

# Motion estimation for crowd analysis using a hierarchical Bayesian approach

E. K. Tzamali      M. D. Plumbley\*  
S. A. Velastin

Department of Electronic Engineering  
King's College London, Strand, London WC2R 2LS, UK  
Email: mark.plumbley@kcl.ac.uk

## Abstract

We consider a probabilistic approach to block matching for motion estimation in images sequences of crowds. We present initial results from a probabilistic block matching approach, and compare with the results of a hardware block match processor. We also consider the use of a joint block probability model, based on the joint probability of  $n$ -best displacements from neighbouring blocks, as a possible improvement to independent block matching. We feel that this approach, combined with estimation of edge probabilities, could offer a practical method for improving motion estimation in crowd image sequences.

## 1 Introduction

Interest in the analysis of visual motion has been increasing over the past decade or so within the computer vision and neural network communities. As well as applications in e.g. traffic monitoring [11], human figures are very important recognition targets for many applications such as the movement in a shopping centre [4, 5, 7] or detection of abnormal activities on a train platform [6].

Previous work in our laboratory has been based on motion estimation as a first step for image processing in a crowd monitoring system [9, 10, 12].

## 2 Block matching

Probably the most popular motion estimation technique is *block-matching*. The digitized image is divided into blocks to be processed individually. For each *reference block* in one image, a search is made over all shifted versions of that block (candidate blocks) within a rectangular region in the next frame, known as the *search window*. The candidate block with the minimum distortion from the reference block gives the estimated motion of the reference block.

Many criteria exist for the distortion measure between reference and candidate blocks. For example, the STi3220 motion estimation processor (SGS-Thomson Microelectronics) used in

---

\*Corresponding author

our laboratory uses mean absolute error (MAE) [3], which offers a good trade-off between complexity and efficiency.

There are other techniques which can be applied to motion estimation. For example, motion estimation along a line only may be suitable for real-time road traffic measurement [1].

For tracking of humans, a more recent approach is the point distribution model (PDM), in which a distribution model is generated by performing a statistical analysis on a set of examples of the objects of interest [8]. Baumberg and Hogg [2] extend this approach by building 2-D contour models automatically, from sequences of training images.

### 3 Probabilistic block matching

In our consideration, we have a sequence of input images  $X(t) = [I_{km}(t)]$  where  $0 \leq k \leq K$  and  $0 \leq m \leq M$  are the  $x$  and  $y$  coordinates of the input image, and  $t = 1, 2, \dots$  is the image frame number. We will have grey level integer pixel values, so  $X_{km} \in \{0, 1, \dots, 255\}$ . In this paper we will be using  $K \times M = 40 \times 40$  pixel images selected from larger  $512 \times 512$  images of a railway station concourse.

Now, our task is to compare blocks in one frame to that in another to find a good match. The candidate block in frame  $t + 1$  is offset from the reference block in frame  $t$  by a distance  $\mathbf{d} = (d_x, d_y)$ .

Suppose the intensity of the reference block in frame  $X(t)$  is  $I_i$ , and the intensity of the candidate block in frame  $X(t + 1)$  is  $I'_i$ , where  $i$  ranges over the  $N$  pixels in the blocks (we use  $N = 8 \times 8 = 64$  pixel blocks). Then we typically measure the distortion between the blocks by a distortion (or error) measure

$$D(I, I') = \frac{1}{N} \sum_{i=1}^N d(I'_i - I_i). \quad (1)$$

For example,  $d_{\text{MAE}}(p, q) = |p - q|$  for mean absolute error (MAE), or  $d_{\text{MSE}}(p, q) = (p - q)^2$  for mean squared error (MSE). For many distortion measures, including MAE and MSE, we can express this as a function of the difference  $\Delta I_i = I'_i - I_i$  between the individual pixel identities.

#### 3.1 Block mismatch modelling

We would like to create a Bayesian model of this distortion. Suppose we wish to find the reference-to-candidate offset  $\mathbf{d}$  with maximum probability density given the difference  $\Delta I$  between the reference block  $I$  and candidate block  $I'$ , i.e.  $\arg \max_{\mathbf{d}} p(\mathbf{d}|\Delta I)$ . From Bayes theorem we know that

$$p(\mathbf{d}|\Delta I) = \frac{p(\mathbf{d}, \Delta I)}{p(\Delta I)} \quad (2)$$

but  $p(\Delta I)$  is independent of  $\mathbf{d}$ , so it suffices to choose  $\mathbf{d}$  to maximize  $p(\mathbf{d}, \Delta I) = p(\Delta I|\mathbf{d})p(\mathbf{d})$ .

