Adaptive LUM Pre- ltering for Af ne Invariant Interest Point Detection

P. Martins, P. Carvalho
Centre of Informatics and Systems
University of Coimbra, Pólo II, 3030-290 Coimbra, Portugal
Email: {pjmm,carvalho}@dei.uc.pt

Abstract **Interest point detection is required by several applications from elds such as computer vision, pattern recognition and content based image retrieval. The performance of the detectors is strictly dependent on the reliability and accuracy with which corresponding points are detected across a wide range of images that are the result of geometric and photometric transformations. This paper presents an adaptive LUM lter and suggests it as the basis of a pre- ltering stage for an af ne interest point detection in order to improve the accuracy and reliability of the detector and to make it suitable for applications that do not require a matching operation.**

Keywords **af ne invariance, feature detection, interest points, LUM lters**

I. Introduction

Image feature detection can be easily found in a wide number of applications from fields such as computer vision, pattern recognition, and content-based image retrieval. Among the several image features that can be detected, local features are indubitably more appropriate to applications involving matching, recognition or reconstruction, since some computation time is saved by using only a spare set of image region points [1]. Furthermore, they can provide robustness to some content transformations such as background clutter and occlusion.

Obviously, some difficulties arise when it is required to the feature detector the identification of repeatable features across a wide range of transformations in the same image, including geometric and photometric transformations and common signal processing operations. In fact, the detection process can become a harder task if it is necessary to provide a high repeatability rate without performing a matching stage, which will be able to remove unmatched features.

This project was partially financed by FCT/PRAXIS XXI (project 36564/2000), and by POSI of the Portuguese Foundation for Science and Technology and the European Union FEDER. The project was performed at the Computacional Perception Laboratory at CISUC.

Among the several local features, interest points (or key points) have shown to be the most suitable to deal with the "invariance issue". There is not a precise definition of interest points; they can be seen as locations where the image content is supposed to be more significant. Typical examples of these descriptors are corners and highly textured regions.

Many interest points detectors have been suggested, following different strategies: gradient based [2][3][4][5][6][7][8][9], morphology based [10][11] or phase-congruency based [12][13].

Among the several proposed methods, the Harris-Stephens corner detector [4] is one the most used due to some reliability when it comes to rotation invariance and slight illumination changes. Recently, some of the presented methods are improved versions of the Harris-Stephens operator. These improvements include an enhanced robustness to illumination changes or invariance to affine transformations [14][15][16], leading to a better repeatability and accuracy of the detected points.

In this paper, the problem of identifying the same points across a wide range of geometric transformations on an image, without performing any matching operation in order to remove unmatched features, is addressed. A pre-filtering operation for affine invariant interest point detection, based on LUM filters [17], is proposed. The inclusion of the proposed solution on an interest point detector makes it suitable for applications such as robust image watermarking, content-based image retrieval and geometric camera calibration.
Recent watermarking schemes, often referred to as second-generation methods [18], are based on the image content. Usually, these algorithms include a feature extraction stage at embedding and extraction phases. In order to obtain an effective robust watermarking method, features should fulfill the following requirements: invariance to noise, covariance to geometrical transformations and localization. Interest points can be applied for this purpose, if a high repeatability rate could be achieved and, consequently, a reduced number of outliers.

Like digital watermarking, content-based image retrieval is an active research field. Some of the most recent solutions propose invariant features to geometric distortions [19][20]. Once again, an effective retrieval method can be designed by using affine invariant interest points as features, without the need of performing a matching operation.

Obviously, this filtering phase can also be applied to other tasks such as gemetric camera calibration or any other computer vision tasks which include a feature matching stage in order to reduce the number of outliers and, consequently, reducing the amount of computation needed to accomplish the matching process.

The remainder of this paper is organized as follows: Section II outlines the principles of Harris and Stephens interest point detector and its adaptation to a more robust detection, in Section III a description of the proposed filtering stage is provided. Results of the proposed method are presented in Section IV. In Section V conclusions are given.

