Improved SIFT-Features Matching for Object Recognition

Faraj Alhwarin, Chao Wang, Danijela Ristić-Durrant, Axel Gräser
Institute of Automation, University of Bremen, FB1 / NW1
Otto-Hahn-Allee 1
D-28359 Bremen
Emails: {alhwarin,wang,ristic,ag}@iat.uni-bremen.de

Abstract:

The SIFT algorithm (Scale Invariant Feature Transform) proposed by Lowe [1] is an approach for extracting distinctive invariant features from images. It has been successfully applied to a variety of computer vision problems based on feature matching including object recognition, pose estimation, image retrieval and many others. However, in real-world applications there is still a need for improvement of the algorithm's robustness with respect to the correct matching of SIFT features. In this paper, an improvement of the original SIFT algorithm providing more reliable feature matching for the purpose of object recognition is proposed. The main idea is to divide the features extracted from both the test and the model object image into several sub-collections before they are matched. The features are divided into several sub-collections considering the features arising from different octaves, that is from different frequency domains.

To evaluate the performance of the proposed approach, it was applied to real images acquired with the stereo camera system of the rehabilitation robotic system FRIEND II. The experimental results show an increase in the number of correct features matched and, at the same time, a decrease in the number of outliers in comparison with the original SIFT algorithm. Compared with the original SIFT algorithm, a 40% reduction in processing time was achieved for the matching of the stereo images.

Keywords: SIFT algorithm, Improved SIFT, Images matching, Object recognition

1. INTRODUCTION

The matching of images in order to establish a measure of their similarity is a key problem in many computer vision tasks. Robot localization and navigation, object recognition, building panoramas and image registration represent just a small sample among a large number of possible applications. In this paper, the emphasis is on object recognition.

In general the existing object recognition algorithms can be classified into two categories: global and local features based algorithms. Global features based algorithms aim at recognizing an object as a whole. To achieve this, after the acquisition, the test object image is sequentially pre-processed and segmented. Then, the global features are extracted and finally statistical features classification techniques are used. This class of algorithm is particularly suitable for recognition of homogeneous (textureless) objects, which can be easily segmented from the image background. Features such as Hu moments [5] or the eigenvectors of the covariance matrix of the segmented object [6] can be used as global features. Global features based algorithms are simple and fast, but there are limitations in the reliability of object recognition under changes in illumination and object pose. In contrast to this, local features based algorithms are more suitable for textured objects and are more robust with respect to variations in pose and illumination. In [7] the advantages of local over global features are demonstrated.

Local features based algorithms focus mainly on the so-called keypoints. In this context, the general scheme for object recognition usually involves three important stages: The first one is the extraction of salient feature points (for example corners) from both test and model object images. The second stage is the construction of regions around the salient points using mechanisms that aim to keep the regions characteristics insensitive to viewpoint and illumination changes. The final stage is the matching between test and model images based on extracted features.

The development of image matching by using a set of local keypoints can be traced back to the work of Moravec [8]. He defined the concept of "points of interest" as being distinct regions in images that can be used to find matching regions in consecutive image frames. The Moravec operator was further developed by C. Harris and M. Stephens [9] who made it more repeatable under small image variations and near edges. Schmid and Mohr [10] used Harris corners to show that invariant local features matching could be extended to the general image recognition problem. They used a rotationally invariant descriptor for the local image regions in order to allow feature matching under arbitrary orientation variations. Although it is rotational invariant, the Harris corner detector is however very sensitive to changes in image scale so it does not provide a good basis for matching images of different sizes. Lowe [1, 2, 3] overcome such problems by detecting the points of interest over the image and its scales through the location of the local extrema in a pyramidal Difference of Gaussians (DOG). The Lowe's descriptor, which is based on selecting stable features in the scale space, is named the Scale Invariant Feature Transform (SIFT). Mikolajczyk and Schmid [12] experimentally compared the performances of several currently used local descriptors and they found that the SIFT descriptors to be the most effective, as they yielded the best matching results. SIFT improving techniques developed recently targeted minimization of the computational time [16, 17, 18], while limited research aiming at improving the accuracy has been done. The work presented in this paper demonstrates increased matching process performance robustness with no additional time costs. Special cases, similar scaled features, consume even less time.

The high effectiveness of the SIFT descriptor has motivated the authors of this paper to use it for object recognition in service robotics applications [5]. Through the performed experiments it was found that SIFT keypoints features are highly distinctive and invariant to image scale and rotation providing correct matching in images subject to noise, viewpoint and illumination changes. However, it was also found that sometimes the number of correct matches is insufficient for object recognition, particularly when the target object, or part of it, appears very small in the test image with respect to its appearance in model image. In this paper, a new strategy to enhance the number of correct matches is proposed. The main idea is to determine the scale factor of the target object in the test image using a suitable mechanism and to perform the matching process under the constraint introduced by the scale factor, as described in Section 4.

The rest of the paper is organized as follows. Section 2 presents the SIFT algorithm. The SIFT-feature matching strategy is presented in Section 3. In Section 4 the proposed modification of the original SIFT algorithm is described and contributions are discussed. A performance evaluation of the proposed technique through the comparison of its experimental results with the results obtained using the original SIFT algorithm is given in Section 5.

