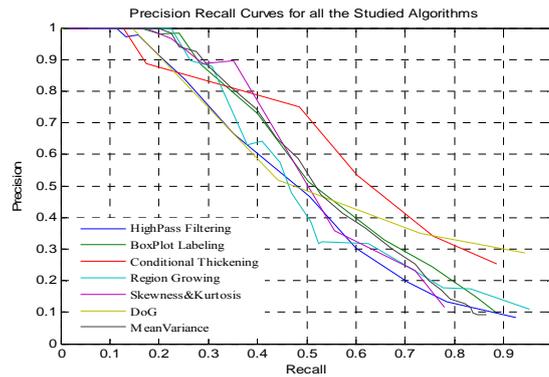
**Figure 9:** Precision Recall curves depicting the performance of the implemented algorithms in the case of the sheltered untreated fluting (illustrated in fig.1).

By observing the Precision Recall curves illustrated in fig. 9, it becomes evident that for lower levels of recall, the Labeling algorithms (Mean Variance and Box Plot) demonstrate a better performance in determining the exact topology of decay areas while the Sub-Region Decomposition, Region Growing and Conditional Thickening follow. The High Pass Filtering Algorithm seems to be less effective especially for higher values of recall. A noteworthy point is the Precision Recall curve of the DoG algorithm. As it can be seen, the DoG algorithm tends to be more effective for higher values of recall. This can be verified from the diagrams, where we observe that as recall levels proceed to higher values the precision of DoG is improved. Regarding the other algorithms, the Conditional Thickening demonstrates better performance for high recall values (>0.4). Furthermore, an important point is also the performance of some algorithms such as the Sub-Region Decomposition. It can be observed that its potential to reduce the false positive rate is quite effective. This observation is closely related to the potential of the local processing approach to segment decay spots while considering the irregularities due to noise and the stone structure. Finally, it is obvious that the Region Growing and the Labeling Algorithms demonstrate better performance than the High Pass Filtering supporting, in this way, our initial claim that the detection procedures should be based on the local background

The objective of presenting the Precision Recall curves is to assess the robust points of each algorithm in the segmentation procedure. Thus, we consider in parallel the performance of each algorithmic scheme illustrated in the ROC (fig 8.) and the Precision Recall curves (fig. 9). An initial assessment regards the performance of the Sub-Region Decomposition Algorithm. According to fig. 9 it performs better than the other algorithms for recall values<0.4; in contrast fig. 8 presents the Sub-Region Decomposition to be less efficient in segmenting corroded areas at their real extent (as it is defined by the GT). This observation indicates that the algorithmic scheme accurately detects the exact locations of decay patterns while distorting their extent by splitting or segmenting them at smaller sizes. The Region Growing was presented in fig.9 to derive quite good performance for low recall values. Its performance in determining the real extent of decay patterns was also estimated to be effective (fig. 8). We can assess thus, that the algorithm provides reliable information for both the size and the topology of degraded areas (especially at low thresholds). A drawback of the algorithm according to the precision recall curves is that it introduces more FPs when adjusting relaxed thresholds. The Conditional Thickening Algorithm

is revealed to be the most efficient for high recall values. In other words, it attains to minimize the false positive and false negative rates for large thresholds. Its performance though is worse than the Region Growing's and Sub-Region Decomposition for low thresholds. This behavior is mainly associated with the fact that the Conditional Thickening fuses the areas segmented by the DoG and the Morphological Detectors and thus the areas segmented for strict thresholds are generally small in extent and the fusion leads to the determination of null areas. Finally, the High Pass Filtering Algorithm demonstrates the worst performance in approaching the decay patterns' topology while it presents a quite better potential in determining the size of corroded areas. In an effort to exploit the algorithm with the best performance at extracting both the correct topology and the size of the segmented areas, the experts should pose the criteria. This means that they should state whether their priority is to define accurately the topology while distorting the size, or the opposite. A further question that should also be answered (by the experts) is whether a larger false positive rate is more preferable than a large false negative rate or the opposite. In any case, according to the authors' judgment, the Conditional Thickening along with the Region Growing Algorithms segment areas that illustrate most effectively the decay state of the surface.

4.3.3.   Algorithms Performance regarding the image obtained by the digital camera

Further to validating the algorithms' efficiency in segmenting decay areas on FOM images, we also investigate their potential in determining corrosion effects on surfaces screened by other monitoring modalities (digital camera system). Such responses are also evaluated through the ROC curves. Figure 10 illustrates the performance curves. We can observe that the High Pass Filtering Algorithm demonstrates a greater efficiency in detecting decay patterns. This assessment is expected if by considering the low-resolution levels provided by the digital camera, which results in an inability to monitor details of the stone structure. Thus, a global processing algorithm seems to be efficient, as it does not emphasize on information related to dynamically varying local features. The Conditional Thickening and the Region Growing algorithms demonstrate worse performance than the High-Pass Filtering algorithm. A significant observation that can also be drawn is that all the algorithms' performances tend to converge for specificity levels<0.5 (1-specificity>0.5).
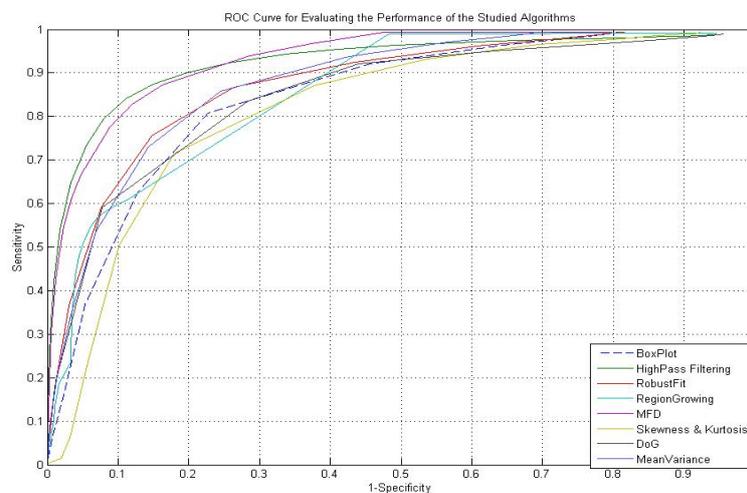
Figure 10: ROC curves depicting the performance of the implemented algorithms in the case of the stone material monitored via the digital camera (illustrated in fig.3).

4.3.4. Comparative studies on the Performance Curves

Through the diagrams illustrated below we study the potential of each individual algorithm in segmenting deterioration effects on various background stone structures. At first, figs. 11(a) and (b) study the case of the High Pass Filtering and the Region Growing algorithms respectively when applied to the three studied GTs.
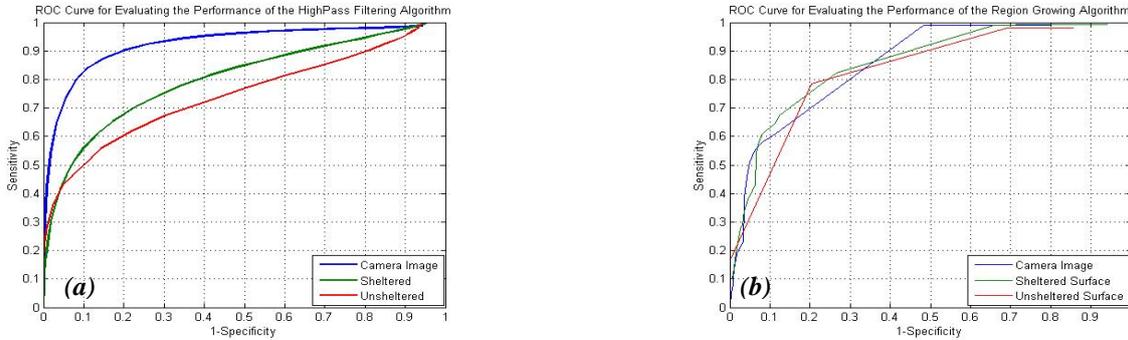
Figure 11: ROC curves illustrating the performance of (a) the high pass filtering algorithm and (b) the Region Growing Algorithms

Figure 11(a), reveals that the High-Pass Filtering Algorithm is more effective when applied on the digital camera image while it seems to be less accurate in performing segmentation on smooth background. As for the Region Growing Algorithm, fig.11 (b) shows that for specificity values>0.65 the Region Growing Algorithm is more efficient when processing the FOM images. This is expectable as the algorithm performs local processing and thus it considers the dynamic variations of gray levels in a pixel's neighborhood. The macroscopical image (digital camera image), does not provide information associated with such local texture variations and thus local processing is not advantageous. A further observation that can also be extracted is that the Region Growing Algorithm seems to be more efficient when it is applied on sheltered surfaces than when applied on un-sheltered (sheltered surfaces demonstrate a more inhomogeneous background). The latter was thoroughly discussed and in the previous subsection

Performance evaluation proceeds by investigating the responses of the labeling algorithms (fig. 12) and the Conditional Thickening and DoG Algorithms (fig.13) when applied on various surfaces.
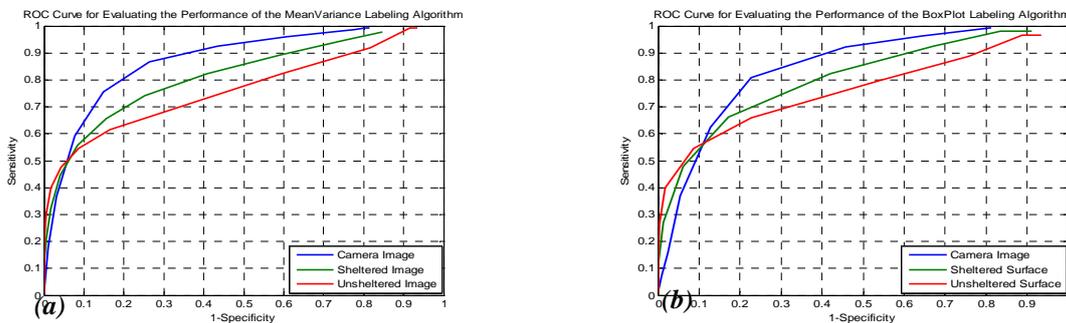
Figure 12: ROC curves illustrating the performance of (a) the Mean Variance Labeling and (b) the Box Plot Labeling Algorithms.

Figure 12(a) and (b) reveal that the discussed algorithms seem to have a better performance when employed to detect decay effects on digital camera images. This observation holds true for specificity levels <90%.
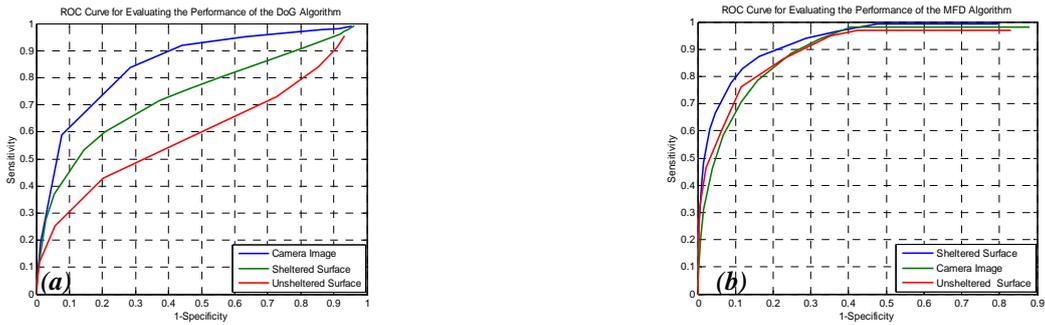
Figure 13: ROC curves illustrating the performance of (a) the DoG detector (b) the Conditional Thickening Algorithms.

Figures 13(a) and (b) reveal that the Conditional Thickening is more reliable than the DoG algorithm. Moreover concerning to the DoG's ROC curves, it can be seen that the algorithm's application on the digital camera image appears greater efficiency at detecting deterioration patterns. A similar comparison for the case of the Conditional Thickening algorithm would be rather dicey, as the ROC curves do not demonstrate considerable discrepancies.

At last, in fig. 14, we present the ROC curves of the Sub-Region Decomposition Algorithm.
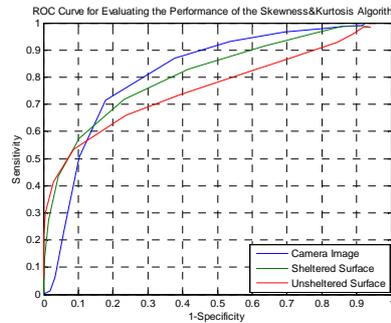
Figure 14: ROC curves illustrating the performance of the Sub-Region Decomposition Algorithm.

It is obvious that the Sub-Region Decomposition algorithm performs better on FOM images for large values of specificity. A further remark derived from the study of figure 14 is that this approach demonstrates better performance when applied to the sheltered surface. This is also exemplified by the fact that the algorithm selects the susceptible blocks (where it performs the segmentation) according to the asymmetry of their histogram. Thus areas demonstrating robust texture irregularities are easier determined.

In an effort to summarize the results illustrated through figures 11-14, we can observe that the vast majority of the recruited algorithms tend to be more efficient when performing segmentation on the digital camera image. This is explained by considering that the digital camera provides low resolution and consequently decay effects are mainly reflected to color alterations of the stone material. Thus, degraded areas are depicted as dark regions on a smooth background, which makes their locations of prevalence almost obvious. The role of the expert is critical at determining the algorithm with the best performance. This person poses the criteria for defining the algorithm with the optimum response. More specifically (s)he should determine the individual characteristics of the most appropriate segmentation scheme. The criteria are mainly associated to the algorithm's potential in providing robust estimation of both their topology and their extent

4.4. Comparison of Population Differences for Various Structural and Cleaning Conditions

As it was discussed earlier in this thesis, one of our aims is to examine the effects of cleaning interventions on stone materials. Through the statistical tests described in this section, we attempt to elicit the particular effects caused by the cleaning methods. More specifically, we study their ability to eliminate crusts' thickness while also investigating reductions on the size of segmented decay areas. Further to estimating the efficiency of the treatments, we also examine (by statistical means) the variations in corrosion effects prevailing surfaces exposed to different weathering conditions. The comparison of the results derived through the implemented segmentation algorithms is also an objective of the current thesis. Through the following two sections, we compare the responses of the three algorithms that present better responses according to the ROC and the Precision Recall curves. These algorithms are namely: the Conditional Thickening, Sub-Region Decomposition and Region Growing. Throughout this comparison we attempt to estimate and explain the observed discrepancies.

Throughout section 4.4 we evaluate the cleaning and structural effects on the studied stone materials. More specifically, after the segmentation of decay areas (with the aid of the 3 segmentation algorithms) some of their features such as their extent or their relative intensities over the background are measured. Such feature values are organized to construct distributions. Subsequently, the tests of statistical significance are employed aiming at assessing whether decay patterns segmented on surfaces of various structural and cleaning states is characterized by feature values that could belong to the same population. We initially introduce T-tests to examine the way that the cleaning interventions are reflected on the relative intensities of corroded areas over the background. According to the experts', the intensity levels encountered at locations of decay particles prevalence is closely associated to the crusts' thickness on the corresponding areas.

4.4.1. T-Tests for Intensity Distributions

At first, we group the images according to their conditions of exposure or their cleaning state and subsequently these groups are compared towards other surfaces to assess the effects of location or cleaning. Besides the estimation of each method's cleaning effects, we are further focused towards assessing differences in the results derived by the implemented algorithms. More specifically, the algorithms tested through this procedure correspond to those exhibiting the best performance in defining either the topology or the extent of decay patterns. The studied algorithms were selected by considering their responses as illustrated through the ROC and Precision Recall curves. Thus, the Region Growing, Sub-Region Decomposition and the Conditional Thickening are further considered here.

The t-test is employed to estimate whether cleaning/and or exposure conditions induce significant alterations on relative intensities of decay areas over the background. Such alterations reflect corresponding alterations on the crusts' thickness according to the chemists. The latter interpretation is based on the fact that thicker crusts tend to absorb greater amounts of impacting luminance, which in turn results in depicting degraded regions as darker (lower intensity values). At this point we should state that the alterations of the intensity levels on regions of decay patterns do not objectively estimate the

thickness of black crusts. A more accurate metric is the relative intensity of locations of degraded areas over the background. In this way the discrepancies induced due to luminance differences are diminished. Thus, prior to the application of the t-tests the set of images are submitted to intensity normalization in order to eliminate the effects of different luminance conditions. At this point we should clarify that before the selection of the test we have verified that the sample means of the examined populations are normally distributed. Table 8 summarizes the results of the T-test when performed to the studied surfaces.

Table 4-8: Comparative study on the significance of intensity values increase[7][8].

