논문 2008-45CI-6-21

푸리에 표현자의 크기와 회전 불변 특징을 에지에 대한 3 차원 정보에 응용한 고효율의 물체 인식

(High Performance Object Recognition with Application of the Size and Rotational Invariant Feature of the Fourier Descriptor to the 3D Information of Edges)

왕 실*, 진 홍 신*, 이 준 호*, 임 해 평*, 김 형 석**, 김 종 만**

(Shi Wang, Hongxin Chen, junho I, Haiping Lin, Hyongsuk Kim, and Jongman Kim)

요 약

3 차원 정보로부터 정확한 에지를 추출하고 푸리 변환하여 물체를 인식할 수 있는 고 효율의 물체 인식방법을 제안하였다. 물체의 윤곽은 인식에 유용한 많은 정보를 포함하고 있지만, 정확한 윤곽정보를 얻기가 어려우며, 정확한 윤곽정보를 얻었다고 하더라도 물체의 크기나 방향 마다 윤곽이 달라지기 때문에 물체 인식에 획기적 대안으로 활용되지 못하고 있다. 제안한 물체 인식 알고리즘은 1) 레이저 스캔 디바이스를 사용하여 얻는 3 차원 물체정보로부터 정밀한 물체 윤곽을 획득하고 2) 크기 및 회전 불변한 푸리에 표시 자를 이용하여 윤곽을 표현함으로써, 필요 데이터 베이스의 크기를 대폭 줄인다. 이렇게 얻어진 물체에 대한 푸리에 표식자 정보는 미리 준비된 푸리에 표식자 데이터 베이스로부터 최적 정합되는 물체를 찾아 인식한다. 이 알고리즘은 MPEG7 Part B의 방대한 영상 데이터 베이스를 대상으로 실험하였으며, 그에 대한 결과를 논문에 포함시켰다.

Abstract

A high performance object recognition algorithm using Fourier description of the 3D information of the objects is proposed. Object boundaries contain sufficient information for recognition in most of objects. Howeveer, it is not well utilized as the key solution of the object recognition since obtaining the accurate boundary information is not easy. Also, object boundaries vary highly depending on the size or orientation of object. The proposed object recognition algorithm is based on 1) the accurate object boundaries extracted from the 3D shape which is obtained by the laser scan device, and 2) reduction of the required database using the size and rotational invariant feature of the Fourier Descriptor. Such Fourier information is compared with the database and the recognition is done by selecting the best matching object. The experiments have been done on the rich database of MPEG 7 Part B.

Keywords : 3-D restoration, Fourier Descriptor, Signature, MPEG 7 Part B

I. Introduction

The object's shape is the primary low-level image feature, and simple for user to describe, either by giving example or by sketching. Research on the object shape is addressed in many areas of applied sciences such as computer vision, artificial intelligence and pattern recognition. To recognize the acquired object, shape analysis and representation is the first but most important issue. There are various shape representation methods exist in the literature, which can be

* 학생회원, ** 정회원, 전북대학교 전자정보공학부
(Div. of Electronic & Information Engineering Chonbuk National University)
접수일자: 2008년6월26일, 수정완료일: 2008년10월23일

divided mainly into two approaches, namely, the region-based[16] and the contour-based[17] approach.

Compared with region-based shape representation, contour-based methods are more popular. The contour-based techniques tend to be more efficient for handling shape that is describable by solely boundary. Among the contour-based methods, three ones are more popular, namely, global shape descriptors [13], shape signatures[1~2], and spectral descriptors[14~15]. Among them, global shape descriptors can only discriminate shapes with large dissimilarities, therefore, it is usually suitable for filtering purpose.

Shape signatures have been developed to facilitate retrieval, however, they are sensitive to noise and require intensive computation. To make the retrieval engine robust, shape signatures should be expressed again with further processing using spectral descriptors such as Fourier Descriptors(FD) and Wavelet Descriptors(WD) which are two most important approaches in spectral descriptors[18]. Among them, Fourier Descriptors are widely used in shape analysis for particular applications[19]. Wavelet Descriptors involve intensive computation in the matching stage since they are not rotation invariant[20].

