# EXTRACTING STRUCTURAL FRAGMENTS FROM IMAGES SHOWING OVERLAPPING PEDESTRIANS

László Havasi*, Csaba Benedek**, Zoltán Szlávik** and Tamás Szirányi**
*Péter Pázmány Catholic University,
H-1052 Budapest Piarista köz 1., Hungary,
e-mail: havasi@digitus.itk.ppke.hu
**Analogic and Neural Computing Laboratory, Hungarian Academy of Sciences,
H-1111 Budapest, Kende u. 13-17, Hungary,
e-mail: {szlavik, benedek, sziranyi}@sztaki.hu

**ABSTRACT**
This paper outlines and demonstrates a new algorithm which is capable of extracting characteristic fragments of the body outline of human figures from video image-sequences, even in the non-ideal case of typical outdoor illumination conditions and camera positions. Our method can derive relevant information regarding the significant body elements from video sequences showing walking people, without the necessity for imposing severe or unusual constraints with regard to the input images.
The proposed algorithm connects featured parts in the image into symmetrical objects, tracks them, and generates derived spatio-temporal statistical features which are used to ensure stable tracking results. Our method is fast enough for use in real-time. Using the grouped dual-point approach outlined here, we can extract biometric information suitable for subsequent analysis of the walking-gait characteristics, even in the case of overlapping and transient image outlines.

**KEY WORDS**
Motion analysis, symmetry tracking, structure of fragments, gait detection, image morphology

## 1. Introduction

Automatic detection of humans, and body-part localisation, are important but challenging problems in computer vision. Human motion analysis and tracking has long been proposed for applications in surveillance [1]. The primary step in analysis and tracking of human motion consists of the modelling of moving people represented in image sequences. Several approaches have been proposed for such modelling: e.g. elliptical cylinders [2], configuration of parameterised primitives [3], or 3-D tracking [4]. However, these methods are too complicated for effective detection of human figures in practical conditions [5]. Other common methods are the shape decomposition method [6], and the skeleton-based representation [7], which has been used to model the topological structure of the body. The contour-based representations can be extended with the use of deformable templates to handle shape deformations [8]. However, a drawback of such shape methods is that the model and the extracted image contour must first be aligned, which is not a trivial task. In addition, these methods cannot model individual parts of the body, so they can handle only a limited variety of shapes. In [9] a blob-based representation is introduced, which is useful in colour images. That method can successfully separate different people in the same image, provided that they are wearing distinct clothes; but it cannot extract detailed information about the various parts of the body.
The foundation of motion analysis is motion tracking. This task is very important because increased precision in tracking brings considerable improvements in recognition accuracy. This improved precision can be achieved by using the above methods in conjunction with spatio-temporal analysis [10]. Kalman filtering [11] is a widely used stochastic modeling method employed to handle occlusion and articulated motion.
Problems commonly arise in situations where the partitions of motions and of people in the input images are not trivial. Nevertheless, in our examples we are able to successfully analyse real images similar to those obtained in practice from city-wide surveillance systems. We used high-resolution (720x576 pixel), wide-angle cameras to observe human figures in busy outdoor locations; and in these quite realistic circumstances the resolution and contrast of the body outline of persons in the image is often rather poor. We also note that most publications focus on cases where only one, or at most a few, people are moving in the scene being analysed [12][17]. Our present method on the other hand is quite successful for multi-person images.
Our ultimate goal is to track moving people in complex scenes, with the help of biometric information derived from the images. The dynamic properties of walking uniquely characterise a moving human figure [13]. In future work we plan to use these attributes for analysis of images obtained in a multi-camera environment [19]. Thus our main thrust in the present paper is not to detect

and track human bodies as such, but rather to extract significant features suitable for subsequent biometric identification e.g. walking-phase and repetition frequency. Necessarily however the identification is sensitive to the accuracy of the derived attributes. Fortunately we can achieve the required precision in the case of tracking small structures on video image-sequences, although of course 100% accuracy is not attainable. Our principal aim in this paper is to isolate and track the legs in groups of humans whose images overlap in the given scene.

