PERSISTENT TRACKING OF MOVING OBJECTS USING DISTRIBUTED CAMERAS

Gang Qian*, Rama Chellappa and Qinfen Zheng
Center for Automation Research
Institute of Advanced Computer Studies
University of Maryland
College Park, MD, 20742-3275

ABSTRACT

In this paper, we present an approach for persistent tracking of moving objects using distributed cameras. Targets can be handed over across multiple cameras. The empirical means of epipolar distances are evaluated to tackle the target cross-sensor association problem.

1. INTRODUCTION

In surveillance applications, detection and continuous tracking of suspicious humans and vehicles is critical. By prompt detection and reliable tracking of intruders, crucial intelligence such as the types/number of vehicles, number of personnel and their current positions can be collected and utilized for taking appropriate actions. Use of multiple cameras will enable persistent detection and tracking of moving humans and vehicles under occlusion and will be useful in wide area surveillance applications. Because the field of view (FOV) of a single camera is limited and also in some circumstances the target being tracked might be occluded by obstacles such as trees, buildings, etc, distributed imaging sensors have to be employed to persistently track targets. By using distributed cameras, when a target is leaving the FOV of one camera or occluded by some obstacles, but still remains in the monitored area, other cameras will be able to pick up this target and continue tracking. There are two major challenges to be addressed here. One is to automatically calibrate a camera network with multiple arbitrarily placed cameras. We refer to this problem as the self-calibration of a camera network. The other is to find correct target associations between video streams from different cameras when multiple targets are present at the same time. Even when only one target is observed from both cameras, they may not correspond to the same target. We refer this problem as target cross-camera hand-over .To solve the first problem, we have developed an algorithm that can automatically calibrate distributed cameras by observing synchronized video streams from different cameras. In this algorithm, 2D trajectories of moving targets in the image planes of different cameras are used as feature correspondences. The posterior distribution of relative position and orientation of cameras is represented by related sample and weight sets by using a sequential Monte Carlo (SMC) (Qian and Chellappa, 2001, 2002) technique. Due to space limitations, the self-calibration algorithm is not discussed in this summary. We assume that the relative positions and orientations among distributed cameras have been estimated and described by a set of properly weighted samples and their associated weights. We describe the algorithm for solving the problem of target cross-sensor hand-over.

2. THE DISTRIBUTED TRACKING ALGORITHM

Assume that a number of targets are tracked in each video stream. The CONDENSATION (Isard and Blake, 1996) or other SMC-based algorithms (e.g. Li and Chellappa, 2000) can be used to reliably track moving objects from monocular video streams. If target ***A*** in the first camera corresponds to target ***B*** in the second camera, points in the motion trajectories of targets ***A*** and ***B*** in image planes satisfy the epipolar constraint (Faugeras and Luong, 2001). Given a pair of target trajectories from two video streams, and using the posterior distribution of the external parameters, the empirical mean of the epipolar distances at different time instants can be evaluated. False matching can be removed by thresholding on residual errors. Let the coordinate systems of camera one and camera two be ***C*** and ***C'***, respectively. Let the rotation matrix and translation from ***C*** to ***C'*** be ***m=(R, T)*** such that for a 3D point ***P*** in ***C***, the coordinate in ***C'*** is ***P'=R(P-T)***. Assume that ***p*** is the projection of ***P*** in the first image plane and that its position is ***(u,v)***. In the second image plane, the epipolar distance from a point ***q*** at ***(u',v')*** to the epiploar line related to ***P*** is given by (1) where the coefficients ***(a,b,c)*** determine the epiploar line and can be solved using (2).

$$d(p,q,m) = \frac{au'+bv'+c}{\sqrt{a^2+b^2}} \quad (1)$$

$$[a,b,c]^T = A'^{-T}[-RT]_\times RA^{-1}[u,v,1]^T \quad (2)$$

A and ***A'*** are the known calibration matrices of the two cameras. $[T]_\times$ is decided by ***T*** such that for any 3D vector ***x***, $[T]_\times x = T \times x$. If ***p*** and ***q*** are related to the

Partially supported by the ARL Collaborative Technology Alliance on Advanced Sensors Contract DAAD 19-01-2-008

Fig. 1. Pictures in the first row show two tracked people from camera one and those in the second row show a tracked person from camera two

same 3D point, the above epipolar distance from (1) is zero. Assume that at time t, the position of target ***A*** tracked in camera one is at position $\boldsymbol{p}_t$. We want to check if target ***B***, now at point $\boldsymbol{q}_t$, tracked using the second coorespondes to target ***A***. If they are the same target, the related epipolar distance must be small over a certain period of time. Since the camera external parameters are expressed by a set of properly weighed samples, we desire to compute the mean of the epipolar distances which can be evaluated by weighted summation. At time t, the mean of the epipolar distances from $\boldsymbol{q}_t$ to the epipolar line related to $\boldsymbol{p}_t$ is given by

$$\bar{d}(q_t,p_t)=\int d(q_t,p_t,m)dp_m=\frac{1}{\sum_{j=1}^{n} w^{(j)}}\sum_{j=1}^{n} d(q_t,p_t,m^{(j)})w^{(j)} \quad (3)$$

where $\{m^{(j)},w^{(j)}\}, j=1,\ldots,n,$ are the samples and weights of relative motion between the two cameras.

3. EXPERIMENTAL RESULTS

In this example, two people are tracked in a sequence from camera one and one person from camera two. The tracking results are displayed in Figure 1, with tracked targets marked by bounding boxes. The question is which person in camera two is the one tracked in camera two. If there is a person that can be matched between these two cameras, there are two possibilities since either person in camera one might match the one in camera two. By using the posterior distribution of external parameters of camera two relative to camera one, the empirical means of epipolar distances of the two possible matching at different time instants are evaluated. The one with smaller epipolar distances is accepted as true matching. If both people result in large errors, none of them match. The epipolar distances related to the two possible matching in this example are shown in Figure 2. We see that one match produces low epipolar distances and it indicates that the person on the left in camera one matches the one in camera two, which is correct. Hence, the tracking of this person is successfully handed over from camera one to camera two.

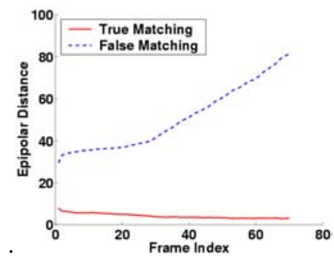

Fig. 2: Empirical mean of the epipolar distances. True target matching produces low epipolar distances.

4. CONCLUSIONS

In this paper, we have proposed an approach for distributed target tracking using multiple cameras. By using multiple cameras, targets can be reliably and persistently tracked in a wide area. The proposed algorithm will be very useful in surveillance systems.

REFERENCES

Faugeras, O. and Luong, Q., *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*, MIT Press, 2001

Isard, M. and Blake, A., "Contour tracking by stochastic propagation of conditional density", in *Proceedings of the European Conference on Computer Vision, Cambridge, UK,* 1996, Part I, pp343-356

Li, B. and Chellappa. R. "Model-based temporal object verification using video," *IEEE Trans. Image Processing*, vol. 10, pp. 897-908, 2001

Qian, G. and Chellappa, R., "Structure from motion using sequential Monte Carlo methods," in *Proceedings of International Conference on Computer Vision*, vol II, pp. 614-621, Vancouver, BC, Canada, 2001

Qian, G. and Chellappa, R., "Bayesian self-calibration of a moving camera," in *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, 2002, Part II, pp. 277-293