Evaluation of Gradient Vector Flow for Interest Point Detection

Julian Stöttinger[1], René Donner[1,2,3], Lech Szumilas[4], and Allan Hanbury[1]

[1] PRIP, Vienna University of Technology, Austria
[2] ICG, Technical University Graz, Austria
[3] CIR, Vienna Medical University, Austria
[4] ACIN, Vienna University of Technology, Austria

Abstract. We present and evaluate an approach for finding local interest points in images based on the non-minima suppression of Gradient Vector Flow (GVF) magnitude. Based on the GVF's properties it provides the approximate centers of blob-like structures or homogeneous structures confined by gradients of similar magnitude. It results in a scale and orientation invariant interest point detector, which is highly stable against noise and blur. These interest points outperform the state of the art detectors in various respects. We show that our approach gives a dense and repeatable distribution of locations that are robust against affine transformations while they outperform state of the art techniques in robustness against lighting changes, noise, rotation and scale changes. Extensive evaluation is carried out using the Mikolajcyzk framework for interest point detector evaluation.

1 Introduction

Interest points are an important tool in many current solutions to computer vision challenges. Fergus et al. [1] point out that frameworks using interest points are heavily dependent on the detector to gather useful features. They detect locations of characteristic structures such as blobs, corners or local image symmetry.

The interest point and interest region detectors allow for the reduction of computational complexity in scene matching and object recognition applications by selecting only a subset of image locations corresponding to specific and/or informative structures. The extraction of stable locations is a successful way to match visual input in images of the same scene acquired under different conditions. As evaluated in [2], successful approaches extracting stable locations rely on corner detection [3,4] or blobs like Maximally Stable Extremal Regions (MSER) [5] and Difference of Gaussians (DoG) [6]. The well known approach of detecting local symmetry [7] is also included in our evaluation.

Donner et al. [8] take the minima of the Gradient Vector Flow (GVF) [9] with one manually adjusted set of parameters to locate centers of local symmetry at a certain scale. We extend their approach and propose a GVF based scale space pyramid and a scale decision criterion to provide general purpose interest points. This multi-scale orientation invariant interest point detector has the

G. Bebis et al. (Eds.): ISVC 2008, Part I, LNCS 5358, pp. 338–348, 2008.

aim of providing stable and densely distributed locations. Due to the iterative gradient smoothing during the computation of the GVF, it takes more surrounding image information into account than other detectors. Its stability against noise, blur, JPEG artifacts, rotation and illumination change makes it a promising approach for many applications in computer vision. For example, low quality images and videos in on-line applications and used in mobile computing suffer from such effects. Medical imaging also often deals with low contrast images.

In the next section, we give an overview of the interest point detectors used. We introduce our approach in detail in Section 3. Experiments and results are given in Section 4. Finally, the conclusion is given in Section 5.

2 Interest Points

We describe the most successful approaches for detecting interest points. For the most stable and broadly used corner detectors, we choose the Harris Laplacian approach for its excellent performance in [2] and describe it in Section 2.1. In Section 2.2, we go into more detail about broadly used blob detectors: DoG and MSER. Symmetry based interest points are covered in Section 2.3.

2.1 Harris Laplacian Detector

The Harris corner detector introduced in [3] provides a corner measure for image data. The pixels are analyzed to result in a one dimensional corner measure, also referred to as Harris energy. It is based on the trace and determinant of the second moment matrix M. An extension of the Harris corner detector, the *scale-adapted Harris detector*, was introduced to achieve scale invariance by Mikolajczyk and Schmid [10]. The resulting patch size is the size of the Gaussian kernel as the scale σ of the corner detector. The second moment matrix of a certain scale is then

$$M = \sigma_D^2 G_{\sigma_I} \otimes \begin{bmatrix} L_x^2(\sigma_D) & L_x L_y(\sigma_D) \\ L_x L_y(\sigma_D) & L_y^2(\sigma_D) \end{bmatrix} \tag{1}$$

where L_x and L_y are respectively derivatives in the x and y direction calculated after smoothing the image by a Gaussian of size σ_D (derivation scale). G_{σ_I} is a Gaussian function of width σ_I (integration scale). As suggested in [10], we set $\sigma_I = 3\sigma_D$.

To detect the characteristic scale, the maxima of the Laplacian of Gaussian function Λ are used [11,4] and is extended to take advantage of different color spaces in [12].

