Spatial Histograms for Region-Based Tracking

Stanley T. Birchfield and Sriram Rangarajan

ABSTRACT—Spatiograms are histograms augmented with spatial means and covariances to capture a richer description of the target. We present a particle filtering framework for region-based tracking using spatiograms. Unlike mean shift, the framework allows for non-differentiable similarity measures to compare two spatiograms; we present one such similarity measure, a combination of a recent weighting scheme and histogram intersection. Experimental results show improved performance with the new measure as well as the importance of global spatial information for tracking. The performance of spatiograms is compared with color histograms and several texture histogram methods.

Keywords—Histogram, spatiogram, visual tracking.

I. Introduction

Histograms have proved to be a powerful representation for the image data in a region. Discarding all spatial information, they have been used in several successful tracking systems which take advantage of their robustness to changing object pose and shape [1], [2]. However, the loss of all spatial information reduces specificity in the model, thus increasing the susceptibility of the tracker to distraction by the background or by other objects. To overcome this limitation, we have recently proposed the concept of a *spatiogram* [3], which augments the histogram bins with the spatial means and covariances of the pixels. Spatiograms have found application in head tracking [3], [4], fusing multi-spectral observations [5], and image and video retrieval [6], [7].

Existing spatiogram trackers utilize the mean shift algorithm [3], [8]. As a hill-climbing approach, mean shift requires a differentiable similarity measure and is susceptible to local

minima. In this work, we present a particle filtering framework for spatiogram tracking to overcome these limitations. Within this framework, we describe a new non-differentiable spatiogram similarity measure that is more robust to noise than the commonly-used Bhattacharyya coefficient [2], [8]. We also compare spatiograms with color histograms and several variations of local texture histograms, showing the importance of global spatial relationships for tracking.

II. Histograms and Spatiograms

Let $I : \mathbf{x} \rightarrow \mathbf{v}$ be an image which maps pixels $\mathbf{x} = [x \quad y]^T \in X$ to values $\mathbf{v} \in V$. The histogram of I is a vector of values given by $h_I(b) = n_b, b = 1, ..., B$, where n_b is the number of pixels whose values are equal to that of the b-th bin, and B is the total number of bins. A histogram captures the number of occurrences of each element in the range of I; that is, $h_I(b)$ is the number of elements $\mathbf{x} \in X$ such that $I(\mathbf{x}) = \mathbf{v}$, where b corresponds to $\mathbf{v}$.

We define the second-order *spatiogram* of I to be a triple of values: $h_I(b) = \langle n_b, \mu_b, \Sigma_b \rangle, b = 1, ..., B$, where μ_b and Σ_b are respectively the mean vector and covariance matrices of the spatial coordinates of the pixels contributing to the b-th bin [3]. Unlike histograms, spatiograms retain some spatial information to enhance specificity.

III. Particle Filtering Framework

The particle filter is a probability propagation model which applies to arbitrary (not necessarily Gaussian) probability distributions [9]. Let $\theta = (x \quad \dot{x})$ be the state of the system, where x is the position and scale of the target. Assuming that the process is first-order Markov $p(\theta_t \mid \theta_{1:t-1}) = p(\theta_t \mid \theta_{t-1})$ and that the measurements z are independent of the process and of each other, $p(z_{1:t} \mid \theta_{1:t}) = \prod_{i=1}^{t} p(z_i \mid \theta_i)$, the problem of

tracking can be formulated as maximizing the *a posterior* probability $p(\theta_t \mid z_{1:t})$, which is proportional to

$$\Phi(z_t \mid \theta_t) \int p(\theta_t \mid \theta_{t-1}) p(\theta_{t-1} \mid z_{1:t-1}) d\theta_{t-1}. \qquad (1)$$

Given the state transition model $p(\theta_t \mid \theta_{t-1})$, the previous posterior distribution $p(\theta_{t-1} \mid z_{1:t-1})$, and the observation likelihood $\Phi(z_t \mid \theta_t)$, the particle filter approximates the current posterior distribution by a set of weighted particles $\{\theta_t^{(j)}, \pi_t^{(j)}\}$, which are updated recursively. The basic steps of sampling, measuring the likelihood, and resampling are described in [9].

