Real-time Pose and Position Estimation of a Camera with Fish-eye Lens

The method enables to get quite accurate pose and position estimation in real-time for a freely moving camera by analyzing its input image. If the camera is attached to a human head, head tracking will be realized in any natural environment without any outer sensors. The base of the method is the pose and position estimation part of the SLAM method described in [1]. We assume that the 3D coordinates have been learned for necessary number of features in advance and stored in the map with those appearances which are small image templates. The inner parameters of a camera is also given. Then if some features can be identified in the input image, the pose and position of a camera can be estimated. The goal of our method is to output pose and position estimation in real-time for a freely moving camera then the procedure which identifies features and calculates new estimation have to be carried out in 1/30 sec in the case of usual video camera rate. The processing time largely depends on the identification methods and the total number of features stored in the map. We use simple cross-correlation with small image templates to identify features. In order to estimate the pose and position of a camera at least 3 features should be viewed from any pose of a camera. Then the necessary number of features in space to allow any pose of a camera relates to the field of a view of a camera. Wider field of view needs less total numbers of features. Using wide field of view brings another advantage for the pose and position estimation of a camera, that is raising accuracy.

We use fish-eye lens to get wide field of view. This drastically reduces the necessary number of features, but the appearances of features largely change while a camera rotates even without translation. In order to cope with the distortion, we introduce the new procedure that predicts feature appearance after camera pose change. The left image patch in the figure 1 shows the original appearance of a feature. The feature appearance is converted as it is looked at the center of view field (middle of Fig.1) and stored in a map with the camera pose that looks at the feature at center. The prediction of the feature appearance from a new pose of a camera is done according to the relation between the camera poses (right of Fig.1).

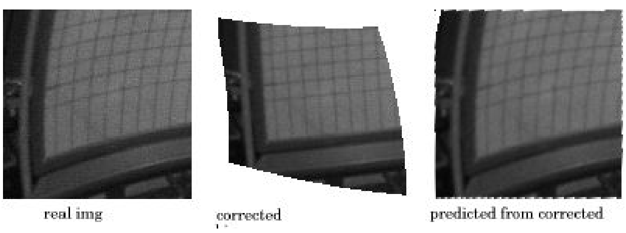

Fig 1

Small image patches surrounded by squares in the figure 2 show the predicted feature appearances while real-time pose and position estimation of a camera. It quite well works while a camera changes its pose with little translation.

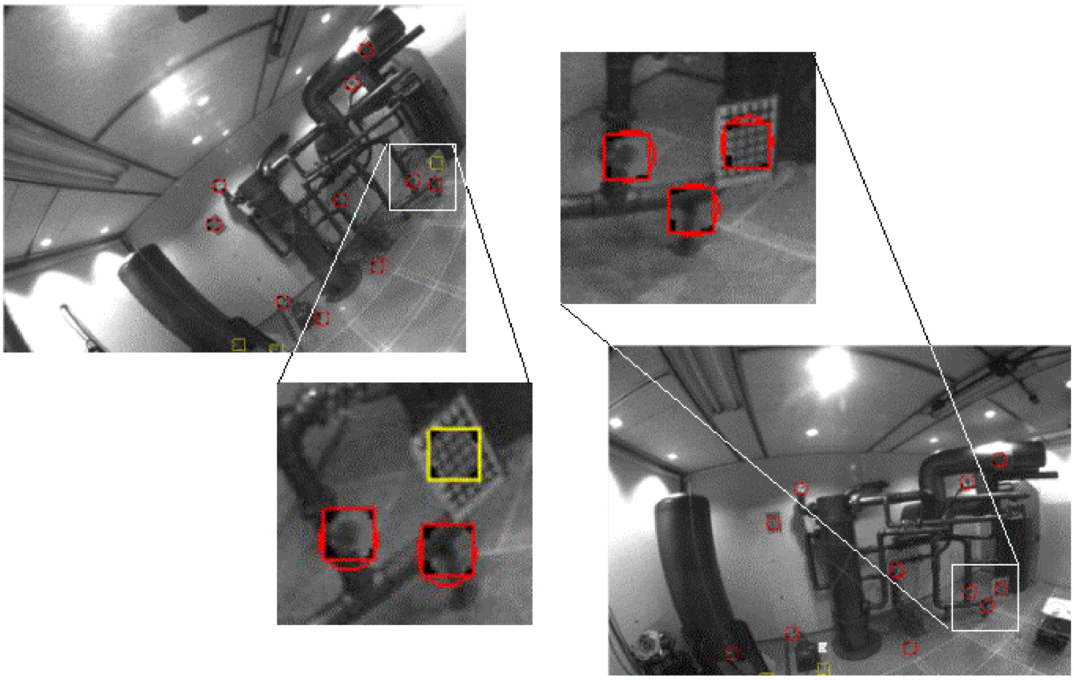

Fig 2

MPEG movie shows one experiment that we have done for a while ago. Three video frames were combined. Upper left frame shows a graphic representation of an obtained estimation of a camera pose and position. The colors of squares mean as follows. Red means successful identification. Yellow means successful identification but not used for a pose and position estimation. Blue means failed identification.

Not shown in the movie, but the method works with some obstacles which may occlude some features, then at the ICCV venue, we will show similar demonstration under quite natural situation.

[1] Andrew Davison: Real-Time Simultaneous Localisation and Mapping with a Single Camera, presented in session 14 of ICCV 2003.