Enhanced SIFT Descriptor Based on Modified Discrete Gaussian-Hermite Moment

Tae-Koo Kang , Huazhen Zhang, Dong W. Kim, and Gwi-Tae Park

The discrete Gaussian-Hermite moment (DGHM) is a global feature representation method that can be applied to square images. We propose a modified DGHM (MDGHM) method and an MDGHM-based scale-invariant feature transform (MDGHM-SIFT) descriptor. In the MDGHM, we devise a movable mask to represent the local features of a non-square image. The complete set of non-square image features are then represented by the summation of all MDGHMs.

We also propose to apply an accumulated MDGHM using multi-order derivatives to obtain distinguishable feature information in the third stage of the SIFT. Finally, we calculate an MDGHM-based magnitude and an MDGHM-based orientation using the accumulated MDGHM.

We carry out experiments using the proposed method with six kinds of deformations. The results show that the proposed method can be applied to non-square images without any image truncation and that it significantly outperforms the matching accuracy of other SIFT algorithms.

Keywords: SIFT, modified discrete Gaussian-Hermite moments (MDGHM), local feature extraction.

I. Introduction

The scale-invariant feature transform (SIFT) was presented by Lowe [1] as a means to extract distinctive invariant features from images that have variations in deformations such as image scale and image rotation. Because the SIFT is computationally efficient, resistant to partial occlusion, and relatively invariant with respect to changes in viewpoint, it was widely used in image mosaics, object recognition, and image retrieval.

Recently, a variety of approaches have been proposed for robust feature extraction. Ke and Sukthankar used principal component analysis (PCA) instead of histograms to reduce the computational time, and they provided some comparisons between the SIFT and PCA-SIFT [2]. Bay and others increased the extraction speed of robust features and used integral images for image convolution and the fast-Hessian detector [3]. Mikolajczyk and Schmid presented a comparative study for several local descriptors [4]. The Gaussian-Hermite moment (GHM) was firstly introduced by Shen [5]. Wang and others [6] applied the moments to fingerprint classification in biometrics. Other applications, such as iris identification [7] and stereo matching based on the GHM, have also been reported recently [8].

The GHM has base functions of different orders having different numbers of zero-crossings and very different shapes. Therefore, it can distinguish image features more efficiently and is less sensitive to noise because of its Gaussian nature. Moreover, the GHM has an orthogonal property, which helps to reduce the computational time. The discrete GHM (DGHM) [9] is a discrete version of the GHM and one of the global feature representation methods that can be applied to square images.

Manuscript received Aug. 24, 2011; revised Mar. 25, 2012; accepted Apr. 6, 2012.

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (Grant No. 2011-0010579).

Tae-Koo Kang (phone: +82 2 3290 4772, tkkang@korea.ac.kr), Huazhen Zhang (zhanghz@korea.ac.kr), and Gwi-Tae Park (grpark@korea.ac.kr) are with the Department of Electrical Engineering, Korea University, Seoul, Rep. of Korea.

Dong W. Kim (dwnkim@inhatc.ac.kr) is with the Department of Digital Electronics and Information, Inha Technical College, Incheon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.12.0111.0538

The DGHM is inappropriate for non-square images because the DGHM functions cause image loss (image truncation) or require extra computations. Accordingly, a new method that can prevent image truncation and can control sampling intervals to reduce the number of computations is needed.

Although the existing SIFT techniques cited above can enhance the performance of feature extraction, they are still sensitive to conditions such as non-rigid deformations, large light changes, and large viewpoint changes.

To solve the aforementioned problems, we propose the concept of a movable mask and a modified DGHM (MDGHM). The MDGHM is the DGHM of the movable mask having controllable sampling intervals to represent the local feature of a non-square image. The entire set of features of an image can be represented by the summation of multiple MDGHMs.

We also propose an MDGHM-based SIFT (MDGHM-SIFT) descriptor, in which an accumulated MDGHM using multi-order derivatives to gain distinguishable feature information is applied to the orientation assignment stage of the SIFT.

The remainder of this paper is organized as follows. Section II reviews the relevant aspects of the conventional SIFT and DGHM. Section III describes our MDGHM and MDGHM-SIFT descriptor for local features. Section IV presents our experiment results and detailed discussions of these. Section V summarizes our conclusions.

II. Conventional SIFT Algorithm and DGHM

Table 1 lists the symbology we use to describe the DGHM and SIFT.

Table 1. Symbols and definitions used to describe DGHM and SIFT.

Symbol	Definition	Symbol	Definition
$I(i,j)$	Digital input image	$f(x,y)$	Continuous image
$D(x,y,\sigma)$	Difference-of-Gaussian (DoG)	$H_p(x)$	Hermite polynomial with p-order derivative
$G(x,y,\sigma)$	Gaussian functions	$\bar{H}_p(x)$	Normalized Hermite polynomial
$L(x,y,\sigma)$	Gaussian filtered image	$\bar{H}_p(x/\sigma)$	Gaussian-Hermite function
σ	Standard deviation	$M_{p,q}$	Continuous GHM
$\mu(x,y)$	Orientation of each keypoint	$\bar{H}_p(x,\sigma)$	Gaussian-Hermite function on the x-axis in the discrete domain
$m(x,y)$	Magnitude of each keypoint	$\bar{H}_q(y,\sigma)$	Gaussian-Hermite function on the y-axis in the discrete domain
p, q	Order of derivative	$\eta_{p,q}$	DGHM

1. SIFT

The SIFT algorithm consists of four major stages: 1) scale-space peak selection, 2) keypoint localization, 3) orientation assignment, and 4) keypoint descriptor computation [1].

The first stage is a scale-space peak selection procedure. Scale-space peak selection is implemented efficiently by constructing a Gaussian pyramid and searching for local peaks (extrema) in a series of difference-of-Gaussian (DoG) images, $D(x,y,\sigma)$, which can be expressed as

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

where k is a scale factor and $I(x,y)$ is an input image.

In the second stage (keypoint localization), the low-contrast and unstable keypoints are removed.

The third stage of the SIFT (orientation assignment) identifies the dominant orientations for each keypoint based on the local image patch. In the orientation assignment stage, an orientation, $\mu(x,y)$, and magnitude, $m(x,y)$, are assigned to the keypoints, based on the local image properties as expressed in (2) and (3):

$$m(x,y)=\sqrt{\begin{aligned}&(L(x+1,y)-L(x-1,y))^2\\&+(L(x,y+1)-L(x,y-1))^2\end{aligned}}\ , \tag{2}$$

$$\mu(x,y)=\tan^{-1}\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\ . \tag{3}$$

The fourth stage of the SIFT algorithm (the keypoint descriptor) builds a descriptor for each keypoint based on a patch of pixels in its local neighborhood.