A final aspect, not mentioned in most publications, is the required computation time; this is very important in real applications. Our method is relatively fast, because it does not use iterative optimisation steps; it consists instead of simple operators that can be implemented using dedicated on-board image-processing hardware such as that described in [14] and [21].

Generally, object detection in videos must be done in two steps. Firstly, some detectable features are needed; secondly, robust tracking of the extracted features must be performed. Detection of humans in video image-sequences can be done in two ways:

- detection of features and their verification;
- detection of features that are characteristic only for humans or for human movements.

In our algorithm, two features are combined: symmetry extraction, and characteristic-points detection. Firstly, the detected object's symmetry-map is extracted; then, based on this feature, characteristic points are detected. The tracking of the features is done by considering the presence of one or two features; these are then used in a predictive fashion to assist in finding the possible location of another feature.

The present paper describes a new method, comprising the following processing steps on the input image:

- Adaptive background filtering
- Finding line segments along ridges
- Detecting the first-level symmetry axis
- Constructing third-level symmetries
- Finding characteristic points along edges
- Pairing points on both sides of first-level symmetries
- Tracking the points: their coming into being, disappearance, and transitions

In the present method first-level symmetries are used to generate dual points associated with symmetrical moving objects; while third-level symmetries are applied to validate these objects as belonging to leg-pairs of the moving persons in the image.

## 2. Adaptive background modelling for change detection

For the detection of changes in video image sequences we have implemented an adaptive background-modelling algorithm, as used by Stauffer et al. [18].

In that paper an adaptive mixture of Gaussians was used for approximation of background changes; this is necessary, because of the following considerations. If each pixel resulted from a single surface element under fixed lighting conditions, a single Gaussian would be sufficient to model the pixel value while accounting for acquisition noise. If only lighting variation was considered to be changed over the time, then a single adaptive Gaussian per pixel would be sufficient. In practice however, multiple surfaces often appear in the view frustum of a particular pixel, and the lighting conditions change as well. Thus, multiple adaptive Gaussians are required. We use an adaptive mixture of Gaussians to approximate this process. Each time their parameters are updated, the Gaussians are evaluated using a simple heuristic method to hypothesis which ones are most likely to be part of the 'background process'. Pixel values that do not match one of the pixel's 'background' Gaussians are grouped using connected components.

In the histogram (Figure 1) showing the grey-values of a single pixel over time we can observe several peaks, and the graph can be approximated well by a mixture of Gaussian functions.

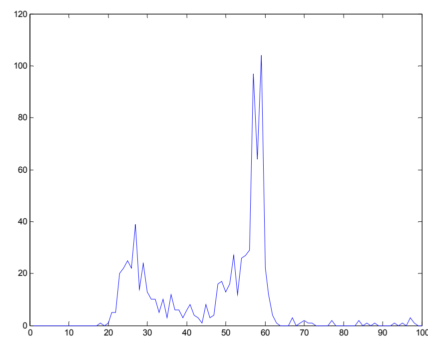



Figure 1: Histogram for grey-values of a single pixel for a short period. (Horizontal axis: pixel brightness; vertical axis: no. of occurrences.)

If we assign probabilities to the pixel values according to the histogram, and $X_t$ denotes a single pixel value at time t, the matching can be expressed as

$$P(X_t) = \sum_{i=1}^{K} w_{i,t} \cdot \eta(X_t, \mu_{i,t}, U_{i,t})$$

where $\eta(X_t, \mu_{i,t}, U_{i,t})$ is a Gaussian function with expected value $\mu_{i,t}$ and deviation $U_{i,t}$. However it is a very time-consuming operation to store and refresh a histogram for every pixel, and the Gaussian matching expectation maximization (EM) process would be correspondingly slow. Stauffer [18] recommended a real-time algorithm in which the parameters of the Gaussians are evaluated using a simple heuristic method to hypothesis which is the most likely to be part of the 'background process'. Every pixel value $X_t$ is checked against the existing $K$ Gaussian distributions until a match is found. A match is defined as a pixel value within 2.5 standard deviations of the distribution mean. If none of the $K$ distributions match the current pixel value, the least probable distribution is replaced with a distribution with the current value as its mean value $a$, an initially high variance, and low prior weight. After the classification of the current pixel value the model is re-estimated. The weight of the matched component will be increased and other weights decreased. The $\mu_{i,t}$ and $U_{i,t}$ parameters are also modified. The ‘raw’ output of the algorithm is a grainy binary-level foreground-background image (Figure 2c). We then use morphological operators to eliminate the grain, and obtain Figure 2d. To reduce the amount of computation, further processing is done only in image regions where moving objects have been detected.