	Cond.Thickening	Sub-Region Decomposition	Region Growing
1) **Sheltered Flutings (Ds) (vs) Sheltered Flutings (Diagnosis)**	Df= 34 t= 25.764 Crit t (1-tail)= 1.691	Df= 34 t= 22.187 Crit t (1-tail)= 1.691	Df= 34 t=24.478 Cr.t(1-tail)= 1.691
2) **Sheltered Flutings (WMB) (vs) Sheltered Flutings (Diagnosis)**	Df= 33 t= 62.410 Crit t (1-tail)= 1.692	Df= 32 T =63.829 Crit t (1-tail)= 1.694	Df= 32 t=59.279 Crit t (1-tail)= 1.694
3) **Sheltered Flutings (BP) (vs) Sheltered Flutings (Diagnosis)**	Df= 27 t= 33.899 Crit t (1-tail)= 1.703	Df= 28 t=40.619 Crit t (1-tail)= 1.701	Df= 28 t=36.787 Crit t (1-tail)= 1.701
4) **Sheltered Reedings (Ds) (vs) Sheltered Reedings (Diagnosis)**	Df= 9 t= 12.591 Crit t (1-tail)= 1.833	Df= 10 t= 14.021 Crit t (1-tail)= 1.812	Df= 10 t=13.712 Crit t (1-tail)= 1.812
5) **Sheltered Reedings (BP) (vs) Sheltered Reedings (Diagnosis)**	Df= 8 t= 12.716 Crit t (1-tail)= 1.860	Df= 10 t= 18.321 Crit t (1-tail)= 1.812	Df= 10 t= 14.436 Crit t (1-tail)= 1.812
6) **Sheltered Reedings (Diag)(vs) Sheltered Flutings (Diag)**	Df= 28 t= 13.443 Crit t (1-tail)= 1.701	Df= 22 t= 13.388 Crit t (1-tail)= 1.717	Df= 28 t= 10.148 Crit t (1-tail)= 1.701
7) **Unsheltered Flutings (Diag) (vs) Sheltered Flutings (Diag)**	Df= 34 t= 47.960 Crit t (1-tail)= 1.691	Df= 34 t= 31.016 Crit t (1-tail)= 1.691	Df= 34 t=41.429 Crit t (1-tail)= 1.691
8) **Unsheltered Flutings (DS) (vs) Unsheltered Flutings (Diagnosis)**	Df= 22 t= 7.749 Crit t (1-tail)= 1.717	Df= 22 t= 12.089 Crit t (1-tail)= 1.717	Df= 22 t= 10.765 Crit t (1-tail)= 1.717
9) **Unsheltered Flutings (Diagnosis) (vs) Sheltered Reedings (Diagnosis)**	Df= 16 t= 16.347 Crit t (1-tail)= 1.746	Df= 16 t= 18.940 Crit t (1-tail)= 1.746	Df= 16 t= 16.487 Crit t (1-tail)= 1.746
10) **Unsheltered Reedings (Diag) (vs) Unsheltered Flutings (Diag)**	Df= 22 t= 4.575 Crit t (1-tail)= 1.717	Df= 22 t= 6.799 Crit t (1-tail)= 1.717	Df= 22 t= 4.090 Crit t (1-tail)= 1.717
11) **Unsheltered Reedings (Ds) (vs) Unsheltered Reedings (Diagnosis)**	Df= 9 t= 7.576 Crit t (1-tail)= 1.833	Df= 10 t= 8.401 Crit t (1-tail)= 1.812	Df= 10 t= 7.004 Crit t (1-tail)= 1.812
12) **Unsheltered Reedings (WMB) (vs) Unsheltered Reedings (Diagnosis)**	Df= 9 t= 8.42 Crit t (1-tail)= 1.833	Df= 10 t= 11.63 Crit t (1-tail)= 1.812	Df= 10 t= 7.770 Crit t (1-tail)= 1.812
13) **Unsheltered Reedings (Diagnosis) (vs) Sheltered Reedings (Diagnosis)**	Df= 16 t= 25.223 Crit t (1-tail)= 1.746	Df= 16 t=30.958 Crit t (1-tail)= 1.746	Df= 16 t= 15.972 Crit t (1-tail)= 1.746

4.4.1.1. Analysis on the Results derived by the T-tests

1. In the first case discussed in Table 8 we investigate whether the employment of a cleaning intervention based on the application of anionic resin on sheltered untreated flutings results in reducing the darkness levels of decay areas. The null and the alternative hypotheses are stated below:

---

[7] According to the statisticians if t>critical value then the probability of the observed difference having occurred by chance is almost *negligible*.
$P<0.05$

[8] The rows highlighted by red fonts correspond to cases at which either a significant difference does not occur or the difference is marginally significant.

H$_0$: The intensity values of the black particles encountered on cleaned sheltered flutings (DS 60min) are laid to lower values than the intensity levels on the corresponding areas encountered on untreated surfaces.

H$_A$: The intensity values of the black particles encountered at sheltered flutings (DS 60min cleaned) are laid to higher values than the corresponding intensity levels encountered at degraded untreated areas.

In order to assess whether we should reject or accept the null hypothesis we perform one-tail t-test. The value of "t" at the 5% level of probability for 34 degrees of freedom (DF), taken from the t-table is 1.691. Thus if the mean level of a population A is significantly larger than the mean level of a population B then the t-value calculated when we compare A towards B will be larger than 1.691 (for 34 DF). The value of "t'" obtained in the case of the Conditional Thickening is 25.764 and is much greater than the value of **'t'** at P=0.05. The probability of obtaining a value of **'t'=** 25.764 by chance, when H$_0$ is truth, is <<0.05 and therefore, we reject H$_0$ to conclude that the decay patterns detected on sheltered untreated flutings tend to have lower intensity levels than the decay patterns segmented on the same areas after the treatment by the DS method. Similar results are also drawn from the responses of the Sub-Region Decomposition and the Region Growing algorithms.

2. At the second experiment we investigate the intensity alterations induced by cleaning sheltered untreated flutings with the WMB method. Throughout this process the hypotheses stated are:

H$_0$: The intensity values of the black particles detected on sheltered flutings, treated by the WMB method, are laid to lower values than the intensity levels encountered on the corresponding particles occurring on untreated sheltered flutings.

H$_A$: The intensity values of the black particles encountered on cleaned sheltered flutings (WMB) are laid to higher values than the intensity levels of the corresponding particles encountered on degraded untreated flutings.

Regarding the Conditional Thickening Algorithm, the value of "t" at the 5% level of probability, for 36 DF is 1.688. The value of "t'" obtained in the case of the Conditional Thickening is 62.410, which is much larger than the **'critical t-value'** at 0.05 probability level. The probability of obtaining a value of **'t'** equal to 62.410 by chance, when H$_0$ is truth, is much lower than 5% (<<0.0005) and therefore, we reject H$_0$ and conclude that the decay patterns detected on sheltered untreated areas are indeed characterized by lower intensity levels than the decay patterns segmented on areas cleaned by the WMB method. For the case of the Sub-Region Decomposition and the Region Growing Algorithms the conclusions are similar, as the critical values of "t" for the one-tail t-test are much smaller than the corresponding critical t-values. Thus, H$_0$ is rejected in favor of H$_A$ and we assess that cleaning with the (WMB) method results in eliminating the crust's thickness.

3. At the 3rd experiment we investigate intensity alterations induced by cleaning the sheltered untreated flutings by the BP method. Throughout this process the hypotheses tested are:

H₀: The intensity values of the black particles encountered on cleaned sheltered flutings (BP) are laid to lower values than the corresponding intensity levels encountered on decay areas of the untreated sheltered flutings.

H_A: The intensity values on black particles occurring on sheltered flutings (BP cleaned) are laid to higher values than the corresponding intensity levels encountered on decay regions detected on untreated flutings.

For the Conditional Thickening Algorithm, the "**critical value of t**" at the 5% level of probability for 27 DF is 1.688. The value of 't' obtained in the case of the Conditional Thickening was 33.899 and is much greater than the **'critical value of t'** at P=0.05. The probability of obtaining a **'t- value'** of 33.899 by chance, when H₀ is true, is much lower than 5% (<<0.0005) and therefore, we reject H₀ and conclude that decay patterns detected on sheltered untreated areas have indeed lower intensity levels than decay patterns segmented on areas cleaned by the BP method. For the case of the Sub-Region Decomposition and the Region Growing Algorithms the conclusions are similar. An assessment, though, that can be drawn is that the response of the Sub-Region Decomposition deviates significantly from the responses of the other two algorithms. This can be explained by considering that the Sub-Region Decomposition Algorithm tends to split areas into smaller segments affecting in this way the extent of the segmented areas and subsequently the distribution of intensities within these areas.

4. Through experiment 6 we examine whether the black crusts prevailing on sheltered untreated reedings tend to be thicker than black crusts occurring on sheltered reedings treated by the DS. The hypotheses made are:

H₀: The mean intensity values of the black particles encountered on sheltered reedings treated by DS are laid to lower values than the corresponding intensity levels encountered at sheltered untreated reedings.

H_A: The mean intensity values of the black particles encountered at untreated sheltered reedings are laid to lower values than the corresponding intensity levels encountered at sheltered reedings treated by the DS.

The t-table reveals that the critical t for the case of the Conditional Thickening (9 df at 0.05) gives a t-value of 1.833. Our calculation of t-value is 10.017, indicating that the likelihood of observed difference in means having occurred by chance given that H₀ is true is (<0.05). Thus H₀ is rejected and we adopt H_A as true. The same conclusions are also drawn for the cases of the Region Growing and the Sub-Region Decomposition Algorithms, where df= 10 (Crit. $t_{df=10}$= 1.812) and t=14.020 and 13.712 respectively. It can be stated, thus, that the DS method is found to be efficient in increasing the mean intensity levels.

5. At the 7th experiment, our objective is to assess whether the application of the (BP) cleaning process on sheltered untreated reedings results in diminishing the black crust's thickness. The hypotheses made are:

H₀: The mean intensity levels within black particles encountered on sheltered reedings treated by BP are higher than the mean intensity levels of black particles detected on sheltered untreated reedings.

H_A: The mean intensity levels of the black particles encountered on sheltered untreated reedings are lower than the mean intensity levels of black particles detected on sheltered reedings treated by the BP.

For the Conditional Thickening Algorithm, the value of "t" at the 5% level of probability for 8 DF, taken from the t-table, is 1.860. The "t- value" obtained in the case of the Conditional Thickening is 12.716 and is larger than the value of 't' at P=0.05. The probability of obtaining a value of 't' of 12.716 by chance, when H_0 is truth, is much lower than 5% (<0.0005) and therefore, we reject H_0 and conclude that the decay patterns detected on sheltered untreated reedings are darker than the decay patterns remained on sheltered reedings after being cleaned by the BP method. For the case of the Sub-Region Decomposition and the Region Growing Algorithms the conclusions are similar as the critical values of "t" for the one-tail t-test (DF= 10) are 1.812 while the t-values derived by the experiment are 14.436 and 18.321 respectively. Thus the probability of these 't' values having arisen by chance given that H_0 is true is <0.0005 and we conclude that H_0 should be rejected in favor of H_A.

6. The aspect studied through test 8 is to assess whether the black crusts located at sheltered untreated flutings are thicker than the black crusts located at sheltered untreated reedings. The hypotheses tested are described below:

H_0: The mean intensity levels within black particles encountered on sheltered untreated reedings are lower the mean intensity levels within black particles detected on sheltered untreated flutings.

H_A: The mean intensity levels within black particles encountered on sheltered untreated reedings are higher than the mean intensity levels of black particles detected on sheltered untreated flutings.

Using the t-table to look up at the critical t for the case of the Conditional Thickening (28 df at 0.05) gives a t-value of 1.701. Our calculation of t-value is 13.443, which is much larger than the critical t-value of the 1-tail t-test indicating that the likelihood of observed differences (in populations' means) having occurred by chance, given that H_0 is true, is extremely low (<0.001). Thus H_0 is rejected in favor of H_A. Similar conclusions are also drawn for the case of the Region Growing and the Sub-Region Decomposition Algorithms, where df= 28 (Crit. $t_{df=28}$= 1.701) and t=14.789 and 17.388 respectively. As it can be seen, the performance of the Sub-Region Decomposition deviates from the performances of the other two algorithms. This discrepancy can be explained by the fact that the former tends to split adjacent decay areas thus providing different distributions of intensity encountered on locations where black spots prevail.

7. Here we are examining whether the black crusts located in sheltered untreated flutings are thicker than the black crusts located at unsheltered untreated flutings. The hypotheses tested are described below:

H_0: The mean intensity levels within black particles encountered on unsheltered untreated flutings are lower than the mean intensity levels characterizing black particles on sheltered untreated flutings.

H_A: The mean intensity levels of the black particles determined on unsheltered untreated flutings are higher than the mean intensity levels of black particles detected on sheltered untreated flutings.

Considering the Conditional Thickening (34 df at 0.05) the t-table gives a critical t-value of 1.691 for the 1-tail test. Our calculation of t-value is 47.960 and is much larger than the critical value of the 1-tail t-test indicating that the likelihood of observed difference in means having occurred by chance given that H_0 is true is extremely low (<0.001). Thus H_0 is rejected in favor of H_A. Similar conclusions are also drawn for the case of the Region Growing and the Sub-Region Decomposition Algorithms, where df= 34 (Crit. $t_{df=28}$= 1.691) and t=41.429 and 31.016 respectively.

8. Through experiment 8 we are attempting to examine whether the black crusts prevailing on unsheltered untreated flutings tend to be thicker than the black crusts encountered on unsheltered flutings treated by the DS. The hypotheses made are:

H_0: The mean intensity values within black particles encountered at untreated unsheltered flutings are laid to higher values than the corresponding intensity levels encountered at unsheltered flutings treated by the DS.

H_A: The mean intensity values of the black particles encountered on untreated unsheltered flutings are laid to lower values than the corresponding intensity levels of black particles encountered on unsheltered flutings treated by the DS.

Using the corresponding look up table, the critical t-value for the case of the Conditional Thickening (22 df at 0.05) gives a t-value of 1.717. Our calculation of t-value is 17.749, and is larger than this critical value indicating that the likelihood of observed difference in means having occurred by chance, given that H_0 is true, is (<<0.05). Thus H_0 is rejected in favor of H_A. Similar conclusions are also drawn for the case of the Region Growing and the Sub-Region Decomposition Algorithms, where DF= 22 and t= 20.765 and 22.089 respectively. It can be stated, thus, that the DS method is found to be efficient in increasing the mean intensity levels at areas of corroded patterns presence.

9. Test 9 investigates whether black crusts located on unsheltered untreated flutings are thicker than the black crusts prevailing on sheltered untreated reedings. The hypotheses tested are described below:

H_0: The mean intensity levels of black particles encountered on unsheltered untreated flutings are lower than the mean intensity levels within black particles detected on sheltered untreated reedings.

H_A: The mean intensity levels of black spots occurring on sheltered untreated reedings are lower than the mean intensity levels of black particles detected on unsheltered untreated flutings.

The t-table reveals that the critical t for the 1-tail t-test (16 df at 0.05) gives a t-value of 1.746. The calculation of t-value regarding the Conditional Thickening is 16.347, much larger than the critical value of the 1-tail t-test, indicating that the likelihood of the observed difference (in populations' means) having occurred by chance, given that H_0 is true, is low (<<0.001). Thus H_0 is rejected in favor of H_A. Similar conclusions are also drawn for the case of the Region Growing and the Sub-Region Decomposition Algorithms, where df= 16 (Crit. $t_{df=28}$= 1.746) and t=16.487 and 18.940 respectively.

10. Test 10 is recruited to assess whether black crusts located on unsheltered untreated flutings are thicker than black crusts located at unsheltered untreated reedings. The hypotheses tested are described below:

H$_0$: The mean intensity levels of black particles encountered on unsheltered untreated reedings are lower than the mean intensity levels of black particles detected on unsheltered untreated flutings.

H$_A$: The mean intensity levels of black particles encountered on unsheltered untreated reedings are higher than the mean intensity levels within black particles detected on unsheltered untreated flutings.

According to the t-table, the critical t (16 df at 0.05) is 1.746. Our calculation of t-value regarding the Conditional Thickening is 3.575 and is marginally larger than the critical value of the 1-tail t-test indicating that the likelihood of observed difference in means having occurred by chance given that H$_0$ is true is marginally lower than 0.05. The conclusions drawn for the case of the Region Growing and the Sub-Region Decomposition Algorithms are similar with df= 16 (Crit. $t_{df=16}$= 1.746) and t=3.090 and 4.799 respectively. This test reveals that a difference between the mean of intensity levels of decay patterns on unsheltered untreated flutings and reedings may occur, but is not of high significance.

11. Through experiment 11 we are attempting to examine whether the black crusts occurring on unsheltered untreated reedings tend to be thicker than the black crusts that prevail on unsheltered reedings treated by the DS. The hypotheses made are:

H$_0$: The mean intensity values of the black particles encountered at unsheltered reedings treated by DS are laid to lower values than the corresponding intensity levels encountered at unsheltered untreated reedings.

H$_A$: The mean intensity values of the black particles encountered at untreated unsheltered reedings are laid to lower values than the corresponding intensity levels encountered at unsheltered reedings treated by the DS.

The t-table derives a t-value (9 df at 0.05) of 1.833 as critical for the 1-tail t-test. Our calculation of t-value regarding the Conditional Thickening is 7.576, which is much greater than this critical value, indicating that the likelihood of observed difference in means having occurred by chance, given that H$_0$ is true, is (<<0.01). Thus H$_0$ is rejected in favor of H$_A$. The same conclusion is also drawn for the cases of the Region Growing and the Sub-Region Decomposition Algorithms, where df= 10 (Crit.$_t$$_{df}$=10= 1.812) and t=7.004 and 8.401 respectively. Thus, it can be stated that the DS method is found to be efficient in increasing the mean intensity levels observed on black particles that occur on unsheltered reedings.

