On-Line Kernel-Based Tracking in Joint Feature-Spatial Spaces

Changjiang Yang, Ramani Duraiswami, Ahmed Elgammal[†] and Larry Davis
Perceptual Interfaces and Reality Laboratory, UMIACS, University of Maryland, College Park, MD
[†] Dept. of Computer Science, Rutgers University, Piscataway, NJ
{yangcj,ramani,lsd}@umiacs.umd.edu [†]elgammal@cs.rutgers.edu

Abstract

We will demonstrate an object tracking algorithm that uses a novel simple symmetric similarity function between spatially-smoothed kernel-density estimates of the model and target distributions. The similarity measure is based on the expectation of the density estimates over the model or target images. The density is estimated using radial-basis kernel functions which measure the affinity between points and provide a better outlier rejection property. The mean-shift algorithm is used to track objects by iteratively maximizing this similarity function. To alleviate the quadratic complexity of the density estimation, we employ Gaussian kernels and the fast Gauss transform to reduce the computations to linear order. This leads to a very efficient and robust nonparametric tracking algorithm. More details can be found in [2]. The system processes online video stream on a P4 1.4GHz and achieves 30 frames per second using an ordinary webcam.

1 Similarity Between Distributions

This demonstration presents a real-time object tracking system running on a PC with an ordinary webcam. The tracking of objects in a video stream is a common task, in which a model image is translated, rotated, and (possibly) deformed to match the given target images. It is important for many computer vision applications such as human-computer interaction, surveillance, smart rooms and medical imaging.

Our approach is based on the matching in the feature spaces (we used RGB color space) which are described by the probability density functions (*pdf*). The *pdf* is estimated in the feature spaces using kernel density estimation (see Figure 1):

$$\hat{p}_x(\mathbf{u}) = \frac{1}{N}\sum_{i=1}^{N} k(\|\frac{\mathbf{u}-\mathbf{u}_i}{h}\|^2). \tag{1}$$

where $k(x)$ is a RBF kernel, usually a Gaussian kernel is adopted.

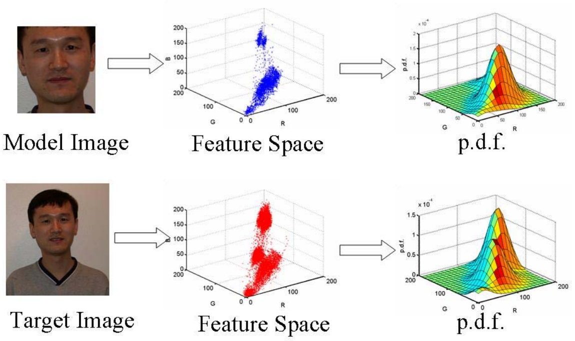

Figure 1: The probabilistic descriptions of the model image and target image.

A similarity between the two *pdfs* is proposed to measure the affinity between two distributions. Given two sets of sample points, the similarity is the expectation of affinities between all pairs of model and target samples:

$$J(I_x, I_y) = \tag{2}$$
$$\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M} w(\|\frac{\mathbf{x}_* - \mathbf{x}_i}{\sigma}\|^2) k(\|\frac{\mathbf{u}_i - \mathbf{v}_j}{h}\|^2) w(\|\frac{\mathbf{y}-\mathbf{y}_j}{\sigma}\|^2).$$

The similarity function (2) is non-metric. However, it can be shown that its negative natural logarithm

$$L(I_x, I_y) = -\log J(I_x, I_y) \tag{3}$$

is a probabilistic distance, provided we have sufficient samples, so that the kernel density estimate converges to the true probability density function [2].

2 Mean-Shift Based Target Localization

Once we have the similarity measure between the model image and target image, we can find the target location in the target image by minimizing the distance (3) with respect to the variable $\mathbf{y}$ using the mean-shift algorithm [1] which has already proved successful in many computer vision applications. The *mean shift* of the smoothed similarity function $\mathbf{m}(\mathbf{y})$ is

$$\nabla L(\mathbf{y}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} \mathbf{y}_j w_i k_{ij} g(\|\frac{\mathbf{y}-\mathbf{y}_j}{\sigma}\|^2)}{\sum_{i=1}^{N}\sum_{j=1}^{M} w_i k_{ij} g(\|\frac{\mathbf{y}-\mathbf{y}_j}{\sigma}\|^2)} - \mathbf{y}, \tag{4}$$

where $w_i = w(\|\frac{\mathbf{x}_*-\mathbf{x}_i}{\sigma}\|^2)$ and $k_{ij} = k(\|\frac{\mathbf{u}_i-\mathbf{v}_j}{h}\|^2)$. $g(x) = -w'(x)$ is also the profile of a RBF kernel.

0-7695-2158-4/04 $20.00 (C) 2004 IEEE

Given the sample points $\{\mathbf{x}_i, \mathbf{u}_i\}_{i=1}^N$ centered at $\mathbf{x}_*$ in the model image, and $\{\mathbf{y}_j, \mathbf{v}_j\}_{j=1}^M$ centered at the current position $\hat{\mathbf{y}}_0$ in the current target image, the object tracking based on the mean-shift algorithm is an iterative procedure which recursively moves the current position $\hat{\mathbf{y}}_0$ to the new position $\hat{\mathbf{y}}_1$ until reaching the density mode according to

$$\hat{\mathbf{y}}_1 = \frac{\sum_{i=1}^N \sum_{j=1}^M \mathbf{y}_j w_i k_{ij} g(\left\|\frac{\hat{\mathbf{y}}_0 - \mathbf{y}_j}{\sigma}\right\|^2)}{\sum_{i=1}^N \sum_{j=1}^M w_i k_{ij} g(\left\|\frac{\hat{\mathbf{y}}_0 - \mathbf{y}_j}{\sigma}\right\|^2)}. \qquad (5)$$

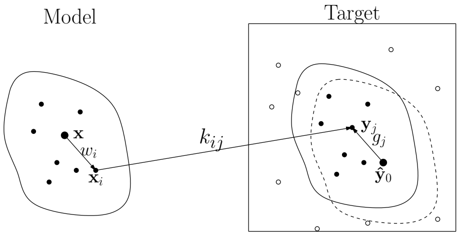

Figure 2: The mean-shift based tracking procedure.

