Stable 2D Feature Tracking for Long Video Sequences

Jong-Seung Park, Jong-Hyun Yoon, and Chungkyue Kim

Department of Computer Science & Engineering, University of Incheon
177 Dohwa-dong, Nam-gu, Incheon, 402-749, Republic of Korea
{jong,jhyoon,ckkim}@incheon.ac.kr

Abstract. In this paper, we propose a 2D feature tracking method that is stable to long video sequences. To improve the stability of long tracking, we use trajectory information about 2D features. We predict the expected feature states and compute a rough estimate of the feature location on the current image frame using the history of previous feature states up to the current frame. A search window is positioned at the estimated location and similarity measures are computed within the search window. Once the feature position is determined from the similarity measures, the current feature states are appended to the history buffer. The outlier rejection stage is also introduced to reduce false matches. Experimental results from real video sequences showed that the proposed method stably tracks point features for long frame sequences.

1 Introduction

Feature tracking is one of the most important issues in the fields of image analysis and computer vision. During the last several decades, video tracking technology has been advanced for the practical uses on the surveillance, analysis and representation of video footage[1]. When an image sequence is given as an input, two sets of feature points are extracted from each pair of two adjacent frames. Then a matching algorithm evaluates feature correspondences between the two sets of feature points. Feature tracking is an important field in various computer vision parts since it provides the fundamental data for the further analysis. They can also be used to estimate the object shape toward the object reconstruction. However, the development of related applications has suffered from the instability of tracking methods. Feature tracking is unstable in nature and false correspondences can be occurred at any unexpected time.

In this paper, we propose a robust feature tracking method for long video sequences. Extraction of good features is also important for the robust feature tracking. Feature points are extracted in stable locations that are invariant with respect to image translation, rotation and scaling[2]. After generating a Gaussian image pyramid, we compute difference of Gaussians and the extrema of the difference of Gaussians are selected as feature locations. If two sets of feature points are extracted, we compute the predicted states of features using the history of the feature points from the first frame up to the immediate previous frame. Then we predict each feature position in the current frame. The predicted position is used as an initial value for searching a more accurate feature position. When we refine the feature location from the predicted location, we estimate the accurate feature position by minimizing the matching error.

In our method, we allow the camera to move freely in the physical space. However, if the target object moves together with the camera, it is difficult to solve the problem due to the occurrence of tracking occlusion and the motion ambiguity even in the case of a slight camera movement[3]. The moving objects must be detected in advance and should be handled separately. To avoid such an ill-posed problem, we assume that all objects in the physical space are static relative to the camera.

2 Improving the Accuracy of Feature Tracking

For the stable and robust feature tracking, each step of the tracking should be processed accurately. The three main steps are the extraction of stable feature points, the prediction of feature states, and the computation of feature correspondences. In this section, we describe each step in detail explaining how to improve the accuracy of feature tracking.

2.1 Feature extraction

The reliable feature extraction is the first fundamental step for the robust feature tracking. Feature points should be unique and distinguishable from other feature points for stable feature matching. Typically, feature points are chosen from the intensity corners on the image. Extraction of good features directly drives the improvement toward the robust feature tracking. Good features could be tracked well regardless of translation, rotation and scale changes in input images.

Our feature extraction is similar to the method of scale-space extrema[4][5]. In the first stage, we determine the locations of feature points that are computed by the difference of Gaussian function[5]. The scale space is computed using the input image $I(x, y)$ and the scale Gaussian $G(x, y, \sigma)$. The scale space is defined by Eq. (1), which is the convolution of G with I:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \ . \qquad (1)$$

In the scale space, the difference of Gaussian (DOG) is used for the computation of keypoint locations[5]. The DOG filter corresponds to $G(x, y, k\sigma) - G(x, y, \sigma)$. The difference of two Gaussian images separated by a factor k are represented by:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \ .$$

The DOG function estimates an approximate value using the scale-normalized Laplacian of Gaussian. If the computed difference is the maximum or the minimum, the location is determined as a feature point location. Such feature points are more stable than feature points from other corner-based feature extraction methods.