Now, let us assume that the difference  $\Delta I$  is caused by additive noise, which for simplicity we will assume is additive gaussian of variance  $\sigma_I^2$ . Then for a given reference-to-candidate offset distance  $\mathbf{d}$  we will have a probability density for the difference between individual pixels of

$$p(\Delta I_i|\mathbf{d}) = \frac{1}{(2\pi\sigma_I^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_I^2}\Delta I_i^2\right) \quad (3)$$

so assuming that the additive noise on each pixel is independent, we get the probability density of the difference between the candidate and reference blocks  $\Delta I$  to be

$$p(\Delta I|\mathbf{d}) = \prod_{i=1}^N p(\Delta I_i|\mathbf{d}) = \frac{1}{(2\pi\sigma_I^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_I^2} \sum_{i=1}^N \Delta I_i^2\right) \quad (4)$$

or

$$-\ln p(\Delta I|\mathbf{d}) = \frac{1}{2\sigma_I^2} \sum_{i=1}^N \Delta I_i^2 + \text{constant} \quad (5)$$

so  $p(\Delta I|\mathbf{d})$  is maximized when  $ND_{\text{MSE}} = \sum_{i=1}^N \Delta I_i^2$  is maximized.

Now we also need a displacement model for the probability density of displacements  $\mathbf{d}$  from one frame to the next. For simplicity, we will use a gaussian of equal variance  $\sigma_d^2$  in the vertical and horizontal directions,

$$p(\mathbf{d}) = \frac{1}{(2\pi\sigma_d^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_d^2} |\mathbf{d}|^2\right). \quad (6)$$

Thus since we have  $p(\Delta I, d) = p(\Delta I|\mathbf{d})p(\mathbf{d})$  we get

$$-\ln p(\Delta I, \mathbf{d}) = \frac{1}{2} \left( \frac{1}{\sigma_I^2} \sum_{i=1}^N \Delta I_i^2 + \frac{1}{\sigma_d^2} |\mathbf{d}|^2 \right) + \text{constant} \quad (7)$$

so we must choose  $\mathbf{d}$  in order to minimize the equation above. This can be interpreted as a weighted sum of the mean squared error between pixels, and the square offset distance between the reference and candidate frames.

## 4 Experimental results

We compared this probabilistic model with the results of the STi3220 hardware motion estimation processor already in use in our laboratory. For the probabilistic model, we used  $40 \times 40$  frames selected from the original  $512 \times 512$  images, with  $8 \times 8$  blocks for the candidate and reference blocks. In this test we used  $\sigma_I^2 = 1^2 = 1$  and  $\sigma_d^2 = 8^2 = 64$ . A full search was carried out over the entire frame, yielding  $33 \times 33 = 1089$  candidate blocks for each reference block. In comparison, the hardware used a  $24 \times 24$  search window, so there are  $17 \times 17 = 289$  candidate blocks.

In each case, the results are presented as a motion vector line drawn in grey from the top left corner of each reference block, to the corresponding position of candidate block with the best match. To help visual interpretation, each motion vector line is drawn in a different shade of grey, depending on the position of the reference block (black for top left, lighter grey for bottom right).

Figure 1 shows a foot movement extracted from an image sequence of passengers on a railway station concourse. We can see that the output from the probabilistic block match (Fig. 1(d)) is generally consistent with the hardware (Fig. 1(c)) in the central area. In particular, the right-and-down movement of the foot has been captured by both the hardware and the probabilistic block match. Also, some movement of the shadow is also detected, leading to a right-and-up movement from the bottom left corner. However, in the probabilistic block

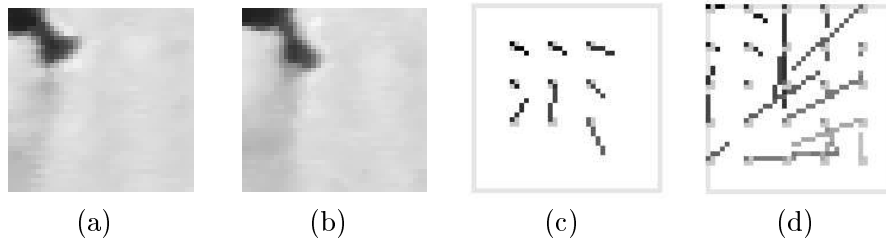


Figure 1: Foot movement showing (a) initial frame containing reference blocks, (b) next frame (containing candidate blocks), (c) results of the hardware motion estimation, and (d) results of the probabilistic block match.

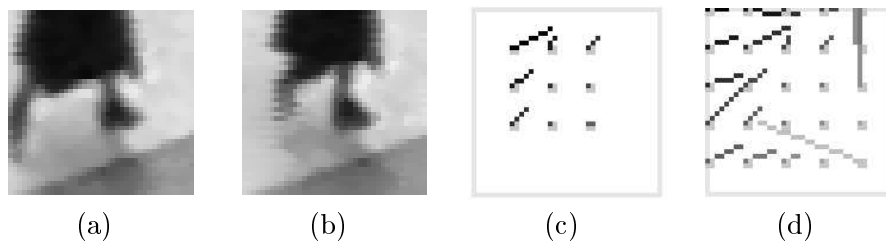


Figure 2: Body movement showing (a) initial frame containing reference blocks, (b) next frame (containing candidate blocks), and the results from (c) hardware motion estimation, and (d) probabilistic block match.

match some of these lead to large displacements ( $> 8$  pixels) which are inconsistent with the actual object movement.

