MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Selection of Object Recognition Methods According to the Task and Object Category

Al Mansur
Department of Information and Computer Sciences, Saitama University,
255 Shimo-Okubo, Sakura-ku, Saitama-shi, Saitama 338-8570, Japan
mansur@cv.ics.saitama-u.ac.jp

Yoshinori Kuno
Department of Information and Computer Sciences, Saitama University,
255 Shimo-Okubo, Sakura-ku, Saitama-shi, Saitama 338-8570, Japan
kuno@cv.ics.saitama-u.ac.jp

Abstract

Service robots need object recognition strategy that can work on various objects in complex backgrounds. Since no single method can work in every situation, we need to combine several methods so that the robots can use the appropriate one automatically. In this paper we propose a scheme to classify situations depending on the characteristics of object of interest and user demand. We classify the situations into four categories and employ different techniques for each one. We use SIFT, kernel PCA (KPCA) in conjunction with Support Vector Machine (SVM) using intensity and Gabor feature for four categories. We show that use of intensity feature or Gabor feature is important for the use of KPCA based techniques on different kinds of objects. Through our experiments, we show that by using our categorization scheme a service robot can select appropriate feature in kernel PCA based techniques and improve its recognition performance considerably.

1 Introduction

Helper robots or service robots have attracted much attention of researchers for handicapped or aged people. We are developing a service robot that can find out a specific or a general class of object ordered by the user. The robot is instructed by user's speech and should be able to conduct two tasks: detecting a specific object and detecting a class. For example, if a user asks a robot to find a 'coke can', his/her demand is for a *specific* object. If he/she asks to find any 'can', his/her demand is for a *class* of object. The robot needs a vision system that can work on various objects in complex backgrounds to carry out the two tasks mentioned above.

There is no single object recognition method that can work equally well on various types of objects and backgrounds perceived by a service robot. It must rely on multiple methods and should be able to select the appropriate one depending on the object characteristics.

SIFT [1] is capable of detecting the exact object that the system has previously seen with an incomparable performance. Unfortunately this method generates very few or no keypoints if the objects are very plain and do not have much detail. Therefore, SIFT is not well suited to recognize such objects. SIFT is also not applicable for class recognition.

In a recent work [2], Serre, Wolf and Poggio proposed the standard model that is suitable for class recognition. Although the results are impressive for some object categories, there are some objects for which detection rate is not good enough.

In [4] Kernel PCA is used in conjunction with SVM (KPCA+SVM) to learn the view subspaces for multi-view face detection and recognition. These methods can be applied for class recognition. When KPCA+SVM is used for object recognition, feature selection is very crucial. To achieve a good recognition performance on particular class of object it is not wise to randomly select intensity or Gabor feature as we show in our experiments.

To develop an integrated object recognition platform for service robots, we split the object recognition problem into several cases depending on the task and object category. In this paper we present scenarios that have been encountered by a service robot to carry out its object recognition task and propose solutions for these challenges.

Our proposed categorization scheme enables the robot to choose an appropriate detection method. Through experiments, we show that a technique selected by the categorization scheme performs better than other techniques. We introduce the categorization scheme in section 2. In section 3 we discuss the recognition framework and feature extraction. Experimental results are shown in section 4 and we draw the conclusion of our work in section 5.

2 Object Categorization

Objects encountered by service robots can be described by their shapes and textures. By 'texture' we mean the pattern (not necessarily regular and periodic) within the object contour. For example, in our notation, the label of a bottle is its texture. Some objects are plain and do not contain any texture. Shape is also a clue to detect them. Objects of this kind will be named as category 1 objects. Recognizing a class of object or any specific object within a class of this category is not possible using SIFT. KPCA+SVM can be used in this case although it uses the same strategy for both class and specific object detection. Some classes of objects have textures although these textures do not characterize them and the texture contents of different members of the class are not the same. Some members of those classes have textureless body. As a result we need to use information of their shapes to describe them. This type of objects will be called category 2 objects. Using SIFT, any specific textured object of this category can be recognized. To recognize a textureless specific object or a class of this category we use KPCA+SVM. Since these objects are shape-based, we have to use Gabor feature because it can efficiently extract the shape feature from differently textured objects. The remaining objects are those whose texture is similar among all members and the texture is required for their recognition. There are two kinds of similar textures. In one type, intensity within the texture varies much. This happens when both dark and light grayscales exist in the texture. Grayscale histograms of these types of textures are wide. In the other kind, intensity does not vary much within the texture. This results in a narrow grayscale histogram. Textured objects with wide grayscale histograms will be named as category 3 and intensity feature based KPCA+SVM can faithfully extract this texture information. The textured objects with narrow grayscale histograms will be named as category 4 objects. Using intensity based KPCA+SVM it is difficult to model the texture with low intensity variation. In this case we use Gabor feature to extract the edges of the texture. Examples of each category are given in section 4.

