Constant tangential angle elected interest points

Ahmed REBAI
INRIA Rocquencourt
Ahmed.Rebai@inria.fr

Alexis JOLY
INRIA Rocquencourt
Alexis.Joly@inria.fr

Nozha BOUJEMAA
INRIA Rocquencourt
Nozha.Boujemaa@inria.fr

ABSTRACT

This paper presents a new interest points detector whose goal is to better catch visual attention than standard detectors. It is based on the generalization of Loy and Zelinsky transform [18] which was originally developed to detect radial symmetry centers. The voting algorithm is here extended to a set of voting orientations which allows the construction of a more complete analyze space. The specificity of this detector is that it enables the detection of points with different topological natures with the same detector. Centers of geometric forms or curves as well as junctions or vanishing points can be detected depending on the image content. The paper describes and illustrates the principle of this new algorithm and discusses some properties and improvements. Preliminary experiments are also given and shows that the detected points are stable under several typical transformations and that they are promising in terms of objects generic subparts detection.

1. INTRODUCTION AND RELATED WORK

To improve automatic processing, scientists are often inspired by natural principles. Like human visual system which focus on some specific points more than others, researchers have used interest points to solve various problems in computer vision such as object recognition [17, 13], wide baseline matching for stereo pairs [1], image retrieval [25], etc. Many different interest points detectors exist in the literature : corners and junctions detectors [9, 10, 7], wavelet-based detectors [2, 11], phase-based detectors [3], Difference of Gaussian based detectors [22, 17], color based detectors [8], etc. To improve the relevance of the detected points, invariance to view point changes has also been widely studied. The scale-space theory [15, 6] has been prolific in the design of scale-invariant detectors [19] and several affine-invariant detectors have been recently compared in [19] and [20]. The evaluation studies typically use two common criteria: repeatability and matching score under viewpoint changes or image transformations [27, 20]. The repeatability means that the points remain stable after changing the imaging conditions or transforming the image and the matching score gives a distinctiveness measure of a given local feature computed around each interest point (it is thus more indicative than quantitative).

These metrics are however not sufficient for several applications and particularly for generic objects recognition or object class recognition methods based on local features [21, 5]. In such applications, generalization properties as well as invariance properties are required. That means that the points should be detected at the same location in two different objects of the same category. Such a criterion is highly related to the human interpretation of the detected points and is thus more difficult to assess. The evaluation is thus classically performed after the learning selection of relevant local features and is thus dependent on the features computed around the interest points and on the learning strategy. It is however well known that only a very few percentage of the initial detected features are used in the final object class model. This tends to prove that most of the initial detected points are unsuitable for such tasks and this amount of undesirable points is problematic both in term of computation time and efficiency. It is thus essential to reduce the input of local features by improving their visual interpretability.

A number of context-free attentional operators have been proposed in that way for automatically detecting points of interest in images. These operators have tended to use local radial symmetry as a measure of interest [18]. This correlates well with psychophysical findings on fixation points of the human visual system. It has been observed that visual fixations tend to concentrate along lines of symmetry [16]. Sela and Levine [28] noted that some psychophysical findings corroborated this, placing the mean eye fixation points at the intersection of lines of symmetry on a number of simple 2D geometric figures. It has also been observed that visual fixations are attracted to centers of mass of objects and that these centers of mass are more readily determined for objects with multiple symmetry axes [18]. Stark and Pritevera [23] compared the responses of a number of artificial region of interest detectors, including Reisfeld's generalized symmetry transform [24], with regions of interest detected by human subjects. By using several different algorithms in conjunction with a clustering procedure, they were able to predict the locations of human-detected regions of interest. The approach developed by Loy and Zelinsky was

inspired by these results, although the final method bares more similarity to the work of Sela and Levine [28] and the circular Hough transform [12]. It determines the contribution each pixel makes to the symmetry of pixels around it, rather than considering the contribution of a local neighborhood to a central pixel. Unlike previous techniques that have used this approach [12, 28, 14], it does not require the gradient to be quantized into angular bins, the contribution of every orientation is computed in a single pass over the image.

In this paper, we propose a generalization of Loy and Zelinsky's method. By adding an orientation parameter to the voting vectors, we can construct a 3-dimensional voting space in which the maxima are expected to be more stable and more topologically diversified than the radial symmetry centers initially detected. We also provide some improvements by adding an alignment rejection criterion avoiding unstable points on contours and a sub-pixel location estimation at low-resolution to detect points corresponding to large objects without strongly degrading the computation time. Section 2 of this paper gives the description of the proposed method and discusses the nature of the detected points. Improvements of the method are presented in section 3. Section 4 relates some promising preliminary experiments and section 5 discusses prospective works and upcoming full experiments.

