Invariant Feature Extraction using 3D Silhouette Modeling

Jaehwan Lee¹, Sook Yoon², and Dong Sun Park³
¹Department of Electronic Engineering, Chonbuk National University, Korea
²Department of Multimedia Engineering, Mokpo National University, Korea
³IT Convergence Research Center, Chonbuk National University,

Abstract - One of the major challenging tasks in object recognition results from the great change of object appearance in the process of perspectively projecting objects from 3-dimensional space onto 2-dimensional image plane with different viewpoints. In this paper, we proposed a method to extract features invariant to limited movements of objects by constructing a 3-D model using silhouettes of objects from images with multiple viewpoints. We investigated several renowned invariant features to find the most appropriate one for the proposed method, including SIFT[5], SURF[6], ORB[7], BRISK[8]. The simulation results shows that all the invariant features tested work well and the SURF performs best in terms of matching accuracy.

Keywords: Invariant Feature, Shape From Silhouette, Intelligence Surveillance System

1 Introduction

Accurate recognition of 3-dimensional objects in 2-dimensional images is the most crucial and difficult task in image understanding. There are many possible factors making the recognition task challenging such as information loss from perspective transformation, illumination effects and various appearance of non-rigid body objects[1]. Especially, movements of non-rigid objects in the 3-D space may significantly change images of objects so that matching models in the database with input object images may experience a large difference.

There have been many techniques to resolve the difficulties by using color information, face recognition, part-based recognition, video based gait recognition[2][3][4], etc. Popular image-based local feature description methods such as SIFT[5], SURF[6], ORB[7], BRISK[8] are used to deal with the movements of objects by designing the features invariant to the appearance changes. These feature description methods may work well for a given situation, however, the matching accuracy may need to be improved for the case of recognizing very flexible objects with various viewpoints.

In this paper, We proposed an invariant feature extraction method using a 3-D modelling based on silhouettes of objects from multiple images with different viewpoints. The method firstly construct 3-D models with shape from silhouette approach and then use these models to extract invariant features applicable for any viewpoint. To determine the best feature description method for the proposed method, we also investigate several state-of-the-art feature description methods including the four image-based local feature description methods mentioned above.

2 Proposed Feature Extraction Method

The overall block diagram of the proposed extraction method is depicted in Fig. 1. It consists of a feature extraction block and a test phase block. The first clock is to construct a 3-D model with multiple images and to extract features after projecting the constructed model onto a 2-D plane according to the angle obtained in the pose detection step. This feature extraction method can generate features for any viewpoint changes that can be used to compare to the actual features from test images.

We reconstructed 3D models using the Shape From Silhouette (SFS), described in the Ref. 15, which requires relatively fewer images than other 3-D modelling techniques. Shape From Silhouette is a shape reconstruction method which constructs a 3D shape estimate of an object using silhouette images of the object[16]. In this step, 3D model is trained using multiple images. A set of reconstructed 3-D models can be stored in a database each representing a object at the training phase. These models can be projected onto any 2-D image plane with a specific viewpoint and to be used to extract local features using one of the existing popular methods such as ORB, BRISK, SIFT and SURF shown at the second and third steps. The two steps are later used to verify the existence of objects for test images with additional information from the test stage.

At the test stage, new test images are presented to the system. A test image is firstly used to segment out object regions and then extract features for the regions. The extracted features from the current input image are compared with those from the 3-D models with a set of viewpoints for
matching. If a maximum matching score above a certain threshold value, we accept the input image containing the specific object with a viewpoint. There can be a series of comparison to find the best possible matching.

In this paper, we focused on selecting the most appropriate feature description method which shows the highest matching accuracy. We used four feature description methods: ORB, BRISK, SIFT and SURF. For this purpose, we use the input data set with known viewpoints representing the target objects and compare these objects to those objects reconstructed from the 3-D models. Each feature description method used in this paper generates two sets of feature points for a target object and a reconstructed object. Then each feature point in a target set searches for a feature point in another set for matching. Matching of a point is defined as true if the relative distance between two points is less than a predefined threshold value. The matching accuracy between the two sets of feature points is then determined as in Eq. 1.

$$\text{Matching Accuracy} = \frac{TP}{TP + FP}$$

Fig. 2 shows an example of true and false matching. The objects on the left and right are from input reference image and reconstructed image with reduced size, respectively. In the figure, a false matching and a true matching between feature points are shown as red and blue line segments.

