Dynamic Building Tracking from UAVs Based on Image Manifold Learning

Peng Zhang
Data Center
National Disaster Reduction Center of China
Beijing, 100124, China
E-mail: zhangpeng@ndrcc.gov.cn

Yuanyuan Ren
Department of Hydraulic Engineering
Tsinghua University
Beijing, 100084, China
E-mail: becci.yam@gmail.com

***Abstract*—Fast and accurate visual tracking of ground buildings can provide unmanned aerial vehicles (UAVs) with rich perceptual information, which is very important for target recognition, navigation and system control. However, when an UAV moves fast, both background and buildings in visual scenes change relatively and rapidly. Consequently, there are no constant features for objects' appearance, which poses great challenges for visual tracking of buildings. In this paper, we first build an image manifold of buildings, which can encode the continuous variation of appearance. We then propose an efficient approach to learn this manifold and obtain more robust feature extraction results. By using a simple tracking framework, we successfully apply the extracted low-dimensional features to real-time building tracking. Experimental results demonstrate the effectiveness of the proposed method.**

Keywords-manifold learning; dynamic visual tracking; unmanned aerial vehicles.

I. Introduction

Along with the advances of technology, vision system has become one of the most important approaches to perceive external environments for unmanned aerial vehicles (UAVs). In dynamic visual perception, dynamic visual tracking plays an important role in providing an UAV with perceptual information. On one hand, visual tracking is not affected by terrain or electromagnetic conditions, hence it can be implemented at any time. On the other hand, efficient visual tracking can promote the understanding of objects' motion, which is crucial to subsequent system control and route design.

So far, various visual tracking methods have been proposed for UAVs. Current methods are mainly based on pattern recognition or computer vision approaches for static video surveillance. Few of them have considered the specific feature of dynamic tracking, that is, difficulties caused by the motion of the system itself. When the system moves, both dynamic and static objects in environment move relatively over time, and the change of objects' appearance would be great if the motion of system is fast. Therefore, there are no constant features for the tracked objects, and tracking methods based on physical or image features would fail.

In this paper, we address the issue of ground building tracking for UAVs, which is a typical task in urban environment or natural disaster site. In such circumstances, vision system mounted on an UAV needs to accurately and rapidly track the target building. Based on our previous works [1][2], we propose an efficient and robust building tracking method by learning an image manifold. We first construct a manifold of building images, which encodes the continuous change of buildings' appearance. Then we introduce a new approach to learn this manifold and extract intrinsic and low-dimensional tracking features, which is more robust to the scale change of buildings in a vision system. Finally, we implement extracted features to a simple tracking framework and achieve real-time and stable tracking of ground buildings from a fast moving UAV.

The rest parts of the paper are organized as follows. Section II reviews related works in the literature. Section III states details of the proposed tracking method. Section IV presents the experimental results. Section V concludes the paper.

II. Related Work

The work in this paper is related to two aspects: one is manifold learning, and the other is visual tracking for UAVs. Related works in these two aspects are reviewed respectively.

The goal of manifold learning methods is to extract intrinsic degrees of freedom underlying high-dimensional inputs which lie on or close to a low-dimensional manifold. They have drawn great research interests since proposed, due to their intuitive motivation and simple implementation. Representative methods include locally linear embedding (LLE) [3], Laplacian eigenmap (LE) [4], locality preserving projections (LPP) [5], and isometric feature mapping (ISOMAP) [6], to name just a few. Recently, manifold learning has been applied to visual tracking and proved to be efficient in extracting intrinsic motion variables of tracked objects, such as head tracking [1], [2] and body tracking [7]. Nevertheless, few works have been done on applying manifold learning to building tracking for UAVs. Although a strategy was proposed in our previous work [2], it does not work well if the scales of a building in successive images change greatly while tracking.

As to visual tracking for UAVs, Campbell and Whitacre [8] proposed an air-ground target tracking method for UAVs. Koch [9] used random matrices to describe objects’ states for air-ground target tracking. Dobrokhodov et al. [10] proposed an object tracking system for unmanned small aircrafts. Zhu and Wang [11] used a bang-bang heading rate controller to achieve circular tracking around the target. Wang et al. [12] presented a compound framework for moving target detection, recognition and tracking based on different altitude UAV captured videos. Zhao et al. [13] proposed an

Published by Atlantis Press, Paris, France.

interacting multiple model Kalman filter for tracking ground moving targets. Current methods are mainly based on physical or image features. Few of them use the intrinsic degrees of freedom underlying objects' appearance for ground building tracking.