II. Interest point detection

A. Harris-Stephens detector

Motivated by the work Moravec [2] and Förstner [3], Harris and Stephens [4] introduced a corner and edge detector which uses the second moment matrix, also called the auto-correlation matrix, in order to describe the gradient distribution in a local neighborhood of a given pixel $\mathbf{x}$. This matrix is given by Equation (1):

$$M_{HS}(\mathbf{x}) = \begin{bmatrix} L_x^2(\mathbf{x}) & L_yL_x(\mathbf{x}) \\ L_xL_y(\mathbf{x}) & L_y^2(\mathbf{x}) \end{bmatrix}, \qquad (1)$$

where L_x and L_y are the image derivatives computed in the x and y direction, respectively. From matrix M_{HS}, the corner measure R_{HS} is obtained:

$$R_{HS}(\mathbf{x}) = det(M_{HS}(\mathbf{x})) - \alpha trace(M_{HS}(\mathbf{x}))^2, \qquad (2)$$

where α is a parameter typically set to 0.04 as suggested by Harris and Stephens in [4]. Local maxima of R_{HS} give the location of interest points.

Most of the versions of this operator convolve the squared image derivatives with a Gaussian kernel in order to reduce interest points detected due to noise.

B. Af ne invariant interest point detection

Mainly due to its reliability concerning robustness to rotation and slight illumination changes, the Harris-Stephens operator has been widely used and it has also been the basis of several other detectors which try to obtain an invariant detection, specially, when it comes to geometric, namely uniform scale changes and photometric transformations like the viewpoint change [21][14][15]. A quite effective solution is the approach suggested by Mikolajczyk and Schmid [15] which tries to detect truly invariant points to any kind of affine transformations. Invariance to scaling (rotation and translation) is achieved by introducing a scale-space representation for the Harris-Stephens operator with pre-selected scales. Locations at which the Laplacian attains a maximum over scales are selected. Invariance to other affine transformations is provided by estimating the affine shape of a pixel neighborhood derived from the second moment matrix. This method can be described as an iterative process that converges to affine interest points by modifying the location, scale and shape of the initial points identified by the Harris-Stephens method.

The second moment matrix M at pixel $\mathbf{x}$ in the affine scale-space is defined by:

$$M(\mathbf{x}, \Sigma_I, \Sigma_D) = \det(\Sigma_D) g(\Sigma_I) * \begin{bmatrix} L_x^2(\mathbf{x}, \Sigma_D) & L_yL_x(\mathbf{x}, \Sigma_D) \\ L_xL_y(\mathbf{x}, \Sigma_D) & L_y^2(\mathbf{x}, \Sigma_D) \end{bmatrix}, \qquad (3)$$

where Σ_I and Σ_D are the covariance matrices which determine the size of the integration and differentiation Gaussian kernels, respectively; L_x and L_y are the image derivatives in the x and y direction, respectively, computed with a Gaussian kernel whose size was determined by Σ_D; and g is a Gaussian kernel determined by Σ_I.

III. Proposed method

As stated early, local features, namely interest points, are suitable for many applications specially due to their robustness to some content changes. However, robustness to geometric and photometric transformations is difficult to obtain. The methods which are able to overcome this problem are usually applied on matching applications. Consequently, the presence of outliers, i. e., points which are only detected in some of the transformed images, is a situation that matching algorithms have to deal with. Although, the presence of such points is critical if it is required to identify the same points across a wide range of transformations on an image, without performing any matching operation due to, for example, the absence of other versions of the same image or, simply, to avoid a higher computational complexity of the process.