2. CONSTANT TANGENTIAL ANGLE ELECTED INTEREST POINTS

2.1 Generalization of Loy and Zelinsky's transform

Loy and Zelinsky's transform is computed over a set of radii $\{1, \cdots, r_{max}\}$ where r_{max} is the maximum radius of the radially symmetric features to be detected. The value of the transform at radius $r \in \{1, \cdots, r_{max}\}$ indicates the contribution to radial symmetry of the gradients a distance r away from each point. While the transform can be calculated for a continuous set of radii, this is generally unnecessary as a subset of radii is normally sufficient to yield a representative result. At each range r, an orientation projection image O_r is formed. It is generated by examining the gradient $\mathbf{g}(\mathbf{p})$ at each point $\mathbf{p}$ of the image from which two corresponding pixels, located at distance r and pointed by the gradient direction, are affected [18].
In our new generalized transform, the direction of this vote is not systematically the gradient orientation and can be tuned by a parameter θ defining the angle between the gradient and the new voting vector $\mathbf{g}_\theta(\mathbf{p})$. An overview of the algorithm is shown in Figure 2. At each radius $r \in \{1, \cdots, r_{max}\}$ and each $\theta \in \{\theta_{min}, \cdots, \theta_{max}\}$, an orientation projection image $O_{r,\theta}$ is formed. It is generated by examining the gradient $\mathbf{g}(\mathbf{p})$ at each point $\mathbf{p}$ of the image from which two corresponding affected pixels are determined as shown in Figure 1. These affected pixels are defined as that the voting vector $\mathbf{g}_\theta(\mathbf{p})$ is pointed to or away from, a distance r away from $\mathbf{p}$. The voting vector is defined as:

$$\mathbf{g}_\theta(\mathbf{p}) = \begin{matrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{matrix} \quad \mathbf{g}(\mathbf{p}) \tag{1}$$

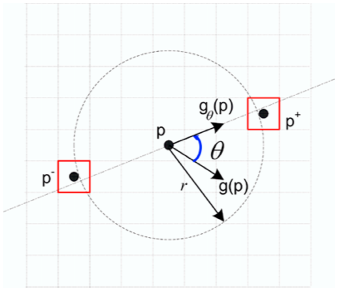


Figure 1: The locations of pixels $\mathbf{p}^+$ and $\mathbf{p}^-$ affected by the elector vector $\mathbf{e}(\mathbf{p})$ for a range of $r = 3$. The dotted circle shows all the pixels which can be affected by the gradient at $\mathbf{p}$ for a range r.

The coordinates of affected pixels $\mathbf{p}^+$ and $\mathbf{p}^-$ are given by

$$\mathbf{p}^+ = \mathbf{p} + round \quad r \cdot \quad \cos(\beta + \theta)\mathbf{x} + \sin(\beta + \theta)\mathbf{y} \tag{2}$$

$$\mathbf{p}^- = \mathbf{p} - round \quad r \cdot \quad \cos(\beta + \theta)\mathbf{x} + \sin(\beta + \theta)\mathbf{y} \tag{3}$$

where $round(.)$ rounds each vector element to the nearest integer, $\mathbf{x}$ and $\mathbf{y}$ are the unit vectors according to the width and the height of the image respectively:

$$\mathbf{x} = \begin{matrix} 1 \\ 0 \end{matrix} \qquad \mathbf{y} = \begin{matrix} 0 \\ 1 \end{matrix}$$

and $\beta = \arctan \frac{\mathbf{g}(\mathbf{p})\cdot\mathbf{y}}{\mathbf{g}(\mathbf{p})\cdot\mathbf{x}}$ is the orientation of the gradient vector.

The orientation image $O_{r,\theta}$ is initially zero. Each affected pixel in the orientation projection image is incremented by 1:

$$O_{r,\theta}(\mathbf{p}^+) = O_{r,\theta}(\mathbf{p}^+) + 1 \tag{4}$$

$$O_{r,\theta}(\mathbf{p}^-) = O_{r,\theta}(\mathbf{p}^-) + 1 \tag{5}$$

The contribution at a radius r and an orientation θ is then defined as the convolution

$$S_{r,\theta} = (O_{r,\theta})^4 * G_r$$

where G_r is a two-dimensional Gaussian. The purpose of G_r is to spread the influence of the affected pixels as a function of radius r. G_r comes to correct the inaccuracy of the gradient direction made by calculation. More than the vote will be further (r increase), more than the standard deviation of the Gaussian kernel will be larger. The purpose of the exponent 4 is to spread out the dynamics of the orientation projection image.

