Journal of Signal Processing, Vol.18, No.4, pp.229-232, July 2014

SELECTED PAPER AT NCSP'14

Human Tracking with Particle Filter Based on Locally Adaptive Appearance Model

Sangeun Lee and Keiichi Horio

Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan
Phone/FAX: +81-93-695-6127
E-mail: lee-sangeun@edu.brain.kyutech.ac.jp, horio@brain.kyutech.ac.jp

Abstract

In previous work, we proposed a human tracking algorithm based on the reliable appearance model (RAM). The RAM is a set of discriminative local image descriptors that is selected by a boosting algorithm to identify a target in the initial frame, and is employed as an observation model in a particle filter. As the appearance model of the target in human tracking constantly changes as time passes owing to changes in pose, it is necessary to adaptively update the RAM to improve the tracking accuracy. In this paper, if necessary, an insufficient local image descriptor for robust tracking is updated. In order to classify whether local image descriptors are suitable or not during tracking, a distance histogram corresponding to a local image descriptor is constructed. When the histogram indicates that the local image descriptor is lacking in tracking performance, then it is updated. The experimental results demonstrate that the adaptive appearance model successfully tracks sport players even when their pose often changes.

1. Introduction

Object tracking has been a hot research topic in computer vision for its importance in applications. In particular, human tracking, where a person's body or face is considered as the target, is important [2]. Furthermore, it has applications in the areas of human-computer interaction, robot vision, surveillance and so forth. However, a person may exhibit a variety of postures as shown in Fig. 1, making tracking complex and difficult. To deal with this problem, a tracking algorithm with a robust appearance model is required. In this paper, we address the problem of tracking a person whose geometric appearance is constantly changing over time.

Traditional tracking methods can be divided into two approaches: sampling-based methods (stochastic approach) and detection-based methods (deterministic approach). The Adaboost detector is a widely known deterministic approach in which the target is tracked by detecting it. The trajectory is estimated using an association-based technique by detecting the target. As a stochastic approach [3, 4], particle filter, in which the posterior distribution is approximated by particles and their weights, can handle non-Gaussianity and multi modality. In the particle filter, it is important to design an appearance model for calculating a likelihood from an observed state. The likelihood indicates the fitting between an observed state and a feature of the target, and it directly affects the tracking accuracy. The appearance model in the image plane is represented as a histogram based on color or shape information. Using a boosting approach, a strong classifier that locally selects weak classifiers is applied to the appearance model, so that partial occlusion can be handled and calculation efficiency can be achieved. Furthermore, the appearance model is updated adaptively to respond to the changing appearance of the target. By updating the appearance model, the tracking performance becomes better than that of conventional methods.

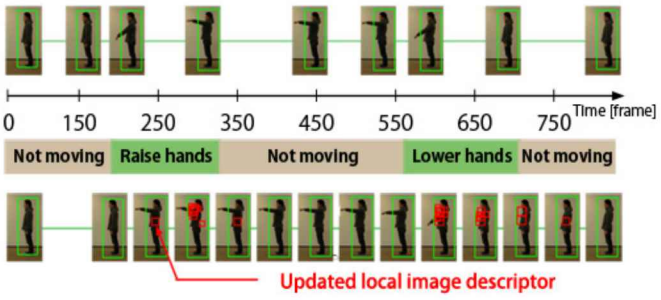

Figure 1: Example of image sequences that contain changes in pose

The contribution of this paper is to present an updating scheme for the evolution of an appearance model based on a local image descriptor. In previous work, we proposed the reliable appearance model (RAM) and employed it as an appearance model to calculate likelihood in the particle filter [5]. The RAM consists of discriminative local image descriptors that are selected by a boosting algorithm. The local image descriptor contains histogram-based color or shape information. The histogram-based information represents the geometric variation to some degree. However, the local image descriptor in the RAM is insufficient for robust tracking when the pose of the target is frequently changing. We

