Visual Tracking via Subspace Motion Model

Jun Wang[1]
jwang1004@gmail.com
Fan Zhong[1]
zhongfan@sdu.edu.cn
Guofeng Wang[1]
wangguofeng13@163.com
Qunsheng Peng[2]
peng@cad.zju.edu.cn
Xueying Qin[1]
qxy@sdu.edu.cn(Correspondence)

[1] School of Computer Science and Technology, Shandong University, Jinan, China
[2] The State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China

The art of visual tracking has been widely studied in the past decades [6]. While most of researches focus on exploring new methods to represent object appearance, little attention has been paid on the description of object motion. In this paper we propose a novel motion model for visual tracking, and in comparison with previous methods, it can better parameterize instantaneous image motion caused by both object and camera movements.

Our approach is inspired by the subspace theory of image motion, that is, for a rigid object imaged by a projective camera, the displacements matrix of its trajectories over a short period of time should approximately lie in a low-dimensional subspace with a certain rank upper bound [2, 5]. We adopt this subspace as the state transition space in particle filtering (PF) [3]. This differs from affine model in two ways: first, the dimension number as well as the sampling weight for each dimension at each moment can be determined by the rank of the subspace automatically; second, the subspace motion model can naturally represent the disparity brought by object or camera rotation. We will show that when compared with the affine model, the subspace motion model is superior in accuracy.

Figure 1 illustrates the procedures of our method. To estimate the motion model, some 2D feature points of the object are first tracked by the standard KLT approach [4]. Assuming that k successive frames have been tracked before the current frame I_t, then the displacements matrix can be built as:

$$U = \begin{bmatrix} u_{11} & u_{22} & \dots & u_{n1} \\ u_{12} & u_{22} & \dots & u_{n2} \\ & & \vdots & \\ u_{1k} & u_{2k} & \dots & u_{nk} \end{bmatrix}, V = \begin{bmatrix} v_{11} & v_{22} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ & & \vdots & \\ v_{1k} & v_{2k} & \dots & v_{nk} \end{bmatrix} \quad (1)$$

where (u_{ij}, v_{ij}) denotes the displacement of point (x_i, y_i) from the reference frame I to frame I_j. Each row in U and V corresponds to a single frame, and each column corresponds to a single point. U, V can be stacked vertically as $[\frac{U}{V}]_{2k\times n}$. Then, according to the subspace theory, the displacements matrix can be expressed as a bilinear product of matrices:

$$\begin{bmatrix} U \\ V \end{bmatrix}_{2k\times n} = \begin{bmatrix} M_U \\ M_V \end{bmatrix}_{2k\times r} P_{(r\times n)} \quad (2)$$

where P is made up of point-dependent column vectors involving only points positions parameters, while M_U and M_V are made up of frame-dependent row vectors involving only camera motion parameters. The decomposition can be performed by SVD [1], from which the actual value of r is determined by the singular values.

We always deem a moving camera with a static object and describe the initial state of the PF in subspace as:

$$\mathbf{x}_{t-1} = \begin{bmatrix} (M_U)_{k-1} \\ (M_V)_{k-1} \end{bmatrix}_{2\times r} \quad (3)$$

where $(M_U)_{k-1}$ and $(M_V)_{k-1}$ are the last rows of M_U and M_V respectively, corresponding to the frame-dependent vectors of frame I_{t-1}.

The prediction stage of PF is implemented as a first-order autoregressive process:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \Delta\mathbf{m} \quad (4)$$

where $\Delta\mathbf{m}$ is a two-row weighted Gaussian random sample drawn from the normal distribution. The sample weight of each column of $\Delta\mathbf{m}$ is decided by the weight of each subspace dimension.

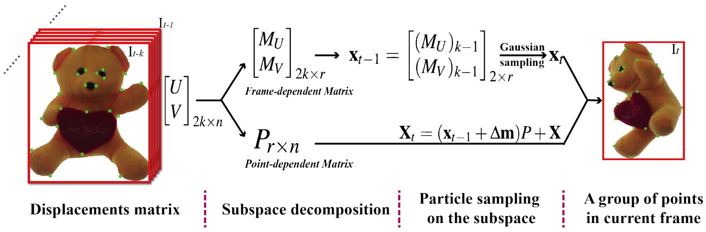

Figure 1: Flow chart of the subspace motion model

Each particle $\mathbf{x}$ in the subspace is corresponding to a group of feature points positions $\mathbf{X}_t$ in the current frame I_t. Let $\mathbf{X}$ be the set of feature points positions in the reference frame I, $\mathbf{X}_t$ is computed by:

$$\mathbf{X}_t = (\mathbf{x}_{t-1} + \Delta\mathbf{m})P + \mathbf{X} \quad (5)$$

The observation state is taken as the point tracking results $\mathbf{Y}_t$ in the current frame I_t. The importance weights w_t^i of each particle is estimated by:

$$w_t^i = \frac{1}{W_t} \exp(-\frac{1}{\eta} \|\mathbf{X}_t^i - \mathbf{Y}_t\|_F^2) \quad (6)$$

where $W_t = \sum_{i=1}^N \exp(-\frac{1}{\eta}\|\mathbf{X}_t^i - \mathbf{Y}_t\|_F^2)$ is the normalization term, and η the smooth term. Then the tracking result $\mathbf{Z}_t$ and tracking error can be represented as:

$$\mathbf{Z}_t = \sum_{i=1}^N w_t^i \mathbf{X}_t^i, \quad err_t = \frac{1}{N}\|\mathbf{Z}_t - \mathbf{Y}_t\|_F^2 \quad (7)$$

The subspace motion model also can facilitate the handling of object appearance change. According to the subspace theory, the rank of the displacement matrix will remain unchanged when a new point belonging to the object is added into the matrix. Based on this idea the appearance model can be updated more robustly by considering only the points that would not change the rank of the displacement matrix.

Comparative experiments of this method are described in the paper, qualitatively and quantitatively. Our conclusion is that, the subspace motion model is capable to describe the image deformation brought by motion of non-plane-like object, and the particle distribution from the PF process could cover the probable object motion states more accordantly. The feasibility of our approach has been effectively demonstrated.

[1] G. H. Golub and C. F. Van Loan. *Matrix Computations 3rd Ed.* Johns Hopkins University Press, 1996.

[2] M. Irani. Multi-frame correspondence estimation using subspace constraints. *Int. J. Computer Vision*, 48(3):173–194, 2002.

[3] M. Isard and A. Blakes. Condensation-conditional density propagation for visual tracking. *Int. J. Computer Vision*, 29(1):5–28, 1998.

[4] J. Shi and C. Tomasi. Good features to track. In *Proc. CVPR*, pages 593–600, 1994.

[5] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int. J. Computer Vision*, 9(2):137–154, 1992.

[6] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):Article No.13, 2006.