Joint 3D object recognition and tracking

Orizio Riccardo,
Rizzini Mattia, Zucchelli Maurizio

Università degli Studi di Brescia

26 luglio 2013

Introduction

The project goal is to realize a joint system of recognition and tracking of some features of an object which is identified in a video. These features are called *label*. The sofware is aimed for execution on low performance devices such as the iPad, therefore the tracking is necessary due to the high computational cost of a pure recognition.

Overall structure

The software is designed for concurrent execution. There is an object of class Manager which, given every frame captured by a webcam or in a video, elaborates it in some way and retuns an object of class Object which contains the actual label positions.

Manager

The managers handles two objects of class Recognizer and Tracker. The two objects are *threads*. Every N frames elaborated, the Manager asks the recognizer to perform a full recognition. In the meantime the tracker keeps tracking the labels starting from the last Object given by the recognizer. When the recognizer has finished it returns an object which contains the positions of the labels in the moment in which the recognition started. These positions are *actualized* by the tracker.

Recognition

To perform recognition, *SURF* descriptors are used in place of *SIFT* ones because they are proved to be lighter in application. The recognition bases itself on a *database*. This database support serialization of the structures used for recognition. At the program start, a database name is provided. If this name corresponds to a serialized database, the database is loaded, otherwise it is trained.

Recognition - database

The database training is performed through the analysis of a group of sample images depicting the object from various points of view. To every image is also associated a text file which contains the absolute positions of the various labels in the referenced sample. Every sample image is loaded, from every image SURF features are extracted, descriptors computed and label positions loaded. Everything is then saved in three structures and serialized.

Recognition - matching

At the end of the training or load phase, a *FLANN* based matcher is trained based on all the samples descriptors. When the manager asks the recognizer to perform a match on a frame, his SURF features are extracted and descriptors computed. The descriptors are then used to find a match in the FLANN index. For every descriptors, the two best matches are given. Then the matches are filtered to exclude the ones in which the distance of the best match are too close to the distance of the second best match.

Recognition - matching

These good matches are analyzed to find out what is the sample who's got more of them. That sample is the only considered and his matching keypoints are extracted from the database. Then an homography matrix is calculated between the sample keypoints and the ones in the frame. The matrix is used to map the positions of the labels in the frame. These labels are added to an Object which is returned to the manager.

Tracking

Starting from the last recognized object

Tracking

Conclusions

It works, so shut up!