Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Extracting Boundaries from Images by Comparing Cooccurrence Matrices

Astrit Rexhepi[1,2], Azriel Rosenfeld[2], and Farzin Mokhtarian[1]

[1] Centre for Vision, Speech and Signal Processing
University of Surrey
Guilford, Surrey, GU2 7XH, UK
[2] Center for Automation Research
University of Maryland
College Park, MD 20742-3275 USA

Abstract. This paper describes methods of extracting region boundaries from the frames of an image sequence by combining information from spatial or temporal cooccurrence matrices of the frames. It summarizes past work on the uses of cooccurrence matrices for image segmentation; qualitatively describes the peaks (clusters of high values) that can be expected to occur in cooccurrence matrices when the image(s) contain smooth regions separated by stationary or moving boundaries; and describes methods of extracting stationary or moving region boundaries from an image by combining information from spatial and temporal cooccurrence matrices.

1 Cooccurrence matrices and their uses

Cooccurrence matrices, originally called gray-tone spatial dependency matrices, were introduced by Haralick et al. [1], who used them to define textural properties of images.

Let I be an image whose pixel gray levels are in the range $0, \ldots, 255$. Let $\delta = (u, v)$ be an integer-valued displacement vector; δ specifies the relative position of the pixels at coordinates (x, y) and $(x + u, y + v)$. A *spatial cooccurrence matrix* M_δ of I is a 256×256 matrix whose (i, j) element is the number of pairs of pixels of I in relative position δ such that the first pixel has gray level i and the second one has gray level j. Any δ, or set of δ's, can be used to define a spatial cooccurrence matrix. In what follows we will usually assume that δ is a set of unit horizontal or vertical displacements, so that M_δ involves counts of pairs of neighboring pixels.*

In addition to their original use in defining textural properties, cooccurrence matrices have been used for image segmentation. Ahuja and Rosenfeld [2] observed that pairs of pixels in the interiors of smooth regions in I contribute to elements of M_δ near its main diagonal; thus in a histogram of the gray levels

* Cooccurrence matrices based on smaller numbers of gray levels can also be used, but our method of combining cooccurrence matrices works better for larger matrices.

of the pixels that belong to such pairs, the peaks associated with the regions will be preserved, but the valleys associated with the boundaries between the regions will be suppressed, so that it becomes easier to select thresholds that separate the peaks and thus segment the image into the regions. In [3], Haddon and Boyce observed that homogeneous regions in I give rise to peaks (clusters of high-valued elements) near the main diagonal of M_δ, while boundaries between pairs of adjacent regions give rise to smaller peaks at off-diagonal locations; thus selecting the pixels that contribute to on-diagonal and off-diagonal peaks provides a segmentation of I into homogeneous regions and boundaries.

Pairs of pixels in the same spatial position that have a given temporal separation in a sequence of images can be used to define *temporal cooccurrence matrices*. Let I and J be images acquired at times t and $t + dt$; thus dt is the temporal displacement between I and J. A temporal cooccurrence matrix M_{dt} is a 256×256 matrix whose (i, j) element is the number of pairs of pixels in corresponding positions in I and J such that the first pixel has gray level i and the second one has gray level j.

Boyce et al. [4] introduced temporal cooccurrence matrices and used them in conjunction with spatial cooccurrence matrices to make initial estimates of the optical flow in an image sequence. They demonstrated that an initial probability of a pixel being in the interior or on the boundary of a region that has smooth optical flow in a given direction in a pair of images could be derived from the positions of the peaks in a spatial cooccurrence matrix of one of the images for a displacement in the given direction, and in the temporal cooccurrence matrix of the pair of images. Borghys et al. [5] used temporal cooccurrence matrices to detect sensor motion in a moving target detection system by comparing the spatial cooccurrence matrix of one of the images with the temporal cooccurrence matrix of the pair of images.

2 The structure of cooccurrence matrices

In the next section we will describe methods of using spatial or temporal cooccurrence matrices to extract stationary or moving region boundaries from the images of a sequence. In this section we describe the peak structures that should be present in spatial and temporal cooccurrence matrices.

