Int. J. Electron. Commun. (AEÜ) 65 (2011) 915–923

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

journal homepage: www.elsevier.de/aeue

Efficient and robust fragments-based multiple kernels tracking

Jiangxiong Fang [a,*], Jie Yang [a], Huaxiang Liu [b]

[a] Shanghai Jiao Tong University, Institute of Image Processing and Pattern Recognition, Shanghai 200240, China
[b] Hunan Normal University, Changsha 410080, China

ARTICLE INFO

Article history:
Received 25 May 2010
Accepted 28 February 2011

Keywords:
Multiple kernels tracking
Object tracking
Adaptive scale selection
Mean shift

ABSTRACT

Representing an object with multiple image fragments or patches for target tracking in a video has proved to be able to maintain the spatial information. The major challenges in visual tracking are effectiveness and robustness. In this paper, we propose an efficient and robust fragments-based multiple kernels tracking algorithm. Fusing the log-likelihood ratio image and morphological operation divides the object into some fragments, which can maintain the spatial information. By assigning each fragment to different weight, more robust target and candidate models are built. Applying adaptive scale selection and updating schema for the target model and the weighting factors of each fragment can improve tracking robustness. Upon these advantages, the novel tracking algorithm can provide more accurate performance and can be directly extended to a multiple object tracking system.

1. Introduction

Robust and real-time tracking has been a challenging problem and has many practical applications, for example in surveillance and monitoring, activity analysis, segmentation. Although there has been some success, it is a still very difficult task to make tracking an object in fast object motion changing appearance, clutter and occlusion. Considerable work has already been done in target tracking to address these challenges.

Many algorithms have been proposed in recent years in object tracking. Most of the tracking algorithms are broadly classified into deterministic methods and stochastic methods. Deterministic methods usually involve a gradient descent search to minimize a cost function, e.g. the snakes model [1], template matching [2], mean shift [3]. In [1], Kass et al. introduced the snakes model is to obtain a tight contour enclosing the object by minimizing an energy function. In [2], the cost function is defined as SSD between the observation candidate and a fixed template. Then the motion parameters are found by minimizing the cost function through a gradient descent search. In [3], mean shift is considered as an approach for estimating the gradient of a density function and applied to visual tracking. in which the cost function between two color histograms is minimized through the mean shift iterations. In contrast, stochastic methods introduce some stochastic factors into the searching process in order to have a higher probability of reaching the global optimum of the cost function, e.g. MAP (maximum a posterior) [4], particle filter [5]. In [4], object tracking is viewed as an online MAP (maximum a posterior) problem, which is solved by randomly generating a large number of particles to find the maximum of the posterior distribution. In [5], the particle filter represents arbitrary probability distributions using sets of samples and potentially handles nonlinear non-Gaussian dynamics. Such simple models cannot match the complexity of human motions. In general, deterministic methods are usually computationally efficient and stochastic methods suffer a large computational load, especially in high-dimensional state space.

The mean shift algorithm is a well-known deterministic tracking algorithm and has the advantage of low complexity, robustness and invariance to object deformations. However, since the target is modeled by color and the histogram features, spatial information of the target is lost, which fails to tracking object and deal with occlusions. Moreover, it is easy to lose the object due to intrinsic local maxima especially when the tracked objects quickly move. To improve the performance of the traditional mean shift, we present an efficient and robust fragments-based multiple kernels tracking algorithm with adaptive scale selection, which is based on describing the target with multiple fragments to help maintain the spatial information. A voting strategy is applied to obtain more reliable object for tracking with high confidence. A scheme to fuse the log-likelihood ratio image and morphological operation divides the object into some fragments and effectively derives a mean shift type algorithm to locate the new object with very low computation. Applying adaptive scale selection and updating schema for the target model and the weighting factors of each fragment can improve tracking effectiveness. The experimental performances demonstrate that it is effective and robust on several challenging sequences.

* Corresponding author.
E-mail addresses: fangchj2002@163.com (J. Fang), huaxiang08@gmail.com (H. Liu).

1434-8411/$ – see front matter © 2011 Elsevier GmbH. All rights reserved.
doi:10.1016/j.aeue.2011.02.013

2. Related work

Recently, numerous improvements to the mean shift tracking algorithm have been proposed and implemented. In [6], one of the preferred solutions involves slowly adapting the target model by taking a combination of both the original target model and the most recent candidate model. In [7], Collins and Liu presented using linear combinations of RGB color channels as a feature pool and selecting the features for tracking based on the log likelihood image. Each likelihood image generates an estimator location using mean shift algorithm. In [8], a new adaptive mixture color model is presented where the original reference model is kept unchanged, but competes constantly with another model that is rapidly updated. Each candidate model is compared with both the original target model and the most recent winning candidate model. A Markov chain decides which of the models is to be used for the comparison in the next time instant. In [9], Baskoro et al. proposes an online tracking algorithm, which consists of a learning stage and estimating stage. The learning stage selects the features and the estimation stage composes a likelihood image to track the object. However, the tracking process needs considerable computational time especially when the object region is very large.