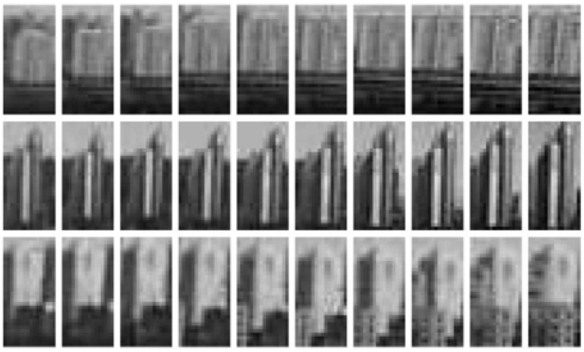

Figure 1. Part of collected training samples.

III. The Tracking Method

A. Manifold Building Process

In this paper, we assume that the change of buildings' appearance in successive frames is continuous and can be parameterized by two intrinsic degrees of freedom, that is, the motion of the vision system and the differences between buildings' appearance. Then image sequences of different buildings are collected, which are recorded by visual cameras mounted on an UAV. Some of the training images are shown in Fig. 1.

Let $\boldsymbol{Q} = \{Q_1, Q_2, \dots, Q_K\}$ in R^n be the set of K sequences of building images. $\boldsymbol{Q}$ contains two subsets $\boldsymbol{Q}_L$ and $\boldsymbol{Q}_R$, which include images of buildings on the left-hand side and right-hand side to the vision system, respectively. Each Q_i contains i_k input samples $\{x_{i1}, x_{i2}, \dots, x_{ik}\}$, which are arranged by time order in the sequence.

For any two sequences Q_i and Q_j ($i, j = 1, 2, \dots, K$), the distance between them is defined as $d(Q_i, Q_j) = (h(Q_i, Q_j) + h(Q_j, Q_i)) / 2$, where $h(Q_i, Q_j)$ is the median Hausdorff distance and given by

$$h(Q_i, Q_j) = \text{median}_l \left(\min_h \| x_{il} - x_{jh} \| \right) .$$

Then the adjacency relationship among input samples is defined by the following principles.

(a) In each sequence Q_i, x_{ij} is adjacent to x_{ij+1}, $j = 1, 2, \dots, k$-1. The edge length between x_{ij} and x_{ij+1} is set to be c_0 where c_0 is a constant.

(b) For any two sequences Q_i and Q_j in $\boldsymbol{Q}_L$ (or $\boldsymbol{Q}_R$), if $d(Q_i, Q_j) < *$, then only x_{i1} , $x_{[ik/2]}$, and x_{ik} are adjacent to x_{j1} , $x_{[jk/2]}$, and x_{jk} respectively, where [•] is the integral part of a real number and $*$ is a given threshold. This is because numbers of samples in Q_i and Q_j are not equal. The edge length between these adjacent samples is set to be $d(Q_i, Q_j)$ / ∀where ∀is a scaling factor.

(c) Let $Q_i^* \in \boldsymbol{Q}_L$ and $Q_j^* \in \boldsymbol{Q}_R$ be the two sequences which have the smallest distance between sequences in $\boldsymbol{Q}_L$ and sequences in $\boldsymbol{Q}$, respectively. Then only x_{i1}^* , $x_{[ik/2]}^*$, and x_{ik}^* are adjacent to x_{j1}^* , $x_{[jk/2]}^*$, and x_{jk}^* respectively. The edge length between these adjacent samples is set to be $d(Q_i^*, Q_i^*)$ / ∃where ∃is a scaling factor.

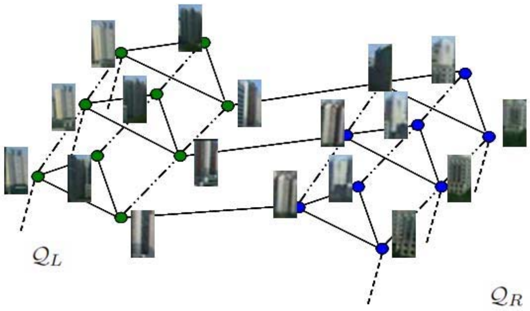

Figure 2. Illustration of the constructed manifold.