To overcome this problem, we propose a filtering operation which precedes the affine interest point detection described in Section II. The solution relies on a filtering stage based on Lower-Upper-Middle filters [17], a well-known class of rank-order-based filters exhibiting a good level of detail preservation and robustness to noise. These filters are characterized by two parameters, besides the window size, responsible for the definition of the levels of image sharpening and smoothing.
Given an image $I(x,y)$ and a window function of size $(2m+1) \times (2m+1) = N$, centered about the pixel (u,v) and

$$I_{(1)} \leq I_{(2)} \leq \ldots \leq I_{(N)}, \qquad (4)$$

the rank-ordered set of the N samples centered about the aforementioned pixel, the output of the LUM filter at the pixel (u, v), with parameters k, l and a $(2m+1) \times (2m+1)$ window size, is

$$y^{\star} = \begin{cases} I_{(k)}, & \text{if } I(u,v) < I_{(k)} \\ I_{(l)}, & \text{if } I_{(l)} < I(u,v) \leq t_{(l)} \\ I_{(N-l+1)}, & \text{if } t_l < I(u,v) \leq I_{(N-l+1)} \\ I_{(N-k+1)}, & \text{if } I_{(N-k+1)} < I(u,v) \\ I(u,v), & \text{otherwise} \end{cases}, \quad (5)$$

where $t_l = \frac{I_{(l)} + I_{(N-l+1)}}{2}$.
By analyzing the filter output, it is seen that LUM filters can perform as pure sharpeners, by setting $k = 1$ or as smoothers, by setting $l = \frac{(N+1)}{2}$. The idea behind the proposed solution is to set adaptively both parameters in order to obtain an invariant interest point detection which can identify a more similar set of points between images representing the same scene. Herein, similar means, besides repeatable and accurate locations, a reduced number of outliers.

The smoothing performed by Gaussian kernels in order to reduce noise can easily corrupt the identification of corners, on the other hand, the interpolation methods employed by the geometric transformations tend to introduce artifacts, specially by less accurate methods such as nearest neighbor and bilinear interpolation [22]. The pre-filtering will try to reduce the effects caused by the Gaussian smoothing and the interpolation. A lower smoothing level will be applied to points which exhibit a higher gradient variation. This will avoid that a large number of possible interest points candidates to be discarded. However, if the variation of the gradient variation in the pixel neighborhood is high, this means that the surrounding region is subject to a less accurate interpolation. Therefore, the sharpening level will be lower in this case.

In order to set the filter parameters according to the aforementioned ideas, the *m-neighborhood* of a pixel (x, y) is defined:

$$\mathcal{N}_m(x,y) = \{(x',y') | 0 < \|(x,y) - (x',y')\|_{\infty} \leq m\}, \quad (6)$$

where $\|\cdot\|_{\infty}$ is the maximum norm. From the *m-neighborhood*, the *m-neighborhood distortion* is derived:

$$\mathcal{D}_m(x,y) = \max_{(x',y') \in \mathcal{N}_m(x,y)} \{|I(x',y') - I(x,y)|\}, \quad (7)$$

and the *m-neighborhood normalized distortion* is defined as

$$\tilde{\mathcal{D}}_m(x,y) = \frac{\mathcal{D}_m(x,y)}{\max_{(x',y') \in \{1,\ldots,n1\} \times \{1,\ldots,n2\}} \{\mathcal{D}_m(x',y')\}}, \quad (8)$$

for a $n1 \times n2$ image I. Let $\tilde{\mathcal{D}}_m^1$ be the matrix containing the *m-neighborhood normalized distortion* of image I and $\tilde{\mathcal{D}}_m^2$ the matrix containing the *m-neighborhood normalized distortion* of $\tilde{\mathcal{D}}_m^1$.

For each pixel in the image, the sharpening parameter is defined according to (9).

$$k(x,y) = \max\{1, round(\tfrac{N+1}{2} \tilde{\mathcal{D}}_m^2(x,y))\}. \quad (9)$$

In a similar way, the parameter l – the smoothing parameter – is defined for each pixel position according to (10).

$$l(x,y) = \max\{1, round(\tfrac{N+1}{2}(1 - \tilde{\mathcal{D}}_m^1(x,y)))\} \quad (10)$$

IV. Experimental Results and Discussion

The experiments focused on geometric transformations due to the severe effects that they have on the performance of the interest point detectors. Three 512x512 test images were used (Fig. 1) to evaluate the performance of the proposed pre-filtering stage on affine interest point detection. For each test image, the following transformations were applied: clockwise rotations of 5, 15 and 25 degrees, and scale changes with scale factors of 1.5, 0.8 and 0.5, using bilinear interpolation. After that, two versions of each of the resulting images and original ones were created: a filtered version using the proposed filter and a non-filtered one. The affine interest point detection was then applied to the whole set of images.

