ROBUST OBSERVATIONS FOR OBJECT TRACKING

Bohyung Han and Larry Davis

Dept. of Computer Science
University of Maryland
College Park, MD 20742, USA
{bhhan, lsd}@cs.umd.edu

ABSTRACT

It is a difficult task to find an observation model that will perform well for long-term visual tracking. In this paper, we propose an adaptive observation enhancement technique based on likelihood images which are derived from multiple visual features. The most discriminative likelihood image is extracted by Principal Component Analysis (PCA) and incrementally updated frame by frame to reduce temporal tracking error. In the particle filter framework, the feasibility of each sample is computed using this most discriminative likelihood image before the observation process. Integral image is employed for efficient computation of the feasibility of each sample. We illustrate how our enhancement technique contributes to more robust observations through demonstrations.

1. INTRODUCTION

Trackers are based on the some measurement of similarity between the target to be tracked and observations, and various observation methods are used to define this similarity. Intensity or color is natural to use in object tracking, and approaches based on templates [1] and histograms [2, 3, 4, 5] are very common. Also, edges [8] and filter responses [9, 6] are important features for object tracking. Various observation strategies have been proposed, but there is no generally superior observation method for general visual tracking algorithms. So, we instead propose an observation enhancement technique based on likelihood images, which can be incorporated into many tracking algorithms.

A likelihood image represents the contrast information between foreground (target region) and background (its surrounding); it is created by comparing histograms of both areas with respect to some features. It was originally suggested in [10] for tracking problems; there, the most discriminative feature selected from a set of likelihood images is directly used for mean-shift tracking. However, this approach may exhibit poor performance in clutter and can lose the target in spite of its visual salience. Also, the likelihood image can be significantly contaminated by temporary tracking errors. There have been some closely related works [7, 6], but they do not provide an adequate solution to these problems.

In this paper, we present an observation enhancement technique using likelihood images obtained from two different feature spaces – RGB and normalized RGB (rgb). Six likelihood images are created, and a most discriminative likelihood image is extracted by PCA. In order to avoid pollution of the extracted likelihood image by tracking error, we update the subspace incrementally on the assumption that the scene around the target changes smoothly. The "brightness" in the most discriminative likelihood image delivers prior information about the target region, and it is employed to compute the *feasibility* of a sample (to be defined rigorously below) before the measurement step in the particle filter. The feasibility is obtained by the summation of values inside the region in the most discriminative likelihood image, and an integral image [12] is employed for efficient computation. The final weight of each particle is determined by the product of the feasibility and the likelihood in observation. As a result, several independent features are merged through the likelihood images and the merged features are utilized for robust observation via the feasibility computation.

This paper is organized as follows. We describe the likelihood images and the feature extraction in section 2 and 3, respectively. In section 4, tracking in a particle filter framework and experimental results are demonstrated.

2. LIKELIHOOD IMAGES

Likelihood images represent the distinctiveness of a target object from background with respect to a given feature or set of features. For the construction of likelihood images, log-likelihood ratios are obtained first from histograms of foreground and background pixels. Then, the salient region in foreground can be detected by identifying high likelihood ratios.

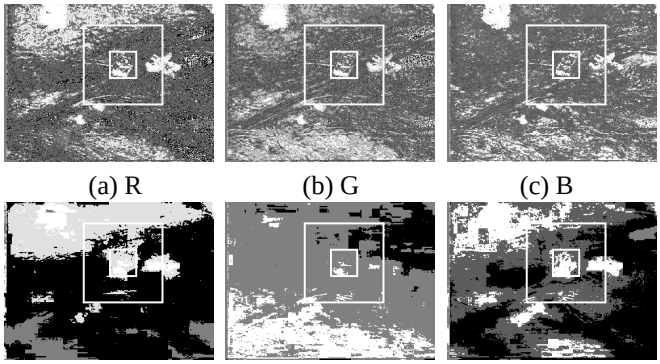
In detail, suppose the foreground is given and the background is regarded as a rectangular region surrounding the

foreground. For a given feature, let $\phi_{fg}(i)$ and $\phi_{bg}(i)$ be the frequency of pixels with value i in the foreground and the background, respectively. The log-likelihood ratio for a feature value i is given by

$$L(i) = \max\left(-1, \min\left(1, \log\frac{\max\left(\phi_{fg}(i), \delta\right)}{\max\left(\phi_{bg}(i), \delta\right)}\right)\right) \quad (1)$$

where δ is a very small number. The likelihood image for each feature is created by backprojecting the ratio into each pixel in the image.

