Improving local features by dithering-based image sampling

Christos Varytimidis, Konstantinos Rapantzikos,
Yannis Avrithis and Stefanos Kollias

National Technical University of Athens
{chrisvar,rap,iavr}@image.ntua.gr, stefanos@cs.ntua.gr

Abstract. The recent trend of structure-guided feature detectors, as opposed to blob and corner detectors, has led to a family of methods that exploit image edges to accurately capture local shape. Among them, the WαSH detector combines binary edge sampling with gradient strength and computational geometry representations towards distinctive and repeatable local features. In this work, we provide alternative, variable-density sampling schemes on smooth functions of image intensity based on dithering. These methods are parameter-free and more invariant to geometric transformations than uniform sampling. The resulting detectors compare well to the state-of-the-art, while achieving higher performance in a series of matching and retrieval experiments.

1 Introduction

Image representation based on local features is often used in many computer vision applications due to the balanced trade-off between sparsity and discriminative power. By ignoring non-salient image parts and focusing on distinctive regions, local features provide invariance, repeatability, compactness and computational efficiency.

Popular detectors like the Hessian-Affine [1] and SURF [2] are based on image gradients, while others like the MSER [3] 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 [4] 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 s. In an attempt to overcome this limitation, we propose a different sampling scheme that relies directly upon image intensity. 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, as shown in [5]. Mikolajczyk *et al.* [6] 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.* [7] 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.* [8] 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 [9] 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.

The recently proposed WαSH detector [4] 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.

In this work, we introduce two sampling methods that are based on the well known Floyd-Steinberg algorithm [10]. 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 bitmap 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 [11] and Zhuand and Fang [12] 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. Nevertheless, we use the initial algorithm because of its computational efficiency and the nature of our problem, which is sampling rather than halftoning.

Our work is also related to the work of Gu et al. [13], 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 [13, 14]. However, we do not detect our sample points as features; we rather use them to initialize the WaSH feature detector.

The main contributions of this work are: (a) the introduction of two image sampling schemes of variable density, and (b) the application to local feature detection, evaluated on image matching and retrieval.

3 Background: the WαSH detector

The WαSH feature detector [4] is based on *α-shapes*, a representation of a point set P in two dimensions, parametrized by scalar α. In fact, α-shapes are a generalization of the convex hull, which is not convex or even connected in general. In the simplest case, α-shapes use an underlying Delaunay triangulation, but *weighted α-shapes* in [4] use the *regular triangulation* instead. The latter is a generalization of Delaunay where each point in P is assigned a non-negative *weight*, hence it can capture more information from the image domain. In practice, weight is a function of image gradient in [4].

A particular *size* is assigned to every simplex (edge or triangle) in the triangulation, as a function of positions and weights of its vertices. Ordering simplices by decreasing size, a *component tree* is used to track the evolution of connected components as simplices are added to form larger regions. Connected components are potentially selected as features during evolution, according to a shape-driven strength measure. The resulting 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 that edge sampling is *uniform*, hence when sampling a contour, 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, a *sampling step* parameter 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 object shape may be lost.

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

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.

For dithering, we use the Floyd–Steinberg algorithm [10], which is fast, requiring only one iteration over the image, and provides reasonable results. In our framework, the 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 needed for the remaining steps of the WαSH detector [4].

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,

$$G = \|\nabla g(\sigma) * I\|. \tag{1}$$

Then, similar to [15], if $\hat{G}(x, y)$ is the gradient strength at point (x, y) normalized to $[0, 1]$, we use the non-linear function

$$s(x, y) = \hat{G}(x, y)^{\gamma} \tag{2}$$

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.* [15] 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 (2), if $\hat{\lambda}_1(x, y)$ is the largest eigenvalue normalized to $[0, 1]$, we use function

$$s(x, y) = \hat{\lambda}_1(x, y)^{\gamma} \tag{3}$$

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. 1f). Compared to the gradient-based sampling, the number of detected features is often lower (see section 5).