2.1 Classification of Situations

In Table 1, we summarize the object categorization as discussed in the previous section. The object recognition problem has been classified into several cases depending on the task and object category.

Table 1. Categorization of an object recognition scenario.

Object category	Specific/ class	Case	Applicability		
			SIFT	KPCA+SVM	
				intensity	gabor
Category 1	specific	1			•
	class				
Category 2	specific (textured)	2	•		
	Specific (texture-less)	3			•
	class	4			•
Category 3	specific	5	•		
	class	6		•	
Category 4	specific	7	•		
	class	8			•

To categorize scenarios into one of the 8 cases, we need two kinds of information: object category and object specificity. We apply the algorithm shown in Fig. 1 to classify an object into category 1, category 2, category 3 or category 4. The robot is trained on all the objects (on which the robot works) using the algorithm prior to recognition. Finally, object specificity will be known from the robot user. Now we deploy appropriate strategies for three groups of cases as follows:

Method 1 (SIFT based): cases 2, 5 and 7
Method 2 (Gabor based KPCA+SVM): cases 1, 3, 4 and 8
Method 3 (Intensity based KPCA+SVM): case 6

To categorize a particular object into one of the four categories using the given algorithm, several sample images of that object class are required. The objects should appear in plain background. This ensures that no keypoint or feature is generated from the background. Note that this is not a recognition step and is done offline. As a result we can use images of objects with plain background.

To find the threshold we collected large number of images of plain objects. Then we extracted SIFT keypoints from each of these images and took a record of these keypoint counts (class label 1). We also counted the SIFT keypoints for non-plain objects (class label 2). Then we estimated the parameters of Gaussian mixture model for given labeled data samples and finally we constructed the decision boundary of a Bayesian classifier. This classifier has the quadratic discriminant function:

$$f(\mathbf{x}) = \langle \mathbf{x} \cdot \mathbf{A}\mathbf{x} \rangle + \langle \mathbf{b} \cdot \mathbf{x} \rangle + c$$

where the classification strategy is

$$q(\mathbf{x}) = 1 \quad if \ f(\mathbf{x}) \geq 0$$
$$= 2 \quad if \ f(\mathbf{x}) < 0$$

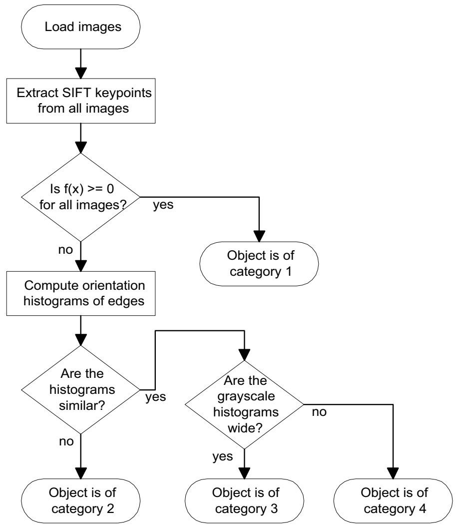
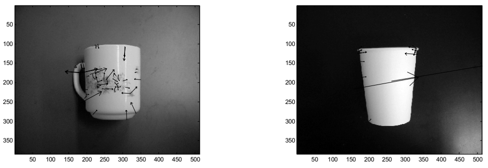


Figure 1. Object categorization algorithm

If the objects have texture, they produce large number of SIFT keypoints and we need further investigation to categorize it. To check the similarity in the texture content of the objects, we compute orientation histograms of the edges with four bins (0°,+45°,-45° and 90°). If the histograms are not similar, they will be marked as category 2. To check the similarity among orientation histograms we used another Bayesian classifier. If the orientation histograms are similar, we compute grayscale histograms and classify them into two categories according to their width. Again a Bayesian classifier is used that had been constructed offline using training examples of textured objects of both low and high contrast.

