Online Selecting Discriminative Tracking Features using Particle Filter

Jianyu Wang[1], Xilin Chen[1,2] and Wen Gao[1,2]
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China.
JDL, Institute of Computing, Chinese Academy of Sciences, Beijing, China.
{jywang, xlchen and wgao}@jdl.ac.cn

Abstract

The paper proposes a method to keep the tracker robust to background clutters by online selecting discriminative features from a large feature space. Furthermore, the feature selection procedure is embedded into the particle filtering process with the aid of existed "background" particles. Feature values from background patches and object observations are sampled during tracking and Fisher discriminant is employed to rank the classification capacity of each feature based on sampled values. Top-ranked discriminative features are selected into the appearance model and simultaneously invalid features are removed out to adjust the object representation adaptively. The implemented tracker with online discriminative feature selection module embedded shows promising results on experimental video sequences.

1. Introduction

One of key factors determining the performance of a 2D blob tracking system is how to choose an appropriate method to represent the object, by which the tracker is robust to background clutters.

For practical applications, object representation methods should mainly satisfy two properties: discriminability and computational efficiency. Considering discriminability, the tracker tends to employ more complicated representation methods to ensure the robustness. On the contrary, computational efficiency requirement limits resources that can be located for evaluating the similarity between the object model and image observations. Therefore, most widely used object representation methods, such as histogram [4], contour [1], template [11] and combining histogram and contour [10], are compromises between these properties.

Previous works mainly select representation modalities experimentally or empirically with the implicit assumption that the selected modality before the task starting can always distinguish interested objects from uncertain and changing background well. Unfortunately, the tracker tends to unstable when this assumption is occasionally violated.

Examining from the viewpoint of classification for foreground/background separation problems, current background patches are the only negative examples. Therefore, background information should be heavily considered when choosing discriminative and efficient object representations. The importance of background information has been approved in [2, 3, 7]. In [3], one from several different color spaces was chosen online to construct the histogram for face tracking. Collins [2] further extended the method to general blob tracking. Through linear combining of R, G and B color channels, 49 kinds of histogram candidates could be adaptively chosen according to their discriminative capacity to background. Above two methods adopted histograms as object model and improved results had been achieved comparing with traditional histogram-based tracking algorithms. Nevertheless, color histograms have limited identification power in many cases (e.g. when some background patches mimic the target in color distribution). In [7], Nguyen proposed to represent the foreground and the background using Garbor filters to cope with large appearance variations of foreground. Although Gabor filters achieve many successes for object recognition tasks, it is not a very good choice for real-time tasks considering computational cost.

In this paper, a method is proposed for maintaining discriminative appearance model by online feature selection. Due to its computational efficiency and strong classification capacity, over-complete Haar wavelet feature dictionary is very suitable for tracking task and a subset of Haar wavelet features is selected to constructing a classifier ensemble for modeling the object appearance. We observe that the stochastic characteristic of particle filter results in many particles corresponding to background areas. When weighting "background" particles using likelihood model, feature

0-7695-2372-2/05/$20.00 (c) 2005 IEEE

values on those background areas are sampled and they are treated as negative examples in this foreground/ background classification task. Then features can be ranked according to their classification power computed on fisher discriminant. Top-ranked discriminative features are selected into the model and simultaneously invalid ones are removed out. This online feature selection procedure is efficient since informative "background" particles naturally exist during the tracking process.

The rest of the paper is organized as follows: The strategy of selecting features by considering background information is described in section 2. In section 3, we point out that there exist informative "background" particles during the tracking process. In section 4, how to maintain the model by online feature reselection are stated. The steps of the whole system are described in section 5. Experimental results are given in section 6 and conclusion is made in section 7.

2. Feature Selection by considering Background Information

In this section, we employ set of simple classifiers to model the appearance of the object and describe the principle of how to choose a group of classifiers from an over-whelming large set by considering background information.

2.1 The Feature Set

In viola's work [4], classifiers based on over-complete Haar basis features demonstrated excellent performance on object recognition task. Furthermore, evaluating Haar feature value is very computational efficient due to the introduction of integral image.