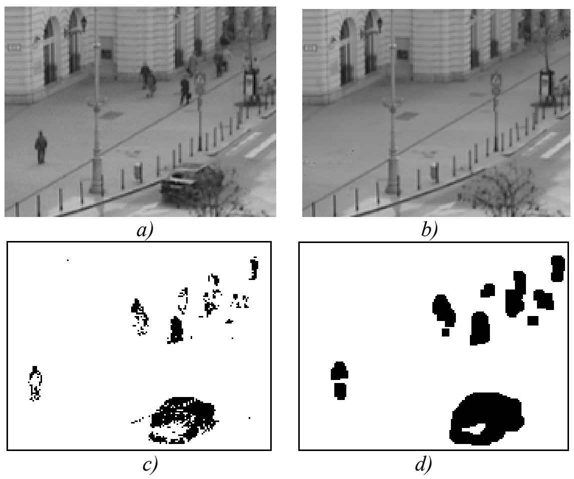



Figure 2: (a) The input image (b) an image composed of the means of the most probable Gaussians in the background model (c) the foreground pixels (d) the foreground pixels after median filtering and using dilations

## 3. Symmetry extraction

There are two main computation methods used for symmetry extraction:

- The Generalized Symmetry Operator applies distance weight function (affected by spacing), a phase weight function (affected by edge direction), and a logarithmic mapping of the intensity of points [15].
- The shock-based method, calculating symmetries by parallel waves propagating from the ridge [16].

In our proposed method [21], based on the latter approach, simulation of wave propagations from the edges and measurement of the collision points of two waves calculates the symmetry map of the object, as demonstrated in Figure 3. Simple morphological operators –dilatations – are used for the simulation of wave propagation. We call the resulting image the *map of Level 1 symmetries*.

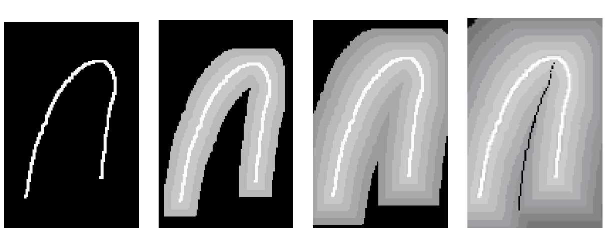


Figure 3: A simple object; and waves of dilatation after 10, 20, and 25 iterations, showing the derived symmetry axis

### 3.1 Extended symmetry extraction

The symmetry operator usually employs as its input the edge map of the object (Canny method). For simple binary edge maps, erroneous symmetry maps may be obtained in case of fragmented object-edges. The method can however be improved by the use of grey-level mode, in which we can add ‘weights’ to the object-edges in the image. The lengths of edge fragments are measured by using a flood-fill algorithm, and then each edge and wave will be weighted by a factor depending on the corresponding edge-fragment length. The computation of grey-level mode, similarly to the simulation of binary wave propagation, can be performed by using a grey-level morphological operator (max filter). The advantage is that the ‘higher’ waves can non-destructively overlap the ‘lower’ ones. With this solution, fragmented edges do not cause so many errors; an example of the improved result is shown in Figure 4. Other discontinuous errors can be minimised using vertical closing operators. This post-processing step is described in the next section, where we define the symmetry levels.