Because SIFT features are local, distinctive, based on the appearance of the object at particular interest points, and robust to partial occlusion, they are invariant to scale changes, image rotations, small illumination changes, and viewpoint changes.

Despite its powerful performance, as mentioned above, the SIFT algorithm is still sensitive to large illumination changes, large viewpoint changes (that is, different keypoints with similar representations in different image maps), and non-rigid changes. A detailed description of the SIFT can be found in [1].

2. DGHM

In [9], the DGHM, which includes a Hermite polynomial, and the GHM are introduced. The p-th degree Hermite polynomial, which is one of the orthogonal polynomials, is given as

$$H_p(x) = (-1)^p \exp(x^2)\frac{d^p}{dt^p}\exp(-x^2). \quad (4)$$

Hermite polynomials satisfy the following orthogonality condition with respect to the weight function $\exp(-x^2)$:

$$\int_{-\infty}^{\infty} \exp(-x^2)H_p(x)H_q(x)dx = 2^p\, p!\sqrt{\pi}\delta_{pq}, \quad (5)$$

where δ_{pd} is the Kronecker delta. To obtain the orthonormal version, the normalized Hermite polynomial $\bar{H}_p(x)$ is calculated by using (5) as

$$\bar{H}_p(x) = \frac{1}{\sqrt{2^p\, p!\sqrt{\pi}\sigma}}\exp(-\frac{x^2}{2})H_p(x). \quad (6)$$

Gaussian-Hermite functions $\bar{H}_p(x/\sigma)$ can be calculated by replacing x in (6) with x/σ:

$$\bar{H}_p(x/\sigma) = \frac{1}{\sqrt{2^p\, p!\sqrt{\pi}\sigma}}\exp(-\frac{x^2}{2\sigma^2})H_p(x/\sigma). \quad (7)$$

Based on (7), the GHM $M_{p,q}$ with order (p, q) of the continuous image $f(x, y)$ can be defined as

$$M_{p,q} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\bar{H}_p(x/\sigma)\bar{H}_q(y/\sigma)dxdy. \quad (8)$$

The GHM is theoretically defined in the continuous domain $(-\infty, \infty)$. To compute the moments for a digital image $I(i, j)$ whose size is $K \times K$ $[0 \le i, j \le K-1]$, a coordinate transformation over the square $[-1 \le x, y \le 1]$ is performed using

$$x = \frac{2i-K+1}{K-1}, \quad y = \frac{2j-K+1}{K-1}. \quad (9)$$

The Gaussian-Hermite functions in the discrete domain can be calculated by using (9) for (7):

$$\begin{cases} \bar{H}_p(x,\sigma) = \dfrac{2}{K-1}\dfrac{1}{\sqrt{2^p\, p!\sqrt{\pi}\sigma}}\exp(-x^2/2\sigma^2)H_p(x/\sigma), \\ \bar{H}_q(y,\sigma) = \dfrac{2}{K-1}\dfrac{1}{\sqrt{2^q\, q!\sqrt{\pi}\sigma}}\exp(-y^2/2\sigma^2)H_q(y/\sigma). \end{cases} \quad (10)$$

From (8) and (10), the DGHM η_{pq}, which is a discrete version of the GHM, can be derived as

$$\eta_{p,q} = \frac{4}{(K-1)^2}\sum_{i=0}^{K-1}\sum_{j=0}^{K-1} I(i,j)\bar{H}_p(x,\sigma)\bar{H}_q(y,\sigma). \quad (11)$$

The DGHM has computational advantages and can be simply implemented. Moreover, it uses all pixels of an image to represent the peculiarity of the image. In other words, it can represent global image features, which can provide a generalized description of an entire object with a single vector [13].

III. MDGHM-Based SIFT Descriptor

The following subsections provide the details of the MDGHM and MDGHM-SIFT algorithm. Table 2 lists the symbols used in our mathematical description of the MDGHM and MDGHM-SIFT descriptor.

1. MDGHM

As shown in (11), the DGHM is applied to a square image. It cannot be applied to non-square images because the discrete Gaussian-Hermite functions, $\bar{H}_p(x,\sigma)$ and $\bar{H}_q(y,\sigma)$, may cause image data loss (image truncation) or require extra

Table 2. Symbols and definitions used to describe MDGHM and MDGHM-SIFT.

Symbol	Definition	Symbol	Definition
$I(i,j)$	Digital input image	$t(u, v)$	Mask image
W	Width of input image	M	Width of mask image
H	Height of input image	N	Height of mask image
k_M	The number of max. samples on the u-axis	m_M	Sampling intervals on the u-axis
k_N	The number of max. samples on the v-axis	m_N	Sampling intervals on the v-axis
k_W	The number of max. samples on the i-axis	m_W	Sampling intervals on the i-axis
k_H	The number of max. samples on the j-axis	m_H	Sampling intervals on the j-axis
$\hat{H}_p(x)$	Modified discrete Gaussian-Hermite functions of the mask on the x-axis	$\hat{H}_q(y)$	Modified discrete Gaussian-Hermite functions of the mask on the y-axis
$\hat{\eta}_{p,q}(i,j)$	MDGHMs of a mask $t(u, v)_{(i,j)}$ at an arbitrary point (i,j) on the image $I(i,j)$	$\hat{\eta}_{p,q}(i,j,m_M,m_N)$	MDGHMs with sampling intervals

computation. Therefore, a new variable that can control the sampling intervals to reduce the number of computations is needed.

The DGHM is a feature representation method for global image features, which means that it cannot be applied to local feature representation [13]. To represent local image features, we apply a mask onto the 2D digital image.

Let $I(i, j)$ be a digital 2D image whose size is $W\times H$ $[0 \le i \le W-1, 0 \le j \le H-1]$, and let $t(u, v)$ be a mask whose size is $M\times N$ $[0 \le u \le M-1, 0 \le v \le N-1]$. The variables k_M, k_N, k_W, and k_H in Table 2 are calculated as

$$k_M = M / m_M,\ k_N = N / m_N,\ \text{ for } t(u,v), \tag{12}$$

$$k_W = W / m_W, k_H = H / m_H,\quad \text{for } I(x,y). \tag{13}$$

The pixel values of the mask $t(u, v)_{(i,j)}$ located at an arbitrary point (i,j) on the input image $I(i,j)$ are obtained using

$$t(u,v)_{(i,j)} = I(u+i-M/2+1, v+j-N/2+1). \tag{14}$$

In (14), $M/2-1$ and $N/2-1$ are offset to center the mask around an arbitrary point (i,j).

