A visual shape descriptor using sectors and shape context of contour lines

Shao-Hu Peng\textsuperscript{a}, Deok-Hwan Kim\textsuperscript{a,}\textsuperscript{*}, Seok-Lyong Lee\textsuperscript{b}, Chin-Wan Chung\textsuperscript{c}

\textsuperscript{a}Department of Electronic Engineering, Inha University, South Korea
\textsuperscript{b}School of Industrial and Management Engineering, Hankuk University of Foreign Studies, South Korea
\textsuperscript{c}Division of Computer Science, Department of EECS, KAIST, South Korea

Abstract

This paper describes a visual shape descriptor based on the sectors and shape context of contour lines to represent the image local features used for image matching. The proposed descriptor consists of two-component feature vectors. First, the local region is separated into sectors and their gradient magnitude and orientation values are extracted; a feature vector is then constructed from these values. Second, local shape features are obtained using the shape context of contour lines. Another feature vector is then constructed from these contour lines. The proposed approach calculates the local shape feature without needing to consider the edges. This can overcome the difficulty associated with textured images and images with ill-defined edges. The combination of two-component feature vectors makes the proposed descriptor more robust to image scale changes, illumination variations and noise. The proposed visual shape descriptor outperformed other descriptors in terms of the matching accuracy: 14.525\% better than SIFT, 21\% better than PCA-SIFT, 11.86\% better than GLOH, and 25.66\% better than the shape context.

1. Introduction

Image matching plays an important role in intelligent systems, such as human face recognition [8,19,21], intelligent cars [1,9] and unmanned air vehicles [2,17]. Some of the most important considerations in image matching are the robustness with respect to the scale image, illumination and noise. Interest region identification and descriptor computation have proven successful in overcoming these problems, and they are used widely in object recognition [4,6,13,18] and image retrieval [3,5,10,16,20] systems.

For image matching, the interest regions are first identified through the use of effective detectors, such as Harris-Laplace [16] and difference-of-Gaussian (DoG) [12]. Local descriptors are then calculated for the interest region to represent the image features. Finding effective local descriptors is very important not only for the matching accuracy but also for the matching efficiency. Several effective descriptors, such as the scale-invariant feature transform (SIFT) [12] and shape context [4], have been developed in recent years.

The SIFT proposed by Lowe is designed for robust image feature detection, and has been proven to be invariant with respect to scale, noise and affine transformations. Several new descriptors based on SIFT have been proposed to improve the matching performance, for example PCA-SIFT [7], GLOH [15], etc. These were evaluated by Mikolajczyk and Schmid, who demonstrated that these new descriptors perform well under scale transformations, illumination variation and noise.

* Corresponding author. Address: #814 High-Tech, Inha University, Yonghyun-dong, Nam-gu Incheon 402 751, South Korea. Tel.: +82 32 860 7424; fax: +82 32 868 3654.
E-mail addresses: pengshaohu@hotmail.com (S.-H. Peng), deokhwan@inha.ac.kr (D.-H. Kim), sllee@hufs.ac.kr (S.-L. Lee), chungcw@kaist.ac.kr (C.-W. Chung).

0020-0255/ - see front matter \textcopyright 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.04.026
This paper presents a visual shape descriptor based on the sectors and shape context of contour lines. Fig. 1 shows a schematic of the process used to compute this descriptor. The scale-space extreme points are first detected and the keypoint is located by removing the unstable ones. The keypoint orientation is then assigned. For feature calculation, a circular region is then set around each keypoint. Finally, a sector-based feature vector and a feature vector based on shape context of contour lines are calculated. Both vectors are then combined to form a new descriptor.