4.3 Examples

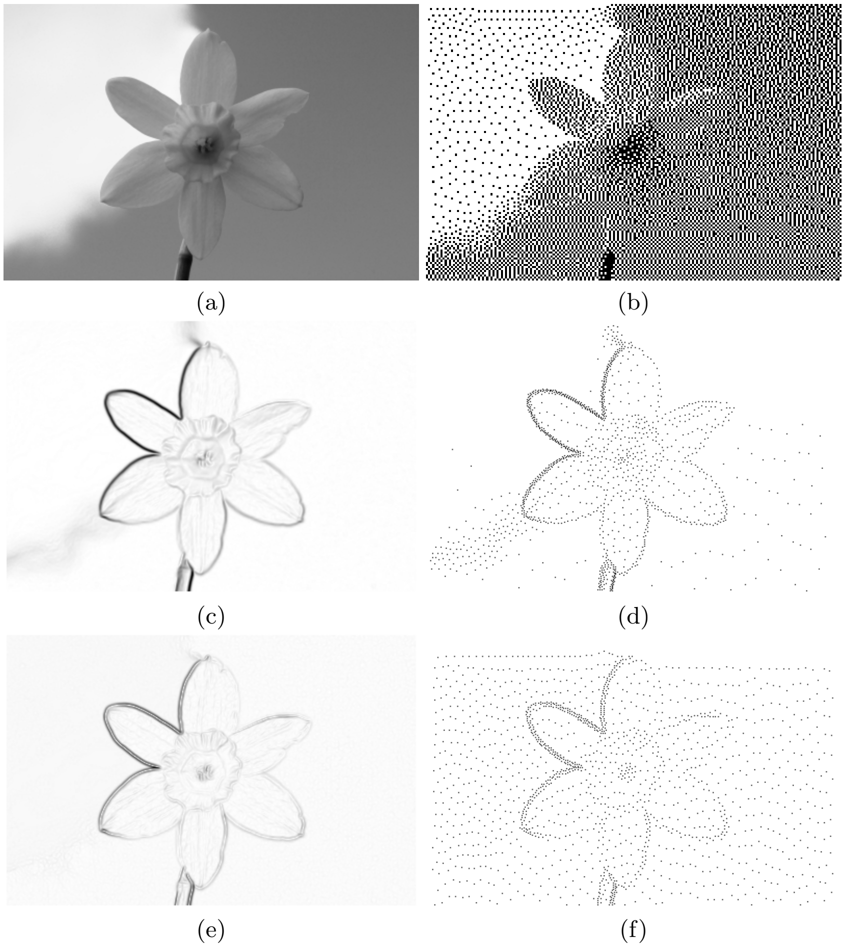
A visual example of the sampling methods is shown in Fig. 1. Fig. 1cd depict the normalized gradient strength $\hat{G}$ and the resulting gradient-based sampling. Notice the sparsity of the samples in smooth areas and the density in structured ones. Fig. 1ef depict the Hessian response $\hat{\lambda}_1$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

Fig. 2 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 sampled on both sides of edges that—depending on the application—may prove useful at enforcing actual edge boundaries.

Examples of the features detected using either the baseline sampling of WαSH or the proposed sampling methods are depicted in Fig. 3, 4. In each example, the number of detected features for each method is approximately the same (around 200 for Fig. 3 and 50 for Fig. 4). In Fig. 3 we present the results on the first image of the *graffiti* dataset of [5]. Both dithering-based samplings detect more detailed regions of the image, and the gradient-based one better captures the correct boundaries of objects. In Fig. 4, the input image comes from the PASCAL VOC 2007 test set [16], a dataset heavily used for evaluating object recognition algorithms. Again the gradient-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



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 1. Dithering-based sampling. (a) Input image and (b) Floyd-Steinberg dithering on (a). (c) Normalized gradient strength $\hat{G}$ and (d) sampling on $\hat{G}$. (e) Hessian response $\hat{\lambda}_1$ and (f) sampling on $\hat{\lambda}_1$. Figure is optimized for screen viewing.