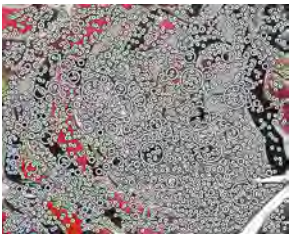
2.2 Blob Detectors

Blob detectors, based on space-scale theory introduced to computer vision by Witkin [13] and extended by Lindeberg [14], rely on differential methods such as Laplacian of Gaussian (LoG), difference of Gaussian (DoG) and Determinant

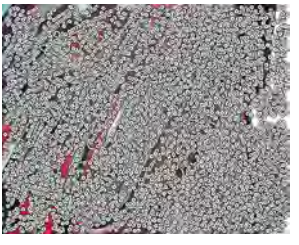
(a) image 1

(b) image 6

Fig. 1. Harris Laplacian detector applied to the *graffiti* testset (see Fig. 7(a)) image 1 (a) and image 6 (b). The size of the circles indicates the size (scale) of the kernel with the highest peak.



(a) image 1



(b) image 6

Fig. 2. DoG applied on image 1 (a) and image 6 (b) of the *graffiti* testset

of Hessian (DoH) [11]. The result of blob detection using either LoG or DoG methods depends on the choice of the scale sampling rate which is analyzed in [6]. Another technique within the class of blob detectors but unrelated to scale-space theory is MSER [5] which is outlined further on.

DoG. As demonstrated in [6], LoG results can be approximated with the DoG at reduced computational complexity.

In this case the Laplacian operator Λ_σ is approximated by the difference between two Gaussian smoothed images. The scale space S_σ is defined by $S_\sigma = G_\sigma \otimes I$. The image pyramid level D_{σ_i} is computed as the difference of the input image I convolved with Gaussian kernels of size $\sigma_i = \left(\sqrt{2}\right)^i$ as $D_\sigma = (G_{\sqrt{2}\sigma} - G_\sigma) \otimes I$. The implementation leads to an early diminished dataset, as the majority of the pixels are discarded at the first scale level.

To discard edges and prioritize corners, the already calculated Hessian matrix is used to process an adapted Harris corner detection algorithm. An example of the resulting interest points is given in Fig. 2.

MSER. [5] are obtained by a watershed-like algorithm. Connected regions of a certain thresholded range are selected if they remain stable over a set of thresholds. The algorithm is efficient both in run-time performance and detection rate. The region priority is measured in the number of thresholds where the region remains stable.

(a) image 1 (b) image 6

Fig. 3. Example of MSER locations applied on the *graffiti* testset image 1 (a) and image 6 (b). The ellipses mark the most stable blobs.

An image region Q is extremal if the intensity of all pixels $q \in Q$ is higher than the intensity of boundary pixels p (adjacent to Q) $I(q) > I(p)$ for maximum intensity regions or lower $I(q) < I(p)$ for minimum intensity regions. Region Q is a contiguous image patch i.e. there is a path S connecting any two pixels $q \in Q$ such that $S \in Q$. Extremal regions are then such that $I(Q_i + \Delta) > I(Q_i) > I(Q_i - \Delta)$. The *maximally stable extremal region* is the one for which variation of the area $q(i)$ has a local minimum at i:

$$q(i) = \frac{|Q_{i+\Delta}| - |Q_{i-\Delta}|}{|Q_i|} \tag{2}$$

Ellipses fitted to the MSER locations can be seen in Fig. 3.

2.3 Symmetry Based Interest Points

The Generalized Symmetry Transform (GST) [15] inspired the *Fast Radial Symmetry Transform* (FRST) by Loy and Zelinsky [16,7]. A pixel of the image contributes to a symmetry measure at two locations called negatively and positively affected pixels. The coordinates of negatively affected $\mathbf{p}_{-ve}$ and positively affected $\mathbf{p}_{+ve}$ pixels are defined by the gradient orientation at pixel $\mathbf{p}$ and a distance n (called *range* in [16]) as follows:

$$\mathbf{p}_{+ve} = \mathbf{p} + \text{round}\left(\frac{\mathbf{g}(\mathbf{p})}{\|\mathbf{g}(\mathbf{p})\|}n\right), \quad \mathbf{p}_{-ve} = \mathbf{p} - \text{round}\left(\frac{\mathbf{g}(\mathbf{p})}{\|\mathbf{g}(\mathbf{p})\|}n\right) \tag{3}$$