The proposed system is designed for robots working in real time that requires fast processing speed. Thus the FD method is utilized to enhance shape signatures instead of WD. Differently from previous approaches, our system acquires the object shape not from images captured by camera or other 2-D approaches, but from the object depth information utilizing a 3-D laser scan device we designed[5~6]. Since depths of objects and their background are significantly different, setting a certain distance as the thresholding, even the tiny details of object shape can be precisely obtained.

For the image retrieval, a rich database is also a very important aspect. Huang et. al used complex coordinates as shape signature and proposed a two-stage image retrieval system[10]: the outline-based image retrieval and the hash-table-based image retrieval. However, due to the too small database composed of only cartoons, the retrieval result is largely irrelevant shapes. Although detailed data about comparison on Precision and Recall are reported[11], the same shortcoming still exists as that in Huang et. al's work, that is the database is too small. To make the database more credible, the MPEG-7 Part B that contains 1400 standard shapes is employed in this paper.

In the section Ⅱ, relevant technologies are described. The proposed shape retrieval system is in the section Ⅲ. Also, the experimental results and the conclusion are presented at the section Ⅳ and Ⅴ respectively.

Ⅱ. Relevant Technologies

1. Fourier Descriptors

In this section, several representative FD methods are described. For the shape signatures referred in this paper, we assume the boundary can be denoted as an ordered sequence of N coordinate points, $((x(t),y(t)),\ t=0,1,\ldots,N-1)$ where t usually means arc length.

(1) Complex coordinates

Complex coordinates is simply the complex number generated from the boundary coordinates:

$$z(t)=x(t)+iy(t) \tag{1}$$

In order to eliminate the effect of bias, shifted boundary coordinates are used:

$$z(t)=[x(t)-x_c]+i[y(t)-y_c] \tag{2}$$

where (x_c, y_c) is the centroid of the shape defined in Eq. 3.

$$x_c=\frac{1}{N}\sum_{t=0}^{N-1}x(t)\ ,\ y_c=\frac{1}{N}\sum_{t=0}^{N-1}y(t) \tag{3}$$

We use the Fourier coefficients, $U(n)$ shown in Eq. 4, to define a set of FD.

$$U(n) = \frac{1}{N}\sum_{t=0}^{N-1} z(t)\exp(\frac{-j2\pi nt}{K}),\ n = 0,1...N-1 \quad (4)$$

The invariant feature vector also given as:

$$f = [\frac{|U(2)|}{|U(1)|}, \frac{|U(3)|}{|U(1)|}, \ldots, \frac{|U(N-1)|}{|U(1)|}]^T \quad (5)$$

Translation invariance is achieved by discarding the first one (n=0), scale invariance is brought by dividing the other coefficients by the second one (n=1), and rotation invariance is achieved by using only absolute values of the coefficients.

(2) Curvature function

The curvature function used in [7] is defined as the differentiation of successive boundary angles by

$$K(t) = \theta(t) - \theta(t-1),\ \text{and}$$
$$\theta(t) = \arctan\frac{y(t) - y(t-w)}{x(t) - x(t-w)} \quad (6)$$

Where w, an integer, is a jump step used in practice. However, $\theta(t)$ can only assume values whose domain is 2π, usually in the domain of $[-\pi,\pi]$ or $[0,2\pi]$ Thus $\theta(t)$ in general contains discontinuities of size 2π. Because of this, a cumulative angular function $\psi(t)$ defined in [8] is introduced to overcome the discontinuity.

$$\phi(t) = [\theta(t) - \theta(0)]\mathrm{mod}(2\pi) \quad (7)$$

Therefore, in this paper we use

$$K(t) = \phi(t) - \phi(t-1) \quad (8)$$

Curvature signature is real valued, there are only N/2 different frequencies in the result of Fourier transformation. Therefore, only half of the FD is needed for the invariant feature vector, and f is the same with Eq. 14.

