

# AN IMPROVED SPATIOGRAM SIMILARITY MEASURE FOR ROBUST OBJECT LOCALISATION

C. Ó Conaire, N. E. O'Connor, A. F. Smeaton

Adaptive Information Cluster, Dublin City University, Ireland

## ABSTRACT

Spatiograms were introduced as a generalisation of the commonly used histogram, providing the flexibility of adding spatial context information to the feature distribution information of a histogram. The originally proposed spatiogram comparison measure has significant disadvantages that we detail here. We propose an improved measure based on deriving the Bhattacharyya coefficient for an infinite number of spatial-feature bins. Its advantages over the previous measure and over histogram-based matching are demonstrated in object tracking scenarios.

**Index Terms**— Image matching, Image pattern recognition, Tracking

## 1. INTRODUCTION

Histograms have been widely used as descriptors for image regions. All spatial information is discarded and only feature occurrence counts are retained. Spatial information may be important for some applications, such as object tracking, and a descriptor that retains some spatial information is desirable. Spatiograms [1] are one such descriptor. As well as object tracking, they have also been shown to be useful in image clustering and data fusion. In this paper, we derive a new spatiogram similarity measure that is superior to the previously proposed measure, and provide theoretical and practical examples to justify its use.

This paper is organised as follows: In section 2, we review related work on spatiograms and other feature-spatial distribution models; section 3 contains a brief overview of spatiograms and the drawbacks of using the original similarity measure; we derive our new measure in section 4; in section 5, we demonstrate experimentally the advantages of our new measure over existing measures and give our conclusions in section 6.

---

This material is based on works supported by Science Foundation Ireland under Grant No. 03/IN.3/I361 and sponsored by a scholarship from the Irish Research Council for Science, Engineering and Technology (IRCSET): Funded by the National Development Plan. The authors would also like to express their gratitude to Mitsubishi Electric Research Labs (MERL) for their contribution to this work.

## 2. RELATED RESEARCH

Many descriptors have been proposed for object modelling in order to provide robust tracking in video sequences. Histograms [2, 3, 4] have commonly been used for tracking, as they discard spatial information and are therefore insensitive to small object pose changes. At the other extreme, image templates [5], which impose rigid spatial constraints on feature layouts, have also been used. In [6], Elgammal et al. provide a parametric feature-spatial distribution model for object modelling. A spatial kernel bandwidth parameter must be supplied, to tradeoff between rigid and no spatial constraints. Spatiograms [1] generalise the histogram to allow higher-order spatial moments to be part of the descriptor. Spatiograms have been shown to be useful in head tracking [1], in fusing features of different modalities for multi-spectral object tracking [7] and in image clustering for retrieval [8].

## 3. REVIEW OF SPATIOGRAMS

### 3.1. Spatiogram models

As well as histograms of feature occurrences, spatiograms also capture higher-order spatial moments. A  $2^{nd}$ -order spatiogram model of an object is identical to a histogram of its features, except that it also stores additional spatial information, namely the mean and covariance of the spatial position of all pixels that fall into each histogram bin.

To compute a normalised histogram for a target image of  $N$  pixels, we write the histogram bin count,  $n_b$ , of bin  $b$  as:

$$n_b = C \sum_{i=1}^N \delta_{ib} \quad (1)$$

where  $\delta_{ib} = 1$  if the  $i^{th}$  pixel falls in the  $b^{th}$  bin and  $\delta_{ib} = 0$  otherwise. The normalising constant,  $C$ , ensures that all bins sum to one (i.e.  $\sum_{i=1}^B n_i = 1$ ). For  $2^{nd}$ -order spatiograms, the spatial mean,  $\mu_b$ , and covariance,  $\Sigma_b$ , also need to be computed for each bin:

$$\mu_b = \frac{1}{\sum_{j=1}^N \delta_{jb}} \sum_{i=1}^N x_i \delta_{ib} \quad (2)$$

$$\Sigma_b = \frac{1}{\sum_{j=1}^N \delta_{jb}} \sum_{i=1}^N (x_i - \mu_b)(x_i - \mu_b)^T \delta_{ib} \quad (3)$$

where  $x_i = [x, y]^T$  is the spatial position of pixel  $i$ . To compare regions of different sizes, it is necessary to map the spatial coordinates to the same scale; for example, by mapping coordinates into the  $[-1, +1]$  range. Additionally, in order to ensure that each  $\Sigma_b$  is invertible, we force them to be diagonal, and set a minimum variance value to one pixel.