We assume that an image I is composed of regions in which (ignoring noise) the gray levels vary smoothly, and that if two regions are adjacent, they meet along a boundary at which the gray level changes significantly. It is well known (see [3]) that in a spatial cooccurrence matrix of I, each region (say having mean gray level g) should give rise to a peak centered on the main diagonal in approximate position (g, g); the sum of the element values in this cluster should be proportional to the area of the region. Similarly, each boundary between two adjacent regions (say having mean gray levels g and h) should give rise to a pair of off-diagonal peaks at approximate positions (g, h) and (h, g), and with value sum proportional to the length of the border.

Figure 1a is a frame of an image sequence showing a person in dark clothes standing in front of a gray screen in a laboratory. Figure 1b shows the histogram of the image. The peak at the low end of the grayscale represents the dark clothes; the peak near the middle of the grayscale represents the screen; and the plateau represents the other regions in the image. Figure 1c shows the spatial cooccurrence matrix of this image for unit displacements in all four horizontal and vertical directions; values of 50 or greater are displayed as white and values less than 50 are displayed as black. In this display, groups of on-diagonal clusters have fused together into elongated clusters, and the small off-diagonal clusters are not visible because the matrix elements in these clusters have values less than 50. The two short elongated on-diagonal clusters represent the dark clothes and the gray screen respectively, and the long elongated on-diagonal cluster represents the other regions in the image. In Figure 1d values greater than 0 in the cooccurrence matrix are displayed as white. The off-diagonal clusters are still not visible because the region boundaries are not very sharp, so pairs of pixels that differ by a unit displacement contribute to cooccurrence matrix elements close to the diagonal; these pairs therefore cannot be distinguished from pairs that belong to on-diagonal clusters. Figure 1e shows the nonzero values in a cooccurrence matrix based on horizontal and vertical displacements of 5. In this matrix, the off-diagonal clusters representing the boundaries between the dark clothes and the gray screen are clearly visible.

Let I and J be consecutive frames of an image sequence acquired by a stationary camera. Suppose the frames show an object moving against a stationary background at a rate of a few pixels per frame. In the temporal cooccurrence matrix of I and J, pairs of pixels that are in a moving region in both images will contribute to an on-diagonal peak, and pairs of pixels that are covered up or uncovered by the motion will contribute to a pair of off-diagonal peaks.

Figures 2a and 2b show the second and tenth frames of the image sequence that had Figure 1a as its first frame. Figures 2c and 2d show the nonzero elements of the temporal cooccurrence matrices of Figures 1a and 2a and of Figures 1a and 2b. In Figure 2c the nonzero values are concentrated near the main diagonal, because relatively little motion took place between Figures 1a and 2a; but between Figures 1a and 2b considerable motion took place, so there are many off-diagonal nonzero values in Figure 2d. There are clusters of these values that correspond to significant covering and uncovering of the background by the person's dark arms. There are also “bridges” joining these clusters to the corresponding on-diagonal clusters (which are visible as bulges in the elongated on-diagonal cluster) because the uncovered background region is composed of parts that have a variety of average gray levels.

3 Extracting boundaries using cooccurrence matrices

As discussed in Section 2, boundaries between contrasting neighboring regions in an image give rise to off-diagonal peaks in a spatial cooccurrence matrix of the image, and motion of an object against a contrasting background between two frames of an image sequence gives rise to off-diagonal peaks in a temporal cooccurrence matrix of the two frames. Thus it should be possible in principle to extract boundaries from an image by detecting off-diagonal peaks in its spatial cooccurrence matrix and identifying the image pixels that contributed to those peaks. Similarly, it should be possible in principle to extract moving boundaries from a pair of successive frames of an image sequence by detecting off-diagonal peaks in the temporal cooccurrence matrix of the two frames and identifying the pixels in either of the frames that contributed to those peaks.

Unfortunately, as we saw in Section 2, off-diagonal peaks are not always easy to detect in cooccurrence matrices. Since the images are noisy, all the elements near the diagonal of a cooccurrence matrix tend to have high values, and the presence of these values makes it hard to detect off-diagonal peaks in the matrix that lie close to the diagonal since these peaks tend to have lower values. If we knew the standard deviation of the image noise, we could estimate how far the high values which are due to noise extend away from the diagonal of the cooccurrence matrix, and we could then look for peaks in the matrix that are farther than this from the diagonal; but information about the image noise level is usually not available.

