

## 8—24

## Probabilistic Model-based Object Recognition using Local Zernike Moments

Sung-Ho Kim and In-So Kweon<sup>1</sup>Dept. of Electrical Eng. & Computer Science  
KAISTIn-Cheol Kim<sup>2</sup>

Agency for Defense Development

## Abstract

Object recognition is one of the most important, yet the least understood, aspect of visual perception. The difficulties originate from the variations of objects such as view position, illumination changes, background clutter, occlusion and etc. In this paper, we present an object recognition paradigm robust to these variations using modified local Zernike moments and the probabilistic voting method. We propose a feature which is robust to scale, rotation, illumination change and background clutter. A probabilistic voting scheme maximizes the conditional probability defined by the features in correspondence to recognize an object of interest. Results from the experiments show the robustness of the proposed system.

## 1 Introduction

One of the most important tasks for an intelligent service robot is to identify objects of interest in indoor environment. If a mobile robot is commanded to bring a certain object, it has to recognize what objects are in the scene in advance.

Object recognition is defined as the process of extracting information such as name, size, position, pose, and functions related to the object. In this paper, we restrict the definition to extracting object's identification and pose information only.

Recently, there has been much development in object recognition, but it still remains on the level of recognizing object in a well controlled environment. This is originated from the object variations such as view angle changes and illumination changes. Further more, it becomes a difficult task when an object is occluded by others or placed in a cluttered environment [1-2]. To solve these problems, various approaches have been proposed using invariant, 3D CAD model and appearance [3-5]. Recently, reflecting the characteristic of the human visual object recognition methods based on local image such as local differential invariants, SIFT (scale invariant feature transform) and eigen window have been suggested [6-9]. But these approaches have shown some limited success to some problems such as illumination change.

We propose a novel framework of object recognition for a service robot to work in indoor environments. To solve the object variations, we propose modified local Zernike moments robust to illumination variations and pose changes.

## 2 Model-based Object Recognition

## 2.1 Proposed object recognition system

The object recognition system is composed of the robust feature extraction part and the feature matching part. Figure 1 shows the overall system of object recognition.

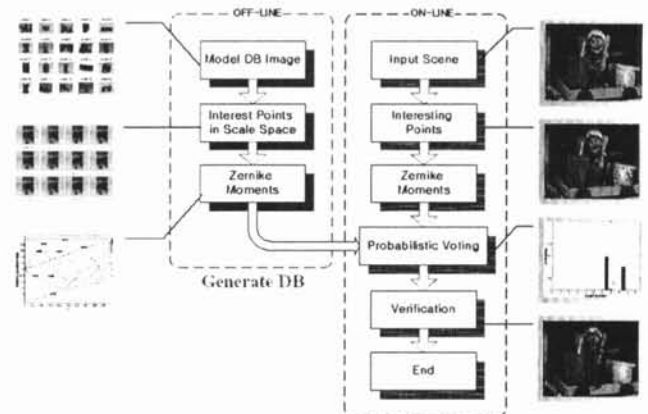


Fig. 1. The proposed object recognition system

In off-line process, Zernike moments are calculated around interest points detected from the scale space of image model and are stored in a database. The image model of an object is the front view of the object to be recognized. We can recognize the object by probabilistic voting of these Zernike moments in on-line process. The locality of the Zernike moment provides some robustness to occlusion and background clutter. We verify the recognition by aligning model features with the input scene. In this process, the homography between the image model and the input scene is calculated. We determine the success of the recognition by the percentage of the outlier which is determined by the distance between scene feature and model feature transformed by homography.

## 2.2 Robust local feature: Zernike moments

The proposed object recognition method is based on the fact that the human visual system concentrates on a certain interesting points during recognition [10]. Interesting points of model object are placed on the same position of scene object. This is called the repeatability of interesting points. We use the Harris corner detector which has shown the superior repeatability [11-12].