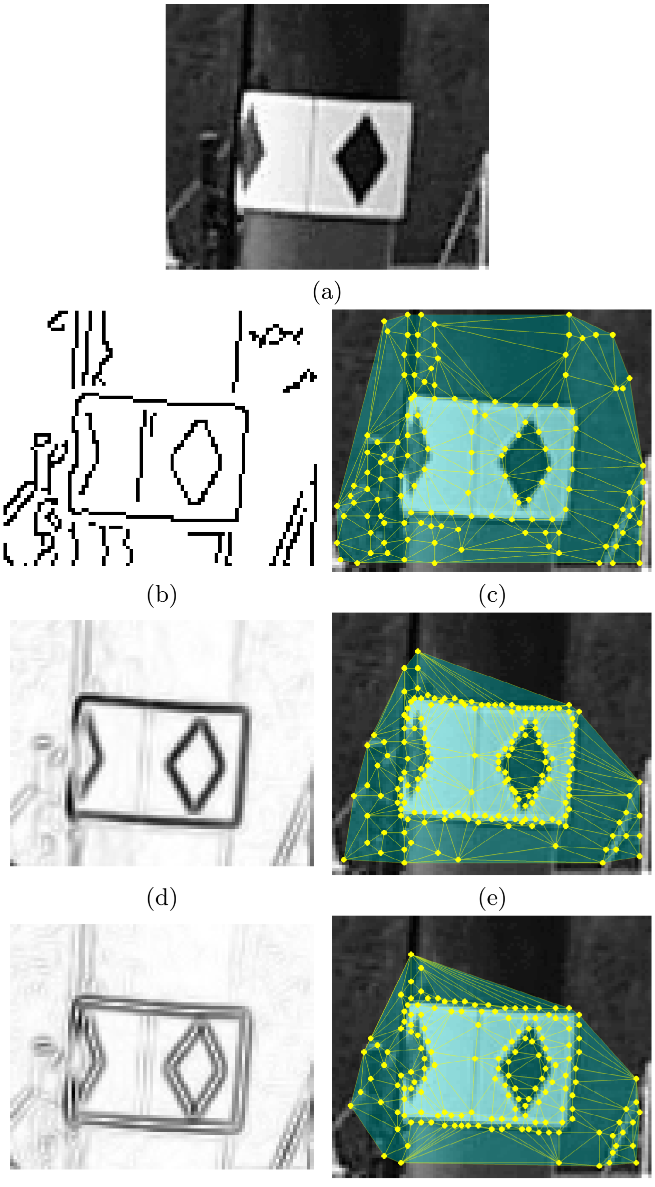


Fig. 2. Example of the different sampling methods and the corresponding triangulations. (a) Input image, a detail of the first image of the boat sequence of [5] (see section 5.1). (b) Binary edge map and (c) uniform sampling on (b). (d) Normalized gradient strength and (e) error-diffusion on (d). (f) Hessian response and (g) error-diffusion on (f). (b,d,f) are shown in negative for better viewing and printing.

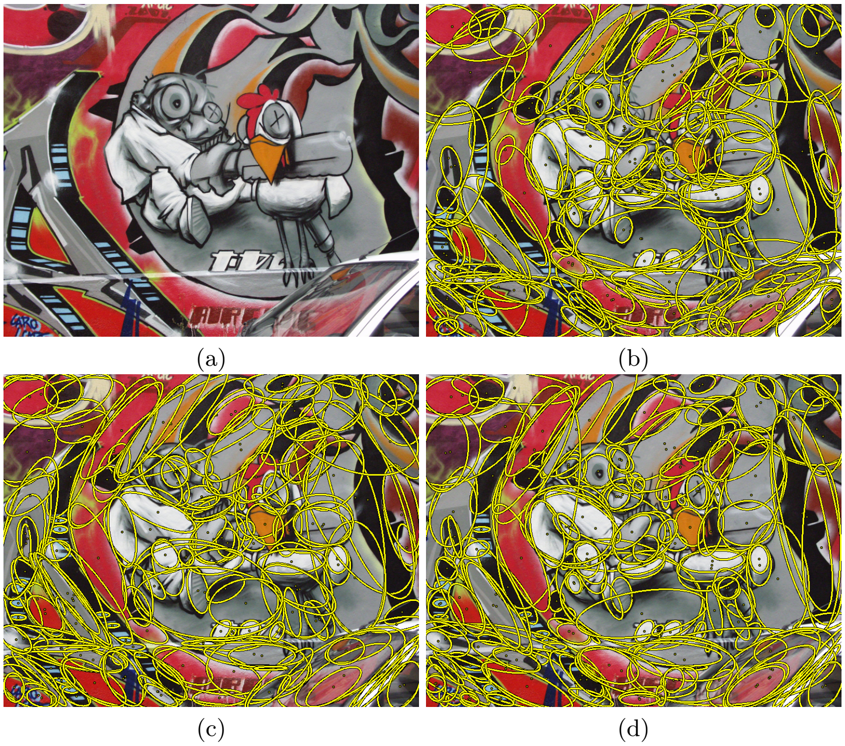


Fig. 3. Example of local features detection. (a) Input image and (b) baseline WαSH results using uniform sampling. (c) Results using the gradient-based sampling and (d) using the Hessian-based sampling.