2. SIFT ALGORITHM

The scale invariant feature transform (SIFT) algorithm, developed by Lowe [1,2,3], is an algorithm for image features generation which are invariant to image translation, scaling, rotation and partially invariant to illumination changes and affine projection. Calculation of SIFT image features is performed through the four consecutive steps which are briefly described in the following:

- *scale-space local extrema detection* - the features locations are determined as the local extrema of Difference of Gaussians (DOG pyramid). To build the DOG pyramid the input image is convolved iteratively with a Gaussian kernel of $\sigma = 1.6$. The last convolved image is down-sampled in each image direction by factor of 2, and the convolving process is repeated. This procedure is repeated as long as the down-sampling is possible. Each collection of images of the same size is called an octave. All octaves build together the so-called Gaussian pyramid, which is represented by a 3D function $L(x, y, \sigma)$. The DOG pyramid $D(x, y, \sigma)$ is computed from the difference of each two nearby images in Gaussian pyramid. The local extrema (maxima or minima) of DOG function are detected by comparing each pixel with its 26 neighbours in the scale-space (8 neighbours in the same scale, 9 corresponding neighbours in the scale above and 9 in the scale below). The search for for extrema excludes the first and the last image in each octave because they do not have a scale above and a scale below respectively. To increase the number of extracted features the input image is doubled before it is treated by SIFT algorithm, which however increases the computational time significantly. In the method presented in this paper, the image doubling is avoided but the search for extrema is performed over the whole octave including the first and thelast scale. In this case the pixel comparing is carried out only with available neighbours.

- *keypoint localization* - the detected local extrema are good candidates for keypoints. However, they need to be exactly localized by fitting a 3D quadratic function to the scale-space local sample point. The quadratic function is computed using a second order Taylor expansion having the origin at the sample point. Then, local extrema with low contrast and such that correspond to edges are discarded because they are sensitive to noise.
- orientation assignment - once the SIFT-feature location is determined, a main orientation is assigned to each feature based on local image gradients. For each pixel of the region around the feature location the gradient magnitude and orientation are computed respectively as:

$$m(x,y)=\sqrt{\left(L(x+1,y,\sigma)-L(x-1,y,\sigma)\right)^2+\left(L(x,y+1,\sigma)-L(x,y-1,\sigma)\right)^2}$$
$$\theta(x,y)=\arctan\left(\left(L(x,y+1,\sigma)-L(x,y-1,\sigma)\right)/\left(L(x+1,y,\sigma)-L(x-1,y,\sigma)\right)\right) \quad (1)$$

  The gradient magnitudes are weighted by a Gaussian window whose size depends on the feature octave. The weighted gradient magnitudes are used to establish an orientation histogram, which has 36 bins covering the 360 degree range of orientations. The highest orientation histogram peak and peaks with amplitudes greater than 80% of the highest peak are used to create a keypoint with this orientation. Therefore, there will be multiple keypoints created at the same location but with different orientations.
- *keypoint descriptor* - the region around a keypoint is divided into 4X4 boxes. The gradient magnitudes and orientations within each box are computed and weighted by appropriate Gaussian window, and the coordinate of each pixel and its gradient orientation are rotated relative to the keypoints orientation. Then, for each box an 8 bins orientation histogram is established. From the 16 obtained orientation histograms, a 128 dimensional vector (SIFT-descriptor) is built. This descriptor is orientation invariant, because it is calculated relative to the main orientation. Finally, to achieve the invariance against change in illumination, the descriptor is normalized to unit length.

3. SIFT FEATURES MATCHING

From the algorithm description given in Section 2 it is evident that in general, the SIFT-algorithm can be understood as a local image operator which takes an input image and transforms it into a collection of local features. To use the SIFT operator for object recognition purposes, it is applied on two object images, a model and a test image, as shown in Figure 1 for the case of a food package. As shown, the model object image is an image of the object alone taken in predefined conditions, while the test image is an image of the object together with its environment.

To find corresponding features between the two images, which will lead to object recognition, different feature matching approaches can be used. According to the Nearest Neighbourhood procedure for each F_1^i feature in the model image feature set the corresponding feature F_2^j must be looked for in the test image feature set. The corresponding feature is one with the smallest Euclidean distance to the feature F_1^i. A pair of corresponding features $\left(F_1^i, F_2^j\right)$ is called a match $M\left(F_1^i, F_2^j\right)$.

To determine whether this match is positive or negative, a threshold can be used.

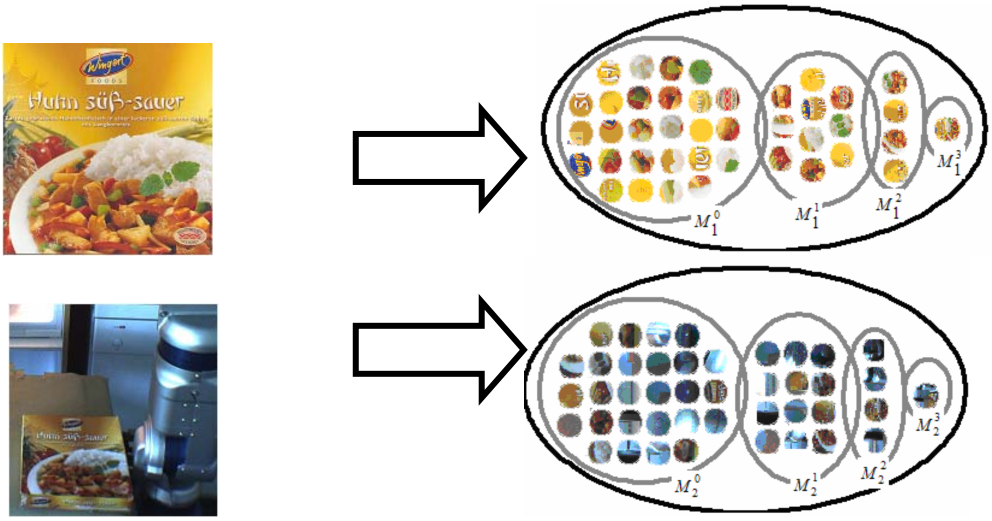


FIGURE 1: Transformation of both model and test image into two collections of SIFT features; division of the features sets into subsets according to the octave of each feature proposed in this paper.