Based on (9) and (10), the coordinate of the mask $t(u, v)_{(i,j)}$ is transformed to be $-1 \le x, y \le 1$ by

$$x = \frac{2u-M+1}{M-1},\ y = \frac{2v-N+1}{N-1}, \tag{15}$$

and the discrete Gaussian-Hermite functions of the mask $t(u, v)_{(i,j)}$ can be written as

$$\begin{cases} \hat{H}_p(x,\sigma) = \dfrac{2}{M-1}\bar{H}_p(x/\sigma) \\ \qquad = \dfrac{2}{M-1}\dfrac{1}{\sqrt{2^p\, p!\sqrt{\pi}\sigma}}\exp(-x^2/2\sigma^2)H_p(x/\sigma), \\ \hat{H}_q(y,\sigma) = \dfrac{2}{N-1}\bar{H}_q(y/\sigma) \\ \qquad = \dfrac{2}{N-1}\dfrac{1}{\sqrt{2^q\, q!\sqrt{\pi}\sigma}}\exp(-y^2/2\sigma^2)H_q(y/\sigma), \end{cases} \tag{16}$$

where $2/M-1$ and $2/N-1$ are scale factors.

From (14) and (16), the MDGHM at the arbitrary point (i,j) on the input image $I(i,j)$ is given as

$$\hat{\eta}_{p,q}(i,j) = \frac{4}{(M-1)(N-1)}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} I(u+i-M/2+1, v+j-N/2+1) \times \hat{H}_p(x,\sigma)\hat{H}_q(y,\sigma). \tag{17}$$

To (17), we then apply the sampling interval m_M, m_N to the position value of the mask $t(u, v)_{(i,j)}$ to reduce the amount of computation. Therefore, the MDGHM with sampling intervals $\hat{\eta}_{p,q}(i,j,m_M,m_N)$ can be written as

$$\hat{\eta}_{p,q}(i,j,m_M,m_N) = \frac{4}{(M-1)(N-1)}\sum_{u=0}^{k_M-1}\sum_{v=0}^{k_N-1} I(i+(m_M u-M/2+1), j+(m_N v-N/2+1))\hat{H}_p(x,\sigma)\hat{H}_q(y,\sigma). \tag{18}$$

Using (18), we can express the whole moment of a non-square image without data truncation by moving the mask on the image, and this moment can be expressed by

$$\begin{aligned}\eta_{p,q} &= \frac{4}{(M-1)(N-1)}\sum_{i=0}^{k_W-1}\sum_{j=0}^{k_H-1}\sum_{u=0}^{k_M-1}\sum_{v=0}^{k_N-1} I((m_W i+M/2-1)+(m_M u-M/2+1), \\ &\quad (m_H j+N/2-1)+(m_N v-N/2+1))\hat{H}_p(x,\sigma)\hat{H}_q(y,\sigma) \\ &= \frac{4}{(M-1)(N-1)}\sum_{i=0}^{k_W-1}\sum_{j=0}^{k_H-1}\sum_{u=0}^{k_M-1}\sum_{v=0}^{k_N-1} I((m_W i+m_M u),(m_H j+m_N v)) \\ &\quad \times \hat{H}_p(x,\sigma)\hat{H}_q(y,\sigma).\end{aligned} \tag{19}$$

Although the quality of a feature image depends on the sampling interval, the MDGHM can present the local moment at an arbitrary point on an input image and can be utilized as an effective tool for identifying local features.

2. MDGHM-Based SIFT Descriptors

In [4], [10], the SIFT is sensitive to conditions such as large light changes and large viewpoint changes because the gradient method used in the orientation assignment of the SIFT does not provide distinctive information to determine the location of each keypoint accurately in a deformed image.

To solve this problem and improve the matching accuracy of the SIFT algorithm, we propose an MDGHM-SIFT algorithm. Figure 1 presents an overview of the proposed MDGHM-SIFT algorithm.

As shown in Fig. 1, we apply the MDGHM to the keypoint candidates in the SIFT and calculate the MDGHM-based magnitude and MDGHM-based orientation using an accumulated MDGHM in the third stage of the SIFT. The MDGHM-based magnitude and MDGHM-based orientation are utilized in structuring the orientation histogram.

Let (i_a, j_a, s) be the locations of a keypoint candidate, where a is the index of the keypoint candidates in a scale space s in the first stage of the SIFT. From (18), an arbitrary point (i, j) corresponds to the location of each keypoint candidate (i_a, j_a, s). The local moment at (i_a, j_a, s) can therefore be calculated by using the MDGHM as