The symmetry measure is a combination of orientation projection O_n and magnitude projection M_n maps, which are obtained through agglomeration of positively and negatively affected pixel contributions. Each positively affected pixel increments the corresponding element of the orientation projection map by 1 and magnitude projection map by $\|\mathbf{g}(\mathbf{p})\|$ while the negatively affected pixel decrements the map by these values:

$$O_n\left(\mathbf{p}_{+ve}(\mathbf{p})\right) = O_n\left(\mathbf{p}_{+ve}(\mathbf{p})\right) + 1, \quad O_n\left(\mathbf{p}_{-ve}(\mathbf{p})\right) = O_n\left(\mathbf{p}_{-ve}(\mathbf{p})\right) - 1 \tag{4}$$

$$M_n\left(\mathbf{p}_{+ve}(\mathbf{p})\right) = M_n\left(\mathbf{p}_{+ve}(\mathbf{p})\right) + \|\mathbf{g}(\mathbf{p})\|, \tag{5}$$

$$M_n\left(\mathbf{p}_{-ve}(\mathbf{p})\right) = M_n\left(\mathbf{p}_{-ve}(\mathbf{p})\right) - \|\mathbf{g}(\mathbf{p})\|$$

(a) image 1

(b) image 6

Fig. 4. Example of Loy symmetry points with a simple scale selection applied on the *graffiti* testset image 1 (a) and image 6 (b). The size of the circles indicates the size of the range with a symmetry peak.

The radial symmetry measure at range n is a combination of normalized orientation and magnitude projection maps, additionally smoothed by a Gaussian kernel:

$$S_n = G_{\sigma_n} \otimes \left(\frac{M_n}{k_n}\right)\left(\frac{|O_n|}{k_n}\right)^{\alpha} \tag{6}$$

where k_n is the scale normalization factor and α is the radial strictness parameter which allows to attenuate the symmetry response from ridges. The orientation projection map used for final calculations is thresholded using k_n. The symmetry measure can be also averaged over a set of ranges $N = \{n_1, ...n_K\}$ to achieve partial scale invariance:

$$S = \frac{1}{K}\sum_{n \in N} S_n \tag{7}$$

An exhaustive discussion of the parameter choice and results are presented in [7]. We refer to these interest points as *Loy Points* (see Fig. 4).

3 Gradient Vector Flow Based Interest Points

To detect points of high local symmetry we use the GVF based interest points (GVF points) as proposed in [8]. The GVF [9], which yields a rotation invariant vector field, was originally proposed to increase the capture range of active contours. It is defined as the vector field $\mathbf{v}(x, y) = (u(x, y), v(x, y))$ which minimizes

$$\mathbf{G} = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + u_y^2) + |\nabla f|^2 |\mathbf{v} - \nabla f|^2 dxdy \tag{8}$$

where f denotes the *edge map* of image I

$$f^2(x, y) = |(G_\sigma(x, y) * I(x, y))| \tag{9}$$

and the parameter μ gives the relation between the first smoothing term (compare with the classic optical flow calculation [17]) and the second term. Its strengths include the ability to detect even weak structures while being robust to high amounts of noise in the image. When $|\nabla f|$ is small, the energy yields a very smooth field.

(a) frontal (b) 20 degree (c) 50 degree (d) 1st scale (e) 2nd scale (f) 3rd scale

Fig. 5. GVF points on image details of the *graffiti* testset. (a)-(c): GVF points under geometric transformation. (d)-(f): GVF points of the first three scales of one image detail.

(a) GVF points on image 1 (b) GVF points on image 6

Fig. 6. GVFpoints applied to image 1 (a) and image 6 (b) of the *graffiti* testset

The field magnitude $|\mathbf{G}|$ is largest in areas of high image gradient, and the start and end points of the field lines of $\mathbf{G}$ are located at symmetry maxima. E.g. in the case of a symmetrical structure formed by a homogeneous region surrounded by a different gray level value the field will point away from or towards the local symmetry center of the structure.

The symmetry interest points are thus defined as the local minima of $|\mathbf{G}|$. In contrast to techniques based on estimating the radial symmetry using a sliding window approach this will yield a sparse distribution of interest points even in large homogeneous regions.

We increase μ iteratively thus smoothing $\mathbf{G}$. As we lose information on local structure and $\mathbf{v}$ takes a gradually larger area into account, a rotation invariant scale space pyramid is built. For the experiments, the parameters $\mu = 0.1$ and scale factor $f = 1.33$ are used. We apply the scale factor five times per image smoothing $\mathbf{G}$ for taking more area into account.

