IPSJ Transactions on Computer Vision and Applications Vol.7 189–200 (Dec. 2015)

[DOI: 10.2197/ipsjtcva.7.189]

Research Paper

Dithering-based Sampling and Weighted α-shapes for Local Feature Detection

Christos Varytimidis[1,a) Konstantinos Rapantzikos[1,b)
Yannis Avrithis[1,c) Stefanos Kollias[1,d)

Received: February 13, 2015, Accepted: September 14, 2015, Released: December 16, 2015

Abstract: Local feature detection has been an essential part of many methods for computer vision applications like large scale image retrieval, object detection, or tracking. Recently, structure-guided feature detectors have been proposed, exploiting image edges to accurately capture local shape. Among them, the WαSH detector [Varytimidis et al., 2012] starts from sampling binary edges and exploits α-shapes, a computational geometry representation that describes local shape in different scales. In this work, we propose a novel image sampling method, based on dithering smooth image functions other than intensity. Samples are extracted on image contours representing the underlying shapes, with sampling density determined by image functions like the gradient or Hessian response, rather than being fixed. We thoroughly evaluate the parameters of the method, and achieve state-of-the-art performance on a series of matching and retrieval experiments.

Keywords: image sampling, dithering, local features

1. Introduction

Local features are found as a core component in many algorithms solving computer vision problems like image retrieval, image classification, object detection, or 3D reconstruction. They provide a sparse representation of images while capturing salient points or regions like corners and blobs. Local feature detectors provide invariance to image transformations, repeatability and computational efficiency compared to dense features, e.g., on a regular grid. Assigning local descriptors (e.g., SIFT [8]) to detected features, creates a compact and robust image representation.

Popular detectors like the Hessian-Affine [10] and SURF [2] are based on image gradients, while others like the MSER [9] are purely based on image intensity. All of them have been successfully applied to a variety of applications, but often the balance between quality and performance remains an issue. For example, the image coverage of the Hessian-Affine detector is limited, since—for a given threshold—multiple detections appear on nearby spatial locations at different scales. The MSER detector is fast, but often extracts sparse regular regions that are not representative enough. SURF is also fast, but detections are often not stable enough.

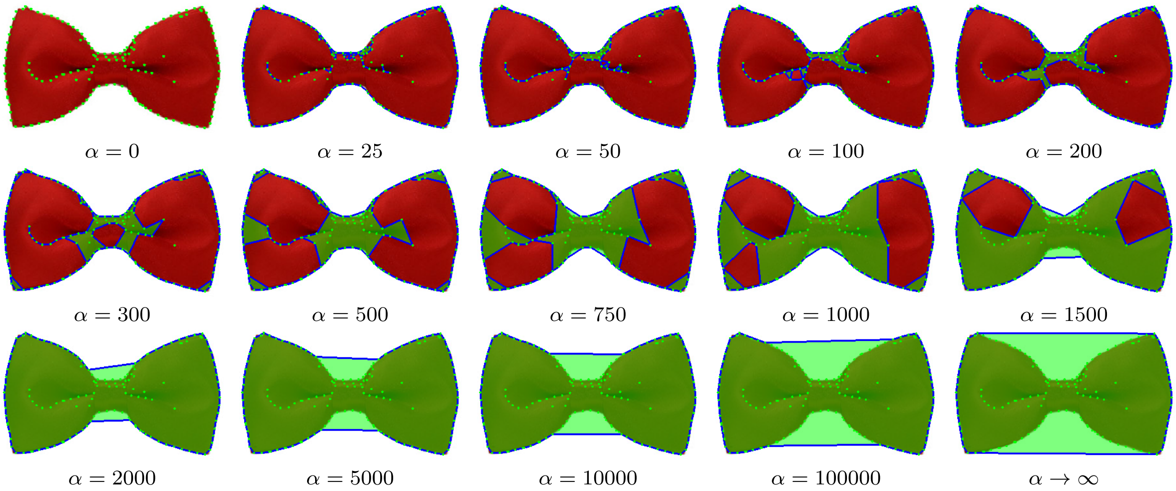
Although not so popular, another family of detectors is based on image edges, which are naturally more stable than gradient, e.g., to lighting changes. The recently introduced WαSH detector [18] belongs to this family and is based on grouping edge samples using the weighted *α-shapes*, a well known representation in computational geometry. A weakness of WαSH is that edge sampling is roughly uniform along edges, with a fixed sampling interval d. In an attempt to overcome this limitation, we propose a different sampling scheme that relies directly upon smooth image functions. We demonstrate its efficiency by common statistics on image matching and retrieval experiments.

2. Related Work and Contribution

Edge-based local features have not become popular due to the lack of stable edges (e.g., under varying viewpoint) and computational inefficiency. One of the earliest attempts, the *edge-based region detector* (EBR), starts from corner points and exploits nearby edges by measuring photometric quantities across them. It is suitable for well-structured scenes (like e.g., buildings), but not for generic matching [11]. Mikolajczyk et al. [12] propose an edge-based detector that starts from densely sampled edge points combined with automatic scale selection and use it for object recognition. Starting also from dense edge samples, Rapantzikos et al. [17] compute the binary distance transform and detect regions by grouping its local maxima, guided by the gradient strength of nearby edges.

Indirectly related to edges are the methods that exploit gradient strength across them by avoiding the thresholding step. Zitnick et al. [23] apply an oriented filter bank to the input image and detect *edge foci* (EF), i.e., points that are roughly equidistant from edgels with orientations perpendicular to the points. The idea is quite interesting, but computationally expensive. Avrithis and Rapantzikos [1] compute the weighted medial axis transform directly from image gradient, partition it and select associated regions as *medial features* (MFD) by taking both contrast and shape into account. Although those methods exploit richer image information compared to binary edges, gradient strength is often quite sensitive to lighting and scale variations.

[1] National Technical University of Athens, Greece
[a) chrisvar@image.ntua.gr
[b) rap@image.ntua.gr
[c) iavr@image.ntua.gr
[d) stefanos@cs.ntua.gr



$\alpha = 0$ $\alpha = 25$ $\alpha = 50$ $\alpha = 100$ $\alpha = 200$

$\alpha = 300$ $\alpha = 500$ $\alpha = 750$ $\alpha = 1000$ $\alpha = 1500$

$\alpha = 2000$ $\alpha = 5000$ $\alpha = 10000$ $\alpha = 100000$ $\alpha \to \infty$

Fig. 1 Example of the different α-shapes created over an image, given a set of points. The first shape (for $\alpha = 0$) is the set of points, growing up to the full convex hull (for $\alpha \to \infty$).

The recently proposed WαSH detector [18] combines edge-sampling and grouping towards distinctive local features supported by shape-preserving regions. It is based on weighted α-shapes on uniformly sampled edges, i.e., a representation of triangulated edge samples parametrized by a single parameter α. WαSH uses a *regular triangulation*, where each sample is assigned a weight originating from the image domain. Despite this rich representation, WαSH is limited by its uniform sampling scheme, which is not stable under varying viewpoint.

Recently, we introduced two sampling methods [19] that are based on the well known Floyd-Steinberg algorithm [6]. The latter was the first of the *error-diffusion* dithering approaches, where the idea is to produce a pattern of pixels such that the average intensity over regions in the output bit-map is approximately the same as the average over the same region in the original image. Error-diffusion algorithms compare the pixel intensity values with a fixed threshold and the resulting error between the output value and the original value is distributed to neighboring pixels according to pre-defined weights. The main advantages of these algorithms are the simplicity combined with fairly good overall visual quality of the produced binary images.

The Floyd-Steinberg algorithm has been extensively studied in the literature. Indicatively, Ostromoukhov [13] and Zhuand and Fang [22] have addressed the limitations of the initial algorithm, like the visual artifacts in highlights/dark areas and the appearance of visually unpleasant regular structures, using intensity-dependent variable diffusion coefficients. Recently, Pang et al. [14] proposed an iterative structure-aware image dithering algorithm that preserves local texture, but involves a computationally prohibitive optimization. Nevertheless, we use the initial algorithm because of its computational efficiency and the nature of our problem, which is sampling rather than halftoning. In addition, we apply dithering on functions other than image intensity, where dithering artifacts are eliminated.

