VF-SIFT: Very Fast SIFT Feature Matching

Faraj Alhwarin, Danijela Ristić–Durrant, and Axel Gräser

Institute of Automation, University of Bremen, Otto-Hahn-Alle NW1,
D-28359 Bremen, Germany
{alhwarin,ristic,ag}@iat.uni-bremen.de

Abstract. Feature-based image matching is one of the most fundamental issues in computer vision tasks. As the number of features increases, the matching process rapidly becomes a bottleneck. This paper presents a novel method to speed up SIFT feature matching. The main idea is to extend SIFT feature by a few pairwise independent angles, which are invariant to rotation, scale and illumination changes. During feature extraction, SIFT features are classified based on their introduced angles into different clusters and stored in multidimensional table. Thus, in feature matching, only SIFT features that belong to clusters, where correct matches may be expected are compared. The performance of the proposed methods was tested on two groups of images, real-world stereo images and standard dataset images, through comparison with the performances of two state of the arte algorithms for ANN searching, hierarchical k-means and randomized kd-trees. The presented experimental results show that the performance of the proposed method extremely outperforms the two other considered algorithms. The experimental results show that the feature matching can be accelerated about 1250 times with respect to exhaustive search without losing a noticeable amount of correct matches.

Keywords: Very Fast SIFT, VF-SIFT, Fast features matching, Fast image matching.

1 Introduction

Feature-based image matching is a key task in many computer vision applications, such as object recognition, images stitching, structure-from-motion and 3D stereo reconstruction. These applications require often real-time performance.

The SIFT algorithm, proposed in [1], is currently the most widely used in computer vision applications due to the fact that SIFT features are highly distinctive, and invariant to scale, rotation and illumination changes. However, the main drawback of SIFT is that the computational complexity of the algorithm increases rapidly with the number of keypoints, especially at the matching step due to the high dimensionality of the SIFT feature descriptor. In order to overcome this drawback, various modifications of SIFT algorithm have been proposed. Ke and Sukthankar [2] applied Principal Components Analysis (PCA) to the SIFT descriptor. The PCA-SIFT reduces the SIFT feature descriptor dimensionality from 128 to 36, so that the PCA-SIFT is size of the SIFT feature descriptor length, which speeds up feature matching by a factor 3 compared to the original SIFT method.

M. Goesele et al. (Eds.): DAGM 2010, LNCS 6376, pp. 222–231, 2010.

Recently, several papers [5, 6] were published addressing the use of modern graphics hardware (GPU) to accelerate some parts of the SIFT algorithm, focused on features detection and description steps. In [7] GPU was exploited to accelerate features matching. These GPU-SIFT approaches provide 10 to 20 times faster processing. Other papers such as [8] addressed implementation of SIFT on a Field Programmable Gate Array (FPGA) achieving about 10 times faster processing.

The matching step can be speeded up by searching for the Approximate Nearest Neighbor (ANN) instead of the exact one. The most widely used algorithm for ANN search is the kd-tree [9], which successfully works in low dimensional search space, but performs poorly when feature dimensionality increases. In [1] Lowe used the Best-Bin-First (BBF) method, which is expanded from kd-tree by modification of the search ordering so that bins in feature space are searched in the order of their closest distance from the query feature and stopping search after checking the first 200 nearest-neighbors. The BBF provides a speedup factor of 2 times faster than exhaustive search while losing about 5% of correct matches. In [10] Muja and Lowe compared many different algorithms for approximate nearest neighbor search on datasets with a wide range of dimensionality and they found that two algorithms obtained the best performance, depending on the dataset and the desired precision. These algorithms used either the hierarchical k-means tree or randomized kd-trees.

In [11] a novel strategy to accelerate SIFT feature matching as a result of extending a SIFT feature by two new attributes (feature type and feature angle) was introduced. When these attributes are used together with SIFT descriptor for matching purposes so that only features having the same or very similar attribute are compared, the execution of the SIFT feature matching can be speeded up with respect to exhaustive search by a factor of 18 without a noticeable loss of accuracy.