Currently, only three kinds of Haar features are used in our tracker (see fig. 1 and [4] for details of how to evaluate feature values). Denote the feature set as $F = \{f_i \mid i = 1, ..., N\}$ and corresponding feature values on image observation z as $V(z) = \{v_i(z) \mid i = 1, ..., N\}$.

2.2 Constructing Appearance Model

The tracking system classifies one local image observation into an object observation or a background patch based on the values of these simple features. For each simple feature, a classifier is constructed. Various kinds of classifiers can be used here. For computational efficiency, we currently adopt a weak classifier,

$$h_i(z) = \begin{cases} 1 & if\ v_i(z) \in [\overline{v}_i^{pos} - \theta_i, \overline{v}_i^{pos} + \theta_i] \\ 0 & otherwise \end{cases}, \qquad (1)$$

where, $\overline{v}_i^{pos} = \overline{v}_i(z^{pos})$ is the mean value of feature f_i evaluated on object observations(we disturb the right object position by up to 2 pixels in all directions for getting more object observations), θ_i is a threshold and $v_i(z)$ is the value of feature f_i computed on local image window z.

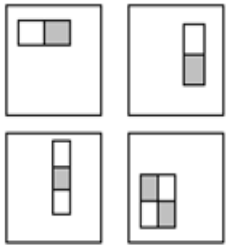

Fig.1. Haar wavelet features used in the current tracker.

Combining these weak classifiers into a strong classifier, AdaBoost has been approved to be a powerful tool with lots of successful applications. Nevertheless, it is computational intensive for online feature selection task. We propose a two-step empirical feature selection method.

Step 1: Selecting features by area filter and information measure. For a 24x24 base size image window, the number of Haar features in the over-complete dictionary is about ninety thousands. The features which the covered image area is less than a threshold (set to 16 in experiments) are filtered out first and left features are sampled spatially with equal spacing. We define a information measure as $E(f_i) = \dfrac{\overline{v}_i^{pos}}{size(f_i)}$. Then remove features f_i with a probability being proportional to $1 - E(f_i) / \max_i(E(f_i))$ to further shrink the feature set. After all above steps, remained features are approximately 4000 and we denote remained feature set as F_{ori}.

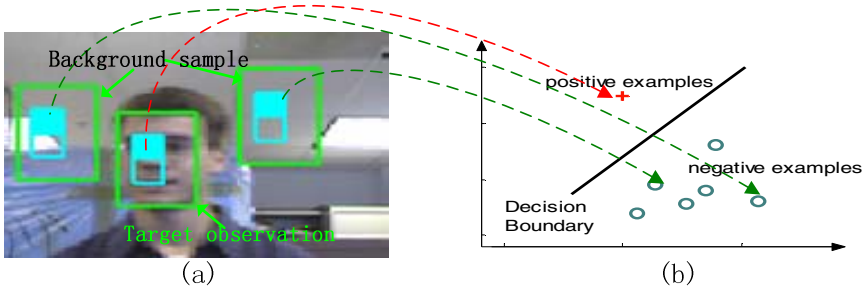

Fig. 2. (a) Demonstrating the process of sampling feature values from object and background and (b) how the sampled feature values are employed for one weak classifier training.

Step 2: Sampling surrounding background to select discriminative features. Suppose sampling *K* different local image windows with *K-p* background patches and p object observations (See fig. (2) for a demonstration on sampling feature values from the object observations and background patchs). Then, for each feature f_i, *K-p* negative examples and *p* positive examples are obtained. Fisher Linear Discriminant (FLD) is adopted to evaluate each feature's classification ability [8]:

$$R_i(f_i) = \frac{|\overline{v}_i^{pos}(f_i) - \overline{v}_i^{neg}(f_i)|^2}{S^{pos}(f_i) + S^{neg}(f_i)}, \qquad (2)$$

where, $\overline{v}_i^{pos}(f_i)$ and $\overline{v}_i^{neg}(f_i)$ are mean feature value of feature f_i from positive and negative examples respectively. $S^{neg}(f_i)$ and $S^{pos}(f_i)$ are the negative and positive class scatters respectively.

