WαSH-ing visual repositories: searching Europeana using local features

Christos Varytimidis, Konstantinos Rapantzikos and Stefanos Kollias
National Technical University of Athens, Greece
{chrisvar,rap}@image.ntua.gr,stefanos@cs.ntua.gr

***Abstract*—Museums, libraries, national archives and art galleries deal with visual objects that must be made accessible to a wide variety of experts or non-experts like researchers, art lovers or interested people. The ability to identify objects sharing some aspect of visual similarity can be useful when trying to trace historical influences or when looking for further examples of paintings, sculptures or other cultural objects appealing to their taste. In this direction, we use our recent detector [1] for an image retrieval task on a subset of Europeana's[1] content. The detector produces distinctive features by grouping sampled image edges according to proper shape stability measures. We evaluate the detector by integrating it into the Visual Search Engine for Europeana (Vieu)[2] tool.**

I. Introduction

Europeana, probably the most significant achievement towards accessibility of digital culture, is becoming the single reference point for European culture online. The collection of Europeana is currently composed of more than 22 million digitized objects, giving access to digital books, maps, newspapers, journals, photographs, sound and video, manuscripts, museum exhibits and paintings. Successful retrieval of visually similar objects is challenging due to the diversity and the inherent complicated structure (e.g. vessel's paintings). The state-of-the-art in visual retrieval continuously scales up and permits searching into huge collections of generic or specific images [2][3]. In this work, we apply WαSH, our recent local visual feature detector [1], to a large collection of images included in the ATHENA content of the Europeana portal [4]. We embed the detector to Vieu that includes about 10% of the ATHENA images, i.e., around 420,000 images retrieved from the Europeana portal and can be evaluated on-line[2].

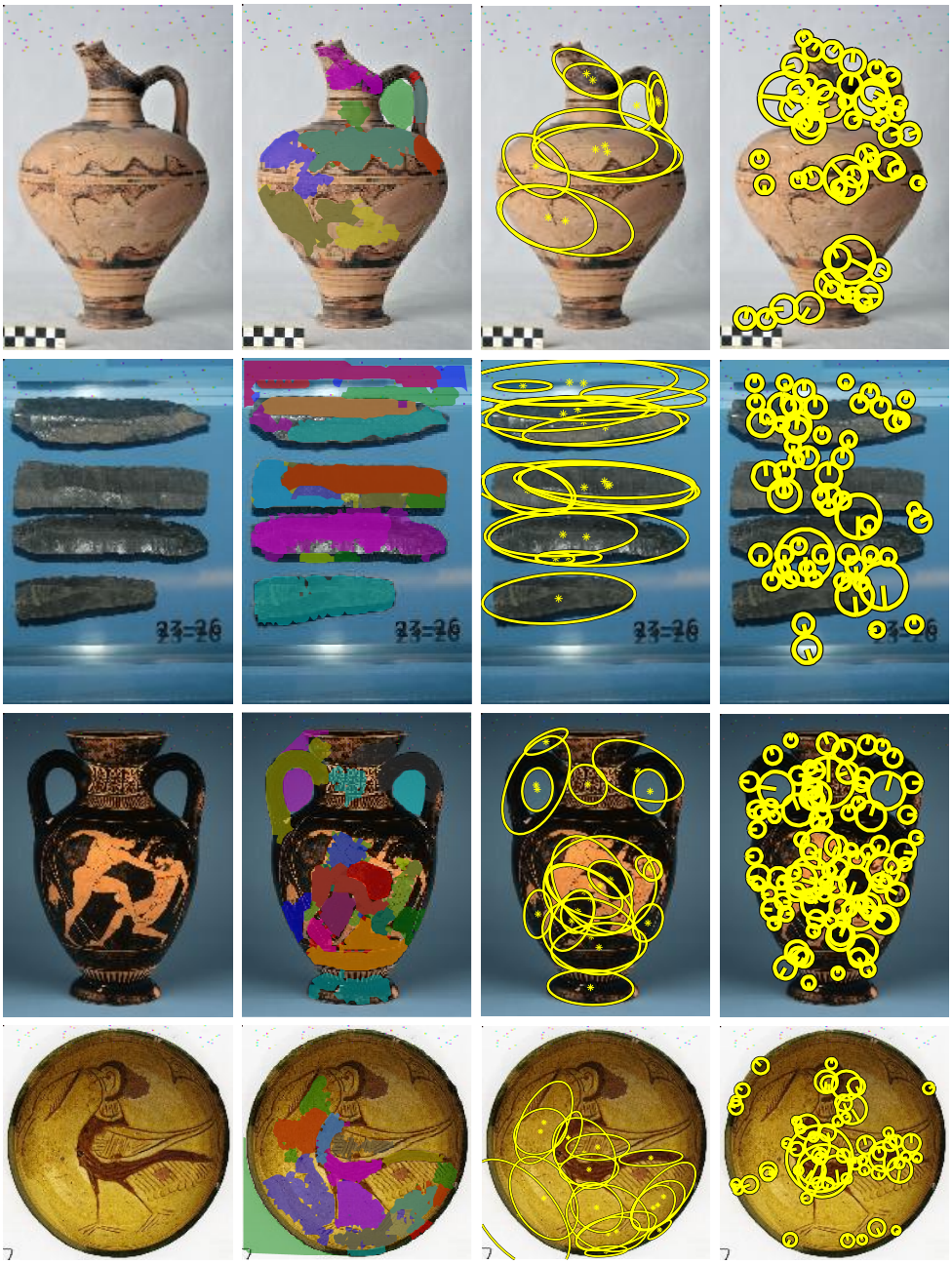
The literature on local feature detection is rich and since the early work of Beaudet [5] and Harris and Stevens [6], based on the Hessian and the second moment matrices respectively, many detectors have been proposed grounding on similar or novel ideas. In his inspiring work, Lindeberg extended detectors by making them scale-invariant [7] and establishing the theoretical foundations for making them affine-invariant [8]. Based on these foundations, Lowe proposed the *scale invariant feature transform* (SIFT) in [9] and Mikolajczyk *et al*. the affine-adapted version of the Harris and Hessian detectors [10], [11]. The *maximally stable extremal regions* detector (MSER) of Matas *et al*. [12] fires on regions of stable intensity and therefore avoids common problems of gradient-based methods like localization accuracy and noise. The recent trend of achieving a good balance between efficiency and performance has led to a group of computationally efficient detectors like SURF [13], an approximate version of SIFT.

