# PEDESTRIAN DETECTION USING DERIVED THIRD-ORDER SYMMETRY OF LEGS

*A novel method of motion-based information extraction from video image-sequences*

László Havasi[1], Zoltán Szlávik[2] and Tamás Szirányi[2]

*[1]Péter Pázmány Catholic University, Piarista köz 1., H-1052 Budapest Hungary; [2]Analogic and Neural Computing Laboratory, Hungarian Academy of Sciences, PO Box 63 H-1518 Budapest Hungary*

Abstract: The paper focuses on motion-based information extraction from video image-sequences. A novel method is introduced which can reliably detect walking human figures contained in such images. The method works with spatio-temporal input information to detect and classify the patterns typical of human movement. Our algorithm consists of easy-to-optimise operations, which in practical applications is an important factor. The paper presents a new information-extraction and temporal-tracking method based on a simplified version of the symmetry which is characteristic for the legs of a walking person. These spatio-temporal traces are labelled by kernel Fisher discriminant analysis. With this use of temporal tracking and non-linear classification we have achieved pedestrian detection from real-life images with a correct classification rate of 96.5%.

Key words: simplified symmetry, pedestrian detection, tracking, surveillance, kernel Fisher discriminant analysis

## 1. INTRODUCTION

In outdoor multi-camera systems such as city-wide distributed monitoring systems in public places, the image-resolution of the surveyed objects is usually comparatively low, while the image-noise originating from lighting conditions and background content is relatively high. In the recognition and tracking of humans by such systems, the first step is target-

detection. Detection of humans in video sequences has been attempted by several different methods, depending on the requirements of the particular application. Model-based methods use matching with *a priori* shapes (Mohan et al., 2001), similarly to the case of optimisation for active contours. Active contour methods (Kass et al., 1988) can in principle handle the detection problem, but the initialisation of weights is sensitive to image deformations and contour-splitting. Optical tracking of image parts and their high-level interpretation can lead to acceptable results (Nguyen et al., 2000), but the method works satisfactorily only in cases where the image contains detailed textures. Song et al. (2003) have presented an unsupervised-learning method for derivation of a probabilistic model of human motion from unlabelled cluttered data. They reported a very promising 4 percent error rate for pedestrian detection, albeit under favourable conditions (in a moderately restricted environment with persons viewed from the side). The periodic character of walking was exploited by Abdelkader et al. (2002); their method can detect humans based on an image sequence covering 5-8 step-periods. The practicability of the use of symmetries for human identification is discussed by Hayfron (2002).

In outdoor environments with practical image-resolution and varied lighting conditions, there is only one well-defined criterion for the recognition of a pedestrian: the walking person must use his two legs. The aim of our paper is to introduce an approach for pedestrian detection in real scenes which can produce a reasonably good false-positive detection rate. We outline a novel feature-extraction and tracking method that can reflect the inherent structural changes of target shape, thus enhancing the method's practical utility.

## 2. FEATURE EXTRACTION AND CLASSIFICATION USING SYMMETRY

Symmetry is a basic geometric attribute, and most objects have a characteristic symmetry-map. These unique and invariant properties lead to the applicability of symmetries in our approach to image-processing. Our method (Havasi and Szlávik 2004) employs a modified shock-based method (Sharvit 1988): it calculates symmetries by propagating parallel waves from the ridge. The general shock-based approach (also called grey-level skeleton) has the limitation that it is sensitive to image noise, and particularly to the presence of discontinuous edges.

Our symmetry-detection method is based on the use of morphological operators to simulate spreading waves from the edges. Each iteration involves one step of spreading; the symmetry points are marked at the

collision-points of the waves. The radius of the extracted symmetry axis corresponds to the number of iterations, to the distance between the collision point and the edge point from both parts. In our approach, we simplify the algorithm by using only horizontal morphological operators; since, in the practical cases we are considering, we essentially need to extract only vertical symmetries. This modification has the advantage that it assists in reducing the sensitivity to fragmentation. Sample outputs of the algorithm can be seen in Figure 1. The symmetry operator normally uses the edge map of the original image as its input; we used the Canny edge-detector algorithm to derive the locations of the edges (ridges). To test the robustness of the algorithm in processing real public-place scenes, in our trials we did not employ any background subtraction or change-detection method. The algorithm described is insensitive to minor edge fragmentations, and a "perfect" definition of the target outline is unnecessary.

*Figure 1.* An idealised outline of a walking person, together with the derived Level 1, Level 2, and Level 3 symmetry maps.