3 Speedup using the Improved FGT

The computational complexity per frame in the above algorithm is $O(PMN)$, where P is the average number of iterations per frame, M and N are the number of sample points in target image and model image respectively. Typically the average number of iterations per frame P is less than ten and $M \approx N$. Then the order of the computational complexity is quadratic. To achieve real-time performance, we rewrite Eq.(5) as

$$\hat{\mathbf{y}}_1 = \frac{\sum_{j=1}^M \mathbf{y}_j f(\mathbf{y}_j)}{\sum_{j=1}^M f(\mathbf{y}_j)}, \qquad (6)$$

where

$$f(\mathbf{y}_j) = \sum_{i=1}^N e^{-\|\mathbf{x}_i - \mathbf{x}_*\|^2/\sigma^2} e^{-\|\mathbf{u}_i - \mathbf{v}_j\|^2/h^2} e^{-\|\mathbf{y}_j - \hat{\mathbf{y}}_0\|^2/\sigma^2} \qquad (7)$$

is a *discrete Gauss transform* of $\mathbf{y}_j$ for $j = 1, \ldots, M$. The computational complexity of (7) can be reduced using the improved fast Gauss transform to linear order even in higher dimensions [3].

References

[1] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603 – 619, May 2002.

[2] C. Yang, R. Duraiswami, A. Elgammal, and L. Davis. Real-time kernel-based tracking in joint feature-spatial spaces. Technical Report CS-TR-4567, UMIACS, College Park, 2003.

[3] C. Yang, R. Duraiswami, N. Gumerov, and L. Davis. Improved fast Gauss transform and efficient kernel density estimation. In *Proc. Int'l Conf. Computer Vision*, pages 464–471, Nice, France, 2003.

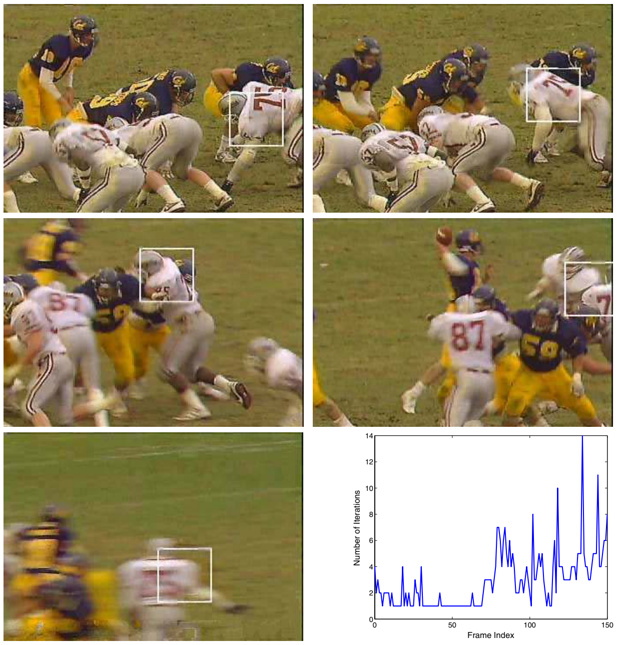

Figure 3: Tracking results of the *Football* sequence. Frames 30, 75, 105, 140, 150, and number of mean-shift iterations *w.r.t.* the frame index are displayed.

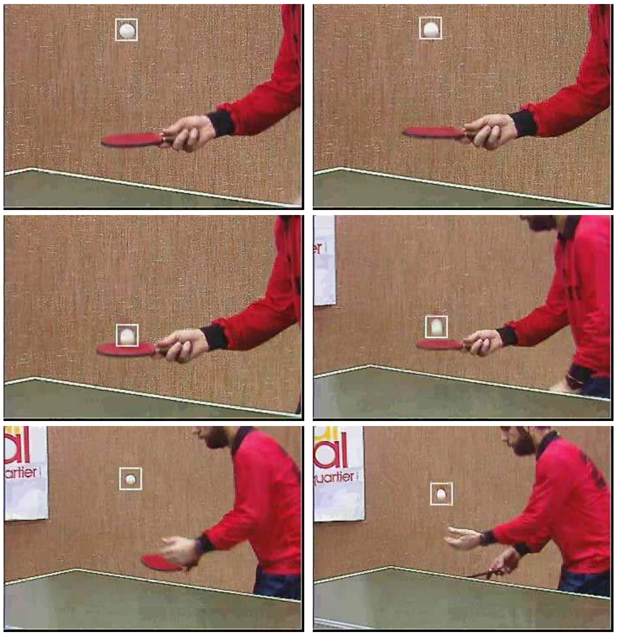

Figure 4: Tracking results of the *Ball* sequence. Frames 3, 16, 26, 40, 48 and 51 are displayed.