If the Euclidean distance between the two features F_1^i and F_2^j is below a certain threshold, the match $M\left(F_1^i, F_2^j\right)$ is labelled as positive. Because of the change in the projection of the target object from scene to scene, the global threshold for the distance to the next feature is not useful. Lowe [1] proposed the using of the ratio between the Euclidean distance to the nearest and the second nearest neighbours as a threshold τ. Under the condition that the object does not contain repeating patterns, one suitable match is expected and the Euclidean distance to the nearest neighbour is significantly smaller than the Euclidean distance to the second nearest neighbour. If no match is correct, all distances have a similar, small difference from each other. A match is selected as positive only if the distance to the nearest neighbour is 0.8 times larger than that from the second nearest one. Among positive and negative matches, correct as well as false matches can be found. Lowe claims [3] that the threshold of 0.8 provides 95% of correct matches as positive and 90% of false matches as negative. The total amount of the correct positive matches must be large enough to provide reliable object recognition. In the following an improvement to the feature matching robustness of the SIFT algorithm with respect to the number of correct positive matches is presented.

4. AN IMPROVEMENT OF FEATURE MATCHING ROBUSTNESS IN THE SIFT ALGORITHM

As discussed in previous section, the target object in the test image is part of a cluttered scene. In a real-world application the appearance of the target object in the test image, its position, scale and orientation, are not known *a priori*. Assuming that the target object is not deformed, all features of the target image can be considered as being affected with constant scaling and rotational factors. This can be used to optimize the SIFT-feature matching phase where the outliers' rejection stage of the original SIFT-method is integrated into the SIFT-feature matching stage.

4.1 Scaling factor calculation

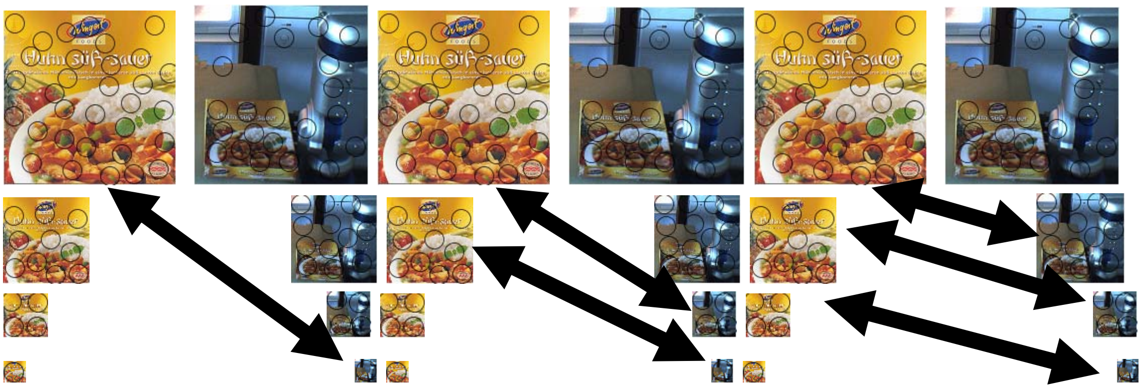
As mentioned in Section 2, using the SIFT-operator, the two object images (model and test) are transformed into two SIFT-image feature sets. These two feature sets are divided into subsets according to the octaves in which the feature arise. Hence, there is a separate subset for each image octave as shown in Figures 1 and 2. To carry out the proposed new strategy of SIFT-features matching, the features subsets obtained are arranged so that a subset of the model image feature set is aligned with an appropriate subset of the test image feature set. The process of alignment of the model image subsets with the test image subsets is indicated with arrows in Figure 2. The alignment process is performed through the (*n*+*m*-1) steps, where *n* and *m* are the total number of octaves (subsets) corresponding to the model and test image respectively. For each step all pairs of aligned subsets must have the same ratio ν defined as: $\nu = 2^{o_1}/2^{o_2}$, where o_1 and o_2 are the octaves of the model image subset and the test image subset respectively. At every step, the total number of positive matches is determined for each aligned subsets pair. The total number of positive matches within each step is indexed using the appropriate shift index $k = o_2 - o_1$. Shift index can be negative (Figures 2a, 2b and 2c), positive (2e, 2f and 2g) or equal to zero (Figure 2d). The highest number of positive matches achieved determines the optimal shift index k_{opt} and consequently the scale factor $S = 2^{k_{opt}}$.