In this paper, a SIFT feature is extended by 4 new pairwise independent angles. These angles are computed from SIFT descriptor. In the SIFT feature extraction phase, the features are stored in 4 dimensional table without extra computational cost. Then, in the matching phase only SIFT features belonging to cells where correct matches may be expected are compared. By exploiting this idea, the execution of the SIFT feature matching can be speeded up by a factor of 1250 with respect to exhaustive search without a noticeable loss of accuracy.

2 Original SIFT Method

The Scale Invariant Feature Transform (SIFT) method takes an image and transforms it into a set of local features extracted through the following three stages, explained here shortly. The more details can be found in [1]:

1. Feature detection and localization: The locations of potential interest points in the image are determined by selecting the extrema of DoG scale space. For searching scale space extrema, each pixel in the DoG images is compared with its 26 neighbors in 3×3 regions of scale-space. If the pixel is lower/larger than all its neighbors, then it is labeled as a candidate keypoint. Each of these keypoints is exactly localized by fitting a 3 dimensional quadratic function computed using a second order Taylor expansion around keypoint. Then keypoints are filtered by discarding points of low contrast and points that belong to edges.

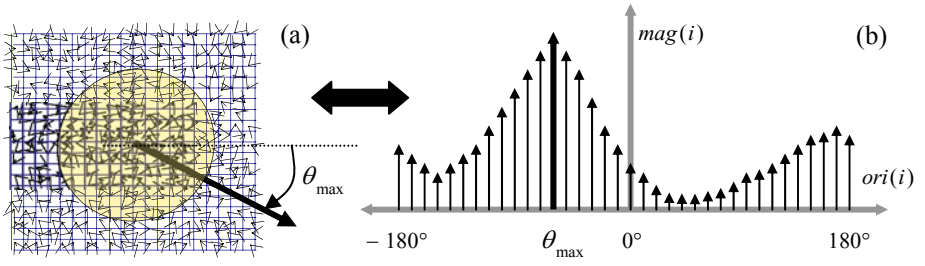

Fig. 1. (a) Gradient image patch around a keypoint. (b) A 36 bins OH constructed from gradient image patch.

2. Feature orientation assignment: An orientation is assigned to each key point based on local image gradient data. For each pixel in a certain image region around the keypoint, the first order gradient is calculated (gradient magnitude and orientation). The gradient data are weighted by scale dependent Gaussian window (illustrated by a circular window on Fig 1a) and then used to build a 36-bin orientation histogram (OH) covering the range of orientations [-180°, 180°] as shown in Fig. 1b. The orientation of the SIFT feature θ_{max} is defined as the orientation corresponding to the maximum bin of the OH as shown in Fig. 1.

3. Feature descriptor: The gradient image patch around keypoint is weighted by a Gaussian window with σ equal to one half the width of the descriptor window (illustrated with a circular window on Fig. 2a) and rotated by θ_{max} to align the feature orientation with the horizontal direction in order to provide rotation invariance (see Fig. 2a). After this rotation, the region around the keypoint is subdivided into 4x4 square sub-regions. From each sub-region, an 8 bin sub-orientation histogram (SOH) is built as shown in Fig. 2b. In order to avoid boundary affects, trilinear interpolation is used to distribute the value of each gradient sample into adjacent histogram bins. Finally, the 16 resulting SOHs are transformed into 128-D vector. The vector is normalized

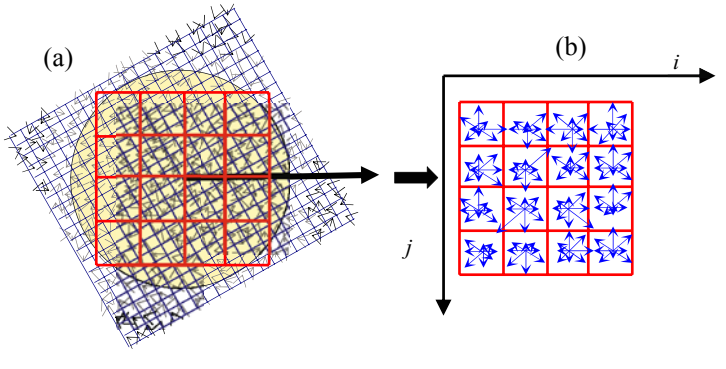

Fig. 2. (a) Rotated gradient image patch with a 4x4 rectangular grid. (b) 16 8-bins SOHs used to build SIFT-descriptor.

to unit length to achieve the invariance against illumination changes. This vector is called SIFT descriptor (SIFT-D) and is used for similarity measuring between two SIFT features.