As illustrated in Figure 1, the symmetry concept can be extended by iterative operations. The symmetry of the Level 1 symmetry map is the Level 2 symmetry; and the symmetry of the Level 2 map is 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 of the pair of legs. In the further processing steps we use only L3Ss. It is obvious however that image noise and edge fragmentation will damage the symmetries, especially the L3S. To minimise such errors an appropriate pre-processing method is to filter the captured image-frame using median filters, applied more than once, to remove small errors. This step has proved effective in processing compressed video images because it retains the essential structure of objects while it removes the irrelevant marks. Our symmetry-extraction method is less sensitive to edge fragmentation than is the original "skeleton" method; but nevertheless the L3Ss 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, it is an important factor when the objects are small and near to one another on the image. The vertically-oriented kernels help to avoid possible confusion with nearby neighbouring symmetries.

## 2.1 Temporal Tracking of Symmetries

The extracted L3Ss for a human target are in practice useful primarily with respect to analysis of images of the legs. The arms do not usually generate significant symmetries, among other reasons because of distortions arising from the perspective view, and because of their small size in proportion to the whole body. Thus the resulting symmetry-image from the arms is typically composed of small fragments which are difficult to distinguish from the noise. However, even the existence of clear symmetries in a single static image does not necessarily provide usable information about the image content; for this, we need to track the changes of the symmetry fragments by temporal comparisons. By using the radii of the symmetries an appropriate mask can be defined, with the aid of which the solution of this task becomes relatively easy. The symmetry fragments and their radii define an outline that can be used as a mask between frames to aid classification of the coherent fragments in successive frames, as illustrated in Figure 2.

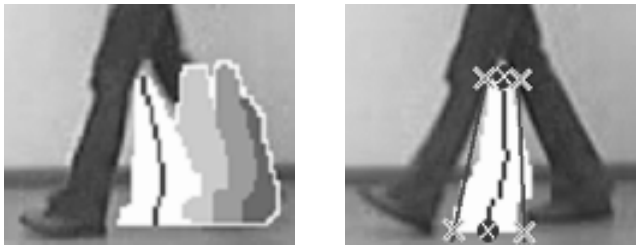

*Figure 2.* Masks of the reconstructed symmetries from successive frames, superimposed on an original image; and the limits (marked by X-symbols) used to define symmetries for the classification task.

The tracking algorithm calculates the overlapping areas between symmetry masks; and as time progresses it constructs the largest overlapping one. The advantage of this simple algorithm is that it is tracking the complete leg movement and the associated structural changes, instead of just tracking selected feature points on the image by means of some optical correlation method. This inherent feature of the method increases the stability and the robustness of the results in cases where the edges of the target are partially “damaged” in some frames. The results of temporal

tracking can be seen in Figure 3, where we demonstrate the resulting symmetry-traces in some real-life situations. With typical pedestrian movement speeds, the tracking algorithm can work correctly when the frame rate is 10 frame/sec or more.

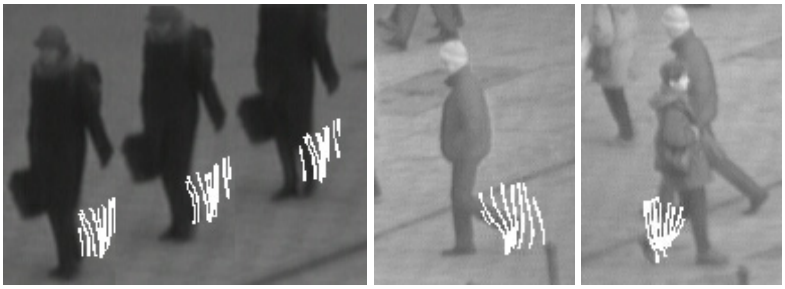

*Figure 3.* Sample symmetry-patterns of real-life pedestrian walking-tracks.

## 2.2 Classification of the Traces

Level 3 symmetries can also appear in other parts of the image, not only between the legs; and the tracking method also collects all of these related symmetries. We can superimpose these traces onto the original images (with information loss where the image-data overlaps), as can be seen in Figure 3. However, although this format is easy to visualise, the detail of these projections is unnecessarily high, which increases the computation time. The other reason why we use an alternative format for the traces is the importance of the radii, which are not very clearly defined in this representation. We therefore reduce the "input space" of the traces by using the following data representation, see Figure 2. One L3S of a frame can be represented by 6 parameters: an upper and a lower position each represented by (x, y) coordinates, and two widths. The coordinates define the positions of the upper and lower ends of the axis (X-symbols in Figure 2), while the widths define the horizontal distance between the two regression lines at the upper and lower points. In fact, the widths correspond to the radii at the upper and lower points of the symmetry axis. This trapezoidal representation can adequately describe both the orientation and the structure properties of the L3S. We collect L3Ss from 8 sequential frames, and thereby have in all 48 parameters for each trace. In the last step of this stage of the process these parameters, including the widths, are normalised in both x and y dimensions; by means of this normalisation the classification becomes invariant with respect to the object size in the image.

