METHOD OF EXTRACTING INTEREST POINTS BASED ON MULTI-SCALE DETECTOR AND LOCAL E-HOG DESCRIPTOR

Manuel Grand-brochier, Christophe Tilmant and Michel Dhome
Laboratoire des Sciences et Matriaux pour l'Electronique, et d'Automatique (LASMEA)
UMR 6602 UBP-CNRS, 24 avenue des Landais, 63177 Beaumont, France

Keywords: Multi-scales analysis, Local descriptor, Robustness to image transformations, Elliptical-HOG.

Abstract: This article proposes an approach to extraction (detection and description) of interest points based Fast-Hessian and E-HOG. SIFT and SURF are the two most used methods for this problem and their studies allow us to understand their construction and extract the various advantages (invariances, speeds, repeatability). Our goal is, firstly, to couple these advantages to create a new system (detector, descriptor and matching) and, secondly, to determine the characteristic points for different applications (image transformation, 3D reconstruction...). Our system must also be as invariant as possible for the image transformation (rotations, scales, viewpoints for example). Finally, we have to find a compromise between a good matching rate and the number of points matched. All the detector and descriptor parameters (orientations, thresholds, analysis shape) will be also detailed in this article.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

There are a large number of applications based on image analysis, especially 3D reconstruction problems, tracking or pattern recognition for example. These applications need data usually extracted with two tools: the detection of interest points (Li and Allison, 2008) and the local description (Li and Allison, 2008) of these. The detector analyses the image to extract the characteristic points (corners, edges, blobs). The neighborhood study allows us to create a local points descriptor, in order to match them. For matched interest points, the robustness of various transformations of the image is very important. To be robust to scale, interest points are extracted with a global multi-scales analysis, we considered the Harris-Laplace detector (Harris and Stephens, 1988; Mikolajczyk and Schmid, 2004b; Mikolajczyk and Schmid, 2002), the Fast-Hessian (Bay and al., 2006) and the difference of Gaussians (Lowe, 1999; Lowe, 2004). The description is based on a local exploration of interest points to represent the characteristics of the neighborhood. In comparative studies (Choksuriwong and al., 2005; Mikolajczyk and Schmid, 2004a; Bauer and al., 2007), it is shown that oriented gradients histograms (HOG) give good results. Among the many methods using HOG, we retain SIFT (Scale Invariant Feature Transform) (Lowe, 1999; Lowe, 2004) and SURF (Speed Up Robust Features) (Bay and al., 2006), using a rectangular neighborhood exploration (R-HOG: Rectangular-HOG). We also mention GLOH (Gradient Location and Orientation Histogram) (Mikolajczyk and Schmid, 2004a; Dalal and Triggs, 2005) and Daisy (Tola et al., 2008), using circular geometry (C-HOG: Circular-HOG). To provide the best possible list of points for different applications, we propose to create a system of detection and local description which is the most robust possible against the various transformations that can exist between two images (illumination, rotation, viewpoint for example). It should also be as efficient as possible as regards the matching rate. Our method relies on a Fast-Hessian points detector, an elliptical exploration and a local descriptor based E-HOG (Elliptical-HOG). We propose to estimate local orientation, with the study of the Harris matrix, in order to adjust the descriptor (rotation invariance) and finally we will normalize (brightness invariance).

Section 2 presents briefly SIFT and SURF, and lists the advantages of each. The various tools and their parameters (orientations, thresholds, analysis pattern) we use are detailed in Section 3. To compare our approach to SIFT and SURF, many tests have been carried out. A synthesis of the different results is presented in Section 4.

Grand-brochier M., Tilmant C. and Dhome M..
METHOD OF EXTRACTING INTEREST POINTS BASED ON MULTI-SCALE DETECTOR AND LOCAL E-HOG DESCRIPTOR.
DOI: 10.5220/0003369300590066
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 59-66
ISBN: 978-989-8425-47-8
Copyright © 2011 SCITEPRESS (Science and Technology Publications, Lda.)