3 Very Fast SIFT Feature

Generally, if a scene is captured by two cameras or by one camera but from two different viewpoints, the corresponding points in two resulted images will have different image coordinates, different scales, and different orientations. Nevertheless, they must have almost similar descriptors which are used to match the images using a similarity measure [1]. The high dimensionality of descriptor makes the feature matching very time-consuming.

In order to speed up the features matching, it is assumed that 4 pairwise independent angles can be assigned to each feature. These angles are invariant to viewing geometry and illumination changes. When these angles are used for feature matching together with SIFT-D, we can avoid the comparison of a great portion of features that can not be matched in any way. This leads to a significant speed up of the matching step as will be shown below.

3.1 SIFT Feature Angles

In [11], a speeding up of SIFT feature matching by 18 times compared to exhaustive search was achieved by extending SIFT feature with one uniformly-distributed angle computed from the OH and by splitting features into Maxima and Minima SIFT features. In this paper the attempts are made to extend SIFT feature by few angles, which are computed from SIFT-D. As described in section 2, for computation of SIFT-D, the interest region around keypoint is subdivided in sub-regions in a rectangular grid. From each sub-region a SOH is built. Theoretically, it is possible to extend a SIFT feature by a number of angles equal to the number of SOHs as these angles are to be calculated from SOHs. In case of 4x4 grid, the number of angles is then 16. However, to reach the very high speed of SIFT matching, these angles should be components of a multivariate random variable that is uniformly distributed in the 16-dimensional space$[-180°, 180°]^{16}$. In order to meet this requirement, the following two conditions must be verified [15]:

- Each angle has to be uniformly distributed in [-180°, 180°] (equally likely condition).
- The angles have to be pairwise independent (pairwise independence condition).

In this section, the goal is to find a number of angles that are invariant to geometrical and photometrical transformations and that meet the above mentioned conditions. First, the angles between the orientations corresponding to the vector sum of all bins of each SOH and the horizontal orientation are suggested as the SIFT feature angles. Fig. 3b presents geometrically the vector sum of a SOH. Mathematically, the proposed angles $\{\theta_{ij}; i, j = 1,..,4\}$ are calculated as follows:

$$\theta_{ij} = \tan^{-1}\left(\sum_{k=0}^{7} mag_{ij}(k) \cdot \sin\left(ori_{ij}(k)\right) \Big/ \sum_{k=0}^{7} mag_{ij}(k) \cdot \cos\left(ori_{ij}(k)\right)\right) \quad (1)$$

where $mag_{ij}(k)$ and $ori_{ij}(k)$ are the magnitude and the orientation of the k^{th} bin of the ij^{th} SOH respectively. Since the angles θ_{ij} are computed from SOHs, from which the SIFT-D is built, they are invariant to geometrical and photometrical transformations. However, these angles must be examined, whether they meet the equally likely and pairwise independence conditions.

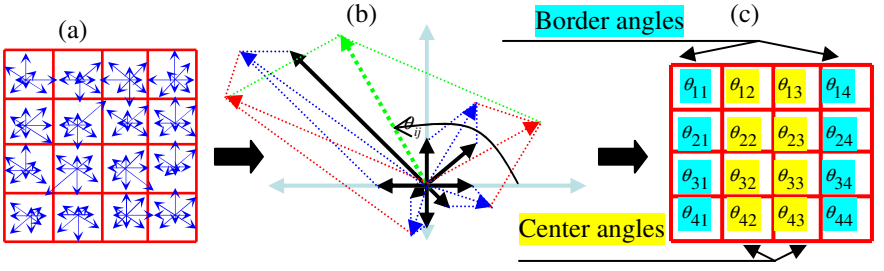

Fig. 3. (a) SOHs, (b):Vector sum of the bins of a SOH, (c) angles computed from SOHs

The equally likely condition: To examine whether the angles θ_{ij} meet the equally likely condition, they are considered as random variables Θ_{ij}. The probability density functions (PDFs) of each angle are estimated from 10^6 SIFT features extracted from 700 test images (500 standard data set images and 200 stereo images from a real-world robotic application). Some examples of used test images are given in Fig 6.

