HAL
open science

SIFT Detectors for Matching Aerial Images in Reduced Space

H. Kamel, Youssef Chahir, Mohamed-Khireddine Kholladi

To cite this version:

H. Kamel, Youssef Chahir, Mohamed-Khireddine Kholladi. SIFT Detectors for Matching Aerial Images in Reduced Space. IEEE International Conference on Computer Integrated Manufacturing - CIP'2007, 2007, Sétif, Algeria. 10 p. hal-00825683

HAL Id: hal-00825683

https://hal.science/hal-00825683v1

Submitted on 24 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SIFT Detectors for Matching Aerial Images in Reduced Space.

Mr. Houari Kamel
University of Oum El Bouaghi, Algeria
hk_houari@yahoo.fr

Dr. Youssef Chahir
University of Caen, France
Group of Research in Data processing, Picture, Instrumentation of Caen, GREYC - URA CNRS 6072
II campus - BP 5186, University of Caen
Tel.: 33/0 2 31 56 73 75 and Fax: 33/0 2 31 56 73 30
chahir@info.unicaen.fr

Dr. Mohamed-Khireddine KHOLLADI
Department chief of Computer Sciences, Sciences faculty of the engineer, University Mentouri of Constantine
Laboratory of the Scientific and Technical information and its role in the national development
Chief of team of research of the New Technologies of information and their applications in national archives institutions
Laboratory LIRE, Member of the team of research "Vision and Computer Graphics"
Tel. & Fax: 213 31 818817 and GSM: 213 70 314924
Email: kholladi@umc.edu.dz, kholladi@yahoo.fr, kholladi@assila.net
Site Web: www.asd-kholladi.org

Abstract. *In this paper we propose a novel approach for matching cartographic images over detecting interest points invariant to scale and affine transformations. Our scale and affine invariant detectors are based on the following recent results: Interest points extracted with the SIFT detector which is adapted to affine transformations and give repeatable results (geometrically stable). This provides a set of distinctive points which are invariant to scale, rotation and translation as well as robust to illumination changes and limited changes of viewpoint. The characteristic scale determines a scale invariant region for each point. The characteristic scale and the affine shape of neighbourhood determine an affine invariant region for each point. We apply an unsupervised classification to reduce the space of sets of interest points by using weighted bipartite graph matching in solving the point correspondence. Diffusion map: projection of the bipartite graph in a reduce space on which we apply K-means to classify the representatives clusters. The performance of our approach detector is also confirmed by excellent matching results.*
Keywords*: Interest points, Local features, Scale invariance, Affine invariance, Matching and Recognition*

1. Introduction

Local features have been shown to be well suited to matching and recognition as well as to many other applications as they are robust to occlusion, background clutter and other content changes. The difficulty is to obtain invariance to viewing conditions witch is the main characteristics of aerial images in cartographic data bases. Different solutions to this problem have been developed over the past few years and are reviewed in Section 2. These approaches first detect features and then compute a set of descriptors for these features. In the case of significant transformations, feature detection has to be adapted to the transformation, as at least a subset of the features must be present in both images in order to allow for correspondences. Features which have proved to be particularly appropriate are interest points. However, the Harris interest point detector is not invariant to scale and affine transformations [29]. There are different descriptors like Gradient Location and Orientation Histogram (GLOH), shape context [12], PCA-SIFT [10], spin images [11], steer able filters [4], differential invariants [9], complex filters [26], moment invariants [28] and cross-correlation of sampled pixel values. Gradient Location and Orientation Histogram (GLOH) is a new descriptor which extends SIFTS by changing the location grid and using PCA to reduce the size.

In this paper we use the SIFT point detector as presented in Section 3. To obtain scale invariant features transform points, for which the localization and scale perfectly reflect the real scale change between two images. The points are then detected at each scale within this range. As a consequence, there are many points, which represent the same structure, but the location and the scale of the points is slightly different. The unnecessarily high number of points increases the probability of mismatches and the complexity of the matching algorithms. In this case, efficient methods for rejecting the false matches and for verifying the results are necessary. For each point detected it is assigned a descriptor for one position,

one scale and one orientation, the descriptor is constituted with a histogram of spatial intensities of gradient orientation weighted by spatial gradient norm. Our approach solves this problem by selecting the points in the multi-scale representation by applying an unsupervised classification to reduce the space of sets of interest points by use of weighted bipartite graph matching in solving the point correspondence. Diffusion map: projection of the bipartite graph in a map reduced space on which we apply K-means to classify the representatives clusters who represent the reduced points matched of input images. To measure the accuracy of our detectors we introduce a repeatability criterion which we use to evaluate and compare our detectors to existing approaches. Section 6 presents the evaluation criteria and the results of the comparison, which shows that our detector performs better then existing ones. Finally, in Section 7 we present experimental results for matching.