For the sector-based feature vector, the region is separated into sectors of equal radii and angles similar to log-polar [22]. The gradient magnitude and orientation of the pixels in each sector are calculated and mapped to a gradient histogram. A feature vector is then generated using the histograms from all sectors around the keypoint. For the feature vector based on the shape context of contour lines, the gray value of the keypoint is projected into the sea level of the geographical map. The altitude of each point in the circular region is then obtained by comparing the gray value of the point to that of the keypoint. Contour lines are then formed by grouping the points of similar altitude and the shape context of each contour line in the region is obtained. Similarly, the feature vector is generated using the shape context histograms of all contour lines. Finally, both feature vectors are combined to form a new descriptor.

The main contribution of the proposed method is as follows:

- The dimensionality of the descriptor is reduced using the sectors to calculate the gradient magnitude and orientation.
- The shape features of the region without edge detection are obtained using the shape context of the contour lines.

A combination of the above feature vectors makes the proposed descriptor more robust with respect to image scale, illumination and noise. This new descriptor can achieve better matching accuracy and lower dimensionality than the SIFT descriptor.

The remainder of this paper is organized as follows: Section 2 presents the background of the paper. Section 3 introduces the new approach. The experiments and conclusions are discussed in Sections 4 and 5, respectively.

2. Background

The SIFT descriptor proposed by Lowe [12] is used extensively for image matching. It includes four stages: scale-space extrema point detection, keypoint localization, orientation assignment, and descriptor formation.

In the first stage, the image is smoothed by the Gaussian kernel of different scales and then down-sampled to construct a Gaussian pyramid (Fig. 2). A DoG pyramid is then built by subtracting the adjacent smoothed images in the same octave. Finally, the value of each point in the DoG image is compared with the other points around it. A point of which value is either the maximum or minimum is determined to be a keypoint.

In the second stage, either the keypoints with low contrast or keypoints close to the edge of the image are pruned to achieve stable keypoints.

In the third stage, one or more consistent orientations are assigned for each keypoint to achieve invariance of image rotation.

![Fig. 1. The process of generating the new descriptor.](image-url)
In the fourth stage, the region around the keypoint is separated into 16 blocks and the descriptor is calculated in these blocks. The gradient magnitude and orientation for each pixel in each block are then calculated. The position and orientation of the point are rotated to the orientation of the keypoint. The descriptor is thus robust with respect to the image rotation. Finally, the gradient magnitude is projected onto an 8-bin histogram for the orientation (Fig. 3).

Since there are 16 blocks in the region and 8 orientation bins for each block, the dimensionality of the SIFT descriptor is 128.

Mortensen et al. introduced a new SIFT descriptor with a global context [14], which combined the SIFT descriptor with curvilinear shape information to improve the matching accuracy. The curvilinear shape information is collected by computing the maximum curvature at each pixel in a region larger than the one used in SIFT. Since the curvilinear information is calculated from a larger region, the level of mismatch is reduced when there are multiple similar descriptors in the same image. However, its dimensionality is 188 compared to 128 for the SIFT descriptor.

The PCA-SIFT descriptor [7] was introduced by Kel and Sukthankar to reduce the dimensionality and improve the matching accuracy. However, experiments showed that its matching accuracy was low in some situations [15]. Mikolajczyk and Schmid proposed a new descriptor based on the gradient location-orientation histogram (GLOH) [15]. They calculated the descriptor in log-polar coordinates with three bins in the radial direction (the radii were set to 6, 11, and 15) and 8 bins in the angular direction, for a total of 17 location bins. The gradients were separated into 16 bins, which led to a descriptor dimensionality of 272. Finally, PCA technology was used to reduce the dimensionality to 128.

### 3. The visual shape descriptor

The proposed local descriptor algorithm uses scale-space DoG extrema detection steps similar to the SIFT detection algorithm [12].