In order to realize the proposed procedure mathematically, a quality-integer function $F(x)$ is defined as:

$$F(x) = \begin{cases} \sum_{j=0}^{j=x} Z(M_1^{n-1-x+j}, M_2^{j}) \ldots if\ (x < n) \\ \sum_{j=0}^{j=n-1} Z(M_1^{x-n-1+j}, M_2^{j}) \ldots if\ (n < x < m) \\ \sum_{j=0}^{j=m+n-2-x} Z(M_1^{x-n+j}, M_2^{x-m+1+j}) \ldots if\ (x \geq m) \end{cases} \qquad m \geq n \quad and \quad x \in [0, m+n-2] \tag{2}$$

where $Z\left(M_1^i, M_2^j\right)$ is the number of positive matches between the *i-th* subset of the model image feature set M_1^i and the *j-th* subset of the test image feature set M_2^j , and x is the modified shift index $x = \operatorname{int}\left(k + \left(\frac{n+m-1}{2}\right)\right)$ introduced for the sake of simplicity of equation 2.

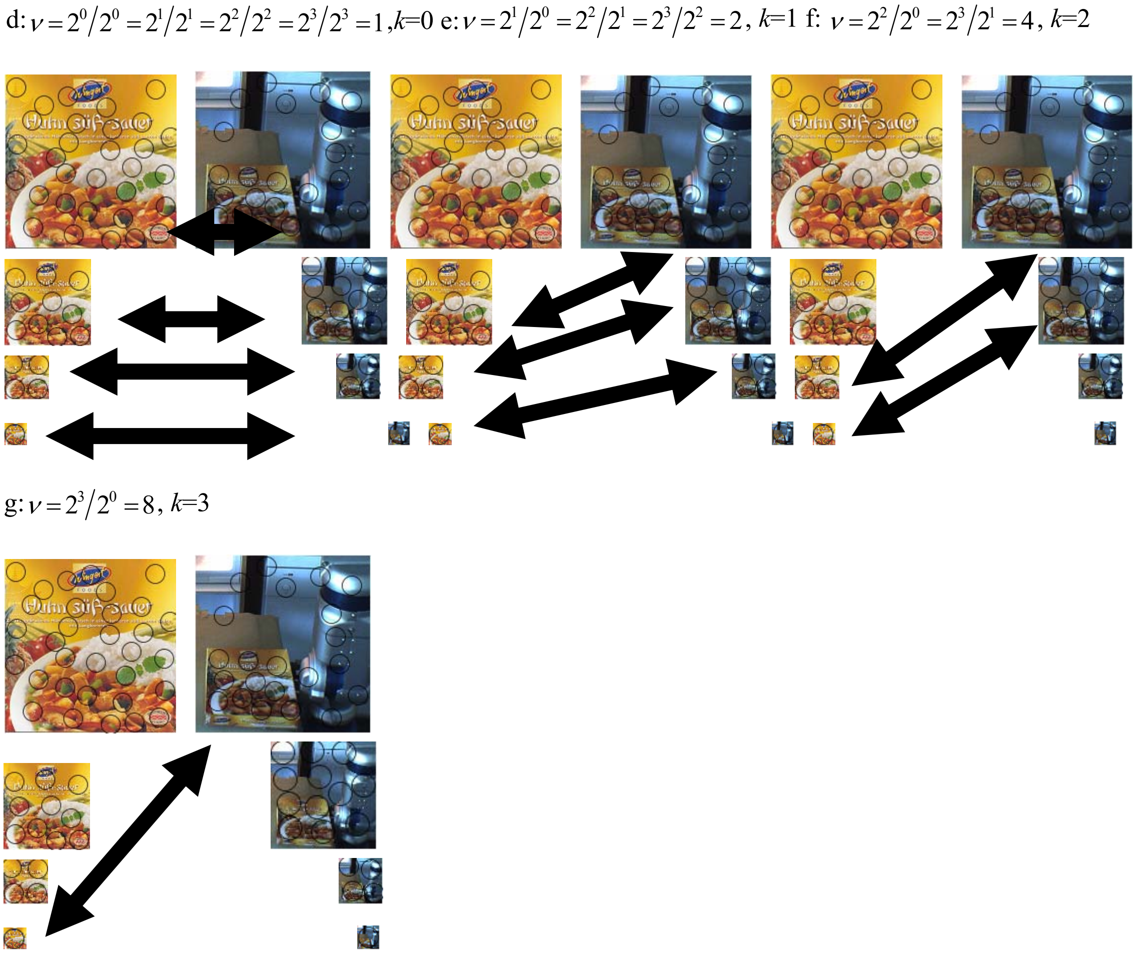
The diagram showing the distribution of $F(k)$ over the range of the shift index k for the example shown in Figure 2 is presented in Figure 3.



a: $\nu = 2^0/2^3 = 1/8$, k=-3

b: $\nu = 2^0/2^2 = 2^1/2^3 = 1/4$ k=-2

c: $\nu = 2^0/2^1 = 2^1/2^2 = 2^2/2^3 = 1/2$, k=-1



d: $\nu = 2^0/2^0 = 2^1/2^1 = 2^2/2^2 = 2^3/2^3 = 1, k=0$ e: $\nu = 2^1/2^0 = 2^2/2^1 = 2^3/2^2 = 2, k=1$ f: $\nu = 2^2/2^0 = 2^3/2^1 = 4, k=2$

g: $\nu = 2^3/2^0 = 8, k=3$

FIGURE 2: Steps of the procedure for scale factor calculation.