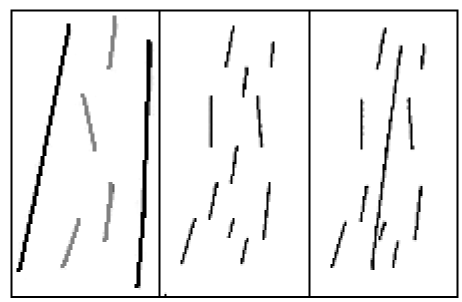


Figure 4: Horizontal symmetries of an artificial input image (left). The centre image is the result of binary wave propagation; while on the right is shown the result of grey-level wave propagation. With the latter method fragmented edges cause fewer errors, and the symmetry axis can be successfully extracted.

### 3.2 Deriving symmetry levels

As illustrated in Figure 5, the symmetry concept can be extended by iterative operations. The symmetry of the Level 1 symmetry map is the Level 2 symmetry map; and the symmetry of the Level 2 map shows the Level 3 symmetry (L3S). The advantage of this approach is that it does not confound the local and global symmetries in the image, so these levels are truly characteristic for the shape-structure. Our symmetry-extraction method is less sensitive to edge fragmentation than is the original “skeleton” method; but nevertheless the L3Ss do contain an accumulation of fragments from the preceding symmetry levels. To reduce this error we use vertical limiting operators at each level of processing. In addition, using vertical kernels (height greater than width) is effective when the objects are small and near to other one on the image. The vertically-oriented kernels help to avoid possible confusion with nearby neighbouring symmetries.

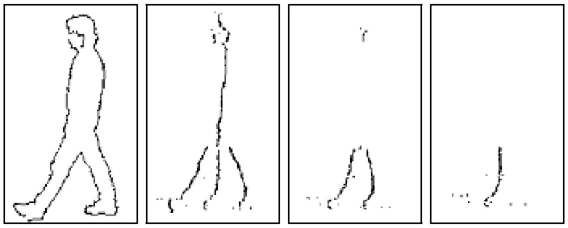


Figure 5: A simplified outline of the human body, and its symmetry-levels. The Level 1 symmetry map (2nd image) is the symmetry map of the extracted edge map. The Level 2 map (3rd image) is the symmetry map of the Level 1 map, and the last image shows the Level 3 symmetry map derived from the Level 2 map.

### 3.3 Applying third-level symmetries

We have found that if higher-level symmetries are calculated and tracked temporally then the third-level symmetries form a pattern, which pattern is uniquely characteristic for human locomotion (walking) [20], see figure 6.

Figure 6: Spatio-temporal patterns formed by the tracking of third-level symmetries of pedestrians.

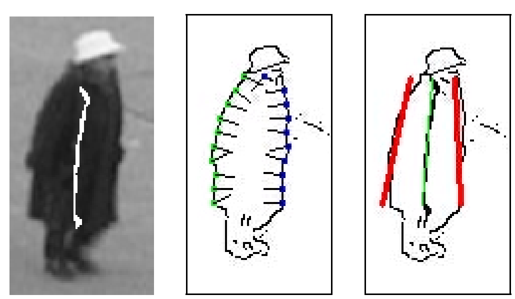
The previously extracted third-level symmetries are primarily presented between the legs, when the image is that of a human. The arms do not usually generate significant symmetries, among other reasons because of distortions arising from the perspective view, and because of their relatively small size in proportion to the whole body.
However, the existence of these symmetries in an image does not provide usable information about the content; for this, we have to track the changes of these symmetry fragments over time.
The method described in [20] can detect pedestrians with a promisingly high success rate (96.5%). In the proposed method we use these classification results to eliminate falsely-detected objects.

## 4. Characteristic points

Our previous work [17] introduced a point-cloud method for stochastic point selection and coupling the points to moving objects. In our experiments, we have found that using some structural information for tracking and grouping points can increase the final accuracy. This new structural information is obtained from the symmetry properties of the objects. The symmetry is a derived feature so its tracking can be achieved by the tracking of its components, which are the contour points. The proposed method uses the symmetry maps and the gradient of the contour in combination.
First we define dual-points around the symmetry axis. Dual-points are defined as a detected feature point on one edge, together with a corresponding point on the opposite side of the symmetry axis. Dual points are connected to each other by a distance parameter, which is the sum of the radii of the extracted first-level symmetry axes.