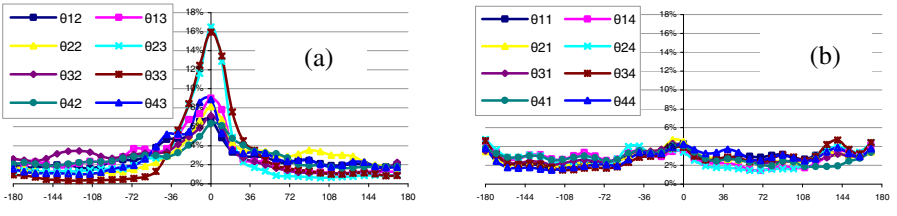

Fig. 4. The PDFs of (a) center and (b) border angles estimated from 10^6 SIFT features extracted from 700 images

The PDFs of Θ_{ij} was computed by dividing the angle space [-180°, 180°] into 36 sub-ranges, where each sub-range cover 10°, and by counting the numbers of SIFT features whose angle θ_{ij} belong to each sub-range. As can be seen from Fig. 4, the angles that are calculated from SOHs around the center of SIFT feature (called center angles), have distributions concerned about 0°, whereas the angles that are calculated from SOHs of the grid border (called border angles), tend to be equally likely distributed over the angle range. The reason of this outcome can be interpreted as follows: On the one hand, the SOHs are computed from the interest region (where OH is computed) after its rotation as shown in Fig. 2a. Therefore the orientations of the

maximum bin of each center SOH tend to be equal 0°. On the other hand, for each SOH, the orientation of the maximum bin and the orientation of the vector sum of all bins are strongly dependent since the vector sum includes the maximum bin that has the dominant influence to the vector sum [11]. In the contrary, the border SOHs and the OH do not share the same gradient data, therefore only border angles meet the equally likely condition. Fig. 3c presents the border and center angles.

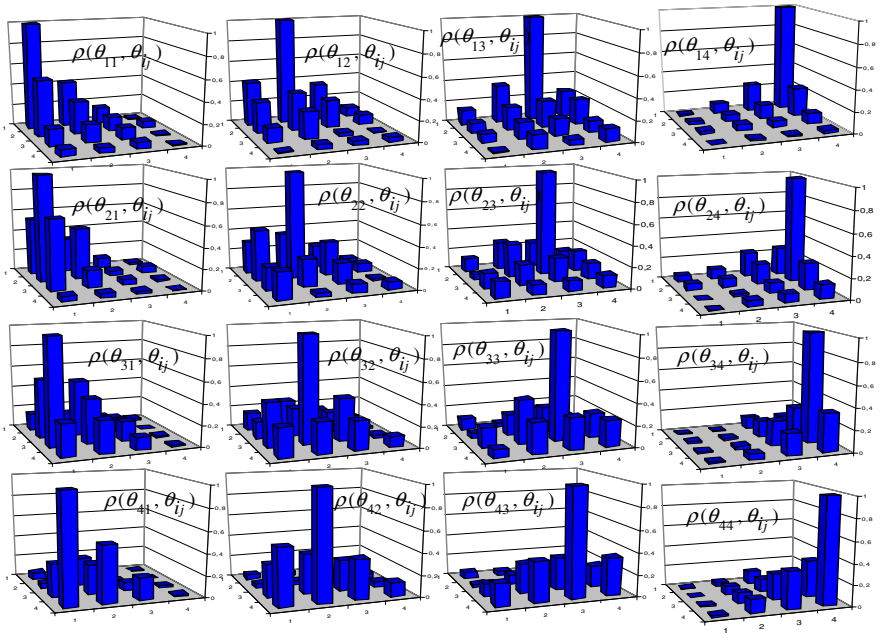

Fig. 5. The correlation coefficients between angles of SIFT features. For example the top left diagram presents correlation coefficients between θ_{11} and all θ_{ij}. The x and y axes present indices i and j respectively while z axis present correlation factor.