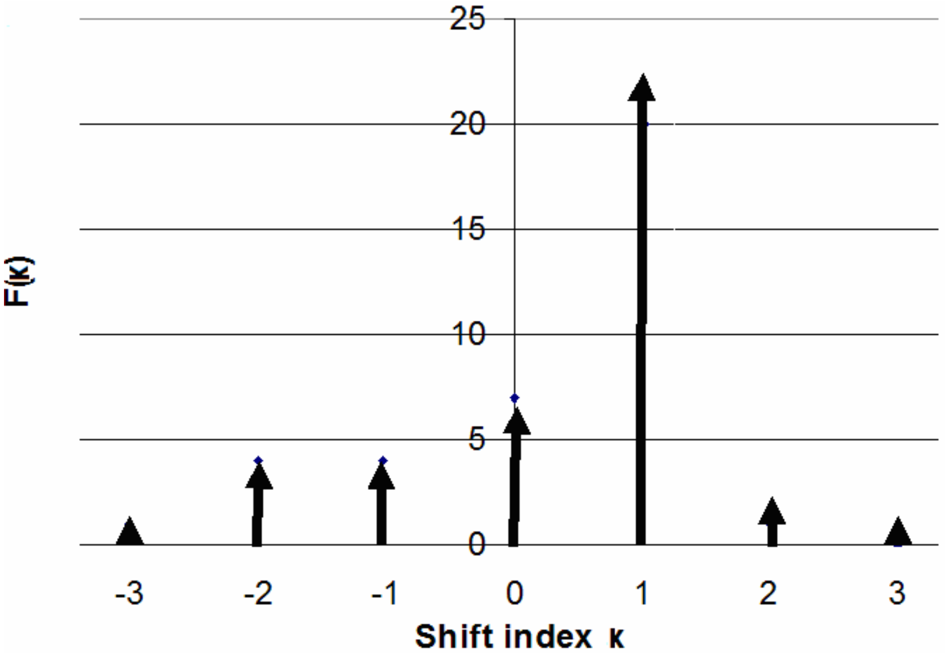


FIGURE 3: The quality-integer function $F(k)$

As evident from Figure 3, the quality-integer function $F(k)$ reaches its maximum $F(k_{opt}) = \max(F(k))$ for the shift index $k = k_{opt} = 1$ which corresponds to the scale factor $S = 2$. The optimal shift index defines a “domain of correct matches”. All matches outside this domain, including positive matches, are excluded. The positive matches from the domain of correct matches are used to determine the affine transformation (rotation matrix, and translation vector) between the two feature sets, using RANSAC method [15]. Once the transformation is calculated, every match, either positive or negative, within the domain of correct matches is examined whether it meets the already calculated transformation. If the match fulfils the transformation, it is labelled as a correct, otherwise as a false match.

4.2 Retrieval of the correct matches

Among all found matches it can happen that a lot of correct matches exceed Lowe's threshold τ. In order to retrieve these correct matches, the ratio between the Euclidean distance to the nearest and the second nearest feature neighbour must be reduced. This can be done either by reducing the smallest distance $\mathrm{D}_1(F_1^i, F_2^{j_0})$ or by increasing the next smallest distance $D_2(F_1^i, F_2^{j_1})$. In practice, the first alternative is impossible while the enlargement of next smallest distance can be achieved by limiting the search area for both the nearest and next nearest feature to the feature F_1^i within a specified domain. For a better explanation of this idea, suppose that a feature F_1^i from the model image feature set is correctly assigned to the feature $F_2^{j_0}$ from the test image feature set. Also, suppose that $F_2^{j_1}$ is the second nearest feature to the F_1^i while $F_2^{j_2}$ is the second nearest feature to it when the search is limited only to the octave in which the $F_2^{j_0}$ is found.

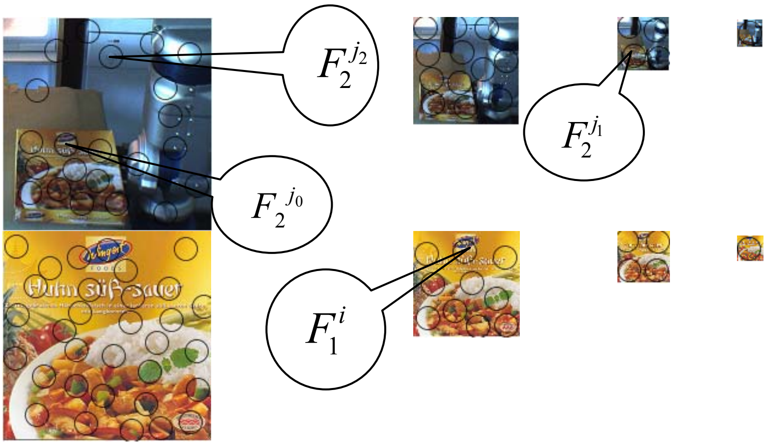


FIGURE 4: Saving the correct matches that may exceed Lowe's threshold.

Since $\mathrm{D}_2(F_1^i, F_2^{j_1}) \le \mathrm{D}_3(F_1^i, F_2^{j_2})$ always holds the following:
$\mathrm{D}_1(F_1^i, F_2^{j_0}) / \mathrm{D}_2(F_1^i, F_2^{j_1}) \ge \mathrm{D}_1(F_1^i, F_2^{j_0}) / \mathrm{D}_3(F_1^i, F_2^{j_2})$ is obtained.
Thus, by reducing the search area it is possible to decrease the ratio related to the feature F_1^i and make it less than threshold τ. In this way the number of correct matches is increased.