Our work is also related to the work of Gu et al. [7], who detect local features as local minima and maxima of the β-stable Laplacian. They combine the local features in order to create a higher level representation, resembling the constellation model [5], [7]. However, we do not detect our sample points as features; we rather use them to initialize the WαSH feature detector.

In our previous work [19] we introduced two sampling methods of variable density, and presented evaluation results on image matching and retrieval applications. In this work, we investigate the impact of sampling parameters to the size of the point set, as well as the representation quality, measured by the performance of WαSH. We also consider the impact of using weights on samples, which changes the form of the constructed triangulation in WαSH.

3. Background

3.1 WαSH Detector

The WαSH feature detector [18] is based on *α-shapes*, a representation of a point set P in two dimensions, parametrized by scalar α. The construction of α-shapes is based on a triangulation $\mathcal{R}$ of P, exploiting geometrical properties of triangles and edges. In fact, α-shapes are a generalization of the convex hull, and are not convex or even connected in general. Starting from the set P for $\alpha = 0$, triangles and edges of the triangulation are added to the shape as α increases (see **Fig. 1**). Finally, the shape converges to the convex hull of the point set P as $\alpha \to \infty$.

In the simplest case, α-shapes use an underlying Delaunay triangulation, but *weighted α-shapes* in WαSH [18] use the *regular triangulation* instead. The latter is a generalization of Delaunay where each point in P is assigned a non-negative *weight*, hence capturing more information from the image domain. In practice, weight is a function of image gradient in WαSH.

The inclusion of simplices σ (edges or triangles) of the triangulation $\mathcal{R}$ in α-shapes is controlled by assigning a *size* quantity $\rho_T \geq 0$ to every simplex σ_T, which is a function of the positions and weights of its vertices $T \subseteq P$. The *weighted α-complex* of P is the subset of triangulation $\mathcal{R}$ containing all simplices up to a given size $\alpha \geq 0$,

$$\mathcal{R}_\alpha = \{\sigma_T \in \mathcal{R} : \rho_T < \alpha\}, \qquad (1)$$

which is neither convex nor connected, in general. Finally, the *weighted α-shape* of P [3] is the union of all such simplices,

$$\mathcal{W}_\alpha = \bigcup_{\sigma \in \mathcal{R}_\alpha} \sigma. \qquad (2)$$

In the evolution of α-shapes, small triangles corresponding to fine details of the image are added first, while large triangles corresponding to coarse parts are added latter.

In order to select regions and extract prominent local features [18], the WαSH detector exploits the *upper α-shapes*. The latter are complementary to α-shapes, having larger triangles and edges included first. Simplices are ordered by decreasing size in order to form the *upper α-complex*

$$\overline{\mathcal{R}}_\alpha = \{\sigma_T \in \mathcal{R} : \rho_T \geq \alpha\}. \qquad (3)$$

For each α value and upper α-shape instance, we define as *connected components* the disjoint parts of the α-shape. A *component tree* is used to track the evolution of connected components as simplices are added to form larger regions. Different connected components of the α-shapes are potentially selected as features during evolution, according to a shape-driven strength measure. The features correspond to blob-like regions that respect local image boundaries. Features are also extracted on cavities of image objects as well as regions that are not fully bounded by edges.

One important limitation of WαSH is the sampling process, which is restricted to points sampled *uniformly* along binary image edges. Even though binary edges often coincide with object contours, noisy edges can lead to sampling on both object boundaries and textured regions. In addition, sampling is uniform, using a fixed *sampling step d* along edges that corresponds to distance between samples measured in pixels, hence the representation scale is fixed. In a single image, objects of diverse scales have different representations: too dense on large objects, and too sparse on small ones. Though this may be partially compensated for by subsequent processes, the sampling step parameter d is still needed to control the density of samples along edges. Further, uniform sampling naturally leads to severe undersampling of highly curved paths, so important details of prominent shapes may be lost.

Using weights on the sampled points brings more information from the image domain to the triangulation. This is expected to provide a more accurate representation of the boundaries, leading to well fitted local features. In the experimental section, we will also evaluate the impact on performance when using either weighted points with the regular triangulation, or unweighted points with a Delaunay triangulation.

In Section 4 we introduce the alternative methods for sampling that apply to smooth functions of the image rather than binary edge maps and provide variable density samples. For the remaining process including the component tree and feature selection, we keep the same choices as in WαSH [18].

3.2 Image Dithering

Dithering is a common method used for image binarization. It has been extensively used to compress images, or display grayscale images in binary monitors. Color image dithering is also possible, where high color depth images are converted to low depth. It is used by monitors capable of displaying a limited number of different colors, or by older image file formats. Image dithering is equivalent to halftoning, a technique used in black and white or color printers in order to accomplish high quality results, despite printing a limited number of different colors.

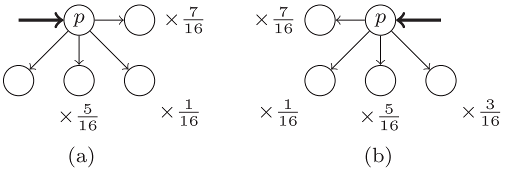


Fig. 2 Error diffusion coefficients used by the Floyd-Steinberg algorithm, when parsing pixels in (a) left-to-right, or (b) right-to-left order.

Image dithering consists of a thresholding step, followed by error diffusion of each thresholded pixel. Each pixel at position $p = (x, y)$ is visited at least once in a specific order and its intensity value $I(x, y)$ is compared to a threshold θ, such that

$$I'(x, y) = \begin{cases} 1 & \text{if } I(x, y) > \theta, \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

The error $e(x, y) = I(x, y) - I'(x, y)$ between the output I' and the input I is diffused to a neighborhood of pixel (x, y).

The most commonly used algorithm for image dithering is the one introduced by Floyd and Steinberg [6]. Image pixels are visited only once, in a serpentine order, left-to-right and right-to-left alternatively. The error $e(x, y)$ is diffused to the 4 pixels of the 8-connected neighborhood of pixel $p = (x, y)$ that are not yet visited, using the coefficients shown in **Fig. 2**. As discussed in Section 2, we use this algorithm for its simplicity and computational efficiency.

4. Dithering-based Sampling

In this section we propose two image sampling methods based on error-diffusion. The goal is to adapt the spatial density of samples over the image and achieve a sparse representation without compromising structure preservation. Removing the limitation of samples belonging to binary edges, we expect to get a triangulated set of sparse samples that fits well with the underlying image structure.

In our framework, the Floyd-Steinberg algorithm is not applied directly to the image intensity, but to a scalar function $s(x, y)$ over the image domain. The two methods we introduce are based on two different choices for $s(x, y)$. In both cases, the extracted samples are the nonzero points of the binary output of the Floyd-Steinberg algorithm.

Each sample point (x, y) is assigned a weight that is proportional to the sampled function $s(x, y)$. These weights are used in the remaining steps of the WαSH detector, in order to create the regular triangulation and weighted α-shapes. In the special case when weights are zero, the triangulation reduces to Delaunay.

4.1 Gradient-based Dithering

The gradient strength G of an image I is obtained by convolving with the gradient of a Gaussian kernel $g(\sigma)$ of standard deviation σ,

$$G = |\nabla g(\sigma) * I|. \qquad (5)$$

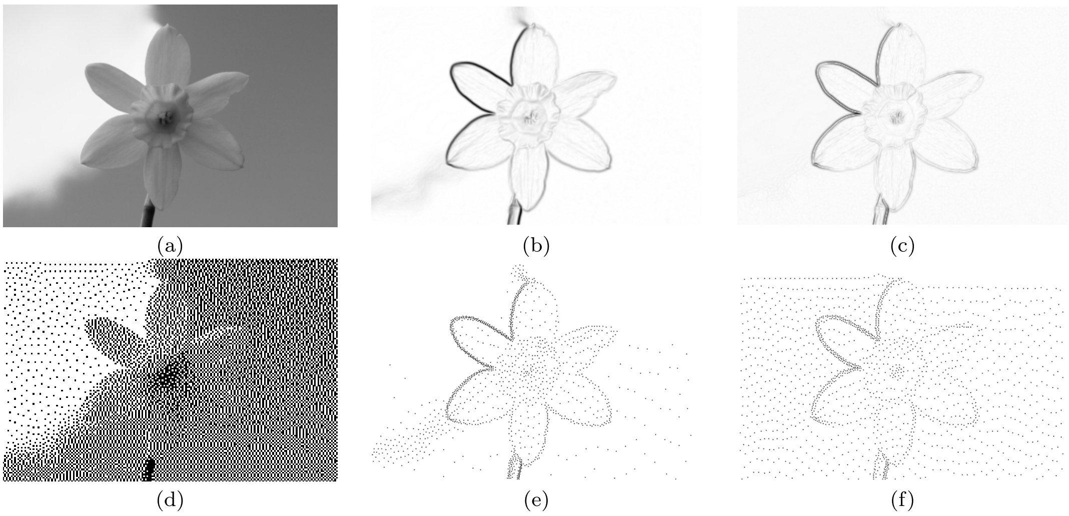


Fig. 3 Dithering-based sampling. (a) Input image and (d) Floyd-Steinberg dithering on (a). (b) Normalized gradient strength $\hat{G}$ and (e) sampling on $\hat{G}$. (c) Hessian response $\hat{\lambda}_1$ and (f) sampling on $\hat{\lambda}_1$. Figure is optimized for screen viewing.