In this section we describe a simple method of suppressing clusters of high-valued elements from a spatial or temporal cooccurrence matrix. As we will see, the suppressed matrix elements tend to lie near the diagonal of the matrix. Hence when the suppression process is applied to a spatial cooccurrence matrix, the image pixels that contributed to the unsuppressed elements of the matrix tend to lie on region boundaries, and when the suppression process is applied to a temporal cooccurrence matrix the image pixels that contributed to the unsuppressed elements of the matrix tend to lie on the boundaries of moving regions.

Our method of suppressing clusters of high-valued elements from a coocurrence matrix takes advantage of two observations:

(1) The matrix elements in the vicinity of a high-valued cluster almost certainly have nonzero values, so that the nonzero values in and near the cluster are "solid". On the other hand, it is more likely that there are zero-valued elements in and near a cluster of low-valued elements, so that the nonzero values in and near such a cluster are "sparse".

(2) As we saw in Section 2, the on-diagonal clusters in a cooccurrence matrix, when arise from regions in the image, can be expected to be symmetric around the main diagonal, and the off-diagonal clusters, which arise from boundaries between contrasting regions in the image, can be expected to occur in pairs whose means are symmetrically located around the main diagonal, since the noise in the image has zero mean. Hence if we have two cooccurrence matrices that are transposes of one another (see below), the clusters in these matrices should occur in the same approximate positions.

We can obtain spatial cooccurrence matrices that are transposes of one another by using symmetric displacements—e.g., unit displacements to the right and downward in one matrix, and unit displacements to the left and upward in the other. Similarly, we can obtain temporal cooccurrence matrices that are transposes of one another by using reverse temporal displacements; i.e., if I and J are successive frames of an image sequence, we can use the temporal cooccurrence matrices of I and J and of J and I. Figures 3a and 3b shows the nonzero elements in two spatial cooccurrence matrices of Figure 1a; unit leftward and upward displacements are used in Figure 3a, and unit rightward and downward displacements are used in Figure 3b. Figure 4a (the same as Figure 2c) shows the nonzero elements in the temporal cooccurrence matrices of Figure 1a and 2a, and Figure 4b shows the nonzero elements in the temporal cooccurrence matrix of Figures 2a and 1a. Evidently, Figures 3a and 3b are transposes of each other, and Figures 4a and 4b are transposes of each other.

Let M and N be two cooccurrence matrices that are transposes of one another. We suppress from M all elements that are nonzero in N (or vice versa). Elements of M that are in or near a "solid" cluster will almost certainly have nonzero values in N; hence these elements will almost certainly be suppressed from M. On the other hand, many of the elements of M that are in or near a "sparse" cluster will have zero values in N because the nonzero elements of these clusters in M and N are not in exactly symmetrical positions; hence many of these elements will not be eliminated by the suppression process.

Figure 5a shows the nonzero elements of Figure 3b that are zero in Figure 3a, and Figure 5b shows the nonzero elements of Figure 3a that are zero in Figure 3b. We see that the "solid" parts of the matrix have been suppressed and the "sparse" parts have survived. Figure 5c shows the pixels of Figure 1a that contributed to the nonzero elements in Figures 5a and 5b. Almost all of these pixels lie on region boundaries in Figure 1a; note, however, that many of the pixels that lie on low-contrast boundaries have been suppressed, because such boundaries contribute to near-diagonal elements of the matrix.

Figure 6a shows the nonzero elements of Figure 4b that are zero in Figure 4a, and Figure 6b shows the nonzero elements of Figure 4a that are zero in Figure 4b. Here too, the "solid" parts of Figures 4b and 4a (respectively) have been suppressed, but the sparse parts have largely survived. Figure 6c shows the pixels of Figure 1a that contributed to the nonzero elements in Figures 6a and 6b. Almost all of these pixels lie on boundaries of the person's body; they are especially strong on the boundaries of the hands and arms, which have the greatest motion.