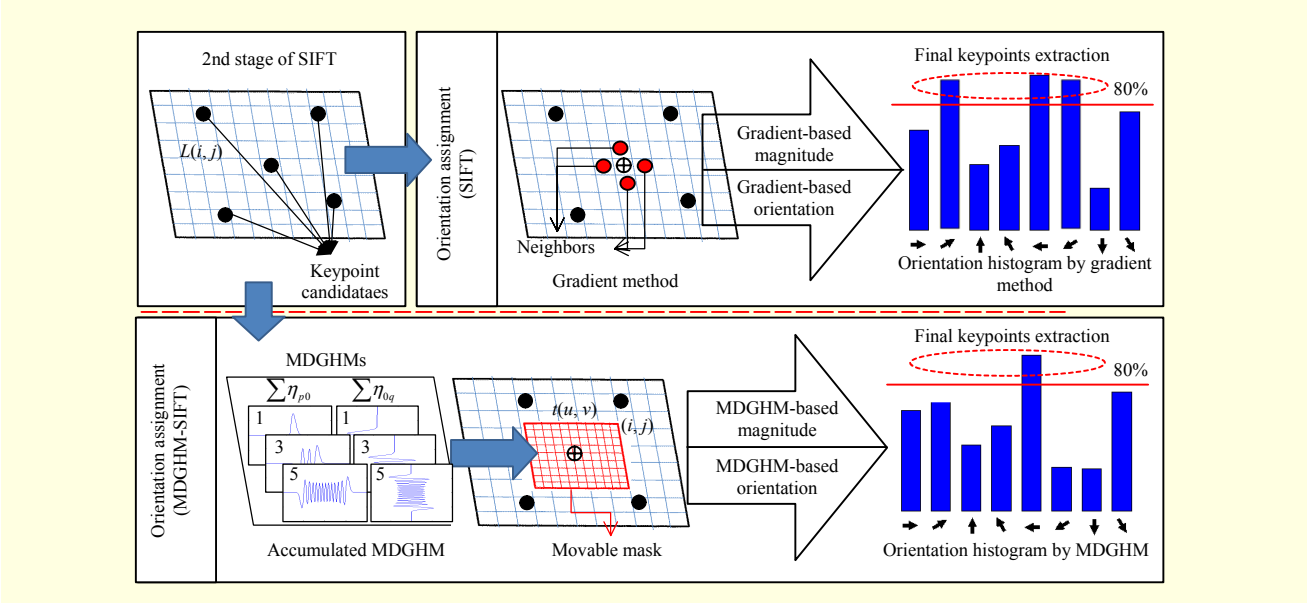


Fig. 1. Overview of MDGHM-SIFT algorithm.

$$\hat{\eta}_{p,q}(i_a, j_a, s) = \frac{4}{(M-1)(N-1)} \sum_{u=0}^{k_M-1} \sum_{v=0}^{k_N-1} L\left(i_a + (m_M u - M/2 + 1), j_a + (m_N v - N/2 + 1), s\right) \hat{H}_p(x,\sigma) \hat{H}_q(y,\sigma), \quad (20)$$

where $L(i_a, j_a, s)$ is the Gaussian filtered image constructed using (1).

In addition, to identify the dominant features, we sum the first, third, and fifth order derivatives (p=q=1, 3, 5) of the MDGHM at a keypoint. The accumulation of multi-order derivatives in the MDGHM (accumulated MDGHM) makes the dominant features more distinguishable than the other features.

From (20), the x-axis component $\hat{\eta}_{p,0}(i_a, j_a, s)$ and y-axis component $\hat{\eta}_{0,q}(i_a, j_a, s)$ of the MDGHM for each keypoint in each scale space can be determined. Therefore, we can calculate the MDGHM-based magnitude $\hat{m}(i_a, j_a, s)$ and the MDGHM-based orientation $\hat{\mu}(i_a, j_a, s)$ using the accumulated MDGHM for each keypoint candidate as

$$\hat{m}(i_a, j_a, s) = \sqrt{\sum_{p=0}^{2} (\hat{\eta}_{(2p+1),0}(i_a, j_a, s))^2 + \sum_{q=0}^{2} (\hat{\eta}_{0,(2q+1)}(i_a, j_a, s))^2}, \quad (21)$$

$$\hat{\mu}(i_a, j_a, s) = \arctan\left(\frac{\sqrt{\sum_{q=0}^{2} (\hat{\eta}_{0,(2q+1)}(i_a, j_a, s))^2}}{\sqrt{\sum_{p=0}^{2} (\hat{\eta}_{(2p+1),0}(i_a, j_a, s))^2}} \right). \quad (22)$$

The size of the MDGHM is defined as

$$\begin{cases} M = 2 \times \text{round}(\max(p)\sigma) + 1, & \text{for } x\text{-axis}, \\ N = 2 \times \text{round}(\max(q)\sigma) + 1, & \text{for } y\text{-axis}, \end{cases} \quad (23)$$

where p and q are the orders of the MDGHM and σ represents the standard deviation. We use p=q=1, 3, 5, and σ=0.3 so that the mask size of the MDGHM is 5×5.

All keypoint candidates that have their own MDGHM-based magnitude and orientation participate in structuring the orientation histogram and are counted to determine the final keypoints. The final keypoints extraction process in the orientation assignment stage classifies keypoint candidates into keypoint candidates with a dominant orientation (dominant keypoints) and keypoint candidates with a minor orientation (minor keypoints). Finally, the dominant keypoints are selected as the final keypoints.

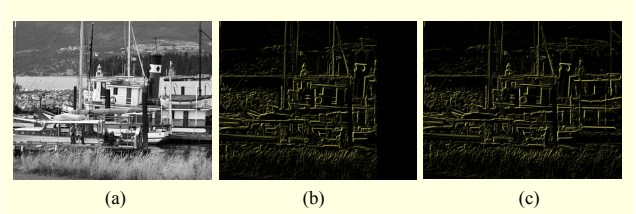
The fourth stage of the MDGHM-SIFT descriptor is similar to the standard SIFT, except that we replace the gradient magnitude and orientation of the descriptor with the MDGHM-based magnitude and orientation.

IV. Experiments and Discussion

We carry out experiments to achieve feature representation of the MDGHM and the performance evaluation of four kinds of SIFT descriptors.

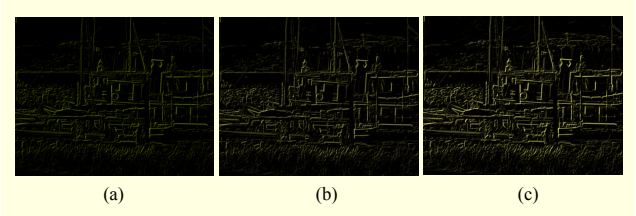
1. Feature Representation in DGHM and MDGHM

We apply the DGHM and MDGHM to a non-square image.



(a) (b) (c)

Fig. 2. Feature images using DGHM and MDGHM for non-square image: (a) initial image, (b) feature image obtained using DGHM, and (c) MDGHM.



(a) (b) (c)

Fig. 3. Feature images using gradient method and MDGHM: (a) gradient, (b) 1st order MDGHM, and (c) 3rd order MDGHM.

Figure 2 shows the results of the experiment.

As shown in Fig. 2(b), the DGHM causes image truncation in the case of the non-square image, resulting in the disappearance of part of the feature image, as demonstrated by the black space on the right side of the image. As shown in Fig. 2(c), the MDGHM represents the features in the entire area of the non-square image. In the case of a square image, both methods produce feature images without any truncation.

We compare the gradient method used in the SIFT with the MDGHM-based method in terms of the amount of feature information produced, and the results of this comparison are shown in Fig. 3. In Fig. 3, the yellow dots represent features. The more bright dots there are, the more distinctive feature information has been extracted. Therefore, the images can be ranked in ascending order of the amount of feature information as Figs. 3(a) < 3(b) < 3(c).

