ISSN : 0974 - 7435 Volume 10 Issue 23

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(23), 2014 [14203-14208]

Non-rigid object tracking via discriminative features

Qian Wang[1], Qingxuan Shi[2]*

[1]Academic Administration, Hebei University, Baoding, (CHINA)

[2]College of Mathematics and Computer Science, Hebei University, Baoding, (CHINA)

E-mail: shiqingxuan@hbu.edu.cn

ABSTRACT

Non-rigid objects are typically complex and difficult to track due to the appearance change caused by geometric changes. In this paper, we model the appearance of non-rigid objects by discriminative features which are adaptively selected according to their descriptive ability. To adapt to the geometric changes, we use a deformable rectangle to represent the object, and use Markov Chain Monte Carlo-based Particle Filter (MCMC-PF) to estimate the state of the object in a restricted four-dimensional space. Experimental results show that the proposed tracking algorithm has ideal performance.

KEYWORDS

Non-rigid object; Tracking; MCMC-PF; Discriminative feature.

INTRODUCTION

Visual object tracking has attracted a great deal of attention in the computer vision community due to its wide applications in motion-based recognition, automated surveillance, video indexing, human-computer interaction[1]. It is essentially the problem of finding the most likely estimation of the object state given a sequence of observations. In real-world, objects are typically complex and difficult to track because of complex object motions, non-rigid or articulated nature of objects, partial or full object occlusions, complex object shapes, scene illumination changes, etc. In this paper, we focus on addressing the difficulty of tracking non-rigid objects whose geometric appearance drastically changes as time goes on. Figure 1 shows some tracking examples of such objects by our method.

Facing extreme geometric appearance changes, an effective appearance model is indispensable. Dividing the object into a set of local parts is a prevalent strategy (e.g.[2-5]). In these approaches, object is represented by a multi-part model in which local parts are loosely connected, allowing some degree of spatial deformation. These part-based models have geometrical deformability, at the same time, offer a more flexible mechanism for updating. Because each local part is a visual model on its own, the updating can be performed by removing the parts that exhibit signs of drifting and adding new parts while keeping the parts that are still performing well intact. While in this kind of methods, the flexibility of the model is hard to control. Cehovin[6] proposed a coupled-layer visual model which combined a set of patches together with global appearance information. The local layer geometrically constrained the changes in the target's appearance, and the global layer was updated using the stable patches from the local layer. Despite the appearance model, online updating the holistic appearance model while tracking is a major method to adapt for the appearance changes[7-9], but when should the visual model be updated and how to update the model or model set are key problems. Wang, Junqiu[10] combined the current model with the initial one according to the similarity between the initial and current appearance of the object.

Color-based features have been proved robust and versatile long time ago[11,12]. They are especially appealing for tracking tasks where the spatial structure of the tracked objects exhibits a dramatic variability. Although tracking based on color histogram appearance models can perform well through partial occlusion and pose variation, tracking success or failure depends primarily on how distinguishable the object is from its surroundings[13]. In real scenes, appearance of both object and background will change as the target object moves. It is the ability to distinguish between object and background that is most important. According to this, in this paper, we model the target object using discriminative color features, which are extracted based on their descriptive ability.

At the end, an effective tracking framework is indispensable. The particle filter[14] has shown efficiency in handling non-gaussianity and multi-modality, but faces with the problem of particle degeneracy after several iterations. To solve this problem, Markov Chain Monte Carlo was proposed[15]. This algorithm samples from posterior distribution directly according to a proper proposal distribution, in which a Markov chain converging to the posterior distribution is built. Compared to sequential Monte Carlo method, MCMC samples from the posterior distribution efficiently, and can approximate the target distribution with fewer particles. In this paper, we adopt MCMC-PF as our tracking framework.

