

Dynamic Texture Based Gait Recognition

Vili Kellokumpu¹, Guoying Zhao¹, Stan Z. Li², and Matti Pietikäinen¹

¹ Machine Vision Group

University of Oulu, P.O. Box 4500, Finland

{kello,gyzhao,mkp}@ee.oulu.fi

² Center for Biometrics and Security Research &

National Laboratory of Pattern Recognition,

Institute of Automation, Chinese Academy of Sciences

95 Zhongguancun Donglu, Beijing 100080, China

szli@nlpr.ia.ac.cn

Abstract. We present a novel approach for human gait recognition that inherently combines appearance and motion. Dynamic texture descriptors, Local Binary Patterns from Three Orthogonal Planes (LBP-TOP), are used to describe human gait in a spatiotemporal way. We also propose a new coding of multiresolution uniform Local Binary Patterns and use it in the construction of spatiotemporal LBP histograms. We show the suitability of the representation for gait recognition and test our method on a popular CMU MoBo dataset. We then compare our result to the state of the art methods.

Keywords: Gait recognition, LBP-TOP, multiresolution LBP, uniform patterns.

1 Introduction

The visual analysis of human movements has become an important area of research in computer vision. It has gained a lot of interest because of its many application domains, such as video indexing, surveillance, human computer interaction, sport video analysis, intelligent environments and gait recognition. Especially gait recognition has recently become a topic of great interest. Compared to other biometrics, gait has an advantage as it can be perceived from a distance and it does not require contact or subject cooperation. Furthermore, gait is difficult to hide or replicate.

Many approaches for human gait recognition have been proposed in the literature [1,2]. Typical approaches are to use either motion [3] or shape information [4,5,6]. However, it may not be a good idea to rely on a single modality. For example, shape based methods are naturally sensitive to changes in appearance like carrying a backpack, whereas motion based methods could have trouble in cases where the pace of walking changes.

Interestingly, very good performance in another human motion analysis application area, human action recognition, has recently been reported by using spatiotemporal analysis that combines both motion and shape cues [7,8]. There has been interest to spatiotemporal analysis in gait recognition as well. BenAbdelkaber et al. [9] used Self similarity plots for gait recognition. Another spatiotemporal pattern proposed for

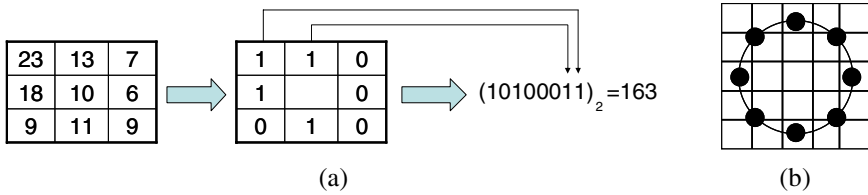


Fig. 1. (a) Illustration of basic LBP operator (b) Circular (8,2) neighborhood. If the sampling point is not in the center of a pixel, the value at that point is bilinearly interpolated from the nearest pixels.

gait recognition is the frieze pattern [10], which was developed into a Shape Variance-Based frieze pattern [11] with good results. Niyogi and Adelson [12] used xyt volume to find out the snakes pattern and then to detect walking people from sequence images. Bounding contours for the walker are recovered and used to define a simple stick figure model, from which several features were extracted and used for individual recognition. Huang and Wang [13] made the recognition based on xt plane energy images extracted from xyt space. Kobayashi and Otsu [14] used Cubic Higher-order Local Autocorrelation to action and person identification.

Dynamic textures are textures with motion, so they inherently combine both appearance and motion. In this paper, we propose a new method for gait recognition that uses dynamic textures for movement description. The method is based on using a dynamic texture descriptor, Local Binary Patterns from Three Orthogonal Planes (LBP-TOP), to represent human movements in a spatiotemporal way. The LBP-TOP features have successfully been used for facial expression [15], visual speech [16] and action recognition [7]. Furthermore, we propose a novel idea for constructing multiresolution LBP histograms and show that it can improve gait recognition performance compared to traditional multiresolution analysis.

The rest of the paper is organized as follows. Section 2 introduces the LBP-TOP features and describes a new method for creating a multiresolution uniform LBP histogram. We show experiment on the CMU database in Section 3 and conclude in Section 4.

2 Gait Description with Spatiotemporal Local Binary Patterns

Dynamic textures are textures with motion. In this section we introduce the LBP-TOP features, which are originally proposed to describe dynamic textures, and propose a novel method for constructing multiresolution uniform LBP histogram. We also show how these can be applied to gait recognition.

2.1 LBP-TOP Features

LBP operator [17] produces a binary code that describes the local texture pattern, which is built by thresholding a neighborhood of pixels by the grey value of its center pixel. The original LBP operator represents a 3×3 pixel neighborhood as a binary number. Figure 1(a) illustrates the basic LBP operator. When LBP operator is applied to an image, the image texture can be described with a histogram of the binary codes.