Then, similar to Yang et al. [21], if $\hat{G}(x, y)$ is the gradient strength at point (x, y) normalized to $[0, 1]$, we use the non-linear function

$$s(x, y) = \hat{G}(x, y)^{\gamma} \qquad (6)$$

to represent image boundaries, where γ is a positive constant. Error-diffusion is performed using the Floyd-Steinberg algorithm on $s(x, y)$ rather than image intensity $I(x, y)$. Increasing the value of γ results in sparser sampling.

In smooth regions of the image, e.g., in the interior of objects or on smooth background, G is low and samples are sparse, resulting in large triangles. Near image edges or corners on the other hand, G is high, samples are dense, and a finer tessellation is generated that captures important details. Variable sample density offers a computational advantage without compromising the descriptive power of the triangulation.

4.2 Hessian-based Dithering

Instead of using the gradient strength as the input to error-diffusion, Yang et al. [21] use the largest eigenvalue of the Hessian matrix at each point. We also explore this option for our sampling.

If $H(x, y)$ is the Hessian matrix at point (x, y), again after filtering with Gaussian kernel $g(\sigma)$, let $\lambda_1(x, y)$ be its largest eigenvalue. It is known that λ_1 is the largest second order directional derivative of I. Similarly to Eq. (6), if $\hat{\lambda}_1(x, y)$ is the largest eigenvalue normalized to $[0, 1]$, we use function

$$s(x, y) = \hat{\lambda}_1(x, y)^{\gamma} \qquad (7)$$

to represent image boundaries, again performing error-diffusion on $s(x, y)$.

The magnitude of the second order derivatives increases near image edges, so the error-diffusion algorithm will favour dense sampling at these regions. However, samples will now appear more scattered at both sides of an edge, making the triangulation more complex. At smooth areas, sampling is sparse, but since the Hessian is more sensitive to noise a grid-like sampling can occur (see **Fig. 3** (f)). Compared to the gradient-based sampling, the number of detected features is often lower (see Section 5).

4.3 Examples – Discussion

A visual example of the sampling methods is shown in Fig. 3. Figures 3 (b), (e) depict the normalized gradient strength $\hat{G}$ and the resulting gradient-based sampling. Notice the sparsity of the samples in smooth areas and the density in structured ones. Figures 3 (c), (f) depict the Hessian response $\hat{\lambda}_1$ and the resulting sampling. Few weak edges are lost within the background noise in this case. For all examples we set $\gamma = 1$.

Figure 4 shows an example on a detail of an image along with different sampling methods and the resulting triangulations. The uniformly sampled edges are sparse and well distributed along the edges, but lose details at the corners and highly curved edge parts. On the other hand, the dithering-based methods are denser, but preserve the underlying structure better. In the Hessian-based approach, points are located around edges that—depending on the application—may prove useful at better reconstructing the underlying image, using only information from the samples. On the other hand, for the gradient-based approach, points are sampled on strong gradients, corresponding to object boundaries. The gradient-based sampling is expected to fit better with WαSH, given that WαSH is biased towards blob-like regions surrounded by object boundaries.

Examples of the features detected using either the baseline sampling of WαSH or the proposed sampling methods are depicted in **Figs. 5**, **6**. In each example, the number of detected features for each method is approximately the same (around 350 for Fig. 5 and 50 for Fig. 6). In Fig. 5 we present the matched features between two images of the *graffiti* dataset of Mikolajczyk et al. [11], using SIFT descriptors. Using the gradient-based sampling method, more detailed regions of the image are captured and matched, while using the Hessian-based sampling results in more matches between bigger blobs. In Fig. 6, the input image comes from the PASCAL VOC 2007 test set [4], a dataset heavily used for evaluating object recognition algorithms. Again the

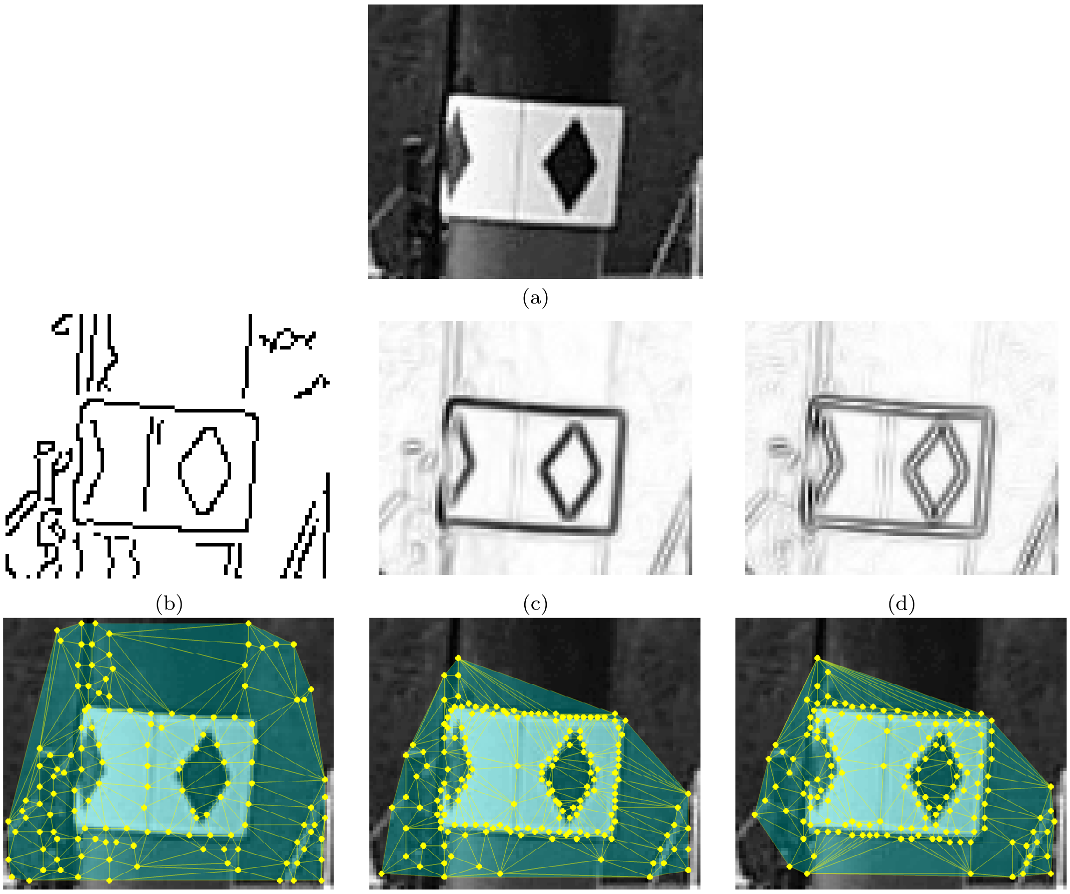


Fig. 4 Example of the different sampling methods and the corresponding triangulations. (a) Input image, a detail of the first image of the boat sequence of Mikolajczyk et al. [11] (see Section 5.2.1). (b) Binary edge map and (e) uniform sampling on (b). (c) Normalized gradient strength and (f) error-diffusion on (c). (d) Hessian response and (g) error-diffusion on (d). (b, c, d) are shown in negative for better viewing and printing.

dithering-based variants capture finer details of the image that can boost the performance in recognition tasks (see the ceiling lamp and the chairs).