For each interest point (x, y) detected on the original test images (filtered and non-filtered versions), the corresponding points after a scaling or a rotation of θ degrees were estimated by means of equations (11) and (12), respectively.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (12)$$

For each one of these operations, we got the transformed points (x^t, y^t). Let $\mathcal{A}$ be the set of interest points detected on the transformed image which corresponds to each one of the aforementioned operations and $\mathcal{B}$ the set of points (x^t, y^t). From these sets, the sub-sets $\mathcal{A}_c$ and $\mathcal{B}_c$ were obtained, with $\#\mathcal{A}_c = \#\mathcal{B}_c = \min\{\#A, \#B\}$ and the sub-set derived from the set with higher number of elements contained the $\min\{\#A, \#B\}$ strongest points according to the corner measure defined in Equation (2). Then for each point (x^t, y^t) of $\mathcal{B}_c$, the distance to the sub-set $\mathcal{A}_c$ was computed by applying Equation (13):

$$d((x^t, y^t), \mathcal{A}_c) = \min_{(x,y) \in \mathcal{A}_c} \|(x^t, y^t) - (x, y)\|, \quad (13)$$

where $\|\cdot\|$ is the Euclidean norm. A point was considered matched if its corresponding distance was inferior to 1.5. The idea of matching only a certain number of points (the strongest ones, according to Harris-Stephens measure) intends to show that the proposed filtering operation has the property of reducing the difference on the number of interest points detected across the several geometrically transformed versions of a given image and, most of all, the sets of strongest points detected in two transformed versions do not differ in too much locations, i. e., the filtering operation reduces the number the outliers and does not have a severe effect on the interest point response defined by Harris and Stephens.

This property also emphasizes that the presented pre-filtering can be suitable for applications, such as image/video watermarking, which use only a reduced number of interest points, usually the most robust ones according to some measure.

Figures 2 and 3 depict the detected points on some of the test images. Tables I-VI present the whole set of results . The column `#Reference points` refers to the number of points detected on the original test images. The number of iterations performed by the affine interest point detection was limited to 20 for each initial interest point detected with a derivation scale set to 1 and the integration scale set to 2. The area of the *m-neighborhood* and the window size were set to 3×3 for smaller images (256×256 pixels) and set to 7×7 for larger images (768×768 pixels). For the remaining images, these parameters were set to 5×5 pixels.
The results given by Tables I-VI show that the inclusion of the proposed pre-filtering reduces the discrepancy on the number of interest points detected across the several geometrically transformed versions of the initial test images and sometimes improving the accuracy of the affine interest detector.

TABLE I.
RESULTS FOR SCALING ON ”F16” IMAGE.

”F16” image				
Scale factor	Filtering	#Points detected	#Reference Points	#Mismatches
1.5	×	86	54	0
	√	43	33	0
0.8	×	34	54	10
	√	36	33	0
0.5	×	12	54	10
	√	12	33	10

TABLE II.
RESULTS FOR ROTATION ON ”F16” IMAGE.

”F16” image				
Angle (degrees)	Filtering	#Points detected	#Reference Points	#Mismatches
5	×	44	54	11
	√	33	33	7
15	×	43	54	0
	√	36	33	0
25	×	46	56	15
	√	40	33	0

V. CONCLUSIONS AND FUTURE WORK

In this paper, a simple adaptive LUM pre-filtering stage for interest point detection has been presented. The main motivation for the inclusion of the proposed pre-filtering operation is to make interest point detectors more appropriate for tasks that do not perform a matching operation, such as image watermarking, by an effective removal of outliers, resulting on a better repeatability of interest points across a set of geometrical transformations over a given image.
Experimental results have demonstrated that the proposed method improves the repeatability of the affine interest point detector, by reducing outliers, and it does not affect the accuracy of the detector.

