Removing Mistracking of Multibody Motion Video Database Hopkins155

Yasuyuki Sugaya[1]
sugaya@iim.cs.tut.ac.jp
Yuichi Matsushita[1]
matusita@iim.cs.tut.ac.jp
Kenichi Kanatani[2]
kanatani2013@yahoo.co.jp

[1] Department of Computer Science and Engineering
Toyohashi University of Technology
Aichi 441-8105 Japan

[2] Okayama University
Okayama 700-8530 Japan

Many mathematical techniques have been presented for classifying feature point trajectories over multibody motion video sequences into different motions, and most are applied to the Hopkins155 database for evaluating their performance. In this paper, we point out that Hopkins155 has problems and that it cannot necessarily evaluate the performance correctly. We create a new database incorrect trajectories from Hopkins155. The basic principle of mistracking removal lies on the fact that correct trajectories all belong to parallel 2-D affine spaces in a high-dimensional space if all motions are translational and that parallel 2-D affine spaces are included in a 3-D affine space. Noting that if the image sequence is divided into short intervals, individual motions can be regarded as approximately translational in each interval, we detect incorrect trajectories by repeated plane fitting in the 3-D using RANSAC.

Suppose we track N feature points $\{p_\alpha\}$ over M frames. Let $(x_{\kappa\alpha}, y_{\kappa\alpha})$, $\kappa = 1,...,M, \alpha = 1,...,N$, be the image coordinates of the αth point p_α in the κth frame. Its motion history is represented by the $2M$-D vector

$$\mathbf{p}_\alpha = (x_{1\alpha}, y_{1\alpha}, x_{2\alpha}, y_{2\alpha}, ..., x_{M\alpha}, y_{M\alpha})^\top, \tag{1}$$

which we simply call the "trajectory" of p_α. We map these $2M$-D points to 3-D points as follows:

1. Compute the centroid $\mathbf{p}_C$ of the trajectories $\mathbf{p}_\alpha$, $\alpha = 1, ..., N$ and the deviations $\tilde{\mathbf{p}}_\alpha$ from $\mathbf{p}_C$:

$$\mathbf{p}_C = \frac{1}{N}\sum_{\alpha=1}^{N}\mathbf{p}_\alpha, \quad \tilde{\mathbf{p}}_\alpha = \mathbf{p}_\alpha - \mathbf{p}_C. \tag{2}$$

2. Compute the singular value decomposition of the $2M \times N$ matrix

$$(\tilde{\mathbf{p}}_1, ..., \tilde{\mathbf{p}}_N) = \mathbf{U}\mathrm{diag}(\sigma_1, ..., \sigma_r)\mathbf{V}^\top, \tag{3}$$

where $r = \min(2M, N)$, $\mathbf{U}$ is a $2M \times r$ matrix with r orthonormal columns, $\mathbf{V}$ is an $N \times r$ matrix with r orthonormal columns, and $\sigma_1 \geq \cdots \geq \sigma_r$ (≥ 0) are the singular values.

3. Let $\mathbf{u}_i$ be the ith column of $\mathbf{U}$, and compute the following 3-D vectors $\mathbf{r}_\alpha$, $\alpha = 1, ..., N$:

$$\mathbf{r}_\alpha = ((\tilde{\mathbf{p}}_\alpha, \mathbf{u}_1), (\tilde{\mathbf{p}}_\alpha, \mathbf{u}_2), (\tilde{\mathbf{p}}_\alpha, \mathbf{u}_3))^\top, \tag{4}$$

where and hereafter we denote the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$ by $(\mathbf{a}, \mathbf{b})$.

Geometrically, we are translating the coordinate system of the $2M$-D trajectory space so that the origin is at the centroid $\mathbf{p}_C$, computing the three vectors (the columns of $\mathbf{U}$) that span the affine space, and expressing all the trajectories as their linear combinations.

To these points in 3-D, we fit multiple planes by the following RANSAC procedure:

1. Randomly choose three from among points $\mathbf{r}_\alpha$, $\alpha = 1, ..., N$.
2. Fit a plane to the selected three points and compute θ by LS.
3. Let S be the number of points $\mathbf{r}_\alpha$ that satisfy

$$\frac{(\mathbf{r}_\alpha, \theta)^2}{\theta_1^2 + \theta_2^2 + \theta_3^2} \leq \sigma^2, \tag{5}$$

where θ_i is the ith component of θ. The left hand side is the square distance of point $\mathbf{r}_\alpha$ from the fitted plane, and σ is the standard deviation of feature point detection accuracy, which is empirically set.

4. Repeat the above computation many times and find the value θ that maximize S.

Figure 1: Three video sequences in Hoplins155. The marks □ and × indicate feature point locations regarded as correctly tracked and incorrectly tracked, respectively, by our procedure.

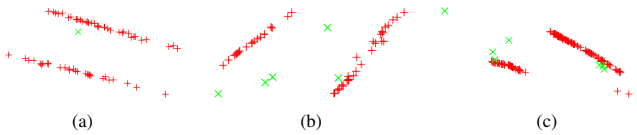

Figure 2: 3-D visualization tracking trajectory in the first five frames in the sequences in Fig. 1.

5. Remove those points $\mathbf{r}_\alpha$ that satisfy

$$\frac{(\mathbf{r}_\alpha, \theta)^2}{\theta_1^2 + \theta_2^2 + \theta_3^2} < \sigma^2 \chi^2_{1;99}, \tag{6}$$

where $\chi^2_{r;a}$ is the ath percentile of χ^2 distribution with r degrees of freedom. This means that we retain those points that cannot be regarded as deviated from the fitted plane by Gaussian noise of mean 0 and standard deviation σ with 1% significance level.

We integrate the result in all intervals by defining the reliability index for the ith interval using the sigmoid function as follows:

$$P(\mathbf{p}_\alpha^{(i)}) = \frac{1}{1 + e^{-(d_\alpha^{(i)} - \sigma^2\chi^2_{1;99})}}. \tag{7}$$

Here, $\mathbf{p}_\alpha^{(i)}$ is the vector that describe the partial trajectory of $\mathbf{p}_\alpha$ over the ith interval, and $d_\alpha^{(i)}$ is the left-hand side of Eq. (6) for the ith interval. However, we regard those points that satisfy Eq. (6) as correct and let $P(\mathbf{p}_\alpha^{(i)}) = 0$. We integrate the results in all the intervals in the following form (the trajectory is more likely to be incorrect if it is larger):

$$L(\mathbf{p}_\alpha) = \Pi^K_{i|P(\mathbf{p}_\alpha^{(i)}) \neq 0} P(\mathbf{p}_\alpha^{(i)}). \tag{8}$$

Here, K is the number of intervals.

Figure 1 shows a video of Hopkins155. We added new feature tracking and divided the sequence into five-frame intervals with one frame overlaps. Five decimated frames are shown in Fig. 1. The marks □ and × indicate feature point locations regarded as correctly tracked and incorrectly tracked, respectively. We set σ =1.0, which we judged to be reasonable according to our experiences. Figure 2 shows 3-D visualization of partial trajectories in the first five-frame intervals. We can clearly see that correct trajectories lie on two planes and that incorrect trajectories stick out from them.

For correct evaluation, we need a reliable database. For this purpose, we created a new database[1] by removing incorrect trajectories using our technique from 35 scenes of Hopkins155, which were natural scenes consisting of two motions.

[1] http://www.iim.cs.tut.ac.jp/T-Hopkins/