Figure 2 shows a body movement extracted from the same image sequence. Here the customer has the foot placed on the floor, and the body moves above it, bringing the rear foot forwards. Both the hardware and the probabilistic block match have correctly interpreted the rightward (forward) body movement and the lifting of the back foot, as well as the stationary foot. However, the larger area covered by the probabilistic match indicates a large displacements for the blocks in the top right and bottom right, which are not what we would expect from the object movement.

These results indicate that refinements are still needed to this approach. For example, we expect that the relative variances of the image pixels  $\sigma_I^2$  and displacement  $\sigma_d^2$  will need adjusting. In addition, in some cases the large displacements were caused by edge effects. Blocks at the far right of the image had ‘nowhere to go’ so the best displacement was a long way back in the wrong direction (see e.g. the motion vector for the bottom left block in Fig 2(d)).

Alternatively, we may be able to improve the results by considering the joint movement of of neighbouring blocks. It is this approach that we consider next.

## 5 Joint block displacement

In the previous probabilistic model, we considered the movement of all reference blocks to be independent. However, in many cases (see e.g. the two previous figures) the movement

vector in one block is similar to the movement vector in a neighbouring block.

Suppose we can express this as a gaussian probability density of the motion vectors  $\mathbf{d}_1$  and  $\mathbf{d}_2$  of the two blocks,

$$p(\mathbf{d}_1, \mathbf{d}_2) = \left[ \frac{1}{(2\pi\sigma_-^2)^{1/2}} \exp\left(\frac{1}{4\sigma_-^2}|\mathbf{d}_1 - \mathbf{d}_2|^2\right) \right] \cdot \left[ \frac{1}{(2\pi\sigma_+^2)^{1/2}} \exp\left(\frac{1}{4\sigma_+^2}|\mathbf{d}_1 + \mathbf{d}_2|^2\right) \right] \quad (8)$$

where

$$\sigma_-^2 = \sigma_d^2(1 - \rho) \quad (9)$$

$$\sigma_+^2 = \sigma_d^2(1 + \rho) \quad (10)$$

and  $\rho$  is the correlation coefficient between the motion of  $\mathbf{d}_1$  and  $\mathbf{d}_2$ . Using this to find the  $\mathbf{d}_1$  and  $\mathbf{d}_2$  that jointly maximize the probability density  $p(\mathbf{d}_1, \mathbf{d}_2|\Delta I^1, \Delta I^2)$ , where  $\Delta I^b$  ( $b = 1, 2$ ) is the difference between reference block  $b$  and the candidate block offset from reference block  $b$  by  $\mathbf{d}_b$ , we must maximize

$$\begin{aligned} -\ln p(\Delta I^1, \Delta I^2, \mathbf{d}_1, \mathbf{d}_2) &= \frac{1}{2} \left( \frac{1}{\sigma_I^2} \left( \sum_{i=1}^N (\Delta I_i^1)^2 + \sum_{i=1}^N (\Delta I_i^2)^2 \right) \right. \\ &\quad \left. + \frac{1}{2\sigma_-^2} |\mathbf{d}_1 - \mathbf{d}_2|^2 + \frac{1}{2\sigma_+^2} |\mathbf{d}_1 + \mathbf{d}_2|^2 \right) + \text{constant} \quad (11) \end{aligned}$$

Now, if we have  $M$  possible candidate blocks for one block ( $M = 33 \times 33 = 1089$  in our small example), in theory we should need to consider  $M^2$  possible combinations of  $\mathbf{d}_1$  and  $\mathbf{d}_2$ . However, this would quickly become computationally intensive, particularly if we were to consider a larger number of blocks. Therefore we propose a method to reduce the number of candidate displacements to e.g. the 3 most probable displacements given by the single-block method. Thus the 3 best candidates for  $\mathbf{d}_1$  from the first block alone are combined with the 3 best candidates for  $\mathbf{d}_2$  for the second block alone, to give 9 possibilities each for  $\mathbf{d}_1 - \mathbf{d}_2$  and  $\mathbf{d}_1 + \mathbf{d}_2$ .

This approach is illustrated in the artificial example of Figure 3. This diagram shows ambiguous motion of the blocks in the first image, which could either be interpreted as moving towards each other, or as a pair downwards. The probabilistic block match from the individual blocks (Fig. 3(c)) interprets this motion as the two parts of the ‘object’ moving to the same final spot, which is the closest to each individual point. For a real object, this could only happen if these were two separate objects, and one had moved behind the other. (Note that the block matching models we consider here do not specifically model occlusion).