2. Related Work

Many approaches have been proposed for extracting scale and affine invariant features. These are reviewed in the following.

2.1. Scale Invariant Detectors

There are a few approaches which are truly invariant to significant scale changes. Typically, such techniques assume that the scale change is the same in every direction, although they exhibit some robustness to weak affine deformations. Existing methods search for local extrema in the 3D scale-space representation of an image (*x, y* and *scale*). This idea was introduced in the early eighties by Crowley [7]. In this approach the pyramid representation is computed using difference-of-Gaussian filters. A feature point is detected if a local 3D extremum is present and if its absolute value is higher than a threshold. The existing approaches differ mainly in the differential expression used to build the scale-space representation.

Lindeberg [21] searches for 3D maxima of scale normalized differential operators. He proposes to use the Laplacian of Gaussian (LoG) and several other derivative based operators. The scale-space representation is built by successive smoothing of the high resolution image with Gaussian based kernels of different size. The LoG operator is circularly symmetric and it detects blob-like structures. The scale invariance of interest point detectors with automatic scale selection has also been explored by Bretzner [3] in the context of tracking.

Lowe [22] proposed an efficient algorithm for object recognition based on local 3D extrema in Scale & Affine Invariant Interest Point Detectors 65 the scale-space pyramid built with difference-of-Gaussian (DoG) filters. The input image is successively smoothed with a Gaussian kernel and sampled. The difference-of-Gaussian representation is obtained by subtracting two successive smoothed images. Thus, all the DoG levels are constructed by combined smoothing and sub-sampling. The local 3D extrema in the pyramid representation determine the localization and the scale of the interest points. The DoG operator is a close approximation of the LoG function but the DoG can significantly accelerate the computation process. A few images per second can be processed with this algorithm. The common drawback of the DoG and the LoG representation is that local maxima can also be detected in the neighbourhood of contours or straight edges, where the signal change is only in one direction. These maxima are less stable because their localization is more sensitive to noise or small changes in neighbouring texture.

A more sophisticated approach, solving this problem, is to select the scale for which the trace and the determinant of the Hessian matrix (*H*) simultaneously assume a local extremum [26]. The trace of the *H* matrix is equal to the LoG but detecting simultaneously the maxima of the determinant penalizes points for which the second derivatives detect signal changes in only one direction. A similar idea is explored in the Harris detector, although it uses the first derivatives. The second derivative gives a small response exactly in the point where the signal change is most significant.

Therefore the maxima are not localized exactly at the largest signal variation, but in its neighbourhood. A different approach for the scale selection was proposed by Kadir [16]. They explore the idea of using local complexity as a measure of saliency. The salient scale is selected at the entropy extremum of the local descriptors. The selected scale is therefore descriptor dependent. The method searches for scale localized features with high entropy, with the constraint that the scale is isotropic.

2.2. Affine Invariant Detectors

An affine invariant detector can be seen as a generalization of the scale invariant detector. In the case of an affine transformation the scaling can be different in each direction. The non uniform scaling has an influence on the localization, the scale and the shape of a local structure. Therefore, the scale invariant detectors fail in the case of significant affine transformations. An affine invariant algorithm for corner detection was proposed by Alvarez [1].

They apply affine morphological multi-scale analysis to extract corners. For each extracted point they build a chain of points detected at different scales, but associated with the same local image structure. The final location and orientation of the corner is computed using the bisector line given by the chain of points. A similar idea was previously explored by Deriche [8]. The main drawback of these approaches is that an interest point in images of natural scenes cannot be approximated by a model of a perfect corner, as it can take any form of a bi-directional signal change. The real points detected at different scales do not move along a straight bisector line as the texture around the points significantly influences the location of the local maxima.