After detecting all keypoints, a visual shape descriptor was constructed. The descriptor consists of a two-component feature vector, which includes the gradient magnitudes and orientations of the sectors as well as the shape context of contour lines around the keypoint. Therefore, the proposed descriptor can be defined as follows:

\[ F = \begin{bmatrix} (1 - w)F_{\text{Sectors}} \\ wF_{\text{ContourLines}} \end{bmatrix}. \]

![Fig. 2. The process of building the Gaussian pyramid.](image_url)

![Fig. 3. The process of the descriptor calculation.](image_url)
where $F_{\text{Sectors}}$ is a 64-dimensional feature vector based on sectors, $F_{\text{ContourLines}}$ is a 48-dimensional feature vector based on the local shape context, and $w$ is the weight of two components according to user’s preference \[11\]. Hence, the proposed descriptor algorithm consists of three parts:

1. Find the feature vector using the gradient magnitude and orientation of pixels in the sectors around the keypoint.
2. Find the feature vector using the shape context of the contour lines in a circular region centered at the keypoint.
3. Combine both feature vectors to construct the visual shape descriptor.

First, a circular region is defined from which a two-component feature vector is extracted. Since the keypoint is detected by the DoG detector \[12\], within different scales, the size of the circular region is defined as being relative to the scale of the keypoint. The keypoint is also set as the center of the circular region because it is the most stable point in that region. Therefore, the radius of the circular region is set by the following equation:

$$R = D \times S,$$

(1)

where $D$ is an experimentally determined multiplier, and $S$ is a smoothing scale of the octave that the keypoint belongs to. Fig. 4 shows the circular region centered at the keypoint.

3.1. Feature vector based on sectors

In this section, the gradient magnitude and the orientation are extracted in Cartesian coordinates. However, instead of using Cartesian coordinates to separate the local region into 4-by-4 sub-regions \[12\], the local region is separated according to the sectors to reduce the dimensionality of the descriptor. Since the points close to the keypoint are more important than those farther away from it, the former can be viewed within small grids, whereas the latter can be viewed within large grids using polar coordinates to separate the circular region. Furthermore, a Gaussian weighting function with $\sigma$ equal to $R$ is assigned to each point to emphasize that the points are close to the keypoint. Therefore, the size of the grid can be enlarged to reduce the dimensionality of the descriptor without degrading its performance.

The circular region is separated into $N$ sectors. A sector is a sub-region divided by the same angle and radius (Fig. 5). The gradient magnitudes and orientations of the points in the sector are then calculated to form a histogram with $M$ bins. The gradient magnitude and orientation can be computed as follows:

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

(2)

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

(3)

where $x$ and $y$ are the Cartesian coordinates of a point, $m(x, y)$ is its gradient magnitude, $\theta(x, y)$ is its orientation, and $L(x, y)$ is its gray value. For each point in the sector, its gradient magnitude is mapped to the histogram bins corresponding to its orientation. Note that the position and orientation of the point are aligned to the orientation of the keypoint, invariance with respect to the image rotation. Finally, all histograms of $N$ sectors around the keypoint are concatenated to form the feature vector $F_{\text{Sectors}}$ with $N \times M$-dimension.

3.2. The shape context of the contour lines

Mikolajczyk and Schmid \[15\] showed that the shape context \[4\] achieved high performance for the images with clear edges. However, its performance suffered when applied to textured scenes or ill-defined edge images. Our innovative approach utilizes the shape context in the circular region without edge detection to overcome the performance problems associated with the problematic edges of textured scenes.

![Fig. 4. The circular region.](image-url)
As discussed in Section 2, the keypoint is detected from the DoG pyramid by comparing its value to the other point values around it. Clearly, the keypoint is the most stable point in that region. Consider that the region near the keypoint is mapped to some terrain when all pixels in the region are projected into a geographical map and the gray value of each pixel represents its altitude in the map. Its altitude increases with increasing value of the point, and vice versa.

Contour lines are useful for representing the terrain characteristics; cf. geographical maps. Therefore, it is important to find the contour lines using the altitudes of the pixels in the circular region. The shape context of the contour lines was then calculated to represent the shape feature of the circular region.