In descending order of their classification abilities, weak classifiers are added into the appearance model until

$$(\sum_{m=1}^{M} R_m(f_i)) > T, \qquad (3)$$

where *T* is a threshold to ensure the classification power of combined weak classifiers (In experiments, we set *T*=0.45).

3. Informative "Background" Particles

Particle Filter is a technique for implementing a recursive temporal Bayesian filter by Monte Carlo simulations. The key idea is to represent the required posterior by a set of random samples with associated weights $\{x_t^k, w_t^k \mid k = 1,...,K\}$ [1].

When applying particle filter to a specific task, there are two key components needed to be carefully defined: dynamic models, which determines how the particles are propagated in the state space, and likelihood models, which weight particles and consequently relate the noisy measurement to the hidden state. In practice, there is infinite number of choices of dynamical models and finding the optimal dynamic model is very difficult if not impossible [9]. Non-optimal dynamic model makes predicted states of many particles lie in background areas. There are many other factors, such as particles may be attracted by clutters and diversity requirements of propagated particles, which all inevitably results "background" particles. In previous works, these background particles often regarded as little usage and contribute little to the final results. We argue that these particles contain rich background information and can be used for assessing application dependent tracking situations.

In the next section, we describe how to make use of these informative "background" particles to adjusting the object appearance model.

4. Adjusting Appearance Model during Tracking

To maintain a discriminative and simple appearance model during the tracking process, two model adjusting modules are embedded into the particle filter process to cope with ambiguities and large object appearance variations respectively.

Suppose at time *t*, denote the appearance model as $H(t, M) = \{h_m \mid m = 1,...,M\}$, where h_m is one weak classifier defined in equation (1). The similarity score between a local window *z* and the model is defined as

$$w(z) = \frac{\sum_{m=1}^{M} h_m(z)}{M} \qquad (4)$$

The object appearance model can be divided into two subsets: $H_1(t, M_1)$ and $H_2(t, M_2)$, where $H_2(t, M_2) = \{h_m(z^{pos}) == 0 \mid m = 1,...,M_2\}$ contai

ns those weak classifiers output zero on current object observations and $H_1(t, M_1) = H(t, M) - H_2(t, M_2)$.

4.1 Reselecting Features to Avoid Ambiguities

Ambiguity is one of the main reasons to cause multi-model posterior [5] and to distract the tracker from right positions.

Suppose after the verification step of the particle filter process, an updated particle set $PT(t) = \{x_t^k, w_t^k \mid k = 0, ..., K\}$ is got. A simplified *K*-means clustering is used to cluster $PT(t)$ according to predicted states x_t^k and their associated weights w_t^k. A set of clusters $CP(t) = \{cp_c \mid c = 1, ..., C\}$ can be obtained (*C*=1 represents the special case of uni-model posterior) and one cp_c corresponds to a subset of particles from $PT(t)$. Denote the particle with largest weight within cp_c as $x(cp_c)$ and $w(cp_c)$ is associated weight of $x(cp_c)$.

In descending order of $\{w(cp_c) \mid c = 1, ...C\}$, corresponding $x(cp_c) \in \{x(cp_c) \mid c = 1, ..., C\}$ is examined sequentially whether it has reasonable dynamics according to the object historical motion. Suppose $x(cp_{c1})$ is the first particle satisfying dynamic constraints and we denote it as cp^{pos}. Then, the object positions at time *t* is estimated using particles within cluster cp^{pos} and particles within other clusters are regarded as corresponding to background areas. WE use $CP^{neg}(t) = \{cp_c^{neg} \mid c = 1, ..., C-1\}$ to denote those "background" clusters.