An intuitive illustration of the manifold building process is shown in Fig. 2.

With adjacency defined as above, a connected graph is built up among all input samples. We define that x_i and x_j are connected if there exists a path between them. In general, the distance d_{ij} between x_i and x_j is the shortest path length between them if they are connected and infinity otherwise.

The similarity S_{ij} between two inputs is defined as

$$S_{ij} = \begin{cases} \exp\left\{-d_{ij}/\sigma^2\right\} & d_{ij} \neq 0 \\ 0 & d_{ij} = \infty \end{cases} ,$$

where Φ is a scaling factor and is set manually. Then the continuity of intrinsic degrees of freedom is encoded by pairwise similarities.

B. Manifold Learning and Feature Extraction Process

The manifold learning process has two goals: first, to compute a set of low-dimensional representations $Y = \{y_1, y_2, \dots, y_N\}$ in R^m, which can best preserve pairwise similarities $\{S_{ij}\}$; second, to compute a mapping relationship from high-dimensional image space to low-dimensional feature space. In our previous work [2], we used LPP [5] to achieve this goal. However, it does not work well if the scale of buildings in visiual scenes changes greatly. In this paper, we switch to Orthogonal Locality Preserving Projections (OLPP) [14], since it is more stable and has more locality preserving power than LPP.

In OLPP, it is assumed that there exists an n by m projection matrix $A = [a_1\, a_2 \cdots a_m]$ such that for any x_i, $i = 1, 2, \dots, N$, its low-dimensional representation y_i satisfies $y_i = A^T x_i$. Here a_i stands for the i-th column vector of A. Then pairwise similarities are best preserved by solving the following optimization problem.

$$\min_{\{y_i\}} \sum_{ij} \left\| y_i - y_j \right\|^2 S_{ij} , \qquad s.t. \quad a_i^T a_j = \delta_{ij} \tag{1}$$

where $*_{ij}$ equals to 1 if $i = j$ and 0 otherwise.

Let X and Y be corresponding data matrices whose columns are data vectors. Substituting the projection assumption into (1) and adding a non-degenerate constraint, Equation (1) is transformed into

$$\begin{array}{ll} \min & Tr\left(A^T XLX^T A\right) \\ s.t. & \begin{array}{l} A^T XDX^T A = I_m \\ A^T A = I_m \end{array} \end{array} ,$$

where $L = D - S$ with $S = (S_{ij})$ and D is a diagonal matrix with $D_{ii} = \Sigma_j S_{ij}$, $i = 1, 2, \ldots, N$.

Let $A^{(k-1)} = [a_1\ a_2 \cdots a_{k-1}]$ and $B^{(k-1)} = [A^{(k-1)}]^T(XDX^T)^{-1}A^{(k-1)}$, $k = 2, 3, \ldots, m$, then orthogonal vectors $\{a_1, a_2, \ldots, a_m\}$ are computed by the following process.

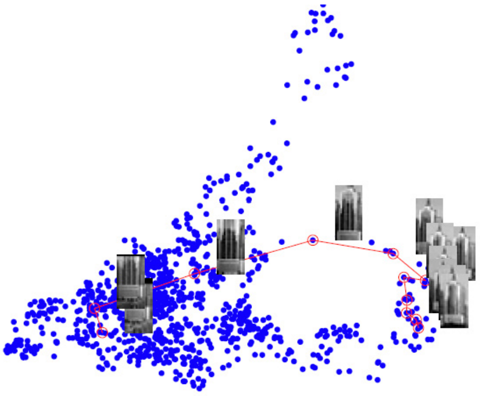

Figure 3. 2D demonstration of the learned embedding.

- a_1 is the eigenvector of $(XDX^T)^{-1}XLX^T$ corresponding to the smallest eigenvalue.
- a_k ($k = 2, 3, \ldots, m$) is the eigenvector of $M^{(k)} = \{I - (XDX^T)^{-1}A^{(k-1)}[B^{(k-1)}]^{-1}[A^{(k-1)}]^T\}(XDX^T)^{-1}XLX^T$ corresponding to the smallest eigenvalue of $M^{(k)}$.