(3) Cumulative angular function

In according with human intuition, a “circle” is shapeless, thus, a normalized cumulative angular function ψ is used for shape signature. From Eq. 7, we can get

$$\psi(t) = \phi(\frac{L}{2\pi}t) \quad (9)$$

Therefore, ψ always starts at zero and ends at -2π. To make ψ a periodic function, a linear term that starts at zero and ends at $+2\pi$ is then added to the cumulative angular function. This periodic function is termed as the cumulative angular deviant function and is defined as

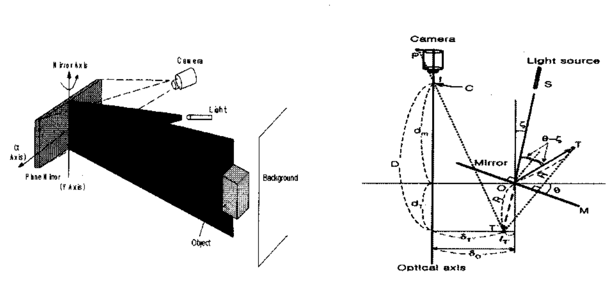
$$\psi^*(t) = \phi(\frac{L}{2\pi}t) + t \quad (10)$$

The function $\psi^*(t)$ measures the way in which the shape in question differs from a circular shape[8]. The periodic cumulative angular function defined in Eq. 10 is itself invariant under translations, rotations and scales, which was proved by Eric Persoon in [9]. Thus the FD derived from this signature can be used directly for shape representation. And only half of the frequency including the DC component is needed for the feature vector f.

$$f = [|U(0)|, |U(1)|, \ldots, |U(N/2)|]^T \quad (11)$$

2. 3-D depth measurement

A laser scan system is used to obtain the shapes in this work. Fig. 1 shows a depth measurement system and its triangulation geometry we proposed in [6]. The system has a single vertical laser stripe projected to the rotating mirror, and then



(a) (b)

그림 1. 레이저 광 투사와 거울 반사애 의한 거리 측정법의 예. (a) 3 차원 거리 측정장치. 2) 거리 측정의 원리

Fig. 1. Illustration for the depth measurement with projection and mirror reflection. (a) a Depth measurement system, (b) Schematic of the depth measurement system.

reflected onto the scene. When the mirror rotates, laser will scan over the object surface. And the laser trace in the image formed by the same mirror will also be captured by the CCD sensor. Without losing the generality, we focus on the image formation of a single light point. To derive equations in 3-D space, let us use the cylindrical coordinate system with the mirror axis as the Z-axis. Assume the light point T has its image on the CCD sensor at P $=(P_x, P_y)$. The distance R from Z-axis to point T can be derived:

$$R=\frac{\frac{f}{\delta_{cell}}\delta_0 - n_x d_m}{\frac{f}{\delta_{cell}}sin\xi + n_x \cos\xi} \tag{12}$$

where the n_x is the pixel number corresponding to P_x, which is the distance from the center of image to P and δ_{cell} is the inter-cell distance on the CCD sensor.

Since all the parameters are fixed, different pixel number n_x corresponds to different distance. Note that the mirror rotating angle is not involved in Eq. 12 for depth computation but the angle ξ must be determined through careful calibration.