proposed by Mikolajczyk *et al.* [5], with 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* [17] dataset. The performance is measured by the *mean average precision* (mAP) of the query results.

Following an initial brief evaluation of the proposed samplings, we set $\gamma = 1$ for all the experiments. For $\gamma > 1$ samplings were sparser and performance slightly dropped, while for $\gamma < 1$ the performance increased, but samplings were denser, increasing the computational cost of the feature detector.

5.1 Repeatability and matching score

In this experiment, we investigate the impact on performance of a matching application, when using the proposed sampling methods on WαSH. We also compare to the state-of-the-art detectors, Hessian-Affine and MSER, for which we use the executables provided by the corresponding authors and default parame-

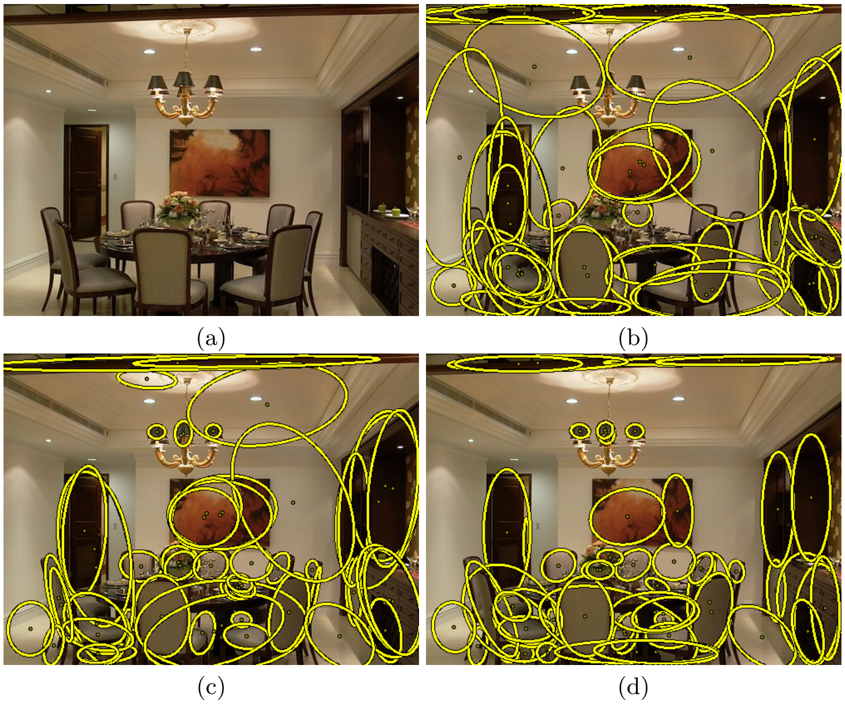

(a) (b) (c) (d)

Fig. 4. 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.