The pairwise independence condition: In order to examine whether suggested angles θ_{ij} meet the pairwise independence condition, it is needed to measure the dependence between each two angles. The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient. It is obtained by dividing the covariance of the two variables by the product of their standard deviations. Assuming that two random variables are given X and Y with expected values μ_x and μ_y and standard deviations σ_x and σ_y then the Pearson product-moment correlation coefficient ρ_{xy} between them is defined as:

$$\rho_{xy} = E\left[\left(X - \mu_x\right)\left(Y - \mu_y\right)\right] / \sigma_x \sigma_y \tag{2}$$

where $E[\bullet]$ is the expected value operator.

The correlation coefficients between each two angles α and β are computed using 10^6 SIFT features extracted from the considered test images.

$$\rho_{\alpha\beta} = 10^6 \cdot \sum_{i=1}^{10^6} \left(\left(\alpha_i - \mu_\alpha \right) \left(\beta_i - \mu_\beta \right) \right) \Bigg/ \left(\sqrt{\sum_{i=1}^{10^6} \left(\alpha_i - \mu_\alpha \right)^2} \cdot \sqrt{\sum_{i=1}^{10^6} \left(\beta_i - \mu_\beta \right)^2} \right) \tag{3}$$

The estimated correlation coefficients are explained in Fig. 5. As evident from Fig. 5, angles that are computed from contiguous SOHs, are highly correlated, whereas there is no or very weak correlations between two angles that computed from non-contiguous SOHs. The reason of this outcome is caused by the trilinear interpolation that distributes the gradient samples over contiguous SOHs. In other words, each gradient sample is added to each SOH weighted by 1-d, where d is the distance of the sample from the center of the corresponding sub-region [1]. Hence from the 16 angles at most 4 angles can meet the pairwise independence condition.

As mentioned above, four angles can be pairwise independent and only border angles can meet the equally likely condition, therefore the best choice are the corner angles: $\phi_1 = \theta_{11}$, $\phi_2 = \theta_{14}$, $\phi_3 = \theta_{41}$, and $\phi_4 = \theta_{44}$ which can be considered as new attributes of the SIFT feature.

3.2 Very Fast SIFT Features Matching

Assuming that two sets of extended SIFT features R and L, containing respectively r and l features, are given,. The number of possible matches is equal to $r \cdot l$.

Among these possible matches a small number of correct matches may exist.

To avoid the check of all possible matches, the introduced angles are exploited.

Assuming that the four angles are considered as components of 4-dimensional random vector of angles $\vec{\Phi} = \Phi_1, \Phi_2, \Phi_3, \Phi_4$. This vector is uniformly distributed in the 4-dimensional range [-180°, 180°]^4 due to its components meet the equally likely and pairwise independence conditions. For possible SIFT matches, a random vector difference can be constructed.

$$\Delta\vec{\Phi} = \vec{\Phi}_r - \vec{\Phi}_l \tag{4}$$

Generally, the difference between two independent uniformly-distributed random angles is uniformly distributed in [-180°, 180°] [12]. Hence, if $\vec{\Phi}_r$ and $\vec{\Phi}_l$ are independent, then $\Delta\vec{\Phi}$ is uniformly distributed in the 4-dimensional range [-180°, 180°]^4.

The behavior of $\Delta\vec{\Phi}$ varies differently according to the type of matches (correct and false matches): For false matches, each two corresponding angles are independent. Hence $\Delta\vec{\Phi}$ is uniformly-distributed. On the other hand, for correct matches, each two corresponding angles tend to be equal, since the features of correct matches tend to have the same SIFT-Ds. Therefore the $\Delta\vec{\Phi}$ tends to have PDF which is concentrated in narrow range around $\Delta\vec{\Phi} = \vec{0}$ (called range of correct matches w_{corr}).

Practically, 95% of correct matches have angle differences in the range [-36°, 36°]^4.

Consider that one of the sets of features R or L (for example R) is stored in a 4-dimensional table of size b^4, so that the $ijfg^{th}$ cell contains only SIFT features whose angles meet the following criteria: $\phi_1 \in [-\pi + (i-1)\cdot 2\pi/b, -\pi + i\cdot 2\pi/b)$, $\phi_2 \in [-\pi + (j-1)\cdot 2\pi/b, -\pi + j\cdot 2\pi/b)$, $\phi_3 \in [-\pi + (f-1)\cdot 2\pi/b, -\pi + f\cdot 2\pi/b)$ and $\phi_4 \in [-\pi + (g-1)\cdot 2\pi/b, -\pi + g\cdot 2\pi/b)$.