This approach cannot be a general solution to the problem of affine invariance but gives good results for images where the corners and multi-junctions are formed by straight or nearly straight step-edges. Our approach makes no assumption on the form of a local structure. It only requires a bi-directional signal change. Recently, Tuytelaars [31] and [32] proposed two approaches for detecting image features in an affine invariant way. The first one starts from Harris points and uses the nearby edges. Two nearby edges, which are required for each point, limit the number of potential features in an image. A parallelogram region is bounded by these two edges and the initial Harris point. Several intensity based functions are used to determine the parallelogram. In this approach, a reliable algorithm for extracting the edges is necessary. The second method is purely intensity-based and starts with extraction of local intensity extrema. Next, the algorithm investigates the intensity profiles along rays going out of the local extremum. An ellipse is fitted to the region determined by significant changes in the intensity profiles. A similar approach based on local intensity extrema was introduced by Matas [23]. They use the water-shed algorithm to find intensity regions and fit an ellipse to the estimated boundaries. Lindeberg [20] developed a method for finding blob-like affine features with an iterative procedure in the context of shape from texture. The affine invariance of shape adapted fixed points was also used for estimating surface orientation from binocular data (shape from disparity gradients). This work provided the theory for the affine invariant detector presented in this paper. It explores the properties of the Mikolajczyk [24] second moment matrix and iteratively estimates the affine transformation of local patterns. The authors propose to extract the points using the maxima of a uniform scale-space representation and to iteratively modify the scale and the shape of points. However, the location of points is detected only at the initial step of the algorithm, by the circularly symmetric, not affine invariant Laplacian measure.

Therefore, the spatial location of the maximum can be slightly different if the pattern undergoes a significant affine deformation. This method was also applied to detect elliptical blobs in the context of hand tracking [18]. The affine shape estimation was used for matching and recognition by Baumberg [2]. He extracts interest points at several scales using the Harris detector and then adapts the shape of the point neighbourhood to the local image structure using the iterative procedure proposed by Lindeberg [21].

The affine shape is estimated for a fixed scale and fixed location, that is the scale and the location of the points are not extracted in an affine invariant way. The points as well as the associated regions are therefore not invariant in the case of significant affine transformations (see section 6.1 for a quantitative comparison). Furthermore, there are many points repeated at the neighbouring scale levels (see figure 1), which increases the probability of false matches and the complexity. Recently, Schaffalitzky [28] extended the Harris-Laplace detector [24] by affine normalization proposed by Baumberg [2]. However, the location and scale of points are provided by the scale invariant Harris-Laplace detector [24] and [26], which is not invariant to significant affine transformations.

3. Scale Invariant Interest Point Detector

The evaluation of interest point detectors presented in Schmid [29] demonstrate an excellent performance of the Harris detector compared to other existing approaches [6], [10], [14] and [15]. However this detector is not invariant to scale changes. In this section we propose a new interest point detector that combines the reliable Harris detector [13] with automatic scale selection [26] to obtain a scale invariant detector. In Section 3.1 we introduce the methods SIFT-Features on which we base the approach. In Section 3.2 we discuss in detail the scale invariant detector and present an example of extracted points.

3.1. Scale Invariant Features Transform

First step of our work is the automatic extraction of image features followed by an automatic matching process in order to assign each image detail a unique number referring to the corresponding object detail. As we deal with aerial images and cartographic databases we use point type image features. As we do not pose any restrictions on the exterior orientation their detection and description should be scale and rotation invariant. We therefore use the point operator proposed by Lowe [22]. The operator detects points in an image pyramid and describes the points by means of rotation and scale-invariant features, so called SIFT-features („Scale-Invariant Feature Transform'').

The descriptor represents the scale dependent window around the point with 16 histograms of the gradient orientations leading to 128 values in the range between 0 and 255. As local projective distortions can often be approximated quite well by a scale and a rotation transformation, using SIFT-features allows handling nearly arbitrary camera positions provided that enough tie points are available. This is an important advantage compared to procedures which employ a conventional point detector and subsequent correlation or least squares matching. Such procedures only work with relative rotations up to approximate 15 degrees.

3.2. SIFT Features

SIFT features were proposed in [22] as a method of extracting and describing key points which are robustly invariant to common image transforms. The Scale Invariant Feature Transform (SIFT) algorithm has 4 major stages.