By comparing Figs. 3(b) and 3(c), we can see that the MDGHM using a higher-order derivative extracts more distinctive feature information than one using a lower-order derivative. We can therefore conclude that the MDGHM method contains more feature information than the gradient method.

2. Experimental Comparison of Four SIFT Algorithms

We evaluate four kinds of SIFT algorithms, as tabulated in Table 3. The SIFT algorithm generally consists of four stages. The four algorithms are classified according to the difference of the third and the fourth stage of each SIFT algorithm. For the experiment, a dataset of six image deformations, as described in [11], is used. They are a) scale change, b) image rotation, c) viewpoint change, d) image blur, e) JPEG compression change, and f) illumination change.

Table 3. Characteristics of four SIFT algorithms tested.

Stage \ Algorithm	SIFT	MDGHM-SIFT	MDGHM-PCA-SIFT (n=20)	PCA-SIFT (n=20)
3rd stage	Gradient	MDGHM	MDGHM	Gradient
4th stage	Histogram	Histogram	PCA	PCA

All necessary metrics for the evaluation of the four SIFT algorithms are measured and calculated during the experiment. We use sampling intervals m_M=m_N=1 for the MDGHM-related SIFT algorithm.

A. Experiment Setup

To evaluate the performance of the methods, we need to define evaluation metrics and a matching method.

a. Evaluation Metrics

We choose three evaluation metrics: recall, 1–precision, and the F-score, as used in [4], [12].

$$\text{recall} = \frac{\text{number of correct-positives}}{\text{total number of positives}}, \tag{24}$$

$$1-\text{precision} = \frac{\text{number of false-positives}}{\text{total number of matches}}, \tag{25}$$

$$\text{F-score}=2\times\frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}. \tag{26}$$

A correct-positive is a match where two keypoints correspond to the same physical location. A false-positive is a match where the two keypoints come from different physical locations. From the correct-positive and the false-positive numbers, we can determine recall and 1–precision. A recall versus 1–precision graph is generated as follows. All pairs of keypoints from different images are examined. If the Euclidean distance between two feature vectors for a particular pair of keypoints falls below a chosen threshold, this pair is termed "a match." Adjusting the threshold selects the appropriate trade-off between correct-positives and false-positives. The recall versus 1–precision graphs are generated by varying the threshold for each algorithm [2].

In addition, we investigate accuracy using the F-score. The F-score is the harmonic mean of the recall and precision, and its highest value is 1, and its lowest value is 0. The F-score is defined as shown in (26).

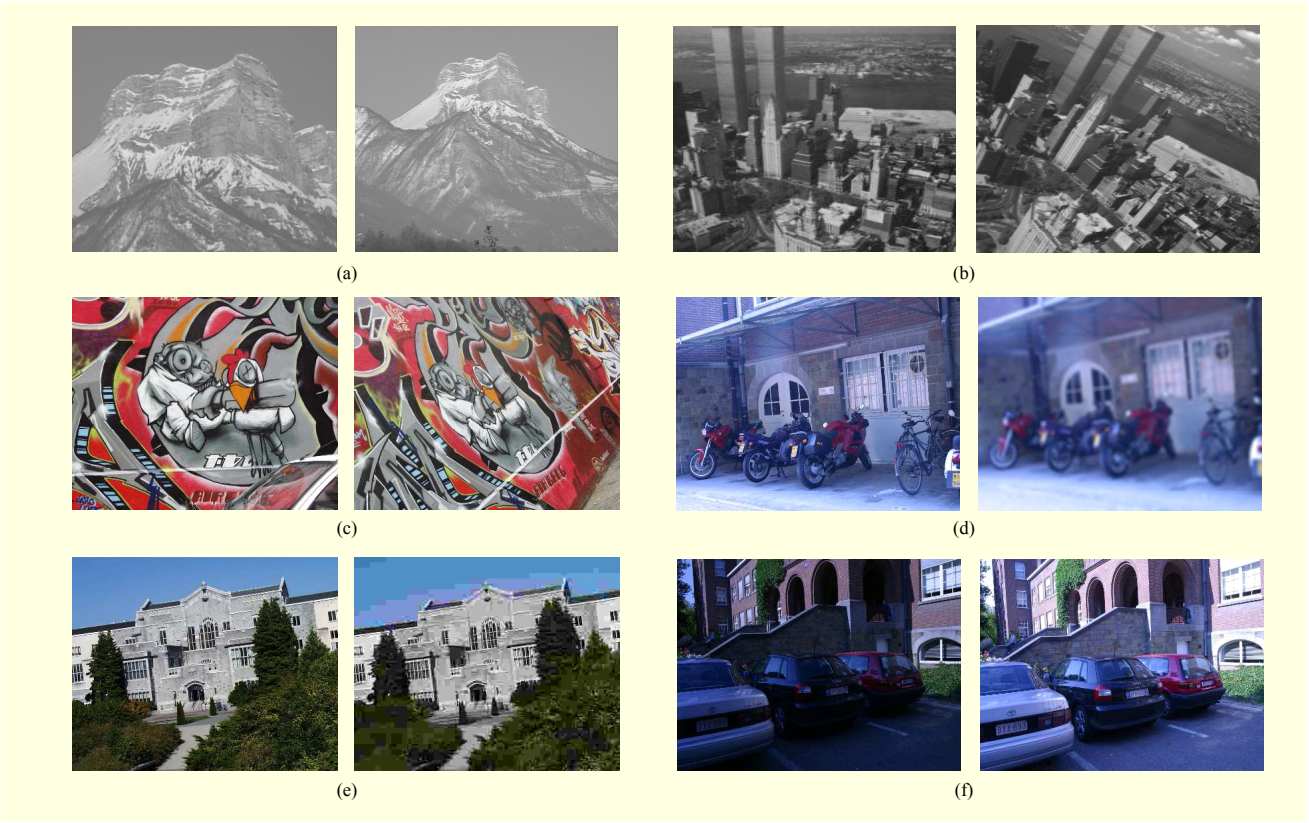


Fig. 4. Examples of images and their deformations used for evaluation: (a) scale change, (b) image rotation, (c) viewpoint change, (d) image blur, (e) JPEG-compression change, and (f) illumination change.

b. Matching Method

The definition of a match depends on the matching method. There are three matching methods [4]. In the nearest neighbor distance ratio (NNDR) matching method, a threshold is applied to the distance ratio between the first and the second nearest neighbors. Therefore, using the NNDR method, only one match is found. Because of these characteristics, we utilize the NNDR matching method. Hereafter, for convenience, we refer to the NNDR simply as the distance ratio (DR).