The final contribution at a given orientation θ is then defined as the sum of the contributions over all the considered radii:

$$S_\theta = \sum_{r=1}^{r_{max}} S_{r,\theta} \tag{6}$$

2.2 Points detection and analysis

At this stage, we have to notice that the transform could be theoretically defined for a continuous set of orientations θ although this is in practice unnecessary as a subset of orientations is normally sufficient to yield a representative

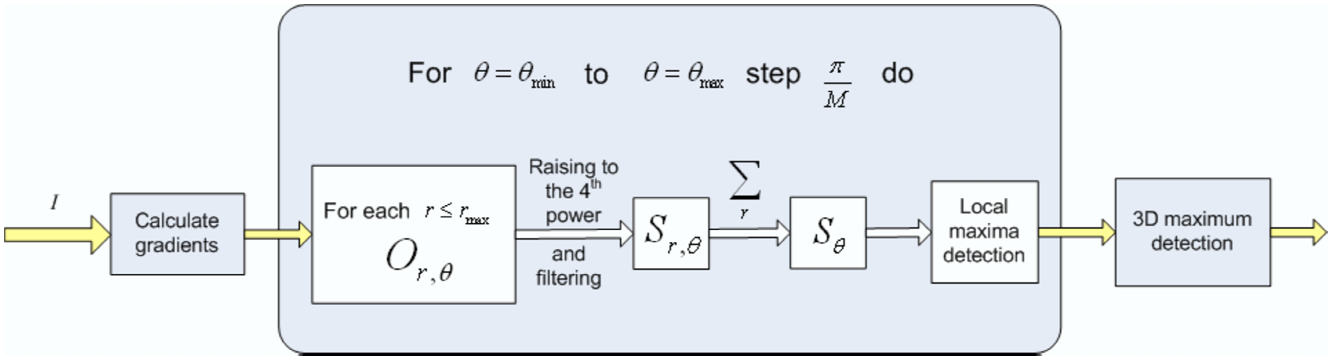


Figure 2: The algorithm process.

result. The definition of this 3-dimensional continuous space (that we call θ-space) is however very interesting since it can be compared to a scale-space although we did not attend to verify the diffusion equation for the moment ($\frac{\delta S}{d\theta} = \Delta S$). Sampled images of such a θ-space are presented in Figure 4.

Given the set of contribution images $\{S_\theta\}$, interest points could be detected by searching for $2D$ local maxima in each of the S_θ images, as in the original version of Loy and al., for which the points were detected as local maxima in the unique contribution image S_0. This would effectively detect interesting points of different topological natures but would also lead to a lot of correlated interest points. One given visually interesting point can indeed correspond to a local maximum at several θ due to the blurring equation 2.1. We thus propose to keep as interest points $\{\mathbf{p}_i\}$, only the points presenting a 3-dimensional local maximum in the θ-space, as usually done in scale-space theory. We can hope these points to be more stables and less correlated. In practice, to speed up the process, the $3D$ maxima are selected among all the $2D$ maxima detected in the contribution images S_θ ($\theta \in \{\theta_{min}, \cdots, \theta_{max}\}$).

The orientation θ_d at which an interest point is detected is an important information about the geometric distribution of its voters. We thus define the **characteristic tangential angle** ψ_d of an interest point as:

$$\psi_d = \frac{\pi}{2} - \theta_d \tag{7}$$

This angle is meaningfull since it determines the constant tangential angle of all the voters that have contributed to a given interest point. It is important to notice that the mathematical object defined by a constant polar tangential angle is a logarithmic spiral whose polar equation is:

$$\rho(\alpha) = k \cdot e^{\cot(\psi)\cdot\alpha} \tag{8}$$

We can thus expect as interest points, centers of logarithmic spirals, or centers of portions of logarithmic spirals, as well as intersections of such curves. When a point is detected with a characteristic tangential angle $\psi_d = \frac{\pi}{2}$, the tangent at the voting points is expected to be perpendicular to the interest point direction and the optimal detected shape will be a circle. When a point is detected with a characteristic tangential angle $\psi_d \equiv 0\,[\pi]$, tangent at the voting points are expected to point toward the interest point, and we can expect convergent contours, such as junctions or vanishing points (see Figure 3).