We use Zernike moments to represent local characteristics of an image segment. Zernike moments are defined over a set of complex polynomials which

<sup>1</sup> Address: 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea. E-mail: {kshrcv, iskweon}@kaist.ac.kr

<sup>2</sup> Address: Yuseong P.O. BOX 35, Daejeon 305-152, Korea. E-mail: cheol@add.re.kr

form a complete orthogonal set over the unit disk  $x^2 + y^2 \leq 1$  [13]. Zernike moments are the projections of the image intensity  $f(x, y)$  onto the orthogonal basis functions  $V_{nm}(x, y)$ .

$$A_{nm} = \frac{n+1}{\pi} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) [V_{nm}(x, y)]^* \quad (1)$$

where

$$V_{nm}(x, y) = R_{nm}(x, y) \exp(jm \tan^{-1}(y/x)) \quad (2)$$

The radial polynomial is defined as

$$R_{nm}(x, y) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} (x^2 + y^2)^{(n-2s)/2} \quad (3)$$

with the conditions  $(n - |m|) : \text{even}, |m| \leq n$ .

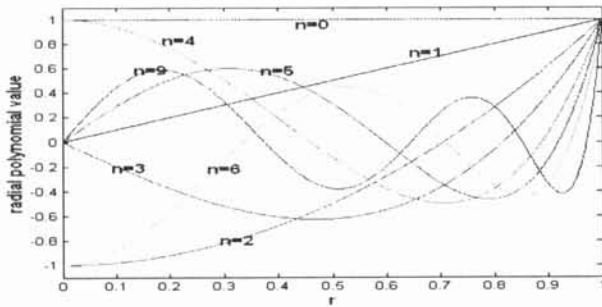
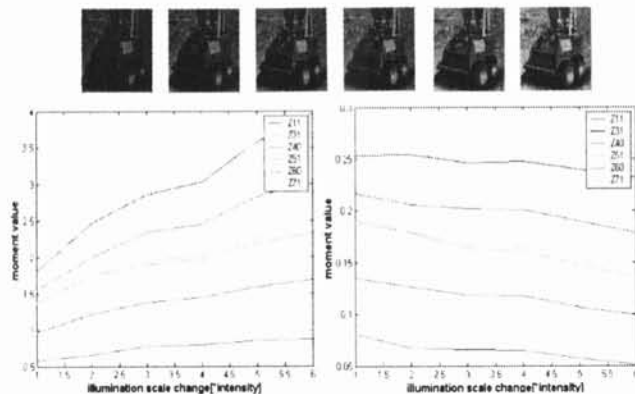


Fig. 2. Radial polynomials of order  $n=0, 1, \dots, 9$

As the Zernike moments are calculated using the radial polynomials shown in figure 2, they have inherent rotation invariant property. Especially, Zernike moments have superior properties in terms of image representation, information redundancy and noise characteristics [14]. But they are sensitive to scale and illumination changes. We reduce the scale problem by applying the scale space theory to image model [15]. This method makes the corner extraction process more effective than using the image pyramid.



(a) Without illumination invariant (b) With illumination invariant

Fig. 3. Zernike moments and illumination changes

The problem of illumination change is simply solved by normalizing the moments by the  $Z_{00}$  moment which is equivalent to the average intensity. Since local illumination change can be modeled as

$$f'(x_i, y_i) = a_L f(x_i, y_i) \quad (4)$$

we define an illumination invariant feature as

$$\frac{Z(f'(x, y))}{m_{f'}} = \frac{a_L Z(f(x, y))}{a_L m_f} = \frac{Z(f(x, y))}{m_f} \quad (5)$$

where  $f(x, y)$  represents the intensity at  $(x, y)$ ,  $f'(x, y)$  is the new intensity after illumination change,  $a_L$  means the rate of illumination change,  $m_f$  denotes the local average intensity and  $Z$  means Zernike moment operator.

Figure 3 shows the robustness of the modified Zernike moments to illumination changes. We can observe that the normalized Zernike moments remain approximately constant even under illumination changes.

### 2.3 Probabilistic voting

In this paper, we propose a recognition method based on probabilistic voting which takes into account the stability of Zernike moments of the image model and the similarity of Zernike moments between the image model and the input scene.

Object recognition using probabilistic voting means finding a model  $M_i$  that maximizes the conditional probability as

$$\arg \max_{M_i} P(M_i | S) \quad (6)$$

where  $S$  represents the input scene.