5. Experiments

We evaluate the proposed sampling methods and compare to the state-of-the-art using two different experimental setups. The first is the matching experiment proposed by Mikolajczyk et al. [11] on the corresponding well-known dataset. We measure the *repeatability* and *matching score* of WαSH when using the proposed sampling methods, and also compare to other state-of-the-art detectors.

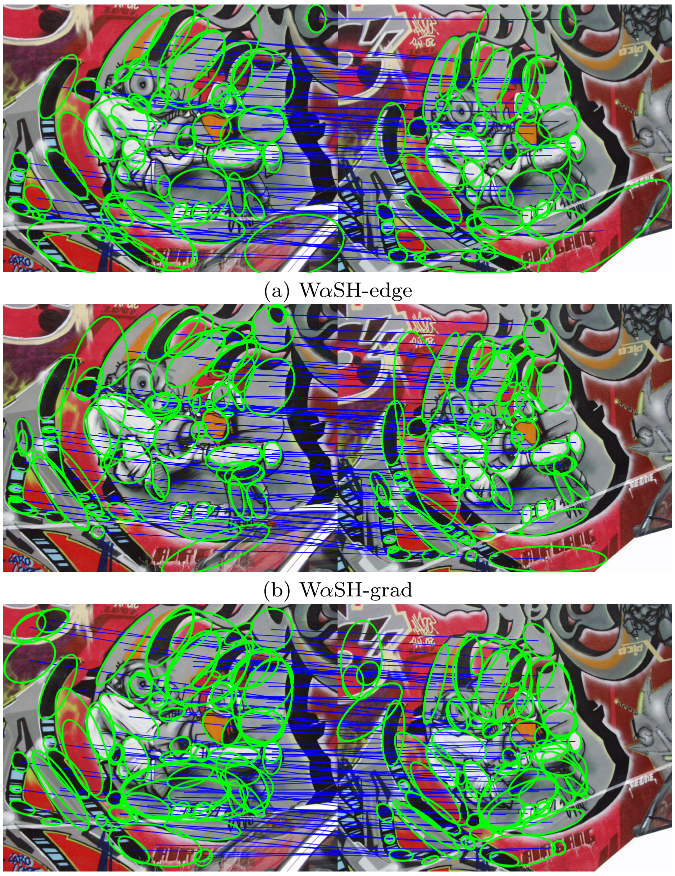
The second experimental setup involves a large scale image retrieval application on the *Oxford 5K* [15] and *Paris* [16] datasets. Both datasets consist of images of buildings, and diverse urban images as distractors. The performance is measured by *mean average precision* (mAP). For each proposed variant of WαSH we adapt the feature selection threshold to extract approximately 7.5×10^6 features for all images of the Oxford dataset, the same number as baseline WαSH [18]. For all different detectors we extract 128-dimensional SIFT descriptors and create visual vocabularies, using approximate k-means. We use the simple bag-of-words (BoW) representation, as well as a spatial re-ranking of the results, using fast spatial matching (FastSM) [15].

Initially, we perform an extensive evaluation of the parameters of the proposed sampling methods. We first investigate the impact of non-linearity γ to (a) the number of samples, (b) the time needed for WαSH algorithm to extract features and (c) the performance of the image retrieval experimen, using the Oxford dataset. Given an optimal value for γ, we also evaluate the effect of threshold θ used in the error diffusion step of the proposed sampling methods. To compare with the edge sampling used in WαSH, we also evaluate the performance when using different sampling steps d. Finally, in an attempt to speed up the algorithm, we examine the use of unweighted samples for WαSH.

After determining the optimal values for the parameters of the proposed methods, we compare to the state-of-the-art feature detectors using both experimental setups. For the image retrieval experiment we use both *Oxford* and *Paris* datasets. We use the Oxford dataset to tune all parameters of the sampling methods. Keeping the parameters and visual vocabularies fixed, we compare to the state-of-the-art on the Paris dataset.

5.1 Parameter Evaluation

In order to investigate the impact of different parameters of the proposed sampling methods described in Section 3.2, we evalu-

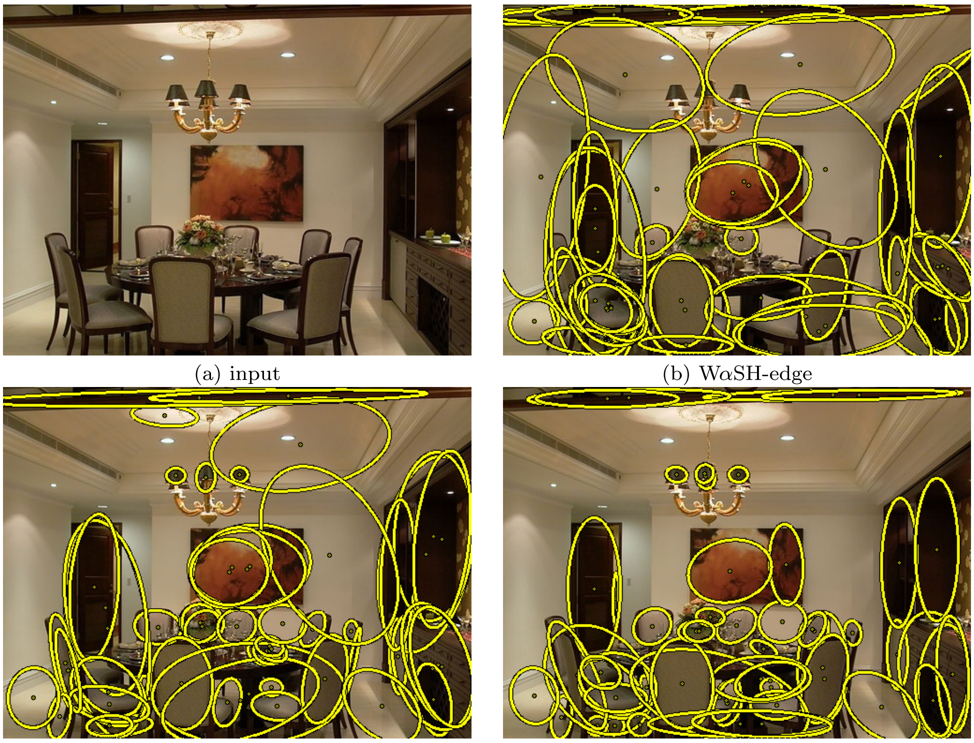


(a) WαSH-edge

(b) WαSH-grad

(c) WαSH-Hess

Fig. 5 Matched features on the *graffiti* dataset using: (a) baseline WαSH with uniform edge sampling, (b) the gradient-based sampling and (c) the Hessian-based sampling.



(a) input

(b) WαSH-edge

(c) WαSH-grad

(d) WαSH-Hess

Fig. 6 Example of local features detection. (a) Input image and (b) baseline WαSH results using uniform sampling. (c) Results using the gradient-based sampling and (d) using the Hessian-based sampling.