### 4.1 Pairing points on both sides of first-level symmetries

The initial step of tracking collects points into two sets. One set contains the ‘left’ points, and the other the ‘right’ points (left and right refer to the corresponding parts of the symmetry). Another useful piece of information is the radius of symmetry; this distance can define a relation between the two parts. Thus we can connect the opposite points using the distance condition, and we call these a ‘dual point’. In the course of tracking we take account of this condition. Thus a dual point denotes one symmetry point of a symmetry axis and two contour points of the shape. The main symmetry axis can be seen in Figure 7a; Figure 7b shows the dual points, and the resulting regression lines can be seen in Figure 7c. During tracking, these points should be moved to preserve the dual and symmetry properties. Since dual points were registered with respect to the contours, their matching property is the gradient of edges, as seen for the data in Figure 7b.
Our aim was to avoid the usual problems of silhouette tracking, namely the problem of occluded objects, partitioned edges etc. We exploit the main characteristics of the silhouettes of pedestrians; parallel behaviour of shapes, and the fact that motion is usually perpendicular to the main axis. On the other hand, we try to avoid the uncertainties resulting from points slipping forwards and backwards on the edges.
One of the main advantages of this approach is that it is able to directly track the points of a given shape with extremely low computation time, because it only has to match the relevant point pairs. The disadvantage is similar to that for other matching methods: searching for the matching points may result in more than one good fitting.



(a) main axis (b) dual points (c) regression lines
Figure 7: Demonstration of dual points

The dual points will be deleted if there is no acceptable matching point, but by employing the other points of the object it is easy to recover from this with a new symmetrical pair of points. To update and improve the objects we check the extension of the collimated axis. The points to be added to the object have to be disposed at a similar radius.
Having the set of feature points, and taking into account the already detected features (symmetry axis, radii, and edges of the legs), a dual-point can be defined on the edge on the opposite side to the symmetry axis.

### 4.2 Tracking the points: 'birth', 'death' and transitions

In the tracking of features, the dual-points are used to assist the tracking of the first-level symmetry axes. Firstly, a gradient-based method searches for dual-points in the current frame. If one point of a dual-point pair has been lost then, using the distance parameter, the algorithm searches for a new point on the edge-map in the current frame. A counter-value is recorded for each dual-point, which counts how many times one of the dual-point pair has been lost. If the value reaches a predefined limit, then the corresponding dual-point will be deleted from the set of valid dual-points. It can be seen in Figure 8 that the algorithm is capable of tracking partially-occluded objects; the dual-points can track an object for some time even if the symmetry axis disappears. If the cardinality of the set of dual-points is less than a predefined parameter value, the algorithm tries to find new characteristic points on the object edges, corresponding to the symmetry map. Therefore, as the object emerges from an occlusion more and more dual-points are found on it, and can be used to track it.

Figure 8: Initialization and updating of dual-points

The proposed algorithm was tested on video image-sequences captured by standard cameras (resolution 720×576 pixels, frame rate 24 frames/sec). In Figure 9 some typical detection results are shown. Five persons appear in this image; the proposed algorithm can successfully detect the legs of four of them, even in the case of occlusion. For the fifth person, it was not possible to extract the edges; the Canny edge detector method could not detect them even if the threshold was set to zero.
Even so, based on experiments on 400 test videos followed by comparisons using human evaluation of the same images, the successful detection rate of legs was found to be about 80%, while the false-detection rate was about 15%.

Figure 9: Detection results

## 3. Conclusion

The above-described method can detect and track symmetrical objects, and even in practical conditions human legs can be recognised with a high rate of success. In determining the accuracy, we have the difficulty that the level of acceptability is somewhat subjective. Nevertheless, in the test videos we measured about an 80% true detection rate and a 15% false detection rate. We have found that the method can track the motion of legs as long as the shapes are visible, and the algorithm works in real-time using images obtained in practical outdoor circumstances.
Our ultimate goal is to use pure biometric registration derived from images obtained from a multi-camera surveillance system [19]. The described feature extraction and tracking method will form the basis of future work on tracking pedestrians by using biometric information derived from real scenes.