For each input feature, we form a set of matching pairs consisting of the corresponding model features whose Zernike moments are similar to that of the input feature. For  $k$ -th input feature, a set of matching pairs is

$$MP_k = \{(Z_k, \hat{Z}_{ki}) | Z_k \in S, \hat{Z}_{ki} \in M, \|Z_k - \hat{Z}_{ki}\| < \varepsilon\} \quad (7)$$

$$i = 1, 2, \dots, N_C$$

where  $Z_k$  denotes the Zernike moments of the  $k$ -th input feature,  $\hat{Z}_{ki}$  means the Zernike moments of the model feature,  $M$  is a set of models and  $N_C$  is the number of corresponding model features.

The hypothesis is then given by

$$H_h = \{(Z_i, \hat{Z}_i) | (Z_i, \hat{Z}_i) \in MP_i, i = 1, 2, \dots, N_S\} \quad (8)$$

where  $N_S$  is the number of interest points of input scene. The set of total feature pairs can be written as

$$H = \{H_1 \cup H_2 \dots \cup H_{N_H}\} \quad (9)$$

where  $N_H$  is the size of product space as  $MP_k \times H_h$ . Since  $H$  is composed of the model Zernike moments corresponding to the input scene  $S$ , we can substitute  $H$  for  $S$ . By the Bayes' theorem, equation (6) becomes

$$P(M_i | S) = P(M_i | H) = \frac{P(H | M_i) P(M_i)}{P(H)} \quad (10)$$

If all the objects are equally probable and independent mutually, equation (10) becomes

$$P(M_i | H) = \frac{\sum_{h=1}^{N_H} P(H_h | M_i) P(M_i)}{\sum_{h=1}^{N_H} P(H_h)} \quad (11)$$

By the theorem of total probability, the denominator can be written as

$$P(H_h) = \sum_{i=1}^{N_M} P(H_h | M_i) P(M_i) \quad (12)$$

The term  $P(H_h | M_i)$  is computed by considering the stability and similarity measure of matching pairs. Basically,  $P(H_h | M_i)$  has to be large when both the stability and the similarity measures are high. The stability measure reflects the incompleteness of repeatability of interest points and the similarity measure reflects the closeness of feature vectors in Euclidean space.

- (1) Stability ( $\omega_S$ ): The stability of Zernike moments is inversely proportional to the sensitivity which is the standard deviation of Zernike moments calculated at four neighboring positions around an interesting point ( $\hat{Z}_{i,j}$ ). The smaller the sensitivity is, the more stable Zernike moments of the image model are.

$$\omega_S(\hat{Z}_{i,j}) = \frac{1}{\sqrt{\frac{1}{4} \left[ \|\hat{Z}_{i,j} - \hat{Z}_{i,j}\|^2 + \|\hat{Z}_{i,j-1} - \hat{Z}_{i,j}\|^2 + \|\hat{Z}_{i,j+1} - \hat{Z}_{i,j}\|^2 + \|\hat{Z}_{i+1,j} - \hat{Z}_{i,j}\|^2 \right]}} \quad (13)$$

- (2) Similarity ( $\omega_D$ ): The similarity of a feature pair is inversely proportional to the Euclidean distance between the Zernike moments of input scene and that of the corresponding image model.

$$\omega_D(Z_j, \hat{Z}_k) = \frac{1}{\|Z_j - \hat{Z}_k\|} \quad (14)$$

Therefore,  $P(H_h | M_i)$  is defined as

$$P(H_h | M_i) = \prod_{j=1}^{N_S} P((Z_j, \hat{Z}_k) | M_i) \quad (15)$$

where

$$P((Z_j, \hat{Z}_k) | M_i) = \begin{cases} \exp\left(-\frac{1}{\omega_S \omega_D \alpha}\right) & \text{if } \hat{Z}_k \in \hat{Z}(M_i) \\ \varepsilon & \text{else} \end{cases} \quad (16)$$

$\alpha$  is the normalization factor for the conditional probability to have proper value and  $\varepsilon$  is assigned as a penalty if the corresponding model feature doesn't belong to a certain model. We use the approximate nearest neighbor search algorithm to find matching pairs [16]. It takes long time for linear search space.