### 3.2. Comparing Spatiograms

To compare two spatiograms,  $S = \{n, \mu, \Sigma\}$  and  $S' = \{n', \mu', \Sigma'\}$ , each with  $B$  bins, the following similarity measure is used in [1]:

$$\rho(S, S') = \sum_{b=1}^B \psi_b \sqrt{n_b n'_b} \quad (4)$$

where  $\psi_b$  is the spatial similarity measure, given by:

$$\psi_b = \eta \exp \left\{ -\frac{1}{2} (\mu_b - \mu'_b)^T \hat{\Sigma}_b^{-1} (\mu_b - \mu'_b) \right\} \quad (5)$$

where  $\eta$  is the Gaussian normalisation term, and  $\hat{\Sigma}_b^{-1} = (\Sigma_b^{-1} + (\Sigma'_b)^{-1})$ , so that the distance between the spatial means is normalised to the average of the two Mahalanobis distances. We will term this similarity measure the *original* measure.

### 3.3. Disadvantages of the original measure

The original similarity measure has two main disadvantages. Firstly, it is not tolerant of small spatial changes of the features. This is clear from the  $\hat{\Sigma}_b$  term, whose value is less than either of the individual variances. This means that small spatial changes are heavily punished by the original measure. Secondly, a good similarity measure should have the property that if an image region spatiogram is compared to itself, the measure should return its maximal value. This is not true of the original measure. With  $N(x; \mu, \Sigma)$  representing a normalised Gaussian evaluated at  $x$ , we write the similarity between a spatiogram and itself:

$$\begin{aligned} \rho(S, S) &= \sum_{b=1}^B \sqrt{n_b^2} [N(\mu_b; \mu_b, (2\Sigma_b^{-1})^{-1})] \\ &= \sum_{b=1}^B \frac{n_b}{2\pi|(1/2)\Sigma_b|^{1/2}} = \sum_{b=1}^B \frac{n_b}{\pi|\Sigma_b|^{1/2}} \end{aligned}$$

This shows that comparing a spatiogram to itself does not equal a constant. Indeed, with the original similarity measure, it is possible for a patch which is **different** from the target patch to be a better match to the target, than the target itself! This is because the normalisation constant of the original similarity measure adds more weight to spatially-tighter feature

clusters. This leads to a non-smooth similarity surface, which often has many spiked peaks. We overcome this difficulty by deriving a new similarity measure in the next section.

## 4. PROPOSED MEASURE

To derive our new similarity measure, we leverage the fact that we can convert the  $2^{nd}$  order spatiogram back to a histogram, adding an extra dimension of space. For bin  $b$ , we divide its contents,  $n_b$ , over an infinite number of spatial bins,  $n_{b,k}$ , where  $k$  is an integer ranging from  $-\infty$  to  $+\infty$ . We express this as:

$$n_{b,k} = \frac{n_b \phi_b(k\Delta\mathbf{w})\Delta\mathbf{w}}{\sum_{i=-\infty}^{+\infty} \phi_b(i\Delta\mathbf{w})\Delta\mathbf{w}} \quad (6)$$

where  $\Delta\mathbf{w}$  is the spatial size of each bin and  $\phi_b$  is a normalised Gaussian with the mean and covariance of bin  $b$ . Since we can now (theoretically) create a histogram from any spatiogram, we can now compare spatiograms using the Bhattacharyya coefficient [2]. This has a relationship with the probability of Bayes error [9], and therefore is more similar to a probability than equation (4). Given two spatiograms (converted to histograms),  $n_{b,k}$  and  $n'_{b,k}$ , we compare them using the Bhattacharyya coefficient as follows:

$$\begin{aligned} \rho(n, n') &= \sum_{b=1}^B \sum_{k=-\infty}^{+\infty} \sqrt{n_{b,k} n'_{b,k}} \\ &= \sum_{b=1}^B \sum_{k=-\infty}^{+\infty} \sqrt{\left( \frac{n_b \phi_b(k\Delta\mathbf{w})\Delta\mathbf{w}}{\sum_{i=-\infty}^{+\infty} \phi_b(i\Delta\mathbf{w})\Delta\mathbf{w}} \right)} \\ &\quad \sqrt{\left( \frac{n'_b \phi'_b(k\Delta\mathbf{w})\Delta\mathbf{w}}{\sum_{i=-\infty}^{+\infty} \phi'_b(i\Delta\mathbf{w})\Delta\mathbf{w}} \right)} \end{aligned}$$