We construct a likelihood image for each color channel in the RGB and *rgb* color spaces, so that 6 different likelihood images are generated for feature extraction. Figure 1 shows example likelihood images derived from each color channel.



(a) R (b) G (c) B

(d) r (b) g (c) b

Fig. 1. Likelihood images in each color channel. For this image, the target in *rgb* space is more distinctive.

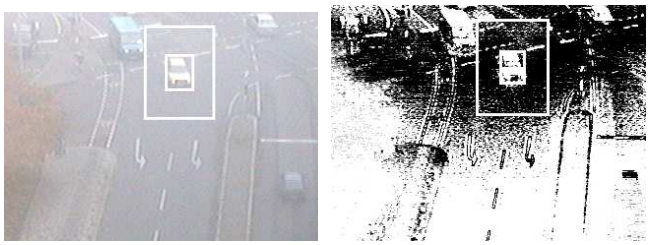
3. FEATURE EXTRACTION BY PCA

3.1. Batch Method

Our objective is to identify the most discriminative likelihood image and measure the feasibility of each particle to improve the observation quality.

There are various feature extraction methods; here, PCA is employed to generate the most discriminative likelihood image. The linear discriminant method may not be appropriate since the histograms of the foreground and the background regions are often multi-modal in the original color image. However, the transformed likelihood image is likely to have positive value pixels in the foreground region while negative value pixels tend to be frequently observed in the background region. Even though the foreground and the background cannot be perfectly separated by a linear hyperplane, we can expect that most pixels would be classified correctly by it. Also, linear methods are much faster than their non-linear counterparts such as Kernel LDA [13] and Kernel PCA [14].

Suppose that S_{fg} and S_{bg} are the set of n-dimensional vectors sampled from the foreground and background area of n likelihood images, and that $\mathbf{m}$ and $\mathbf{V}$ are the $n\times 1$ mean vector and $n \times n$ covariance matrix of these data, respectively. Let $\mathbf{e}_i$ $(i = 1, \ldots, n)$ be the eigenvectors associated with the eigenvalues λ_i which are sorted in non-increasing order. Once $\mathbf{e}_i$ are obtained, the value y projected from the original vector $\mathbf{x}$ to the most discriminative feature space is given by $y = \mathbf{e}_1^T(\mathbf{x}-\mathbf{m})$. The following figure shows an example of the most discriminative likelihood image extracted by PCA.



(a) original image (b) extracted image

Fig. 2. Comparison between original image and the most discriminative likelihood image

The most discriminative likelihood image for the first frame is created by batch processing, but the subsequent ones are constructed by the following incremental method to include previous information.

3.2. Incremental Subspace Update

As described above, the distinctiveness information of the target in likelihood images can be significantly reduced by tracking errors. We alleviate this problem by updating the subspace incrementally rather than using a completely new subspace in each frame.

Since the data added in each frame has as many samples from the foreground or from the background region and the number of dimension is moderate in our application, the standard incremental PCA [15, 11] is not required and a simpler (but accurate) method can be used. Instead of computing eigenvectors without consideration of the full covariance matrix and the matrix decomposition, we just compute the updated mean and covariance with new observations in the current frame and perform SVD to obtain eigenvectors.

Denote by $(\mathbf{m}_{old}, \mathbf{V}_{old})$ and $(\mathbf{m}, \mathbf{V})$ pairs of mean and covariance in the previous and current time step, respectively. Then, the updated mean and covariance $(\mathbf{m}_{new}, \mathbf{V}_{new})$ including the new observations are as follows.

$$\mathbf{m}_{new} = (1-\alpha)\mathbf{m}_{old} + \alpha\mathbf{m} \quad (2)$$

$$\mathbf{V}_{new} = (1-\alpha)\mathbf{V}_{old} + \alpha\mathbf{V} + \alpha(1-\alpha)(\mathbf{m}_{old} - \mathbf{m})(\mathbf{m}_{old} - \mathbf{m})^T \quad (3)$$

where α is the learning rate whose value is between 0 and 1. The derivations of the above equations are shown in equation (7) and (8) in [15].

In each time step, an incremental subspace update is performed to obtain the most discriminative likelihood image. This method is more efficient than the batch method since we do not need to store the data from the previous frames; instead, we only need the mean and covariance matrix.

4. TRACKING BY PARTICLE FILTERING

In this section we will show how to incorporate feature extraction technique for robust observations into the particle filter framework.

The particle filter [8] is a stochastic framework to propagate the conditional density to the next step. The state variable $\mathbf{x}_t$ ($t = 0 \dots n$) is characterized by its probability density function estimated from the sequence of measurements $\mathbf{z}_t$ ($t = 0 \dots n$). The density function is represented with a set of samples and their weights which enable us to describe an arbitrary probability density function effectively.