In our classification method we used kernel-based (non-linear) Fisher discriminant analysis (Mika et al., 1999). The original linear FDA method classifies two sample sets by maximising the between-class scatter while minimising the intra-class scatter of the features. The aim is to find linear projections that optimise the distinctiveness of the classes. When the problem is not linearly separable however, the solution given by FDA may not be satisfactory. In our case, using the linear method we find that we can achieve a 10% error rate, but with a false-positive rate of 8%, which is rather high. In the non-linear extension of the method, we tested several kernel functions, and concluded that only the Gaussian radial and the inverse multiquadratic kernels produced acceptable classification rates. Practical test results are summarised below.

## 3. EXPERIMENTAL RESULTS

To evaluate the proposed method, we derived "walking" and "non-walking" traces from a considerable number of real-life outdoor video sequences representing a variety of different walk directions, viewing distances and surrounding situations. There were in all 1000 samples, and according to our manual classification these comprised 300 "walking" and 700 "non-walking" samples. In the experiments our main goal was to reliably detect human movements, but at the same time with a false-positive detection rate as small as possible.

Before considering the numerical results, we summarise some practical limitations of the symmetry-tracking method which we noted. The L3Ss can be evolved only if the leg-opening is visible. In our tests we found that this meant that the direction of movement had to be at more than about 70° from the viewing axis; but this is not a serious limitation when more than one camera is monitoring the area (Szlávik, Havasi and Szirányi, 2004). Crowds, and some other specific "overlap" situations are the main cases which cause problems, although "overlap" does not always prevent successful tracking. The most common problematic cases were as follows: subject wearing long coat; subject carrying large bag etc. in the hand nearest to the camera; full masking of the legs by another person in the perspective view; partial masking by another person moving on a parallel track, with synchronised step periods. All in all, the proportion of such "problem" cases in the processed real-life video sequences was some 15%.

For training the KFDA algorithm, we used 50 "walk" and 100 "non-walk" traces representing all situations of the data set. In our evaluation we found that both types of kernel function can achieve a 96% classification rate. At the same time we chose kernel parameters at points where the false-

positive detection rate is zero, to keep the false detection error rate to a minimum. The good detection rates achieved confirm the power of the data representation introduced in Section 2.2. The final choice between the two kernel functions can be based on analysis of the between-class distances, and using this criterion we chose the Gaussian kernel function. In practice we achieved an approximate 96.5% classification rate, with a 1.6% false-positive detection rate.

## 4. CONCLUSIONS

The method we describe can detect pedestrians in image-sequences obtained in outdoor conditions in real-time. Considering even a single step-period, a very low false-detection rate is obtainable. To achieve this, we used a novel feature-extraction and tracking method that can reflect the natural structural changes of human leg-shape; the method seems promising for the purpose of providing a useful "understanding" of image-content. Through experiments using a data set derived from real-life video sequences we found that the Gaussian kernel function is a good choice for the classification of traces. The low classification error-rate achieved demonstrates the power of our spatio-temporal data-representation method. The method appears suitable for the detection of human activity in images captured by video surveillance systems such as those typically used in public places.

## 5. REFERENCES

Abdelkader, C., Cutler, R., and Davis, L., 2002, Motion-based recognition of people in eigen-gait space, *Proc. of the 5th Int. Conference on Automatic Face and Gesture Recognition*

Hayfron, A. J., Nixon, M. S. and Carter, J. N., 2002, *Human identification by spatio-temporal symmetry*, ICPR, 632-635

Kass, M., Witkin, A., and Terzopoulos, D., 1988, Snakes active contour models, *International Journal of Computer Vision*, 321-331

Mika, S., Rätsch, G., Weston, J., Schölkopf, B., and Müller, K.-R., 1999, Fisher Discriminant Analysis With Kernels. *Neural Networks for Signal Processing IX*, 41-48

Mohan, A., Papageorgiou, C., and Poggio, T., 2001, Example-based object detection in images by components, *IEEE Trans. PAMI*, 23(4), 349-361

Nguyen, H. T., Worring, M., and Dev., A., 2000, Detection of moving objects in video using a robust motion similarity measure, *IEEE Trans. on Image Processing*, 9(1), 137-141

Sharvit, D., Chan J., Tek H. and Kimia B.B., 1988, Symmetry-based indexing of image data-bases, *J. Visual Comm. And Image Representation*, vol. 9 no. 4, 366-380

Song, Y., Goncalves, L., and Perona, P., 2003, Unsupervised learning of human motion, *IEEE Trans. PAMI*, Vol. 25, 814-828

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