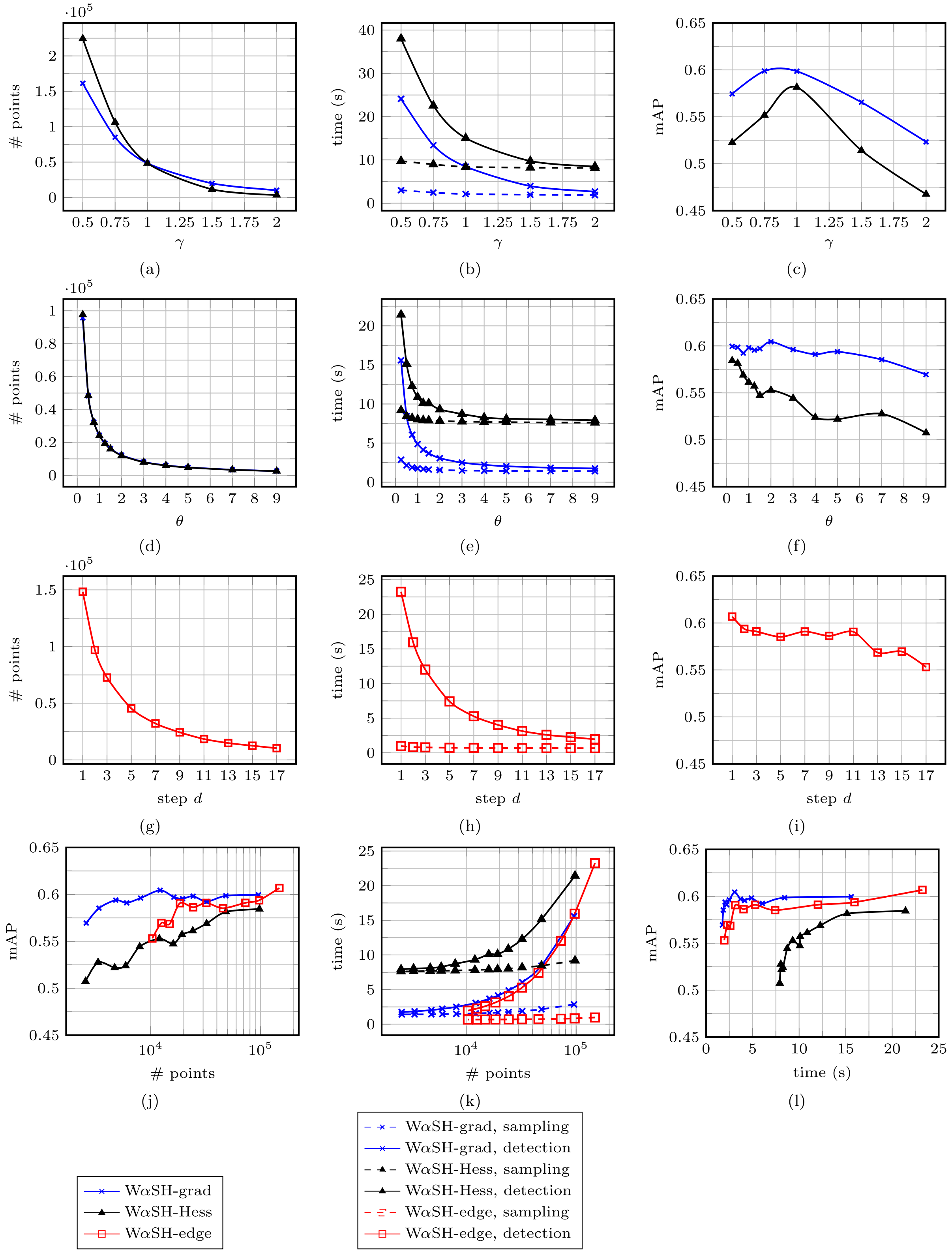
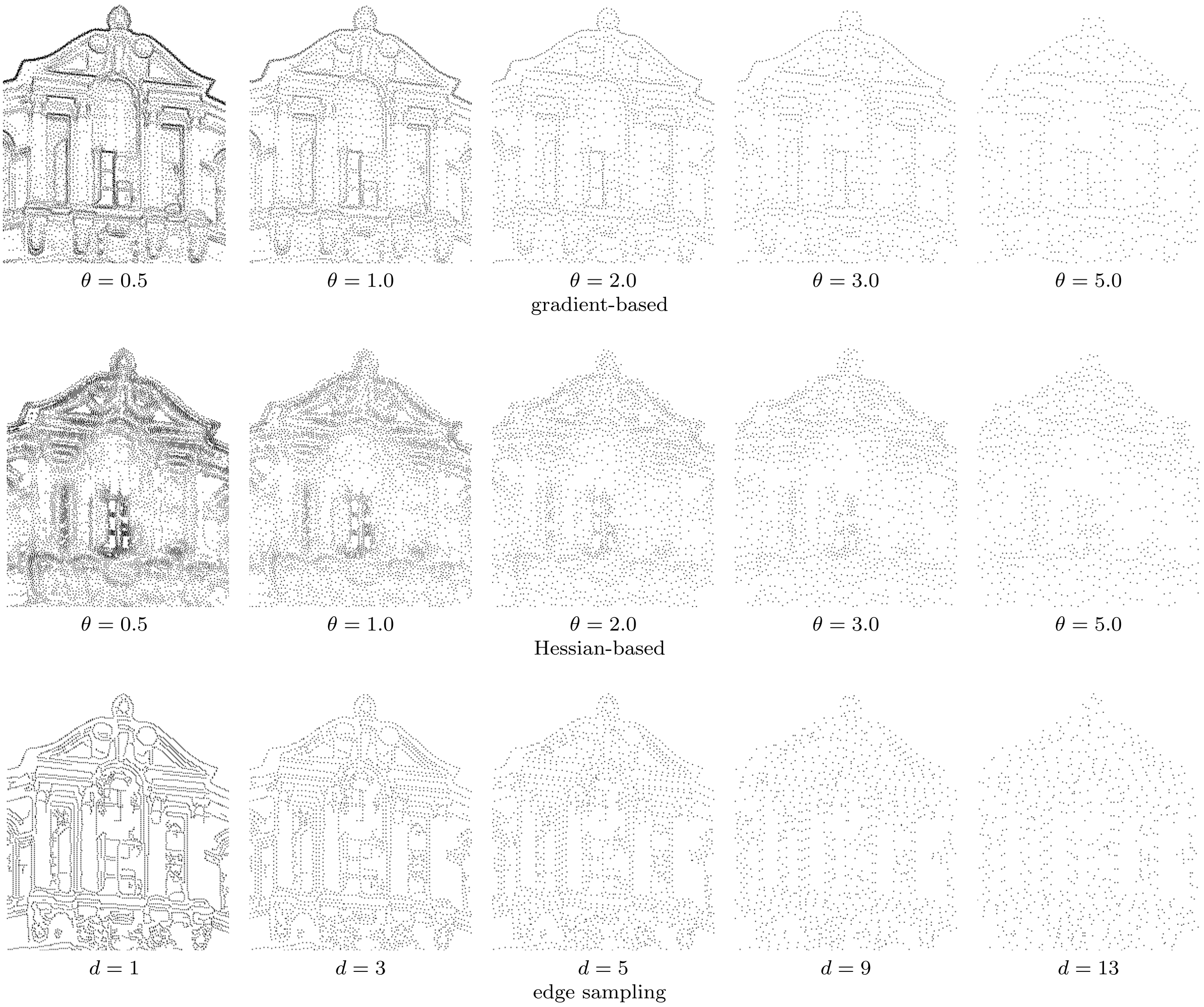


Fig. 7 Parameter evaluation of the different sampling methods. For the proposed methods, we evaluate the impact of γ (first row) and θ (second row), while for edge sampling the sampling step d (third row). In the last row we compare the performance and time complexity of the different sampling methods, based on the number of sample points.

ate the performance of WαSH on the large scale image retrieval experiment, using the Oxford dataset. In this setup we create vocabularies of 200K visual words and represent images with the BoW model.

Non-linearity γ in Eqs. (6), (7) affects the behavior of function $s(x, y)$. For $\gamma > 1$, low values of $s(x, y)$ decrease further, since $\hat{G}(x, y)$ and $\hat{\lambda}_1(x, y)$ are normalized in $[0, 1]$. This makes $s(x, y)$ more selective and less smooth, looking rather like a binary image map. Less points are sampled for a given threshold, concentrated near image edges. On the other hand, low values of



$\theta = 0.5$ $\theta = 1.0$ $\theta = 2.0$ $\theta = 3.0$ $\theta = 5.0$
gradient-based

$\theta = 0.5$ $\theta = 1.0$ $\theta = 2.0$ $\theta = 3.0$ $\theta = 5.0$
Hessian-based

$d = 1$ $d = 3$ $d = 5$ $d = 9$ $d = 13$
edge sampling

Fig. 8 Examples of samplings using the proposed methods for different threshold values. For uniform edge sampling we adjust the sampling step d.

$s(x, y)$ increase for $\gamma < 1$; $s(x, y)$ is smoother and more points are sampled over the entire image plane.