12. At the 12^h experiment our objective is to assess whether the application of the (WMB) cleaning process on unsheltered untreated reedings results in diminishing the black crust's thickness. The hypotheses made are:

H$_0$: The mean intensity levels of the black particles encountered on unsheltered reedings treated by the WMB are lower than the mean intensity levels of black particles detected on unsheltered untreated reedings.

H$_A$: The mean intensity levels of the black particles encountered on unsheltered untreated reedings are lower than the mean intensity levels of black particles detected on unsheltered reedings treated by the WMB.

As it is derived by the t-table the critical t-value for the 1-tail t-test at 5% level of probability for 9 df, is 1.833. The value of 't' obtained in the case of the Conditional Thickening is 8.42 and appears to be larger than the value of **t'** at P=0.05. The probability of obtaining a value of **'t'** of 8.42 by chance, when H_0 is truth, is lower than 5% (<0.005) and therefore, we reject H_0 and conclude that the decay patterns detected on unsheltered untreated reedings have indeed lower intensity levels than the decay patterns segmented on unsheltered reedings cleaned by the WMB method. For the case of the Sub-Region Decomposition and the Region Growing Algorithms the conclusions are similar as the critical values of "t" for the one-tail t-test (DF =10) are 1.812 while the t-values derived by the experiment are 11.63 and 7.77 respectively. Thus the probability of these **'t- values'** to have arisen by chance, given that H_0 is true, is <<0.0005 and we conclude that H_0 should be rejected and we should accept H_A.

13. Here we are testing whether the black crusts located on sheltered untreated reedings are thicker than the black crusts located at unsheltered untreated reedings. By H_0 and H_A we denote the null and the alternative hypotheses employed:

H_0: The mean intensity levels of the black particles encountered on unsheltered untreated reedings are lower than the mean intensity levels of black particles detected on sheltered untreated reedings.

H_A: The mean intensity levels of the black particles encountered on unsheltered untreated reedings are higher than the mean intensity levels of black particles detected on sheltered untreated reedings.

The t-table reveals that the critical t-value (16 df at 0.05) equals 1.746. The calculation of t-value in the case of the Conditional Thickening is 25.223, which is much larger than the critical value of the 1-tail t-test, indicating that the likelihood of observed difference in means having occurred by chance given that H_0 is true is extremely low (<<0.001). Thus H_0 is rejected in favor of H_A. Identical conclusions are also drawn for the cases of the Region Growing and the Sub-Region Decomposition Algorithms, where df= 10(Crit. $t_{df=16}$= 1.746) and t=25.972 and 30.9584, respectively. In this test we can assess that the results derived by the Sub-Region Decomposition depart significantly from the results derived by the other 2 algorithmic processes. This can be attributed to the fact that the number of decay areas segmented on untreated reedings is low and makes the conclusions rather dicey.

14. Through experiment 14 we are trying to assess whether DS performs better removal of crust's thickness compared to the WMB. The tests concern unsheltered surfaces located at the reedings of the columns. The hypotheses made are:

H_0: The intensity values of the black particles encountered at unsheltered reedings (cleaned by the WMB) are laid to lower values than the corresponding intensity levels encountered at areas treated by the DS.

H_A: The intensity values of the black particles encountered at unsheltered reedings (cleaned by the WMB) are laid to higher values than the corresponding intensity levels encountered at areas cleaned by the DS.

Using the t-table at the critical t (8 df at 0.05) gives a t-value of 1.860. Our calculation of t-value concerning the Conditional Thickening is 3.46 and is marginally larger than the critical t-value, indicating

that the likelihood of observed difference in means having occurred by chance, given that H_0 is true, is low (<0.05). Thus H_0 is rejected (marginally) in favor of H_A. The conclusion drawn for the case of the Region Growing and the Sub-Region Decomposition Algorithms appear to be quite similar. More specifically, df= 10 (Crit. $t_{df=10}$= 1.812) and t=2.897 and 3.226 respectively. Thus for the two latter algorithms, H_0 can only marginally be rejected.

4.4.1.2. Summarization on the results derived from the T-tests

The tests discussed above are focused towards investigating the potential and the limitations of each of the employed chemical intervention methods in eliminating the crusts thickness. As it is revealed, the cleaning procedures succeed in eliminating the depth of black crusts. This is reflected to the studied images by an increase of the relative intensities of degraded regions over the background. The tests regarding the DS cleaning approach, however, revealed that this treatment demonstrates poorer performance than the other cleaning procedures. Further tests were carried out with the aim of examining the occurrence of significant differences between the mean intensity levels observed on DS cleaned surfaces and surfaces cleaned by other treatments. Another objective of the t-tests was to estimate whether structural effects are reflected in the relative intensities of decay regions over the background. According to the derived results, sheltered flutings generally demonstrate darker decay patterns than sheltered reedings. This assessment is in accordance with the experts' judgment as corrosion materials are more easily deposited on sheltered flutings [54, 55]. Regarding the unsheltered areas, a similar relation is not so obvious. The latter observation can be explained by considering that the effects of weathering are almost identical for both the unsheltered flutings and reedings and, thus, we can expect that no significant discrepancies on the intensity values observed on decay areas would occur. Moreover tests concerning variations on the thickness of black crusts between sheltered and unsheltered untreated areas revealed that on sheltered areas the black crusts are always thicker. This is mainly due to the fact that the crusts prevailing on unsheltered areas are always characterized by minor thickness due to the water's fluency that removes the deposited corrosion materials. Finally, the t-tests offer the opportunity to draw some important conclusions regarding the algorithmic responses. Thus, we can observe that Sub-Region Decomposition algorithm provide results demonstrating some discrepancies from the results derived by the other two algorithms. This is mainly arisen due to the fact that it tends to split areas that are segmented as compact by the others. This behaviour affects the distribution of intensity levels of areas segmented as decay patterns.

4.4.2. Mann-Whitney U-test

As it was mentioned in the 1st chapter of this thesis, we have observed that an association occurs between the extent of each individual degradation spot and the severity of degradation. More specifically, it was assessed that we tend to segment larger in size decay patterns on areas, which are considered as more degraded by the experts. In order to establish the existence of such an association we employ a non-parametric statistical test, which takes as input the areas (size) covered by deterioration particles at different surfaces and studies the significance of the observed differences. The employment of the

specific statistical test was decided because according to our observations the distribution of the median levels of the area sizes (considering each category of the studied areas separately) depart significantly from the normal distribution. Due to the small size of the samples we decomposed each image in 6 sub-images of equal dimensions (288x256 pixels) and the median size of the black particles detected in each sub-image was extracted. Table 9 summarizes the Mann-Whitney U-test when performed on the studied surfaces.

Table 4-9: Study on the significance of decay patterns size alteration[9].

	MFD	Region Growing	Skew & Kurtosis
1. Sheltered Flutings (Ds) (vs) Sheltered Flutings (Diag)	$N_1=12$ $N_2=24$ $U=36$ $U_{crit}=74$ $P=5.92*10^{-5}$	$N_1=12$ $N_2=24$ $U=0$ $U_{crit}=74$ $p=7.98x10^{-10}$	$N_1=12$ $N_2=24$ $U=0$ $U_{crit}=74$ $p=7.98x10^{-10}$
2. Sheltered Flutings (Diag) (vs) Sheltered Flutings (WMB)	$N_1=24$ $N_2=6$ $U=144$ $U_{crit}=27$ $P=1.68x10^{-6}$	$N_1=24$ $N_2=10$ $U=288$ $U_{crit}=74$ $p=7.98x10^{-10}$	$N_1=24$ $N_2=18$ $U=432$ $U_{crit}=124$ $p=10^{-6}$
3. Sheltered Flutings (Diag) (vs) Sheltered Flutings (BP)	$N1=24$ $N_2=6$ $U=144$ $U_{crit}=27$ $P=1.68x10^{-6}$	$N_1=24$ $N_2=6$ $U=144$ $U_{crit}=27$ $p=1.68x10^{-6}$	$N_1=24$ $N_2=6$ $U=144$ $U_{crit}=27$ $p=1.68x10^{-6}$
4. Sheltered Reedings (Diag) (vs) Sheltered Reedings (DS)	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $P=10.8x10^{-4}$	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $p=10.8x10^{-4}$	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $p=10.8x10^{-4}$
5. Sheltered Reedings (Diag) (vs) Sheltered Reedings (BP)	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $P=10.8x10^{-4}$	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $p=10.8x10^{-4}$	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $p=10.8x10^{-4}$
6. Sheltered Reedings (Diagnosis) (vs) Sheltered Flutings (Diagnosis)	$N_1=6$ $N_2=24$ $U=0$ $U_{crit}=27$ $P=1.68x10^{-6}$	$N_1=6$ $N_2=24$ $U=18$ $U_{crit}=27$ $p=16.7x10^{-4}$	$N_1=6$ $N_2=24$ $U=18$ $U_{crit}=27$ $p=16.7x10^{-4}$
7. Unsheltered Flutings (Diagnosis) (vs) Sheltered Flutings (Diagnosis)	$N_1=12$ $N_2=24$ $U=70$ $U_{crit}=74$ $P=7.98x10^{-10}$	$N_1=12$ $N_2=24$ $U=70$ $U_{crit}=74$ $p=7.98x10^{-10}$	$N_1=12$ $N_2=24$ $U=36$ $U_{crit}=74$ $p=5.92*10^{-5}$
8. Unsheltered Flutings (Diag) (vs) Unsheltered Flutings (Ds)	$N_1=12$ $N_2=12$ $U=144$ $U_{crit}=31$ $P=3.69x10^{-7}$	$N_1=12$ $N_2=12$ $U=144$ $U_{crit}=31$ $p=3.69x10^{-7}$	$N_1=12$ $N_2=12$ $U=127$ $U_{crit}=31$ $p=4.28x10^{-4}$
9. Unsheltered Flutings (Diagnosis) (vs) Sheltered Reedings (Diagnosis)	**$N_1=12$ $N_2=6$** **$U=62$ $U_{crit}=9$** **$P=6.7x10^{-3}$**	**$N_1=12$ $N_2=6$** **$U=55$ $U_{crit}=9$** **$p=4.1x10^{-2}$**	**$N_1=12$ $N_2=6$** **$U=54$ $U_{crit}=9$** **$p=5.1x10^{-2}$**
10. Unsheltered Flutings (Diag) (vs) Unsheltered Reedings (Diagnosis)	$N_1=12$ $N_2=6$ $U=72$ $U_{crit}=9$ $P=5.38x10^{-5}$	$N_1=12$ $N_2=6$ $U=72$ $U_{crit}=9$ $p=5.38x10^{-5}$	$N_1=12$ $N_2=6$ $U=65$ $U_{crit}=9$ $p=2.3x10^{-3}$
11. Unsheltered Reedings (Diag) (vs) Unsheltered Reedings (Ds)	**$N_1=6$ $N_2=6$** **$U=24$ $U_{crit}=3$** **$P=0.19$**	$N_1=6$ $N_2=6$ **$U=24$ $U_{crit}=3$** **$p=0.19$**	**$N_1=6$ $N_2=6$** **$U=26$ $U_{crit}=3$** **$p=0.12$**
12. Unsheltered Reedings (Diag) (vs) Unsheltered Reedings (WMB)	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $P=10.8x10^{-4}$	$N_1=6$ $N_2=6$ $U=31$ $U_{crit}=3$ $p=2.05x10^{-2}$	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $p=10.8x10^{-4}$
13. Sheltered Reedings (Diag) (vs) UnSheltered Reedings (Diagnosis)	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $P=10.8x10^{-4}$	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $p=10.8x10^{-4}$	$N_1=6$ $N_2=6$ $U=36$ $U_{crit}=3$ $p=10.8x10^{-4}$

4.4.2.1. Analysis on the results derived by the Mann-Whitney U-Test

1. In the first case discussed in Table 9 we investigate whether the employment of a cleaning intervention based on application of anionic resin onto sheltered untreated flutings results in a reduction of size in the individual black particles.

H₀: The median levels of the black particles sizes detected on sheltered flutings treated by the DS method are greater than the median levels of the black particles sizes segmented on sheltered untreated flutings.

---

[9] *The rows highlighted in red fonts correspond to cases at which either a significant difference does not occur or the difference is marginally significant.*

H_A: The decay areas that prevail on sheltered untreated flutings (their median levels) tend to be greater in extent than black particles detected on the sheltered flutings treated by the DS.

In order to assess whether we should reject or accept the null hypothesis we perform the one-tail Mann Whitney U-test. The value of "U" at the 1% level of probability for $n_1=12$ and $n_2=24$ taken from the t-table is $U_{crit}=74$. The value of "U'" obtained in the case of the Conditional Thickening is 36, which is lower than the value of **'U'** at P=0.01. The probability of obtaining a value of **'U'** of 36 by chance, when H_0 is truth, is much lower than 1%. Therefore, we reject H_0 and conclude that the black particles detected on sheltered untreated areas tend to e larger in extent than the black particles segmented on areas cleaned by the DS method. H_A is also presented to be true according to the segmentation results derived by the Region Growing and the Sub-Region Decomposition. As it is seen in the 1st case, the Sub-Region Decomposition and the Region Growing Algorithms differ from the Conditional Thickening's performance, especially for the surfaces that have been cleaned by the DS Method (anionic resin in combination with de-ionized water). The observed variation is mainly due to the fact that the images depicting sheltered flutings that have been cleaned by the DS method present strains on the stone material with colour alterations. The Conditional Thickening Algorithm does not detect decay patterns on these regions, while Sub-Region Decomposition and the Region Growing Algorithm segment decay patterns small in extent. The occurrence of these regions mainly affects the observed deviation on the results derived by the three algorithms.

2. Through the second experiment, we attempt to assess whether cleaning with the use of the WMB manages to reduce the size of the remaining black particles. The hypotheses tested are:

H_0: The median levels of the black particles sizes detected on untreated sheltered flutings is lower than the median levels of the black particles sizes segmented on sheltered flutings treated by the WMB cleaning process.

H_A: The median levels of the black particles sizes detected on sheltered flutings treated by the WMB method is lower than the median levels of the black particles sizes segmented on sheltered untreated flutings.

Using the U-table, the critical **U** for the case of the Conditional Thickening (for n1=6 and n2=24 at p=0.01) gives a U critical-value of 27. Our calculation of U-value is 144, which is larger than the critical value, indicating that the likelihood of observed difference in medians having occurred by chance given that H_0 is true is very low ($p=1.68 \times 10^{-6}$) ($p \ll 0.001$). Thus H_0 is rejected in favor of H_A. The conclusions drawn for the case of the Region Growing and the Sub-Region Decomposition Algorithms appear to deviate from the Conditional Thickening. More specifically, for the Region Growing we can assess that H_0 is rejected as the probability of the specific U-value (U=288 for n1=12 and n2 =24) to have arisen by chance, given that H_0 holds true, is $p=7.98 \times 10^{-10}$ ($p \ll 0.001$). H_0 is rejected in favor of H_A and we can state that the decay patterns detected on sheltered flutings after the application of the WMB methods are smaller in size.

3. Similarly to experiment 2, we now aim at investigating the effect of BP cleaning on the size of the black particles that have not been removed. In order to assess whether a significant reduction of the black particles sizes takes place after the cleaning with the BP, the hypotheses tested are:

H_0: The median levels of the black particles sizes detected on untreated sheltered flutings is lower than the median levels of the black particles sizes segmented on sheltered flutings treated by the BP cleaning process.

H_A: The median levels of the black particles sizes detected on sheltered flutings treated by the BP method is lower than the median levels of the black particles sizes segmented on sheltered untreated flutings.

As it is revealed by a look-up in the U-Table considering the Conditional Thickening Algorithm, the likelihood of U's value to be equal to 144 given that $n_1=6$ and $n_2=24$ is $p=1.68 \times 10^{-6}$ ($p<0.0001$). Thus the difference between the medians of the two populations is highly significant and we can conclude that H_0 is rejected in favor of H_A. The same conclusions are also true when considering the decay areas segmented by the Sub-Region Decomposition and the Region Growing Algorithms. Thus, the test also proves that cleaning intervention applied on sheltered untreated flutings manages to reduce the extent of the remaining decay patterns sizes.

4. Through test 4 we examine the alteration of decay patterns sizes detected on sheltered untreated reedings after the surfaces have been cleaned by the DS method. The hypotheses tested to estimate whether a significant alteration occurs are:

H_0: The median levels of the black particles sizes detected on untreated sheltered reedings are lower than the median levels of the black particles sizes segmented on sheltered reedings treated by the DS cleaning process.

H_A: The median levels of the black particles sizes detected on sheltered reedings treated by the DS method are lower than the median levels of the black particles sizes segmented on sheltered untreated reedings.

From table 9 we assess that the probability of H_0 to be true with $n_1=n_2=6$ and $U=36$ is $p=10.8 \times 10^{-4}$ ($p<0.01$). Thus, we can assess that H_0 can be rejected in favor of H_A. Such a result indicates that, according to the three studied algorithms, the DS method when applied to sheltered untreated regions results in a significant diminishment of the decay patterns sizes.

5. Test 5 performs an identical test to the preceding one, but here we examine the efficiency of BP in eliminating the decay patterns sizes. The hypotheses stated are:

H_0: The median levels of the black particles sizes detected on untreated sheltered reedings are lower than the median levels of the black particles sizes segmented on sheltered reedings treated by the BP cleaning process.