The manifold learning result is shown in Fig. 3, where one sequence of building images are shown with red lines connecting adjacent samples. It can be seen that the continuous change of building scale and appearance are preserved.

C. Visual Tracking Process

The key problem in tracking is to determine where the object should be in the next frame according to current and previous states of the object. Let (O^x_t, O^y_t) be the coordinates of the center of the detection window in Frame t. Then in Frame $t + 1$, positions of candidate windows are acquired by increasing/decreasing (O^x_t, O^y_t) with equal intervals.

Assume that I_t is the detected object image from Frame t and that I^c_{t+1} is a candidate object image from Frame $t + 1$. Then their low-dimensional representations are computed as follows:

$$\begin{array}{l} y_t = A^T I_t \\ y^c_{t+1} = A^T I^c_{t+1} \end{array} .$$

In Frame $t + 1$, the image which has the shortest distance to I_t in the low-dimensional space is considered as the optimal object image.

IV. Experimental Results

In this section, we apply the proposed tracking method to building tracking missions on an UAV. We conduct two experiments to demonstrate the validity of extracted features. In both experiments, the feature dimension is 9 and $\Phi = 1$, under which tracking results are the most stable.

In the first experiment, the vision system needs to track a distant building while rapidly flying over an overpass. There are many similar buildings close to the target one, which may interfere in the tracking process. Experimental results are shown in Fig. 4. The proposed tracking method has successfully tracked the building during the whole process.

In the second experiment, the proposed tracking method is used to track a static building while an UAV flies over an avenue in high speed. The main challenges consist of great changes of both the relative scale/position of building with respect to the camera and the appearance of building. Experimental results are shown in Fig. 5. The proposed tracking method has successfully tracked the building during the whole process.

V. Conclusion

In this paper, we proposed an efficient ground building tracking method for unmanned aerial vehicles (UAVs) based on image manifold learning. We built a low-dimensional manifold which encodes the intrinsic degrees of freedom of buildings' appearance. We introduced a more stable approach to learn this manifold and extracted tracking features that are robust to scale changes of buildings. We successfully applied the extracted features to ground building tracking tasks, which were validated by challenging experiments.

Acknowledgment

This work was partly supported by the NNSF of China under Grant nos. 41201552 and no. 41174013.

References

[1] H. Qiao, P. Zhang, B. Zhang, and S. Zheng, “Learning an intrinsic variable preserving manifold for dynamic visual tracking,” IEEE Trans. Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 40, no. 2, pp. 868-880, 2000.

[2] H. Qiao, P. Zhang, B. Zhang, and S. Zheng, “Tracking feature extraction based on manifold learning framework,” Journal of Experimental & Theoretical Artificial Intelligence, vol. 23, no. 1, pp. 23-38, 2011.

[3] S. Roweis, and L. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” Science, vol. 290, no. 5500, pp. 2323-2326, 2000.

[4] M. Belkin, and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.

[5] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, “Face recognition using Laplacianfaces,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328-340, 2005.

[6] J. Tenenbaum, V. Silva, and J. Langford, “A global geometric framework for nonlinear dimensionality reduction,” Science, vol. 290, no. 5500, pp. 2319-2323, 2000.

[7] A. Elgammal, ”Tracking people on a torus,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 3, pp. 520-538, 2009.

[8] M. Campbell, and W. Whitacre, “Cooperative tracking using vision measurements on SeaScan UAVs,” IEEE Trans. on Control Systems Technology, vol. 15, no. 4, pp. 613-626, 2007.

[9] J. Koch, “Bayesian approach to extended object and cluster tracking using random matrices,” IEEE Trans. on Aerospace and Electronic Systems, vol. 44, no. 3, pp. 1042-1059, 2008.

[10] V. Dobrokhodov, I. Kaminer, K. Jones, and R. Ghabcheloo, “Vision based tracking and motion estimation for moving targets using unmanned air vehicles,” Journal of Guidance Control and Dynamics, vol. 31, no. 4, pp. 907-917, 2008.

[11] S. Zhu, and D. Wang, ”Adversarial ground target tracking using UAVs with input Constraints”, Journal of Intelligent & Robotic Systems, vol. 65, pp. 521-532, 2012.