For different values of $\gamma \in [0.5, 2.0]$, we examine the average number of sample points extracted from the images of the dataset (see **Fig. 7** (a)), as well as the time taken by WαSH to extract local features (see Fig. 7 (b)). We also perform an image retrieval experiment for each value of γ, in order to measure the influence on the performance of WαSH (see Fig. 7 (c)).

The impact of γ to the number of samples, as well as the computation time of WαSH, is significant, although sampling time is not severely affected itself. Low values lead to sampling a large number of points and creating more complex representations (triangulation and α-shapes).

It turns out that the performance of WαSH as measured by mAP in the image retrieval experiment with the Oxford dataset is maximised for $\gamma = 1$. This is the linear case, where the sparsity of the representation and the complexity of the detector are balanced. The results of the gradient-based and the Hessian-based sampling schemes agree in this respect. For the following experiments we set $\gamma = 1$ for both. However, we keep the nonlinear term since it may be useful in boosting performance on other datasets, or in sampling for applications other than WαSH.

Threshold θ of the dithering algorithm directly controls the number of samples extracted, making sampling sparser as θ increases (see Fig. 7 (d)). The sample density significantly affects the feature detection time (see Fig. 7 (e)).

For the Hessian-based sampling, performance drops as sample density decreases (see Fig. 7 (f)). We select $\theta = 0.5$ in order to maintain the high performance, despite the computational cost. For the gradient-based method performance is not highly affected until $\theta = 5$, having a maximum for $\theta = 2$, which we select for the rest of the experiments.

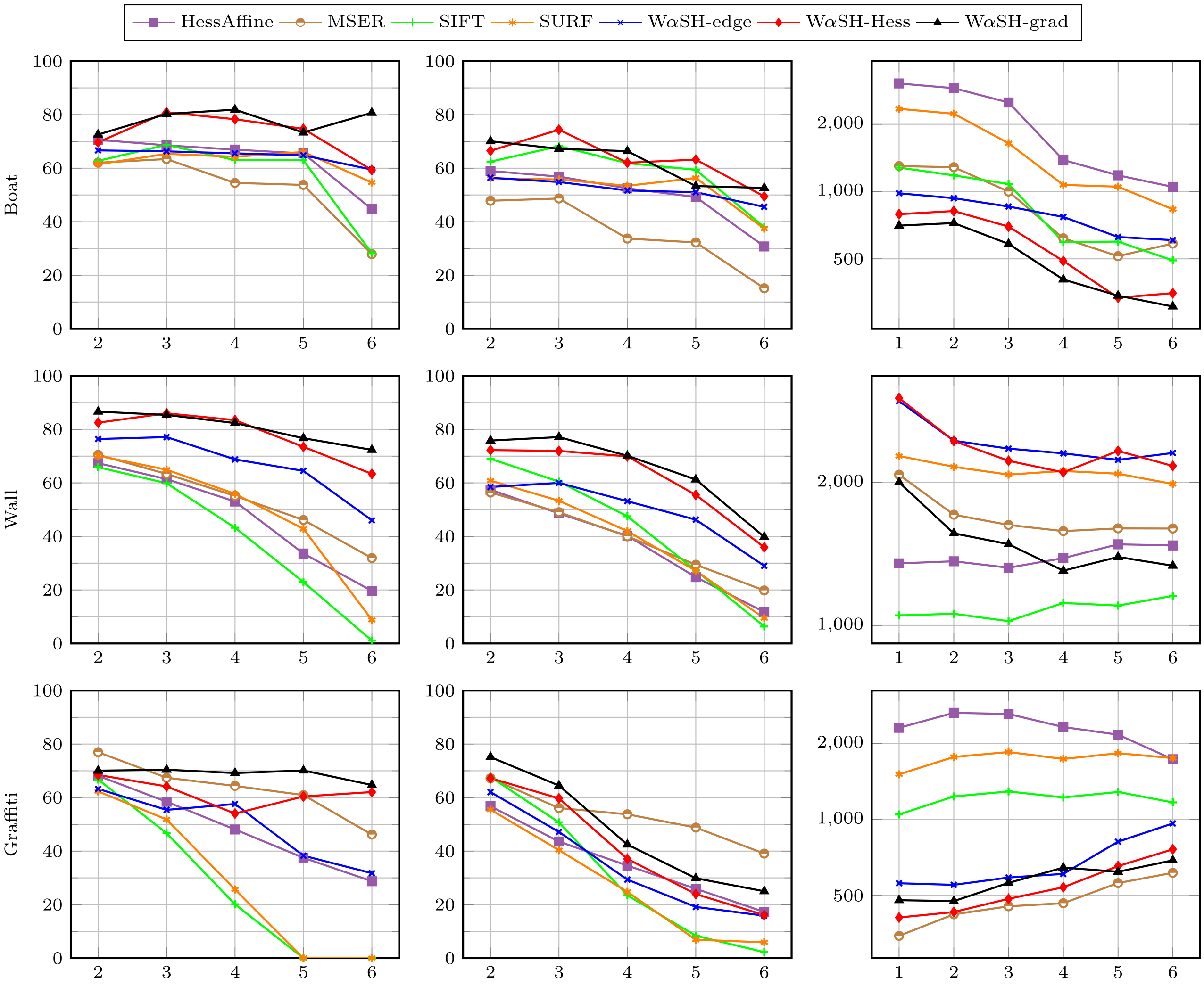
Figure 8 shows different samplings in a detail of an image of *Oxford 5K* dataset. For all methods, sampling density is decreased from left to right. Using gradient-based sampling, structures depicted in the image remain prominent even for sparse samplings.

Comparison to edge sampling. In order to examine the impact of the number of samples, we compare the proposed sampling methods to the original uniform edge sampling used in WαSH. For the latter, we let the sampling step vary in the interval [1, 17].

The number of samples decreases exponentially with the sampling step d (see Fig. 7 (g)), which affects the feature detec-

Table 1 Results of the image retrieval experiment on the *Oxford* dataset, using 3 different vocabularies, the Bag-of-Words model and spatial reranking of the results, measuring mean Average Precision.

detector		features	detection	Bag-of-Words (mAP)			ReRanking (mAP)		
		($\times 10^6$)	time (s)	50K	100K	200K	50K	100K	200K
HessAff		29.02	6.54	0.483	0.539	0.573	0.518	0.577	**0.607**
MSER		13.33	**0.40**	0.487	0.534	0.565	0.519	0.569	0.595
SIFT		11.13	5.24	0.422	0.465	0.495	0.441	0.486	0.517
SURF		**6.84**	0.43	0.465	0.526	0.574	0.509	0.573	0.603
weighted	WαSH-edge	7.66	3.14	**0.542**	**0.583**	0.591	0.530	0.573	0.590
	WαSH-grad	7.59	3.07	0.532	0.575	**0.605**	**0.543**	**0.581**	0.599
	WαSH-Hess	7.30	15.14	0.507	0.559	0.582	0.515	0.555	0.570
unweighted	WαSH-edge	7.41	3.04	0.507	0.547	0.583	0.507	0.552	0.581
	WαSH-grad	7.42	2.91	0.537	0.569	0.598	0.539	0.565	0.591
	WαSH-Hess	7.30	14.14	0.506	0.545	0.569	0.499	0.535	0.564