Further research directions are aimed at the development of an automatic selection of the window size.

TABLE III.
RESULTS FOR SCALING ON ”FISH” IMAGE.

”Fish” image				
Scale factor	Filtering	#Points detected	#Reference Points	#Mismatches
1.5	×	22	19	0
	√	14	15	0
0.8	×	29	19	0
	√	16	15	0
0.5	×	45	19	0
	√	31	15	0

TABLE IV.
RESULTS FOR ROTATION ON ”FISH” IMAGE.

”Fish” image				
Angle (degrees)	Filtering	#Points detected	#Reference Points	#Mismatches
5	×	19	19	5
	√	16	15	0
15	×	15	19	2
	√	20	15	0
25	×	15	19	6
	√	15	15	0

REFERENCES

[1] B. Triggs, “Detecting Keypoints with Stable Position, Orientation and Scale under Illumination Changes,” in *European Conf. on Computer Vision*, May 2004, vol. 4, pp. 100–113.

[2] H. P. Moravec, “Towards automatic visual obstacle avoidance,” in *Proc. of the 5th International Joint Conference on Artificial Intelligence*, 1977, p. 584.

[3] W. Förstner and E. Gülch, “A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Circular features,” in *Proceedings of the Intercommission Conference on Fast Processing of Photogrametric Data*, 1987, pp. 281–305.

[4] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proc. of the 4th Alvey Vision Conference*, 1988, pp. 147–151.

[5] C. Schmid, R. Mohr, and C. Bauckhage, “Evaluation of interest point detectors,” *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.

[6] O. Zuniga and R. Haralick, “Corner detection using the facet model,” in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 1983, pp. 30–37.

[7] L. Kitchen and A. Rosenfeld, “Gray-level corner detection,” *Patt. Rec. Lett.*, vol. 1, pp. 95–102, 1982.

[8] R. Deriche and G. Giraudon, “A computational approach for corner and vertex detection,” *Int. J. Computer Vision*, vol. 10, no. 2, pp. 101–124, 1993.

[9] J. A. Noble, “Finding corners,” *Image & Vision Computing*, vol. 6, no. 2, pp. 121–128, 1988.

[10] S. M. Smith and J. M. Brady, “SUSAN - A New Approach to Low Level Image Processing,” *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[11] R. Laganière, “Morphological corner detection,” in *Int. Conf. Computer Vision*, 1998, pp. 280–285.

TABLE V.
RESULTS FOR SCALING ON "WATER MILL" IMAGE.

"Water mill" image				
Scale factor	Filtering	#Points detected	#Reference Points	#Mismatches
1.5	×	32	26	0
	√	16	13	0
0.8	×	20	26	8
	√	17	13	0
0.5	×	13	26	5
	√	13	13	5

TABLE VI.
RESULTS FOR ROTATION ON "WATER MILL" IMAGE.

"Water mill" image				
Angle (degrees)	Filtering	#Points detected	#Reference Points	#Mismatches
5	×	29	26	0
	√	14	13	0
15	×	28	26	0
	√	13	13	0
25	×	30	26	0
	√	23	13	0

[12] P. Kovesi, "Phase Congruency Detects Corners and Edges," in *Proc. of The Australian Pattern Recognition Society Conference: DICTA 2003*, 2003, pp. 309–318.

[13] D. Reisfeld, "The constrained phase congruency feature detector: Simultaneous localization, classification,and scale determination," *Patt. Rec. Lett.*, vol. 17, pp. 1161–1169, 1996.

[14] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *European Conf. on Computer Vision*, 2002, vol. 1, pp. 128–142.

[15] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[16] Y. Dufournaud, C. Schmid, and R. Horaud, "Matching Images with Different Resolutions," in *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*. June 2000, pp. 612–618, IEEE Computer Society Press.