H_A: The median levels of the black particles sizes detected on sheltered reedings treated by the BP method are lower than the median levels of the black particles sizes segmented on sheltered untreated reedings.

The results derived by the three algorithms indicate that the parameters of the test are: $n_1=6=n_2=6$ and $U=36$. Thus the likelihood of the U value to have arisen by chance given that H0 is true is $p=10.8 \times 10^{-4}$ ($p<0.01$). This in turn means that H_0 should be rejected and we accept H_A to be true. Experiment 5 also reveals that the BP cleaning methodology affects the size of the remaining decay patterns and succeeds in eliminating then.

6. Further to estimating the effects of cleaning methods in the size of decay patterns, through the current approach we are also focused towards investigating whether conditions of exposure affects the size of decay patterns. Through experiment 6 we examine if the size of the black particles that prevail on sheltered untreated reedings differs significantly from the size of the decay patterns segmented on sheltered untreated flutings. The hypotheses stated are:

H₀: The median levels of the black particles sizes detected on untreated sheltered reedings are greater than the median levels of the black particles sizes segmented on sheltered untreated flutings.

H_A: The median levels of the black particles sizes detected on sheltered untreated flutings are laid to higher values than the corresponding values concerning the sheltered untreated reedings.

Table 9 reveals that considering the Conditional Thickening algorithm we obtain a U-value of $U=0$ and $U<<U_{crit}$. Thus the probability of the value of U (while $n_1=6$ and $n_2=24$) to have arisen by chance given that H_0 holds true is $16.8 \times 10^{-5}<0.001$ and, thus, H_0 is rejected to accept H_A. Similar results are also obtained for the cases of Sub-Region Decomposition and the Region Growing, despite the fact that the level of significance between the studied populations is different.

7. Test 7 is quite similar to 6, except for the fact that we study the differences between the sizes of black particles encountered on unsheltered untreated flutings and those detected on sheltered untreated flutings. The hypotheses tested are:

H₀: The median levels of the black particle sizes detected on untreated unsheltered flutings are greater than the median levels of the black particles sizes segmented on sheltered untreated flutings.

H_A: The median levels of the black particles sizes detected on sheltered untreated flutings are laid to higher values than the corresponding values concerning the unsheltered untreated flutings.

The conclusions drawn from this test is that regarding the 2 first algorithms (Conditional Thickening and Region Growing), we can see that $U<U_{crit}$ ($U=70$, $U_{crit}=74$) thus H_0 should be rejected in favor of H_A. However, the results of the Sub-Region Decomposition seem to depart from the results derived by the other two algorithms. This variation is mainly due to the sample flutingx50-4 (diagnosis). This surface demonstrates decay patterns whose spatial distribution is quite different from the spatial arrangement of the other studied degradation effects. More specifically, in this surface we can observe the existence of larger in size decay patterns located at close distances between them, in specific neighborhoods of the image. Thus, the size of the decomposition window used (dimension 60x60) exhibits a difficulty in discriminating decay pattern locations from the background stone structure. The application of windows of larger dimensions has proven to provide results for the Sub-Region Decomposition approach that are more comparable to those derived by the other two algorithms.

8.   Test 8 examines whether a significant elimination of black particle sizes has taken place after the application of DS cleaning method on unsheltered untreated flutings. The hypotheses used are:

H_0:   The median levels of the black particles sizes detected on untreated unsheltered flutings are lower than the median levels of the black particles sizes segmented on unsheltered flutings treated by the DS.

H_A: The median levels of the black particles sizes detected on unsheltered untreated flutings are laid to higher values than the corresponding sizes concerning the unsheltered flutings treated by the DS.

Table 9 reveals that H_0 should be rejected considering all the algorithms studied here, as $U>U_{crit}$ and thus the likelihood of obtaining such values of U for the specific population sizes (n1, n2) and given that H_0 is true is $<<0.01$. The results derived by the Sub-Region Decomposition though differ from those obtained by the other two methods. This behavior of the Sub-Region Decomposition algorithm is also explained through figures (7-10) where we have concluded that the algorithm exhibits a quite good performance at extracting the topology of degradation particles prevalence (fig. 9) while it performs poor in segmenting decay effects at their real extent (figs. 7, 8).

9.   Through experiment 5 we investigate whether the black particles segmented on unsheltered untreated flutings tend to be larger in size than the corresponding particles encountered on sheltered untreated reedings. The hypotheses tested are:

H_0:   The median levels of the black particles sizes detected on untreated unsheltered flutings are greater than the median levels of the black particles sizes segmented on sheltered untreated reedings.

H_A: The median levels of the black particles sizes detected on sheltered untreated flutings are laid to higher values than the corresponding values concerning the sheltered untreated reedings.

By studying table 9 we conclude that the difference between the medians of the studied populations is either marginally significant or insignificant depending on the algorithm. Thus H_0 cannot be rejected and we can state that any differences (if exist) in size between the black particles segmented on sheltered untreated reedings and the unsheltered untreated flutings have arisen by chance.

10.   At a further step, we also examine the occurrence of significant differences between the decay patterns sizes that prevail on unsheltered untreated flutings and on the unsheltered untreated reedings. The hypotheses used are:

H_0:   The median levels of the black particles sizes detected on untreated unsheltered flutings are lower than the median levels of the black particles sizes segmented on unsheltered untreated reedings.

H_A: The median levels of the black particles sizes detected on unsheltered untreated flutings are laid to higher values than the corresponding values concerning the unsheltered untreated reedings.

According to the statistical test, the probability of U=9 (while n_1=6 and n_2=6/ n_2=12) to have arisen by chance given that H_0 holds true is very low (p<0.001) and, thus, H_0 is rejected to accept H_A.

11-12   Through these tests we examine the alteration of decay pattern sizes when unsheltered untreated reedings are cleaned either by DS or by WMB. The hypotheses stated are:

H₀: The median levels of the black particles sizes detected on untreated unsheltered reedings are lower than the median levels of the black particles sizes segmented on unsheltered treated reedings (either by DS or WMB).

Hₐ: The median levels of the black particles sizes detected on unsheltered untreated reedings are laid to higher values than the corresponding values concerning the unsheltered reedings treated (either by DS or WMB).

Table 9 reveals that regarding the DS method, the likelihood of U-values (for the three algorithms) to be equal to 0.19, 0.19 and 0.12 for n1=n2=6 given that H_0 is true, is greater than 0.1. Thus H_0 cannot be rejected. As for the WMB cleaning method, H_0 is rejected as the previous discussed likelihood is >0.05 and thus the difference in the median values obtained from the studied populations are indeed significant.

13. Finally we investigate whether the decay patterns size segmented on sheltered or unsheltered reedings are significantly different. The hypotheses studied are:

H₀: The median levels of the black particles sizes detected on untreated unsheltered reedings are higher than the median levels of the black particles sizes segmented on sheltered untreated reedings

Hₐ: The median levels of the black particles sizes detected on untreated unsheltered reedings are lower than the median levels of the black particles sizes segmented on sheltered untreated reedings.

Table 9 reveals that considering the Conditional Thickening algorithm we obtained U=36 ($U >> U_{crit}$). Thus the probability of the value of U (while $n_1=6$ and $n_2=6$) to have arisen by chance given that H_0 holds true is $10.8 \times 10^{-4} < 0.001$ and thus H_0 is rejected to accept H_A. Similar results are also obtained for the cases of Sub-Region Decomposition and the Region Growing, despite the fact that the level of significance between the studied populations is different.

4.4.2.2. Summarization of the Results of Mann-Whitney U-test

Through the Mann-Whitney U-test we attempt to investigate whether the decay areas that remain after the application of cleaning interventions are eliminated in extent. The selection of the specific statistical test was made because, according to our experiments, the distributions of pattern sizes depart significantly from the normal distribution and thus a non-parametric robust statistical test was employed. An initial assessment that can be drawn by the results reported in table 9 is that the cleaning processes manage to eliminate the size of black particles. This observation is valid for almost all tests, except for the case where unsheltered reedings are cleaned by the DS method. At these areas we can assess the occurrence of decay patterns large in extent after cleaning. This supports the conclusions derived by the chemical analysis [54, 55], where DS performs mild cleaning and is preferable for the unsheltered reedings, as the cracks that prevail and the texture irregularities prescribe the use of methods that minimize the abrations in order to prevent material loss. Another objective of these tests was to elucidate whether the different conditions of exposure affect the size of decay areas segmented on the corresponding surfaces. According to the results obtained through our investigations, the black particles detected on sheltered flutings are always larger than the black particles detected on any other of the studied surfaces. The sheltered untreated surfaces, in general, demonstrate decay patterns of larger extent

comparing to the unsheltered untreated surfaces. This is expected as at sheltered areas crusts of greater thickness prevail. Comparing the size of decay patterns that occur on sheltered reedings and flutings, we can state that on sheltered flutings decay areas of larger size prevail. A similar observation is also valid for the unsheltered areas, but the difference is less significant in this case.

4.5. Shape Features Analysis under different Structural and Cleaning Conditions

As it was discussed earlier in this thesis, an objective of the current work is to examine how the cleaning and structural conditions are reflected onto the shape of the segmented decay areas. Throughout this chapter, we study the shape features derived using the decay patterns boundaries. In order to simplify the analysis and make more effective the comparison of the results we organize the study as follows. At first, we examine whether an alteration of the shape features takes place after the application of cleaning treatments. Subsequently, we examine the variations on shape features induced by the different conditions of exposure.

All data presented in this sub-section refer to areas segmented by the Conditional Thickening Algorithm; because according to the ROC and the Precision Recall curves it presents to be the most reliable both at determining the topology of prevalence and the real extent of decay regions. In the diagrams reported below we illustrate the distribution of values of the shape features. More specifically we try to assess the structural and cleaning effects through studying the eccentricity, compactness, 1st and 2nd central moment of the segmented degraded areas. The number of hole-regions that prevail in the body of the detected decay areas as well as their roughness are also considered as complementary measures indicative of the corrosion effects. In the following subsections we study the alterations of corrosion patterns shape through inspecting the shape feature distribution values. All distributions are described through the mean, median and the quartiles (both lower and upper quartile) since the formed distributions depart significantly from the normal.

4.5.1. Study of the decay patterns eccentricity

Figure 15 illustrates the distribution of the eccentricity values measured on decay areas before and after cleaning.


(a)

Figure 15: Eccentricity values regarding black (a), (c) and white (b), (d) particles located at sheltered ((a), (b)) and unsheltered ((c), (d)) areas respectively.

A brief study of the diagrams, reveals that decay patterns segmented on untreated surfaces tend to have similar eccentricity values. The same observation is valid both for the black and white particles located either at sheltered or unsheltered areas. Regarding the black particles detected on sheltered areas, we can see that reedings are characterized by higher eccentricity values. A further study of the diagrams shows that the black particles segmented on unsheltered surfaces present lower eccentricity values. A final conclusion that can be drawn is that the eccentricity metric is not altered significantly after cleaning.

4.5.2. Study of the Decay Patterns Compactness

Figure 16 investigates the compactness of decay patterns detected on surfaces with different severity of degradation.

(a)

(b)

(c)

(d)

Figure 16: Compactness values regarding black (a), (c) and white (b), (d) particle located at sheltered ((a), (b)) and unsheltered ((c), (d)) areas respectively.

Figure 16(a) and (b) verify that after the application of cleaning interventions a significant elimination of the decay patterns compactness values takes place. The median level of the feature values is considered to indicate reliably the distribution of the metric. Considering the unsheltered surfaces, we can assess that decay patterns of higher compactness values are encountered on the flutings surfaces in comparison to the reedings. A further assessment that can also be drawn is that on unsheltered reedings,

white particles are characterized by higher compactness values than black particles. The opposite observation holds true for the case of the flutings. Finally, it is obvious that corroded areas segmented on sheltered untreated regions are characterized by higher compactness values than the corresponding decay areas (both black and white particles) detected on untreated unsheltered surfaces. Through these comparisons we attempt to investigate whether the different types of corrosion damage are reflected on the shape features of the segmented decay areas.

4.5.3. Study of the Decay Patterns 1st Moment

Figure 17 illustrates the distribution of the 1st moment values measured on surfaces before and after cleaning.

(a)

(b)

(c)

**1st Moment Values Measured on Unsheltered Areas (White Spots)**

(d)
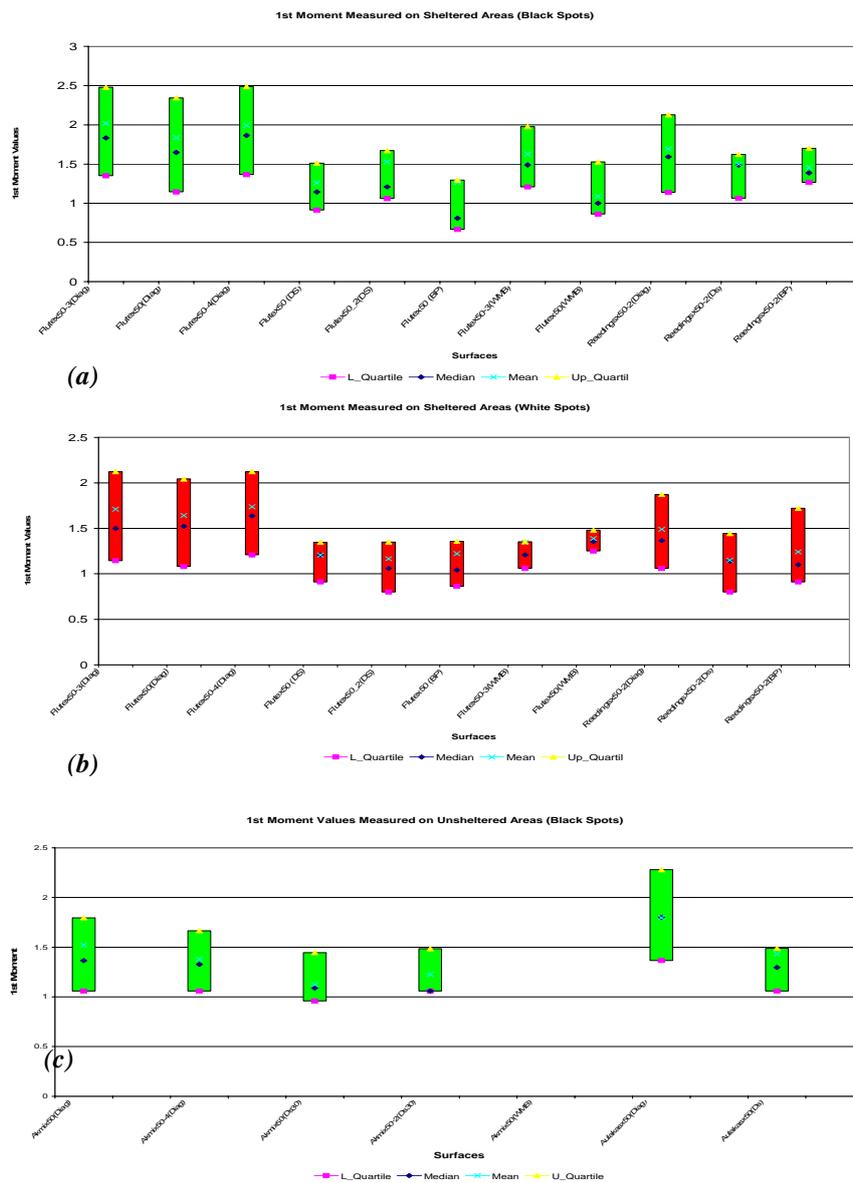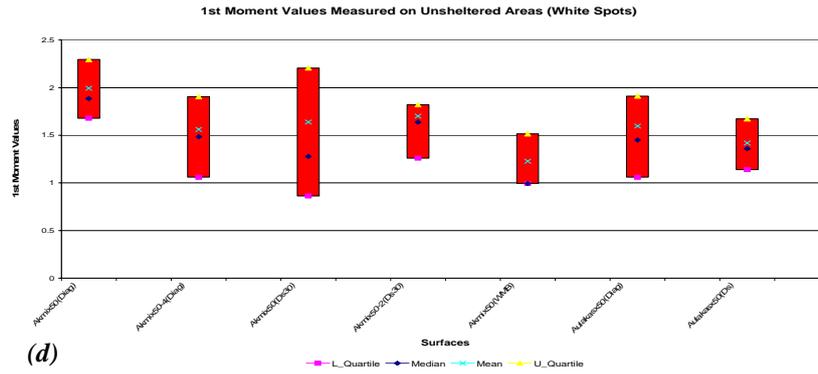
Figure 17: 1st Moment values regarding black (a), (c) and white (b), (d) particle located at sheltered and unsheltered areas respectively.

The 1st moment metric represents the mean Euclidean distance from the centroid to the boundary of segment. A brief study of the data reported in fig. 17(a) and (b) reveals that the patterns' 1st moment values are diminished after chemical treatments. This effect is observable in all types of surfaces (both reedings and flutings), although it is not so clear in the reedings case. Moreover, it can be seen that the 1st moment values of white particles tend to be slightly lower than the corresponding values of the black particles on the sheltered surfaces. The opposite holds true in the case of the unsheltered areas. A further assessment that can be drawn by studying figs. 17(a), (b) and (c) is that the decay patterns segmented at the same surfaces e.g. sheltered flutings or sheltered reedings (untreated), are characterized by similar shape features. This assessment is obvious not only when studying the 1st moment values but to the other shape features as well. Finally, at this point it should be said that patterns located in close areas in the image tend to have similar shape features values. This practically means that we can define neighborhoods in the image where decay patterns of similar feature values prevail. The final assessment may comprise a point of further study.