3 Experiments and Discussions

Two data sets, Visual Geometry Group data set[13] and Yasutaka Furukawa and Jean Ponce data set[14], are used for the experiment. The Visual Geometry Group data set contains 36 720x576 images for an object all with different viewpoints. The Yasutaka Furukawa and Jean Ponce data set also contains 24 200x1500 images for another object. Fig 3 shows two example images from each data set. We used the shape from silhouette(SF5)functions introduced by Lore Shure[11] to perform the 3D modeling and the projection of the constructed 3D model. We used feature description methods
implemented in opencv library to extract local features for ORB, BRISK, SIFT and SURF.

### 3.1 3D Silhouette Modeling

A 3-D model of an object is reconstructed with different number of images, using the SFS. For this experiment, we tested with 4, 8 or 36 images for the reconstruction. The angles between two images become 90°, 45°, and 10° for 4, 8, and 36 training images, respectively.

Fig. 4 shows the 3D modeling results of the Visual Geometry Group data set. Three 3D models are reconstructed first with three different number of images and the models are projected for two different viewpoints. For the original images as targets with two different viewpoints, shown in Fig. 4a, the projected images from three 3-D models shown in Fig. 4 (b)-(d). As we can expect, the more images with different viewpoints are used, the better the reconstructed image quality is. Fig. 5 shows the 3-D modelling results the Yasutaka Furukawa and Jean Ponce data set.

### 3.2 Matching Accuracy of Feature Extraction Methods

Before testing the matching accuracy of the invariant features from 3-D modelling, we executed a simple effectiveness and accuracy test to each the feature extraction method. To know the performance of each method, a reference image is modified with a gaussian smoothing and a resizing operations and the matching between the original and the modified versions are performed. Fig. 6 shows the images used for this experiment. Fig. 6a-c shows the 256×256 original image, and the modified images with Gaussian smoothing and the resizing to half size.
Table 1 shows the accuracy measurement results. In the experiment, a matching is defined as true if the distance between locations of two feature points is less than 10 pixels. As we can see in the table, BRISK extracts the less number of feature points and the accuracy is very low. For the case of SURF, we assume that it produces enough number of feature points but the accuracy is not high enough. The number of extracted features of ORB, SURF and SIFT are abundant, the accuracy is relatively high. Thus we assume that ORB and SIFT are good features to use for these types of modifications. Especially, the SURF is very robust to smoothing operation and the SIFT is robust to resizing operation.

Fig. 7 shows an example of feature matching between a projected image from its 3D model reconstructed with 8 images and the corresponding target image. Experimental results show rather lower accuracy than the simple image example. In case of comparison between the actual image and the projected image from a 3D model, SIFT has best accuracy. In this experiment, we measure the accuracy between images from a 3D model with more viewpoints and images from a 3D model with a less number of viewpoints. The reason of this comparison is to produce target images with more details than test images. In case of comparison between images from a 3D model and another 3D model, SURF has best accuracy, over ORB and SIFT.

4 Conclusion

Identifying an 3-dimensional object appeared in different angles is a very challenging task in computer vision, even if
we exclude external factors making the recognition even worse. In this paper, we used the shape from silhouette technique to reconstruct 3-D models of objects with multiple images. The reconstructed 3-D model is used to produce a 2-D projected image with a specific viewpoint for comparison using renowned feature description methods, including ORB, BRISK, SIFT and SURF. The reconstructed 3D models from multiple images contains more details as the more training images are used. When a better reconstructed model is used for testing, the matching accuracy becomes higher. Although there are some positive evidences for automatic generation of invariant features using 3-D modelling, generally speaking, matching performance of feature points are not good enough for the current set of feature extraction methods and different types of features should be developed for this purpose.

We will further search for better features and 3-D models to automatically generate invariant features using 3D models.

5 Acknowledgement

This work was supported by the Brain Korea 21 PLUS Project, National Research Foundation of Korea and by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A2013778).

6 References


[14] Yasutaka Furukawa and Jean Ponce, "3D Photography Dataset", Beckman Institute and Department of Computer Science University of Illinois at Urbana-Champaign