IV. Comparing Spatiograms

The observation likelihood Φ computes the similarity between two spatiograms as the weighted sum of the similarity between the two histograms:

$$\rho(h, h') = \sum_{b=1}^{B} \psi_b \rho_n(n_b, n_b'). \qquad (2)$$

In the original spatiogram approach [3], the weights are

$$\psi_b = N\left(\mu_b; \mu_b', \Sigma_b'\right) N\left(\mu_b'; \mu_b, \Sigma_b\right), \qquad (3)$$

where $N(a; \mu, \Sigma)$ is a (μ, Σ) Gaussian evaluated at a. However, it was noted in [8] that this measure is overly sensitive to small spatial changes, thus motivating an improved measure that overcomes this problem:

$$\psi_b = \eta N\left(\mu_b; \mu_b', 2(\Sigma_b + \Sigma_b')\right), \qquad (4)$$

where the factor η ensures that $0 \le \rho \le 1$ and $\rho(h, h) = 1$ for any h [8].

Since mean shift tracking [2] requires the similarity measure to be differentiable, the Bhattacharyya coefficient $\rho_n(n_b, n_b') = \sqrt{n_b n_b'}$ is often used to compare the two histograms [2], [8]. However, particle filtering allows us to relax this constraint and instead use non-differentiable similarity measures. As a result, we combine the improved weights of (4) with the non-linear technique of histogram intersection [1] $\rho_n(n_b, n_b') = \min(n_b, n_b') / \sum_{j=1}^{B} n_j$ to yield a new similarity measure which is less sensitive to outliers.

V. Experimental Results

To determine the importance of global versus local spatial relationships, we implemented trackers based on spatiograms (CS), color histograms (CH), color co-occurrence matrices (CCM), log-Gabor histograms (LGH), Haar wavelet

Fig. 1. Tracking results for four image sequences showing the best (red ellipse with knots) and worst (blue ellipse) algorithms for each sequence. From top to bottom, the worst are LGH, EOH, EOH, and HH; in every sequence the best algorithm is CS (color spatiograms).

histograms (HH), edge orientation histograms (EOH), and foreground/background histograms (FBH) consisting of a pair of concentric histograms. CH, SH, and CCM utilized a red-green, green-blue, intensity color space [1] with 8, 8, and 4 bins, respectively. A separate co-occurrence matrix was computed for each of the three color channels, in the horizontal direction. The LGH algorithm utilized 4 scales and 8 orientations, HH used 3 scales with 3 orientations (horizontal, vertical, and diagonal), and EOH was computed with 3 levels and 8 orientations. The FBH algorithm involved two color histograms, one for the object itself and one surrounding the object with a radius 50% larger.

The algorithms were tested on four image sequences, given in Fig. 1, which show a person's head ("Stan"), another head ("Mike"), a football player ("Football"), and a person walking ("Sriram"), respectively. The images were labeled by hand to determine ground truth in *x* and *y*. For the head sequences, all algorithms were augmented with an intensity gradient module to increase tracking reliability [1].

The mean error for each sequence and each algorithm is shown in Table 1. In all four sequences, spatiograms exhibited the fewest errors, with color histograms showing the fewest errors on "Sriram." An error plot versus time for two of the algorithms (CS and CH) for "Stan" is shown in Fig. 2. Overall, there are fewer errors with CS except for frame 300, which is shown in the top-left of Fig. 1.

Table 1. Mean error in pixels for the algorithms over four sequences, with bold indicating the smallest errors.