4.5.4. Study of the Decay Patterns 2nd central moment

Figure 18 presents the distribution of the 2nd central moment values measured on surfaces before and after cleaning.


2nd Central Moment Measured on Sheltered Areas (Black Spots)

(a)

(b)

(c)

(d)
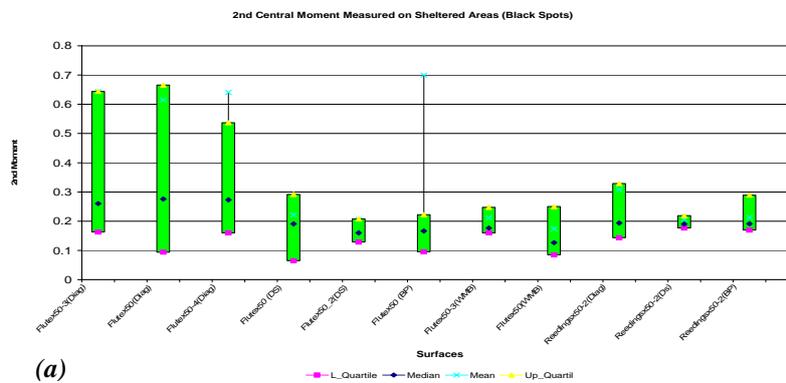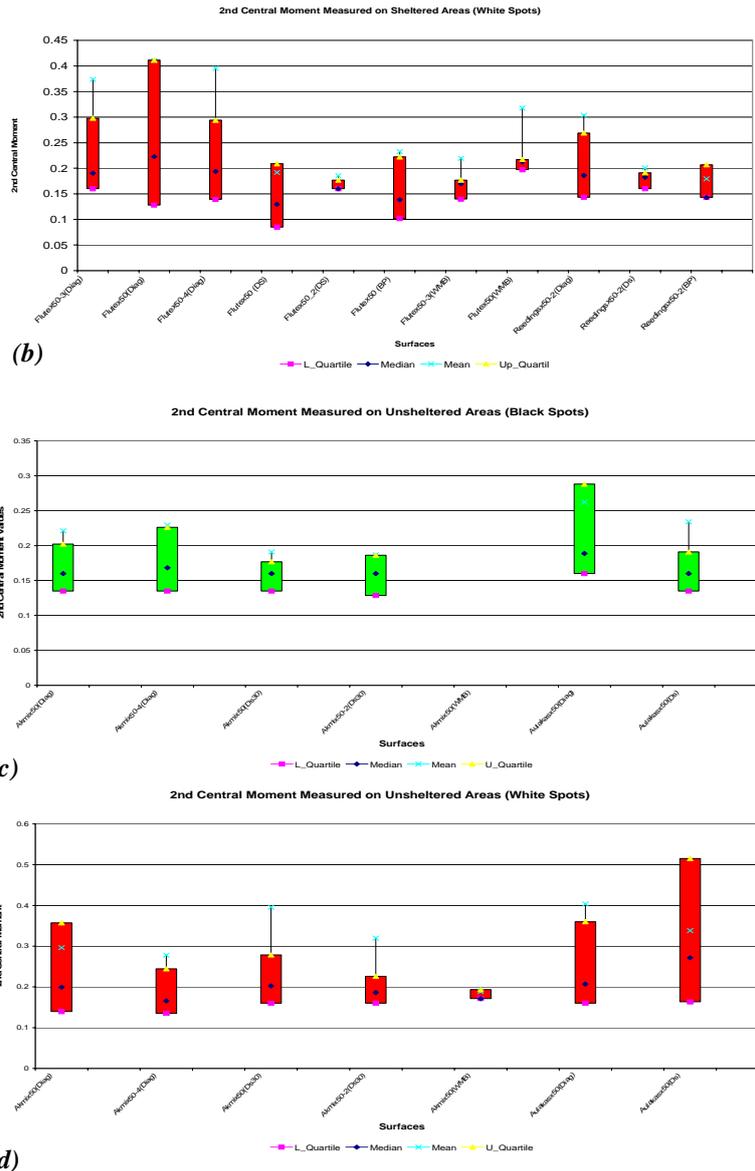
Figure 18: 2nd Moment values regarding black (a), (c) and white (b), (d) particle located at sheltered and unsheltered areas respectively.

The 2nd central moment is another norm recruited in shape boundary representations. Its physical interpretation is not clear. Higher order boundary moments though are mentioned in the literature as translation and rotation invariant shape descriptors and their application is popular in object recognition. Regarding figs. 18(a) and (b) it is revealed that the black particles' 2nd central moment is diminished after the application of the cleaning methods. Such a result of the cleaning method seems to take also place in the case of white particles, but it is not so obvious. Furthermore, we can see that black particles' 2nd central moments higher values compared to white particles.

4.5.5. Study of the Decay Patterns Roughness (F_3-F_1 Metric)

Figure 19 illustrates the distribution of the (F_3-F_1) metric values measured on surfaces before and after cleaning.

(a)

(b)

(c)

(d)
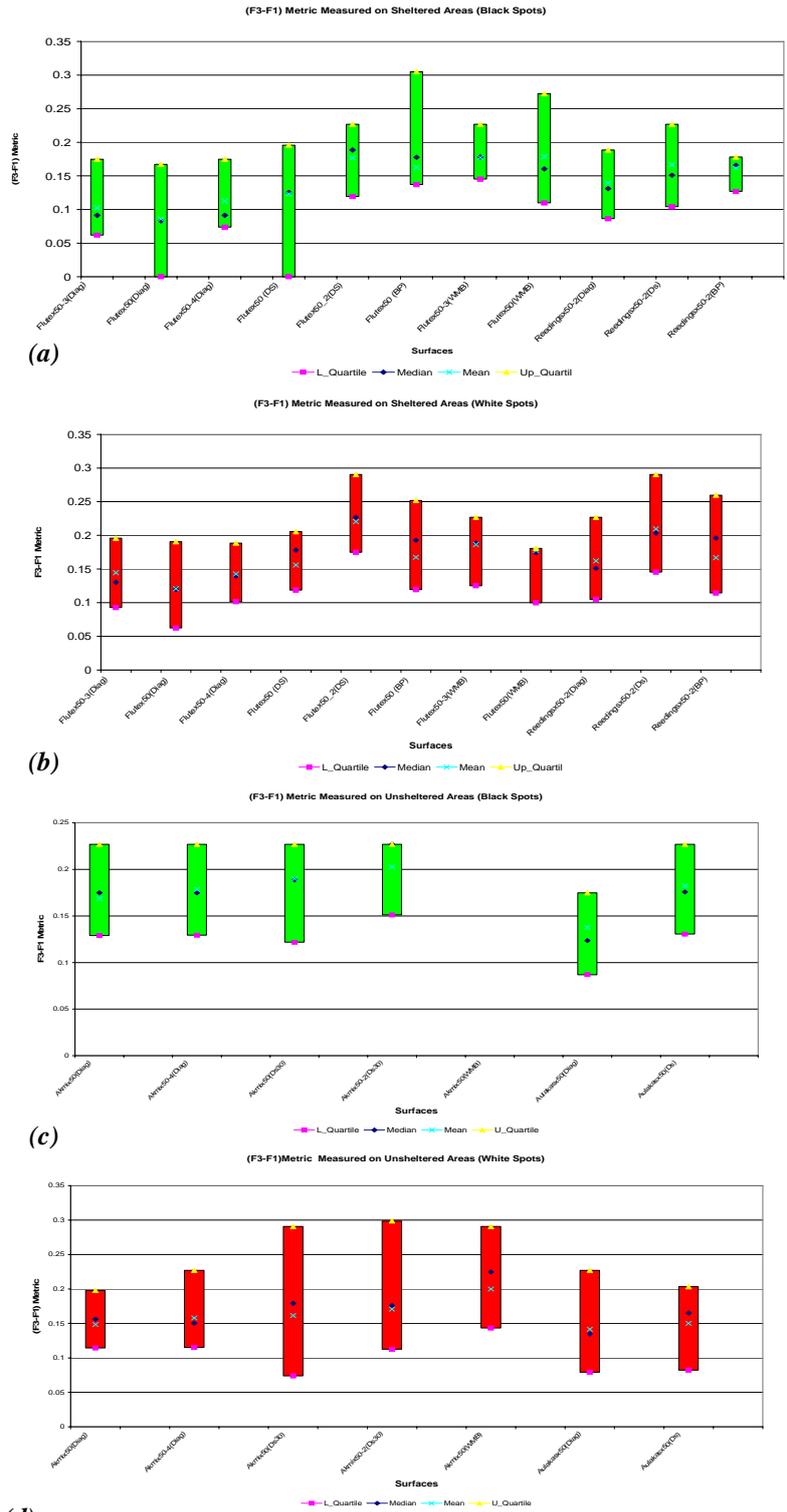
Figure 19: (F$_3$-F$_1$) metric values regarding black (a), (c) and white (b), (d) particle located at sheltered and unsheltered areas respectively.

The (F$_3$-F$_1$) moment value has proven to be a quite accurate roughness metric used in biomedical imaging. A brief observation of the results summarized in figs. 19(a) and (b) reveals that the (F$_3$-F$_1$) values are laid to higher levels in the cleaned surfaces comparing to the un-cleaned. As for the white particles, a similar relation in the distribution of (F$_3$-F$_1$) values can also be drawn. A further study of figs 19(a) and (b) reveals that the distribution of (F$_3$-F$_1$) is laid to higher values in the case of un-cleaned reedings than on

un-cleaned flutings. On the other hand, on figures 19 (c) and (d) it can be seen that the (F_3-F_1) results are laid to higher values in cleaned surfaces compared to uncleaned surfaces. This deviation is more apparent in the detected black spots (Flutingx50 (Diagnosis) vs Flutingx50 (Ds 30 min)). Similar assessments can also be drawn for the white particles.

4.5.6. Study of orientation of Decay Patterns

In the following sub-section, we investigate the orientation of decay patterns as a measure for studying the variability of corrosion effects. More specifically, the axis of least inertia is extracted for each segmented object and then its orientation is calculated. Subsequently the data are processed statistically in order to define the mean and the standard deviation of the distribution of orientations. The use of the mean level of orientations in as an indicative measure of the directionality of segmented decay areas becomes rather dicey, because it introduces faults related to the procedure of monitoring. More specifically movements of the sample during the imaging process affect the metric of mean orientation. On the other hand, according to the author, the variance of the orientation values comprises an indicative measure potential to determine whether the decay patterns tend to have their elongation axis directed to a specific orientation. Through this approach, we attempt to examine whether the conditions of exposure or the type of degradation (chemical composition of the decay patterns) are reflected to the orientation of the decay areas. Table 10 summarizes the variance of the orientations distribution regarding the sheltered studied surfaces.

Table 4-10: Standard deviation of the orientations of degradation patterns segmented on sheltered areas.

Sheltered Surfaces		
	Standard Deviation of Orientations	Mean of Orientations
Flutingx50-1 (Diag) (Black Spots)	1.086	-1.1066
Flutingx50-1 (Diag) (White Spots)	0.657	-2.0206
Flutingx50-2 (Diag) (Black Spots)	1.155	-2.0605
Flutingx50-2 (Diag) (White Spots)	0.747	-0.0949
Flutingx50-3 (Diag) (Black Spots)	1.008	-1.1085
Flutingx50-3 (Diag) (White Spots)	0.563	-1.0702
Flutingx50-4 (Diag) (Black Spots)	1.295	-1.0853
Flutingx50-4 (Diag) (White Spots)	0.743	-1.1204
Reedingx50 (Diag) (Black Spots)	1.082	-1.1533
Reedingx50 (Diag) (White Spots)	0.497	-2.1524
Flutingx50-1(WMB) (Black Spots)	0.000	0.0000
Flutingx50-1(WMB) (White Spots)	0.121	-0.8111
Flutingx50-3(WMB) (Black Spots)	0.447	-0.8730
Flutingx50-3(WMB) (White Spots)	0.528	0.3586
Reedingx50 (BP) (Black Spots)	0.265	-1.0760
Reedingx50 (BP) (White Spots)	0.093	-0.0933
Flutingx50-2 (Ds) (Black Spots)	0.305	1.0426
Flutingx50-2 (Ds) (White Spots)	0.118	1.3227
Flutingx50-1 (Ds) (Black Spots)	0.450	-0.9333
Flutingx50-1 (Ds) (White Spots)	0.520	-0.8161
Reedingx50 (DS) (Black Spots)	0.309	-0.7229
Reedingx50 (DS) (White Spots)	0.118	-0.9134

The data presented in Table 10 indicate that the standard deviation of the orientation values (in ^{0}) is greater for black particles compared to white particles detected on the same areas. This could lead to the assessment that white spots appear to be more oriented at specific direction than black spots. This should be investigated in greater extent to conclude whether an association occurs between the orientation of the decay spots and the origin and the formation of black crusts. In contrast, the segments' orientation after

the application of cleaning methods seems to be more random. Table 11 provides similar information regarding the unsheltered areas.

Table 4-11: Standard deviation of the orientations of degradation patterns segmented on unsheltered areas.

Unsheltered Surfaces		Standard Deviation of Orientations	Mean of Orientations
Reedingx50 (Diag)	(Black Spots)	0.746	-0.0105
Reedingx50 (Diag)	(White Spots)	0.356	0.1164
Reedingx50-4 (Diag)	(Black Spots)	0.770	0.3250
Reedingx50-4 (Diag)	(White Spots)	0.331	-0.0674
Fluting left x50 (Diag)	(Black Spots)	0.464	-0.0273
Fluting left x50 (Diag)	(White Spots)	0.283	-0.0905
Fluting x50 right (Diag)	(Black Spots)	0.441	0.0187
Fluting right x50 (Diag)	(White Spots)	0.234	
Fluting right x50 (DS)	(Black Spots)	0.088	-0.0150
Fluting right x50 (DS)	(White Spots)	1.888	0.6844
Reedingx50 (DS)	(Black Spots)	0.781	0.1522
Reedingx50 (DS)	(White Spots)	0.150	0.0621
Reedingx50-2_(DS)	(Black Spots)	0.233	0.0564
Reedingx50-2_(DS)	(White Spots)	0.301	-0.1077
Reedingx50 (WMB)	(Black Spots)	-	-
Reedingx50 (WMB)	(White Spots)	0.225	-0.1876

By studying the data presented in Table 11, we can assess that similar observations regarding the directionality of white and black spots (untreated) occur. A further conclusion that can be drawn by studying in parallel the data presented in tables 10 and 11 is that the standard deviations of the orientations in the untreated surfaces are generally lower on the unsheltered areas. This may comprise an effect of the rain action, which results in developing decay effects at a specific direction (according to the water fluency).

4.5.7. Study of the Hole regions that prevail in the body of decay areas

A final feature of the decay areas, also considered in this analysis, is the occurrence of hole-regions within the body of segments. According to the experts this is a remarkable characteristic, which reflects the prevalence of discontinuities in the body of black crusts. More specifically the humidity on masonry affects the structure of black crusts due to the dissolution of the gypsum. This phenomenon can be observed (in the microscopical scale) by the occurrence of white spots within the body of black crusts. Through this subsection we attempt to investigate associations between the occurrence of hole-regions and the exposure of the stone material/or its cleaning state. At first, in table 12 we present the fraction of decay areas that present hole-regions. This measure provides an initial general view of the phenomenon.

Table 4-12: Percentage of the segmented decay patterns containing holes into their areas

	Sheltered untreated Flutings	Sheltered Flutings Cleaned by DS	Sheltered Flutings Cleaned by BP	Sheltered Flutings Cleaned by WMB	Sheltered untreated Reedings	Sheltered Cleaned Reedings (DS)	Sheltered Cleaned Reedings (BP)	Unsheltered Untreated Flutings	Unsheltered Flutings Cleaned by DS	Unsheltered Untreated Reedings	Unsheltered Cleaned Reedings (DS)	Unsheltered Cleaned Reedings (WMB)
Percentage of labels containing holes (%)	4.97	0.67	0	0	2.29	0	0	0.88	0	0	0	0

It can be observed that more hole-regions occur within decay areas detected on sheltered flutings and reedings. On the reedings case, though the phenomenon is less severe. As it regards to the other studied surfaces we either do not detect hole regions or their occurrence is almost negligible. These results are in accordance to the experts' judgment as the dissolution of gypsum mainly takes place on sheltered areas where thicker black crusts prevail (sheltered untreated flutings correspond to these areas). Similar examination was conducted for white particles and we did not detect hole-regions in the body of white spots.