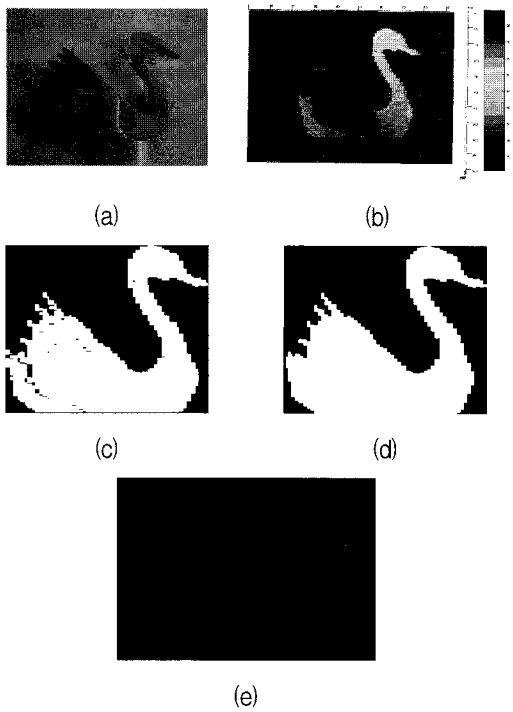


그림 2. 물체 윤곽 획득 과정 (a) 테스트 한 물체 (b)3 물체의 3 차원 구조 (c) 영역화 및 2 진화 (d) 얻어진 윤곽선

Fig. 2. Process to get the boundary. (a) Object for testing, (b) 3-D restored shape, (c) Segmentation and binarisation, (d) Component labeling,(e) Contour tracking.

III. Proposed Depth-based Shape Retrieval

1. Depth-based boundary detection

The precise and integrated boundary is so important that all the following steps such as matching and indexing in the retrieval system are based on it. Fig. 2 shows the process to obtain the object boundary in our system. The first step is to get the depth information and do 3-D restoration as shown in Fig. 2(b). Then just a certain distance as thresholding is needed to segment the object out of the background as shown in Fig. 2(c). At last, we obtain the boundary that contains all the details through contour tracking as shown in Fig. 2(e). The advantage of our system is that: since the laser trace keeps bright and obvious at all the time, no matter in dim environment or similar with the background, even when the image gets blurred, it still performs well.

2. Shape representation

Suppose we have already obtained the precise object shape, and then the FD derived from centroid distance signature are used to represent object shape. Our basic idea is to, first, represent a closed curve as a periodic function of one variable, and then express it using FD.

The centroid distance function is expressed by the distance of the boundary points from the centroid of the shape. Obviously, this conversion makes the shape representation invariant to translation.

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2} \tag{13}$$

We assume the signal $r(t)$ is periodic with period N, and then apply Eq. 4 on it. Because the signal is real, there are only N/2 different frequencies in the result of Fourier transformation (magnitude of the frequency is symmetric). Scale invariance is obtained by dividing the magnitude values of the FD coefficients by the DC component (n=0), and rotation invariance is achieved by taking only the magnitude values of the FD. Therefore, the invariant feature vector f just consists of the first few coefficients in the following,

$$f = [\frac{|U(1)|}{|U(0)|}, \frac{|U(2)|}{|U(0)|}, \ldots, \frac{|U(N/2)|}{|U(0)|}]^T \quad (14)$$

Till now, we have already transformed the 2-D contour to a sequence of 1-D signal, and just comparing the signals, one can easily judge the similarity between different objects. Assume a model shape indexed by FD feature f_m and a data shape indexed by FD feature f_d, since both features are normalized as to translation, rotation and scale, the Euclidean distance between the two feature vectors can be used as the similarity measurement.

$$d = (\sum_{i=0}^{N_c} |f_m^i - f_d^i|)^{1/2} \quad (15)$$

where N_c is the truncated number of harmonics needed to index the shape.

IV. Experimental Results

1. Evaluation of different FD methods

First we give the evaluation of different FD methods derived from complex coordinate, curvature function, cumulative angular function, and centroid distance function referred above. As known, to give the convincing performance measure, a standard and rich database is indispensable. In recent years,

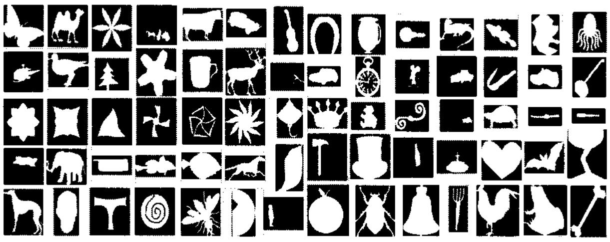


그림 3. MPEG-7 Part B 의 데이터 베이스 예
Fig. 3. Illustration of Database MPEG-7 Part B.