Examples of the resulting images are shown in Fig. 5: (d)-(f) show the distributions of the resulting points for increasing scale; (a)-(c) show the interest points on geometric transformations. Further examples are shown in Fig. 6.

4 Results

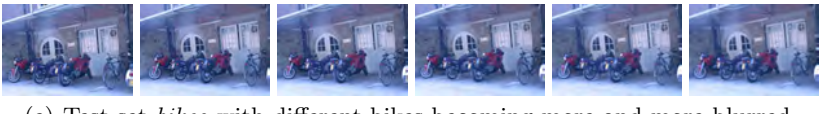
In this section, a performance evaluation of the GVFpoints is given. We show that they outperform current approaches for invariant interest point locations in several important tasks.

(a) Test-set *graffiti* depicts a painted wall under heavy viewpoint changes.

(b) Test-set *boat* changes the viewpoint and the zoom level while rotating the scene.

(c) Test-set *cars* provides a natural scene at different daytimes.

(d) 6 out of 20 images of the test-set *toy*. It provides a natural scene under different lighting directions. Main challenge is the stability against shadowing effects.



(e) Test-set *bikes* with different bikes becoming more and more blurred.

(f) Test-set *ubc* adds JPEG compression to a natural scene.

Fig. 7. One data-set per challenge used in the experiments

We give an overview of the experimental setup and then give a detailed review of the results of the experiments. Mikolajczyk and Schmid [2] suggest a test for the quality of local features. They measure the repeatability of local features under different image transformations. These tests consist of a set of images, where one acts as the reference image and the other images show the same scene under predefined changes like blur, rotation, zoom, viewpoint change, JPEG compression or lighting changes[1].

Having a correct homography H between two images L and L', we can project a point $\mathbf{x}$ in L into the transformed image $\mathbf{x}'$ in L'. Regarding the area of interest points, the overlap of the projected area is estimated. Areas are normalized so

[1] http://lear.inrialpes.fr/people/mikolajczyk/Database

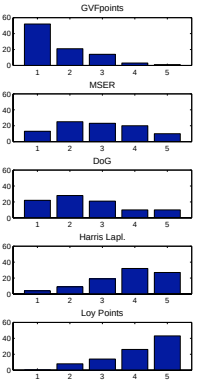
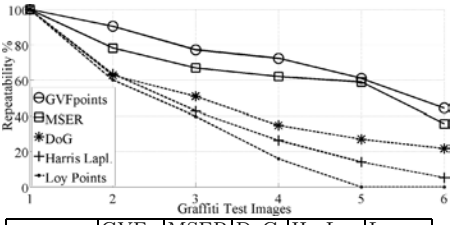


Fig. 8. Histogram of the ranks of the compared algorithms. For each of the 91 test images the algorithms were sorted according to their performance.



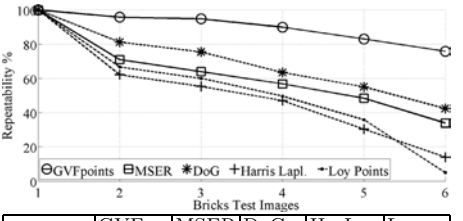
	GVFp	MSER	DoG	HarLap	Loy
ø(rep)	74.2	66.9	49.5	42.0	36.0
corresp	7361.7	215.8	1984	1016.8	474.5
ø(area)	2496.2	1565.7	65.1	4490.5	684.4
std(area)	4273.5	3930.7	149.5	17926.1	1013.8
nr. pts	17349	533	5479	3116	1690

Fig. 9. Repeatability experiment test-set *graffiti* – viewpoint transformation of colourful patterns. GVFpoints perform best for each of the test images.

that each radius is set to 30 pixels. If the overlapping error is below 40% to the nearest neighbour, the interest point is repeated. The repeatability rate is defined as the ratio between the number of detected correspondences and the number of regions that occur in the area common to both images. Feature detectors tend to have higher repeatability rates when they produce a richer description. Twelve commonly used data-sets have been chosen for this evaluation. Examples of the images are shown in Fig.7.