Although naturally meaningful, image edges have attracted less attention in local feature detection, with the reasons being mainly related to the lack of stable edges (e.g. across different illumination conditions, image blurring or affine transformations) and the computational inefficiency. The WαSH detector we proposed in [1] builds on recent work on edge-based local features [14] [15] and overcomes previous drawbacks. Edge points are grouped based on location, gradient strength and local shape by exploiting α-shapes [16], which can be considered as a generalization of the *convex hull*, being parametrized by scalar $\alpha \geq 0$. Starting from the convex hull of a point set for $\alpha = \infty$, α-shapes reduce to the set itself at the other extreme, for $\alpha = 0$. This collection of different sets of points, edges and triangles of the full triangulation of a point set is called the α-filtration of the set. This representation captures the different subsets of the convex hull constructed by varying α, which plays the role of a scale parameter, selecting different levels of detail.

In our work, we exploit *Weighted α-shapes* [17], that provide a richer description of the input, since for a single value of α, *i.e*. for a single scale, they capture different levels of spatial details. Binary image edges are appropriately sampled and shape is used as the main feature selection criterion, a choice that bears similarities to [14], although the geometric representation is entirely different. The evolving topology of local regions in the α-filtration is captured by applying a stability measure on the nodes of a *component tree* representing the filtration, selecting prominent regions acquired for different α-values. Intuitively, the method provides an efficient way to overcome the main weakness of the baseline α-shapes, namely the automatic selection of a single value of α that best represents the underlying point set. The entire method does not rely on a "clean" edge map, i.e. noisy edges do not reduce the efficiency, and is controlled by a simple and intuitive parameter related to region shape. Some detection examples are given in Fig. 1c.

The remaining of the paper is organized as follows: In section II we provide a short description of WαSH and in section III we describe the experimental framework and comment on the results. The conclusions follow in section IV.

[1]http://www.europeana.eu
[2]http://vieu.image.ntua.gr



(a) (b) (c) (d)

Fig. 1. Local feature detection on images of cultural items. (a) Original images. (b) WαSH regions in color and (c) corresponding WαSH features. (d) SURF features.

II. WαSH-Detector

In the following we provide a short description of the detector. More details and thorough evaluation can be found in [1].