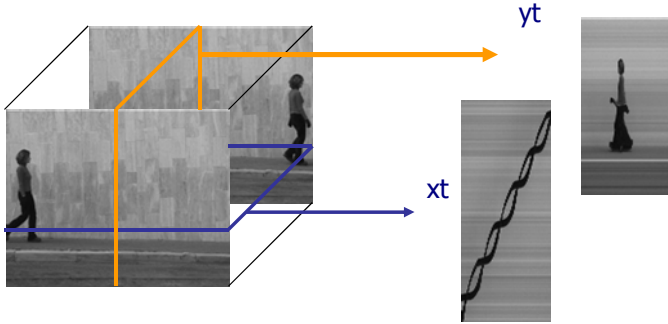


Fig. 2. Illustration of a person walking and the corresponding xt and yt planes from a single row and column. The different frames correspond to the xy planes.

The LBP operator has also been extended to different kinds of neighborhoods. With a circular neighborhood and bilinear interpolation of pixels, any radius and number of sampling points in the neighborhood can be used:

$$LBP(x_c, y_c) = \sum_p^{P-1} s(g_p - g_c) 2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (1)$$

where g_c is the gray value of the center pixel (x_c, y_c) and g_p are the gray values at the P sampling points. Figure 1(b) shows an example of the circular (8,2) neighborhood that has 8 sampling points and radius of 2.

The LBP operator was extended to a dynamic texture operator by Zhao and Pietikäinen [15], who proposed to form their dynamic LBP description from three orthogonal planes (LBP-TOP) of a space time volume. Figure 2 shows the spatiotemporal volume of a person walking from left to right. It also illustrates the resulting xt and yt planes from a single row of and column of the volume as well as the first and last xy planes that are the frames themselves. The LBP-TOP description is formed by calculating the LBP features from the planes and concatenating the histograms.

The original LBP operator was based on a circular sampling pattern but different neighbourhoods can also be used. Zhao and Pietikäinen proposed to use elliptic sampling for the temporal planes: $(x_c - R_x \sin(2\pi p/P_{xt}), y_c, t_c - R_t \cos(2\pi p/P_{xt}))$ for xt plane and similarly $(x_c, y_c - R_y \sin(2\pi p/P_{yt}), t_c - R_t \cos(2\pi p/P_{yt}))$ for yt plane. R_d is the radius of the ellipse to direction of the axis d (x, y or t). As the xy encodes only the appearance, i.e., both axes have the same meaning, circular sampling is suitable. The values g_p for points that do not fall on pixels are estimated using bilinear interpolation.

2.2 Multiresolution Uniform LBP

Multiple kernels can be used together to improve the LBP description. Traditionally, LBP histograms with different radii are concatenated to form the final histogram, as using a joint histogram is infeasible because of the resulting histogram size. However, a lot of information about the underlying structure is lost when only marginal distributions are used to model the texture statistics.

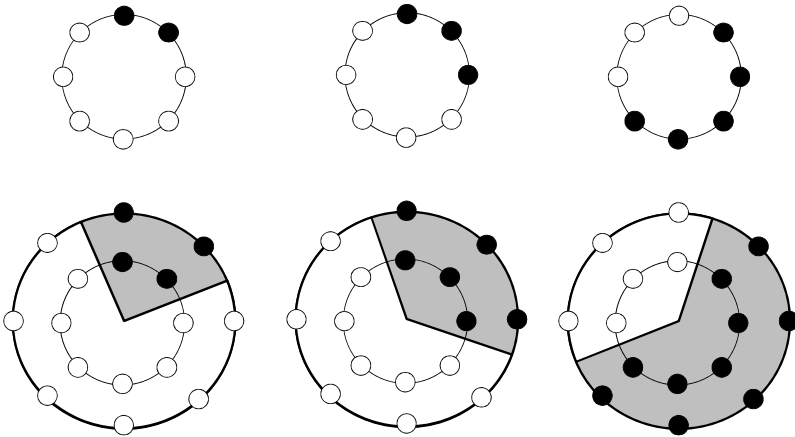


Fig. 3. Illustration of three uniform patterns (top row) and their corresponding multiresolution versions

One way to improve the performance of LBP based methods and shorten the feature histogram is to use the “uniform” subset of LBP codes. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular [17]. It is observed that there are a limited number of transitions or discontinuities in the circular presentation of the 3×3 texture patterns; according to Ojala, these uniform patterns are fundamental properties of local image texture and they provide a vast majority amount all patterns.

By the definition, uniform patterns are codes that consist of at most two homogeneous regions on a circle. From this definition, they can be thought of as representing sectors on a sampling circle. From this insight a new way of calculating multiresolution LBP representation efficiently can be derived.

Figure 3 illustrates a few uniform patterns and their corresponding multiresolution versions. It can be seen that if the multiresolution sampling points are ordered according to the sampling angle, they will also produce codes that satisfy the bit transition condition. It should also be noted that any number of sampling points can be used on different LBP kernels.