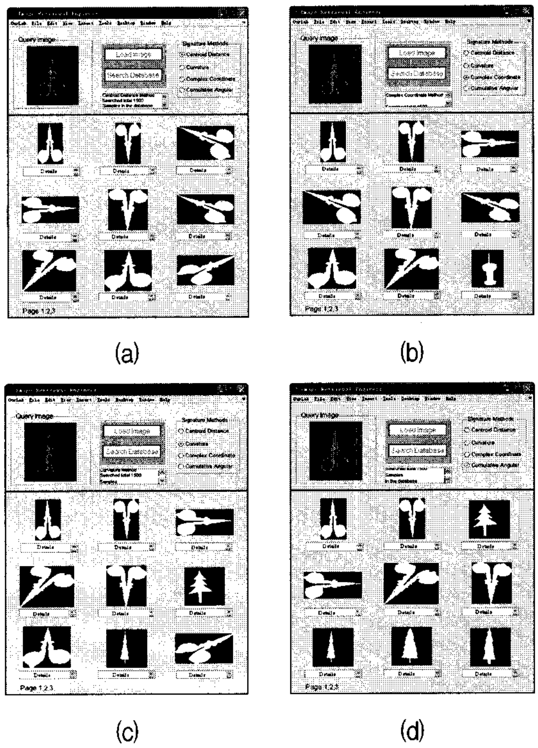


그림 4. 4 종류의 푸리에 지시자들에 의해 표현된 물체들에 대한 인식 예 (a)중심으로부터 거리 방법 (b)복소 좌표축 방법 (c) 곡률 함수 방법 (d) 누적 각도 함수 방법
Fig. 4. Retrieval example of the objects with 4 different FD methods. (a) Centroid distance function, (b) Complex coordinates, (c) Curvature function, (d) Cumulative angular function.

especially after 2005, some shape retrieval such as [12] used MPEG-7 Part B as the standard database. MPEG-7 develops a hierarchical structure of audio-visual descriptors. The Part B contains 1400 reference shapes that are big enough as a database for image retrieval. In this paper, we use these 1400 binary shapes grouped into 70 classes, each class contains 20 similar shapes as shown in Fig. 3.

In addition, 100 shapes of 5 classes presented in Fig. 6 are also added into the database, thus the database created in this way contains 1500 shapes of 75 different categories, makes the evaluation much more convincing. Fig. 4 shows certain retrieval results with different FD methods. Notice that the query image, a scissors on the top left corner, is the real object given for intuitionistic feeling, the 3-D model obtained by our laser scan system and the corresponding boundary are acquired during the preprocessing step in the background.

Compared with these four figures, they show different performance: the retrieval objects are not all scissors. There are some shapes looked similar with the query image such as the nail and trees appeared in the result. Therefore, Precision and Recall are used for quantitative measuring.

Precision is the fraction of a retrieval output that is relevant for a particular query. In our study, Precision, P, is defined as:

$$P = \frac{Sum\ of\ the\ retrieved\ relevant\ shapes}{Sum\ of\ the\ retrieved\ shapes} \times 100\% \quad (16)$$

Recall is the fraction of the shapes that are relevant to the query that are successfully retrieved. It requires knowledge not just of the relevant and retrieved but also those not retrieved. Therefore, Recall, R, is defined as:

$$R = \frac{Sum\ of\ the\ retrieved\ relevant\ shapes}{Sum\ of\ the\ relevant\ shapes\ in\ the\ database} \times 100\% \quad (17)$$

Precision can be seen as a measure of exactness or fidelity, whereas Recall is a measure of completeness.