4.3 Complexity and cost of time

An additional result of the research presented in this paper is consideration of the improvement of the original SIFT algorithm with respect to the processing time. As first, it can be shown that the original SIFT procedure and the procedure developed in this paper complete the matching procedure in the same time.

Assuming that the number of features in the model object image is: $h = h_0 + h_1 + ... + h_{n-1}$, and in the test image: $l = l_0 + l_1 + ... + l_{m-1}$, where n and m are the total number of octaves corresponding to the model and test image respectively.

Thus, the complexity of original SIFT-matching procedure is proportional to the product $P_1 = l.h$ .The complexity of the proposed approach, which can be seen from Figure 2, is proportional to the following sum of the products:

$$\begin{array}{l} P_2 = l_{m-1} \cdot h_0 + \\ l_{m-1} \cdot h_1 + l_{m-2} \cdot h_0 + \\ . \\ . \\ l_{m-1} \cdot h_{n-1} + l_{m-2} \cdot h_{n-2} + l_{m-3} \cdot h_{n-3} + ...+ l_1 \cdot h_1 + l_0 \cdot h_0 + \\ \qquad l_{m-2} \cdot h_{n-1} + l_{m-3} \cdot h_{n-2} + ...+ l_1 \cdot h_2 + l_0 \cdot h_1 + \\ . \\ . \\ \qquad\qquad\qquad l_1 \cdot h_{n-1} + l_0 \cdot h_{n-2} + \\ \qquad\qquad\qquad\quad + l_0 \cdot h_{n-1} = \sum_{i=0}^{i=n-1} h_i \sum_{j=0}^{j=m-1} l_j \end{array} \tag{3}$$

Substituting $\sum_{j=0}^{j=m-1} l_j = l, \ \sum_{i=0}^{i=m-1} h_i = h$ in (3) one obtains:

$$P_2 = \sum_{i=0}^{i=n-1} h_i \sum_{j=0}^{j=m-1} l_j = l. \sum_{i=0}^{i=n-1} h_i = l.h \tag{4}$$

which is equal to the product P_1 corresponding to the complexity of the original SIFT matching procedure.

The above condition represents the complexity of the proposed matchin procedure when no *a-priory* information about the scaling factor of corresponding features is available, that is when the procedure consist of all (*n*+*m*-1) steps as explained in Section 4.1. However, in some applications the complexity is reduced. For example,, if the two images to be matched are images of stereo camera system with small baseline, all corresponding features should have the same scale. Hence the proposed matching procedure is carried out with only one step corresponding to the shift index $k = 0$. In this case, the complexity of the proposed procedures is reduced, since it is proportional to the sum of the following products:

$$P_3 = l_0.h_0 + l_1.h_1 + ... + l_{n-1}.h_{n-1} \tag{5}$$

In order to determine the amount of reduced processing time in comparison to original SIFT procedure, it is assumed that the number of extracted features in the lower octave with respect to the higher octave is decreased 4 times due to the down-sampling by the factor of 2 in both image directions. Hence, it is assumed that:

$$l_{i-1} \approx 4.l_i \, , h_{i-1} \approx 4.h_i \ . \tag{6}$$

Substituting (6) in both products P_2 and P_3, defined with (4) and (5) respectively, one obtains:

$$P_3 = l_0.h_0 + (1/4)^2 .l_0.h_0 + (1/4)^4 .l_0.h_0 .... + (1/4)^{2(n-1)} .l_0.h_0$$

$$P_3 = l_0.h_0.\sum_{i=0}^{n-1} (1/4)^{2i} \qquad (7)$$

$$P_2 = l.h = (l_0 + l_1 + ... + l_{n-1}).(h_0 + h_1 + ... + h_{n-1})$$

$$P_2 = (l_0 + (1/4).l_0 + ... + (1/4)^{(n-1)}.l_0).(h_0 + (1/4).h_0 + ... + (1/4)^{(n-1)}.h_0) \qquad (8)$$

$$P_2 = l_0.h_0.(\sum_{i=0}^{n-1} (1/4)^i)^2$$

From (7) and (8) the ratio P_2 / P_3 is given as:

$$P_2 / P_3 = l_0.h_0.\left(\sum_{i=0}^{n-1} (1/4)^i\right)^2 \Big/ l_0.h_0.\sum_{i=0}^{n-1} (1/4)^{2i} \approx \left(\sum_{i=0}^{\infty} (1/4)^i\right)^2 \Big/ \sum_{i=0}^{\infty} (1/4)^{2i} \qquad (9)$$

It is known that $\sum_{i=0}^{\infty} x^i = 1/(1-x) \; if \; |x| < 1$ (10)

Substituting (10), the ratio (9) becomes:

$$P_2 / P_3 \approx \left(\sum_{i=0}^{\infty} (1/4)^i\right)^2 \Big/ \sum_{i=0}^{\infty} (1/4)^{2i} = \left(\frac{1}{1-1/4}\right)^2 \Big/ \left(\frac{1}{1-(1/4)^2}\right) = \frac{15}{9} = 1.67$$

Hence, the matching time cost in the case of matching stereo images is reduced 1.67 times in comparison to the original SIFT method.