## 2.4 Recognition verification

Recognition results are verified using feature pairs. We find optimal feature pairs by rejecting outliers using the area ratio which is preserved under the affine transformation.

For given four points ( $P_1, P_2, P_3, P_4$ ) shown in figure 4, we calculate the area ratio  $S_2/S_1$  in the image model and  $S_2'/S_1'$  in the input scene. If the two ratios differ with by a

predetermined threshold, we reject the fourth feature point. We assume the first three points are matched.

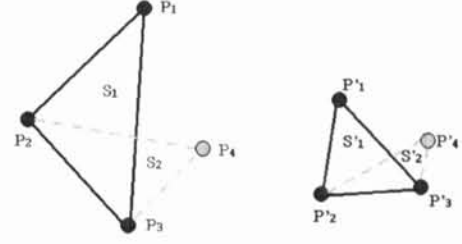


Fig. 4. Rejection of outliers using area ratio: (a) Local feature of model, (b) Local feature of scene

Then we calculate an initial homography based on these optimal feature pairs randomly selected. A LMedS based method selects an optimal homography from feature pairs. We can also determine the percentage of outliers by applying this homography to the remaining image model points.

## 3 Experimental results

Figure 5 shows the image model consisting of twenty objects.



Fig. 5. Image models used for object recognition

We have tested the proposed system using various models shown in figure 5. Figure 6 shows the results of object recognition for each stage.

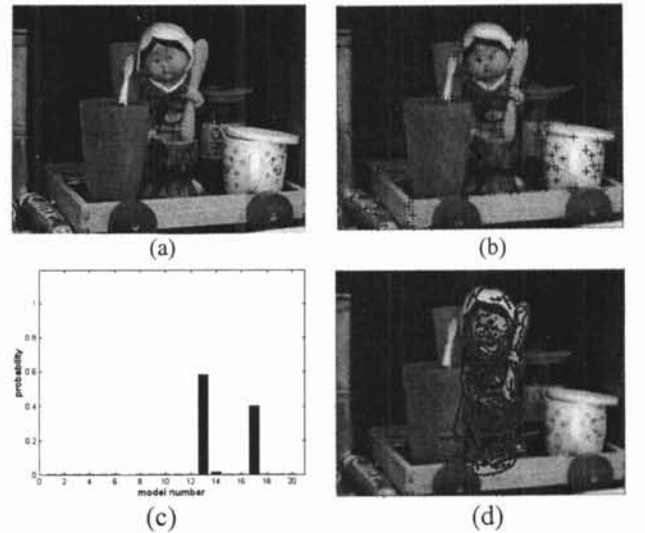


Fig. 6. Object recognition process for model-13: (a) input scene, (b) features by the Harris corner detector, (c) probabilistic voting, (d) model alignment

Figure 7 shows experimental results when cluttered background, illumination change and occlusion exist.