B. Experiment and Results

a. Image Deformation

We carry out our test of the four SIFT algorithms identified in Table 3 by imposing six deformations: a) scale change, b) image rotation, c) viewpoint change, d) image blur, e) JPEG-compression change, and f) illumination change.

For the deformation test, we prepare six pairs of images as shown in Fig. 4, including a reference image and one image of a deformation. A deformation in a dataset consists of five to six continuously varying images. The image on the left side is a reference image and the image on the right side is one image among five to six continuously varying images. To deform the reference images, we use a) zoom out, b) rotation by 30º to 45º, c) viewpoint change by 50º to 60º, d) Gaussian blur by zooming and focusing, e) quality degradation by 5% through JPEG compression, and f) illumination change or brightening of the image.

b. Experiment Result Data

Table 4 shows the resultant data of the SIFT and MDGHM-SIFT methods after imposing the six deformations. Each result in Table 4 represents the average value of the test results of five to six continuously varying images. All necessary metrics for the evaluation of the four SIFT algorithms are products of the experiment. However, for this paper, we plot for comparison only the results of applying the SIFT and the MDGHM-SIFT methods.

During the experiment, DR is increased from 0 to 1 in intervals of 0.2. The keypoints values represent the total number of keypoints extracted by each of the SIFT algorithms. True and false matches are obtained during the experiment by varying the DR used as a matching threshold. Recall and 1–precision (or false-positive rate) scores are calculated using the obtained data. The number of total matches, correct-positives,

Table 4. Experiment results of SIFT and MDGHM-SIFT.

Algorithm	Deformation (Keypoints)	DR	Recall	1–precision (FPR)	True match	False match	Total match	Algorithm	Deformation (Keypoints)	DR	Recall	1–precision (FPR)	True match	False match	Total match
SIFT	Scale (4,900)	0.2	0.001	0.000	3	0	3	MDGHM-SIFT	Scale (2,695)	0.2	0.000	0.000	1	0	1
		0.4	0.016	0.000	80	0	80			0.4	0.015	0.000	40	0	40
		0.6	0.053	0.004	259	1	260			0.6	0.068	0.005	183	1	183
		0.8	0.096	0.082	469	42	511			0.8	0.117	0.060	316	20	336
		1.0	0.283	0.716	1,385	3,500	4,885			1.0	0.307	0.693	827	1,867	2,694
	Rotation (1,974)	0.2	0.000	0.000	13	0	13		Rotation (984)	0.2	0.000	0.000	5	0	5
		0.4	0.043	0.000	162	0	162			0.4	0.035	0.000	81	0	81
		0.6	0.179	0.000	410	0	410			0.6	0.180	0.000	210	0	210
		0.8	0.358	0.014	719	10	729			0.8	0.377	0.003	367	1	368
		1.0	0.628	0.367	1,248	723	1,971			1.0	0.700	0.293	694	288	982
	Viewpoint (5,021)	0.2	0.000	0.000	0	0	0		Viewpoint (3,611)	0.2	0.000	0.000	0	0	0
		0.4	0.004	0.000	22	0	22			0.4	0.002	0.000	8	0	8
		0.6	0.016	0.000	81	0	81			0.6	0.018	0.000	65	0	65
		0.8	0.067	0.037	338	13	351			0.8	0.086	0.025	310	8	318
		1.0	0.436	0.561	2,187	2,790	4,977			1.0	0.541	0.454	1,954	1,622	3,576
	Blur (513)	0.2	0.133	0	68	0	68		Blur (514)	0.2	0.146	0	75	0	75
		0.4	0.222	0	114	0	114			0.4	0.267	0	137	0	137
		0.6	0.263	0	135	0	135			0.6	0.317	0	163	0	163
		0.8	0.353	0.037	181	7	188			0.8	0.424	0.031	218	7	225
		1.0	0.659	0.341	338	175	513			1.0	0.685	0.315	352	162	514
	JPEG-compression (907)	0.2	0.287	0.000	260	0	260		JPEG-compression (957)	0.2	0.295	0.000	282	0	282
		0.4	0.465	0.007	422	3	425			0.4	0.496	0.006	475	3	478
		0.6	0.581	0.019	527	10	537			0.6	0.631	0.019	604	12	616
		0.8	0.628	0.087	570	54	624			0.8	0.681	0.073	652	51	703
		1.0	0.656	0.344	595	312	907			1.0	0.710	0.290	679	278	957
	Illumination (203)	0.2	0.099	0.000	20	0	20		Illumination (172)	0.2	0.134	0.000	23	0	23
		0.4	0.172	0.000	35	0	35			0.4	0.297	0.000	51	0	51
		0.6	0.227	0.000	46	0	46			0.6	0.366	0.000	63	0	63
		0.8	0.325	0.043	66	3	69			0.8	0.488	0.023	84	2	86
		1.0	0.512	0.477	104	95	199			1.0	0.616	0.380	106	65	171

and false-positives increases as the DR increases and abruptly increases in the range between DR=0.8 and DR=1.0.

For most of the deformed images, the number of total matches, correct-positives, and false-positives obtained using the SIFT is greater than that obtained using the MDGHM-SIFT, but this pattern is reversed in the case of image blur and illumination change. Therefore, it is difficult to evaluate which algorithm is best using only these figures. We will discuss our evaluation of this in subsection IV.3. Besides these numbers, there are also fewer keypoints detected by the MDGHM-SIFT than by the SIFT. This comes from the fact that the number of dominant orientations is reduced in the MDGHM-SIFT, unlike in the SIFT.

3. Evaluation and Discussion

The performance of the four SIFT algorithms listed in Table 3 is illustrated by the 1–precision versus recall graphs shown in Fig. 5 for six deformations. Markers on the line to identify each algorithm are DR points at intervals of 0.05.