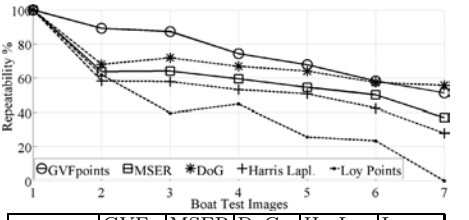
The histograms in Fig. 8 provide a summary view of the ranks of the individual algorithms. Each of the 91 reference image / test image pairs was treated as a separate experiment. For each of these, the algorithms were ranked according to their repeatability from 1 to 5. In 57.1% of the cases the GVFpoints exhibited the best performance (rank=1), while in 80.2% they performed either best or second best (rank$\leq$2). Harris Laplacian and Loy's symmetry points show far lower performance, with Loy having the lowest performance (rank=5) in 47% of cases. MSER and DoG display mixed results: while leading in same cases they fare badly in others, exhibiting an average performance overall.

For one testset per challenge, the repeatability graphs and numerical results are given. These statistics give the means of the repeatability rate, number of correspondent regions, area of the circles in the images in pixel^2, standard deviation of the area and the average number of interest points in the image. GVFpoints show to be repeatable under geometric transformation (Fig.10). Elongated structures like the ones found in the *graffiti* testset (see Fig. 7(a) and Fig. 5) are centered precisely. This works also for MSER, having very well defined blobs on the wall. Therefore, DoG performs also better than the Harris Laplacian as the corners are heavily transformed during the challenge. For Loy points, no repeatability is found for the last two test images. Note the high number of GVFpoints compared to the other approaches because of the elongated structure of



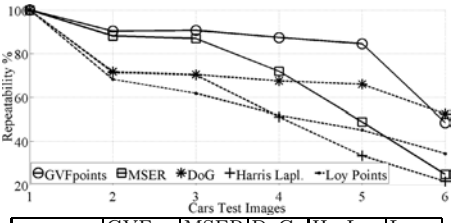
	GVFp	MSER	DoG	HarLap	Loy
ø(rep)	89.9	62.4	69.6	51.5	52.9
corresp	15828.8	1026	8270.8	3036.2	776.5
ø(area)	2481.8	507.8	52.8	1714.9	669.2
std(area)	4252.9	1459.3	65.0	2789.4	1007.6
nr. pts	28093	2363	13604	7599	1600

Fig. 10. Repeatability experiment test-set *bricks* – viewpoint transformation on a highly structured plane



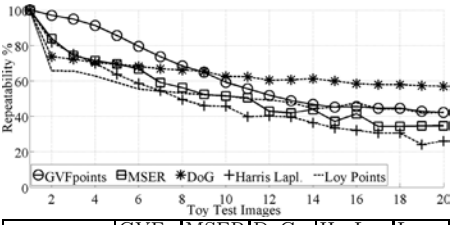
	GVFp	MSER	DoG	HarLap	Loy
ø(rep)	75.5	61.3	69.1	55.8	42.2
corresp	6990.6	497	4969.7	1628.4	502.3
ø(area)	2422.3	1083.8	51.6	3857.0	816.6
std(area)	4206.8	4320.7	33.9	11852.8	1077.8
nr. pts	17625	1524	13797	4961	1803

Fig. 11. Repeatability experiment test-set *boat* for zoom and rotation



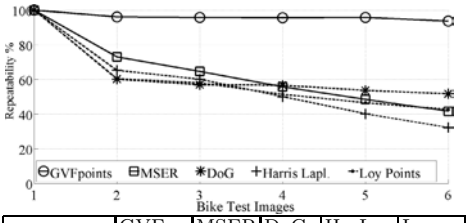
	GVFp	MSER	DoG	HarLap	Loy
ø(rep)	83.5	70.1	71.3	57.9	60.2
corresp	10467.2	90.2	8623	421.5	950.7
ø(area)	2369.7	1395.7	51.4	2345.5	576.4
std(area)	4168.9	3824.4	32.8	4916.2	934.0
nr. pts	14524	172	12477	813	2106

Fig. 12. Repeatability experiment test-set *cars* with increasing illumination



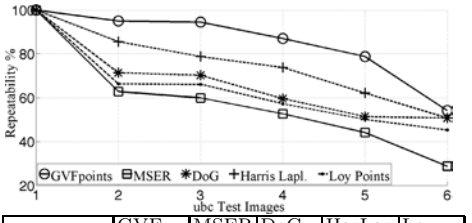
	GVFp	MSER	DoG	HarLap	Loy
mean(rep)	64.1	54.0	65.4	48.9	53.8
corresp	3804.2	56.3	3742.2	348.8	538.4
mean(area)	2493.1	350.8	54.0	3671.8	263.5
std(area)	4264.8	625.5	76.6	11198.4	621.3
nr. points	6995	117	5732	799	1238