4 Concluding remarks

Spatial and temporal cooccurrence matrices can be combined in other ways to extract pixels that lie on boundaries. Figure 7a shows the nonzero elements in a (symmetric) spatial cooccurrence matrix of Figure 1a; Figure 7b shows the elements in Figure 7a that are also nonzero in the temporal cooccurrence matrix of Figure 2c; and Figure 7c shows the pixels that contributed to the nonzero elements of Figure 7b. We see that these pixels nearly all lie on boundaries in

Figure 1a; they provide an even stronger representation of these boundaries than we had in Figure 5, because Figure 7b has many more nonzero elements than Figures 5a-b.

Because of space limitations, we have made no attempt to provide a theoretical analysis of our methods or to compare them with other methods of cooccurrence-based segmentation or of spatial or temporal boundary detection. However, we should point out that since our methods involve only Boolean comparisons of cooccurrence matrices, they are quite inexpensive computationally.

References

1. Haralick, R.M., Shanmugam, R., Dinstein, I.: Textural features for image classification. IEEE Trans. on Systems, Man, and Cybernetics **3** (1973) 610–621
2. Ahuja, N., Rosenfeld, A.: A note on the use of second-order gray-level statistics for threshold selection. IEEE Trans. on Systems, Man, and Cybernetics **8** (1978) 895–898
3. Haddon, J.F., Boyce, J.F.: Image segmentation by unifying region and boundary information. IEEE Trans. on Pattern Analysis and Machine Intelligence **12** (1990) 929–948
4. Boyce, J.F., Protheroe, S.R., Haddon, J.F.: A relaxation computation of optic flow from spatial and temporal cooccurrence matrices. International Conference on Pattern Recognition **3** (1992) 594–597
5. Borghys, D., Verlinde, P., Perneel, C., Acheroy, M.: Long-range target detection in a cluttered environment using multi-sensor image sequences. Proc. SPIE **3068** (1997) 569–578