- Scale-space extrema detection: The first stage searches over scale space using a Difference of Gaussian function to identify potential interest points.
- Key point localisation: The location and scale of each candidate point is determined and key points are selected based on measures of stability.
- Orientation assignment: One or more orientations are assigned to each key point based on local image gradients.
- Key point descriptor: A descriptor is generated for each key point from local image gradients information at the scale found in stage 2.

An important aspect of the algorithm is that it generates a large number of features over a broad range of scales and locations. The number of features generated is dependent on image size and content, as well as algorithm parameters. A typical image of 500x500 pixels will generate approximately 2000 features however in our indoor examples a similar size image will typically only generate 300 features. In figure 1, typical indoor office environment and the extracted SIFT features with their locations represented by arrows. The length of the arrow represents the scale of the extracted key point and the direction represents the orientation of the descriptor. The SIFT feature algorithm is based upon finding locations within the scale space of an image which can be reliably extracted. The first stage finds scale-space extrema located in D(x, y, θ), the Difference of Gaussians (DOG) function, which can be computed from the difference of two nearby scaled images separated by a multiplicative factor k:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1)$$

Where L(x, y, σ) is the scale space of an image, we built by convolving the image I(x, y) with the Gaussian kernel G(x, y, σ). Points in the DOG function which are local extrema in their own scale and one scale above and below are extracted as key points. Generation of extrema in this stage is dependent on the frequency of sampling in the scale space k and the initial smoothing σ_0. The key points are then filtered for more stable matches and more accurately localised to scale and sub pixel image location using methods described in Brown [5].

Before a descriptor for the key point is constructed, the key point is assigned an orientation to make the descriptor invariant to rotation. This key point orientation is calculated from an orientation histogram of local gradients from the closest smoothed image L(x, y, σ). For each image sample L(x, y) at this scale, the gradient magnitude m(x, y) and orientation θ(x, y) is computed using pixel differences:

$$m(x, y) = ((L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2)^{1/2} \quad (2)$$

$$\theta(x, y) = \tan^{-1}(L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y)) \quad (3)$$

The orientation histogram has 36 bins covering the 360 degree range of orientations. Each point is added to the histogram weighted by the gradient magnitude, m(x, y), and by a circular Gaussian with σ variance that is 1.5 times the scale of the key point. Additional key points are generated for key point locations with multiple dominant peaks whose magnitude is within 80% of each other. The dominant peaks in the histogram are interpolated with their neighbours for a more accurate orientation.

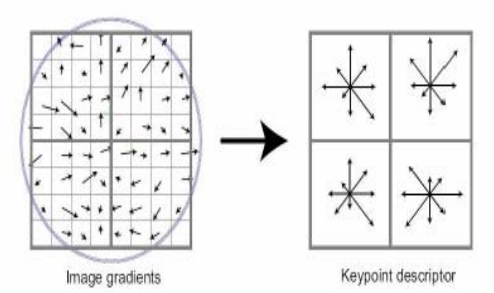

Figure 1. A key point descriptor is created using the gradient magnitude, m(x, y) and orientation, θ(x, y) around the key point. These are weighted by a circular Gaussian window indicated by the overlaid circle. Each orientation histogram is calculated from a 4x4 pixel support window and divided over 8 orientation bins. Figure from [22]

The local gradient data from the closest smoothed image L(x, y, σ) is also used to create the key point descriptor. This gradient information is first rotated to align it with the assigned orientation of the key point and then weighted by a Gaussian with variance that is 1.5 times the scale of the key point. The weighted data is used to create a nominated number of histograms over a set window around the key point. Typical key point descriptors use 16 orientation histograms aligned in a 4x4 grid. Each histogram has 8 orientation bins each created over a support window of 4x4 pixels. The resulting feature vectors are 128 elements with a total support window of 16x16 scaled pixels. For a more detailed discussion of the key point generation and factors involved see [22]. Like other feature-based matching systems, it sequentially finds key points in an image, determines a feature vector for each, and matches those features to a database. Each of these steps is described as bellow:

1. Detect key points based scale-space extremas
2. Remove unstable key points
3. Determine key point orientation
4. Construct key point descriptor

- Construct a scale-space representation

Very similar to the standard Laplacian pyramid (see figure 2).