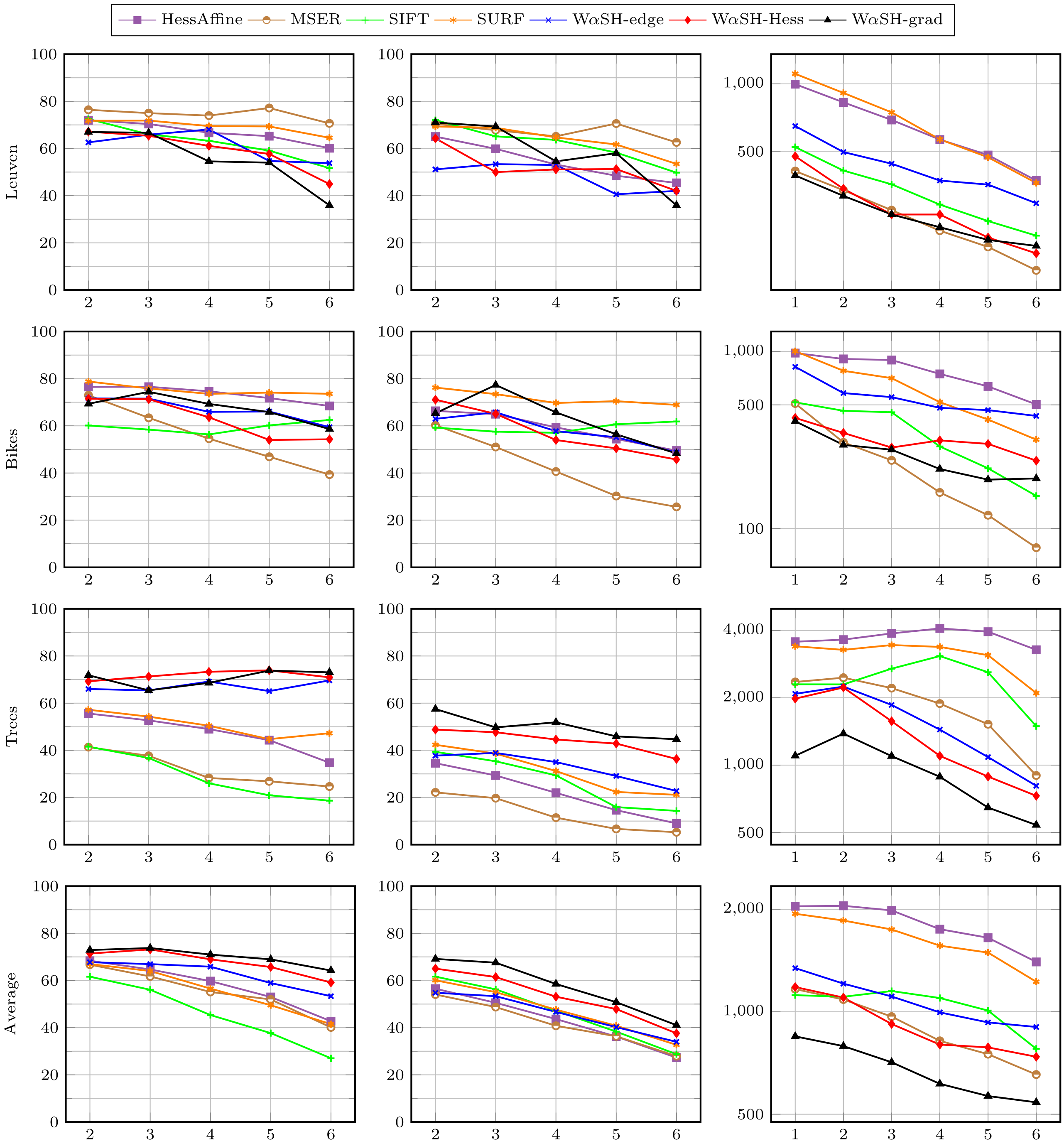
(a) repeatability (b) matching score (c) #features

Fig. 9 Comparison of our proposed sampling methods to baseline WαSH and the state-of-the-art in sequences *boat, wall* and *graffiti*. #features: number of features detected per image. Hess: Hessian-based dithering; grad: gradient-based dithering.

tion time accordingly, though not the sampling time itself (see Fig. 7 (h)). Performance of WαSH is quite stable until $d = 11$ and drops after that (see Fig. 7 (i)). In order to obtain high performance in a reasonable time, we select $d = 11$ for the remaining experiments.

In order to compare all sampling methods, we examine how the performance and detection time of WαSH are affected by changing the number of sample points (see Fig. 7 (j), (k)). We also examine the performance *vs.* feature detection time in Fig. 7 (l). Gradient-based sampling outperforms the other two methods, while keeping the computational cost low. On the other hand, Hessian-based sampling is inferior and significantly slower.

Sample weights. **Table 1** shows retrieval results for different sampling methods and weighted *vs.* unweighted samples. The cost of WαSH is higher when using weighted samples, as more computations are required to constructing both the regular triangulation and the α-shapes. However, the increase is not higher than 5%. On the other hand, the additional information of the



(a) repeatability (b) matching score (c) #features

Fig. 10 Comparison of our proposed sampling methods to baseline WαSH and the state-of-the-art in sequences *leuven, bikes* and *trees*, together with the averaged values over the dataset.

weights yields a performance gain of up to 2% for dithering-based sampling and up to 4% for edge-based uniform sampling.

5.2 Comparison to the State-of-the-art

In this section, we evaluate the performance of the proposed detectors against the state-of-the-art in both experimental setups.

5.2.1 Repeatability and Matching Score

We evaluate the performance on the matching experiment using the proposed sampling methods on WαSH. We also compare to the state-of-the-art detectors, Hessian-Affine, MSER, SIFT and SURF, for which we use the executables provided by the corresponding authors and default parameters, apart from SIFT where we use the implementation provided by *VLFeat* [20]. The image sets used, evaluate the impact of changes in viewpoint, rotation, zoom, blur and illumination. For the matching score we use 128-dimensional SIFT descriptors for all detectors, apart from SURF, which performs best using the corresponding descriptor.

The results of the evaluation are depicted in **Figs. 9**, **10**. The last row of Fig. 10 shows the average scores for the 6 datasets. Along with the repeatability and matching score, we also provide the number of features detected. Overall, the gradient-based sampling performs best, followed by the Hessian-based one.

Table 2 Results of the image retrieval experiment, on the *Paris* dataset. Performance is evaluated by mean Average Precision.

detector	features	Bag-of-Words (mAP)			ReRanking (mAP)		
	($\times 10^6$)	50K	100K	200K	50K	100K	200K
HessAff	36.14	0.467	0.491	0.507	0.479	0.500	**0.517**
MSER	17.33	0.465	0.485	0.497	0.480	0.499	0.503
SIFT	25.54	0.476	0.492	0.492	0.457	0.457	0.476
SURF	**8.56**	0.458	0.479	0.487	0.471	0.486	0.493
WαSH-edge	9.01	0.454	0.459	0.457	0.449	0.455	0.455
WαSH-grad	9.35	**0.497**	**0.509**	**0.511**	**0.498**	**0.506**	0.510
WαSH-Hess	9.32	0.477	0.474	0.478	0.468	0.469	0.476

5.2.2 Image Retrieval

In this experiment, we compare the proposed variants of WαSH to the state-of-the-art, on the image retrieval application using the *Oxford 5K* and *Paris* datasets. Similarly to the matching experiment, we compare against Hessian-affine, MSER, SIFT and SURF, using the corresponding executables with default parameters and SIFT descriptors for all detectors apart from SURF.

Similarly to the parameter tuning experiments, for the Oxford dataset we adapt the selection thresholds for the different versions proposed, in order to extract approximately the same number of features as baseline WαSH. For all detectors we create 3 different vocabularies of size 50K, 100K and 200K visual words. We compare performance on both the bag-of-words baseline and the spatial reranking of the results. The results are shown in Table 1.

The number of features extracted by each detector is critical for the large scale retrieval applications, affecting the indexing time and memory needed to store the inverted files, while using a lower number of features typically drops performance. SURF extracted the least number of features, followed by the baseline WαSH and our variants. Despite the low number of features, SURF and baseline WαSH perform comparably to Hessian-affine. Increasing the size of the vocabulary boosted the performance of all detectors. The gradient-based variant we propose outperformed all other detectors when using the spatial verification step, a result that verifies the previous findings. Without the spatial verification step performance is comparable with the edge-based WαSH.

Finally, we compare the same detectors using the Paris dataset. All detector parameters as well as the feature selection thresholds are kept fixed to the values used for the Oxford dataset. Visual words are extracted using the vocabularies created for the Oxford dataset. We only evaluate WαSH using weighted samples, as it outperformed the unweighted case in the parameter evaluation. The results are shown in **Table 2** and confirm the previous findings. SURF features were the least for the whole dataset, with the number of WαSH features being very close. Using the Hessian-based sampling outperformed the uniform sampling, while the gradient-based method performed best, exceeding the state-of-the-art.

6. Conclusions

In this paper we extend the recently introduced WαSH detector by proposing different image sampling methods. Image sampling is the first step of the algorithm and changes the qualities of the detected features, together with the overall performance of the detector. We propose two different image sampling methods that build on ideas from image halftoning. In that direction, we sample points based on error diffusion of smooth image functions. We provide a thorough parameter evaluation for the proposed methods, and compare to state-of-the-art feature detectors in a matching and an image retrieval experiment.