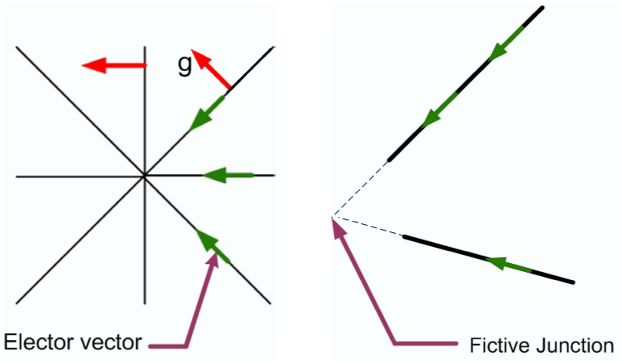


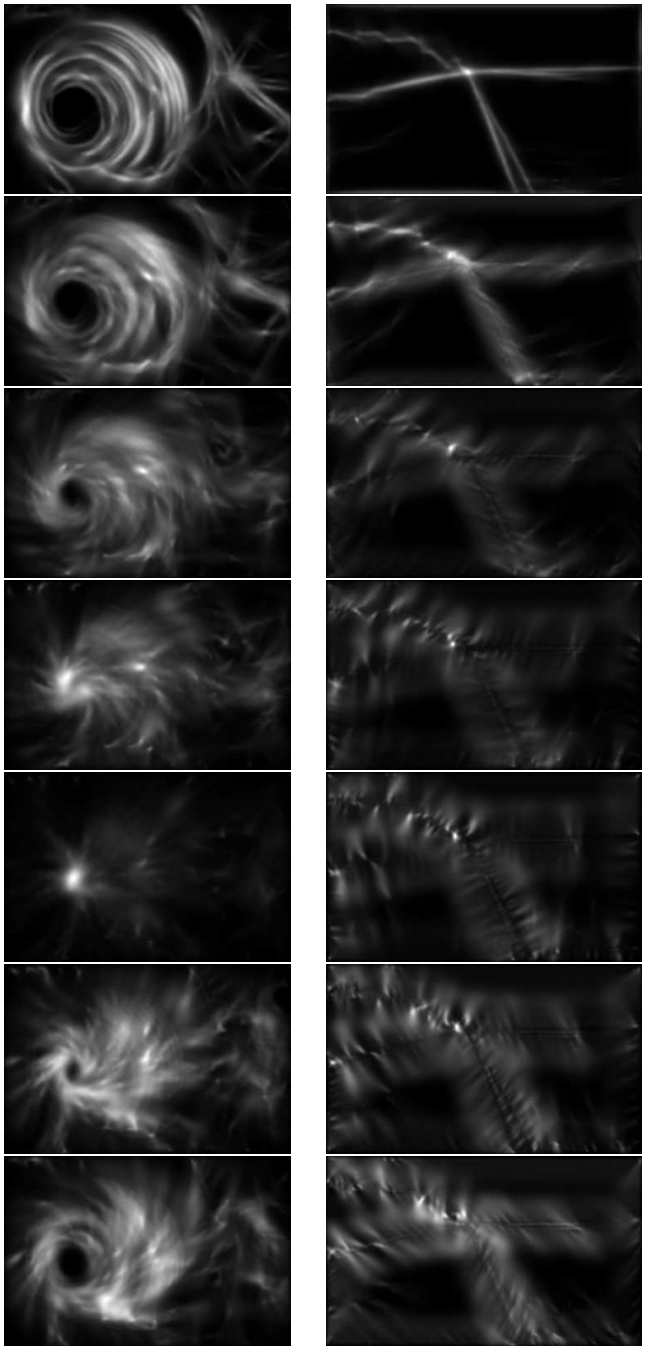
Figure 3: Detection of junctions or vanishing points

Figure 4 shows sampled images of the θ-space of two real images, illustrating the different nature of the points when the characteristic tangential angle changes. One can easily note the smooth displacement of the $2D$ local maxima in the θ-space and the energy concentration at θ_d ($\theta_d = 0$ in the left image and $\theta_d = \frac{\pi}{2}$ in the right image). The interest point with the highest response is also displayed on Figure 4 for each image and corresponds well to the main visual attraction centers of these images although they are topologically very different. Main elected points detected in other images and at other characteristic tangential angles are displayed on Figure 5 to illustrate other typical detected features.

2.3 Robustness to occlusion

Figure 6 illustrates another interesting property of the proposed interest points, that is their robustness to occlusions. We see that the vanishing point is detected as the main elected interest point even after the insertion of a hidding object. This property is due to the fact that most of the voting vectors are not necessary in the direct neighborhood of the interest point. This is quite original compared to usual interest points which mainly focus on the local content itself. In our case, interest points are selected according to how they are seen by other points in the image more than according to their local neighborhood.

(a)



(b)

Figure 4: (a): Original images and best elected interest point (b): θ-space sampled images (from top to bottom: $\theta \in [\frac{\pi}{2}, \frac{3\pi}{8}, \frac{\pi}{4}, \frac{\pi}{8}, 0, -\frac{\pi}{8}, -\frac{\pi}{4}]$).

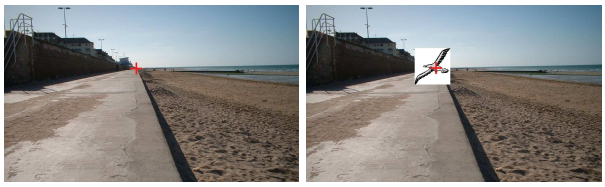


Figure 6: Best elected point detected without and with occlusion.