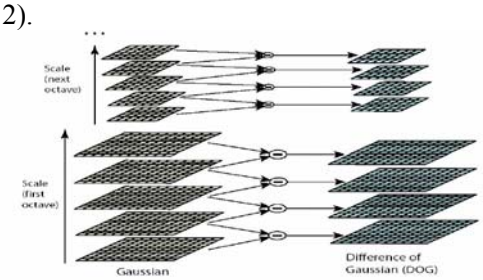

Figure 2. Octaves are separated by 2*σ in scale space

- Gaussian and Difference of Gaussians Example (see figure 3).

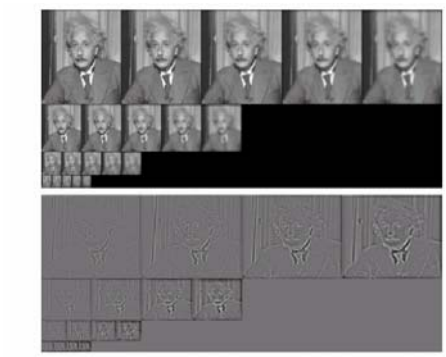

Figure 3. Difference Gaussians example

- Scale-space Extrema Detection (see figure 4).

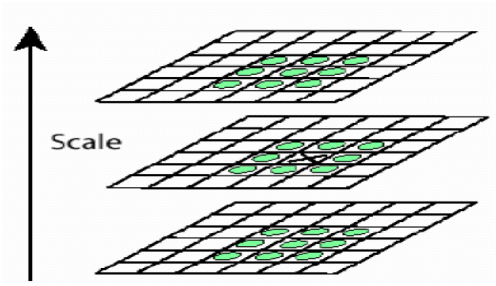

Figure 4. Pixel marked with "X" is compared to 26 neighbors in a 3x3x3 window that spans adjacent pixels and scales

- Accurate key point localization: Our scale-space representation is **desecrated**; hence the detected key points are only close to true minima / maxima in the scale-space. What can we do to add accuracy to key point detection:
  - Select finer decartelization (costly)
  - Fit a polynomial (quadratic) function to the discrete pixel values, and look for a minimum/maximum in the fitted function (details omitted)
- Removing unstable points:
  - Key points that correspond to regions of low contrast are removed
  - Not enough distinct information
  - Key points along the edges are removed
  - Helps with scale and rotation invariance

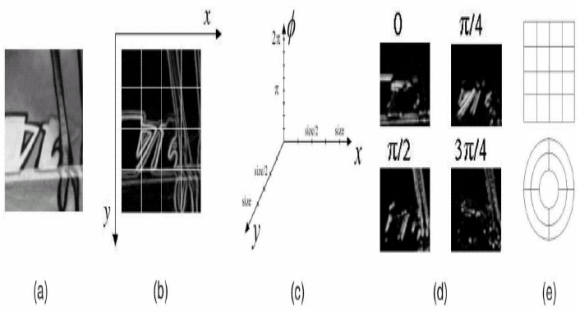

Figure 5. SIFT descriptor. (a) Detected region. (b) Gradient image and location grid. (c) Dimensions of the histogram. (d) Four of eight orientation planes. (e) Cartesian and the log-polar location grids. The log-polar grid shows nine location bins used in shape context (four in angular direction).

1. Identify candidate feature key points
2. Finding extrema across scales. Gaussians blur the scales with a bunch of stepped-variance Gaussians, and take the difference to approximate a Laplacian. Find maxima and minima that are maximal and minimal across different levels of blur, and different scales, effectively finding scale-invariant illumination extrema.
3. Removing undesirable extrema. First, each extremum was adjusted to be the maximum around the original extremum of the derivative of a Taylor expansion of a difference of Gaussians. Bad extrema (as determined from a criterion based on this adjustment) were thrown out. Next, extrema corresponding to otherwise-un-noteworthy locations along important edges were removed. Without this, very small changes to an image could lead to very different points along an important edge (that isn't around other features) being chosen as key points. To get rid of these, they got rid of key points where the ratio of the Eigen values of a Hessian of the difference of Gaussians (by space) exceeded a threshold.