A similar approach to contour lines is adopted, as seen in geographical maps, and the contour line is defined as the distribution of points of similar altitudes in a given region. Therefore, the contour lines are allowed to intersect, and the points are not strictly requested to connect to each other. In addition, the line formed by the points does not need to be closed. Fig. 6 shows three contour lines of an image local region, similar to geographical maps.

Fig. 7 highlights the similarity between the gray values of the pixels and the altitudes of the terrain model. The image can be well-modeled by the “terrain”. The terrain model of Fig. 7a is shown in Fig. 7c, where the x- and y-axes denote the pixel position. The z-axis denotes the gray value of each pixel in the image. Similarly, the terrain model of a local region focused in Fig. 7b is shown in Fig. 7d.

The process of calculating the shape context of the contour lines includes four stages:

3.2.1. Sea level mapping for the keypoint

As mentioned earlier, the keypoint is the most stable point in a circular region. Therefore, the gray value of the keypoint is set to sea level in the circular region for the terrain model.
3.2.2. Altitude formation for the circular region

After setting the sea level, the altitude of each point in the circular region can be calculated by comparing their gray values to that of the keypoint. The altitude can be calculated using Eq. (4)

$$A = |Pv - KeyPv|,$$

where $A$ is the altitude of the point in the region, $KeyPv$ and $Pv$ are the gray values of the keypoint and the other point in the circular region, respectively.

3.2.3. Contour lines grouping

After obtaining the altitudes of all points, the altitudes are grouped into $b$ levels and the pixels in the same level are considered to have similar altitudes. Therefore, the $b$ contour lines can be obtained in the circular region. The contour lines in this paper are not the same as geographical contour lines because they may not be closed and they can intersect with each other. An example of these contour lines is shown in Fig. 8. Fig. 8b shows an enlarged local region of the original image, and Fig. 8c illustrates the contour lines of that local region. The contour lines for the local region are represented by the color as shown in Fig. 8c. The contour lines are not closed and they can intersect with each other, such as the second contour line (green color) and the first contour line (blue color).

While constructing the Gaussian pyramid, the image is smoothed by Gaussian functions at different intervals and different octaves, which means that the neighboring pixel values are not too dissimilar. Therefore, the altitudes of the pixels in the circular region are close to the low level. In Fig. 8, the first and last contour lines denote the lowest and highest altitudes, respectively. The lower altitude areas (the first, second, and third contour lines) are much larger than the higher altitude areas (the fourth, fifth, and the last contour lines). Fig. 9 shows the distribution of the altitudes based on the local regions. The distribution is calculated from 128,511 circular regions of 210 different images, and the size of the local region is determined from (1).

As shown in Fig. 9, the lower altitudes have more points. The altitudes are grouped to make the number of points in each group uniform. This makes the proposed descriptor more distinctive. Therefore, let $F(x)$ be a function corresponding to the distribution shown in Fig. 9. After grouping the altitudes into $b$ levels, the average number of points for each level can be calculated as follows:

$$Avg = \frac{\int_0^{255} F(x)dx}{b}.$$
Let the range of the first level be $[0, G_0)$, where $G_0$ is an altitude value that satisfies the following equation:

$$\int_0^{G_0} F(x)dx = Avg \quad (0 \leq G_0 \leq 255).$$

(6)

$G_0$ can be determined by solving (6). Let the range of level $i$ be $[G_{i-1}, G_i)$. $G_i$ can be found by solving the following equation:

$$\int_{G_{i-1}}^{G_i} F(x)dx = Avg \quad (1 \leq i < \beta, 0 \leq G_i \leq 255).$$

(7)

This grouping strategy is illustrated in Fig. 10.

3.2.4. Shape context histogram generation

After grouping the altitudes into $\beta$ contour lines, all points in the same contour line are selected and the angle values are calculated according to their positions in the region. In the next step, the gradient magnitude of these points is mapped to the shape context histogram according to their angles. Fig. 11 outlined the process of mapping the gradient magnitude of the point to the shape context histogram.