3. ALGORITHM IMPROVEMENTS

3.1 Ignoring small gradients and reducing the number of convolutions

As suggested in [18], the detection process has been speed up by two refinements:

1. Ignoring small gradients by thresholding the gradient magnitude before voting in the contribution images
2. Reducing the number of convolutions (Equation 2.1) by convolving directly partial sums over the different radii

3.2 Low resolution detection and sub-pixel precision

As explained before, the maximum size of the detected features depends mainly on the maximum radius r_{max}. High values for r_{max} are however not practicable since the computation time of the convolutions defined in Equ. 2.1 becomes prohibitive. One way to detect large structures without degrading the computation time is to reduce the resolution of the image before applying the transform. However, this method will degrade the precision of the points location in the original image. The distance between the wrong and the right coordinates increases significantly when using very small scales. To face this problem, we estimate the coordinates with a sub-pixel approach. If (x, y) is a local maximum in a contribution image S_θ then estimate the interest point coordinate as the maximum of interpolation on the grey levels values of neighboring pixels. We will consider two simple quadratic interpolations. The first interpolation is between the three pixels values we have in X-axis and the second one is between the three pixels values we have in Y-axis. The method used here is more developed in [4]. The question is to find the x position corresponding to the maximum of the parabola passing through the three points (x_1, a), (x_2, b) and (x_3, c) such as $b \geq a$ and $b \geq c$. It can be easily seen that $x = x_2 + m$ and m is given by

$$m = \frac{a - c}{2(a - 2b + c)}$$

3.3 Alignment rejection criterion

According to some carried out experiments, we notice, that aside junctions, we detect undesirable points along contours (mainly at $\theta_d = \pm\frac{\pi}{2}$). In fact, local maxima are formed along the contours' lines in the orientation projection image. This result is due to the fact that all pixels of lines contribute to each other resulting in high scores all along the line. Local maxima are then formed due to small variations or side effects. A solution consists in keeping only the points whose voters are located at various orientations. To

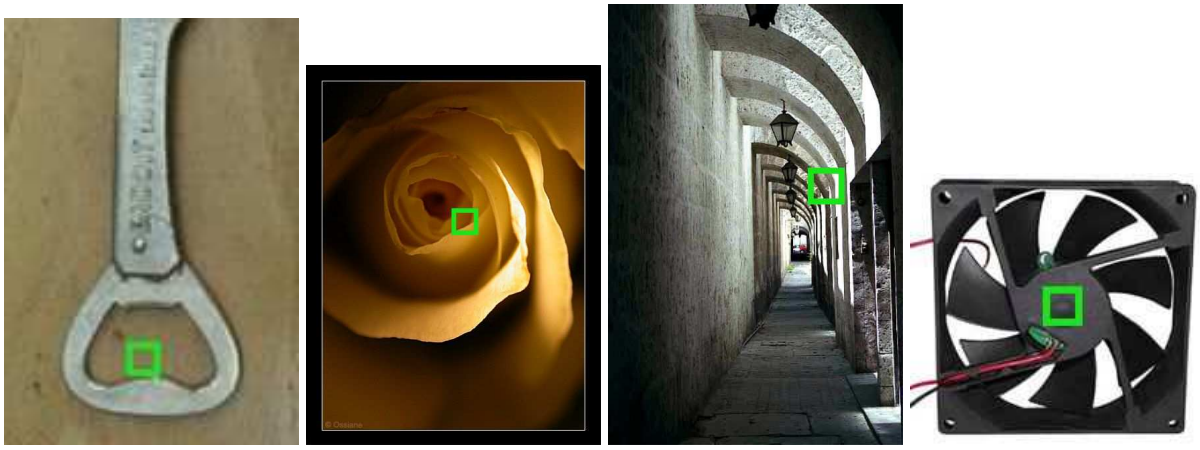


Figure 5: Best elected points detected at different θ_d - from left to right: $\theta_d = 9°$, $\theta_d = 27°$, $\theta_d = 45°$, $\theta_d = 67°$

this end, we used central moments. The moments of an image are very useful to describe the objects which it contains. Moreover, these moments give a single characterization of the signal. Given moments to all orders, we can correspond only one image to these moments and conversely. For a two-dimensional signal f, the moment of order $(p+q)$ is defined by the following formula

$$M_{pq} = \sum_x \sum_y x^p y^q f(x,y)$$

From image moments, we can define the central moments as follows

$$\mu_{pq} = \sum_x \sum_y (x-\overline{x})^p (y-\overline{y})^q f(x,y)$$

where

$$\overline{x} = \frac{M_{10}}{M_{00}} \qquad \text{and} \qquad \overline{y} = \frac{M_{01}}{M_{00}}$$

are the coordinates of the centroid. Information about orientation of an object can be derived by only using the second order central moments to construct a covariance matrix. The eigenvectors of this matrix constitute the predominate axes of the object, and orientation can thus be extracted from the angle of the eigenvector associated with the largest eigenvalue. The eigenvalues of the covariance matrix can easily be shown to be

$$\lambda_i = \frac{\mu_{20} + \mu_{02}}{2} \pm \frac{\sqrt{4{\mu_{11}}^2 + (\mu_{20} - \mu_{02})^2}}{2}$$

which are proportional to the squared length of the eigenvector axes.