2 RELATED WORK

In order to propose a robust method and giving many interest points, Lowe propose a new approach, SIFT (Lowe, 1999; Lowe, 2004), consisting of a difference of Gaussians (DoG) and R-HOG analysis. The detector is based on an approximation of the Laplacian of Gaussian (Lindeberg, 1998) and interest points are obtained by maximizing the DoG:

$$\begin{aligned} D(x,y,\sigma) =&(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y) \\ & L(x,y,k\sigma)-L(x,y,\sigma). \end{aligned} \quad (1)$$

The descriptor uses an orientation histogram, based on equation 2, to determine the angle of rotation θ to be applied to the mask analysis.

$$\theta(x,y)=tan^{-1}(\frac{(L(x,y+1)-L(x,y-1)}{(L(x+1,y)-L(x-1,y)}) \quad (2)$$

It then uses R-HOG, formed by local gradients in the neighborhood, previously smoothed by a Gaussian. Finally, the descriptor is normalized to be invariant to illumination changes.

An extension of SIFT, GLOH (Mikolajczyk and Schmid, 2004a; Dalal and Triggs, 2005), has been proposed to increase the robustness. It amounts to the insertion of a grid in log polar localization. The mask analysis of this descriptor is composed of three rings (C-HOG), whose two largest are divided along eight directions. More recently, the descriptor Daisy (Tola et al., 2008) has been proposed. It is also based on a circular neighborhood exploration and constructs convolved orientation maps.

SIFT has not a fast computational speed. SURF (Bay and al., 2006) proposes a new approach, whose main objective is to accelerate the various image processing steps. The first problem was to choose the detector method. The various tests (Juan and Gwun, 2009; Bay and al., 2006) show that the Fast-Hessian has the best repetability rate. It is based on the Hessian matrix:

$$H(x,y,\sigma)=\begin{bmatrix} L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma) \\ L_{xy}(x,y,\sigma) & L_{yy}(x,y,\sigma) \end{bmatrix}, \quad (3)$$

with $L_{ij}(x,y,\sigma)$ the second derivative in the directions i and j of L. The maximization of its determinant (Hessian) allows us to extract the coordinates of interest points in a given scale. The second step, the local description, is based on Haar wavelets. These estimate the local orientation of the gradient, allowing the construction of the descriptor. Finally, SURF studied the sign of the wavelet transform to increase the quality of results.

The presented methods use similar tools: multiscale analysis (Fast-Hessian or DoG), local description based HOG, local smoothing and descriptor normalization. For matching they use a minimization of either the Euclidean distance between descriptors (SURF) or the angle between vectors descriptors (SIFT). Many tests (Mikolajczyk and Schmid, 2004a; Juan and Gwun, 2009; **?**) can establish a list of different qualities of each. It follows that SURF, with its detector, has the best repeatability for viewpoint changes, scale, noise and lighting. It is also faster than SIFT, however it has a higher precision rate for rotations and scale changes. It has also a higher number of detected points for all transformations. It might be interesting to combine these two methods.

3 METHOD

The method we propose is divided into three parts: a Fast-Hessian multi-scale detector, a local E-HOG descriptor and an optimized matching. This section describes the different steps of our method and parameters used. The detector Fast-Hessian provides a list of interest points, characterized by their coordinates and local scale. Our descriptor is based on the Harris matrix interpretation, and the construction of E-HOG. Matching is based on an approximation of the nearest neighbors and removing duplicates. These issues will be detailed below.

3.1 Detection

The Fast-Hessian is an approximated method of Hessian matrix (equation 3), to reduce the computing time. This detector uses integral images (Figure 1), therefore it takes only three additions and four memory accesses to calculate the sum of intensities inside a rectangular region of any size. The Fast-Hessian relies on the exploitation of the Hessian matrix (equation 3), whose determinant is calculated as follows:

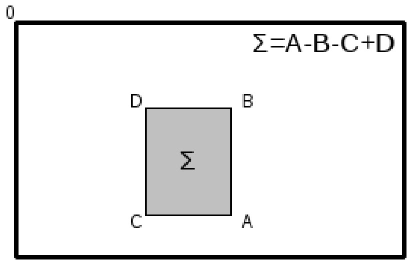


Figure 1: Determination of integral image.

$$\begin{aligned} & det(H(x,y,\sigma)) = \\ & \quad \sigma^2(L_{xx}(x,y,\sigma)L_{yy}(x,y,\sigma)-L_{xy}^2(x,y,\sigma)), \end{aligned} \quad (4)$$

where $L_{xx}(x,y,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2}g(\sigma)$ with the im-

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

age I in point (x,y), and similarly for $L_{xy}(x,y,\sigma)$ and $L_{yy}(x,y,\sigma)$. Gaussians are optimal for scale-space analysis and the Fast-Hessian provides an approximation of these second order derivative (Figure 2).