The $x$-axis is separated into $\alpha$ bins according to the angle of the point, and the $y$-axis denotes the accumulation of the gradient magnitude of the points with respect to the points sharing the same angle and contour line level. According to Fig. 11, for example, if the altitude of $P_2$ belongs to contour line 1, and the angle of the point belongs to $A_3$, then the gradient magnitude of this point is mapped to the histogram bin $A_3$ for contour line 1. $\beta$ contour lines are assumed. Similar to Section 3.1, another feature vector $F_{\text{ContourLines}}$ is formed using the shape context of the contour lines. When there are $\alpha$ angle bins and $\beta$ contour lines, its dimensionality is $\alpha \times \beta$.

Finally, both the vectors are normalized and combined to form a descriptor of dimensionality $N'M + \alpha \times \beta$. 

Fig. 8. The contour lines of the circular region.

Fig. 9. The distribution of the altitudes.
4. Experimental evaluation

The performance of the proposed approach was evaluated by comparing the matching accuracy of the proposed descriptor with those of the SIFT, PCA-SIFT, GLOH, and shape context.

4.1. Image collection

The Frontal Face Dataset (M. Weber, California Institute of Technology) was used as the test image set. The dataset consisted of multiple facial images of 24 people. In addition to the differences between people, the images varied according to the facial expression, illumination and background. For these experiments, the following five image sets were selected:

Set 1: for each person, five images were selected randomly from the dataset and a dataset consisting of 120 images was formed. This dataset was used mainly to test the robustness of the descriptors under illumination changes.
Set 2: the sizes of the images in the first dataset were changed according to the scale factors 0.5, 0.8, 1.5, and 2.0. This second (size-scaled) dataset was used to test the robustness of the descriptors under image size scaling and illumination changes.

Set 3: the third dataset was formed by adding noise to the first dataset to test the performance of the descriptors under illumination changes and noisy images.

Set 4: the size of the first dataset was changed and noise was added. These images were compounded with illumination changes, size scale changes and noise.

Note that all the selections above are random and the changes are artificial. The matching accuracy of the proposed descriptor method was then evaluated using these sets.

Finally, to further test and evaluate the proposed method against other methods, a fifth dataset (Set 5) was formed using a generic 3D object categorization dataset consisting of thousands of images in 10 object categories with various backgrounds: bicycles, cars, cell phones, head, irons, monitors, mouse, shoes, staplers, and toasters. Twenty groups were collected randomly from the dataset. Each group consisted of five images of the same object but from different viewing angles, viewing heights, and viewing distances. Some samples of these images are shown in Fig. 12.

4.2. Experimental methodology

The “matching accuracy” was evaluated to probe the performance of the proposed descriptor as follows:

4.2.1. Calculate the distance between the descriptors in both images

Since the descriptor is normalized, a Euclidean distance measure was used to compare two image keypoints. For point \( P_i \) in image \( I \) and point \( P_j \) in image \( J \), the Euclidean distance between them can be expressed as

\[
D_{ij} = \sqrt{\sum_{k=1}^{N} (F_{ik} - F_{jk})^2},
\]

where \( D_{ij} \) is the distance between points \( P_i \) and \( P_j \), \( F_i \) and \( F_j \) are the features of \( P_i \) and \( P_j \), respectively. \( N \) is the dimension of the feature. The more similar two descriptors are, the smaller the distance.
4.2.2. Calculate the number of matching image pairs

For a single keypoint \( P_i \) in image \( I \), the distance values between \( P_i \) and all keypoints in the other image \( J \) were calculated. The distances were ranked in order of closest to farthest. To determine if \( P_i \) is matched to any point in image \( J \), an ambiguity rejection metric is used, which is the ratio of the closest to the second closest distance. If the ratio is smaller than 0.8, the image pair \( I \) and \( J \) are considered to be matched. Therefore, to match the two images, \( I \) and \( J \), the number of matching point pairs between the images is counted. The more matching point pairs found, the more similar the two images.