Fig. 5 is the evaluation of these four methods using Precision and Recall.

From Fig. 5, obviously, the retrieval performance of

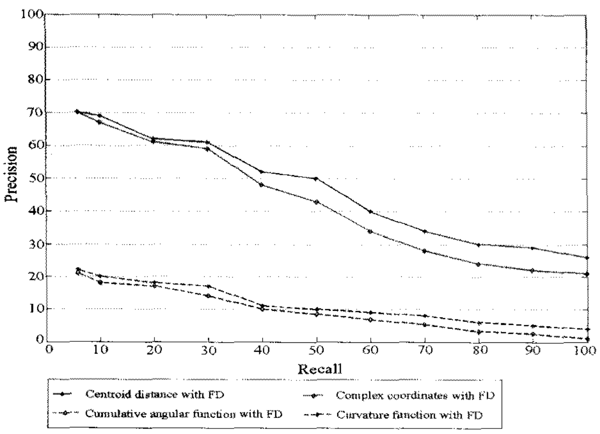


그림 5. 4 종류의 푸리에 지시자들로 표현된 물체에 대한 인식성능 비교
Fig. 5. Comparison of the evaluation performance of 4 FD methods.

centroid distance with FD is the best one among the four. (And this is the reason why we choose it for our application.) The second is the complex coordinates method, and it has comparable performance with centroid distance method. However, the last two, curvature function and cumulative angular function, give the incompetent result. As referred, centroid distance is an approach among contour-based shape representation. Whereas, it also has the character of region-based shape representation, since the distance from the centroid to boundary samples can be also seemed as regional information. However, curvature function and cumulative angular function, they are derived only from the boundary and just capture local feature of the shape.

1. Comparisons between the conventional 2-D and the 3-D depth-based methods

Since our system is designed for indoor robots, 100 shapes with 5 different classes are added. They are cup, scissors, toy bird, mouse and mobile telephone which are familiar objects indoors. Fig. 6 shows their binary shapes. Please note that each

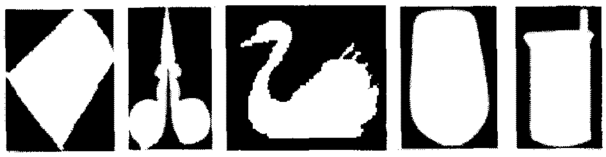
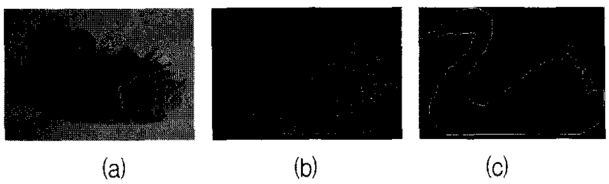


그림 6. 사용된 실내 물체들
Fig. 6. Classes of indoor objects added.



(a) (b) (c)

그림 7. 제안한 3D 정보를 사용한 방법과 기존의 2D 영상을 사용한 방법들에 의한 에지 검출 성능 비교 (a) 원 영상 (b) 2D 영상을 이용한 에지 검출 (c) 3D 거리정보를 이용한 에지 검출
Fig. 7. Comparison of edge detection performance between the proposed 3-D depth-based method and the conventional 2-D image-based one. (a) Original image, (b) the method with conventional 2-D image-based one, (c) the method with the 3-D image-based one.

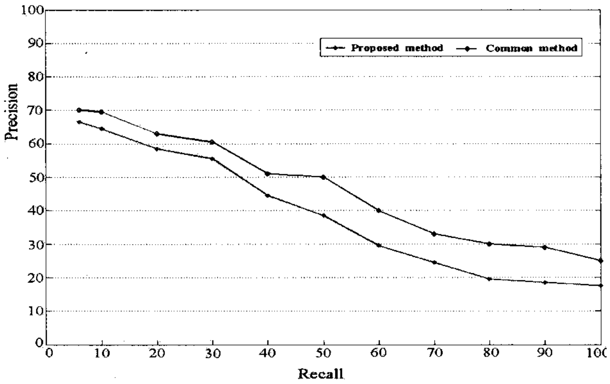


그림 8. 2D 영상정보와 제안한 3D 정보를 이용한 방법들에 의한 인식률 비교

Fig. 8. Comparison of the recognition performance between with 2-D image-based and with 3-D depth-based.

class contains 20 similar shapes with different poses.