2.2 Categorization Example

When the categorization algorithm is applied to ‘cup’ (Figs. 2 (a) and 2 (b)) it has been found that some of the sample images of this category have very few keypoints while the others have large number of keypoints. As a result the discriminant function returns positive as well as negative values. The orientation histograms of different ‘cup’ images are not similar. Consequently ‘cup’ has been classified as a category 2 object. Application of the categorization algorithm to ‘keyboard’ resulted that it is not a category 1 object. Moreover orientation histograms of different ‘keyboard’ examples are similar and its grayscale histograms (256 bins) are also narrow (Fig. 3). As a result this object has been categorized into category 4.



(a) (b)

Figure 2. (a) One example of a cup have 67 keypoints (b) another example have 17 keypoints

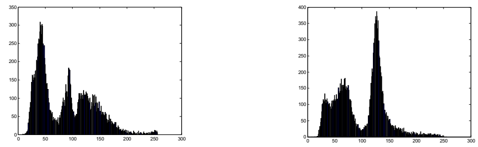


Figure 3. Narrow grayscale histograms of two ‘keyboard’.

3 Recognition Framework

In the recognition framework we use three different methods for four different object categories. Method 1 is based on SIFT keypoints. For methods 2 and 3, we use a kernel machine learning based approach for extracting nonlinear features from the object images. Kernel PCA [3,4,5] is applied to a set of labeled images to learn nonlinear view-subspaces.

3.1 Method 1

We followed [1] in this method. First, the original image is progressively filtered using Difference of Gaussian filters with σ in a band from 1 to 2 resulting in a series of Gaussian blurred images. This processing produces a scale space representation. Then, these images are subtracted from their direct neighbors (by σ) to produce a new series of images. Each pixel in the images is compared to its 8 neighbors and the 9 pixels each of the other pictures in the series. Then keypoints are chosen from the extrema in scale space. To derive the SIFT keypoint descriptors for each keypoint, histograms of gradient directions are computed in a 16x16 window using bilinear interpolation.

3.2 Method 2

We apply a battery of Gabor filters to each of the training and test images to extract the edges oriented in different directions. These filters come in 4 orientations with 8 scales in each orientation. Let (p_1, p_2,...., p_m) be the positive images and (n_1, n_2,...., n_m) be the negative images provided for training. These images are resized to 140x120 pixels. 4x8 Gabor filters are applied to each of the positive and negative training images. We take max over the scales to provide scale invariance. Now we have 4 Gabor response maps. Each map contains edges in a particular direction determined by the orientation of the Gabor filter. These maps are normalized and augmented into a single column vector. We obtain KPCA based feature vectors by computing principal component projections of each orientation map of training sample onto the nonlinear subspaces of positive and negative samples. These features are used to train a SVM classifier.

3.3 Method 3

All the test and training images are resized to 140x120 pixels. These resized images are then normalized to compensate the effect of varying illumination. Finally they are converted into column vectors and KPCA features are derived. Then a support vector classifier is trained to build the classifier.

4 Experimental Results

In first few experiments, we evaluated our object recognition techniques using objects from Caltech database (available at www.vision.caltech.edu). We carried out the experiments as follows: each of the dataset was split randomly into two sets – training set and test set. The first set was used for training and the second one for testing. The negative training and test sets were randomly generated from the background images as in [2]. From the results shown in Fig. 4, we can conclude that for category 2 objects (1) when intensity feature is used, detection rate is poor if the number of KPCA components is kept below 30. It rises with the increase of the number of components. Frustratingly, the false positive rate also increases simultaneously. Therefore, intensity feature is not suitable for the detection of ‘car’. (2) When Gabor feature is used, both detection rate and false positive rate are excellent. Detection rate is almost flat and false positive rate degrades if the number of KPCA components is increased beyond 30. Use of small number of KPCA components is desirable because it minimizes training and recognition time.

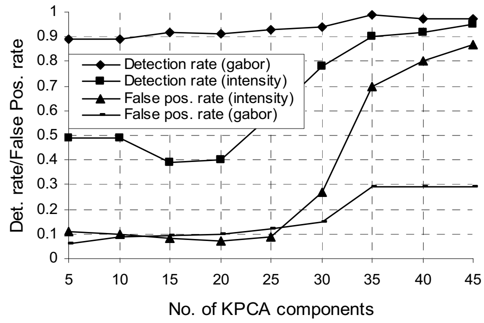


Figure 4. Recognition performance on ‘car’ object.