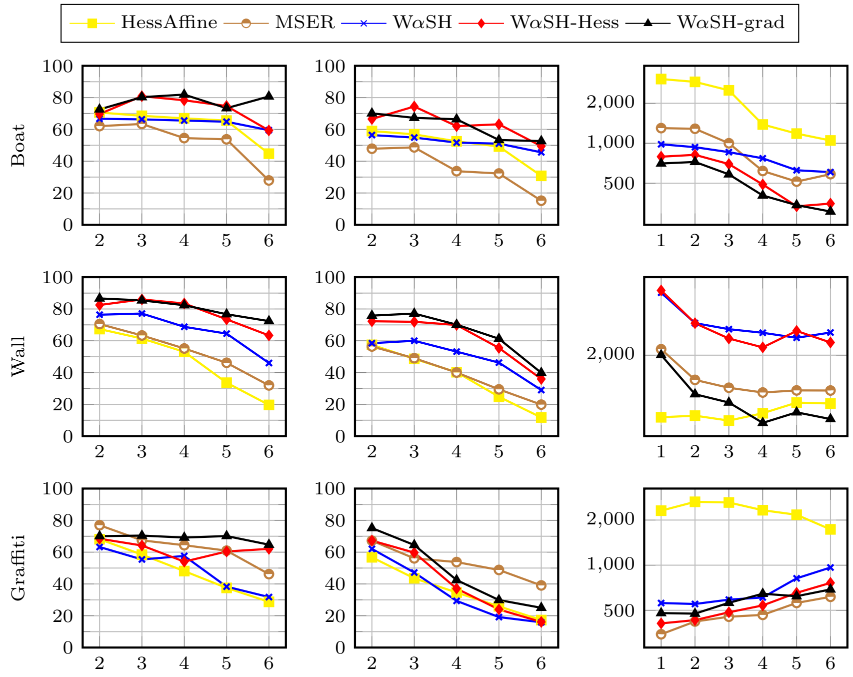
ters. The image sets used, evaluate the impact of changes in viewpoint, rotation, zoom, blur and illumination. For the matching score we use 128-dimensional SIFT descriptors for all detectors.

The results of the evaluation are depicted in Fig. 5-6. The last row of Fig. 6 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.

5.2 Image retrieval

In this experiment, we evaluate the proposed variants of WαSH on an image retrieval application. The dataset is the *Oxford 5K*, consisting of images of buildings as queries, and other urban images as distractors. We compare against Hessian-affine, MSER, SIFT and SURF, using the corresponding executables and default values. For all detectors we extract SIFT descriptors, apart from SURF, which performs best using the corresponding descriptor.

For the different versions proposed, we adapt the selection threshold to extract approximately the same number of features as the baseline WαSH. For all



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

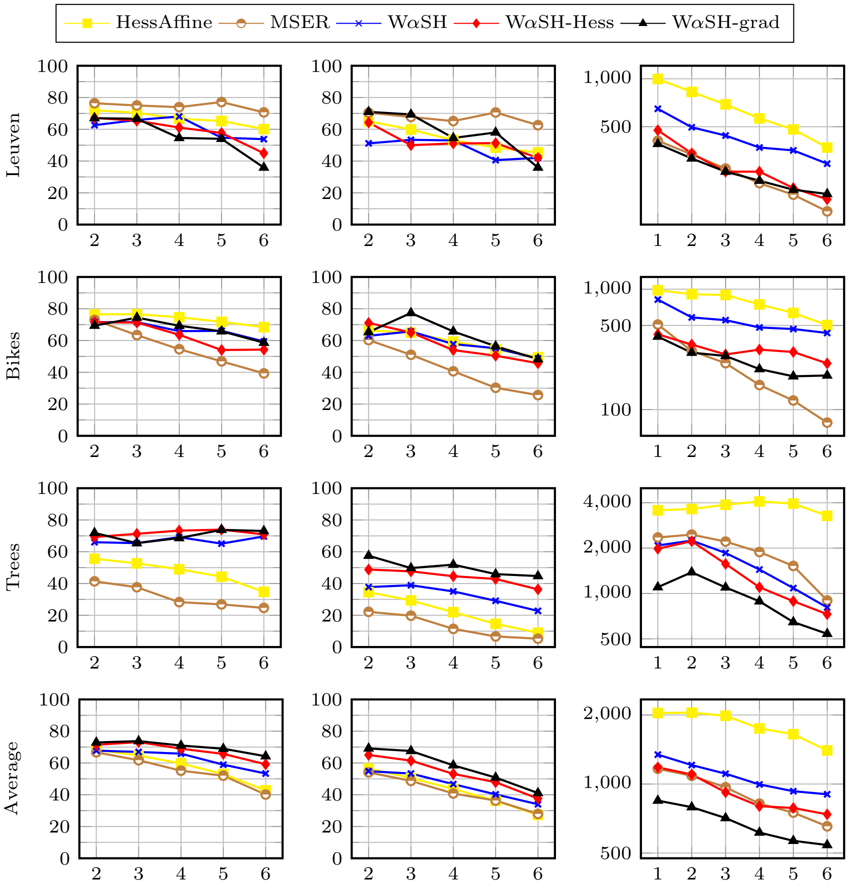
Fig. 5. Comparison of our proposed sampling methods to baseline WαSH and the state-of-the-art in sequences *boat, wall* and *graffiti*. #features: number of features detected per image. Hess: Hessian-based dithering; grad: gradient-based dithering.

