A Memory Reduction Method for 3D Object Recognition Based on Selection of Local Features

Katsufumi Inoue†, Hiroshi Miyake† and Koichi Kise†

†Department of Computer Science and Intelligent Systems, Osaka Prefecture University

1-1 Gakuencho, Naka, Sakai, Osaka, 599-8531, Japan

E-mail: {inoue, miyake}@m.cs.osakafu-u.ac.jp, kise@cs.osakafu-u.ac.jp

Extended Abstract

The task of object recognition can be classified into two categories: generic and specific. Generic object recognition is to recognize *classes* of objects such as “a chair” and “a car”. Specific object recognition, on the other hand, is for identifying object *instances* such as a specific type of chair and car, in other words “the chair” and “the car”. This paper concerns the latter, especially methods which employ local features such as SIFT (scale-invariant feature transform)[1] for modeling and recognizing 3D objects. As a recognition method, we focus here on a simple one based on voting by matching local features. Although such a simple method offers high recognition rate, it poses the following problems caused by a large number of local features. First, matching of local features (finding their nearest neighbors) requires a long processing time. Second, many local features need immense amounts of storage.

For the first problem, it is necessary to speed up the process of nearest neighbor (NN) search for matching. Fortunately, it has already been known that *approximate* NN search enables us to improve the efficiency drastically while keeping the accuracy of recognition [2]. For the second problem, a possible approach is to select local features for modeling objects. Fortunately, it is not necessary for the recognition by voting to find correct NNs for all local features; What is required is that the number of correct votes is largest. From this viewpoint, we consider that the voting strategy is also advantageous to reduce the number of local features in the object models.

In this paper, we propose a simple method to reduce the required memory space by selecting local features in the object models. The goal is to find the selection such that 3D objects are correctly recognized in high accuracy with the minimum number of local features. However, the selection is the solution of an intractably large combinatorial optimization problem. Thus we pursue it by a greedy method.

In the voting process, each local feature in the object models has two effects: positive and negative. A local feature p in an object model (called a *model local feature*) receives a positive vote when the nearest neighbor of a local feature q in the query image (a *query local feature*) is p whose object ID is the same as q’s. On the other hand, p receives a negative vote if the nearest neighbor of q is p with a different object ID. For reducing the number of local features in the object models, it is required both to employ local features that receive positive votes, and to eliminate local features that receive negative votes, as many as possible. In the proposed method, we estimated positive and negative votes for model local features through the process called the *simulation of object recognition* using images for model construction (model images). In the simulation, one image from the model images is selected as a current query image, and the rest are employed as images for constructing the object models. Query local features are employed for estimating positive and negative votes of model local features in the current object models. By altering the role of a query image within the model images, we accumulate positive and negative votes from many query local features. Note that a single model local feature can receive many positive votes from different query local features. Thus, it is better for reducing the size of the object models to keep a local feature that receives many positive votes. For example, a local feature with 10 positive votes is equivalent to 10 local features each of which is with a single vote. Thus to include a local feature with 10 positive votes is 10 times memory efficient than to employ 10 local features with single positive votes. Note also that a single model local feature has both positive and negative votes from various query local features. Thus in the proposed greedy selection, the total balance is utilized as a criterion to select model local features. However, this simple criterion of selection alone fails because many model local features tend to be selected from limited objects and model images. As a result, for example, objects with no model local features cannot be recognized. In the proposed method, therefore, local features are selected *evenly* from all model images of all objects.

The proposed method was implemented using feature vectors obtained by PCA-SIFT [3] and evaluated using the following two data sets: 30-objects and COIL-100 (Columbia Object Image Library-100) [4]. The data set 30-objects was prepared by ourselves using 30 various objects, some of which are shown in Fig.1. Model images were captured by rotating each object on a turn table using a high-vision video camera whose size was 740 × 480. COIL-100 is a data set of images of 100 objects taken by rotating them in increments of 5º. The size of images is 128 × 128. Figure 2 shows example images of COIL-100. From experimental results for 30-objects, we achieved the recognition rate of 98.8% with about 10% of local features extracted from all model images. Note that the recognition rate with this model (10% local features) decreased only 1.1% as compared to the model including *all* local features. From the model images, in total 900 thousand local features were extracted and their total size was 68.3 MB. Experimental results with COIL-100 show that 1/6 of the total local features allowed us the recognition rate of 96.8%, which is only 2% lower than the recognition rate obtained by the model including all local features. The total number of local features was 180 thousand, and their total size was 13.6 MB. In order to evaluate the effectiveness of our greedy selection, we also applied random selection of local features. Figs. 3 and 4 show the results for 30-objects and COIL-100, respectively. As shown in these figures, the proposed method is superior to the random selection of local features. For the dataset of 30-objects, the advantage was small; this is due to an object (packaging tape) that had so many negative votes. For the dataset of COIL-100, on the other hand, the advantage of the proposed method was clear; the difference of the recognition rate became larger as the model size smaller.

Future work is to evaluate the method with more objects, as well as to decrease the number of required images to construct object models.

This work was supported in part by the Grant-in-Aid for Scientific Research (B) (19300062) from Japan Society for the Promotion of Science (JSPS).

Fig.1. Example of images for 30 objects.

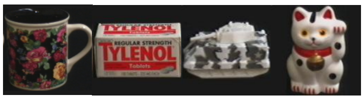

Fig.2. Example of images for COIL-100.

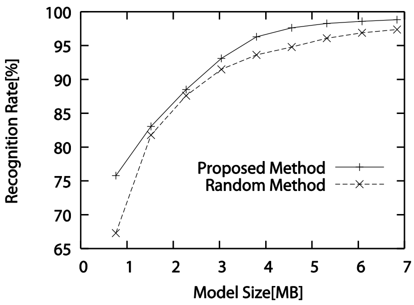

Fig.3. Experimental results for 30objects.

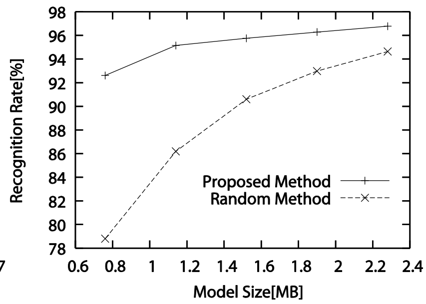

Fig.4. Experimental results for COIL-100.

Key words 3D object recognition, local features, reduction of memory size, COIL.

References

[1] Lowe D., “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.

[2] Kise, K., Noguchi K., and Iwamura M. “Simple Representation and Approximate Search of Feature Vectors for Large-Scale Object Recognition”, In: Proc. BMVC2007, pp. 182-191, 2007.

[3] Ke Y., and R. Sukthankar, R., “PCA-SIFT:A More Distinctive Representation for Local Image Descriptors, In: Proc. of CVPR2004, Vol. 2, pp. 506-613, 2004.

[4] http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php