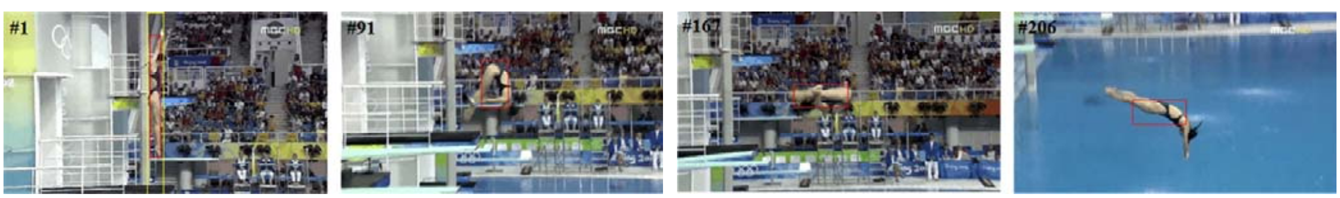

Figure 1 : Example of tracking results in *diving* seq. In this sequence, the geometric appearance of the object changes drastically

The rest of this paper is organized as follows: The discriminative features used in our model are described in Section 2. Section 3 gives the MCMC-PF tracking frame. Section 4 discusses the experimental implementation. Section 5 concludes this paper.

DISCRIMINATIVE COLOR FEATURE

Based on the philosophy of[13], like[16], We select the most discriminative color features to model the target object. Given the object area in image, we first get the color histograms of the object(H_{obj}) and the background(H_{bg}) in *HSV* color space with N bins. The background's histograms are built in the region around the target. The size of the background region depends on the size of the target. The ratio of the two sizes(width and height) is 1.5:1 in this work. The log-likelihood ratio of these two histograms can then be easily calculated with respect to each bin:

$$L_i = log\left(\frac{H_{obj}(i)+\epsilon}{H_{bg}(i)+\epsilon}\right) \quad \textbf{(1)}$$

Where L_i is the log-likelihood ratio of the ith color bin, i ranges from 1 to N. ϵ is a very small constant (which is set to 0.001 in this work) to avoid infinity when H_{obj} or H_{bg} approaches zero. The nonlinear log-likelihood ratio maps object/background distributions into positive values for colors distinctive to the object and negative for colors associated with the background. For colors that are shared by both object and background, their L_i values tend towards zero. It's clear, when a color mainly appears in the foreground and rarely appears in the background, its L_i value would be high, and vice versa. Accordingly, we consider the *i*th color bin as a distinctive color feature for the foreground if L_i is larger than a threshold TL, and consider the color bin as a distinctive color feature for the background if $L_i < -TL$. A discriminative color feature list of the target object *Fobj* is then maintained, which contains the most distinctive colors of the object against the background.

Let $Fobj(i)$ be the *i*th discriminative color feature, the descriptive ability of $Fobj(i)$ can be evaluated by its *L* value. That is, color features with higher log-likelihood ratio have stronger descriptive ability. Figure 2 shows us the discriminative color features and the confidence map corresponding to them. In Figure 2, (a) is the original image of the first frame, where red rectangle is the target object and the yellow rectangle is the background area, (b) is the target mask extract based on *Fobj*, (c) is the confidence map corresponding to the discriminative color features, (d) shows the tracked object in red rectangle.

We observe that such a simple description based on discriminative features is very powerful at distinguishing the object from the background.

MCMC-PF FOR TRACKING

Bayesian formulation

In this paper, we formulate visual tracking as a dynamic Bayesian inference task with a hidden Markov model. Let X_t denote the object state at time t. Given a series of observations $Z_{1:t} = \{Z_1, Z_2, ..., Z_t\}$ up to time t, our aim is to estimate the hidden state variable X_t. According to Bayesian rule, the filtering distribution can be recursively estimated by:

$$p(X_t \mid Z_{1:t}) \propto p(Z_t \mid X_t) \int p(X_t \mid X_{t-1}) p(X_{t-1} \mid Z_{1:t-1}) dX_{t-1} \qquad \textbf{(2)}$$

Where $p(X_t \mid X_{t-1})$ is the prior model describing the temporal evolution of the state variable which predicts the next state X_t based on the previous state X_{t-1}, and $p(Z_t \mid X_t)$ is the observation likelihood that measures the similarity between the observation at the estimated state and the given model.