For each "background" cluster cp_c^{neg}, compute

$$\Delta w_c = | w(cp^{pos}) - w(cp_c^{neg}) | . \quad (5)$$

If Δw_c is below a threshold T (*T*=0.17 in experiments), then the background area associated with $x(cp_c)$ is signed as a threat (possible ambiguity), which may distract the tracker in future if the similarity score on object observation degrades.

All feature values collected on threats are used as negative examples to reselect features. First remove $H_2(t, M_2)$ from $H(t, M)$. Then remove features in ascending order according to their classification power till up to only 70% features remained in the origianl appearance model. Replace those removed features from the feature set F_{ori} with the feature selection method described in section 2. The difference is that only consider those threats as background samples instead of all sampled background windows.

4.2 Reselecting Features for Violent Appearance Variations

Only updating model parameters may not follow violent variations of the object appearance. In those cases, reselecting features to maintain the appearance model should be considered.

Suppose at time *t*, if mean weight on object observations $\overline{w}(z^{pos}) = \frac{1}{p}\sum w(z^{pos})$ is below some threshold (in experiments, set it to 0.75), this indicates that the model cannot explain the appearance very well. Then remove subset $H_2(t, M_2)$ and re-select M_2 features from the original feature set F_{ori}.

5. Steps of the Whole Algorithm

Up to discussed here, steps of the proposed method can be described as follows.

Pre-processing: Using procedure described by feature selection step 1 in section 2 to get feature set F_{ori}.

Initialization: Define z^{pos} in the reference image and scale F_{ori} from base size to actual object size. Employ particle filter to sample background patches z^{neg} s and use the feature selection method described in section 2 to constructing initial object model $H(0, M)$ from the feature dictionary.

Prediction: Use dynamic model $P(x_t \mid x_{t-1})$ propagating particles to generate *K* predicted states (In our experiments, nearly constant velocity motion model is adopted as the dynamic model.);

Verification: Compute similarity scores (or un-normalized weights) $W(t)$ using likelihood model $P(y_t \mid x_t)$ (see equation (4)), simultaneously store every observation z_i 's feature values $V(z_i)$.

Background Assessment: As described in subsection 4.1, find those threats in the neighborhood of the object. If threats exist, replace some features to update $H(t, M)$ ensuring that the tracker will not be confused in future.

Model Updating: As described in subsection 4.2, if similarity scores on object observations $\overline{w}(z^{pos}) < T$, replace some features to update target model $H(t, M)$.

Re-sampling: Normalize the weights and compute the covariance of the normalized weights. If this variance exceeds some threshold, then construct a new set of samples by drawing, with replacement, K samples from the old set, using the weights as the probability that a sample will be drawn.

6. Experimental Results

The tracker has been evaluated on some video sequences and representative results are reported.

Some results of tracking people are shown in fig. 3. The sequence is publicly available at [12] and is named as "ThreePastShop2cor.mpg". It has 1527 frames of 384×288 pixels and one person out of three is successfully tracked by the proposed tracker until he is occluded. The main challenges of the sequence arise from that two nearby persons who are very similar to the interested person walk with him together and they change their relative positions during walking (Please note that the tracker input is intensity images and see http://www.jdl.ac.cn/user/jywang/projects/HaarTracking.htm for details and more examples). Tracking results show that the tracker can separate the interested person from the other twos by adjusting the appearance model online. The effectiveness of feature reselection procedure can be shown by fig. 4 and 5. When a person approaches the interested person (see frame 469 and 472 of fig. 3), there appear some threats that may confuse the tracker (see fig. 4). By examining background information, the tracker can detect these threats and reselect features to lower the similarity scores on the threats to distinguish them from the object (see fig. 5 for the result similarity scores after feature reselection procedure).

Fig. 6 shows some results of tracking cars. The sequence has 2803 frames and is recorded with handheld digital camera at the resolution of 320×240 pixels. Additional to the problems of ambiguity, more uncertain factors need to be carefully considered, such as illumination condition (e.g. the shadow cast on the car, see frame 1913 in fig. 6), scale, temporary occlusion and viewpoint variations, in this outdoor sequences. From the tracking results, we can conclude that the tracker can handle these uncertain factors well.