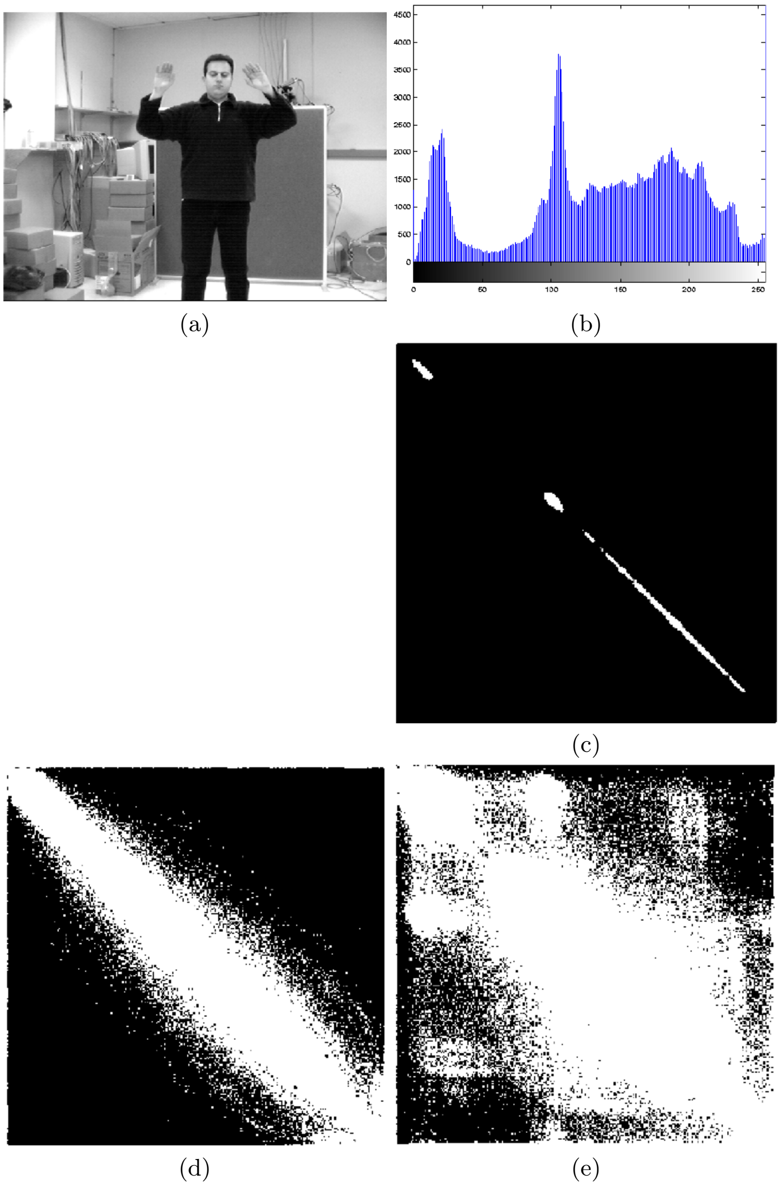
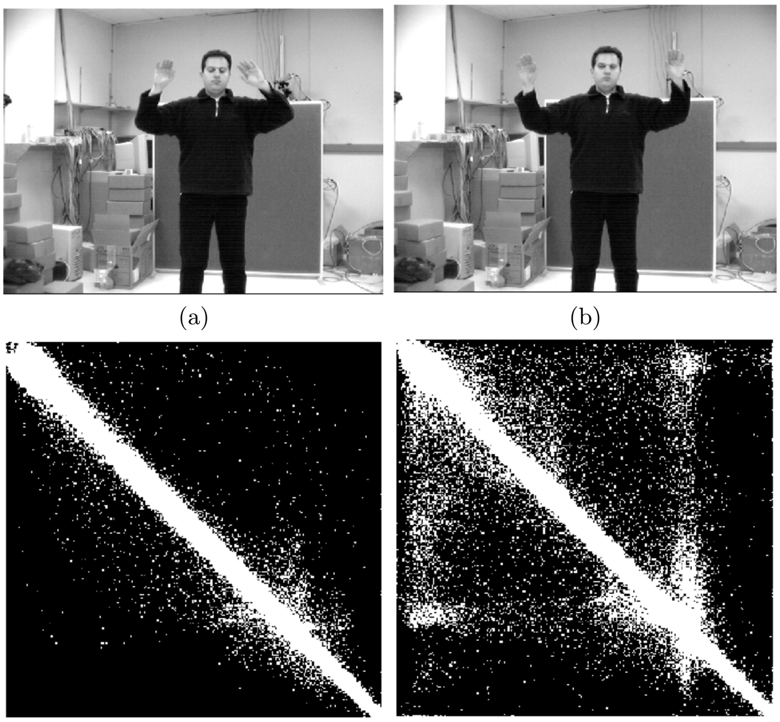


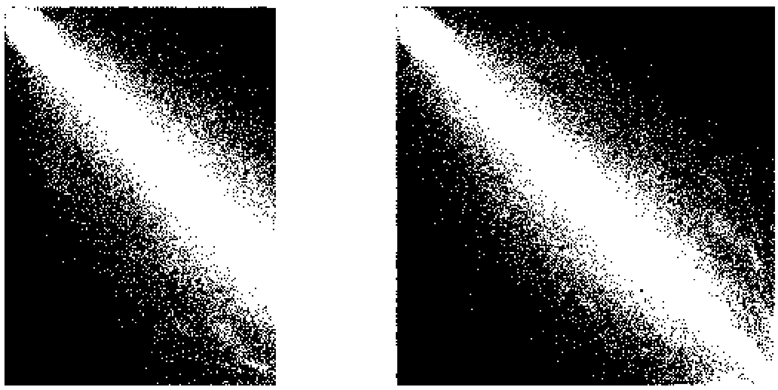
Fig. 1. (a) A real image. (b) Its histogram. (c) Its spatial cooccurrence matrix for one-pixel horizontal and vertical displacements in all four directions; values of 50 or greater are shown as white. (d) The same matrix with values greater than 0 shown as white. (e) The analogous matrix for five-pixel displacements.



