Switching Local and Covariance Matching for Efficient Object Tracking

Junqiu Wang and Yasushi Yagi
The Institute of Scientific and Industrial Research, OSAKA University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
jerywangjq@gmail.com, yagi@am.sanken.osaka-u.ac.jp

Abstract

The covariance tracker finds the targets in consecutive frames by global searching. Covariance tracking has achieved impressive successes thanks to its ability of capturing spatial and statistical properties as well as the correlations between them. Nevertheless, the covariance tracker is relatively inefficient due to its heavy computational cost of model updating and comparing the model with the covariance matrices of the candidate regions. Moreover, it is not good at dealing with articulated object tracking since integral histograms are employed to accelerate the searching process. In this work, we aim to alleviate the computational burden by selecting appropriate tracking approaches. We compute foreground probabilities of pixels and localize the target by local searching when the tracking is in steady states. Covariance tracking is performed when distractions, sudden motions or occlusions are detected. Different from the traditional covariance tracker, we use Log-Euclidean metrics instead of Riemannian invariant metrics which are more computationally expensive. The proposed tracking algorithm has been verified on many video sequences. It proves more efficient than the covariance tracker. It is also effective in dealing with occlusions, which are an obstacle for local mode-seeking trackers such as the mean-shift tracker.

1. Introduction

Object tracking in video sequences is challenging under uncontrolled conditions. Tracking algorithms have to estimate the states of the targets when variations of background and foreground exist, occlusions happen, or appearance contrast becomes low. Trackers need to be efficient and can track variant targets. Target representation, similarity measure and localization strategy are essential components of most trackers. The selection of components leads to different tracking performance.

The mean-shift algorithm [2] is a non-parametric density gradient estimator which finds local maxima of a similarity measure between the color histograms (or kernel density estimations) of the model and the candidates in the image. The mean-shift algorithm is very fast due to its searching strategy. However, it is prone to failure in detecting the target when the motion of the target is large or when occlusions exist since only local searching is carried out.

The covariance tracker [7] represents targets using covariance matrices. The covariance matrices fuse multiple features in a natural way. They capture both spatial and statistical properties of objects using a low dimensional representation. To localize targets, the covariance tracker searches all the regions; and the region with the highest similarity to the target model is taken as the estimation result. The covariance tracker does not make any assumption on the motion. It can compare any regions without being restricted to a constant window size. Unfortunately, the Riemannian metrics adopted in [7] are complicated and expensive. Since it uses a global searching strategy, it has to compute distances between the covariance matrices of the model and all candidate regions. Although an integral image based algorithm that requires constant time is proposed to improve the speed, it is still not quick enough for real time tracking. It is difficult for the covariance tracker to track articulated objects since computing covariance matrices for articulated objects is very expensive.

In this work, we propose a tracking strategy that switches between local tracking and global covariance tracking. The switching criteria are determined by the tracking condition. Local tracking is carried out when the target does not have large motion. When large motion or occlusions happen, covariance tracking is adopted to deal with the issue. The switching between

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

local and covariance matching makes the tracking efficient. Moreover, it can deal with sudden motions, distractions, and occlusions in an elegant way. We compute covariance matrices only on those pixels that are classified as foreground. Therefore we can track articulated objects.

To speed up the global searching process, we use Log-Euclidean metrics [1] instead of the Riemannian invariant metrics [6, 7] to measure the similarity between covariance matrices. The model update in covariance tracking [7] is also expensive. We update the model by computing the geometric mean of covariance matrices based on Log-Euclidean metrics. The computation is simply Euclidean in the logarithmic domain, which reduces the computational costs. The final geometric mean is computed by mapping back to the Riemannian domain with the exponential. Log-Euclidean metrics provide results similar to their Riemannian affine invariant equivalent but takes much less time.

This paper is structured as follows. In Section 2, we introduce the local tracking method based on foreground likelihood computation. In Section 3, we apply Log-Euclidean metrics in covariance tracking. The switching criteria for the local and global tracking strategies are described in Section 4. Experimental results are given in Section 5. Section 6 concludes the paper.

2. Local Tracking

The local tracking is performed based on foreground likelihood. The foreground likelihood is computed using the selected discriminative color and shape-texture features [10]. The target is localized using mean-shift local mode seeking on the integrated foreground likelihood image.

2.1. Foreground Likelihood

We compute foreground likelihood based on the histograms of the foreground and background with respect to a given feature. The frequency of the pixels that appear in a histogram bin is calculated as $\zeta_f^{(b_{in})} = p_f^{(b_{in})}/n_{fg}$ and $\zeta_b^{(b_{in})} = p_b^{(b_{in})}/n_{bg}$, where n_{fg} is the pixel number of the target region and n_{bg} the pixel number of the background.

The log-likelihood ratio of a feature value is given by

$$L^{(b_{in})} = \max(-1, \min(1, \log \frac{\max(\zeta_f^{(b_{in})}, \delta_L)}{\max(\zeta_b^{(b_{in})}, \delta_L)})), \quad (1)$$

where δ_L is a very small number. The likelihood image for each feature is created by back-projecting the ratio into each pixel in the image [8, 10].