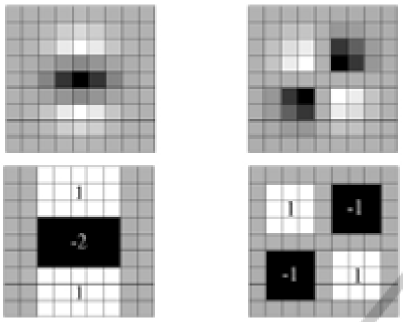


Figure 2: Top row: Gaussian second order partial derivatives, bottom row: approximation for the second order Gaussian partial derivatives.

By looking for local maxima of the determinant, we establish a list of K points associated with a scale, denoted $\{(x_k,y_k,\sigma_k);\ k \in [\![0;K-1]\!]\}$, where:

$$(x_k,y_k,\sigma_k) = \underset{\{x,y,\sigma\}}{argmax}(det(H(x,y,\sigma))). \quad (5)$$

The number of interest points obtained depends on the space scale explored and thresholding of local maxima. We have to find a compromise between scale space exploration and relevance of extracted points. The influence of this one will be detailed in Section 4.

3.2 Description

As with SIFT and SURF, our method is based on HOG, yet our analysis window will consist of ellipses (E-HOG). Different tools will also be necessary to adjust and normalize our descriptor.

3.2.1 Determining the Local Orientation Gradient

To be as invariant as possible for rotations, estimating the local orientation gradient of the interest point is necessary. This parameter allows us to adjust the E-HOG, giving an identical orientation for two corresponding points. For this, we use the Harris matrix, calculated for each point (x_k,y_k) and defined by:

$$M_H(x_k,y_k) = \begin{bmatrix} \sum\limits_{V(x_k,y_k)} [I_x(x_k,y_k)]^2 & \sum\limits_{V(x_k,y_k)} I_x(x_k,y_k)I_y(x_k,y_k) \\ \sum\limits_{V(x_k,y_k)} I_x(x_k,y_k)I_y(x_k,y_k) & \sum\limits_{V(x_k,y_k)} [I_y(x_k,y_k)]^2 \end{bmatrix} \quad (6)$$

where $V(x_k,y_k)$ represents the neighborhood of the interest point, I_x and I_y are the first derivatives in x and y of image, calculated using the Canny-Deriche operator. The properties of this matrix can study the information dispersion. The local analysis of its eigenvectors ($\overrightarrow{v_1}$ and $\overrightarrow{v_2}$) associated with corresponding eigenvalues can extract an orientation estimate:

$$\Delta = \text{Trace}(M_H(x_k,y_k))^2 - 4\text{Det}(M_H(x_k,y_k))$$

$$\lambda_1 = \frac{\text{Trace}(M_H(x_k,y_k)) + \sqrt{\Delta}}{2}$$

$$\overrightarrow{v_1} = \begin{pmatrix} \dfrac{(\sum\limits_{V(x_k,y_k)} [I_y(x_k,y_k)]^2) - \lambda_1}{\sum\limits_{V(x_k,y_k)} I_x(x_k,y_k)I_y(x_k,y_k)} \\ 1 \end{pmatrix}$$

$$\theta_k = \arctan(\overrightarrow{v_1}). \quad (7)$$

3.2.2 Descriptor Construction

The initial shape of our descriptor relies on a circular neighborhood exploration of the interest point. The seventeen circles used, are divided into three scales (Figure 3) and are adjusted by θ_k (equation 7).

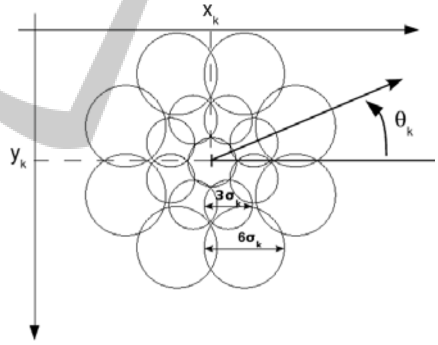


Figure 3: Initial mask analysis of our descriptor, centered at (x_k,y_k) and oriented by an angle θ_k.