There are also a number of literatures on the problem based on fragments of object tracking. In [10], Adam et al. presented fragment-based tracking which accounts for partial occlusions. It uses a computationally and memory expensive technique called integral histograms. Through exhaustively searching, there is no formal framework by which we can selectively use certain fragments. In [11], Jaideep et al. proposed a weighted fragment-based approach. The weights are derived from the difference between the foreground and background colors. The coordinates of the winning fragment of the object location using the most confident estimate is used to obtain the object position. However, there is no self-driven ways by which we can get the most confident estimate of the object. Moreover, the intrinsic feature selection would result in a tracking drift or decreasing the weights of some target blobs. In [12], a fusion scheme has been proposed to fuse multiple spatially distributed fragments. Under the fusion scheme, a mean shift type algorithm which allows efficient target tracking with very low computational overhead. However, the weight of each fragment in occlusion will result in draft. Multipart tracking is also described in [13], but the robustness to occlusion is not explored and there is no mechanism to selectively discard badly tracked fragments.

In this paper, the remaining of this paper is organized as follows. In Section 3, the formulated framework is described and the fragments-based mean shift algorithm is derived. The fragmentation method, including foreground feature separation, fragments divided, and specifying the weighting factors are presented in Section 4. The implementation issues are described in Section 5. Experimental results illustrating the performances of the tracker are discussed in Section 6. Finally concluding remarks and future works are given in Section 7.

3. Problem formulation

3.1. Formulated framework

Let a target template T and the current frame I, we want to find the position and the scale of a region in the image I which is closest to the template T in some sense. As noted in Adam et al. [10], the fragments described cannot overlap (Fig. 1).

Initially, the target to be tracked is divided into some fragments by fusing the log-likelihood ratio image and morphological operation and the weight for each fragment is assigned, which will be discussed in detail in Section 4.3. Each fragment is considered as

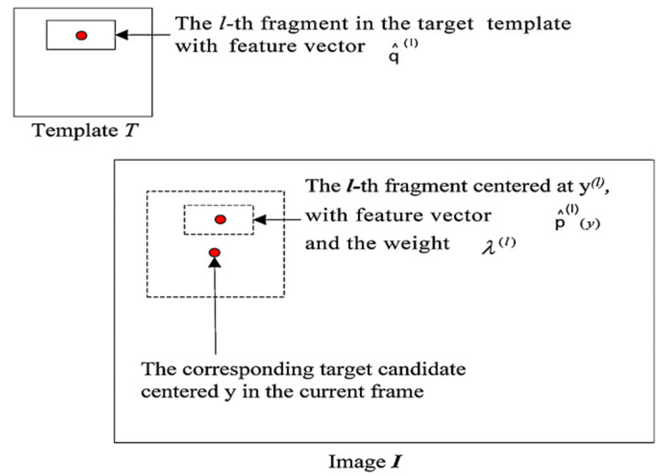

Fig. 1. The target template T and the corresponding target candidate fragment with a hypothesized position centered at y.

a tracker with a kernel function and the object to be tracked is divided with multiple kernels. That is why it is called fragment-based multiple kernels tracking. A target model for each fragment is built using feature histograms $\{\hat{q}^{(l)}\}_{l=1,2,...L}$ (where L is the number of fragments). Each $\hat{q}^{(l)}$ is built by taking a histogram after centering an Epanechnikov kernel in each image fragment. Once a new frame from the video sequence arrives, the target candidate for each fragment centered at position centered at $y^{(l)}$ is represented by $\{\hat{p}^{(l)}(y^{(l)})\}_{l=1,2,...,L}$, where $\hat{p}^{(l)}(y^{(l)})$ is built in the same manner as the target model. The similarity function $\hat{\rho}^{(l)}(y^{(l)})$ for each fragment based on Bhattacharyya coefficient between the template and the target candidate is defined. The fragments-based mean shift algorithm is applied to find which regions in the frame are most similar to the target model. To improve the robustness of mean shift algorithm, the updating schema for the target model and the weighting factors of each fragment are applied, which also can contribute to the performance of foreground and background separation. In view of these advantages, the proposed tracking algorithm can provide more accurate performance.

With the models described above, the similarity function $\hat{\rho}(y^{(l)})$ is the sum of each fragment with different weight and defined.

$$\hat{\rho} = \lambda^{(1)}\hat{\rho}^{(1)} + \lambda^{(2)}\hat{\rho}^{(2)} + \cdots + \lambda^{(L)}\hat{\rho}^{(L)} = \sum_{l=1}^{L}\lambda^{(l)}\hat{\rho}^{(l)} \tag{1}$$

where $\lambda^{(l)}$ denotes the importance level of l-th fragment and is set $\sum_{l=1}^{L}\lambda^{(l)} = 1$. The higher the value is, the more reliable and effective the corresponding fragment is. The l-th fragment similarity function $\hat{\rho}^{(l)}(y^{(l)})$ and the distance distribution based on the Bhattacharyya coefficient between the template and the target candidate are defined as follows, respectively.

$$\hat{\rho}^{(l)}(y^{(l)}) = \rho[\hat{p}^{(l)}(y^{(l)}), \hat{q}^{(l)}] = \sum_{u=1}^{m}\sqrt{p_u^{(l)}(y^{(l)})q_u^{(l)}} \tag{2}$$

$$d^{(l)}(y^{(l)}) = \sqrt{1 - \rho[\hat{p}^{(l)}(y^{(l)}), \hat{q}^{(l)}]} \tag{3}$$

where m is the l-th fragment number of bins of the color histogram. So the object distance distributions should be the sum of multiple fragments.

$$d = \sum_{l=1}^{L}\sqrt{1 - \rho[\hat{p}^{(l)}(y^{(l)}), \hat{q}^{(l)}]} \tag{4}$$