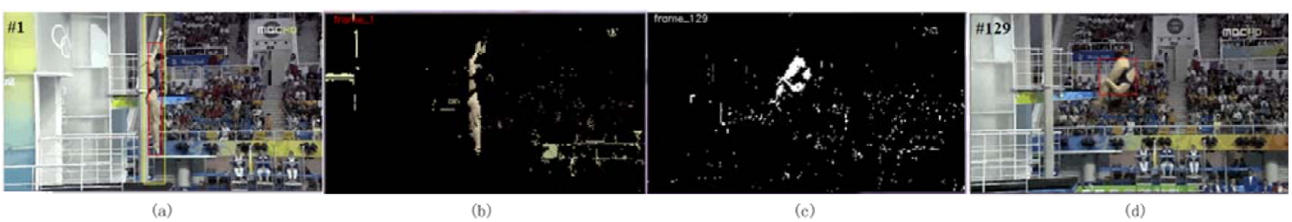

Figure 2 : Tracking the object based on the discriminative color features

Particle filter (PF)[17] approximates the filtering distribution(2) by a set of weighted particles. Given the particles set $X_{t-1}^{(i)}$ and their respective weights $w_{t-1}^{(i)}$ at time step $t-1$, the filtering distribution can be recursively approximated by

$$p(X_t \mid Z_{1:t}) \propto p(Z_t \mid X_t) \sum_{i=1}^{n} w_{t-1}^{(i)} p(X_t \mid X^{(i)}) \qquad \textbf{(3)}$$

An mixture proposal density $\sum_{i=1}^{n} w_{t-1}^{(i)} p(X_t \mid X^{(i)})$ is often employed to sample n particles $\{X_t^{(i)}\}_{i=1}^{n}$ at time step t, and each particle $X_t^{(i)}$ is then weighted by its observation likelihood $p(Z_t \mid X_t)$. In many tracking applications, it is typical to estimate the object state at each time step based on the MAP estimation, or sample expectation on the weighted particles set. With the posteriori probability $p(X_t \mid Z_{1:t})$ computed by the observation model and the transition model, we obtain the Maximum a Posteriori (MAP) estimate over the N samples at each time t.

$$\hat{X}_t = \arg\max_{X_t} p(X_t \mid Z_{1:t}) \qquad \textbf{(4)}$$

MCMC-PF

To achieve a good approximation to the filtering distribution, MCMC algorithm can be used to simulate a Markov chain that converges to a stationary distribution, i.e., filtering distribution. Given the target distribution $p(X_t \mid Z_{1:t})$, an Metropolis–Hastings (MH) sampling step includes drawing a candidate sample X' based on the current sample X_t, by using a proposal $Q(\cdot; X_t)$. With these notations, a Markov chain grows and moves to the new sample with the acceptance probability, i.e.,

$$\alpha(X'; X_t) = \min\{1, \frac{p(X' \mid Z_{1:t})Q(X_t; X')}{p(X_t \mid Z_{1:t})Q(X'; X_t)}\} \quad \textbf{(5)}$$

Otherwise remains at X_t.

In this paper, we replace the traditional importance sampling step in the particle filter with an MCMC sampling step. The MCMC-PF algorithm is outlined in Algorithm 1.

Algorithm 1 MCMC_PF

Input: **Particles set** $\{X_{t-1}^{(i)}\}_{i=1}^{N}$ **}, Observation** Z_t **;**
Output: **Particles set** $\{X_t^{(i)}\}_{i=1}^{N}$ **}, MAP estimation** X_t^{MAP} **;**
Select one sample $X_{t-1}^{(j)}$ from particle set $\{X_{t-1}^{(i)}\}_{i=1}^{N}$ randomly;
Generate initial sample $X^{(0)_t} \sim p(X_t \mid X_{t-1}^{(j)})$;
For $i = 1$ to N
Propose a candidate sample $X' \sim Q(\cdot; X_t^{(i-1)})$;
Calculate the acceptance ratio α according to (5);
Accept X' with probability α, and let $X_t^{(i)} = X'$,
else let $X_t^{(i)} = X_t^{(i-1)}$;
End For
MAP estimate X_t^{MAP} according to (4).