Summarizing the results derived through the shape feature analysis we can state that these along with the patterns' extent may provide a tool for classifying the corrosion damage effects. More specifically, we can observe alterations on the (F_3-F_1) metric values when decay patterns located at surfaces of different exposure or cleaning state are studied. Thus, it came out that the distribution of values of the (F_3-F_1) metric is tends towards lower levels for corroded areas segmented on sheltered untreated flutings, while decay areas segmented on sheltered untreated reedings and unsheltered untreated flutings follow. After the application of cleaning interventions the (F_3-F_1) levels are diminished. Furthermore it was revealed that the severity of degradation is closely associated to the levels of 1st and 2nd central moments. Thus, we can assess that cleaning interventions result in a drastic reduction of the metrics' values. A further observation relative to this metric's values is that their values for white spots are generally lower than for the black particles. In respect to the eccentricity and the compactness metrics, we conclude that in general, decay patterns segmented on areas more exposed to rain's action seem to be less compact. This assessment is in accordance to the experts' judgment regarding the effects of weathering on stone's corrosion damage. Furthermore, the decay patterns' compactness values are altered after cleaning treatments. Finally, two other features extracted on the segmented labels are the orientation and the Euler number. As orientation of a label (segmented area) we define the orientation of its axis of least inertia (the extraction and the definition of this term are analyzed in chapter 3). Through the study of the labels orientation, we attempt to estimate whether the decay patterns tend to be oriented to specific directions, which in turn would provide information relative to the procedure of crust formation and maybe the type of corrosion patterns presenting this property. As a means to assess it, we used the standard deviation measured on the distribution of orientations (in ^{0}). Through the results reported in Tables 10 and 11, we can observe that (in general) the decay patterns segmented on unsheltered areas tend to be more oriented than the corresponding patterns segmented at sheltered areas while a similar conclusion can also be drawn concerning the orientation of both black and white particles. The directionality of corroded areas detected on unsheltered areas may reflect an effect of water's fluency. As it concerns to the Euler number the remarkable results are that regarding the sheltered flutings the fraction of the segmented black particles that present holes reaches the 4.97% while the corresponding fraction regarding the unsheltered flutings is 0.88%. Moreover, Euler number for the 1st case falls in the interval [-2, 1] while the corresponding interval for the 2nd case is [0 1]. According to the authors, there is a relation between the size of decay patterns and the Euler number, which can be considered as an effect of splitting. This is in accordance to the experts' judgment that the presence of holes in the detected black

particles is associated to the discontinuities observed in the body of black crusts. The latter are formed due to the dissolution of gypsum by humidity of the walls. Thus the phenomenon mainly prevails on sheltered surfaces, as the crusts' thickness on those areas is larger.

4.6.   Comparison of Monitoring Modalities

Further to validating the efficiency of the implemented segmentation algorithms, this work also investigates the potential and the limitations of various monitoring systems (digital camera, reflectography (Vis/ NIR/ IR) and FOM) via IP techniques in nondestructive qualitative and quantitative evaluation of degradation on marble surfaces. The segmentation algorithm recruited in this procedure is the Conditional Thickening Algorithm, because, according to the ROC and Precision Recall Curves discussed earlier in this chapter, it is considered to be the most accurate in extracting both the locations of decay patterns and their shape and size features. Due to the quite general specifications employed in its design, the proposed algorithm can be applied for the evaluation of decay areas formed on a broad range of background surfaces. Further to its applicability on the FOM images, it can also be used on images providing a more macroscopically description of the corrosion damage. Thus, it can be employed for estimating the degradation state of a stone material as it is illustrated by monitoring systems (Digital Camera and Reflectography) providing macroscopic information, at the same or similar scale as the human eye.

The algorithm is applied to images obtained by different monitoring methods, depicting a variety of degradation patterns. The particular imaging modalities were selected in order to assess how they converge in the determination of decays extent and features. Reflectography in the visible and infrared spectral bands has been extensively used as a diagnostic tool of old paintings damage [53]. The monitoring system used in the current work was implemented by Balas et. al. and provides information on the texture of the studied crusts.

The studied specimens correspond to marble surfaces where adjacent regions of cleaned and un-cleaned crusts prevail. The cleaning process was conducted by a Nd:YAG laser system used to partially remove the crust. The energy fluency of the Nd:YAG laser was fixed at 6.3 J/cm². Throughout the cleaning process, some parameters such as the laser pulses are modified resulting in the removal of crust layers differing in thickness. Each cleaned strip was obtained by increasing the number of laser pulses per spot from one up to six; a 40% area overlap was recorded between adjacent spots. Fiber Optics Microscope (FOM), Digital Camera and a reflectography system were used to depict the degradation effects in the visible, infrared and near infrared bands of the spectrum. The images obtained by the predefined monitoring systems were used as input images in the Conditional Thickening Algorithm.

The evaluation of the derived results is performed by means of both qualitative and quantitative inspection. Qualitative inspection is conducted by the experts in order to estimate the accuracy of the employed approaches in segmenting decay patterns. Through this process, further to the efficiency of the algorithmic approach, the performance of the cleaning interventions is also assessed. The adopted approach is tested on a variety of images obtained through the three monitoring systems. At first, we present results derived by studying the surface illustrated in Fig. 1(a). More specifically, the afore-

mentioned surface is depicted via the FOM in 33 images and via the Reflectography (vis, ir & nir) and digital camera by 3 and one image respectively. These images are studied to validate the accuracy of each imaging modality and to estimate the degree to which result from each modality converge. Our evaluation is conducted in both qualitative and quantitative means.

The qualitative validation is carried out by experts, by optically inspecting the detection results and the original images. Further to visual estimation, the detection results are also compared to assessments derived by the chemical study of the specimen under consideration. The above analysis is essential for validating the correctness of the detected areas in terms of their location of prevalence, as well as their size and shape characteristics. The statistical results obtained through the application of the automated approach further exemplify the efficiency of this approach and elucidates the potential of the proposed monitoring systems as means of revealing and depicting corrosion effects.

4.6.1. Qualitative Results

In this section several results are presented demonstrating the effectiveness of the implemented approach under different monitoring systems. Fig. 11 (a) illustrates the stone surface under consideration as it is depicted by the digital camera, while (b) and (c) shows the segmented black and white spots. As it is discussed in the experimental section, the stone specimen was partially cleaned by using a laser cleaning method with modified pulse intensities. In Fig. 11(a) we can observe the co-existence of untreated and treated stripes on the specimen; these can be discriminated by the colour alterations, so that areas that exhibit a darker surface correspond to uncleaned regions.
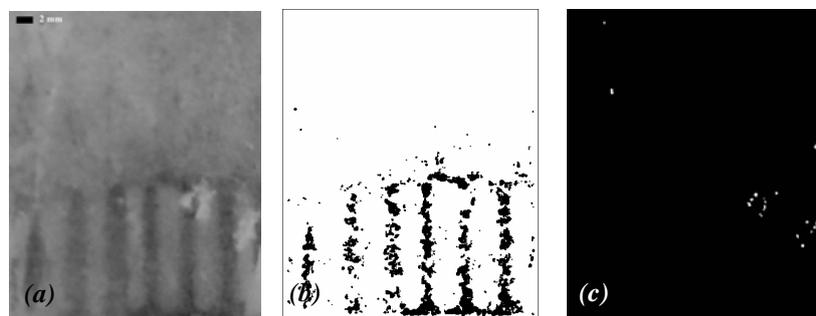

Figure 20:(a) Stone material (monitored by a digital camera) demonstrating cleaned and un-cleaned stripes, (b) Black particles detected, (c) white particles detected.

As it is observed, the segmented degraded areas illustrate regions quite large in extent. The estimation process on the image from a digital camera does not reflect reliable information concerning the structure and the micro-formation of black crusts. The latter can be further explained by considering the low resolution of the digital camera and the fact that it cannot sufficiently depict the structure of the crust, as it does not provide information related to texture characteristics. Thus, the segmentation procedure is carried out by considering only the intensity alteration, which is a secondary effect introduced by the presence of deterioration. Nevertheless, it should be noticed that the determination of the topology of treated and untreated regions is quite effective and in accordance to the experts judgment. For example, it is obvious that only a few black particles are segmented in areas depicting cleaned stripes. Regarding the segmentation of white particles, it should be noticed that the detected areas are associated

with the ablation of stone material and do not reflect the existence of gypsum, re-crystallized $CaCO_3$ or any other decay effect. The latter is associated to the fact that the inter-particle distance between the black spots in the crust matrix is not discernible under the digital camera monitoring. Thus, the prevalence of regions demonstrating bright abrupt changes is considered as white particles.

Reflectography is one of the non-destructive testing methods adopted in the analysis and diagnosis of artworks. The various modalities, (IR) Infrared Reflectography, (NIR) Near Infrared reflectography and (VR) Visible Reflection, allow carrying out tests using different options of illumination. Reflectography provides an effective tool to discover invisible details and texture irregularities. Fig. (21), (22) and (23) depict the same stone specimen as presented in Fig. (20), imaged via a reflectography system that acquire images operating in the spectral band of the visible near-infrared and infrared light respectively. The detected particles are depicted in Figs. (b) and (c).
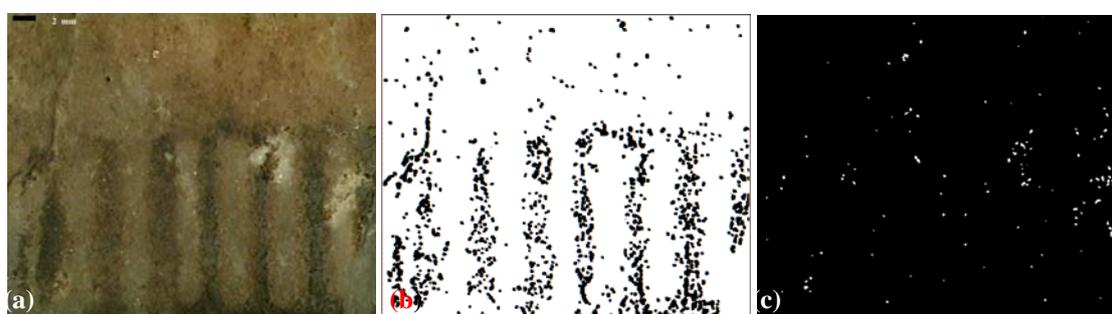

Figure 21: (a) Stone material (monitored by a refectography system (vis)) demonstrating cleaned and un-cleaned stripes, (b) Black particles detected, (c) white particles detected.
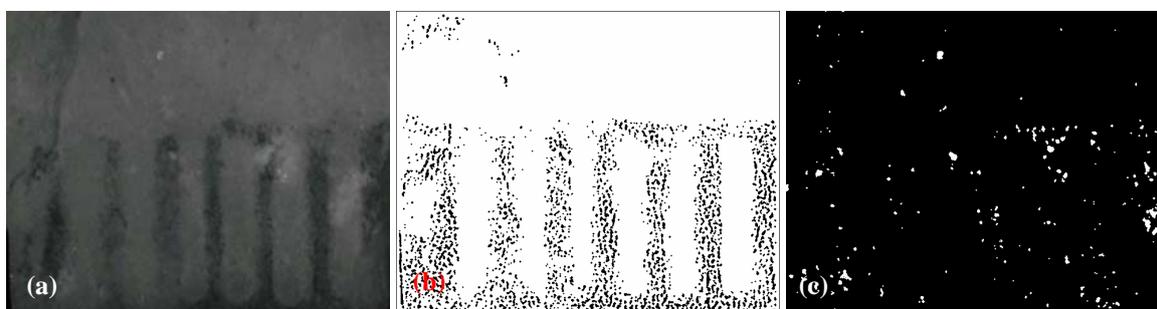

Figure 22: (a) Stone material (monitored by a refectography system (NIR)) demonstrating cleaned and un-cleaned stripes, (b) Black particles detected, (c) white particles detected.
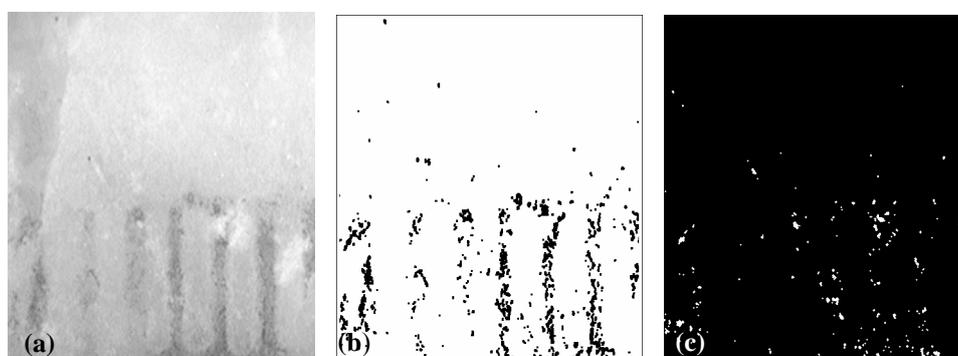
Figure 23: (a) Stone material (monitored by a refectography system (IR)) demonstrating cleaned and un-cleaned stripes, (b) Black particles detected, (c) white particles detected.

Optical inspection by the experts on Figs. (21), (22) and (23) verifies that the topology of the detected black particles, their spatial distribution as well as their shape and size closely resemble their own judgment of sporadic particle presence. A deeper inspection of Fig. 21(c) though, reveals the detection of white decay spots even on treated regions. According to the experts' assessment, these spots are associated to regions of material loss. The above false positive induction illustrates the inability of the system to distinguish between areas where gypsum prevails and other areas where material loss occurs. This limitation arises from its low magnification as it becomes difficult to view the inter particle area between adjacent black spots and thus to accurately locate areas of gypsum or $CaCO_3$ presence. A further explanation of the false positive detection of white spots is associated to the operation of the reflectography screening system. Thus, according to the literature [], when visible light impacts on a black surface or an opaque surface (cavity), it is almost totally absorbed producing black colored surface. In contrast, when the light impacts on a smooth and brilliant surface, it is reflected resulting in the formation of white regions. In our case, the areas corresponding to gypsum prevalence do not express smooth surfaces. In contrast, texture irregularities can be observed leading to diffusion of the impacting luminance and, as a consequent, only a fraction of the light is reflected resulting to the illustration of white areas. This phenomenon provides a further explanation for the inaccurate segmentation of white particles. In an attempt to consider in parallel the detection results depicted in figs. (21), (22) and (23) we can assess that infrared reflectography (IR) manages to segment decay patterns that are generally smaller in extent and arranged at closer distances. This modality tends to provide information close to the information provided by the FOM images. Some important features regarding the size and the compactness of the decay areas segmented through the different monitoring systems are presented in the following sub-section. Furthermore, visual inspection of the results derived through the three operations of reflectography system, reveals that the IR tends to approach more accurately the topology of prevalence of white particles. The latter characteristic of the IR reflectography may be associated to the wavelength of the impacting light and the inter-particle distance between the white particles.

As it is discussed in the introduction of this section, one of our aims is to investigate the potential of using optical monitoring methods in the description of corrosion damage. To validate this potential, the results of such methods are compared with the results derived by imaging the same surface via a FOM system. Fig. 24(a) shows a part of the stone material of 20(a) monitored by the FOM system while (b) and (c) depict the black and white particles detected superimposed on the original image.

Figure 24: (a) Stone material (monitored by a FOM system (magnificationx25)) demonstrating cleaned and un-cleaned stripes, (b) The detected black particles overlaid on the original surface, (c) The detected white particles overlaid on Fig. 12(a).

Experts' evaluation of the results reported in Fig. (24) reveals that the proposed automated method approaches accurately the topology of black spots while preserving information concerning their size and shape. Even the spatial distribution of deterioration patterns is in accordance with their own judgment of the sporadic presence of decay patterns into the matrix of encrustation.

4.6.2. Quantitative Results

The corrosion effects encountered are subsequently quantified by measuring the number of spots, the percentage of area covered by such spots and their average size and spatial distribution. In order to make the results comparable regarding the spatial distribution of spots obtained through the various monitoring systems, we measure the mean number of black spots encountered per cm² of the cleaned and un-cleaned areas of the stone material. Fig. 16 depicts the number of black particles detected on the surfaces studied in the previous subsection.
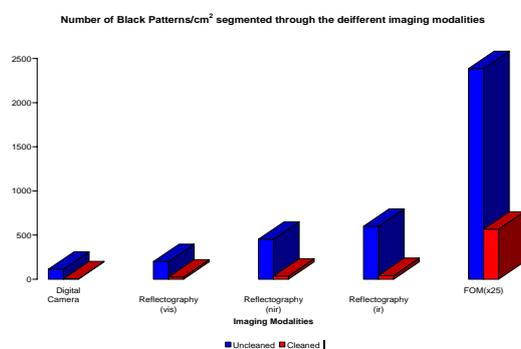

Figure 25: Number of black particles/cm² of stone surface as it was determined by the results of the different monitoring methods.

An obvious result that can be drawn from Fig. 25 is that the number of black patterns/cm² of stone surface encountered via the aid of the three monitoring methods is significantly reduced on cleaned areas. Moreover, it can be observed that FOM reveals a significantly higher number of detected black patterns. This can be explained by taking into account the high resolution provided by the FOM system, as well as its higher magnification rates. In other words, the FOM system demonstrates better discernibility in detecting the individual patterns prevailing in the structure of black crusts, while the other monitoring systems tend to merge adjacent degraded areas. The above assessment provides also a reason of the disability of the digital camera and reflectography to precisely determine the locations of white particles prevalence. A further noteworthy conclusion that can be drawn from figure 25 is that the reflectography system when operating at the NIR and IR spectral bands tends to derive results more close to the FOM. This characteristic may be associated to a better discriminability provided by the IR as the diffusion reflectance takes place at the sub-layer and not at the outer surface.