There are many approaches to detect object edge in 2-D domain. However, if images are taken in extraordinary poor environment, like similar color with background or dim environment with shadows as shown in Fig. 7(a), it is hard to get the complete boundary. Fig. 7 shows the comparison of the previous edge detection technique to the proposed 3-D depth information-based method.

In Fig. 7(b), edge misses at the left side due to the object incorporation with shadows, while the complete boundary is detected using our depth-based method as shown in Fig. 7(c). Precision and Recall are given again for quantitative measuring the performance as in Fig. 8.

It is clear from Fig. 8, our performance is better than the conventional 2-D methods. Since this recognition technology is based on object shape, the more accurate shape means the better retrieval performance.

V. Conclusions

In this paper, we proposed a shape retrieval system using a laser scan equipment to obtain the query shape. For engineering applications, our retrieval engine as well as the laser scan equipment can be easily installed into indoor robots. Compared with conventional 2-D image-based methods, our proposed approach enhances the robustness for robots to index and recognize the query object against poor working environment. Moreover, we have measured the performance of four mostly used shape signatures expressed with FD. Our database built with MPEG-7 Part B and certain real objects makes the evaluation much more convincing.

References

[1] E.R. Davies, Machine Vision : Theory, Algorithms, Practicalities, Academic Press, 1997.

[2] P.J. van Otterloo, A Contour-Oriented Approach to Shape Analysis, Prentice-Hall International Ltd, 1991.

[3] Quang Minh Tieng and W. W. Boles, "Recognition of 2D Object Contours Using the Wavelet Transform Zero-Crossing Representation," IEEE Trans. on PAMI 19(8), pp. 910-916, Aug. 1997.

[4] Hee Soo Yang, Sang Uk Lee, and Kyoung MuLee, "Recognition of 2D Object Contours Using Starting-Point-Independent Wavelet Coefficient Matching," Journal of Visual Communication and Image Representation, vol. 9, no. 2 , pp. 171-181, Jun. 1998.

[5] Hyongsuk Kim, Chun-Shin Lin, Chang-Bae Yoon, Hye-Jeong Lee, and Hongrak Son, "A Depth Measurement System with the Active Vision of the Striped Lighting and Rotating Mirror,"Lecture Note in Computer Sciences: Progress in Pattern Recognition, Image Analysis and Applications, vol. 3287, pp. 108-115, Oct. 2004.

[6] Hyongsuk Kim, Chun-Shin Lin, Jaehong Song, and Heesung Chae, "Distance Measurement Using a Single Camera with a Rotating Mirror,"International Journal of Control, Automation, and Systems, vol. 3, no. 4, pp. 542-551, Dec. 2005.

[7] Hannu Kauppinen, Tapio Seppanen and Matti Pietikainen, "An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 17, no. 2, pp. 201-207, 1995.

[8] Charles T. Zahn and Ralph Z. Roskies, "Fourier Descriptors for Plane closed Curves," IEEE Trans. on Computer, vol. 21, pp. 269-281, 1972.

[9] Eric Persoon and King-sun Fu, "Shape

Discrimination Using Fourier Descriptors," IEEE Trans. on Systems, Man and Cybernetics, vol. Mar, pp. 170-179, 1977.

[10] Chung-Lin Huang and Dai-Hwa Huang, "A Content-based image retrieval system," Image and Vision Computing, vol. 16, no. 3, pp. 149-163, 1998.

[11] Atul Sajjanhar. A Technique for Similarity Retrieval of Shapes. Master thesis, Monash University, Australia, 1997.