EXPERIMENTS

Implementation Details

***Object state*:** In this paper, we represent the object as a rectangular region. The state of the object is defined by its position in the scenario and the scale compared with the original size, i.e., $X = \{x, y, s_x, s_y\}$. Where (x, y) is the center position of the object and s_x, s_y are the horizontal direction scale and the vertical direction scale respectively.

***Appearance model*:** We use a discriminative feature set $Fobj = (Fobj(i), w_i), i = 1, 2, ..., N_F$ to define the appearance model, where N_F is the number of features, w_i weighting the *i*th feature according to its descriptive ability (which is evaluated by its L value). To extract the color feature, we convert the RGB space into the HSV space, and quantize it into 110 bins ($N_h = N_s = N_v = 10$). After we get the color histograms of the object and the background, L value is calculated according to (1).

***Observation*:** Based on the appearance model, we define the observation likelihood function for the observation Z_t in the image given the state X_t as:

$$p(Z_t \mid X_t) = \sum_{i=1}^{N_F} w_i \sum_{j=1}^{N_P} I(HSV(p_j)) = Fobj(i)) \quad \textbf{(6)}$$

Which describes the distance between the target model and the candidate. In this definition N_P is the number of the pixels in the observation region, w_i donates the weight of *i*th feature $Fobj(i)$, I is indicator function, and HSV function returns the HSV bin of pixel p_j.

***Motion model*:** The motion model between frames is defined bellow:

$$p(X_t \mid X_{t-1}) = \begin{cases} \mathrm{N}\ (\mu_p, \sigma_p^2) & objectposition \\ \mathrm{N}\ (\mu_{s_x}, \sigma_{s_x}^2) & objectscale \end{cases} \quad \textbf{(7)}$$

Where $N(\mu_p, \sigma_p^2)$ is a normal distribution on the 2-D space with mean $\mu_p = A_1 * \hat{X}_{t-1}^{p} + A_2 * \hat{X}_{t-2}^{p}$ and variance σ_p^2, A_1 and A_2 are the constants in this $AR2$ model. $N(\mu_{s_x}, \sigma_{s_x}^2)$ is a normal distribution with mean $\mu_{s_x} = A_1 * \hat{X}_{t-1}^{s_x} + A_2 * \hat{X}_{t-2}^{s_x}$ and variance $\sigma_{s_x}^2$. The other component of scale s_y is get by $s_y = 1/s_x$. So, the four-dimensional object state space is constraint by $s_x * s_y = 1$.

Experimental results

We evaluate our tracking algorithm on several challenging video sequences which are sports video clips. In these video sequences, the tracked objects all contain large amount of extreme geometric appearance changes. For each sequence, the position of the tracked target is manually labeled in the first frame.

To extract discriminative color features, the color space is divided into 110 HSV color bins, and the threshold TL is set to 0.35. In Sampling period, we use 600 samples to approximate the posterior distribution of the object. We sample three video sequences (*diving*, *tennis*, *gymnastics-woman*), and show the tracking results in Figure 1, Figure 3 and Figure 4 respectively.

CONCLUSION

In this paper, we use discriminative color features to model the object's appearance, and embed it to the MCMC-PF tracking framework. Experiments on the particular sequences demonstrate that we can successfully track the objects even they are experiencing extreme geometric appearance changes. Although only color features are considered in this paper, the proposed approach can be extended to other appearance cues represented as histograms of feature values. At the same time, the appearance model used here can be introduced to other tracking frameworks.

ACKNOWLEDGEMENTS

This work is supported in part by National Natural Science Foundation of China (No: 61375075) and Natural Science Foundation of Hebei Province (No: F2012201020).