The proposed sampling methods, combined with the α-shapes grouping, result in a more accurate representation of the image structures. The detected features capture finer image structures, while keeping the high image coverage of the baseline method. Using the gradient-based scheme, samples are extracted on strong image gradients that capture object boundaries, boosting performance without increasing the computational cost. The method based on Hessian response provides competitive performance, but is computationally more expensive.

References

[1] Avrithis, Y. and Rapantzikos, K.: The medial feature detector: Stable regions from image boundaries, *International Conference on Computer Vision* (*ICCV*), pp.1724–1731 (2011).

[2] Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L.: Speeded-Up Robust Features (SURF), *Computer Vision and Image Understanding* (*CVIU*), Vol.110, pp.346–359 (2008).

[3] Edelsbrunner, H.: Alpha Shapes — A Survey, *Tessellations in the Sciences: Virtues, Techniques and Applications of Geometric Tilings*, Springer-Verlag (2010).

[4] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J. and Zisserman, A.: The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results, available from ⟨http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html⟩ (2003).

[5] Fergus, R., Perona, P. and Zisserman, A.: Object Class Recognition By Unsupervised Scale-Invariant Learning, *Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.Ii–264, IEEE (2003).

[6] Floyd, R.W. and Steinberg, L.: An adaptive algorithm for spatial grayscale, *Proc. Society of Information Display*, Vol.17, pp.75–77 (1976).

[7] Gu, S., Zheng, Y. and Tomasi, C.: Critical Nets and Beta-Stable Features for Image Matching, *European Conference on Computer Vision*, pp.663–676, Springer Berlin Heidelberg (2010).

[8] Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision* (*IJCV*), Vol.60, No.2, pp.91–110 (2004).

[9] Matas, J., Chum, O., Urban, M. and Pajdla, T.: Robust Wide-Baseline Stereo From Maximally Stable Extremal Regions, *Image and Vision Computing*, Vol.22, No.10, pp.761–767 (2004).

[10] Mikolajczyk, K. and Schmid, C.: An affine invariant interest point detector, *European Conference on Computer Vision* (*ECCV*), pp.128–142, Springer (2002).

[11] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Gool, L.: A Comparison of Affine Region Detectors, *International Journal of Computer Vision* (*IJCV*), Vol.65, No.1, pp.43–72 (2005).

[12] Mikolajczyk, K., Zisserman, A. and Schmid, C.: Shape Recognition with Edge-Based Features, *British Machine Vision Conference* (*BMVC*), Vol.2, pp.779–788 (2003).

[13] Ostromoukhov, V.: A simple and efficient error-diffusion algorithm, *Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp.567–572, ACM (2001).

[14] Pang, W.-M., Qu, Y., Wong, T.-T., Cohen-Or, D. and Heng, P.-A.: Structure-Aware Halftoning, *ACM Trans. Graphics* (*TOG*), Vol.27, No.3, p.89, ACM (2008).

[15] Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A.: Object Retrieval With Large Vocabularies and Fast Spatial Matching, *Computer Vision and Pattern Recognition* (*CVPR*), pp.1–8 (2007).

[16] Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A.: Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases, *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp.1–8, IEEE (2008).

[17] Rapantzikos, K., Avrithis, Y. and Kollias, S.: Detecting Regions from Single Scale Edges, *Intern. Workshop on Sign, Gesture and Activity* (*SGA*), *European Conference on Computer Vision* (*ECCV*), Lecture Notes in Computer Science, Vol.6553, pp.298–311, Springer Berlin Heidelberg (2010).

[18] Varytimidis, C., Rapantzikos, K. and Avrithis, Y.: WαSH: Weighted α-Shapes for Local Feature Detection, *European Conference on Computer Vision* (*ECCV*), Florence, Italy, pp.788–801, Springer Berlin Heidelberg (2012).

[19] Varytimidis, C., Rapantzikos, K., Avrithis, Y. and Kollias, S.: Improving local features by dithering-based image sampling, *Proc. Asian Conference on Computer Vision* (*ACCV 2014*), Singapore (2014).

[20] Vedaldi, A. and Fulkerson, B.: VLFeat: An Open and Portable Library of Computer Vision Algorithms, available from ⟨http://www.vlfeat.org/⟩ (2008).

[21] Yang, Y., Wernick, M. and Brankov, J.: A Fast Approach for Accurate Content-Adaptive Mesh Generation, *IEEE Trans. Image Processing*, Vol.12, No.8, pp.866–881 (2003).

[22] Zhou, B. and Fang, X.: Improving mid-tone quality of variable-coefficient error diffusion using threshold modulation, *ACM Trans. Graphics* (*TOG*), Vol.22, No.3, pp.437–444, ACM (2003).

[23] Zitnick, C. and Ramnath, K.: Edge foci interest points, *International Conference on Computer Vision* (*ICCV*), pp.359–366 (2011).

Christos Varytimidis was born in Katerini, Greece in 1983. He received the Diploma degree from the School of Electrical and Computer Engineering of the National Technical University of Athens (NTUA) in 2008. He is currently pursuing his Ph.D. in image processing and computer vision and is a member of the Image, Video and Multimedia Systems Laboratory (IVML) of NTUA. His research interests are visual feature detection, object recognition and detection, in images and video. He has published 7 articles in proceedings of international conferences and international journals.

Konstantinos Rapantzikos received his Ph.D. degree (2008) from the School of Electrical and Computer Engineering of the National Technical University of Athens, the M.Sc. degree (2002) from the Department of Electronic and Computer Engineering of the Technical University of Crete and the Diploma (2000) from the same department. His interests include visual feature extraction, modeling of visual saliency, action recognition, and optical flow estimation. He has published 14 articles in international journals and books and 25 in proceedings of international conferences.

Yannis Avrithis was born in Athens, Greece in 1970. He received his Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in September 1993, the M.Sc. degree in Communications and Signal Processing (with Distinction) from the Department of Electrical and Electronic Engineering of the Imperial College of Science, Technology and Medicine, University of London, UK, in October 1994, and the Ph.D. degree from the School of Electrical and Computer Engineering (ECE) of NTUA, in March 2001. He is currently a senior researcher at the Image, Video and Multimedia Systems Laboratory (IVML) of the National Technical University of Athens (NTUA), carrying out research on image and video analysis, computer vision and machine learning, and teaching in NTUA. His research interests include visual feature detection, representation of visual appearance and geometry, image matching and registration, image indexing and retrieval, clustering, nearest neighbor search, object detection and recognition, scene classification, image/video segmentation and tracking. He has been involved in 15 European and 9 National research projects, he has co-supervised 8 Ph.D. theses and 12 Diploma theses, and he has published 3 theses, 3 edited volumes, 25 articles in journals, 99 in conferences and workshops, 8 book chapters and 7 technical reports in the above fields. He has contributed to the organization of 19 conferences and workshops, and is a reviewer in 15 scientific journals and 15 conferences.

Stefanos Kollias, Fellow IEEE, has received his B.Sc. Electrical Engineering from National Technical University of Athens, Greece, 1979, the M.Sc. Communication Engineering, from UMIST, England, 1980, the Ph.D. from Computer Science Division, NTUA, 1984. He has been Research Scientist at Center Telecommunications Research, Columbia University, NY, USA, 1987–1989, Ass. Professor (1990–1996) and Professor of the Computer Science Division of NTUA, 1997– and Director of Image, Video & Multimedia Systems Lab, ECE School, NTUA, 1990–. He has been a Member of the: Executive Committee of the European Neural Network Society, 2007–, European Member State Expert Group (MSEG) 2004–, High level Expert Group of EUREKA, 2012–. His research interests include: Analysis and Search of Multimedia Content, Intelligent Systems & Artificial Neural Networks, Knowledge Technologies, Human Machine Interaction and Affective Computing, Learning Technologies, Serious Games, Education and training of persons with special needs. He has been leading more than 100 projects in these research fields. He has more than 100 publications in International Scientific Journals, 200 presentations in International Conferences, being Co-Editor of the book "Multimedia and the Semantic Web", Wiley, 2005 and General Chair of 5 International Conferences.

(Communicated by *Hideo Saito*)