A. Shape representation from edges

The input is a set of edge points $P \subseteq \mathbb{R}^2$ that is generated by applying a binary edge detector on the input (grayscale) image and sampling the resulting edge map. Sampling is roughly uniform along edges with a fixed *sampling interval*. Although sparsely sampled along the edges, the points in P capture the shape of the detected boundaries. In practice, the binary edge map is obtained by the *Canny* edge detector [18], modified to provide the samples online while following the edges and applying the hysteresis thresholds. We do not rely on a clear edge map and therefore the high and low hysteresis thresholds of the detector are kept fixed.

The definition of α-shapes is based on the Delaunay triangulation. Weighted α-shapes are based on its generalization, the *regular triangulation*, formed by replacing the Euclidean

distance by the power to weighted points. For each point $p \in P$, we define its *weight* $w(p) \geq 0$ to be proportional to its gradient strength. A point $p \in P$ along with its weight $w(p) \geq 0$ makes up a pair $(p, w(p))$ that is called a *weighted point* and can be seen as a *circle* centered at p, with squared radius $w(p)$. The collection of all regular triangles over P is called the *regular triangulation* of P. Observe that, if $w(p) = 0$ for all $p \in T$, then the triangulation reduces to *Delaunay*.

The collection of all triangles and their faces (line segments and points) are called *simplices*. If we define a *size* $\rho_T \geq 0$ for each simplex, then the *weighted* α*-complex* of P is the subset containing all simplices up to a given size $\alpha \geq 0$, Finally, the *weighted* α*-shape* of P [17] is the union of all such simplices.

B. Feature selection

All simplices are typically ordered by ascending size to obtain what is called a *weighted* α*-filtration* [17]. We deviate from this standard setting by considering only *triangles* and their *edges* (line segments) discarding points $p \in P$ themselves. For any point set, there is a finite number of critical α values that, compared to the size ρ of faces, produce different α-shapes. Starting from the largest element of size ρ_1 and decreasing the value of α towards ρ_n, the upper α-complex evolves from the set containing the largest faces to the full convex hull. This way we model the growing *cavities* of the original shape. To capture its evolving topology, we construct a *component tree* [19]. The component tree organizes the connected components for all different critical values of α in an efficient representation. We employ a strict neighborhood system, where two triangles are separated by their common edge to retain image boundaries between components (more details in [1]). Starting with all simplices being individual components, we process them in descending order of size, joining them with their neighbors to end up with the convex hull of the points (Fig. 2).

While tracking the evolution of the connected components of the upper α-complex, we measure the significance of changes in its topology so as to decide on stability and distinctiveness of the corresponding image regions. Whenever a new component is created by joining two adjacent components, these two components are individually considered as potential features. We choose to base the decision on a simple scale-invariant measure (*strength*) that favors large components with small (or no) openings on their boundary. The final features are blob-like and include regions that are not extremal in the intensity domain or regions determined by cavities of boundary shape (not completely bounded by edges). Example images to evaluate the type of produced features can be found in Fig. 1. Overall, WαSH captures blob-like regions that are either structural parts of the depicted objects or parts of the background that form prominent cavities in the objects' shape.

III. Experimental evaluation

Searching based on visual similarity is essential for providing hints to people searching cultural items for historical influence or for browsing based on visual interest. Towards this goal, we have created Vieu, a large scale image retrieval portal based on [3] to test the efficiency of state-of-the-art retrieval methods on the Europeana content. In this section we conduct experiments based on the WαSH detector and provide quantitative and qualitative results.

Experimental evaluation is based on the online digital content of the Europeana portal and specifically on a subset of the ATHENA collection, namely a set of ∼420,000 images. A set of 50 queries was selected randomly from the images and the results were labeled by an expert as correct or wrong. The performance was measured using the widely used mean Average Precision metric. For comparison purposes we evaluate WαSH against the SURF detector that is part of the baseline method in the VIEU tool. We use the default parameters for both detectors and extract SIFT descriptors for WαSH (SURF comes with its own descriptor).

The Bag-of-Words (BoW) model is used to represent the images using a vocabulary of 1,000 visual words. To obtain the visual vocabulary, we create clusters in the space of descriptors and assign each feature to the closest centroid (i.e., visual word). We use a fast variant of the k-means algorithm that uses approximate nearest neighbor search, i.e. nearest cluster centers at each iteration are assigned using randomized kd-trees [20]. Specifically, we use the FLANN library of Muja and Lowe [21] both in vocabulary creation and in assigning visual words to image features. Having each local feature assigned to a visual word, we can represent each image in terms of the visual words it contains. A histogram of constant-length is then constructed for each image, containing the appearance frequencies of each visual word (BoW histogram). Following this, visual representation of all images has been organized in an index structure to allow for fast retrieval.