The number of features of the set R can be expressed as:

$$r = \sum_{i=1}^{b}\sum_{j=1}^{b}\sum_{f=1}^{b}\sum_{g=1}^{b} r_{ijfg} \tag{5}$$

Because of the uniformly distribution of $\vec{\Phi}_r$ in the 4-dimensional range [-180°,180°], the features are almost equally distributed into b^4 cells. Therefore, it can be asserted that the feature numbers of each cell are almost equal to each other.

$$\forall i,j,f,g \in \{1,2,...b\} : r_{ijfg} \cong r/b^4 \tag{6}$$

To exclude matching of features that have angle differences outside the range [-a°, a°][4], each SIFT feature of the set L is matched to its corresponding cell and to n neighbour cells to the left and n neighbour cells to the right side for each dimension. In this case the matching time is proportional to the term:

$$T = l \cdot \sum_{o=i-n}^{i+n}\sum_{p=j-n}^{j+n}\sum_{s=f-n}^{f+n}\sum_{t=g-n}^{g+n} r_{opst} = l \cdot r \cdot (2n+1/b)^4 \tag{7}$$

Therefore, the achieved speedup factor with respect to exhaustive search is equal to:

$$SF = (b/2n+1)^4 \tag{8}$$

The relation between n, a and b is as follows:

$$(2n+1)\cdot 360°/b = 2a \tag{9}$$

Substituting of (9) into (8) yields:

$$SF = (360/2a)^4 \tag{10}$$

To exclude matching of features that have angle differences outside the range [-36°,36°][4] we chose n=1 and b=15, then the matching is speeded up by a factor of 625. When this modification of original SIFT feature matching is combined with the split SIFT features matching explained in [11], the obtained speedup factor reaches 1250 without losing a noticeable portion of correct matches.

4 Experimental Results

The proposed method (VF-SIFT) was tested using a standard image dataset [13] and real-world stereo images. The used image dataset consists of about 500 images of 34 different scenes (some examples are shown in Fig. 6a and 6b). Real-world stereo

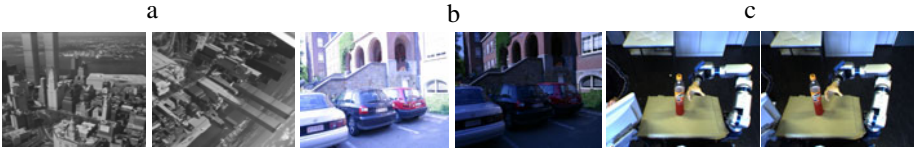

Fig. 6. Examples of the standard dataset (a,b) and real world stereo images (c)

images was captured using robotic vision system (A Bumblebee 2 stereo camera with the resolution of. 1024X768 pixels), an example is shown in Fig. 6c.

In order to evaluate the effectiveness of the proposed method, its performance was compared with the performances of two algorithms for ANN (Hierarchical K-Means Tree (HKMT) and Randomized KD-Trees (RKDTs)) [10]. Comparisons were performed using the Fast Library for Approximate Nearest Neighbors (FLANN) [14]. For all algorithms, the matching process is run under different precision degrees making trade off between matching speedup and matching accuracy. The precision degree is defined as the ratio between the number of correct matches returned using the considered algorithm and using the exhaustive search, whereas the speedup factor is defined as the ratio between the exhaustive matching time and the matching time for the corresponding algorithm.

For both ANN algorithms, the precision is adjusted by the number of nodes to be examined, whereas for the proposed VF-SIFT method, the precision is determined by adjusting the width of the range of correct matches w_{corr}.