Fig. 13. Repeatability experiment test-set *toy* with changing lighting direction



	GVFp	MSER	DoG	HarLap	Loy
mean(rep)	96.1	63.9	63.1	57.9	59.8
corresp	13795.7	193.2	6893	1096.7	898.3
mean(area)	2476.8	580.2	51.4	5894.8	669.2
std(area)	4248.7	1726.0	32.0	25058.1	1007.6
nr. points	19648	616	14382	2111	1600

Fig. 14. Repeatability experiment test-set *bikes* challenging blur



	GVFp	MSER	DoG	HarLap	Loy
mean(rep)	84.9	58.1	67.2	75.2	64.2
corresp	11743.2	415.7	5160.2	1482.3	983.8
mean(area)	2403.7	451.8	51.4	2839.4	669.2
std(area)	4194.2	1708.6	36.8	8376.9	1007.6
nr. points	15388	770	8417	2053	1600

Fig. 15. Repeatability experiment test-set *ubc* for JPEG compression

the blobs, which increases the repeatability rate. On small, often repeated structures like in test-set *bricks*, GVFpoints are able to estimate correspondences for over 75% of all locations, even after 60 degree of transformation (Fig. 9). The experiment on the testset *boat* (see Fig. 7(b)) shows that GVFpoints exhibit higher repeatability at small details, being more invariant to rotational change than other approches. As shown in Fig. 12 and Fig. 13, GVF based points are more stable against changing illumination than all other interest point detectors. Linear illumination change does not affect the GVF to the same degree as the other interest point detectors. For heavy change of lighting, MSER provide slightly more stable locations than GVF. Fig. 14 shows the GVFpoints are almost perfectly invariant to blur. Local noise like the JPEG compression artefacts in testset *ubc* is evaluated in Fig. 15. We show that the GVFpoints provide more stable locations up to the point where the extrema are significantly shifted by the newly introduced structures. Surprisingly, MSER turn out to be very unstable to this kind of noise, where Loy points provide better results. Harris Laplacian points are obviously more stable and perform almost comparably to the DoG.

5 Conclusion

We showed that for the majority of challenges, interest points based on GVF provide more stable locations than the well known and broadly used corner or blob detectors. They give a rich and well-distributed description for diverse visual data. Especially the invariance against linear and arbitrary lighting and illumination changes as well as viewpoint transformations makes the proposed interest points well suited for many problems dealing with rotation, noise, low contrast or heavy compression. These effects often occur in on-line and mobile applications. Medical imaging often deals with low contrast images, where the GVFpoints have advantages compared to corner or blob detectors. The GVF deals with those challenges in a very stable way, being almost invariant to blur and more repeatable towards JPEG compression than other detectors.

Acknowledgment

This work was partly supported by the NB project 12537 (COBAQUO) and the Austrian Research Promotion Agency (FFG) project 815994.

References

1. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: CVPR, pp. 264–271 (2003)
2. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. IJCV 60, 63–86 (2004)
3. Harris, C., Stephens, M.: A combined corner and edge detection. In: AVC, pp. 147–151 (1988)

4. Mikolajczyk, K., Schmid, C.: Indexing based on scale invariant interest points. In: ICCV, pp. 525–531 (2001)
5. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In: BMVC, pp. 384–393 (2002)
6. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60, 91–110 (2004)
7. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. PAMI, 959–973 (2003)
8. Donner, R., Mičušík, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: Proc. BMVC (2007)
9. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. In: PR (1998)
10. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 128–142. Springer, Heidelberg (2002)
11. Lindeberg, T.: Feature detection with automatic scale selection. IJCV 30, 79–116 (1998)
12. Stöttinger, J., Hanbury, A., Sebe, N., Gevers, T.: Do colour interest points improve image retrieval? In: ICIP, pp. 169–172 (2007)
13. Witkin, A.P.: Scale-space filtering. In: IJCAI, pp. 1019–1022 (1983)
14. Lindeberg, T.: Effective scale: A natural unit for measuring scale-space lifetime. In: ISRN KTH (January 1994)
15. Reisfeld, D., Wolfson, H., Yeshurun, Y.: Context free attentional operators: the generalized symmetry transform. JCV (1994)
16. Loy, G., Zelinsky, A.: A fast radial symmetry transform for detecting points of interest. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 358–368. Springer, Heidelberg (2002)
17. Horn, B., Schunck, B.: Determining optical flow. Artiflcial Intelligence 17, 185–204 (1981)