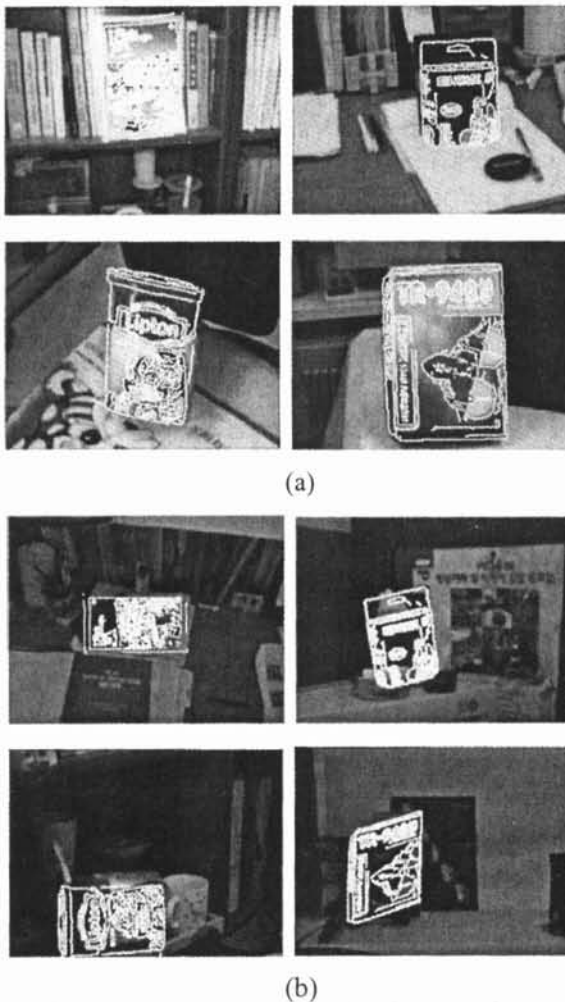


Fig. 7. Object recognition results for images taken under various conditions: (a) pose + background clutter changes, (b) pose + background clutter + occlusion + illumination changes

Table 1 shows the statistical results of the object recognition for model-10. It has failed to recognize some scenes where high specularity, blurring, background clutter and low illumination exist.

Table 1. Statistical results of the object recognition for model-10

Model #	The number of trials	The number of success	Recognition rate [%]
10	118	105	88.9

#### 4 Conclusions

In this paper we have proposed a practical object recognition system. Main contributions of the proposed system are two folds:

First, the normalized local Zernike moments has shown the robustness to view position, illumination change, occlusion and background clutter.

Second, a probabilistic voting method recognizes objects based on the stability and similarity of Zernike moments

The experimental results demonstrate that robust object recognition is feasible by introducing the robust Zernike moments feature into the novel probabilistic model-based recognition framework.

#### Acknowledgments

This research was supported by grants from ADD and HWRS-ERC.

#### References

- [1] S. Ullman, "High-level Vision: Object Recognition and Visual Cognition", *The MIT Press*, 1997.
- [2] C. Olson, "Fast Object Recognition by Selectively Examining Hypothesis", *Doctor's Thesis, University of California*, 1994.
- [3] K. Roh and I. Kweon, "2-D Occluded and Curved Object Recognition using Invariant Descriptor and Projective Refinement", *IEICE Transactions on Information and Systems*, vol. E81-D, No. 5, 1998.
- [4] M. Stevens, J. Beveridge, "Precise Matching of 3-D Target Models to Multisensor Data", *CSU*, Jan. 20, 1997.
- [5] D. Keren, M. Osadchy, C. Gotsman, "Antifaces: A Novel, Fast Method for Image Detection", *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 23, No.7, pp.747-761, 2001.
- [6] C. Schmid, A. Zesserman, R. Mohr, "Integrating Geometric and Photometric Information for Image Retrieval", *In International Workshop on Shape, Contour and Grouping in Computer Vision*, 1998.
- [7] D. Lowe, "Object Recognition from Local Scale-Invariant Features", *Proc. International Conference on Computer Vision, IEEE Press*, 1999.
- [8] K. Ohba, K. Ikeuchi, "Detectability, Uniqueness, and Reliability of Eigen-Windows for Robust Recognition of Partially Occluded Objects", *IEEE Pattern Analysis and Machine Intelligence* 19(1997), no.9, 1043-1048, 1997.
- [9] D. Jugessur, "Robust Object Recognition using Local Appearance based Methods", *Master Thesis, McGill Univ.*, 2000.
- [10] A. Treisman, "Features and Objects in Visual Processing", *Scientific American*, November 1986.
- [11] C. Schmid, R. Mohr, C. Bauckhage, "Comparing and Evaluating Interest Evaluation of interests", *In ICCV*, pp. 230-235, 1998.
- [12] C. Harris and M. Stehens. "A Combined Corner and Edge Detector", *In Alvey Vision Conference*, pages 147-151, 1988.
- [13] S. Abdallah, B. MengSc, "Object Recognition via Invariance", *Doctor's Thesis, The Univ. of Sydney*, 2000.
- [14] C. The, R. Chin, "On Image Analysis by the Methods of Moments", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, 1988.
- [15] T. Lindeberg, "Scale-space theory: A Basic Tool for Analysing Structures at Different Scales", *Journal of Applied Statistics*, 20, 2, pp. 224-270, 1994
- [16] S. Arya, et al., "An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions", *J. ACM* 45(6):891-923, Nov 1998.