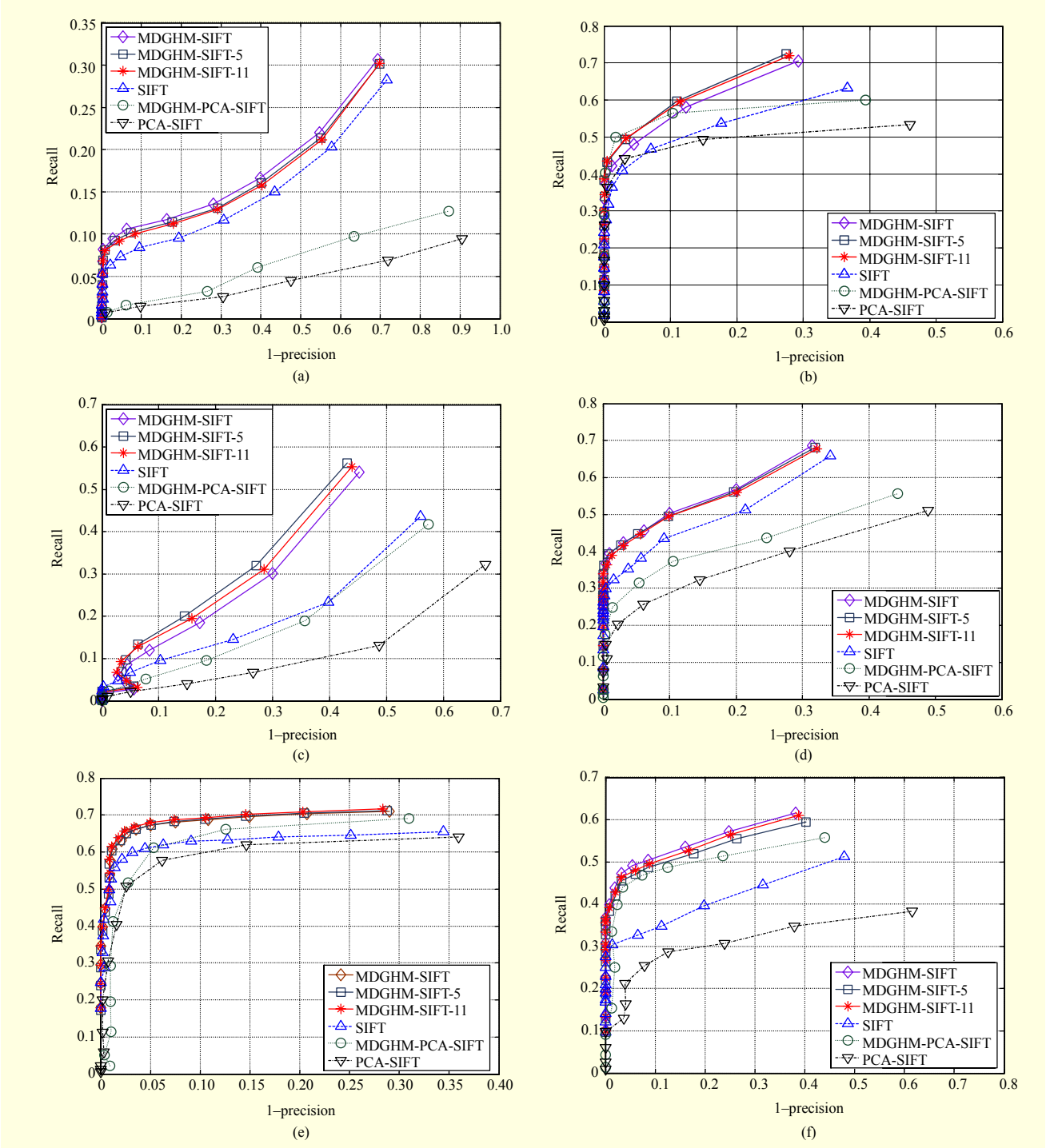


Fig. 5. Comparison of four SIFT algorithms in response to six distortions: (a) scale change, (b) image rotation, (c) viewpoint change, (d) image blur, (e) JPEG-compression change, and (f) illumination change.

The DR value of the rightmost marker (last marker) on each performance line is 1.0. We can see that the four SIFT algorithms performed well at DR values greater than 0.8. The MDGHM-SIFT displays the best performance among the four SIFT algorithms for all deformations. In addition, the MDGHM-PCA-SIFT performs better than the SIFT in response to a viewpoint change and illumination change and displays an almost equal performance to the SIFT in image rotation and JPEG-compression. These facts tell us that the MDGHM has a positive effect on the performance of the PCA-

Table 5. Performance results.

Algorithm / Metric / Deformation	MDGHM-SIFT			SIFT		
	Recall	Precision	F-score	Recall	Precision	F-score
Scale	0.307	0.307	0.307	0.283	0.284	0.283
Rotation	0.700	0.707	0.703	0.628	0.633	0.630
Viewpoint	0.541	0.546	0.544	0.436	0.439	0.437
Blur	0.685	0.685	0.685	0.659	0.659	0.659
Compression	0.710	0.710	0.710	0.656	0.656	0.656
Illumination	0.616	0.620	0.618	0.512	0.523	0.517
Algorithm / Deformation	MDGHM-PCA-SIFT			PCA-SIFT		
Scale	0.127	0.128	0.128	0.095	0.095	0.095
Rotation	0.600	0.606	0.603	0.534	0.539	0.537
Viewpoint	0.417	0.426	0.422	0.322	0.287	0.303
Blur	0.556	0.558	0.557	0.511	0.512	0.511
Compression	0.691	0.691	0.691	0.641	0.641	0.641
Illumination	0.558	0.561	0.560	0.384	0.381	0.383

SIFT.

Subsection III.2. provides the calculation of the accumulated MDGHM using multi-order derivatives to obtain dominant feature information. To compare the influence of the derivatives on the MDGHM-SIFT performance, we carry out a performance test on three MDGHM-SIFTs, that is, the MDGHM-SIFT (p=q=1, 3, 5), MDGHM-SIFT-5 (p=q=5, 7, 9), and the MDGHM-SIFT-11 (p=q=11, 13, 15). The upper three lines, designated by MDGHM-SIFT, MDGHM-SIFT-5, and MDGHM-SIFT-11, of each graph in Fig. 5 demonstrate the influence of the derivatives on the performance.

The three methods demonstrate very similar performances, with each being superior for at least one of the deformations. The MDGHM-SIFT-5 shows a slightly superior performance. However, because the increase in the order of derivatives results in an increase of the number of computations required, the MDGHM-SIFT-5 does not result in a large improvement in performance compared to the MDGHM-SIFT. The MDGHM-SIFT-11 exhibits the worst performance among the MDGHM-SIFTs. This is because as the order of derivatives becomes very high, the mask size becomes excessively large, and the possibility of the mask overlapping increases. As a result, both the distinctiveness of the keypoints and the matching accuracy decrease.