In our experiments, the state variable is a 3-dimensional vector (x, y, s) where (x, y) is 2D location of an object and s is a scale parameter, and the target is represented with a rectangular region. A random walk is assumed for the process model since it is not desirable to assign any specific motion model before observation. Since we employ SIR filter, the weights of particles are equal until the prediction step; then they are updated twice by the feasibility and the likelihood in observation.

4.1. Feasibility for Particle

Feasibility is meant to capture how the region represented by a sample is salient with respect to the background, and it is computed by the summation of values inside the region in the most discriminative likelihood image.

Formally, suppose that the value at (x, y) in the most discriminative likelihood image is $MD(x, y)$. Since we use rectangular regions for the observation, the feasibility w_f is

$$w_f(x_i, y_i, s_i) = \sum_{x_i \le x \le x_i + w_i, y_i \le y \le y_i + h_i} MD(x, y) \qquad (4)$$

where $x_i \le x \le x_i + w_i, y_i \le y \le y_i + h_i$ is the area associated with the i-th particle. The integral image (II) proposed in [12] is defined to be

$$II(x, y) = \sum_{x' \le x, y' \le y} MD(x', y'), \qquad (5)$$

so that the feasibility can be computed by only 4 table look-up operations using the integral image. After computing the feasibility, the sample weight is updated with this value.

This strategy is reasonable since the regions containing many high likelihood-ratio pixels are selected target candidates based on multiple visual cues and the observation process described below can compensate for the disadvantage of likelihood images – poor performance in clutter.

4.2. Observation

Color-based tracking is employed in our experiments. The likelihood of each step is based on the similarity of the RGB histogram between the target and the candidates. The histogram of the target is denoted by $c^\star(i)$ ($i = 1 \dots N$), where N is the number of bins in the histogram and $\sum_{i=1}^{N} c^\star(i) = 1$. The Bhattacharyya distance in equation (6) is used to measure the similarity between two histograms

$$D[c^\star, c(\mathbf{x}_t)] = \left(1 - \sum_{i=1}^{N} \sqrt{c^\star(i) c(\mathbf{x}_t; i)}\right)^{1/2} \qquad (6)$$

and the final measurement function including feasibility at time t is given by

$$p(\mathbf{z}_t | \mathbf{x}_t) \propto w_f(\mathbf{x}_t) \exp\left(-\lambda D^2[c^\star, c(\mathbf{x}_t)]\right) \qquad (7)$$

where λ is a constant.

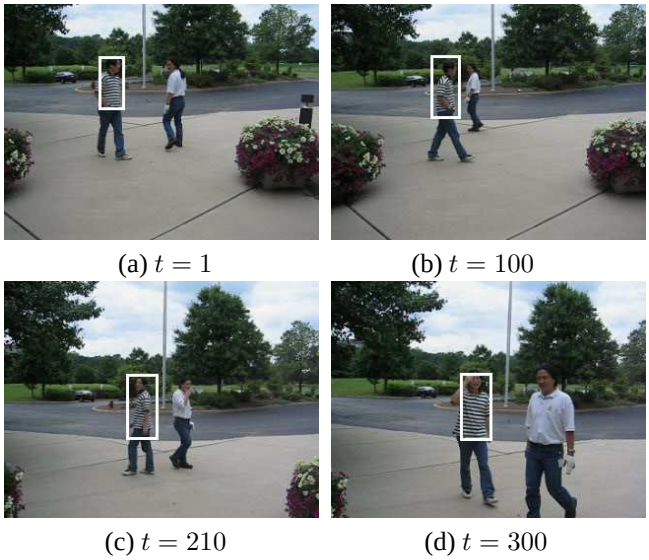
4.3. Results

Two different video sequences were used to test the performance of our tracker – *people* and *vehicle* sequence. In the *people* sequence, the target is not so distinctive in likelihood images due to clutter, so tracking only with likelihood images is not successful. However, the combination of feasibility and likelihood in observation can track a person for the whole sequence with 100 particles. The tracking results are shown in figure 3.

A car is moving in the severe fog in the second video, which is downloaded from Universität Kahlsruhe homepage (`http://i21www.ira.uka.de/image_sequences`). Even with white pixels due to the fog in the background, the white car in the foreground is identified clearly in the most discriminative feature space as seen in figure 2, and tracking was also successful with 100 particles.