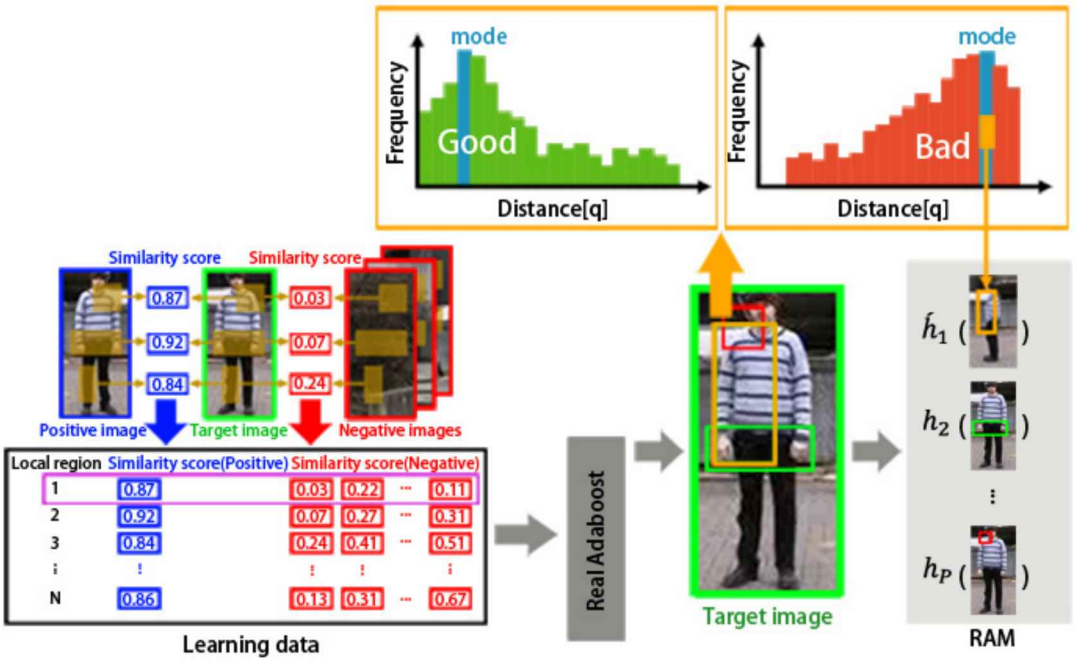

Figure 2: Scheme of proposed method

have devised a scheme that reselects the RAM by performing boosting when the tracking accuracy gradually decreases. The conventional approach has adequate time for reselecting and reemploying the RAM. Finally, it reads to an extra delay for tracking with online.

In this paper, we describe a new scheme in which the local image descriptor is updated adaptively if necessary. For the decision of updating, a histogram corresponding to a local image descriptor is constructed during tracking. If the histogram is classified as lacking in robustness, the local image descriptor is updated by one characterized up to the present time.

2. RAM-Based Human Tracking

We present details of the RAM scheme shown in Fig. 2 and the tracking algorithm in this section.

2.1 Particle filter

The visual tracking problem is efficiently formulated by considering Bayesian filtering. Given a state, the posterior probability $p(X_t|Y_{1:t})$ is computed by

$$p(X_t|Y_{1:t}) \propto$$

$$p(Y_t|X_t) \int p(X_t|X_{t-1})p(X_{t-1}|Y_{1:t-1})dX_{t-1} \quad (1)$$

where X_t is the state at time t and $Y_{1:t}$ consists of all the observations up to time t. Since the visual tracking problem is non linear and non-Gaussian, Eq. (1) is approximated by m particles with weight w_t^m and is written by

$$p(X_t|Y_{1:t}) \approx p(Y_t|X_t) \sum_i w_{t-1}^{(m)} p(X_t|X_{t-1}^{(m)}) \quad (2)$$

The motion model $p(X_t|X_{t-1})$ denotes the characteristic of the target motion at time t obtained by predicting the next state X_t based on the previous state X_{t-1}. The appearance model $p(Y_t|X_t)$ describes the appearance of the target at time t while measuring the likelihood between the target appearance and the observation.

2.2 Reliable appearance model (RAM)

The appearance model used for calculating likelihood in the particle filter is designed with discriminative local image descriptors classified by a boosting algorithm. A local image descriptor

$$d_{i,j} = I(r_i, \lambda_j) \quad (3)$$

is established as a histogram-based feature λ_j at region r_i. The feature is calculated by HUE in HSV color model and HOG [1], where HUE and HOG represent color and shape information, respectively. In the initial frame, a target image and learning images are manually generated. Here, a positive image is a mirror-reversed image of the target image. The

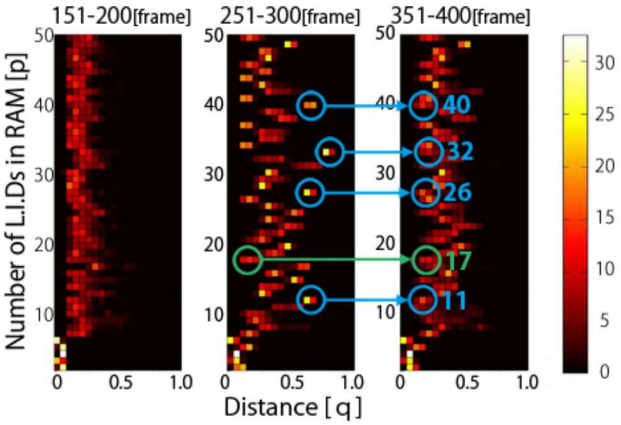

Figure 3: Difference mode of RAM showing a set of local image descriptors in Fig. 1 over time

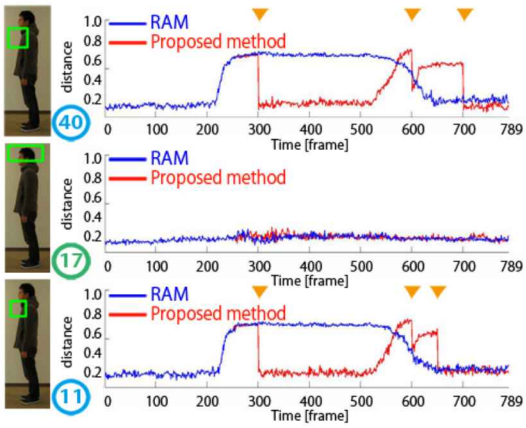

Figure 4: Comparison of local image descriptors between RAM and proposed method: The yellow represents the updating point.