In order to overcome the different size of particles detected and rather focus on the overall effect on stone surfaces, figure (26) depicts the percentage of surface covered by black particles in the same cases as in Fig. (25).
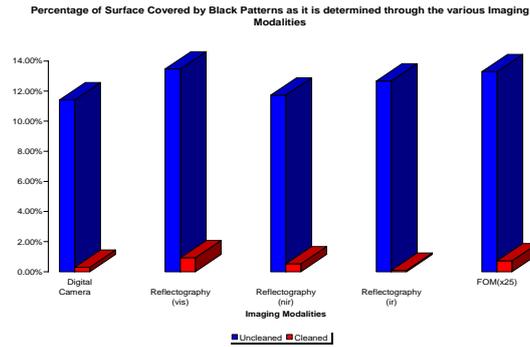
Figure 26: Percentage of surface covered by black patterns, as it is determined through the various monitoring methods.

The results in Fig. (26) reveal that the fraction of the stone material covered by black particles is determined to be similar through the various monitoring systems. More specifically, the FOM system as well as the Reflectography (vis/ ir/ nir) have only little deviation in the percentage of stone covered by black particles (13.29% & 13.47% 11.73% 12.67 respectively), while monitoring by the digital camera derives a fraction of the stone specimen equal to 11.42% to be covered by black patterns. By comparing together the results illustrated in Figs 25 & 26, it becomes obvious that despite the greater number of black particles detected via the FOM system, the percentages of surface associated with corrosion damage tend to converge for all the monitoring modalities. This observation reflects the potential of the FOM system to discriminate small adjacent decay areas, while these areas tend to appear as larger merged regions when they are viewed by other imaging systems. Furthermore, Fig. 26 reveals that cleaning achieves considerable elimination of black particles. The effect of cleaning is reflected by the significant reduction of both number and size of decay areas on the treated regions of the studied surfaces.

Another interesting quantitative measure concerning the size distribution of decay particles is summarized in Table 12. The distribution in terms of the median measure estimator rather than its mean measure is presented, since the actual size distribution on several images tested is heavily tailed. In particular, the measures computed are defined as follows:

- Median: The particle size that is greater than the 50% of the sizes detected on the image.

- Lower and Upper Quartiles: The particle sizes that are greater than 25% and 75% of sizes detected on the image, respectively.

Table 4-13: Size distribution of decay patterns encountered on the studied surfaces

	Lower Quartile (μm^2 of stone surf)	Median (μm^2 of stone surf)	Upper Quartile (μm^2 of stone surf)
Digital camera image (Uncleaned)	363.93×10^2	647×10^2	1294×10^2
Digital camera image (Cleaned)	242.62×10^2	364×10^2	485.25×10^2
Reflectography (vis) (Uncleaned)	186.27×10^2	479×10^2	878.16×10^2
Reflectography (vis) (Cleaned)	213.09×10^2	293×10^2	559.36×10^2
Reflectography (nir2) (Uncleaned)	203.69×10^2	331×10^2	636.53×10^2
Reflectography (nir2) (Cleaned)	76.38×10^2	178.23×10^2	280.07×10^2
Reflectography (ir2) (Uncleaned)	168.75×10^2	270×10^2	506.25×10^2
Reflectography (ir2) (Cleaned)	101.25×10^2	168.75×10^2	168.75×10^2
FOM images (Black Spots) (Uncleaned)	26.83×10^2	47.71×10^2	74.46×10^2
FOM images (Black Spots) (Cleaned)	20.92×10^2	34.74×10^2	52.11×10^2
FOM images (White Spots)(Uncleaned)	23.21×10^2	40.37×10^2	66.05×10^2

FOM images_(White Spots) (Cleaned)	14.17×10^2	22.4×10^2	42.49×10^2

The results reported in Table 13 indicate that the size distribution of black particles detected on uncleaned areas is spread to greater values than that detected on cleaned areas. This observation holds true for the surfaces monitored by the digital camera as well as for the surfaces monitored by the reflectography (vis, NIR and IR) and the FOM system. The above assessment indicates that the deterioration effects that cannot be removed by the cleaning process are, at least, reduced in extent. A significant point studying the results of Table 1 is the variation in the size distribution of black patterns through the different imaging systems. Thus, it can be observed that detection on the FOM images segments corroded areas that correspond to smaller in extent surfaces on the stone material. This also supports the previous statements regarding the differences in the ability of the employed monitoring systems to discriminate the prevalence of small black particles onto the body of black crusts. In essence, the variation in the size of detected areas arises from the magnification of the systems considered, which reflects their individual potential to discern two contiguous decayed areas as two individual patterns, without merging them. A further observation that can be drawn concerns the variation in the decay patterns sizes when operating the reflectography system at different spectral bands. Thus it is obvious that the size distribution of the decay patterns segmented in the IR case tends towards lower levels, while the NIR follows. This feature may be explained by taking under consideration the reflectance behaviour of the IR luminance structure. According to the literature, infrared light enters the body of the crust and thus the reflection takes place at an inner sub-layer and not exclusively at the surface layer thus resulting in providing more accurate information regarding the spatial arrangement of the particles. Figure 18 provides an illustration of the behavior of reflections.
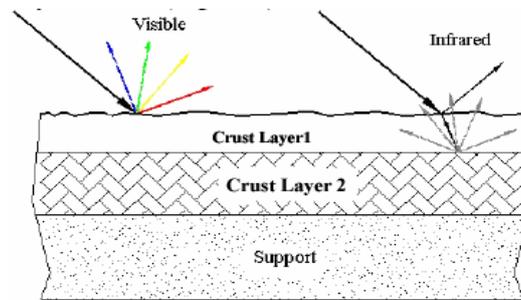

Figure 27: Behaviour of the reflections

Up to this point, we consider only area measurements on the detected decay patterns. Intensity information from the original image on these patterns is also of great concern, since it relates with the depth of the crust accumulated on the surface. In this study, the aspect of crust thickness is approached in a rather qualitative point of view, in that darker formation implies more light absorption and, thus, thicker formation of black crust. Since the intensity distribution does not reflect heavy tails and resembles better the normal distribution, the reported measures present the mean and the upper and lower quartiles of the intensity over the particles of all images of the same type considered. The intensity levels derived by studying the decay patterns segmented on images obtained by the IR and NIR reflectography images

are not considered in this study because of the severe intensity alteration caused due to the illumination at the infrared spectral band.

Table 4-14: Intensity distributions of black particles detected at various surfaces

	Lower Quartile	Mean	Upper Quartile
Digital camera image (Uncleaned)	76	83.50	88
Digital camera image (Cleaned)	100	103	124
Reflectography image (Uncleaned)	59	78	82
Reflectography image (Cleaned)	72	98	88
FOM images (Uncleaned)	72	81	92
FOM images (Cleaned)	86	97	102

The results presented in Table 14 indicate the change in intensity distribution before and after chemical cleaning. It is verified that the employed automated approach derives intensity distributions shifted to lower values when applied to surfaces with black crusts of higher thickness (untreated areas). Moreover, after chemical cleaning the intensity distribution of the detected particles is increased, since such areas are diminished and appear brighter and less disturbing in a macroscopic point of view. This result also indicates that, even though cleaning does not completely eliminate all decay formations, it manages to reduce the thickness of remaining crust patterns relative to their original state. A comparison regarding the intensity distribution derived by the various monitoring methods reveals similar results with small variations, mainly arising from the different conditions of illumination of each modality.

As it was mentioned before, this work aims not only at studying the efficiency of the various monitoring modalities in defining corrosion damage, but also at assessing the extent and the severity of the degradation effects. Such a study is conducted by quantifying the deterioration state on parts of the stone surface that have been submitted to laser cleaning with different parameters. Fig. (28) depicts the studied stone material. The treated and the untreated stripes are marked in the image. The (T) symbol indicates that the area was submitted to laser cleaning while (U) indicates that no cleaning intervention was performed. The treated surfaces were cleaned by different laser cleaning pulses resulting to more severe cleaning from left to right. We provide information concerning the cleaning efficiency at each application of the recruited intervention methods at each stone region through some statistical parameters. More specifically, the number of decay patterns detected per cm², the percentage of the surface covered by decay as well as the distribution of decay pattern sizes encountered on the studied surfaces is used as means for quantifying the degradation state and estimating the cleaning efficiency at each region of application. Subsequently, figs. (29) and (30) illustrate the corrosion state encountered on the stripes defined in Fig. 28. In particular, Fig. 29 presents the percentage of decay coverage of each of the studied stripes either cleaned or uncleaned. The results are presented in a logarithmic scale diagram, where the data points illustrated at the left correspond to cleaned areas while those depicted at the right correspond to uncleaned areas. A diagram of similar structure is also presented regarding the number of decay patterns (Fig. 20). As it can be observed, both diagrams provide information concerning both the treated and the untreated stripes, in order to assess the crusts homogeneity prior to the cleaning and[2] the cleaning efficiency at each level of cleaning parameters.
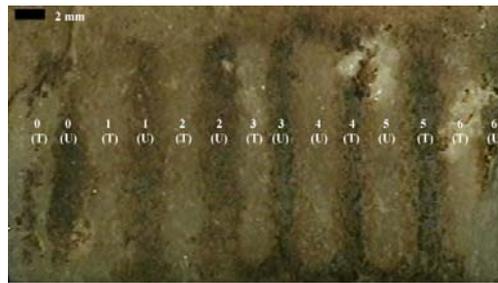
**Figure 28:** The studied specimen having its treated and untreated stripes labeled.
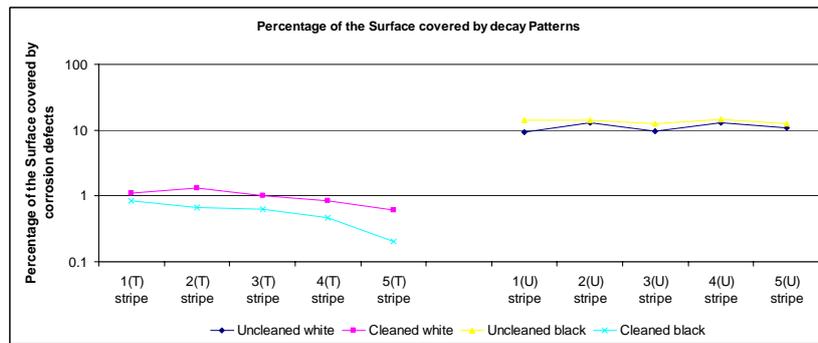
Figure 29: Percentage of the stripes covered by corrosion damage both in cleaned (stripe ia 1<=i<=5) and the uncleaned areas (stripe ib 1<=i<=5).
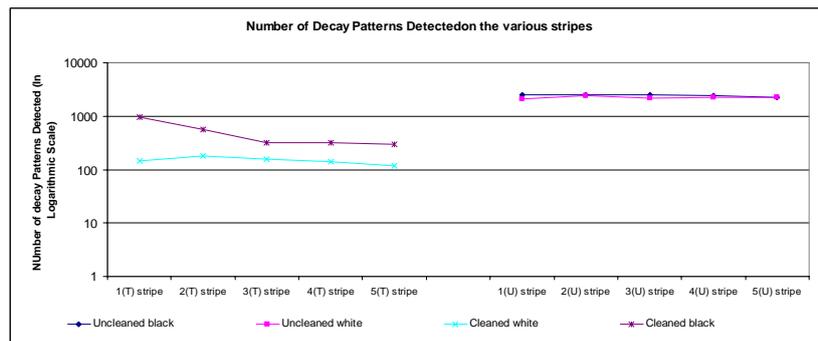

Figure 30: Number of decay patterns encountered on the treated (stripe i(T) 1<=i<=5) and the untreated (stripe i(U) 1<=i<=5) areas of the stone specimen.

An initial study of Figs (29) and (30) shows a significant reduction of the decay effects after the application of cleaning intervention. This assessment holds true for both the cases of black and white particles encountered and is reflected on the number and the percentage of surface covered by corroded regions. A further observation of the results demonstrated in Figs. (29) and (30) also verifies that the treated stripes demonstrate a different degree of degradation. As we move from stripes 0 and 1 to the subsequent stripes, a subsistent elimination of corrosion occurs reflected by the reduction of number and percentage of black particles. This quantitative estimation is in accordance with the experts' judgment regarding the effects of the cleaning, as these stripes were cleaned by the application of increasingly higher laser pulses. Furthermore, it can be observed in the diagrams, that prior to the cleaning a homogeneous crust layer prevails on the stone surface. This is verified by the similar statistical measures employed to quantify corrosion in uncleaned stripes (right part of diagrams). The size distribution of corrosion patterns is employed as a further means to estimate the cleaning effects or the degree of cleaning. In figure (31) we investigate associations between the degree of cleaning and the extent of the decay patterns encountered.
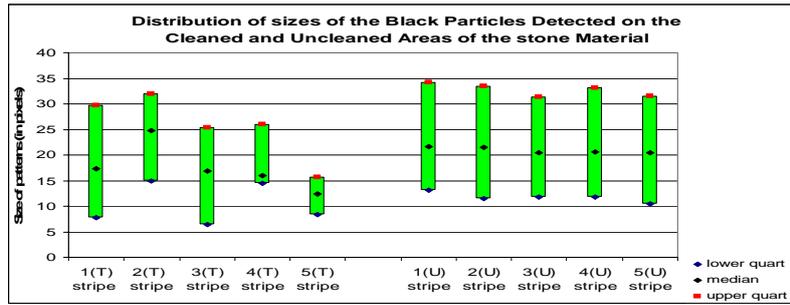
Figure 31: Distribution of the surfaces of black spots encountered in both cleaned (stripe ia 1<=i<=5) and the uncleaned areas (stripe ib 0<i<=5).

Figure (22) depicts a substantial shifting of the size distribution to lower values as we pass from stripes 1(T) and 2(T) to the consecutive stripes. The above diagram reveals that there might indeed exist an association between the size of the decay patterns and the cleaning state of the surface under consideration. In particular, it seems that the corrosion effects encountered on cleaned surfaces tend to be smaller in extent than those detected on untreated areas. An analogous observation can be drawn by observing the results concerning decay patterns on differently cleaned surfaces. Thus, we can detect larger in size patterns on surfaces less intensely cleaned. The fact that the distribution of decay patterns detected on the untreated regions display almost the same distribution of sizes supports the claim of homogeneous corrosion state before any treatment. The above Figs. (29, 30 and 31) illustrate the potential of the employed algorithmic approach to assess the corrosion rate as it can provide reliable results concerning even slight variation of the corrosion effects.

A further investigation on the segmentation results obtained by the different imaging modalities regards the compactness of the detected decay areas. Visual inspection of the segmentation results reveals that the decay patterns detected on the IR reflectography image tend to be less compact, in the sense that these areas usually contain holes. Through the subsequent section, we investigate this feature by measuring the distribution of the Euler number of spots detected in the entire image.

4.7. Observations regarding the patterns shape

The aim of this sub-section is to study the aspect of decay patterns' nestedness. The motivation of this investigation is the presence of hole-regions in the decay areas segmented in the digital camera image. In an attempt to investigate this aspect, we measure the Euler number on all the labels depicted on each image and subsequently extract statistical measures on the Euler number associated to segmented parts of the image. Namely we use the mean, standard deviation, median, quartiles and finally the percentage of the segmented areas that contain holes. Such a study provides further information on the potential of the monitoring systems to be used in an accurate detection procedure. At first, in table 15 we present the fraction of decay areas of each category (FOM, Digital Camera, Reflectography) that contain holes. This measure provides a general description of shape features in decay areas.

Table 4-15: Percentage of the segmented decay patterns containing holes into their areas

	2am_sample (Digital Camera Image)	2am sample Reflectography (Vis)	2am sample Reflectography (IR)	2am sample FOM (x25)	2am sample FOM (x50)	Flutingsx50_Sheltered (Diagnosis)
Percentage of labels depicting holes	9.53%	12.31%	16.52%	10.45%	7.22%	4.97%

By observing the results reported in table 15, we can conclude that black particles detected on the (IR) reflectography image tend to contain more holes (16.52% of the decay areas contain holes) than the corresponding black particles segmented on the reflectography (Vis) (12.31%) and the digital camera image (9.53%). Another assessment that can be drawn from table 15, is that black particles detected on the surface monitored under the FOM (x25) presents more surfaces that contain holes than the corresponding surfaces detected by the FOM (x50). A further assessment extracted from table 15 is that the decay patterns detected on sample 2am via the FOM(x50) imaging system tend to be less compact. In other words, more hole-regions are detected on these areas (7.22%), while the corresponding fraction for the flutingx50 (untreated) case is 4.97%. The latter feature may be associated with the type of decay that prevails on these surfaces. Such a conclusion though requires more elaborate study by the chemists and a larger collection of (FOM (x50) 2am sample) images.

In order to provide more detailed information regarding the occurrence of hole-regions on the studied decay areas, we present measures associated with the distribution of the Euler number concerning each surface. We remind that N(H) denotes the number of hole regions detected into an area R. The Euler Number regarding an area R is given by $R_{Euler\ Number} = 1 - N(H)$ and the mean Euler number concerning L areas is $Mean(R_{Euler\ Number}) = \frac{1}{L}\left\{\sum_{l=1}^{L}(1 - N(H))\right\}$.