In our case, we want to keep only the points whose voters are distributed according to varied directions. Consequently, we will eliminate all the points whose eigenvalues are very distant one from the other (i.e. there is a predominant direction for the object). To this aim, we chose a threshold noted s. If $|\frac{\lambda_1}{\lambda_2}| < s$ or $|\frac{\lambda_2}{\lambda_1}| < s$ then the point is rejected. The choice of s is strongly related to the visual significance that we attach to a point.

Normally, central moments are computed after segmentation process in order to retain scene's objects. In our case, we don't need to use any segmentation. In fact, the centroid $(\overline{x}, \overline{y})$ corresponds to the elected point. So, when applying the transform, we have to compute three other images in the same way we compute the orientation projection image. The only difference lies in the manner of voting. In the first image, we make votes using $(x-\overline{x})^2$ value. In the second image, votes are made using $(y-\overline{y})^2$ value and finally, in the third image, votes are made using $(x-\overline{x})(y-\overline{y})$ value. Thanks to these images, we can determine the eigenvalues of the local maxima points detected in S_θ image. After that, we apply the rejection criterion.

Figure 7: Elected points detected with and without alignment rejection criterion

4. PRELIMINARY EXPERIMENTS

We present in this section some preliminary experiments, mainly qualitative results and quantitative control of stability to image transformations. More exhaustive and

Figure 8: Detection examples

relevant experiments will be performed in next works and are discussed in the last section of this paper. The images used in these experiments come from **ImagEval** benchmark (**http //www.imageval.org/e_presentation.html**), a french initiative related to the evaluation of technologies of image filtering, content-based image retrieval and automatic description of images in large-scale image databases.

4.1 Detection examples

We present here some images showing the qualitative results of the detector. We have used various symbols to represent the points detected at different characteristic tangential angles. Table 1 defines these symbols. The points were detected at low resolution (30% of original size) with $r_{max} = 50$ and with 19 values of θ varying from $-90°$ to $80°$. Alignment rejection threshold s was set to $s = 0.1$. Only the 20 best points are displayed for each image. The average real number of $3D$ maxima, computed on 50 images is equal to $\overline{n_p} = 34, 2$.

Symbol	θ values
○	$[-5°, 5°]$
□	$[-45°, -5°[\cup]5°, 45°]$
×	$[-85°, -45°[\cup]45°, 85°]$
+	$[-90°, -85°] \cup [85°, 90°]$

Table 1: Points symbols table

Observing these results, we can remark that the detected points are not correlated and that the points detected at different characteristic tangential angles are complementary since they do not detect the same object parts. It is also often possible to recognize the expected topological nature of the detected features for each type of points. The red crosses mainly fall on large line junctions whereas the red crosses often correspond to curves junctions; green squares are mainly located at centers of curved-based objects and the blue circles correspond to circular object centers. Globally the detected points correspond to visually attractive and interpretable points of the images and we think that this is promising for the fixed objective, that is detecting points catching human visual attention.

4.2 Repeatability to transformations

The aim of this experiment is to test the stability of our detector towards different image transformations. The images and their transformed versions used in this experiment come from the first task of **ImagEval** benchmark whose goal is to index and retrieve transformed images. The 10 transformations studied in this experiment include both geometrical and chromatic transformations and five of them are illustrated on Figure 9. Note that the parameters of the transformations are randomly selected and two images are not necessary attacked by the same transformations. This experiment was carried out on 10 randomly selected original images and we computed the Constant Tangential Angle elected points (**CTA** points) on the $10 \times 10 = 100$ transformed images (and on the original images). The points were detected at low resolution (40% of original size) with $r_{max} = 50$ and with 19 values of θ varying from $-90°$ to $80°$. Alignment rejection threshold s was set to $s = 0.1$. The average number of interest points was $\overline{n_p} = 43, 7$.
We made a comparison between CTA points and Harris points [9] since the harris detector is known to be one of the most point detector to such image transformations [26].

Figure 9: Some of the studied transformations for the repeatability measurements - from left to right: Rotation, 3D Projection, Negative, Floyd-Steinbeck transform, Resize in a new image

To achieve better comparison, we computed the Harris detector in the same low resolution image and kept only the 50 best points. The scale parameter of the Harris filter was set to $\sigma_h = 1, 2$ which is a typically used value.
We computed the commonly used ϵ-repeatability metric between the original image and the transformed ones to evaluate the stability to transformations. ϵ was fixed to 6 pixels, which is quite large compared to other evaluations, but more appropriated to our low resolution detection that leads to quite large detected features for both detectors ($\epsilon = 6$ in the original image corresponds to $\epsilon_l = 0.4 \times \epsilon = 2.4$ pixels in the low resolution image which is a more standard value). Experiments' results are presented in table 2.