2.2. Model Updating for Local Tracking

The local tracker needs adaptivity to handle appearance changes. The model is computed by mixing the current model with the initial model which is considered as correct [10]. The mixing weights are generated from the similarity between the current model and the initial model [10]. The initial model works in a similar way to the stable component in [4]. But the updating approach in [10] takes less time.

3. Improving Covariance Tracking

The covariance tracker [7] describes objects using covariance matrices. The covariance matrix fuses different types of features and modalities with small dimensionality. Covariance tracking searches all the regions and guarantees a global optimization (Up to the descriptive ability of the covariance matrices). Despite of these advantages, covariance tracking is relatively expensive due to the distance computation and model updating in Riemannian manifold. We speed up the global searching and the model updating by introducing Log-Euclidean metrics.

3.1. Target Representation

The target is described by covariance matrices that fuse multiple features. We adopt the features used in [7], which consist of pixel coordinates, RGB colors and gradients. The region R is described with the $d \times d$ covariance matrix of the feature points in R

$$C_R = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{z}_k - \mu)(\mathbf{z}_k - \mu)^T, \quad (2)$$

where μ is the mean of the points.

The covariance of a certain region reflects the spatial and statistical properties as well as their correlations of a region. However, the means of the features are not taken into account for tracking. We use the means by computing the foreground likelihoods and incorporate them into the covariance computation.

3.2. Similarity Measuring For Covariance Matrices

The simplest way for measuring similarity between covariance matrices is to define a Euclidean metric, for instance, $d^2(C_1, C_2) = \mathrm{Trace}((C_1 - C_2)^2)$ [1]. However, the Euclidean metric can not be applied to measure the similarity due to the fact that covariance matrices may have null or negative eigenvalues which are meaningless for the Euclidean metrics [3]. In addition, the Euclidean metrics are not appropriate in terms of symmetry with respect to matrix inversion, e.g., the multiplication of covariance matrices with negative scalars is not closed for Euclidean space.

Since covariance matrices do not lie on Euclidean space, affine invariant Riemannian metrics [3, 6] have been proposed for measuring similarities between covariance matrices. To avoid the effect of negative and null eigenvalues, the distance measure is defined based on generalized eigenvalues of covariance matrices:

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^{n} \ln^2 \lambda_i(C_1, C_2)}, \qquad (3)$$

where $\{\lambda_i(C_1, C_2)\}_{i=1\ldots n}$ are the generalized eigenvalues of C_1 and C_2, computed from

$$\lambda_i C_1 \mathbf{x}_i - C_1 \mathbf{x}_i = 0, i = 1 \ldots d, \qquad (4)$$

and $\mathbf{x}_i \neq 0$ are the generalized eigenvectors. The distance measure ρ satisfies the metric axioms for positive definite symmetric matrices C_1 and C_2. The price paid for this measure is a high computational burden, which makes the global searching expensive.

In this work, we use another Riemannian metrics – Log-Eucliean metrics proposed in [1]. When only the multiplication on the covariance space is considered, covariance matrices have Lie group structures. Thus the similarity can be measured in the domain of logarithms by Euclidean metrics:

$$\rho_{LE}(C_1, C_2) = \| \log(C_1) - \log(C_2) \|_{Id}. \qquad (5)$$

This metric is different from the classical Euclidean framework in which covariance matrices with null or negative eigenvalues are at an infinite distance from covariance matrices and will not appear in the distance computations.

Although Log-Eucliean metrics are not affine-invariant [1], some of them are invariant by similarity (orthogonal transformation and scaling). It means that the Log-Euclidean metrics are invariant to changes of coordinates obtained by a similarity [1]. The properties of Log-Euclidean make them appropriate for similarity measuring of covariance matrices.

3.3. Model Updating

Covariance tracking has to deal with appearance variations. Porikli et al. [7] construct and update a temporal kernel of covariance matrices corresponding to the previously estimated object regions. They keep a set of previous covariance matrices $[C_1 \ldots C_T]$. From this set, they compute a sample mean covariance matrix that blends all the previous matrices. The sample mean is an intrinsic mean [7] because covariance matrices do not lie on Euclidean spaces. Since covariance matrices are symmetric positive definite matrices, they can be formulated as a connected Riemannian manifold. The structure of the manifold is specified by a Riemannian metric defined by collection of inner products. The model updating is computationally expensive due to the heavy burden of computation in Riemannian space.

In this work, we use the Log-Euclidean mean of T covariance matrices with arbitrary positive weights $(w_i)_{i=1}^T$ such that $\sum_{i=1}^T w_i = 1$ is a direct generalization of the geometric mean of the matrices. It is computed as

$$C_m = \exp(\sum_{i=1}^{T} \log(C_i)). \qquad (6)$$

This updating method need much less computational costs than the method used in [7].