As  $\Delta\mathbf{w} \rightarrow 0$ , the denominators of both fractions disappear, since  $\phi_b$  and  $\phi'_b$  are both normalised Gaussians, therefore:

$$\sum_{i=-\infty}^{+\infty} \phi_b(i\Delta\mathbf{w})\Delta\mathbf{w} \approx \int_{-\infty}^{+\infty} \phi_b(x) dx = 1$$

This gives:

$$\rho(n, n') = \sum_{b=1}^B \sqrt{n_b n'_b} \int_{-\infty}^{+\infty} \sqrt{\phi_b(x) \phi'_b(x)} dx \quad (7)$$

This can be simplified further by noting that the product of two Gaussians is Gaussian, and also that the square-root of a Gaussian is Gaussian. The resulting Gaussians are not necessarily normalised, however, and therefore do not usually integrate to one. We are left with constant terms which can be thought of as weights for each bin comparison. Given that  $\sqrt{N(x; a, A)} = qN(x; a, 2A)$ , with  $q = 2(2\pi)^{m/4}|A|^{1/4}$

for  $m$  dimensions, and  $N(x; a, A)N(x; b, B) = zN(x; c, C)$  with  $z = N(a; b, A + B)$ , we can simplify equation (7) to produce our compact new measure:

$$\begin{aligned}
\rho(n, n') &= \sum_{b=1}^B \sqrt{n_b n'_b} \int_{-\infty}^{+\infty} \sqrt{z_b \hat{\phi}_b(x)} dx \\
&= \sum_{b=1}^B \sqrt{n_b n'_b} \sqrt{z_b} \int_{-\infty}^{+\infty} \sqrt{\hat{\phi}_b(x)} dx \\
&= \sum_{b=1}^B \sqrt{n_b n'_b} \sqrt{z_b} \int_{-\infty}^{+\infty} q_b \hat{\phi}_b(x) dx \\
&= \sum_{b=1}^B \sqrt{n_b n'_b} [q_b \sqrt{z_b}] \\
&= \sum_{b=1}^B \sqrt{n_b n'_b} \left[ q_b \sqrt{N(\mu_b; \mu'_b, \Sigma_b + \Sigma'_b)} \right] \\
&= \sum_{b=1}^B \sqrt{n_b n'_b} \left[ q_b Q_b N(\mu_b; \mu'_b, 2(\Sigma_b + \Sigma'_b)) \right]
\end{aligned}$$

where  $\hat{\phi}_b$  and  $\hat{\phi}'_b$  are normalised Gaussians. Terms  $q_b$  and  $Q_b$  are given by

$$\begin{aligned}
q_b &= 2\sqrt{2\pi} \left| \Sigma_b + \Sigma'_b \right|^{1/4} \\
Q_b &= 2\sqrt{2\pi} \left| (\Sigma_b^{-1} + (\Sigma'_b)^{-1})^{-1} \right|^{1/4}
\end{aligned}$$

Noting that  $q_b Q_b = 8\pi |\Sigma_b \Sigma'_b|^{1/4}$ , our final similarity measure becomes:

$$\rho = \sum_{b=1}^B \sqrt{n_b n'_b} \left[ 8\pi |\Sigma_b \Sigma'_b|^{1/4} N(\mu_b; \mu'_b, 2(\Sigma_b + \Sigma'_b)) \right] \quad (8)$$

#### 4.1. Analysis of the new measure

Firstly, comparing equations (4) and (8), of the old and new measures respectively, it is clear that the new measure is more tolerant of small spatial changes, as the covariance is equal to twice the sum of individual covariances. Secondly, If we compare two identical spatiograms using the new measure, we obtain:

$$\begin{aligned}
\rho &= \sum_{b=1}^B \sqrt{n_b^2 8\pi |\Sigma_b|^{1/2}} N(\mu_b; \mu_b, 4\Sigma_b) \\
&= \sum_{b=1}^B 8\pi n_b \frac{|\Sigma_b|^{1/2}}{2\pi |4\Sigma_b|^{(1/2)}} = \sum_{b=1}^B n_b = 1
\end{aligned}$$

This shows that any spatiogram compared to itself will always receive a similarity score of 1 using the new measure, which is its maximal value.