4.2.3. Get the matching accuracy

One of the image sets was chosen to determine the matching accuracy, as described in Section 4.1. One of the images was selected randomly from the image set and used to match the other images. The number of point pairs matching the other images can be obtained for a given selected image (Fig. 13).

In Fig. 13, \( I_s \) is the selected image, while \( I_1, I_2, I_3, \ldots, I_n \) are the other images in the image set. \( N_1, N_2, N_3, \ldots, N_n \) are the corresponding matching point pairs between image \( I_s \) and images \( I_1, I_2, I_3, \ldots, I_n \). The other images were then ranked from greatest to smallest according to the number of matching point pairs \( (N_1, N_2, N_3, \ldots, N_n) \). Images that are similar to the selected image, which means that they are in the same group as the selected image, will rank near the top if the matching is correct.

The total number of correct matches \( correctN \) can be determined by repeating this process. If the total number of similar images corresponding to the selected images in the image set is \( similarN \) then the matching rate can be defined as follows:

\[
rate = \frac{correctN}{similarN}.
\]  

(9)

All experiments were run on a dual core (TM)2 2.4 GHz machine with 2 GB of main memory running Microsoft Windows XP.

4.3. Experimental parameters

Seventy-two images (three images for each person) were selected from Set 1 to experimentally test the parameters. The weight of the feature was set to 0.5, and the value of the multiplier in (1) was fixed to 10. First, detail experiments were performed to determine if the number of sectors and number of orientation bins separately affects the matching accuracy. Fixing the number of sectors at 8, the matching accuracy did not increase (68.05%) when the number of orientation bins was 8 or greater (Fig. 14). The number of orientation bins was then fixed to 8. Under this condition, the matching accuracy was examined as a function of the number of sectors (Fig. 15). The matching accuracy was best (68.05%) when the number of sectors was 8 or more. The number of sectors was then fixed to 8. Note that the results shown in Figs. 14 and 15 were obtained using the feature vector \( F_{\text{Sectors}} \) only without adding the feature vector based on the shape context of the contour lines.

Other experiments were performed to determine if the number of contour lines and number of the angle bins for the feature vector based on the shape context of contour lines affects matching accuracy. Fig. 16 shows the matching accuracy as a function of the number of contour lines when the number of sectors, orientation, and angle bins for the contour lines are all set to 8. The matching accuracy was best (73.306%) when the number of contour lines was 8 or greater. Therefore, the number of contour lines was fixed to 8. Fig. 17 shows the matching accuracy with respect to the number of angle bins when the number of contour lines, sectors and orientation bins is 8. The best matching accuracy was obtained for 6 angle bins. Therefore, the number of angle bins was fixed to 6.

Another experiment was performed to probe the sensitivity of the matching accuracy to the weight of the feature vector that is based on the shape context of the contour lines. The best matching accuracy was achieved by setting the weight to 0.6 (Fig. 18).

Experiments were also performed to determine the satisfactory value for the multiplier \( D \) in (1). For \( D = 10 \), the performance was best within the range of \( D \) probed in this work (Fig. 19). Therefore, the value of \( D \) was fixed to 10 for all remaining experiments.

4.4. Matching accuracy

The matching accuracy of the proposed descriptor was evaluated by comparing it with those of SIFT [12], PCA-SIFT [7], GLOH [15], and the shape context descriptors [4]. For the PCA-SIFT and GLOH descriptors, since the PCA technique was
adopted, the covariance matrix for PCA was estimated on 128,511 patches collected from 210 images taken from the Frontal Face Dataset. For the shape context, the faces were first extracted from Set 1 - Set 4 to improve its performance.