WαSH outperformed SURF by 7.8%, as shown in Table I, while using a smaller number of features, which is a significant improvement. The detection examples shown in Fig. 1 show that WαSH better captures the spatial structure of the object and is less affected by non-distinctive regions like the ones on the physical external boundary of the objects (e.g. small-scale features on the exterior of a round plate may look quite similar to the corresponding ones on a round vessel). The top 10 results for some queries are shown in Fig. 3 and Fig. 4 for WαSH and SURF respectively.

TABLE I
RESULTS OF THE RETRIEVAL EXPERIMENT

Detector	WαSH	SURF
Features/image	63.0	78.5
mAP	0.595	0.552

IV. Discussion

Visual information offered to users of digital libraries, has not gained much attention (especially in formation of the Europeana portal) due to problems, on the one hand, with intellectual property rights of the associated cultural objects

Fig. 2. Evolution of α-shapes. Starting from connected components created for large α values on the left, we end up with almost all the triangles of the triangulation for a small α value on the rightmost image (better viewed in color and under magnification).

and, on the other hand, with the inherent difficulty of analysing complex images often offered as low resolution versions of the originals.

Based on our recent WαSH detector we set up a visual retrieval method of images related to cultural items of the Europeana's collection. It seems that the type of the detected features fits well with this type of images. The exploitation of edges, the efficient grouping based on weighted α-shapes and the shape-based selection criterion produce blob-like regions, cavities and pockets that adequately represent the structural complexity of the input. Furthermore, the computational complexity is low and makes the detector suitable for large-scale evaluation. By integrating the detector into the VIEU platform we allow for such large-scale evaluation and enable additional services like tag aggregation (when an image has been identified as similar to a tagged one) from external sites (e.g. Wikipedia). In order to improve performance, we plan to incorporate color in the process that seems to play an important role in the specific dataset (e.g. bronze vs. gold items). Hence, we will consider exploiting color descriptors (extracted from WαSH regions) to further improve the results.

Acknowledgment

This work is supported by the National GSRT-funded project 09SYN-72-922 "IS-HELLEANA: Intelligent System for HELLEnic Audiovisual National Aggregator", http://www.helleana.gr, 2011-2014.

References

[1] C. Varytimidis, K. Rapantzikos, and Y. Avrithis, "Wαsh: Weighted α-shapes for local feature detection," in *Proceedings of European Conference on Computer Vision (ECCV 2012)*, Florence, Italy, October 2012.

[2] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Computer Vision and Pattern Recognition (CVPR)*, 2007.

[3] Y. Kalantidis, G. Tolias, Y. Avrithis, M. Phinikettos, E. Spyrou, P. Mylonas, and S. Kollias, "Viral: Visual image retrieval and localization," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 555–592, 2011.

[4] "Proceedings of the athena conference cultural institutions online," 2011.

[5] P. Beaudet, "Rotationally invariant image operators," in *Proceedings of the International Joint Conference on Pattern Recognition*, 1978, pp. 579–583.

[6] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of Alvey Vision Conference*, 1988.

[7] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision (IJCV)*, vol. 30, no. 2, pp. 79–116, 1998.

[8] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure," *Image and Vision Computing*, vol. 15, no. 6, pp. 415–434, 1997.

[9] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004.

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

[11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision (IJCV)*, vol. 65, no. 1, pp. 43–72, 2005.

[12] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[13] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding (CVIU)*, vol. 110, pp. 346–359, June 2008.

[14] Y. Avrithis and K. Rapantzikos, "The medial feature detector: Stable regions from image boundaries," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2011.

[15] K. Rapantzikos, Y. Avrithis, and S. Kollias, "Detecting regions from single scale edges," in *Proc. of Intern. Workshop on Sign, Gesture and Activity (SGA), European Conference on Computer Vision (ECCV)*, September 2010.

[16] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 551–559, 1983.

[17] H. Edelsbrunner, "Alpha shapes—a survey," in *Tessellations in the Sciences: Virtues, Techniques and Applications of Geometric Tilings*, R. van de Weijgaert, G. Vegter, J. Ritzerveld, and V. Icke, Eds. Springer Verlag, 2010.