Table 4-16: Features related to the distribution of Euler values

	2am_sample(Digital Camera	2am sample Reflectography	2am sample Reflectography (ir)	2am sample FOM(x25)	2am sample FOM(x50)	Flutingsx50_Sheltered(Diagnosis)
Min Euler Number	-2	-3	-5	-2	-2	-2
Max Euler Number	1	1	1	1	1	1
Mean Euler Number	0.885	0.790	0.687	0.871	0.919	0.958
Standard deviation of Euler Numbers	0.393	0.731	0.833	0.230	0.308	0.251

An observation of the results reported in table 15 reveals that the decay patterns detected on the reflectography (IR) image tend to present areas with more hole-regions while the reflectography (Vis) case follows. These variations may depict the discrimination ability of the specific monitoring modality.

5. Conclusions And Further Work

This thesis is geared towards investigating aspects of non-destructive detection and quantification of corrosion damage on stone surfaces. Through this approach, the studied surfaces are monitored via the aid of several imaging modalities. We use a Fiber Optics Microscope (FOM), a Digital Camera, and a Reflectography system operating at the visible, infrared, and near infrared spectral bands. Several algorithms are tested for segmentation of decay patterns. The segmented degradation patterns are classified into two broad categories identified under the terms "black particles" and "white particles". This discrimination is performed in accordance also to the chemical composition of the degraded areas. "Black particles" are associated with the presence of carbonaceous particles, alumino-silicates, dust, metal oxides and other pollutants embedded in the gypsum cavities. On the other hand, "white particles" represent decay patterns associated with gypsum crystals and re-crystallized $CaCO_3$.

One of the initial objectives of this work is to study the efficiency of the implemented algorithms in accurately determining the exact location of prevalence of decay patterns, as well as their size and shape features. The performance of the algorithmic schemes is assessed through studying the ROC and the Precision-Recall curves. The performance curves reveal that the efficiency of each algorithm in accurately extracting the location or the extent of decay areas is closely related to the individual characteristics of the studied image. Thus, it is assessed that some of the algorithms demonstrate better performance when processing images depicting texture irregularities, while the opposite occurs for other algorithmic approaches. According to the performance curves (both ROC and Precision-Recall) the Conditional Thickening along with the Region Growing algorithms approaches more accurately these aspects of segmentation (location, size and shape). The determination of an algorithm though that provides the most reliable information regarding the prevalence of corroded areas requires the employment of criteria that should be posed by the experts. Further to validating the performance of the algorithms, this work also aims at investigating the features of decay patterns that are either segmented by all the algorithms or they are detected by each algorithm separately and not by others (SDAs). According to this analysis the SDAs usually correspond to degraded areas that are smaller in extent and more isolated than the CDAs.

The efficiency and the effects of pollution and cleaning interventions was also an aspect of high importance for this work. Thus, we employ statistical tests in order to estimate whether the cleaning approaches attain to reduce the crusts' thickness or to eliminate the extent of corroded areas. More specifically, we study whether the intensity of black particles depends on the cleaning state or the location of the studied stone surfaces on the material structure. Through the statistical tests we measure the relative intensities of corroded areas over the background. According to the experts' judgment darker decay areas correspond to regions where black crusts of greater thickness prevail. The results derived from the statistical tests indicate that all the cleaning methods result in diminishing the crusts thickness and the extent of the

segmented corrosion damage. The DS cleaning approach though (anionic resin in combination with de-ionized water), seems to induce color alterations, as an effect of the removal of original stone material. Furthermore it is revealed that crusts of greater thickness are encountered on sheltered untreated flutings. This conclusion is in accordance to the experts' opinion.

Another significant characteristic, relative to the type of degradation particles, is their shape features. Throughout this work we measure shape features of the decay patterns in order to investigate whether these are altered due to structural or cleaning effects. According to shape descriptors, less compacted (their shape departs significantly from the circle) decay spots are encountered on unsheltered areas. These areas are also characterized by lower 1st and 2nd central moment values while they reflect rougher decay spots. The same shape features are also varying between cleaned and uncleaned surfaces. Further to the above, we also investigated whether the decay patterns segmented on the various studied surfaces tend to be oriented towards specific directions. In order to define the orientation of decay patterns, their *axis of least inertia* is extracted. The experiments reveal that the corroded areas segmented on unsheltered areas tend to be more oriented to specific directions. Finally the occurrence of hole-regions in the segmented degradation patterns is also investigated. According to the experts, this effect is closely related to the discontinuities encountered in black crusts and is arising due to the dissolution of gypsum by the walls' humidity. Our study indicated that decay areas with more hole-regions prevail on the sheltered untreated flutings, while the sheltered untreated reedings and the unsheltered untreated flutings preserve less decay regions with holes. This is in accordance to the experts' judgment since on these sheltered untreated flutings crusts of greater thickness prevail.

Finally we investigate the potential of various imaging modalities in accurately determining corrosion damage. The corrosion characteristics vary depending on the imaging modality examined. Nevertheless, the amount of degraded stone surface derived from our analysis scheme on any modality tested converges to almost the same percentage. Thus, for degradation evaluation purposes even a simple camera imaging modality can be used with quite accurate results. Besides the quantification of degradation effects, our analysis scheme can be used for an evaluation of the cleaning efficiency. As it is assessed by the results derived, the number of the black particles is significantly lower on cleaned areas. The comparison of results derived by the different monitoring systems provides an overall assessment of the potential and the limitations of optical inspection (provided by the digital camera and the reflectography systems) in the reliable estimation of the corrosion damage.

Further investigations should be focused towards classifying the segmented degraded areas according to their type and their location of prevalence. This would be an extension of the current work, as we have already assessed the occurrence of associations between the degraded patterns on specific surfaces with some parameters of their shape. Such approaches are applicable not only to FOM images but also to images that provide macroscopic information (Reflectography, CCD cameras etc). Thus, the various types of corrosion damage i.e flaws; material loss, black crusts etc could be classified automatically. The issue of tracking decay

changes through the time course is also an aspect that can be approached through computational methods, as it would provide a non-destructive in-situ method for assessing the degradation process. Finally, the implementation of automated approaches for measuring (numerically) the crust thickness also becomes an aspect for further study.

References

[1]   D. Camuffo, Deterioration processes of historical monuments, in: T.S. Schneider (Ed), Acidification and its Policy Implications, Elsevier, Amsterdam (1986), pp. 189–211.

[2]   G. Zappia, C. Sabbioni, G. Gobbi, Non-carbonate carbon content on black and white areas of damaged stone monuments, Atmospheric Environment A 27 (1993). 1117–1121.

[3]   S.C. Jimenez, Deposition of airborne organic pollutants on historic buildings, Atmospheric Environment B 27 (1993). 77–85.

[4]   A.G. Nord, T. Ericsson, Chemical analysis of thin black layers of building stone, Studies in Conservation 38 (1993). 22–35.

[5]   J.R. Cheng, J.R. Hwu, J.T. Kim, S.M. Leu, Deterioration of marble structures, Analytical Chemistry 59 (1987).104–106.

[6]   C. Sabbioni, G. Zappia, G. Gobbi, M.G. Pauri, Deterioration of ancient and modern buildings materials due to environmental factors, in: Structural Repair and Maintenance of Historical Buildings III, Proceeding (1993), pp. 235–242.

[7]   A.J. Hutchinson, J.B. Johnson, G.E. Thompson, G.C. Wood, The role of fly-ash particulate material and oxide catalysts in stone degradation, Atmospheric Environment A 26 (1992). 2795–2803.

[8]   P. Ausset, J.L. Crovisier, M.Del. Monte, V. Furlan, F. Girardet, C. Hammecker, D. Jeannette, R. Lefevre, Experimental study of limestone and sandstone sulphation in polluted realistic conditions: the Lausanne Atmospheric Simulation Chamber (LASC)., Atmospheric Environment 30 (1996). 3197–3207.

[9]   Baer, N. S. Sabbioni, C. and Sors, A. I. (1991) *Science, Technology and Europeun Cultural Heritage.* Butterworth-Heinemann, Oxford.

[10]   Leysen, L., Roekens E. , and Van Grieken, R. (1989) Air pollution-induced chemical decay of a sandy-limestone cathedral in Belgium. *Science qf the Total Environment 78, 263-287.*

[11]   Sabbioni. C., Zappia, G., Ghedini, N. and Gobbi, G. Carbon due to atmospheric deposition on stone monuments and historical buildings. *Proceedings* of the *8th International Congress on Deterioration and Conservation of Stone,* Berlin. 30 September-4 October 1996, pp. 333-337*.*

[13]   Primerano P., Pasquale Di. S., Mavilia L., Coriglaino F. Possible Alteration of monuments caused by particles emitted into the atmosphere carrying strong primary acidity.

[14]   Warscheid, Th., Krumbein, W.E., Biodeterioration of inorganic nonmetallic materials-general aspects and selected cases. In: Heitz, H., Sand, W., Flemming, H.C. (Eds.), Microbially Induced Corrosion of Materials. Springer, Berlin (1996), pp. 273-295.

[15]   Moropoulou A., Avdelidis N., "IRT in the investigation of buildings and historic structures" Thermosense XXVI, vol. 5405,

[16]   Lebrun V, Bonino E, Nivart JF, Pirard E. Development of specific acquisition techniques for field imaging-applications to outcrops and marbles. In: Proceedings of the International Symposium of Geovision on Imaging Applications in Geology. Belgium: Liege, 1999. p. 165-168.

[18]   Gelli D., Vitulano D., "Speed Up of Shape from Shading Using Graduated Non-convexity" 11th International Conference, DGCI 2003, Naples, Italy, November 19-21, 2003. Proceedings Editors: Ingela Nyström, Gabriella Sanniti di Baja, Stina Svensson

[19]   Moltedo L, Mortelliti G, Salvetti O, Vitulano D. Computer aided analysis of buildings. J Cult Herit 2000; 1(1): 59-67.

[20]   Boukouvalas C, De Natale F, De Toni G, Kittler J, Marik R, Mirmehdi M, Petrou M, Le Roy P, Salgari R, Vernazza G. ASSIST: Automatic System for Surface Inspection and Sorting of Tiles. J Mater Process Tech 1998; 82 (1-3):179-188.

[21]   Mahadevan S., Casasent D., "Detection of triple junction parameters in microscope images", Proc. SPIE Vol. 4387, p. 204-214

[22]   Pappas M, Pitas I. Old painting digital color restoration. In: Noblesse Workshop on Non-Linear Model Based Image Analysis. Scotland: Glasgow, 1998. p. 188-192

[23]   V. Kokla, A. Psarrou, and V. Konstantinou, "Image processing a helpful tool in the art conservation" In: Art 2002.

[24]   Crinzato E., Vavilov V., "Corosion evaluation by thermal image processing and 3D Modeling", Rev. Cen. Therm., vol. 37, pp. 669-679, (1998).

[25]   Gros X.E., Bousique J., Takahashi K., " NDT data Fusion at pixel level", NDT &E International, vol. 32, pp. 283-292, (1999).

[26]   Choi KY, Kim SS. Morphological analysis and classification of surface corrosion damage by digital image processing, Corros. Sci. 2005; 47(1): 1-15

[27A]   Kim K.M.: Design of a Binary Decision Tree for Recognition of the Defect Patterns of Cold Mill Strip Using Generic Algorithm. In: Innovations in Applied Artificial Intelligence: 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE 2004, Ottawa, Canada, May 17-20, 2004; Volume 3029

[27]   Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, "Machine Vision" McGraw-Hill International Editions, Computer Science Series

[28]   Dengler J, Behrens S, Desaga JF. Segmentation of micro-calcifications in mammograms. IEEE Trans. Medical Imaging 1993; 12: 634-642.

[29]   H.Christensen and W. Forstner, "Performance characteristics of vision Algorithms" Mach. Vis. Applicat., vol. 9, no. 5-6, pp. 215-218, 1997.

[30]   H. Cristensen and P. Philips, " Performance Characterization in Computer Vision". Singapore: World Scientific 2002

[31]   K. Bowyer and P. Philips, "Empirical Evaluation Techniques in Computer Vision". Los Alamitos, CA: IEEE Comp. Soc. Press, 1998.

[32]   M. Viergever, "Performance Characterization and Evaluation of Computer Vision Algorithms." Boston, MA: Kluwer, 2000.

[33]   P. Courtney, N. Thacker, A. Clark, "Algorithmic modeling for performance evaluation," Mach. Vis. Applicat., vol. 9, pp. 219-228, 5-6, 1997.

[34]   M. Heath, S. Sakar, T. Sanocki, and K. Bowyer, "A robust visual method for assessing the relative performance of edge detection algorithms," IEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 1338-1359, Dec. 1997.

[35]   Dengsheng Zhang, Guojun Lu, "Review of shape Representation and Description techniques." Pattern Recognition, vol. 37, pp. 1-19, 2004.

[36]   H. Soltanian-Zadeh, F. Rafiee-Rad, S.Pourabdollah-NejadD, "Comparison of multiwavelet, wavelet, Haralick and shape features of microcalcification classification in mammograms.", Pattern Recognition, vol. 37, pp. 1973-1986, 2004.

[37]   K. Abter, W.E. Snyder, H. Burkhardt, G. Hirzinger, "Application of affine-invariant Fourier descriptors to recognition of 3-d objects.", IEEE Trans. Pattern Anal. Mach. Intell. 12 (7) (1990) 640-647

[38]   E.L. Brill, "Character recognition via Fourier descriptors.", WESCON, Session 25, Qualitative Pattern Recognition Through Image Shaping, Los Angeles, CA, (25/3) 1-10, 1968,

[39]   G. Eichmann, et al., "Shape Representation by Gabor Expansion", Hybrid Image and Signal Processing II, SPIE vol. 1297, Orlando, Florida, USA 1990, pp. 86-94.

[40]   C.C. Lin, R. Chellappa, "Classification of partial 2-D shapes using Fourier descriptors.", IEEE Trans. Pattern Anal. Mach. Intell. 9 (5) (1987) 686-690.

[41]   O.R. Mitchel, T.A. Grogan, "Global and partial shape discrimination for computer vision.", Opt. Eng. 23 (5) (1984) 484-491.

[42]   T.W. Rauber, "Two-dimensional shape description.", Technical Report: GR UNINOVA-RT-10-94, University Nova de Lisboa, Portugal, 1994.

[43]   C.W. Richard, H. Hemami, "Identification of 3-dimensional objects using Fourier descriptors of the boundary curve", IEEE Trans. System Man Cybernet. SMC-4 (4) (1974) 371-378.

[44]   Y.Rui, A.C. She, T.S. Huang, "A modified Fourier descriptor for shape matching in MARS", In: A.W.M Smeulders, R. Jain(Eds.), Image Databases and Multimedia Search, World Scientific Publishing Co., Singapore, 1997, pp. 165-177.

[45]   D.S. Zhang, G.Lu, "A comparison of shape retrieval using Fourier descriptors and short-time Fourier descriptors", in: Proceedings of the Second IEEE Pacific-Rim Conference on Multimedia (PCM01), Beljing, China, October 24-26, 2001, pp. 855-860.

[46]   J.R. Ohm, F.B. Bunjamin, W. Liebsch, B. Makai, K. Muller, A. Somlie, D. Zier, "A set of visual feature descriptors and their combination in a low-level descriptor scheme", Signal Process, Image Commun. 16 (2000) 157-179.

[47]   Q.M. Tieng, W.W. Boles, "Recognition of 2D object contours using the wavelet transform zero-crossing representation", IEEE Trans. Pattern Anal. Mach. Intell. 19 (8) (1997) 910-916.

[48]   H.S. Yang, S.U. Lee, K.M. Lee, "Recognition of 2D object contours using starting-point-independent wavelet coefficient matching", J. Visual Commun. Image Represent. 9 (2) (1998) 171-178.

[49]   H.Freeman, "On the encoding of arbitrary geometric configurations", IRE Trans. Electron. Comput. EC-10 (1961) 260-268.

[50]   H. Freeman, A. Saghri, "Generalized chain codes for planar curves", in: Proceedings of the Fourth International Joint Conference on Pattern Recognition, Kyoto, Japan, November 7-10, 1978, pp. 701-703.

[51]   J. Iivarinen, A. Visa, "Shape recognition of irregular objects", In: D.P. Casanent (Ed.), Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, Proc. PIE 2904 (1996) 25-32

[52]   A. Hoover, G. Jean-Baptiste, X. Jiang, P. Flynn, H. Bunke, D. Goldof, K.Bowyer, D. Eggert, A. Fitzgibbon and R. Fisher, "An experimental comparison of range image segmentation algorithms", IEEE Trans. Pattern Anal. Machine Intell, vol. 18, pp. 673-689, July 1996.

[53]   Balas, C., Papadakis, V., Papadakis, N., Papadakis, A., Vazgiouraki, E., Themelis, G. 2003. A Novel Hyper-Spectral Imaging Apparatus for the Non-Destructive Analysis of Objects of Artistic and Historic Value. J. of Cultural Heritage 4(S1): 330-337.

[54]   Moropoulou A, Koui M, Delegou ET, Bakolas A, Karoglou M, Rapti E, Kouris S, Mauridis Th, Ypsilanti E, Avdelidis NP, Ntae D. Conservation interventions planning for the main façade of the National Archaeological Museum. In: Final Research Working Paper, Lab of Materials Science and Engineering, National Technical University of Athens, August, 2002.

[55]   Koui M, Delegou ET, Dai D, Rapti E, Mavridis T, Moropoulou A. FTIR for the assessment of cleaning interventions on Pentelic marble surfaces. In: 6th International Symposium on the Conservation of Monuments in the Mediterranean Basin. Lisbon, CD-Rom Proc., 2004.