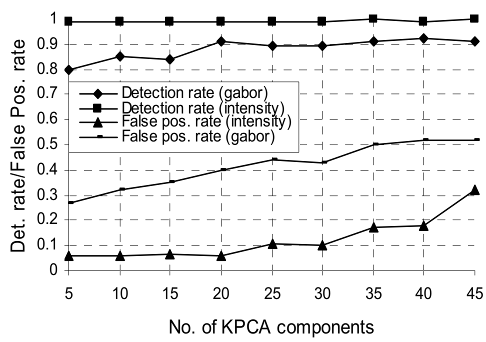


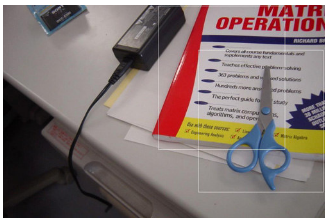
Figure 5. Recognition performance on ‘leopard’ object.

Table 2. Class recognition performance.

Object	Category	Method 3 (intensity based)		Method 2 (Gabor based)		No. of KPCA components	No. of Test/Training images
		Detection Rate	False Positive Rate	Detection Rate	False Positive Rate		
car	2	0.41	0.06	0.91	0.1	20	50
leopard	3	0.99	0.06	0.91	0.4	20	50
sunflower	3	0.84	0.31	0.9	0.4	20	42
umbrella	2	0.7	0.3	0.86	0.21	20	37
scissors	2	0.83	0.51	0.79	0.3	20	25
pizza	2	0.7	0.59	0.93	0.4	20	25
keyboard	4	0.73	0.52	0.79	0.15	20	22
can	2	0.77	0.34	0.71	0.14	20	30

Table 3. Comparison with [2].

Object (all are category 2)	Serre's method, 800 features (Training time = 1200 sec for 25 images. Det. time = 6 sec/image)		Method 2, 20 features (Training time = 20 sec for 25 images. Det. Time is less than 0.1 sec/image)	
	Detection rate, %	False positive rate, %	Detection rate, %	False positive rate, %
Watch	85	13	88	12.5
Ewer	79	20	81.5	22.5
Lamp	80.5	18	77	28.5
Chair	57.5	25	68	24.6



(a)

(b)

Figure 6. Recognition of (a) class using Gabor based KPCA+SVM and (b) specific object using SIFT.

From the same series of experiments (Fig. 5) for category 3 object ('leopard') we notice that: (1) Use of intensity feature results in flat and very high (99 %) recognition rate. It also results in a low false positive rate when the number of KPCA components is kept below 40. Therefore we can safely use a small number of KPCA features. (2) When Gabor feature is used, the detection rate lies above 80% but the false positive rate becomes high. Moreover, the false positive rate becomes worse with the increase of number of KPCA features. In Table 2, class recognition performances of two methods have been compared.

Finally in Table 3, we compare method 2 with Serre's work [2]. These four objects (category 2) are included in the ten worst case categories in [2]. We also experimented with daily objects placed in home scenes. In Fig. 6 (a) Gabor based KPCA+SVM is used to recognize a scissors (class) because it is a category 2 object. Recognition of a specific textured cup using SIFT is shown in Fig. 6 (b).

5 Conclusion

To make a service robot's vision system work well in various situations, we have integrated several methods so that robot can use the appropriate one. We have proposed a scheme to classify the situations depending on the characteristics of object of interest and user's demand. It has been shown that it is possible to categorize the objects into four categories and to employ suitable techniques for each category. Our categorization scheme enables a service robot to automatically select the appropriate feature and detection method to use. SIFT and KPCA in conjunction with SVM have been employed for different categories of objects. We also applied the categorization scheme to select intensity or Gabor feature to use in the KPCA based technique to achieve the better recognition results. Our experimental results confirm the advantage of categorization.

References

[1] D. Lowe: "Distinctive Image Features from Scale-invariant Keypoints," *International Journal of Computer Vision*, vol.60, no.2, pp. 91-110, 2004.

[2] T. Serre, L. Wolf, and T. Poggio: "A New Biologically Motivated Framework for Robust Object Recognition," Ai memo 2004-026, cbcl memo 243, MIT, 2004.

[3] S. Z. Li et. al.: "Kernel Machine Based Learning for MultiView Face Detection and Pose Estimation, " *ICCV*, 2001.

[4] C. Liu: "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.5, pp.572 – 581, 2004.

[5] B. Schölkopf, A.J. Smola, and K.-R. Muller: "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol.10, no.5, pp.1299-1319, 1998.