[12] J. Wang, Y. Zhang, J. Lu, and W. Xu, ”A framework for moving target detection, recognition and tracking in UAV videos,” Advances in Intelligent and Soft Computing, vol. 137, pp. 69-76, 2012.

[13] W. Zhao, W. Chen, G. Zheng, K. Huang, K. Zhao, and Y. Li, ”Study on UAV video reconnaissance based adaptively tracking algorithm for the ground moving target,” Lecture Notes in Computer Science, vol. 6838, pp. 282-289, 2012.

[14] D. Cai, X. He, J. Han, and H. Zhang, “Orthogonal Laplacianfaces for face recognition,” IEEE Trans. on Image Processing, vol. 15, no. 11, pp. 3608-3614, 2006.

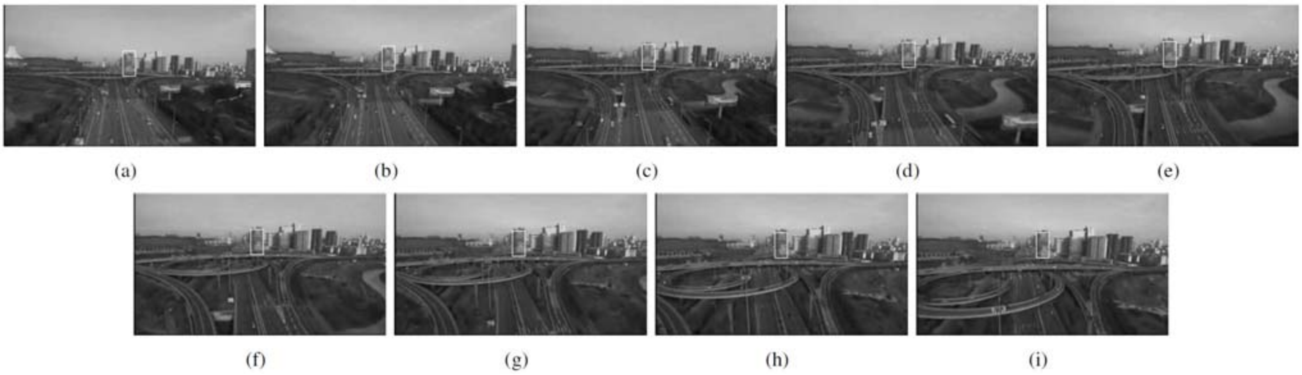

Figure 4. Building tracking results (white box) with the proposed new tracking method. The UAV rapidly flies over an overpass.

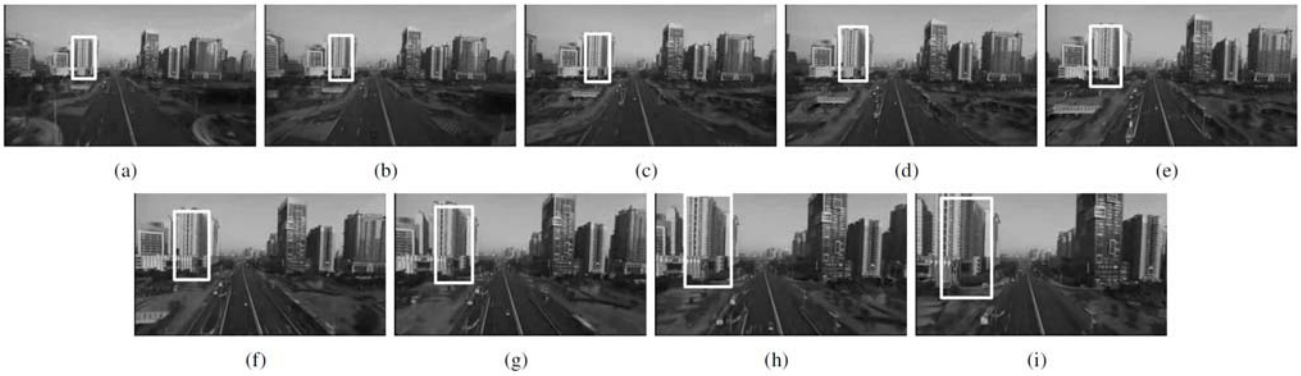

Figure 5. Building tracking results (white box) with the proposed new tracking method. The size of building’s appearance changes rapidly in the camera.