## 5. EXPERIMENTAL RESULTS

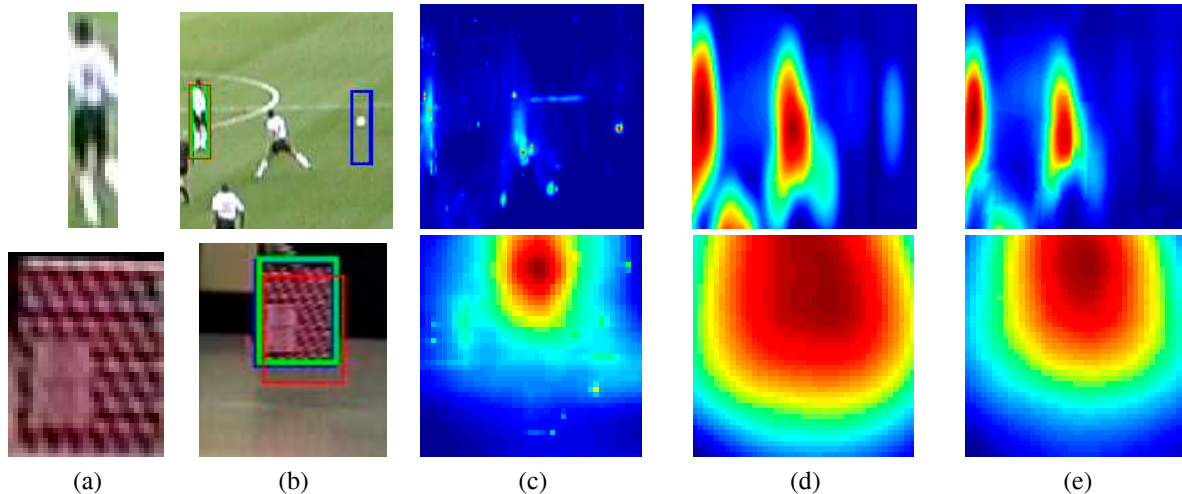
The plot in figure 1 shows the effect of noise on our proposed spatiogram similarity measure, compared to the original measure and to histogram similarity. The original similarity score is normalised by dividing by its maximum value so that it reaches a value of one at zero noise. The image we used (top right corner of figure 1) was selected as it contained pixel clusters with different variances. Adding Gaussian noise causes a sharp decrease in the similarity score of the original measure, due to its intolerance of small spatial changes. Histogram based matching is quite insensitive to noise and even returns a relatively high matching score when the original image is lost in the noise signal. Our proposed measure has a linear response within a large window of added noise.

Figure 2 shows the similarity surfaces generated by (c)the original measure, (d)histogram matching and (e)our proposed measure for two tracking examples, corresponding to localising a rigid highly-textured object (a book) and a non-rigid human object (a football player). In the top row of fig 2(c) the original measure produces a *spiked* similarity surface and finds the best match on the ball, instead of the player. This is because it weights bins with low variances higher, such as the tightly clustered white pixels of the ball. Its surface in fig 2(c), bottom row, is also not very smooth, but it correctly localises the book. Figure 2(d) shows the histogram similarity surfaces. Since it contains no spatial information, it provides poor object localisation, as can be seen by its poor lock on the book in fig 2(b). Our proposed measure's surfaces, shown in fig 2(e), are smooth, as in histogram matching, but also have good object localisation, indicated by the tighter peaks.