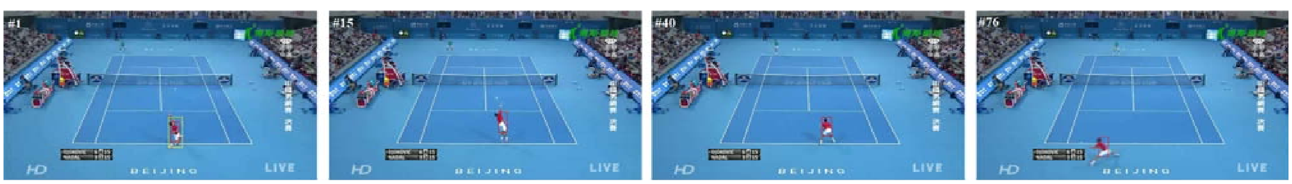

Figure 3 : Tracking results on *tennis* seq

Figure 4 : Tracking results on *gymnastics-woman* seq

REFERENCES

[1] Yilmaz, Alper, Omar Javed, Mubarak Shah; Object tracking: A survey, Acm computing surveys (CSUR), **38(4)**, 13 **(2006)**.

[2] Chang, Wen-Yan, Chu-Song Chen, Yi-Ping Hung; Tracking by parts: A bayesian approach with component collaboration, Systems, Man, Cybernetics; Part B: Cybernetics, IEEE Transactions on, **39(2)**, 375-388 **(2009)**.

[3] Kwon, Junseok, Kyoung Mu Lee; Highly nonrigid object tracking via patch-based dynamic appearance modeling, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **35(10)**, 2427-2441 **(2013)**.

[4] S.M.Nejhum, Jeffrey Ho, Ming-Hsuan Yang; Online visual tracking with histograms and articulating blocks, Computer Vision and Image Understanding, **114(8)**, 901-914 **(2010)**.

[5] Godec, Martin, Peter M.Roth, Horst Bischof; Hough-based tracking of non-rigid objects, Computer Vision and Image Understanding, **117(10)**, 1245-1256 **(2013)**.

[6] Cehovin, Luka, Matej Kristan, Ales Leonardis; Robust visual tracking using an adaptive coupled-layer visual model, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **35(4)**, 941-953 **(2013)**.

[7] Ross, A.David et al.; Incremental learning for robust visual tracking, International Journal of Computer Vision, **(77)**, 1-3, 125-141 **(2009)**.
[8] Babenko, Boris, Ming-Hsuan Yang, Serge Belongie; Robust object tracking with online multiple instance learning. Pattern Analysis and Machine Intelligence, IEEE Transactions on, **33(8)**, 1619-1632 **(2011)**.
[9] Kalal, Zdenek, Jiri Matas, Krystian Mikolajczyk; Pn learning: Bootstrapping binary classifiers by structural constraints. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, **(2010)**.
[10] Wang, Junqiu, Yasushi Yagi; Integrating color and shape-texture features for adaptive real-time object tracking, IEEE Transactions on Image Processing, **17(2)**, 235-240 **(2008)**.
[11] Pérez, Patrick et al.; Color-based probabilistic tracking, Computer vision—ECCV 2002, Springer Berlin Heidelberg, 661-675 **(2002)**.
[12] Comaniciu, Dorin, Visvanathan Ramesh, Peter Meer; Real-time tracking of non-rigid objects using mean shift, Computer Vision and Pattern Recognition(CVPR), 2000 IEEE Conference on, **2**, IEEE, **(2000)**.
[13] Collins, T.Robert, Yanxi Liu, Marius Leordeanu; Online selection of discriminative tracking features, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **27(10)**, 1631-1643 **(2005)**.
[14] Isard, Michael, Andrew Blake; ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework, Computer Vision—ECCV 1998, Springer Berlin Heidelberg, 893-908 **(1998)**.
[15] Choo, Kiam, David J.Fleet; People tracking using hybrid Monte Carlo filtering, Computer Vision, 2001 Eighth IEEE International Conference on, **2**. IEEE, **(2001)**.
[16] Fan, Jialue, Xiaohui Shen, Ying Wu; Scribble tracker: A matting-based approach for robust tracking, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **34(8)**, 1633-1644 **(2012)**.
[17] Isard, Michael, Andrew Blake; Condensation—conditional density propagation for visual tracking, International Journal of Computer Vision, **29(1)**, 5-28 **(1998)**.