4. Calculate feature value at each key point

Now that we know where our features are, we calculate them. First normalizing for feature-neighbourhood orientation, we create a grid of orientation histograms for local areas surrounding the key point. This seems to work pretty well. Intuitively, I wonder about orientation stability at finer scales (not when everything's blurred out, but pixel-scale features might not survive rotation well). By allowing a little bit of sliding of position within much localized areas, the system attempts to use the same trick the visual system does for local location independence in primary visual cortex complex cells. SIFT descriptors are computed for normalized image patches with the code provided by Lowe [22]. A descriptor is a 3D histogram of gradient location and orientation, where location is quantized into a 4*4 location grid and the gradient angle is quantized into eight orientations. The resulting descriptor is of dimension 128. The figure 5 illustrates the approach. Each orientation plane represents the gradient magnitude corresponding to a given orientation. To obtain illumination invariance, the descriptor is normalized by the square root of the sum of squared components. After this first issue, namely, feature

selection by extracting valid and invariant set of features in both images. The second issue, namely, feature matching is very difficult problem and it is unlikely that any single algorithm will provide a solution enough to address all application and in particular in cartographic data bases image. So in this paper we address the issue of feature matching by use of weighted bipartite graph as detailed in the section below.

4.1. Weighted bipartite graphs

A weighted bipartite graphs whose vertex set can be divided into two disjoint sets U and V such that all edges run between U and V. Formally, such a graph is denoted by G = (U, V, E, W) where U and V are disjoint sets of vertices, R U x V is a set of edges and W: E →R is a function that associates a real number called a weight to each edge. For simplicity, let us assume that $|U| = |V| = n$ and that E = U x V (i.e., every vertex in U is connected to every vertex in V). Given any subset A of edges, the weight of A is defined to be the sum of weights of the edges in A. we define a matching to be a set M of edges that (1) $|M| = n$ and (2) each vertex in the graph is incident on exactly one edge in M. if an edge (u, v) belongs to M then vertex u is said to be matched to vertex v and vice versa. A matching with a minimum weight is said to be the minimum weight matching or simply a minimum matching. in figure 6 shows a weighted bipartite graph with n = 3 and a minimum weight matching of weight 11.

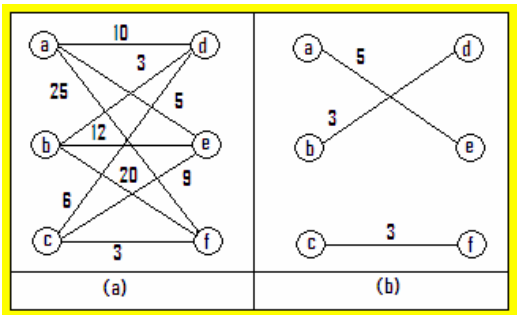

Figure 6. An example of a weighted bipartite graph (a) and a minimum weight matching (b).

There are two elegant algorithms for finding a minimum weight matching in a weighted bipartite graph called Hungarian method and greedy algorithm.

Despite the fact that the greedy algorithm does not always provide a minimum weight matching, It is useful to use also the Hungarian method in combination with it or a hybrid method (greedy and Hungarian). For a detailed discussion about these methods, the reader is referred to Refs [8], [9] and [10]. These three algorithms give as results a set of minimal weight matching between images point descriptors.

5. Computational Complexity

The complexity and efficiency of a feature detector is an important issue in particular when applying the detectors to image sequences or large image databases as Cato graphics one. The comparison done in related works shows that the computation time required by the detectors. Here, each detector is applied to an image of size 800 × 640 Detection is done on a Pentium 4 1.3 GHz. The run time is the computational time required by a Pentium 4 1.3 GHz to detect features in a 800×640 image. This time can slightly vary depending on the number of features in the image. The fastest detector is SIFT since it only smoothes subtracts and samples the image. The scale selection (4 the best one and σ = 2) All these Simplifications can significantly reduce the detection time but at the cost of accuracy.

6. Applications

In this section we present an example application for our interest point detectors and show how they can be used to match image pairs with significant scale or viewpoint changes. For examples of other applications the reader is referred to Lazebnik [19], Rothganger [27], and Schaffalitzky [28]. In Section 6.1 we describe our matching approach. Section 6.2 shows the results for scale and affine invariant features.

6.1. Matching Algorithm

Given an image we detect a set of interest points and compute the point descriptors. The descriptors are then compared with a similarity measure. The resulting similarity is used for finding the corresponding points.