5 RESULTS

In this section a performance evaluation of the proposed improvement of the Lowe's SIFT feature matching algorithm is presented. Since the goal is to achieve a trade-off between the increasing the number of correct matches and minimizing the number of false matches for an object image pair consisting of test and model object images, the performance of the proposed method is evaluated using the popular Recall-Precision metric [14]. As mentioned in Section 3, two SIFT features F_1^i and F_2^j are matched when the SIFT descriptor of the feature F_2^j has the smallest distance to the descriptor of feature F_1^i among distances corresponding to all other extracted features. If the ratio between the Euclidian distances to the nearest neighbour and to the second nearest neighbour is below a threshold τ, the match is labelled as positive, otherwise as negative. Among positive and negative labelled matches, correct as well as false matches can be found. Thus there are four different possible combinations through the following confusion matrix:

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

TABLE 1: The confusion Matrix

During the matching of an image pair the elements of the confusion matrix are counted. The value of τ is varied to obtain the Recall versus 1-Precision curve, with which the result are presented.

Recall and 1- Precision are calculated based on the following definitions [14]:

$$Recall = TP/(TP+FN), 1 - \Pr ecision = FP/(TP+FP) . \qquad (11)$$

The algorithms were tested by matching real images of the scenes from working scenarios of the robotic system FRIEND II. containing different target objects to be recognized (bottles, packages, and etc), acquired with the stereo camera system of FRIEND II robot.

Two main types of experiments were run to discuss the difference between the original SIFT and the proposed optimized SIFT matching algorithm. In the first experiment, the model images of two different objects, a bottle of the "mezzo mix" drink and a coffee filters package, were matched with the corresponding test object images using the original and proposed improved SIFT matching algorithm. The experimental results are illustrated in Figure 6. As evident, the appearance of the target objects in the test images is different from their appearance in model images due to different conditions such as illumination during the image acquisition, viewpoint, partial occlusion etc. the advantage of the proposed matching technique over the original SIFT matching technique is evident from Figure 6. Beside the examination of the results illustration in Figure 6, performance evaluation can be done by examination of the recall versus 1-precision curve shown in Figure 5. the curves are obtained by varying the threshold from 0.5 till 1.0.
In the second experiment images of a scene from the robot FRIEND II environment, captured by the robot stereo camera system, were matched to evaluate the optimizing of the computational matching time of the proposed approach with respect to the original SIFT. The experimental results are given in the Table 2. The experimentally obtained ratios of the processing time of original SIFT and processing time of proposed technique slightly differ from the ratio derived in section 4 because the assumption assumed the proof does not necessarily hold. The matching process was carried out using a Pentium IV 1GH processor with, images of size 1024X768 pixels.

Key-points number in stereo images		Original SIFT matching		Improved SIFT matching	
left	right	Matching time (sec)	Number of inliers	Matching time (sec)	Number of inliers
217	229	0.140	111	0.025	133
777	640	0.790	284	0.230	325
3014	2233	10.760	605	4.950	683
6871	6376	69.810	751	47.790	856

TABLE 2: Comparison of the stereo images matching time.

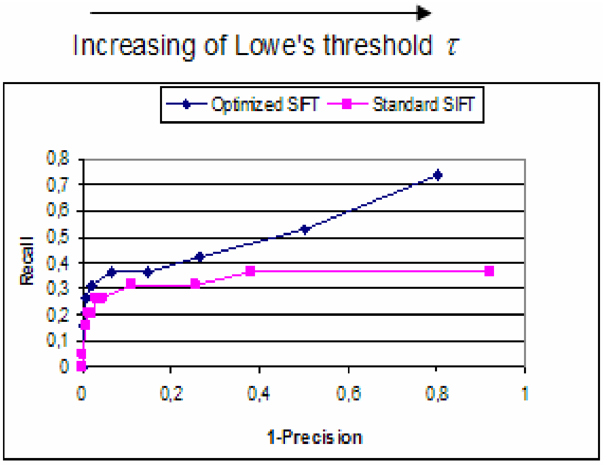


FIGURE 5: Recall versus 1-Precision curves for the original and optimized SIFT matching methods