[17] R. C. Hardie and C. Boncelet, "LUM filters: a class of rank-order-based filters for smoothing and sharpening," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 41, no. 3, pp. 1061–1076, March 1993.

[18] M. Kutter, S. K. Bhattacharjee, and T. Ebrahimi, "Towards second generation watermarking schemes," in *1999 Int. Conf. on Image Processing, ICIP 99*. 1999, vol. 1, pp. 320 – 323, IEEE Computer Society Press.

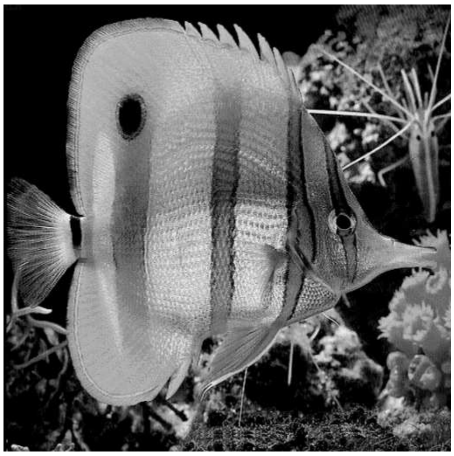
[19] M.-C. Lee and Chi-Man Pun, "Rotation and scale invariant wavelet feature for content-based texture image retrieval," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 1, pp. 68–80, January 2003.

[20] S. Derrode, M. Daudi, and F. Ghorbel, "Invariant content-based image retrieval using a complete set of Fourier-Mellin descriptors," in *Proc. of the IEEE Int. Conf. on Multimedia Computing and Systems*, 1999, vol. 2, pp. 877–881.

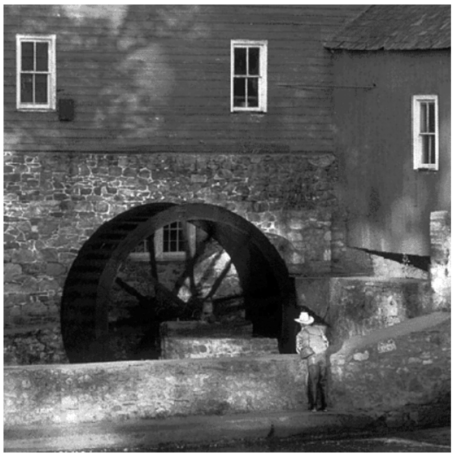
[21] T. Tuytelaars and L. V. Goll, "Content-based image retrieval based on local affinely invariant regions," in *Proc. Int. Conf. on Visual Information Systems*, 1999, pp. 493–500.

[22] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: interpolation methods in medical image processing," *IEEE Trans. on Medical Imaging*, vol. 18, no. 11, pp. 1049–1075, November 1999.

(A)

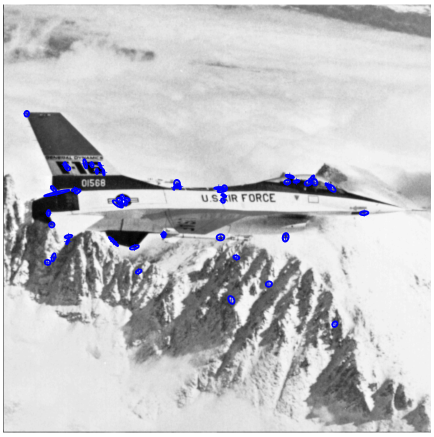


(B)

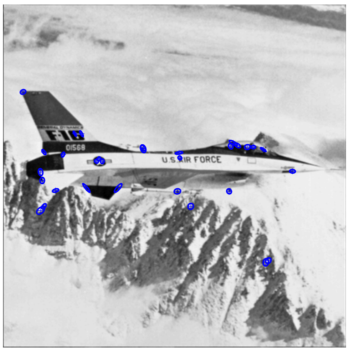


(C)

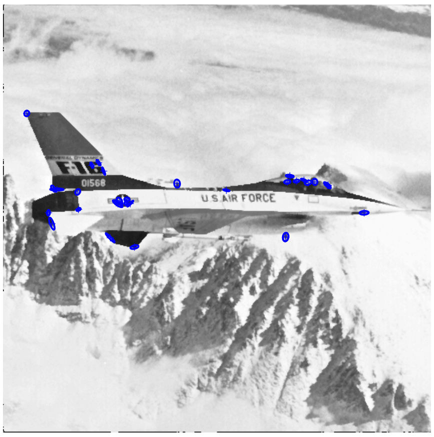
Fig. 1. INITIAL TEST IMAGES: (A) "F16", (B)"FISH", (C)"WATER MILL".