Algorithm	Stan	Mike	Football	Sriram
CS	**4.8**	**6.3**	**1.4**	**1.1**
CH	14.6	12.7	1.6	**1.1**
CCM	5.0	10.2	3.7	1.8
LGH	15.9	12.1	6.6	3.3
HH	15.4	9.5	7.0	3.7
EOH	12.0	28.9	7.9	1.9
FBH	9.3	13.3	2.7	2.9
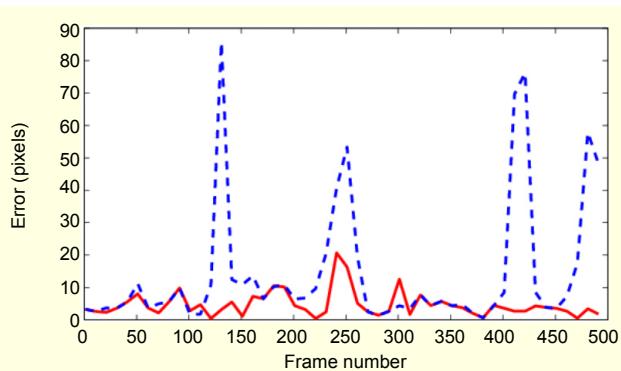
Fig. 2. Error in pixels versus time for CS (solid red) and CH (dashed blue) for "Stan." CS shows fewer errors.
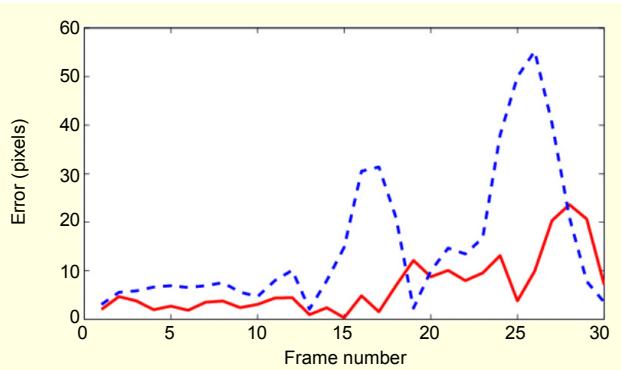
Fig. 3. Error in pixels for the proposed similarity measure (solid red) and the measure in [8] (dashed blue) for "Mike."

Figure 3 displays the performance of the proposed similarity measure compared with the measure described in [8] for the "Mike" sequence. Overall, the proposed measure exhibits fewer errors. In particular, the algorithm handles the fast motion around frame 17 and is not distracted by the other person in the scene in frame 26. Occasionally, however, the algorithm performs slightly worse, as in frames 19, 29, and 30.

VI. Conclusion

We have presented a particle filtering framework for region-based tracking using spatiograms. The framework overcomes two limitations of the popular mean shift approach. It is not susceptible to local minima, and it does not require a differentiable similarity measure. By combining histogram intersection with the recent weighting scheme introduced in [8], we developed a novel, non-linear similarity measure for spatiograms which is more robust to outliers. Experiments demonstrate the improved performance resulting from the new measure. Comparison with color histograms and local texture histogram techniques confirms the importance of global spatial information for tracking objects whose spatial color configuration remains relatively constant in different poses.

References

[1] S. Birchfield, "Elliptical Head Tracking Using Intensity Gradients and Color Histograms," *Proc. CVPR*, 1998, pp. 232-237.

[2] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, May 2003, pp. 564-577.

[3] S.T. Birchfield and S. Rangarajan, "Spatiograms Versus Histograms for Region-Based Tracking," *Proc. CVPR*, June 2005, pp. 1158-1163.

[4] Y. Li, H. Ai, C. Huang, and S. Lao, "Robust Head Tracking with Particles Based on Multiple Cues," *Proc. ECCV Workshop on HCI*, 2006.

[5] C. O'Conaire, N.E. O'Connor, and A. Smeaton, "Thermo-Visual Feature Fusion for Object Tracking Using Multiple Spatiogram Trackers," *Mach. Vis. App.*, 2007.

[6] C. O'Conaire, N.E. O'Connor, A. Smeaton, and G.J.F. Jones, "Organising a Daily Visual Diary Using Multi-Feature Clustering," *19th Annual Symp. on Elec. Imaging*, 2007.

[7] A. Ulges, C.H. Lampert, and D. Keysers, "Spatiogram-Based Shot Distances for Video Retrieval," *Proc. of TRECVID Workshop*, 2006.

[8] C. O'Conaire, N.E. O'Connor, A.F. Smeaton, "An Improved Spatiogram Similarity Measure for Robust Object Localization," *Proc. ICASSP*, 2007.

[9] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Non-linear/Non-Gaussian Bayesian Tracking", *IEEE Trans. on Signal Processing*, vol. 50, no. 2, 2002, pp. 174-188.