[12] Bin Wang and Chaojian Shi, "A Novel Fourier Descriptor for Shape Retrieval," Lecture Note in Computer Sciences: Fuzzy Systems and Knowledge Discovery, vol. 4223, pp. 822-825, Sep. 2006.

[13] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, G. Taubin, The QBIC project: querying image by content using color, texture and shape, in: Proceedings of SPIE Storage and Retrieval for Image and Video Databases, vol. 1908, pp. 173-187, 1993.

[14] C.T. Zahn and R.Z. Roskies, Fourier descriptors for plane closed curves, IEEE Transactions on Computer c-21, pp. 269-281, 1972.

[15] Q.M. Tieng and W.W. Boles, Recognition of 2D object contours using the wavelet transform zero-crossing representation, IEEE Transactions on PAMI 19, Aug. 1997.

[16] W.-Y. Kim and Y.-S. Kim, A region-based shape descriptor using Zernike moments, Signal Processing: Image Communication 16, pp. 95-102, 2000.

[17] K. L. Kashyap and R. Chellappa, Stochastic models for closed boundary analysis: representation and reconstruction, IEEE Transactions on Information Theory 27, pp. 627-637, 1981.

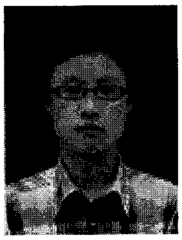
[18] B. Dorner, Hand Shape Identification and Tracking for Sign Language Interpretation, IJCAI Workshop on Looking at People, 1993.

[19] Kindratenko V.V, On Using Functions to Describe the Shape, Journal of Mathematical Imaging and Vision, Volume 18,Number 3, pp. 225-245(21), May 2003.

[20] I. Kunttu, L. Lepisto, and A. Visa, Efficient Fourier shape descriptor for industrial defect images using wavelets, Optical Engineering, vol. 44, 2005.

저 자 소 개

왕 실(학생회원)
2004년 Wuhan University of Technology, China, 졸업(학사)
2006년 Huazhong University of Science and Technology, China, 졸업(석사)
2006년～현재 전북대학교 일반대학원 전자정보공학부 박사과정
<주관심분야: 멀티미디어와 통신 시스템, 로봇비젼, 원격제어>



진 홍 신(학생회원)
2004년 Wuhan University of Technology, China, 졸업(학사)
2007년 Wuhan University of Technology, China, 졸업(석사)
2007년～현재 전북대학교 일반대학원 전자정보공학부 박사과정
<주관심분야: 멀티미디어와 통신 시스템, 로봇비젼, 원격제어>

이 준 호(학생회원)
1998년 전북산업대학교 기계공학과 졸업(학사)
2000년 원광대학교 트라이볼로지 졸업(석사)
2004년～현재 전북대학교 학과간 협동 메카트로닉스공학과 박사과정
2007년～현재 진우SMC 회사 경영
2008년～현재 전북대학교 산학커플링사업단 겸임교수
<주관심분야: 유압 최적화 회로 설계, 로봇비젼>

임 해 평(학생회원)
2005년 South-central University for Nations, China, 졸업(학사)
2007년 전북대학교졸업(석사)
2007년～현재 전북대학교 전자정보공학부 박사과정
<주관심분야: 로봇비젼, 아날로그 병렬처리 회로>

김 형 석(정회원)
2002년 University of Missouri, Columbia, 졸업(박사)
2003년～현재 전북대학교 전자정보공학부 교수
<주관심분야: 로봇비젼, 로봇센서시스템, 아날로그 병렬처리 회로 >

정 길 도(정회원)
1992년 미국 Texas A&M University 기계공학 박사.
1993년～1995년 영남대학교 전임강사.
1995년～현재 전북대학교 전임강사, 부교수, 교수.
<주관심분야: Time-Delay, 실시간 멀티미디어 전송, Web 기술, Robotics.>