This angle allows us to be robust to image rotation. The circle diameter is proportional to σ_k, thus accentuating the scale invariance. To manage the problem of viewpoint changes and anisotropic transformations, we propose to modify the shape of our descriptor. The goal is to get local information more consistently. We propose to use an elliptical exploration to describe the neighbor of interest points (Figure 4).

An analysis of the properties of affine detectors (Mikolajczyk and Schmid, 2004b; Mikolajczyk and al., 2005) allows us to determine the ratio r_k between the axes of ellipses. For example for the Harris affine detector, two scales are used and are bound by the following equation: $\sigma_D^k = s\sigma_I^k$ where σ_D is the differentiation scale, σ_I is the integration scale and s is the ratio. It is noted that s is generally between 0.5 and

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

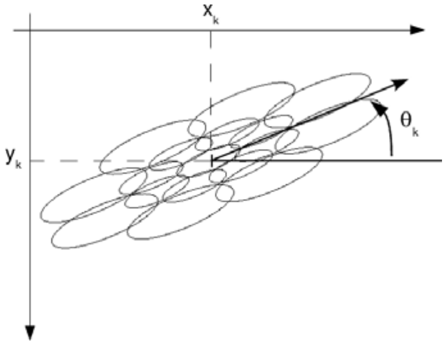


Figure 4: Final mask analysis of our descriptor, centered at (x_k, y_k) and oriented by an angle θ_k.

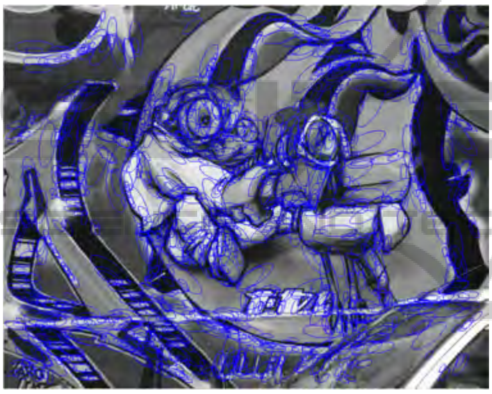


Figure 5: Example of local description of interest points.

0.75. By the parallelism between these two scales and our ellipses, it is possible to determine the best ratio r_k. Based on many tests, the ratio giving the best result is equal to 0.5. Figure 5 illustrates the construction of our ellipses (for better visualization, we only show the central ellipse of each descriptor).

We construct a HOG eight classes (in steps of 45) for each ellipse. Our descriptor, we note $des_I(x_k, y_k)$, belongs to $\mathbb{R}^{136}$ (17 ellipses × 8 directions). To be invariant for brightness changes, histogram is normalized and we use also a threshold for E-HOG to remove the high values of gradient.

3.3 Matching

The objective is to find the best similarity (corresponding to the minimum distance) between descriptors of two images. Euclidean distance, denoted d_e, between two descriptors is defined by:

$$d_e(des_{I_1}(x_k,y_k), des_{I_2}(x_l,y_l)) = \sqrt{[des_{I_1}(x_k,y_k)]^T \cdot des_{I_2}(x_l,y_l)}. \qquad (8)$$

The minimization of d_e, denoted d_{min}, provides a pair of points $\{(x_k, y_k); (x_{\tilde{l}}, y_{\tilde{l}})\}$:

$$\tilde{l} = \underset{l\in[\![0;L-1]\!]}{argmin}\,(d_e(des_{I_1}(x_k,y_k), des_{I_2}(x_l,y_l))), \qquad (9)$$

$$d_{min} = d_e(des_{I_1}(x_k,y_k), des_{I_2}(x_{\tilde{l}},y_{\tilde{l}})). \qquad (10)$$

To simplify the search for this minimum distance, we propose to use an approximative nearest neighbor search method (a variant of k-d tree) (Arya and al., 1998). The principle is to create a decision tree based on descriptors components of the second image (Figure 6). So, for each new descriptor of the first image, all components are tested and the nearest neighbor is defined. Research is therefore faster, without sacrificing precision. To have a more robust matching, thresholding is applied to this distance, to find a ``high˝minimum. The pair of points is valid if:

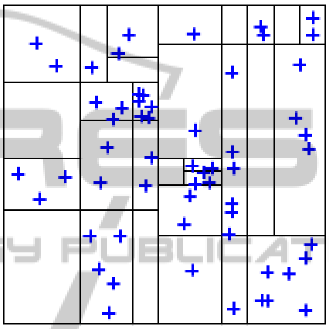


Figure 6: Example of decision tree to extract the nearest neighbors.

$$d_{min} \leq \alpha \times min(d_e(des_1(x_k,y_k), des_2(x_l,y_l))), \qquad (11)$$

for $l \in [\![0; N-1]\!] \backslash \tilde{l}$ and with α the threshold selection (detailed in Section 4). We do not allow a point to match with several other points, and a final step is to remove duplicates.

4 RESULTS

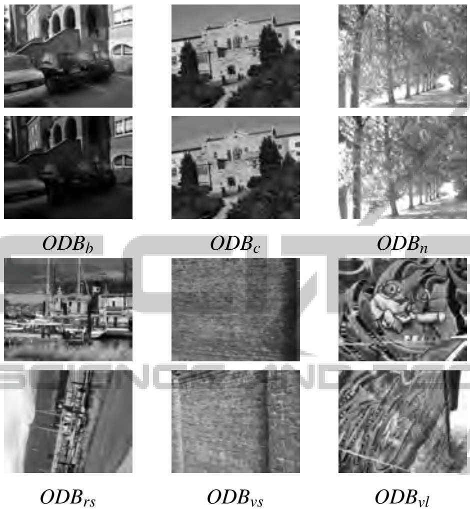
We are going to compare our method with SIFT and SURF. These two methods are the most used and give the best results. We propose to study the matching rate and the precision of each of them. We will also study the $recall = f(1 - precision)$ curves and the estimation error of the transformed image.

4.1 Databases

To validate our method, we chose two databases:

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

- The first one, noted *ODB* and extracted from the Oxford [1] database, proposes scene transformations with an access to the matrix of homography. Transformations studied are brightness changes (ODB_b), modified jpeg compressions (ODB_c), blur (ODB_n), rotations and scales (ODB_{rs}), and small and large angle viewpoint changes (respectively ODB_{vs} and ODB_{vl}). Figure 7 illustrates this database.


ODB_b ODB_c ODB_n

ODB_{rs} ODB_{vs} ODB_{vl}

Figure 7: Examples of images used for transformations: (top: left to right) brightness changes, modified jpeg compressions, blur, (bottom: left to right) rotations + scales, viewpoint changes (small angle), and viewpoint changes (large angle).

- A second database, noted *SDB*, composed of a set of synthetic image transformations (Figure 8 and Figure 9). These transformations are rotations 45 (SDB_r), scales (SDB_s), anisotropic scales (SDB_{as}), rotations 45 + scales (BS_{rs}) and rotations 45 + anisotropic scales (BS_{ras}).

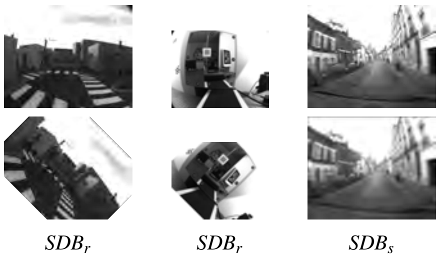

SDB_r SDB_r SDB_s

Figure 8: Examples of laboratory images (board cameras) used for synthetic transformations.

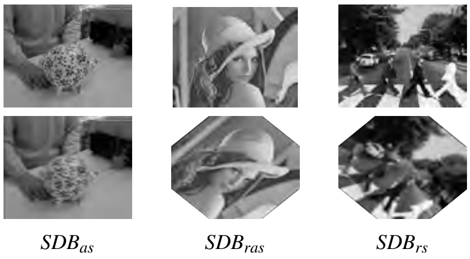

SDB_{as} SDB_{ras} SDB_{rs}

Figure 9: Examples of internet images (Pig, Lena, Beatles) used for synthetic transformations.