With the additional consideration of the joint displacement of the two blocks, the second-best match candidates (not visible in Fig. 3(c)) come in to play. Using a correlation coefficient of  $\rho = 0.8$  the common downwards motion (shown in grey in Fig. 3(d)) becomes more probable than the different cross-wise motion (shown in black in Fig. 3(d)).

## 6 Extension to more blocks

This approach could be extended to 4 blocks, 8 blocks, etc. in a hierarchical manner to handle motion of larger objects. However, the larger our collection of blocks, the more likely we are

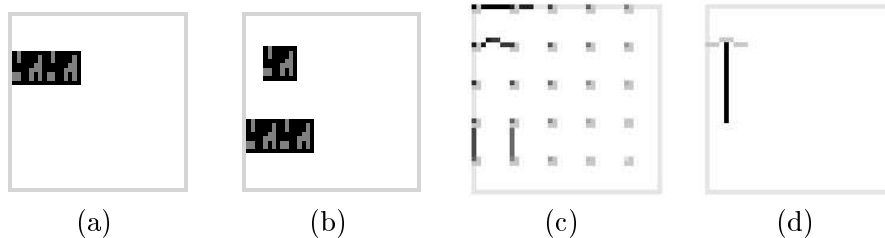


Figure 3: Ambiguous movement of a block pair showing (a) initial frame containing reference blocks, (b) frame containing candidate blocks, (c) probabilistic block match on individual blocks, (d) common motion extracted from motion of block pair.

to have an edge within the block, leading to motion vectors which will not be in the direction of the mean block motion.

One way to tackle this would be to explicitly model the probability of an edge. We could extend our Bayesian approach via the formula

$$p(\mathbf{d}_1, \mathbf{d}_2, e_{12}) = p(\mathbf{d}_1, \mathbf{d}_2 | e_{12}) P(e_{12}) \quad (12)$$

where  $P(e_{12})$  is the probability of a motion edge (i.e. a discontinuity in motion vectors) between blocks 1 and 2. We could either allow the edge decision  $e_{12}$  to be estimated from global image statistics, or use an image edge detector to give an improved estimate of the existence of a motion edge (i.e. the likelihood of a motion edge is increased if a luminance edge exists at that point in an image).

It may also be that a joint model of motion edges and luminance edges could improve object segmentation based on luminance images alone. We believe that this will be an interesting approach for further study.

## 7 Conclusions

We have considered a probabilistic approach to motion estimation in crowd images, including the use of a probability density model for displacement, and a method for the use of joint block displacements to refine motion estimates.

Initial results are promising, and generally consistent with existing hardware, although they indicate that further refinement is needed.

## References

- [1] M. Aoki. Detection of moving objects using line image sequence. In *Seventh International Conference on Pattern Recognition*, volume 2, pages 784–786, 1984.
- [2] A. Baumberg and D. Hogg. Learning deformable models for tracking the human body. In *Motion-Based Recognition*, pages 39–60. Kluwer Academic Publishers, The Netherlands, 1997.
- [3] B. A. Boghossian. Real time motion detection in video signals. MSc thesis, Department of Electronic Engineering, King’s College London, 1997.

- [4] A. W. J. Borgers and H. J. P. Timmermans. City centre entry points, store location patterns and pedestrian route choice behaviour: A micro-level simulation model. *Socio. Econ. Plan. Sci.*, 20:25–31, 1986.
- [5] A. W. J. Borgers, H. J. P. Timmermans, and P. Waerden. Behaviour of shopping pedestrians in city centres: Model development and empirical tests. *Planning*, 34:15–25, 1988.
- [6] S. Bouchafa, D. Aubert, and S. Bouzar. Crowd motion estimation and motionless detection in subway corridors by image processing. *Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC'97)*, pages 332–337, 1997.
- [7] S. Butler. Modelling pedestrian movements in central Liverpool, 1978. Institute for Transport Studies, University of Leeds.
- [8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: Their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.
- [9] A. C. Davies, J. H. Yin, and S. A. Velastin. Computer-based image processing for the monitoring of crowds. *Civil Protection (UK Home Office)*, January 1995.
- [10] A. C. Davies, J. H. Yin, and S. A. Velastin. Crowd monitoring using image processing. *Electronics and Communication Engineering Journal*, 7:37–47, February 1995.
- [11] D. Koller, K. Daniilidis, and H. H. Nagel. Model based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10:257–281, 1993.
- [12] J. H. Yin. *Automation of Crowd Data Acquisition and Monitoring in Confined Areas Using Image Processing*. PhD thesis, Department of Electronic and Electrical Engineering, King's College London, September 1996.