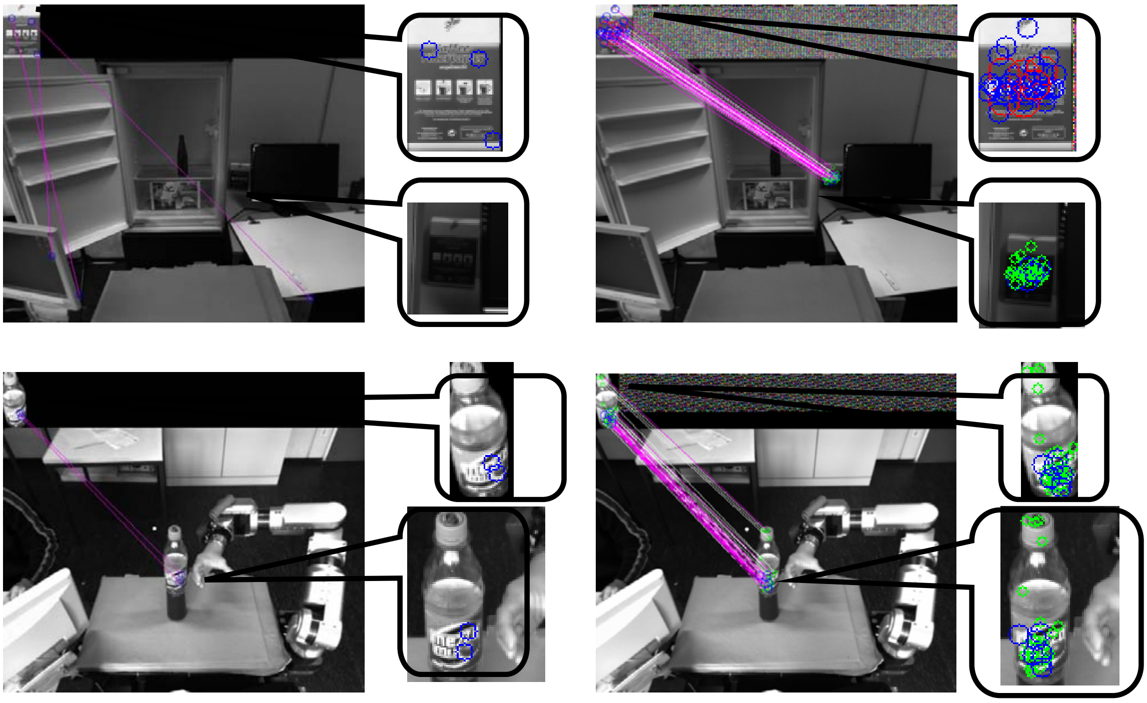


FIGURE 6: (left column) matching result with original SIFT, (right column) matching result with improved SIFT.

6 CONCLUSIONS

In this paper an improvement of the original SIFT-algorithm developed by Lowe was proposed. This improvement corresponds to enhancement of feature matching robustness, so the number of correct SIFT features matches is significantly increased while nearly all outliers are discarded. Also the matching time cost for the case of extracted features into subsets corresponding to different octaves. The new proposed approach was tested using real images acquired with the stereo camera system of FRIEND II robotic system. The presented experimental results show the effectiveness of the proposed approach.

REFERENCES

[1] David G. Lowe. : Object recognition from local scale-invariant features, International Conference on Computer Vision, Corfu, Greece (September 1999), pp. 1150-1157.

[2] David G. Lowe. : Local feature view clustering for 3D object recognition, IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii (December 2001), pp. 682-688.

[3] David G. Lowe. : Distinctive image features from scale-invariant key-points, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[4] T. Brox and R. Deriche. : Active Unsupervised Texture Segmentation on a Diffusion Based Feature Space, IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2003, Madison, Wisconsin, USA.

[5] Sai K. Vuppala, Sorin M. Grigorescu, Danijela Ristic, and Axel Gräser. : Robust color Object Recognition for a Service robotic Task in the System FRIEND II, 10th International Conference on Rehabilitation Robotics - ICORR'07, 2007.

[6] Yongjin Lee, Kyunghee Lee, and Sungbum Pan. :Local and Global Feature Extraction for Face Recognition, Springer-Verlag Berlin Heidelberg 2005.

[7] Ke, Y., Suthankar, R., Huston L.: Efficient Near-Duplicate Detection and Sub image Retrieval, ACM International Conference on Multimedia. (2004) 869–876.

[8] H. P. Moravec. : Towards Automatic Visual Obstacle Avoidance, Proc. 5th International Joint Conference on Artificial Intelligence, pp. 584, 1977.
[9] C. Harris and M. Stephens. : A Combined Corner and Edge Detector. Proc. Alvey Vision Conf., Univ. Manchester, pp. 147-151, 1988.
[10] C. Schmid, R. Mohr. : Local Greyvalue Invariants for Image Retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 1997.
[11] K . Mikolajczyk and C. Schmid. : Scale & Affine Invariant Interest Point Detectors, International Journal of Computer Vision 60(1), 63–86, 2004.
[12] K. Mikolajczyk and C. Schmid. : A Performance Evaluation of Local Descriptors, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 27, NO. 10, OCTOBER 2005.
[13] S. Lazebnik, C. Schmid, and J. Ponce. : Sparse Texture Representation Using Affine-Invariant Neighborhoods, Proc. Conf. Computer Vision and Pattern Recognition, pp. 319-324, 2003.
[14] J. Davis, M. Goadrich. : The Relationship Between Precision-Recall and ROC Curves, University of Wisconsin-Madison, 1210 West Dayton Street, Madison, WI, 53706 USA.
[15] M. A. Fischler and R. C. Bolles, ”Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography”, Communication Association and Computing Machine, 24(6), pp.381-395, 1981.
[16] H. Bay, T. Tuytelaars, L. Van Gool: SURF, Speeded Up Robust Features", Proceedings of the ninth European Conference on Computer Vision, May 2006.

[17] G. Michael, G. Helmut, B. Horst : Fast Approximated SIFT, Conference on Computer Vision, Hyderabad, India , Springer, LNCS 3851, pages 918-927,2006.

[18] Y. Ke, R. Sukthankar: PCA-SIFT: A more distinctive representation for local image descriptors. In: Proc. CVPR. Volume 2. (2004) 506–513
[19] F. Alhwarin: DAS TITEL, Proc. of 29th Colloquium of Automation, Salzhausen, Germany, 2008 (to be published)
[20] Grigorescu, S. M., Ristić-Durrant, D., Vuppala, S. K., Gräser, A.: Closed-Loop Control in Image Processing for Improvement of Object Recognition, 17th IFAC World Congress, 2008