negative images are sampled by sliding windows in the background of the target image. The learning sample

$$l_{i,j} = S(I_1(r_i, \lambda_j), I_2(r_i, \lambda_j)) \tag{4}$$

for the boosting algorithm is computed by the Bhattacharyya distance as a similarity score for corresponding regions in the target image and the positive or negative images. The discriminative regions and features are classified as a weak classifier $h_p(x)$. We use Real Adaboost as the boosting algorithm.

$$H(x) = \sum_{p=1}^{P} h_p(x) \tag{5}$$

A set of classifiers is employed as the appearance model to calculate the likelihood of a response in the particle filter. This method is called the RAM in this study.

3. Online Update of Appearance Model

The RAM is only selected at the initial frame. The appearance model is suitable for partial and full occlusion, whereas it is lacking in dynamic poses of the target. To cope with this problem, we propose an approach based on the construction of a histogram in this section.

3.1 Construction of distance histogram

We utilize the m th weight $w_t^{(m)}$ and the distance $dis_{t,m,p}$ of corresponding local image descriptors between the RAM and the observed state to build distance histograms at time t. The distance histogram $his_{p,q}$ of p th local image descriptor is quantized using q bins and is written as

$$his_{p,q} = \sum_t \sum_m w_t^{(m)} \quad \text{if} \quad \mathrm{th}_q \le dis_{t,m,p} < \mathrm{th}_{q+1} \tag{6}$$

The distance histogram is generated with an interval of 50 frames in this work. The parameters of the number of local image descriptors and the quantitation rate are 50 and 20, respectively.

3.2 Decision and online update

The decision to update a local image descriptor is determined by considering the histogram distribution at that time. We simply calculate the mode, the number which appears most often in a set of numbers. The mode in Fig. 3 indicates whether or not the local image descriptor is adequate. If the mode is located at a long distance, the local image descriptor has poor matching. In other words, it is recognized as lacking in accuracy of tracking and is updated adaptively. A feature, which is the maximum weighted image descriptor in the mode bin, is identified using the adaptive feature of the local image descriptor in the RAM. This is because the maximum weighted image descriptor has a high matching score between a set of local image descriptors in the RAM and the observed state, even if a local image descriptor in the RAM has significantly poor matching as compared with other descriptors.

Table 1: Comparison of tracking accuracy

Sequence	VTD*	Superpixel	RAM	Proposed method
Skating1^L[4]	9	-	18	16
Skating2[4]	19	-	15	15
Basketball[4]	10	6	71	8
Bolt[3]	-	6	14	14
Girl[3]	-	21	14	14

Figure 5: Tracking result of five image sequences $Skating1^L$, $Skating2$, $Basketball$, $Bolt$ and $Girl$ [3, 4]: Violet rectangles demonstrate the ground truth provided by [3, 4]. Blue and green rectangles denote the results of the RAM and the proposed method, respectively.

4. Experimental Results

We tested five video sequences. It was found that our method can handle a cluttered background, large movements, heavy occlusion and changes in light. Our method was compared with three different methods: Superpixel [3], VTD* [4], and RAM [5]. The rectangle obtained from our proposed method in Fig. 5 is wider than the ground truth. This is because our method is designed to extract shape information from the skeleton image. Thus, the average error of the center location in a pixel is employed for fair comparison and is summarized in Table 1. The proposed method achieves improved performance compared with the conventional RAM. For the sequences $Skating2$ and $Girl$, the proposed method was qualitatively superior to the other state-of-the-art methods.

5. Conclusions

In this paper, a particle filter based on local adaptive image descriptors was proposed for human tracking. As shown in Fig. 1, local image descriptors in a region with changed appearance are adaptively updated online when a change in pose occurs. Consequently, the performance of our method is superior to that of the RAM. The experimental results demonstrated that the proposed method outperformed conventional tracking algorithms in terms of accuracy.

References

[1] N. Dalal and B. Triggs: Histograms of oriented gradients for human detection, Proc. IEEE Int. Conf. on Comput. Vision and Pattern Recognition, pp.886–893, 2005.

[2] A. Yilmaz, O. Javed and M. Shah: Object tracking: A survey, ACM Computing Surveys (CSUR), Vol.38, No.4, pp.1–45, 2006.

[3] S. Wang, H. Lu, F. Yang and M.-H. Yang: Superpixel tracking. Proc. IEEE Int. Conf. on Comput. Vision, pp.1323–1330, 2011.

[4] J. Kwon and K. M. Lee: Tracking by sampling and integrating multiple trackers, IEEE Trans. Pattern Anal. Mach. Intell., Vol.36, Issue 7, pp.1428–1441, 2013.

[5] S. Lee and K. Horio: Human tracking using particle filter with reliable appearance model, Proc. SICE Annual Conference 2013 (SICE2013), pp.1418–1424, 2013.