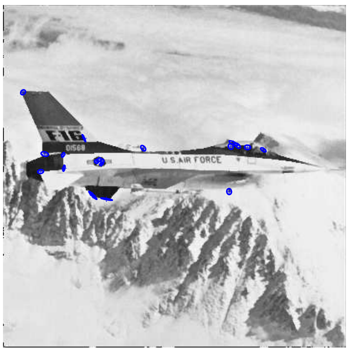
(A)



(B)

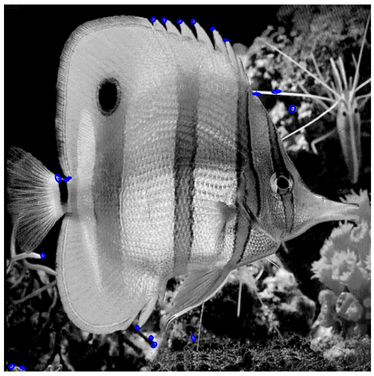


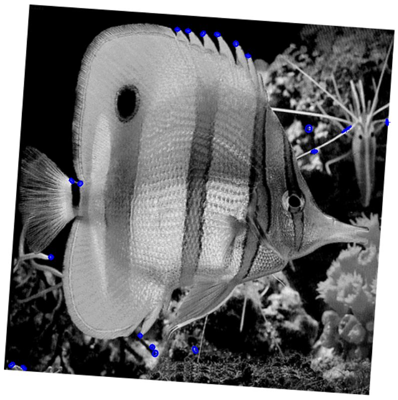
(C)



(D)

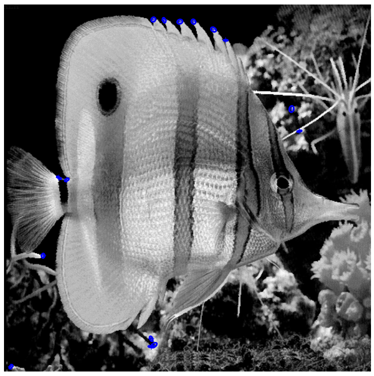
Fig. 2. "F16" IMAGE AND ITS HARRIS-AFFINE INTEREST POINTS: (A) ORIGINAL; (B) RESIZED TO 410X410 PIXELS, (C) ORIGINAL (PRE-FILTERING WITH PROPOSED FILTER), (D) RESIZED TO 410X410 PIXELS (PRE-FILTERING WITH PROPOSED FILTER).

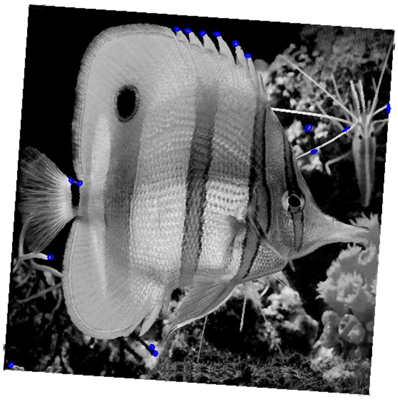




(A)

(B)





(C)

(D)

Fig. 3. "FISH" IMAGE AND ITS HARRIS-AFFINE INTEREST POINTS: (A) ORIGINAL; (B) ROTATED 5 DEGREES, (C) ORIGINAL (PRE-FILTERING WITH PROPOSED FILTER), (D) ROTATED 5 DEGREES (PRE-FILTERING WITH PROPOSED FILTER).