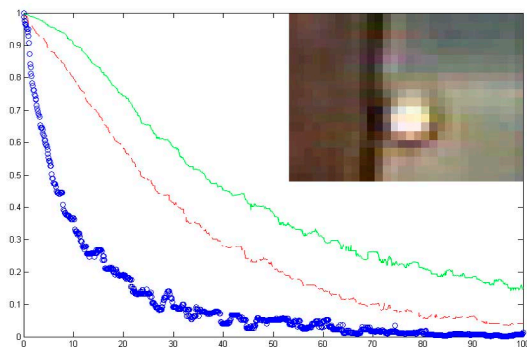
Our final experiment in figure 3 shows results from typical object-tracking scenarios. For this we use the multi-feature spatiogram-based mean-shift tracking framework described in our previous work [7]. We compare the original measure to our proposed measure for person tracking (top row) and head tracking (bottom row) using thermal infrared and colour pixel features. The original measure (shown in red) fails quickly when the tracked object moves, due to its narrow similarity surface peak. The mean-shift procedure generally fails when an object moves faster than the peak width. The proposed measure (shown in yellow) successfully tracks both objects. Videos online at: <http://www.eeng.dcu.ie/~oconaire/icassp07>

## 6. CONCLUSIONS

In this work, we have detailed the limitations of the original spatiogram comparison measure and have proposed an improved measure, derived from the Bhattacharyya coefficient. Its advantages, both in more accurate object localisation and in giving a smooth similarity surface, were demonstrated theoretically and on real video data. Future work will examine how to update object spatiogram models during tracking, while minimising model drift.



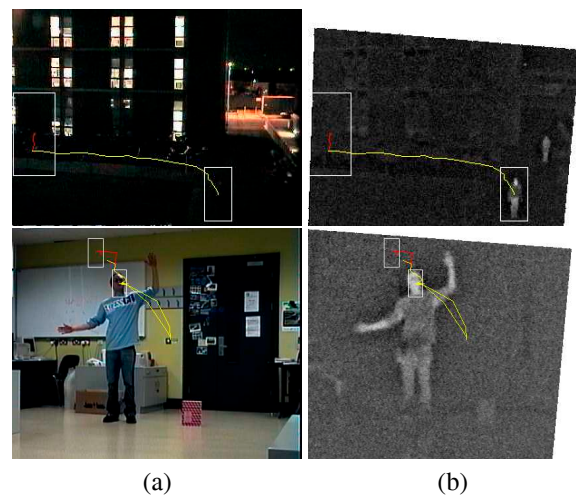
**Fig. 2.** (a) Target objects, (b) Search areas with best matches shown: Original measure (blue), Histogram matching (red), Proposed measure (green). Similarity surfaces: (c) Original measure, (d) Histogram matching and (e) Proposed spatiogram measure



**Fig. 1.** The effects on the similarity measures of adding Gaussian noise to a target image (top right): Histogram similarity (blue), original spatiogram similarity measure (green), proposed measure (red). Added noise RMS shown on x-axis.

## 7. REFERENCES

- [1] S. T. Birchfield and S. Rangarajan, "Spatiograms versus histograms for region-based tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2005, pp. 1158–1163.
- [2] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000, pp. 142–151.
- [3] R. T. Collins, "Mean-shift blob tracking through scale space," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003, vol. 2, pp. 234–240.
- [4] Z. Zivkovic and B. Krose, "An em-like algorithm for color-histogram-based object tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [5] I. Matthews, T. Ishikawa, and S. Baker, "The template update problem,"



**Fig. 3.** Multi-feature mean-shift tracking results using: (in red) Original measure, (in yellow) Proposed Measure. (a) Visible images and (b) Infrared images.

*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, June 2004.

- [6] A. Elgammal, R. Duraiswami, and L. S. Davis, "Probabilistic tracking in joint feature-spatial spaces," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2003.
- [7] C. Ó Conaire, N. E. O'Connor, and A. Smeaton, "Thermo-visual feature fusion for object tracking using multiple spatiogram trackers," *Machine Vision and Applications (submitted to)*, 2006.
- [8] C. Ó Conaire, N. E. O'Connor, A. Smeaton, and G. J. F. Jones, "Organising a daily visual diary using multi-feature clustering," in *Proc. of 19th annual Symposium on Electronic Imaging*, 2007.
- [9] K. Tumer and J. Ghosh, "Estimating the bayes error rate through classifier combining," in *In Proceedings of the International Conference on Pattern Recognition, Vienna, 1996*, pp. 695–699.