Descriptors and Similarity Measure

Our descriptors are Gaussian derivatives computed in the local neighborhood of interest points. Invariance to rotation is obtained by Scale & Affine Invariant Interest Point Detectors 81 "steering" the derivatives in the direction of the gradient [11]. To

obtain a stable estimate of the gradient direction, we use the average gradient orientation in a point neighbourhood [25] and [26]. Invariance to affine intensity changes is obtained by dividing the higher order derivatives by the first derivative. We obtain descriptors of dimension 12 by using derivatives up to 4th order. To measure the similarity between the descriptors we use the Mahalanobis distance. The covariance matrix is estimated over a large set of images and incorporates signal noise, variations in photometry as well as inaccuracy of the interest point location.

Matching

To robustly match the images, we first reduce the space of Interest points given by SIFT algorithm, we project theses sets of features in a reduced space (i.e., two dimensional), this is done by comparing each point to its eight neighbour in each image, in second phase, we determine point-to-point correspondences using the similarity measure. We select for each descriptor in the first image the most similar descriptor in the second image using the Mahalanobis distance. If the distance is below a threshold the match is potentially correct. A set of initial matches is obtained. In the second step of verification we apply cross-correlation, which rejects low-score matches. Finally, a robust estimation of the transformation between the two images based on Random Sample Consensus (RANSAC) enables the selection of the inliers. In our experiments the transformation is either a homographic or a fundamental matrix. A model selection algorithm [17] and [30] can be used to automatically decide which transformation is the most appropriate.

6.2. Experimental Results for Matching

In this section, we present matching results in the presence of scale and viewpoint changes. The results are obtained with the Harris-Laplace and the Harris-Affine detector. We show the matched points which are inliers to the estimated transformations. The number of correctly matched descriptors is limited by the number of corresponding features provided by the detector and depends on the accuracy of the detectors. The matching approach is based on the distance measure e between the descriptors and RANSAC. If the fraction of inliers among the initial matches is too small then RANSAC fails. Note that there are points which are correctly detected but are rejected by the distance measure. However, these points could be matched by using a more distinctive descriptor or by applying semi-local constraints.

7. Conclusion and Future Work

SIFT features provide a distinct and accurate means of matching digital images for image retrieval and vision based localisation. In this paper we have presented a reduction to the traditional SIFT feature by projecting the set of descriptors in reduced space of two dimension to improve their performance. This reduction uses the structure of an indoor environment to remove the need for translation invariance of the features in different scale. From the results obtained we have shown that this reduction has a minimal affect on the retrieval rate of images and significantly reduces the size of the image descriptors and the time to needed to generate and match them. Furthermore we have used the scale information of the SIFT features to improve location discrimination. Future work with the reduced SIFT feature will look at further reducing the size of the image description by filtering key points with low matching likelihood. A future area of work will also be the use of the proposed approaches in different applications, as for example, shot matching in a video sequence, recognition of object classes and tracking.

8. References

[1] Alvarez, L. and Morales, F. 1997. Affine morphological multiscale analysis of corners and multiple junctions. *International Journal of Computer Vision*, 2(25):95–107.

[2] Baumberg, A. 2000. Reliable feature matching across widely separated views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA, pp. 774–781.

[3] Bretzner, L. and Lindeberg, T. 1998. Feature tracking with automatic selection of spatial scales. *Computer Vision and Image Understanding*, 71(3):385–392.

[4] Brown, M. and Lowe, D.G. 2002. Invariant features from interest point groups. In *The 13th British Machine Vision Conference*, Cardiff University, UK, pp. 253–262.

[5] Brown, M. and Lowe, D.G. 2002 Invariant features from interest point groups. In British Machine Vision Conference, BMVC 2002, Cardiff, Wales (September 2002), pp. 656-665.

[6] Cottier, J. 1994. Extraction et appariements robustes des points d'intérêts de deux images non étalonnés. Crowley, J. 1981.Arepresentation for visual information.PhDthesis, Carnegie Mellon University.

[7] Crowley, J. and Parker, A. 1984. A representation for shape based on peaks and ridges in the difference of low pass transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):156–170.

[8] Deriche, R. and Giraudon, G. 1993. A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 10(2):101–124.

[9] Förstner,W. and Gülchm, E. 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, pp. 281–305.

[10] Förstner, W. 1994. A framework for low level feature extraction. In *Proceedings of the 3rd European Conference on Computer Vision*, Stockholm, Sweden, pp. 383–394.

[11] Freeman, W. and Adelson, E. 1991. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906.