In all experiments, no image pre-processing module is employed and the frame data is directly input to the tracker. Typically the number of features selected to represent object is approximately 120-180. 1089 particles are employed to track the object and the system can run at approximately 10Hz at a Pentium 1.5GHz PC with 256M RAM without specialized optimization.

7. Conclusion

A novel method is proposed for tracking 2D image blobs and experimental results confirm that the method is effective and computational efficient. Main contributions of the work can be concluded as follows:

1) An online feature selection method that can adaptively choose discriminative features from a large set with the aid of "background" particles during the tracking process.
2) Haar features dictionary is introduced into visual tracing applications for efficient and scalable object representation.

Generally, there are two kinds of mainstream methods to deal with visual tracking problems. One is based on foreground matching and another is based on background modeling. The novel idea presented in this paper can be regarded as an attempt to enjoy the advantages of the above two kinds of methods to track the object corporately.

References

[1] M. Isard, A. Blake, CONDENSATION-Conditional Density Propagation for Visual Tracking, *Int. Journal of Computer Vision*, 29(1), pp.5--28, 1998.

[2] R. Collins and Y. Liu. On-line selection of discriminative tracking features. *Proc. IEEE Conf. on Computer Vision*, pp.346-352, 2003.

[3] Stern, H. and Efros, B., Adaptive color space switching for face tracking in multi-colored lighting environments, *Int. Conf. On Automatic Face and Gesture Recognition*, pp.236-241, 2002.

[4] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. IEEE Proc. On Computer Vision and Pattern Recognition, vol II: pp.142–149, 2000.

[5] J. Vermaak, A. Doucet, P. Perez, Maintaining Multi-Modality through Mixture Tracking. *Int. Conf. on Computer Vision*, pp.1110-1116, 2003.

[6] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *IEEE Proc. Computer Vision and Pattern Recognition*, pp.511-518, Dec. 2001.

[7] H.T. Nguyen and A.W.M. Smeulders, Tracking Aspects of the Foreground against the Background, *European Conference on Computer Vision*, vol II: pp. 446-456, 2004.

Fig. 3. Some people tracking results by the proposed tracker. The frame no. is: 370, 472, 506, 523, 544 (top row) and 576, 578, 582, 621 and 746 (bottom row).

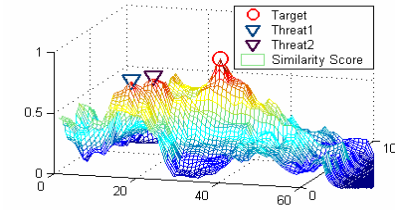

Fig. 4. Distribution of similarity scores on frame 472. There are two threats appears.

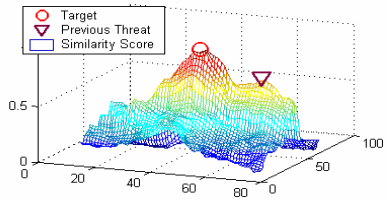

Fig. 5. After feature reselection procedure, the distribution of similarity scores of frame 506.

Fig. 6. Some car tracking results by the proposed tracker. The frame no is: 2, 224, 244, 494, 1035 (top row) and 1913, 2194, 2317, 2488, 2673 (bottom row).

[8] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification (2nd Edition), *Wiley-Interscience Press*, ISBN: 0471056693, October, 2000.

[9] Y. Rui, Y. Chen, Better Proposal Distributions: Object Tracking Using Unscented Particle Filter. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp.786-793, 2001.

[10] S. Birchfield, Elliptical Head Tracking Using Intensity Gradients and Color Histograms, *IEEE Proc. on Computer Vision and Pattern Recognition,* pp.232-237, 1998.

[11] G. Hager and P. Belhumeur, Efficient Region Tracking with Parametric Models of Geometry and Illumi-nation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1025-1039, 1998.

[12] CAVIAR Test Case Scenarios at: http://homepages.inf.ed.ac.uk/rbf/CAVIAR.