4. Switching Criteria

The local tracking strategy is adopted when the tracker runs in steady states. When sudden motion, distractions or occlusions happen, local tracking strategy tends to fail due to its limited searching region. We switch to the global searching strategy based on the improved covariance tracker described in the previous section. Motion prediction techniques such the Kalman filter have been used to deal with occlusions. However, when the prediction is far away from the true location, a global searching is preferred to recover from tracking failure.

The detection of sudden motion and distraction is performed using the effective methods proposed in [9]. Occlusions are announced when the objective function value of the local tracking is lower than some threshold t_l. The threshold for switching between local and covariance tracking is computed by fitting a Gaussian distribution based on the similarity scores (Bhattacharyya distances) of the frames labeled as occlusion. The threshold is set to $3\sigma_t$ from the mean of the Gaussian. The covariance tracking is applied when the above threats are detected.

5. Experiments

We verify our approach by tracking different objects in some challenging video sequences.

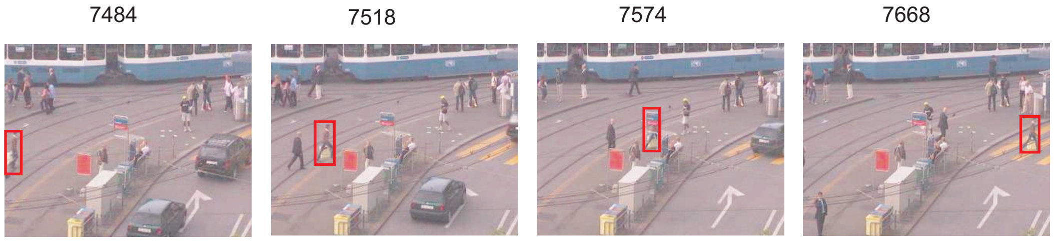

Figure 1. Tracking pedestrian in the complex background. No background subtraction is applied in the tracking.

Table 1. Tracking percentages of the proposed and other trackers.

Algorithm	Seq1	Seq2	Seq3
Meanshift	72.6	78.5	35.8
Covariance	89.7	90.4	78.8
TheProposed	91.3	88.1	83.1

In Figure. 1, we show the tracking results on the street sequence [5]. Pedestrians are articulated objects which are difficult to track. The occlusions in frame 7574 brings more difficulty to the tracking. The proposed tracker successfully tracks through the whole sequence. We compare the proposed tracker with the mean-shift and covariance trackers. Different objects in the three sequences [5] are tracked and the tracking percentages are given in Table. 1. The proposed tracker provides higher or similar correct ratio.

5.1. Computation Complexity

The tracking is faster when the local tracking method is applied since the searching of local tracking is only performed on certain the regions. It takes less than 0.02 seconds to process one frame.

The covariance tracking is also sped up thanks to the efficiency of Log-Euclidean distance computation adopted in this work. The iterative computation of the affine invariant mean leads to heavy computational cost. In contrast, the Log-Euclidean metrics are computed in a closed form. The computation of mean based on Log-Euclidean distances takes less than 0.02 seconds, whereas the computation based on Riemannian invariant metrics takes 0.4 seconds.

6. Conclusions

We propose a novel tracking framework taking the advantages of local and global tracking strategies. The local and global tracking are performed by using the mean-shift and covariance matching. The proposed tracking algorithm is efficient because local searching strategy is adopted for most of the frames. It can deal with occlusions and large motions for the switching from local to global matching. We adopt Log-Euclidean metrics in the improved covariance tracking, which makes the global matching and model updating fast.

References

[1] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the log-euclidean framework. In *Proc. MICCAI'05*, pages 115–122, 2005.

[2] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.

[3] W. Forstner and B. Moonen. A metric for covariance matrices. *Technical report, Dept. of Geodesy and Geoinformatics*, 1999.

[4] A. D. Jepson, D. J. Fleet, and T. EI-Maraghi. Robust online appearance models for visual tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(10):1296–1311, 2003.

[5] B. Leibe, K. Schindler, and L. V. Gool. Coupled detection and trajectory estimation for multi-object tracking. In *Proc. of Int'l Conf. on Computer Vision*, pages 115–122, 2007.

[6] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *Intl. Journal of Computer Vision*, 66:41–66, 2006.

[7] F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on lie algebra. In *Proc. of Intl Conf. on Computer Vision and Pattern Recognition*, pages 728–735, 2006.

[8] M. Swain and D. Ballard. Color indexing. *Intl. Journal of Computer Vision*, 7(1):11–32, 1991.

[9] J. Wang and Y. Yagi. Discriminative mean shift tracking with auxiliary particles. In *Proc. 8th Asian Conference on Computer Vision*, pages 576–585, 2007.

[10] J. Wang and Y. Yagi. Integrating color and shape-texture features for adaptive real-time tracking. *IEEE Trans. on Image Processing*, 17(2):235–240, 2008.