First, image Set 1 (illumination variation) was used, and matching accuracy of the proposed descriptor and the others was compared (Fig. 20). In terms of the matching accuracy, the new descriptor ("proposed method") outperformed the others. Its
matching accuracy was about 14% higher than that of the SIFT descriptor. GLOH with 128-dimension performs second best. PCA-SIFT and GLOH with 64-dimension achieved similar performance. The PCA-SIFT with lower dimensions (38 and 20) achieved lower matching accuracy than the SIFT descriptor, GLOH and PCA-SIFT with 64-dimension. The shape context method achieved the lowest matching accuracy because the human faces in this dataset had similar shapes.
Fig. 21 shows the matching accuracy of the descriptors using image Set 2 (illumination and scaling size variation). Changing the image size reduces matching accuracy for all descriptors. Interestingly, the new descriptor ("proposed method") outperformed all others. Its matching accuracy was about 16% higher than that of SIFT and about 13% higher than that of the GLOH with 128-dimension. PCA-SIFT and GLOH with 64-dimension achieved similar matching accuracy. GLOH with 128-dimension was the second best and achieved about 3% higher matching accuracy than that of the SIFT descriptor. The matching accuracy of the PCA-SIFT with 20 and 38-dimension, and the shape context was lower than those of the others.

![Fig. 20. The matching accuracy in terms of illumination change.](image1)

![Fig. 21. The matching accuracy in terms of the illumination change and image size change.](image2)

![Fig. 22. The matching accuracy in terms of the illumination change and image noise.](image3)
Fig. 22 shows the matching accuracy of the descriptors using image Set 3 (illumination variation and noise). Since the SIFT descriptor only uses the gradient magnitude and orientation information for its feature vector, it is more sensitive to noise than the proposed descriptor. The new descriptor outperformed the SIFT matching accuracy by about 19%. Note that GLOH and PCA-SIFT apply the PCA technique to remove high frequency components from the feature vector. It thus follows that GLOH with 64 and 128-dimension and the PCA-SIFT with 64-dimension perform better than the SIFT descriptor, as observed in these experiments. The shape context yields similar matching accuracy to that of the PCA-SIFT with 20-dimension.

Fig. 23 shows the matching accuracy of the descriptors using image Set 4 (illumination variation, size variation, and noise). The matching accuracy of all descriptors was degraded for this image set. Notwithstanding, GLOH with both 128 and 64-dimension performed better than the SIFT descriptor. SIFT and PCA-SIFT with 64-dimension achieved similar matching accuracy. PCA-SIFT with 38-dimension and 20-dimension and the shape context yielded lower matching accuracy than the other methods. The new descriptor still outperformed the others. It showed about 21% higher matching accuracy than SIFT.

Fig. 24 shows the matching accuracy of the descriptors using the 3D dataset (Set 5). The new descriptor outperformed the others. The SIFT descriptor performed second best, while PCA-SIFT with 64 and 38-dimension performed better than the GLOH. Since the shapes of the objects in this dataset are different, the shape context achieved higher matching accuracy than those of the GLOH and PCA-SIFT.

5. Conclusion

The visual shape descriptor proposed in this paper is based on the sectors and shape context of the contour lines. The gradient magnitude and orientation feature based on the sectors were first calculated to reduce the dimensionality of the descriptor. Next, a new approach was developed to obtain the local shape feature of the region around the keypoint by using the shape context of the contour lines. Since this approach can acquire the shape feature without edge detection, it overcomes the problem of obtaining the edges from textured images and images with problematic edges. The new descriptor
combines the gradient magnitude, orientation and shape context feature of the local region around the keypoint. Therefore, it represents richer feature information of the local region with lower dimension of 112, whereas the dimension of the SIFT descriptor is 128. Experimental results demonstrate that the new visual shape descriptor is more robust than the others, namely SIFT, PCA-SIFT, and GLOH, in terms of the matching accuracy with respect to illumination variations, scale variations, and noise.

Acknowledgement

The authors thank the Editor and the anonymous reviewers for their valuable comments in improving the quality of this paper. This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2009-0081365), in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (R01-2008-000-20685-0) and in part by the Ministry of Knowledge Economy (MKE) and Korea Institute for Advancement in Technology (KIAT) through the Workforce Development Program in Strategic Technology.

References