Transformation	Repeatability of **CTA** points (%)	Repeatability of Harris detector (%)
Negative	83.3	83.2
Black and White	81.2	83.4
Gaussian blur	82.0	88.3
Desaturation	76.9	84.6
JPEG compression	66.3	69.2
Floyd-Steinbeck transform	60.7	63.4
Rotation	52.2	59.3
Random Noise	56.9	63.6
3D Projection	22.4	19.8
Resize in a new image	29.3	21.3

Table 2: Repeatability to several transformations (ϵ=5).

We notice that the repeatability rate of the **CTA** points is very good although it is worse than Harris one in all chromatic transformations. It is however better in the two stronger geometric transformations, that are: insertion in a new image after resizing (between 40% and 70%) and $3D$ projection on an inclined plane. Thus, **CTA** points seem to be stable enough according to the fact the main expected property is not a better repeatability to transformations but a better adequacy to points catching visual attention.

4.3 Detections on a generic class of objects

One good way to evaluate the visual relevance of an interest points detector would be to estimate its ability to detect the same objects parts in a several images representing objects of a same class. Defining a correct metric for such a task is not trivial and would necessitate the manual construction of a ground truth or at least a manual analysis of detected points. We will certainly address this problem in future works and we relate here only a small qualitative experiment. We computed our **CTA** points in images representing objects of a same class labelled as *sun glasses*. The used images come from the fourth task of **ImagEval** benchmark whose goal is to recognized objects categories. The interest points were detected at low resolution (50% of original size) with $r_{max} = 50$ and with 19 values of θ varying from -90° to 80°. Alignment rejection threshold s was set to $s = 0.1$.
The results are presented on Figure 10. They show the ability of our interest points detector to focus on the main subject of the images and also to detect similar objects part in different instances of the object class, such as the bridge between the two glasses, the centroid(s) of the glasses themself or the interface region between the nose and the glasses.

5. CONCLUSION AND FUTURE WORKS

This paper presents a new interest points detector whose goal is to catch visual attention better than standard detectors. It is based on a generalization of Loy and Zelinsky's transform [18]. This transform was originally developed to detect radial symmetry centers, which are known to correlate well with psychophysical findings on fixation points of the human visual system. The initial voting algorithm is here extended to a set of voting orientations which allow the construction of a more reach analyze space. The proposed interest points are then detected as 3-dimensional maxima in this space and are expected to be more distinctive and stable. The specificity of this detector is that it enables the detection of points with different topological natures with the same detector. Each point is detected at a specific tangential angle which can be used as a topological characteristic associated to each interest point. Experiments showed that the detected points present a good stability to a lot of image transformations and qualitative preliminary experiments are very promising about their visual relevance.

Nevertheless, more reliable experiments on largest datasets have to be carried out to confirm these results. This will be the main topic of our future works. Visual relevance metrics have to be designed either by comparing the detected points to eye-tracking systems or by manually constructing ground truth of attractive object parts.

The definition of descriptors based on these interest points will be an other important task. Topological local descriptors could be for example derived from several geometrical moments computed on the voters locations. Invariance to geometric transformations could be then easily obtained. The definition of global image descriptor based on the distribution of the characteristic tangential angles is also a prospect. It could be useful to characterize the global topology of an image, and for example to distinguish images containing a lot of circular and smoothed forms from those con-