To evaluate the proposed method two experiments were run. In the first experiment image to image matching was studied. SIFT features were extracted from 100 stereo image pairs and then each two corresponding images were matched using HKMT, RKDTs and VF-SIFT, under different degrees of precision. The experimental results are shown in Figure 7a. The second experiment was carried out on the images of the dataset [13] to study matching image against a database of images. SIFT features extracted from 10 query images are matched against database of 100000 SIFT features using all three considered algorithms, with different degrees of precision. The experimental results are shown in Figure 7b. As can be seen from Figure 7, VF-SIFT extremely outperforms the two other considered algorithms in speeding up of feature matching for all precision degrees. For precision around 95%, VF-SIFT gets a speedup factor of about 1250. For the lower precision degree speedup factor is much higher. Through comparison between Fig. 7a and Fig. 7b, it can be seen that the

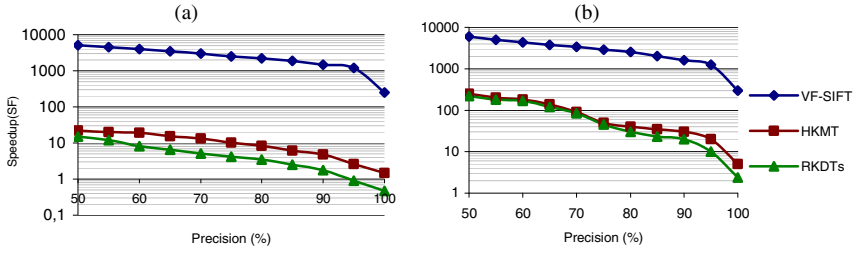

Fig. 7. Trade-off between matching speedup (SF) and matching precision

proposed method performs similarly for both cases of image matching (image to image and image against database of images), whereas ANN algorithms are more suitable for matching image against database of images.

6 Conclusion

In this paper, a new method for fast SIFT feature matching is proposed. The idea behind is to extend a SIFT feature by 4 pairwise independent angles, which are invariant to rotation, scale and illumination changes. During extraction phase, SIFT features are classified based on their angles into different clusters. Thus in matching phase, only SIFT features that belong to clusters where correct matches may be expected are compared. The proposed method was tested on real-world stereo images from a robotic application and standard dataset images. The experimental results show that the feature matching can be speeded up by 1250 times with respect to exhaustive search without lose of accuracy.

References

1. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision 60(2), 91–110 (2004)
2. Ke, Y., Sukthankar, R.: PCA-sift: A more distinctive representation for local image descriptors. In: Int. Conf. on Computer Vision and Pattern Recognition, pp. 506–513 (2004)
3. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on pattern analysis and machine intelligence 27(10) (2005)
4. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding 110(3), 346–359 (2008)
5. Sinha, S.N., Frahm, J.-M., Pollefeys, M., Genc, Y.: GPU-based video feature tracking and matching. Technical report, Department of Computer Science, UNC Chapel Hill (2006)
6. Heymann, S., Miller, K., Smolic, A., Froehlich, B., Wiegand, T.: SIFT implementation and optimization for general-purpose gpu. In: WSCG 2007 (January 2007)
7. Chariot, A., Keriven, R.: GPU-boosted online image matching. In: 19th Int. Conf. on Pattern Recognition, Tampa, Florida, USA (2008)
8. Se, S., Ng, H., Jasiobedzki, P., Moyung, T.: Vision based modeling and localization for planetary exploration rovers. In: Proceedings of International Astronautical Congress (2004)
9. Firedman, J.H., Bentley, J.L., Finkel, R.A.: An algorithm for finding best matches in logarithmic expected time. ACM Transactions Mathematical Software, 209–226 (1977)
10. Muja, M., Lowe, D.G.: Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In: Int. Conf. on Computer Vision Theory and Applications (2009)
11. Alhwarin, F., Ristic Durant, D., Gräser, A.: Speeded up image matching using split and extended SIFT features. In: Conf. on Computer Vision Theory and Applications (2010)
12. Simon, M.K., Shihabi, M.M., Moon, T.: Optimum Detection of Tones Transmitted by a Spacecrft, TDA PR 42-123, pp.69–98 (1995)
13. Image database: `http://lear.inrialpes.fr/people/Mikolajczyk/Database/index.html`
14. FLANN: `http://people.cs.ubc.ca/~mariusm/index.php/FLANN/FLANN`
15. Fleischer, K.: Two tests of pseudo random number generators for independence and uniform distribution. Journal of statistical computation and simulation 52, 311–322 (1995)