[12] Goodrum, A. Image Information Retrieval: An Overview of Current Research. In Informing Science, Volume 3, No 2, 2000.

[13] Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In *Alvey Vision Conference*, pp. 147–151.

[14] Heitger, F., Rosenthaler, L., Von der Heydt, R., Peterhans, E., and Kuebler, O. 1992. Simulation of neural contour mechanism: From simple to end-stopped cells. *Vision Research*, 32(5):963–981.

[15] Horaud, R., Skordas, T., and Veillon, F. 1990. Finding geometric and relational structures in an image. In *Proceedings of the 1st European Conference on Computer Vision*, Antibes, France, pp. 374–384.

[16] Kadir, T. and Brady,M. 2001. Scale, saliency and image description. *International Journal of Computer Vision*, 45(2):83–105,

[17] Kanatani, K. 1998. Geometric information criterion for model selection. *International Journal of Computer Vision*, 26(3):171–189.

[18] Laptev, I. and Lindeberg, T. 2001. Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features. In *Proceedings of Scale Space and Morphology Workshop*, Vancouver, Canada, vol. 2106. Lecture Notes in Computer Science, pp. 63–74.

[19] Lazebnik, S., Schmid, C., and Ponce, J. 2003. Sparse texture representation using affine-invariant neighbourhoods. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, pp. 319–324.

[20] Lindeberg, T. and Garding, J. 1997. Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2- D brightness structure. *Image and Vision Computing*, 15(6):415–434.

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

[22] Lowe, D.G. 1999. Object recognition from local scale-invariant features. In *Proceedings of the 7th International Conference on Computer Vision*, Kerkyra, Greece, pp. 1150–1157.

[23] Matas, J., Chum, O., Urban, M., and Pajdla, T. 2002. Robust wide baseline stereo from maximally stable extremal regions. In *The 13th British Machine Vision Conference*, Cardiff University, UK, pp. 384–393.

[24] Mikolajczyk, K. and Schmid, C. 2001. Indexing based on scale invariant interest points. In *Proceedings of the 8th InternationalConference on Computer Vision*, Vancouver, Canada, pp. 525–531.

[25] Mikolajczyk, K. 2002. Interest point detection invariant to affine transformations. PhD thesis, Institut National Polytechnique de Grenoble.

[26] Mikolajczyk, K. and Schmid, C. 2002. An affine invariant interest point detector. In *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, vol. I, pp. 128–142.

[27] Rothganger, F., Lazebnik, S., Schmid, C., and Ponce, J. 2003. 3D Object modelling and recognition using affine-invariant patches and multi-view spatial constraints. In *Proceedings of the Conference on Computer Vision and Pattern*

Recognition, Madison, Wisconsin, USA, pp. 272–277.

[28] Schaffalitzky, F. and Zisserman, A. 2002. Multi-view matching for unordered image sets. In *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, pp. 414–431.

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

[30] Triggs, B. 2001. Joint feature distributions for image correspondence. In *Proceedings of the 8th International Conference on Computer Vision*, Vancouver, Canada, pp. 201–208.

[31] Tuytelaars, T. and Gool, L.V. 1999. Content-based image retrieval based on local affinely invariant regions. In *Int. Conf. on VisualInformation Systems*, pp. 493–500.

[32] Tuytelaars, T. and Van Gool, L. 2000.Wide baseline stereo matching based on local, affinely invariant regions. In *The 11th British Machine Vision Conference*, University of Bristol, UK, pp. 412–425.

Mr. Houari Kamel is a assistant professor in university of Oum El Bouaghi, Algeria.

Dr Youssef Chahir is Master of conferences (associate professor) in University of Caen, France. He is member of Research Group in Data processing, Picture, Instrumentation of Caen (GREYC - URA CNRS 6072).

Dr. Mohamed-Khireddine KHOLLADI is a Doctor of state in Data processing (graduate Doctor of the INSA of Lyon). He is currently a head of Computer science department of the Faculty of Science of the engineer of the Mentouri university of Constantine and president of association sciences and development of the university. He is a head of team of research at the documentary research laboratory of the department of biblio-economics and member of the vision team & computer graphics of the research laboratory LIRE of the computer sciences department. He is member of the network Cassini in France and member of network ACIT International. Its work concerns the geographical information systems, the computer graphics, the treatment and synthesis of image, national's archives, knowledge bases, the new technology of information and communication "NTIC", etc.