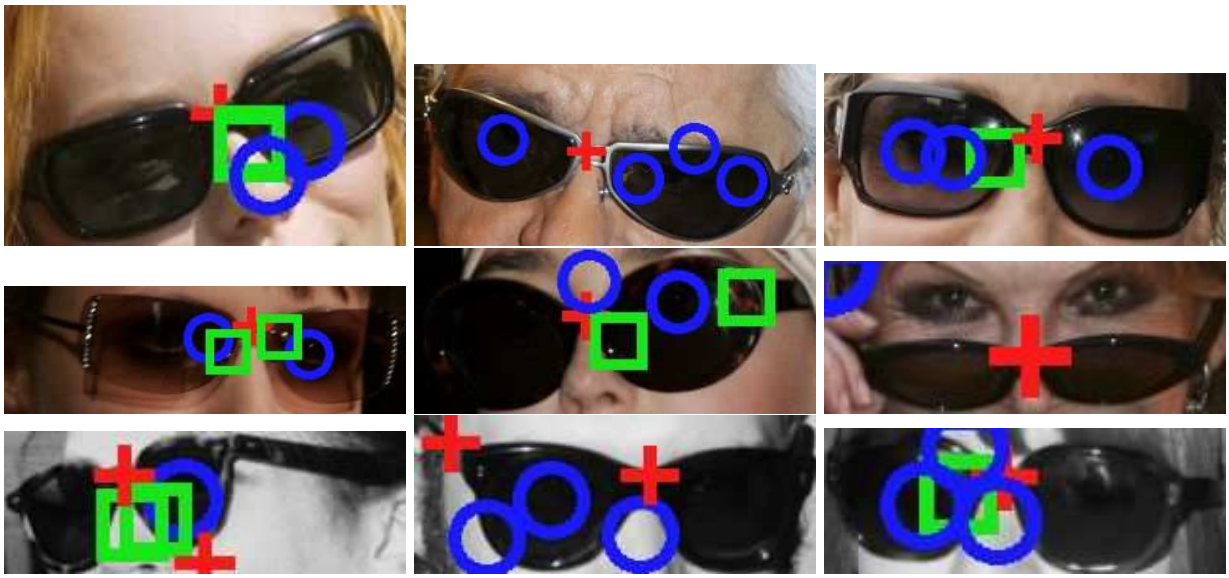


Figure 10: Detection on the images of the ImagEval class labelled as SUN GLASSES

taining a lot of abrupt junctions.

Once the interest points will be associated to efficient descriptors, the relevance of the resulting local features could be evaluated thanks to a multiple instances boosting procedure such as the one described in [21]. The main advantage of such a learning procedure is that the learner selects clearly the features being the more reliables for a given class and that heterogeneous features can be learned at the same time. It is thus possible to compare several local features by the frequency at which they are selected by the learner.

A last prospective work is to study the links between the scale-space theory and the 3-dimensional θ-space constructed by the proposed transform. If we show some common properties such as the diffusion equation, a lot of results could be automatically applied.

6. REFERENCES

[1] A. Baumberg. Reliable feature matching across widely separated views. In *CVPR*, pages 1774–1781, 2000.

[2] S. Bres and J. M. Jolion. Detection of interest points for image indexing. In *Proc. of Int. Conf. on Visual Information Systems*, pages 427–434, 1999.

[3] G. Carneiro and A. D. Jepson. Phase-based local features. In *Proc. of European Conf. on Computer Vision*, pages 282–296, 2002.

[4] F. Devernay. *A Non-Maxima Suppression Method for Edge Detection with Sub-Pixel Accuracy.* RR 2724, INRIA, November 1995.

[5] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2003.

[6] L. Florack and A. Kuijper. The topological structure of scale-space images. *Journal of Mathematical Imaging and Vision*, 12(1):65–79, February 2000.

[7] W. Förstner. A framework for low level feature extraction. In *Proc. of European Conf. on Computer Vision*, pages 383–394, 1994.

[8] V. Gouet and N. Boujemaa. Object-based queries using color points of interest. In *Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 30–36, 2001.

[9] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. of Alvey Vision Conf.*, pages 147–151, 1988.

[10] F. Heitger, L. Rosenthaler, R. Heydt, and O. Kubler. Simulation of neural contour mechanisms: From simple to end-stopped cells. *Vision Research*, 32(5):963–981, 1992.

[11] N. K. J. Fauqueur and R. Anderson. Multiscale keypoint detection using the dual-tree complex wavelet transform. In *accepted in Proc. of Int. Conf. on Image Processing*, 2006.

[12] C. Kimme, D. Ballard, and J. Sklansky. Finding circles by an array of accumulators. *Communications of the Association for Computing Machinery*, 18(2):120–122, Feb. 1975.

[13] S. Lazebnik, C. Schmid, and J. Ponce. Semi-local affine parts for object recognition. In *British Machine Vision Conference*, volume volume 2, pages 779–788, 2004.

[14] C.-C. Lin and W.-C. Lin. Extracting facial features by an inhibitory mechanism based on gradient distributions. *Pattern Recognition*, 29(12):2079–2101, 1996.

[15] T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Applied Statistics*, 21(2):224–270, 1994.

[16] P. J. Locher and C. F. Nodine. Symmetry Catches the Eye. In A. Levy-Schoen and J. K. O'Reagan, editors, *Eye Movements: From Physiology to Cognition*, pages 353–361. Elsevier Science Publishers B. V. (North Holland), 1987.

[17] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. of Int. Conf. on Computer Vision*, pages 1150–1157, 1999.

[18] G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8), August 2003.

[19] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[20] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005.

[21] A. Opelt, M. Fussenegger, and P. Auer. Generic object recognition with boosting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):416–431, 2006.

[22] A. C. Parker. A representation for shape based on peaks and ridges in the difference of low pass transform. *IEEE Transactions on PAMI*, 6(2), mar 1984.

[23] C. Privitera and L. Stark. Algorithms for defining visual regions-of-interest : Comparison with eye fixation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):970–982, 2000.

[24] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: the generalized symmetry transform. *Computer Vision*, 14(2):119–130, 1995.

[25] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[26] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *International Conference on Computer Vision, Real-Time Imaging*, pages 230–235, January 1998.

[27] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *Computer Vision*, 37(2):151–172, 2000.

[28] G. Sela and M. D. Levine. Real-time attention for robotic vision. *Real-Time Imaging*, 3(3):173–194, 1997.