detectors we create 3 different vocabularies of size 50K, 100K and 200K visual words. We use the simple Bag-of-Words approach, as well as a spatial reranking of the results, using fastSM [17]. Performance is measured using the *mean Average Precision* (mAP) metric, and 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 with and without the spatial verification step, a result that verifies the findings of section 5.1.

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 al-



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

Fig. 6. 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.

gorithm 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 thoroughly evaluate the performance of the proposed methods 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, the performance of WαSH increases in both applications, exceeding the state-of-the-art. In the fu-

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

detector	features	Bag-of-Words (mAP)			ReRanking (mAP)		
	($\times 10^6$)	50K	100K	200K	50K	100K	200K
HessAff	29.02	0.483	0.539	0.573	0.518	0.577	0.607
MSER	13.33	0.487	0.534	0.565	0.519	0.569	0.595
SIFT	11.13	0.422	0.465	0.495	0.441	0.486	0.517
SURF	**6.84**	0.465	0.526	0.574	0.509	0.573	0.603
WαSH	7.19	0.529	0.569	0.590	0.537	0.569	0.585
WαSH, grad	7.63	**0.531**	**0.580**	**0.605**	**0.543**	**0.578**	**0.609**
WαSH, Hess	7.29	0.518	0.553	0.582	0.511	0.557	0.584

ture, we will further investigate the effect of the scaling factor γ applied on both proposed sampling methods, as well as evaluate the performance on different applications of the feature detector.

Acknowledgement. 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. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: European Conference on Computer Vision (ECCV), Springer (2002) 128–142
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Computer Vision and Image Understanding (CVIU) **110** (2008) 346–359
3. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing **22** (2004) 761–767
4. Varytimidis, C., Rapantzikos, K., Avrithis, Y.: Wαsh: Weighted α-shapes for local feature detection. In: European Conference on Computer Vision (ECCV), Florence, Italy, Springer Berlin Heidelberg (2012) 788–801
5. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision (IJCV) **65** (2005) 43–72
6. Mikolajczyk, K., Zisserman, A., Schmid, C.: Shape recognition with edge-based features. In: British Machine Vision Conference (BMVC). Volume 2. (2003) 779–788
7. Rapantzikos, K., Avrithis, Y., Kollias, S.: Detecting regions from single scale edges. In: Intern. Workshop on Sign, Gesture and Activity (SGA), European Conference on Computer Vision (ECCV). Volume 6553 of Lecture Notes in Computer Science., Springer Berlin Heidelberg (2010) 298–311
8. Zitnick, C., Ramnath, K.: Edge foci interest points. In: International Conference on Computer Vision (ICCV). (2011) 359–366
9. Avrithis, Y., Rapantzikos, K.: The medial feature detector: Stable regions from image boundaries. In: International Conference on Computer Vision (ICCV). (2011) 1724–1731

10. Floyd, R.W., Steinberg, L.: An adaptive algorithm for spatial gray-scale. In: Proceedings of the Society of Information Display. Volume 17. (1976) 75–77
11. Ostromoukhov, V.: A simple and efficient error-diffusion algorithm. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM (2001) 567–572
12. Zhou, B., Fang, X.: Improving mid-tone quality of variable-coefficient error diffusion using threshold modulation. In: ACM Transactions on Graphics (TOG). Volume 22., ACM (2003) 437–444
13. Gu, S., Zheng, Y., Tomasi, C.: Critical nets and beta-stable features for image matching. In: European Conference on Computer Vision, Springer Berlin Heidelberg (2010) 663–676
14. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Volume 2., IEEE (2003) Ii–264
15. Yang, Y., Wernick, M., Brankov, J.: A fast approach for accurate content-adaptive mesh generation. Image Processing, IEEE Transactions on **12** (2003) 866–881
16. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html (2003)
17. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: Computer Vision and Pattern Recognition (CVPR). (2007) 1–8