[18] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence (PAMI)*, vol. 8, pp. 679–698, 1986.

[19] L. Najman and M. Couprie, "Building the component tree in quasi-linear time," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3531–3539, 2006.

[20] C. Silpa-Anan and R. Hartley, "Optimised KD-trees for fast image descriptor matching," in *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, 2008.

[21] M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2009.

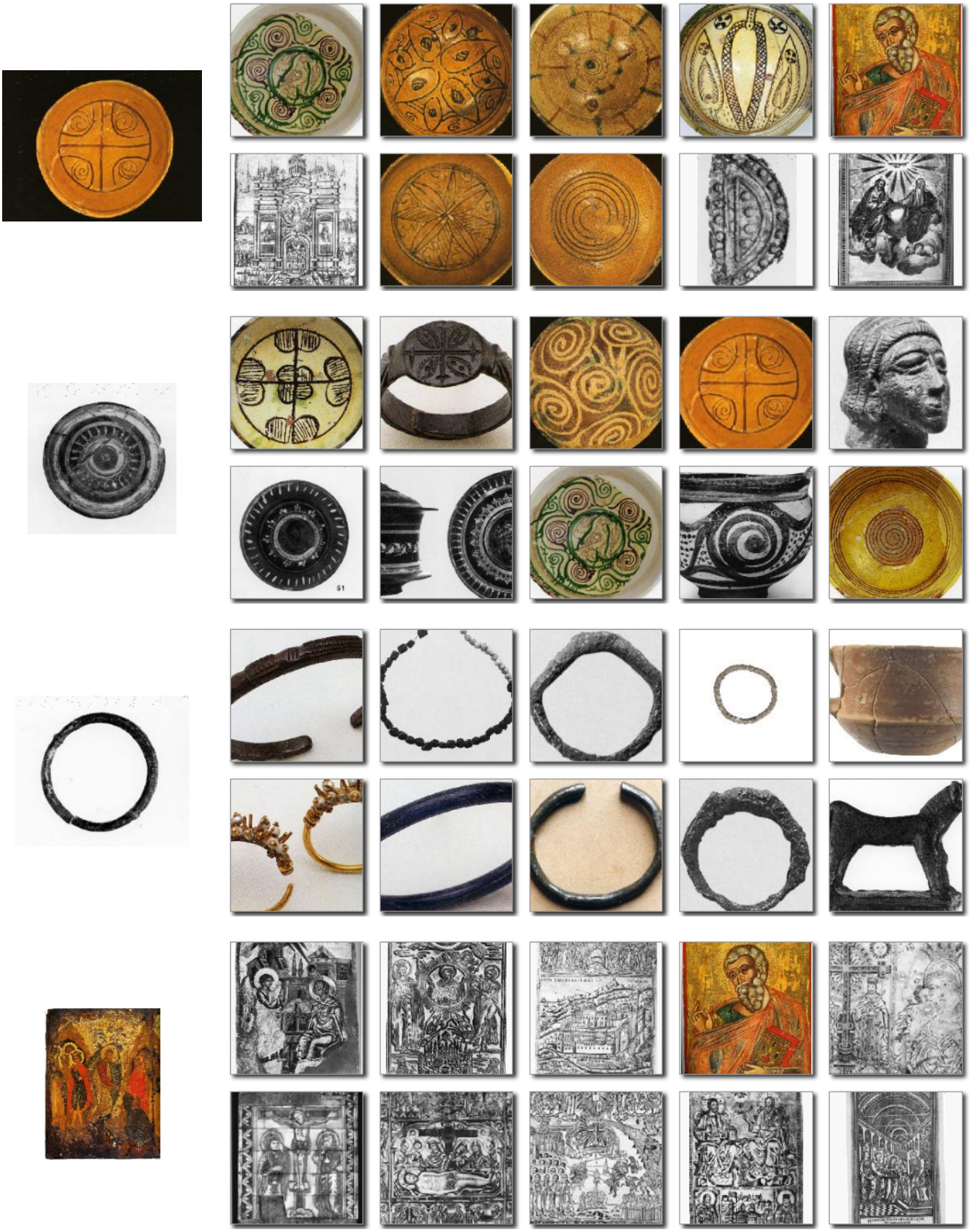


Fig. 3. Retrieval results using the WαSH detector. For each query we present the top 10 results.

Fig. 4. Retrieval results using the SURF detector. For each query we present the top 10 results.