4.2 The Parameters of our Method

4.2.1 Influence of the Number of Octave (Detection Parameter)

To get the best compromise between relevance and number of extracted points, we use the following curves:

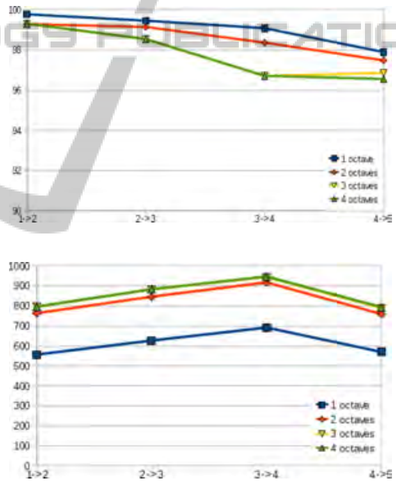


Figure 10: These graphs represent, for ODB_{vl} images: (top) the correct matching rate according to the number of octave and (bottom) the number of points that can be matched.

It represents, on one hand the matching precision for ODB_{vl} images, on the other hand the number of points that can be matched. The best compromise is two octaves, it has better precision than three or four octaves while keeping an almost identical number of points. One octave has a better precision than two octaves, but loses a lot of points. Therefore we choose two octaves instead of four used by SURF.

4.2.2 Threshold Selection

The threshold selection α used in the equation 11 is determinated by analysing the curves of Figure 11. This threshold allows us to increase the selectivity,

[1] http://www.robots.ox.ac.uk/∼vgg/data/data-aff.html

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

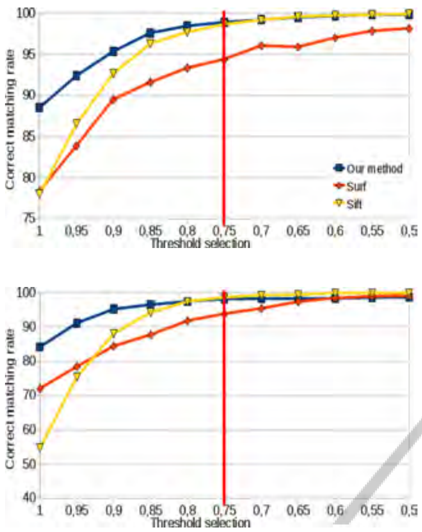


Figure 11: These graphs represent the correct matching rate according to the threshold selection. Top is for viewpoint changes (grafiti 1 → 2) and bottom is for rotation + scale (boat 1 → 3).

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

but the consequence is a reduction in the number of matched points.

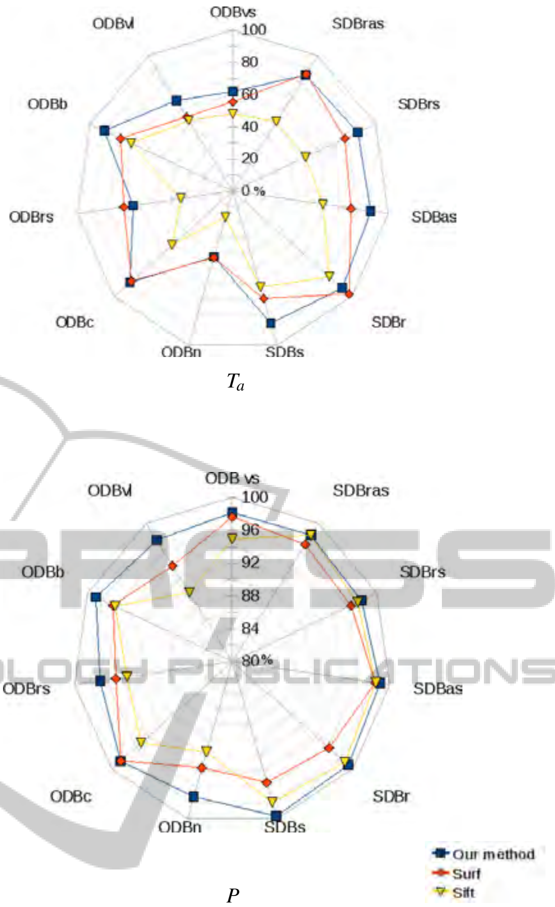
If the threshold goes away from 1, the matching becomes more selective and therefore fewer points are used. The problem is to find the best compromise between the correct matching rate and the number of matched points. SIFT recommands a threshold of 0.8 and SURF a threshold of 0.7. By analysing curves of Figure 11, we choose a threshold $\alpha = 0.75$.