(a) (b)

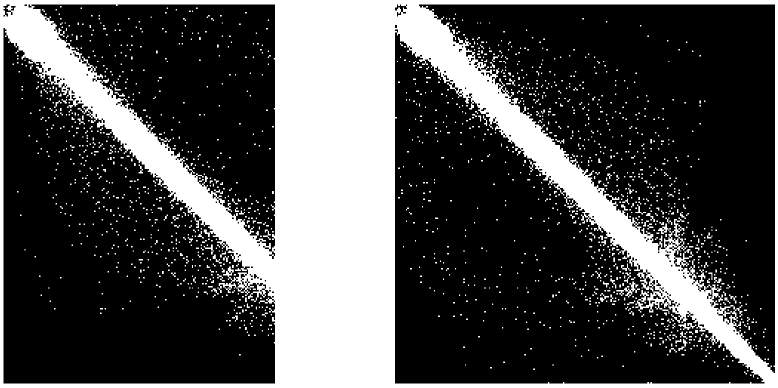
(c) (d)

Fig. 2. (a,b) Two other frames of the image sequence containing Figure 1a; Figures 1a,2a, and 2b are frames 1,2, and 10 of the sequence. (c-d) Nonzero elements in the temporal cooccurrence matrices of Figures 1a and 2a and of Figures 1a and 2b.



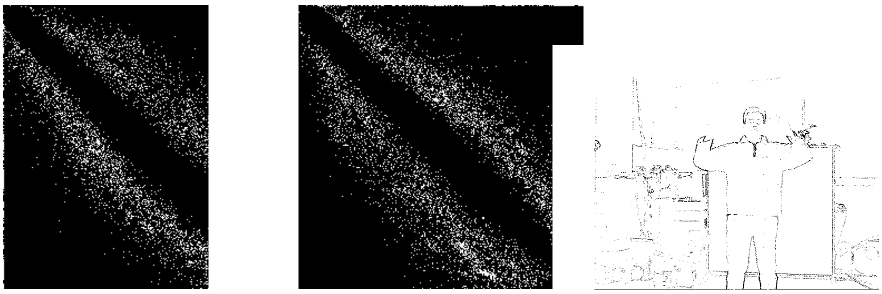
(a) (b)

Fig. 3. Nonzero elements in the spatial coocurrence matrices of Figure 1a for (a) rightward and downward displacements (b) leftward and upward displacements.



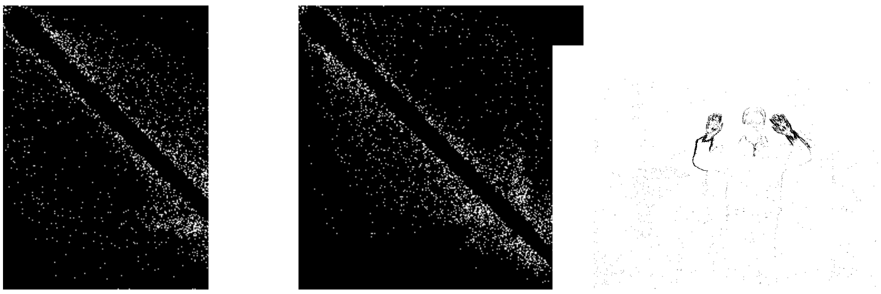
(a) (b)

Fig. 4. Nonzero elements in the temporal cooccurrence matrices of (a) Figures 1a and 2a (same as Figure 2c) and (b) Figures 2a and 1a.



(a) (b) (c)

Fig. 5. (a-b) Nonzero elements of Figure 3b that are zero in Figure 3a and vice versa. (c) Pixels of Figure 1a that contributed to the nonzero elements in Figures 5a-b.



(a) (b) (c)

Fig. 6. (a-b) Nonzero elements of Figure 4b that are zero in Figure 4a and vice versa. (c) Pixels of Figure 1a that contributed to the nonzero elements in Figures 6a-b.

(a) (b) (c)

Fig. 7. (a) Nonzero elements in the spatial cooccurrence matrix of Figure 2a for unit displacements in all four directions (same as Figure 2d). (b) Nonzero elements of Figure 7a that are zero in the temporal cooccurrence matrix shown in Figure 3c. (c) Pixels of Figure 1a that contributed to the nonzero elements in Figure 7b.