## 4. Acknowledgements

The authors wish to acknowledge the support received from the Hungarian National Research and Development Programme, TeleSense project grant (NKFP) 035/02/2001.

## References:

[1] T. Boult, R. Micheals, A. Erkan, P. Lewis, C. Powers, C. Qian and W. Yin, Frame-rate multi-body tracking for surveillance, *Proc. of the DARPA Image Understanding Workshop*, Monterey, CA, 20-23 November 1998, 305-313.

[2] D. Hogg, Model-Based Vision: A Program to See a Walking Person, *Image and Vision Computing*, 1(1), 1983, 5-20.

[3] J. O'Rourke and N. I. Badler, Model-Based Image Analysis of Human Motion Using constraint propagation, *IEEE Trans. PAMI*, 2(6), 1980, 522-536.

[4] J. M. Rehg and T. Kanade, Model-based tracking of self-occluding articulated objects, *Proc. of the Int. Conf. on Computer Vision*, Cambridge, MA, 20-23 June 1995, 618-623.

[5] Liang Zhao, Dressed Human Modeling, Detection and parts Localization, PHD Thesis (Carneige Mellon University, Pittsburgh, 2001).

[6] F. Mokhtarian and A. K. Mackworht, ATheory of Multi-Scale, Curvature-Based Shape Representation for Planar Curves, *IEEE Trans. PAMI*, 14(8), 1992, 789-805.

[7] S. C. Zhu and A. L. Yuille, Forms: A Flexible Object Recognition and Modeling System, *Int. Journal of Computer Vision*, 20(3), 1996.

[8] A. Baumberg, D. Hogg, Learning Flexible Models from Image Sequences, *Proc. European Conf. on Computer Vision*, 1994, 299-308.

[9] C. Wren, A. Azarbayejani, T. Darrel and A. Penltand, Pfinder: Real time Tracking of the Human Body, *IEEE Trans. PAMI*, 19(7), 1997, 780-885.

[10] D.-S. Jang and H.-I. Choi, Active Models for Tracking Moving Objects, *Pattern Recognition*, 33(7), 2000, 1135-1146.

[11] Y.-S. Yao and R. Chellappa, Tracking a Dynamic Set of Feature Points, *IEEE Trans. on Image Processing*, 4(10), 1995, 1382-1395.

[12] I. Haritaoglu, D. Harwood and L. S. Davis, $W^4$: Real-Time Surveillance of People and Their Activities, *IEEE Trans. PAMI*, 22(8), 2000, 809-830.

[13] Hayfron-Acquah, J., Nixon, M. S. and Carter, J. N., Human Identification by Spatio-Temporal Symmetry, *Proceedings of International Conference on Pattern Recognition*, Quebec, 2002, 632-635.

[14] Bi-I: Bio-inspired Real-Time Very High Speed Stereo or Mono Image Processing System, www.analogic-computers.com/cgi-bin/sub_pages/products/bi-i.php

[15] D. Reisfeld, H. Wolfson, and Y. Yeshurun, Context-free attentional operators - The Generalized Symmetry Transform, *Int. Journal of Computer Vision*, 17, 1995, 119-130.

[16] D. Sharvit, J. Chan, H. Tek and B.B. Kimia, Symmetry-Based Indexing of Image Databases, *J. Visual Comm. And Image Representation*, 9(4), 1998, 366-380.

[17] L. Havasi and T. Szirányi, Motion Tracking Through Grouped Transient Feature Points, *Proc. ACIVS*, Ghent, 2003.

[18] C. Stauffer, W. Eric L. Grimson, Learning patterns of activity using real-time tracking, *IEEE Trans. PAMI, 22(8),* 2000, 747-757.

[19] Z. Szlávik, L. Havasi, T. Szirányi, Estimation of common groundplane based on co-motion statistics, ICIAR, Lecture Notes on Computer Science, 2004, accepted

[20] L. Havasi, Z. Szlávik, T. Szirányi, Pedestrian detection using derived third-order symmetry of legs, ICCVG, 2004, under reviewing

[21] L. Havasi and Z. Szlávik, Symmetry feature extraction and understanding, *Proc. CNNA*, Budapest, 2004.