4.3 Evaluation Tests and Results

4.3.1 Matching Rate and Precision

We propose to compare the matching rate T_a, as well as the precision P of every method. T_a is defined by the number of correct matches divided by the number of possible matches. P is defined by the number of correct matches divided by the number of matches performed. A synthesis of the results obtained is proposed in Figure 12.

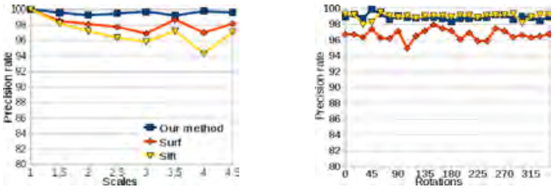
Our method presents results better or as good as SIFT and SURF. Our matching rate remains better than the two other methods with the exception of the databases ODB_{rs} and SDB_r transformations. Nevertheless the difference between SURF and our method for this type of transformation is lower than 4%. For other type of transformation, the biggest differences are observed for rotation 45 + scale (≈ 10% between our method and SURF and 37% with SIFT) and for large angle viewpoint changes (≈ 18% with SURF and SIFT). Our matching precision is also better and remains constantly above 95%. The biggest difference is obtained for large angle viewpoint changes (4% for SURF and 8% for SIFT). To detail the precision curves of different methods, we propose graphs in Figures 13 to 16.



T_a

P

Figure 12: These graphs represent at top: a matching rate and at bottom: a matching precision. Matching rate is the ability to match points and matching precision is the match quality. The goal is to have the highest rate of correctly matched points (with better precision).

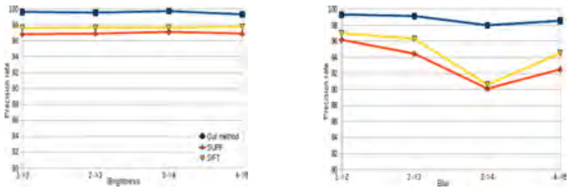


(a)

(b)

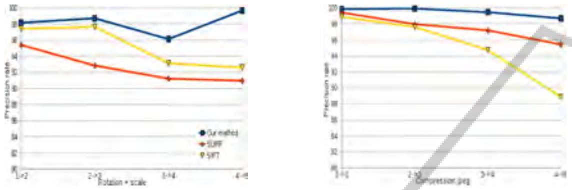
Figure 13: (a) precision rate for scales changes (SDB_s) and (b) precision rate for rotations (SDB_r).

These curve show a better precision rate for our method, or similar for rotation transformations (figure



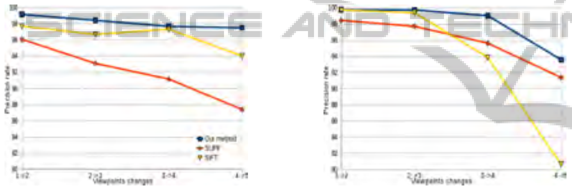
(a) (b)

Figure 14: (a) precision rate for brightness (ODB_b) and (b) precision rate for blur (ODB_n).



(a) (b)

Figure 15: (a) precision rate for rotation and scale (ODB_{rs}) and (b) precision rate for compression jpeg (ODB_c).



(a) (b)

Figure 16: (a) and (b) precision rate for viewpoints changes respectively (ODB_{vl}) and (ODB_{vs}).

13.b). Another observation can be made through the transformations studied, concerning the stability of our method. Indeed, our curves decrease more slowly than SIFT and SURF, implying a precision rates more constant.

4.3.2 Influence of Data

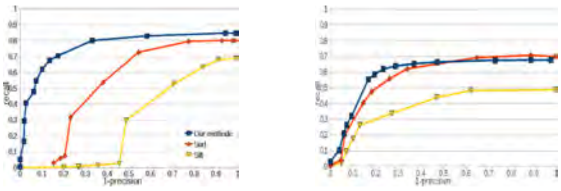
To observe the influence of data on different methods, we use the $recall = f(1-precision)$ curves (Mikolajczyk and Schmid, 2004a) with two components come from:

$$recall = \frac{\text{number of correct matches}}{\text{number of possible correct matches}},$$

and

$$1 - precision = \frac{\text{number of false matches}}{\text{number of (correct matches + false matches}}.$$

We propose to analyse two curves (Figure 17) and we can observe that our method is more stable than SIFT and SURF. Therefore, our method is more robust to the deterioration of data for these transformations.