5. DISCUSSION

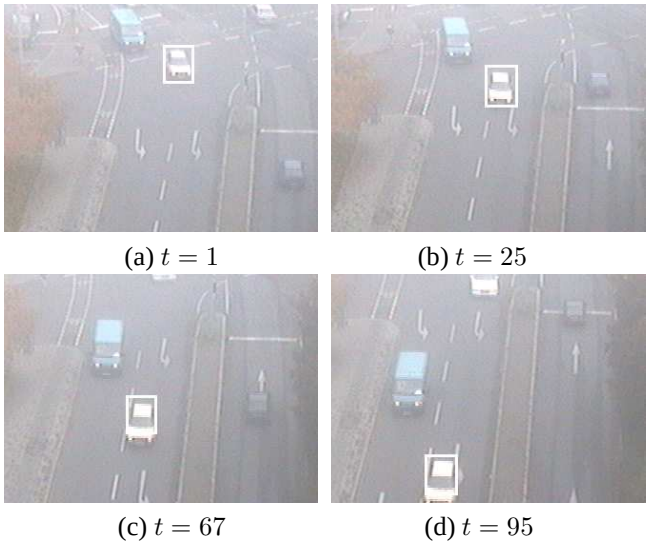
We described a method to improve the robustness of observations for object tracking using the most discriminative likelihood image. This likelihood image is obtained from the combination of multiple independent features and updated incrementally. This technique is incorporated into a particle filter, and tracking is performed based on the original image as well as the combined likelihood image.



(a) $t = 1$ (b) $t = 100$

(c) $t = 210$ (d) $t = 300$

Fig. 3. Tracking results with *people* sequence



(a) $t = 1$ (b) $t = 25$

(c) $t = 67$ (d) $t = 95$

Fig. 4. Tracking results with *vehicle* sequence

In particle filter tracking , the quality of sampling is critical to its overall performance. So, we can achieve better results with a small number of samples if the particles with low feasibility are rejected and new samples are used. Currently, only color information is used, and other visual features should be tested in our framework.

6. REFERENCES

[1] I. Matthews, T. Ishikawa and S. Baker, "The Template Update Problem," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 4, pp. 810–815, 2004

[2] D. Comaniciu, V. Ramesh and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head, SC, vol. 2, pp. 142–149, June, 2000

[3] P. Perez, C. Hue, J. Vermaak and M. Cangnet, "Color-Based Probabilistic Tracking," *Proc. European Conference on Computer Vision*, Copenhagen, Denmark, vol. 1, pp. 661–675, 2002

[4] K. Nummiaro, E. Koller-Meier and L. Van Gool, "An Adaptive Color-Based Particle Filter," *Image and Vision Computing*, vol. 21, no. 1, pp. 99–110, 2003

[5] S.J. McKenna, Y. Raja and S. Gong, "Tracking Colour Objects Using Adaptive Mixture Models," *Image and Vision Computing Journal*, vol. 17, pp. 223-229, 1999

[6] H.T. Nguyen and A. Smeulders, "Tracking Aspects of the Foreground against the Background," *European Conf. on Computer Vision* Prague, Czech Republic, 2004

[7] H.T. Chen, T.L. Liu and C.S. Fuh, "Probabilistic Tracking with Adaptive Feature Selection," *Proc. of the 17th Intl. Conf. on Pattern Recognition* Cambridge, UK, vol. 2 736-739, Aug, 2004,

[8] M. Isard and A. Blake "Condensation - Conditional Density Propagation for Visual Tracking", *Intl. Journal of Computer Vision*, vol. 29, no. 1, 1998

[9] A.D. Jepson, D.J. Fleet and T.F. El-Maraghi, "Robust Online Appearance Models for Visual Tracking," *Proc. 8th Intl. Conf. on Computer Vision*, Vancouver, Canada, vol. 1, pp. 415–422, 2001

[10] R. Collins and Y. Liu, "On-Line Selection of Discriminative Tracking Features," *Proc. 9th Intl. Conf. on Computer Vision*, Nice, France, Oct., 2003

[11] J. Weng, Y. Zhang and W.S. Hwang, "Candid Covariance-Free Incremental Principal Component Analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 8, pp. 1034–1040, 2003

[12] P. Viola and M. Jones, "Rapid Object Detection Using A Boosted Cascade of Simple Features," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001

[13] S. Mika, G. Rätsch, J. Western, B. Schölkoph, and K.R. Müller, "Fisher Discriminant Analysis with Kernels," *Neural Networks for Signal Processing IX*, pp. 41–48, 1999

[14] B. Schölokph, A. Smola and K.R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problems," *Neural Computation*, vol 10, pp. 1299–1319, 1998

[15] P. Hall, D. Marshall and R. Martin, "Merging and Splitting Eigenspace Models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 9, pp. 1042–1048, 2000