Table 5 shows the recall, precision, and F-score needed to judge which algorithm is superior. As shown in Table 5, according to the F-score, the MDGHM-SIFT exhibits an enhancement of approximately 2.4% in the case of scale change, a 7.3% enhancement in the case of image rotation, a 10.7% enhancement in the case of viewpoint change, a 5.4% enhancement in the case of image blur, a 2.6% enhancement in the case of JPEG-compression change, and a 10.1% enhancement in the case of illumination change compared to the performance of the SIFT. Even though the number of total keypoints identified by the MDGHM-SIFT is smaller than that of the SIFT as shown in Table 4, the performance of the MDGHM-SIFT is still better than that of the SIFT according to the F-score. This means that the MDGHM-SIFT is the most effective of the four SIFT algorithms tested.

These differences in performance stem from the MDGHM-based dominant orientation and magnitude in the orientation assignment stage of the SIFT algorithm. In the orientation assignment stage, the MDGHM-SIFT uses the accumulated MDGHM with multi-order derivatives, which makes dominant keypoints more dominant and minor keypoints weaker, which results in a decrease in the number of final keypoints and an increase in the matching accuracy.

The MDGHM-SIFT can extract more distinguishable final keypoints because more minor keypoints can be discarded. The MDGHM-SIFT can therefore be an efficient alternative approach, as it performs with a higher matching accuracy and lower computation cost than other algorithms.

V. Conclusion

We proposed a modified discrete Gaussian-Hermite moment (MDGHM) and the MDGHM-based scale-invariant feature transform (MDGHM-SIFT) descriptor. We devised and tested the concept of a movable mask. The MDGHM is the DGHM with a movable mask and controllable sampling intervals that allow the local features of a non-square image to be represented. The total features of an image can then be represented by the summation of multiple MDGHMs.

We also proposed the application of the accumulated MDGHM using multi-order derivatives to obtain distinguishable feature information in the third stage of the SIFT. Finally, we calculated the MDGHM-based magnitude and the MDGHM-based orientation using the accumulated MDGHM.

We carried out the experiment for the proposed method using six kinds of deformations: a) scale change, b) image rotation, c) viewpoint change, d) image blur, e) JPEG-compression change, and f) illumination change. The result shows that the proposed method can be applied to non-square images without any image truncation and that the proposed method outperforms other SIFT algorithms by significantly improving the matching accuracy.

An algorithm for adaptively tuning sampling interval

parameters and the application of the proposed method to particular areas such as image stitching and environment recognition of robots are left for future study.

References

[1] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Computer Vision*, vol. 60, 2004, pp. 91-110.

[2] Y. Ke and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," *Proc. Int. Conf. Computer Vision Pattern Recognition*, 2004, pp. II: 506-513.

[3] H. Bay, T. Tuytelaars, and L.V. Gool, "SURF: Speeded Up Robust Features," *9th European Conf. Computer Vision*, 2006, pp. 404-417.

[4] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, 2003, pp. 1615-1630.

[5] J. Shen, "Orthogonal Gaussian-Hermite Moments for Image Characterization," *Proc. SPIE Intelligent Robots Computer Vision XVI*, 1997, p. 224.

[6] L. Wang, Y. Wu, and M. Dai, "Some Aspects of Gaussian-Hermite Moments in Image Analysis," *Proc. Int. Conf. Natural Computation*, 2007, p. 450.

[7] L. Ma et al., "Local Intensity Variation Analysis for Iris Recognition," *Pattern Recognition*, vol. 37, no. 6, 2004, pp. 1287-1298.

[8] W. Shen and Y. Xiao, "Stereo Matching Based on Orthogonal Gaussian-Hermite Moments," *Proc. SPIE Int. Symp. Multispectral Image Process. Pattern Recognition*, 2009.

[9] B. Yang and M. Dai, "Image Analysis by Gaussian-Hermite Moments," *Signal Process.*, vol. 91, no. 10, 2011, pp. 2290-2303.

[10] L. Juan and O. Gwun, "A Comparison of SIFT, PCA-SIFT and SURF," *Int. J. Image Process.*, vol. 3, no. 4, 2009, pp.143-152.

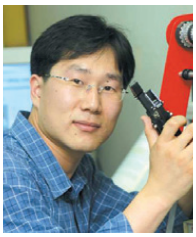
[11] http://www.robots.ox.ac.uk/~vgg/research/affine

[12] C.J. van Rijsbergen, *Information Retrieval*, Butterworth-Heinemann, London, UK, 1979.

[13] D. Lisin et al., "Combining Local and Global Image Features for Object Class Recognition," *IEEE Workshop Learning Computer Vision Pattern Recognition*, 2005.

Tae-Koo Kang received his BS in applied electrical engineering and his MS in visual image processing in 2001 and 2004, respectively, and is currently working toward his PhD in electrical engineering from Korea University, Seoul, Rep. of Korea. His research interests include computer vision, robotics, artificial intelligence, and machine learning.

Huazhen Zhang received his BS in electrical engineering from North China University of Technology, Beijing, China, in 2003. He received his MS in electrical engineering from Korea University, Seoul, Rep. of Korea, in 2009. His research interests include computer vision, robotics, artificial intelligence, and machine learning.



Dong W. Kim received his PhD in electrical engineering from Korea University, Seoul, Rep. of Korea, in 2007. Dr. Kim was a post-doctoral research scholar at BISC (Berkeley Initiative in Soft Computing), University of California, Berkeley, Berkeley, CA, USA, in 2008 and the AHMCT (Advanced Highway Maintenance and Construction Technology Research Center), University of California, Davis, Davis, CA, USA, in 2009. He is now a professor in the Department of Digital Electronics, Inha Technical College. His research interests include the intelligent humanoid robot, autonomous multi-mobile robot navigation and robot intelligence based on the neuro-fuzzy system.

Gwi-Tae Park received his BS, MS, and PhD in electrical engineering from Korea University, Seoul, Rep. of Korea, in 1975, 1977, and 1981, respectively. He was a technical staff member in the Korea Nuclear Power Laboratory and an Electrical Engineering faculty member at Kwangwoon University, in 1975 and 1978, respectively. He joined Korea University in 1981 where he is currently a professor in the School of Electrical Engineering. He was a visiting professor at the University of Illinois in 1984. He is a fellow of the Korean Institute of Electrical Engineers (KIEE), the Institute of Control, Automation, and System Engineers, Korea (ICASE) and advisor of the Korea Robotic Society. He is also a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Korea Fuzzy Logic and Intelligent Systems Society (KFIS).