(a) (b)

Figure 17: (a) a recall versus 1-precision for brightness (ODB_b: 1→ 4) and (b) a recall versus 1-precision for rotation + scale (ODB_{rs}: 2→ 4).

4.3.3 Estimation of the Image Transformation

We propose a final study on the estimation of the homography matrix. For 3D reconstruction, for example, the estimate of this matrix is very important, and the goal is to have the best estimation (the error as low as possible) with maximum points validating this matrix. The estimation is based on the matches and the Ransac algorithm. We studied three transformations and the results are:

- viewpoints changes (ODB_{vl}):

	error (%)	valid points (%)
Our method	0.06	96.72
SURF	3.46	96.34
SIFT	2.98	92.35

- rotation and scale (ODB_{rs}):

	error (%)	valid points (%)
Our method	0.54	93.23
SURF	1.69	86.03
SIFT	2.06	88.24

- blur (ODB_n):

	error (%)	valid points (%)
Our method	0.61	97.13
SURF	0.66	92.1
SIFT	0.58	96.56

Our method obtains, for the first two transformations, an error rate below SIFT and SURF and a higher rate of valid points. For the last transformation, the results are similar for all three methods.

5 CONCLUSIONS

In this article we presented a method based on, on the one hand the avantages of SIFT and SURF (repetability, invariances) and on the other hand the use of tools such as the Harris matrix, the HOG or the decision tree. Detection relies on Fast-hessian detector that we

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

thresholded. Description interprets the Harris matrix and uses an elliptical shape to adapt the descriptor to the image transformation. Finally, matching creates decision tree and removes duplicates. To validate our method, we compared it to SIFT and SURF. Our method has better matching rate and better precision for most transformations. It is also robust to data degradation problems and provides an estimation of homography matrix more reliable, keeping good rate of valid points. Therefore data extracted from images are better and will result in an improvement of the applications referred (3D reconstruction or pattern recognition for example).

Our prospects are a generalization of our method, with application to a spatio-temporal analysis. We will add a temporal variable to the Hessian matrix (equation 3). We also transform our descriptor shape to obtain an neighborhood exploration ellipsoidal (for tracking for example). An other prospect is to integrate a third dimension to use it in medical imaging.

SCIENCE AND TECHNOLOGY PUBLICATIONS

REFERENCES

Arya, S., Mount, D., Netanyahu, N., Silverman, R., and Wu, A. (1998). An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *J.ACM*, 45:891–923.

Bauer, J., Snderhauf, N., and Protzel, P. (2007). Comparing several implementations of two recently published feature detectors. *Intelligent Autonomous Vehicles*.

Bay, H., Tuylelaars, T., and Gool, L. V. (2006). Surf : Speeded up robust features. *European Conference on Computer Vision*.

Choksuriwong, A., Laurent, H., and Emile, B. (2005). Etude comparative de descripteur invariants d'objets. *ORASIS*.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE Conference on Computer Vision and Pattern Recognition*.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. *Alvey Vision Conference*, pages 147–151.

Juan and Gwun (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing*, 3(4):143–152.

Li, J. and Allison, N.M. (2008). A Comprehensive Review of Current Local Features for Computer Vision. *Neurocomputing*, 71(10-12):1771–1787.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116.

Lowe, D. (1999). Object recognition from local scale-invariant features. *IEEE International Conference on Computer Vision*, pages 1150–1157.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Mikolajczyk, K., and Schmid, C. (2002). An affine invariant interest point detector. *European Conference on Computer Vision*, 1:128–142.

Mikolajczyk, K. and Schmid, C. (2004a). A performance evaluation of local descriptors. *IEEE Pattern Analysis and Machine Intelligence*, 27(10):1615–1630.

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

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

Tola, E., Lepetit, V., and Fua, P. (2008). A fast local descriptor for dense matching. *IEEE Conference on Computer Vision and Pattern Recognition*.