We will not give formulation of the new multiresolution uniform patterns here, but we describe the basic idea and show its applicability to the gait recognition problem. Extensive experiments need be done to verify its generalization capability to different problems. The method does, however, have a lot of potential to increase the performance of the many LBP based methods.

In our preliminary setup, sampling points are ordered by going clockwise around the sampling circle. As the representation of uniform patterns is circular, sampling can be started from any point. Figure 4 illustrates the idea in a case where two eight point sampling kernels are used. It is easy to see that the new method preserves the structural information better than the original method. Also the length of the histogram does not increase too much. For two eight point sampling kernels the histogram length is the same as the regular uniform pattern histogram for a sixteen bit kernel, i.e., 243. If an eight and a sixteen point kernels were used, the histogram length would be 555, the same as for a regular 24bit uniform pattern histogram.

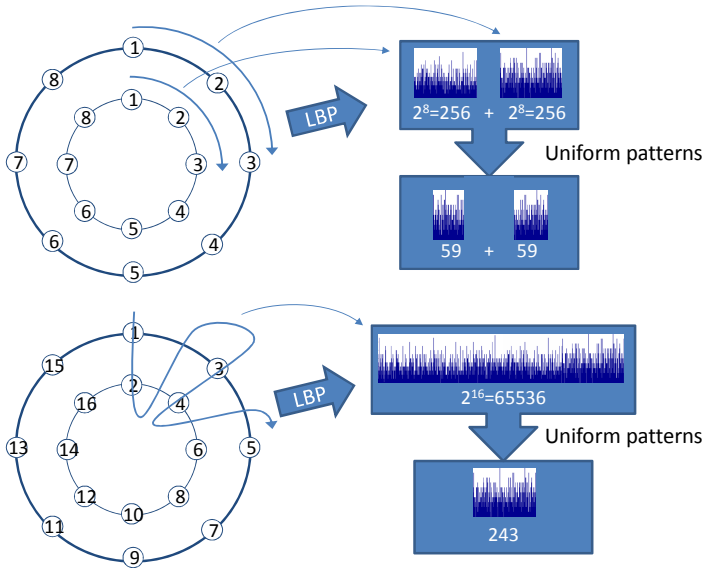


Fig. 4. Illustration of the sampling and formation of the feature histogram. The top image represents the traditional histogram formation where the different kernels are sampled individually. The lower image shows the proposed method. Note the arrows and numbering of the sampling points illustrating the sampling order.

2.3 Gait Description

The dynamic LBP features calculated over the whole xyt volume encode the local properties of the movements without any information about their local or temporal locations. For this reason the volume is partitioned into subvolumes and the feature histogram is formed by concatenating the subvolume histograms. Using the subvolume representation, motion and shape are encoded on three different levels: pixel-level (single bins in the histogram), region-level (subvolume histogram) and global-level (concatenated subvolume histograms).

To obtain a rough spatial definition of human movements, we divide the xyt volume into four regions through the centroid of the silhouette. This division roughly separates the hands and legs of the person. Using more blocks would of course allow a more detailed description but would also produce more local histograms and make the whole histogram longer. Using too many blocks could also possibly make the system too sensitive for natural intraperson variance in the gait.

The subvolume division and the formation of our feature histogram are illustrated in Figure 5. The LBP-TOP features are calculated from the whole duration of a sequence and concatenate all the subvolume histograms on each plane.

Furthermore, as we use silhouette data in the experiments, we prune the histogram by removing the bins that represent monotonous areas (all black/white) from the histogram of each plane. We also observed that removing the collection bin, the bin that represents the non-uniform patterns, improves the performance when uniform patterns are used.

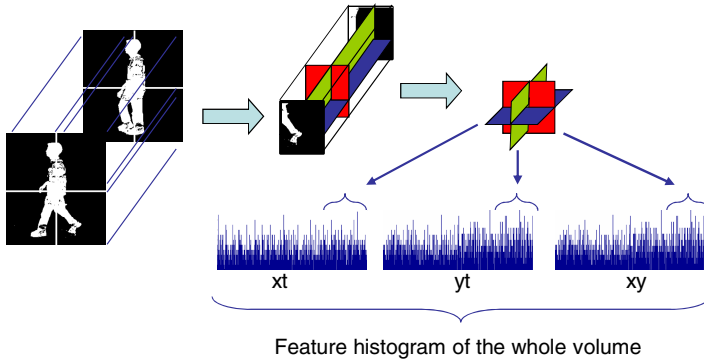


Fig. 5. Formation of the feature histogram

As there is only one appearance plane xy and two temporal planes xt and yt , the motion will have more weight in the concatenated histogram. If histogram of each plane is normalized individually and double weight is given to the appearance plane, no bias is given to either appearance or motion.

Various methods can be used to measure the similarity of LBP-TOP histograms. In this work we have chosen to use the histogram intersection:

$$\text{Similarity} = \sum \min(h_i, h_j) \quad , \quad (2)$$

where the summation is taken over the bins. In classification, an unknown gait example is assigned as being from the same person as the most similar example in the training set.

3 Experiments

We demonstrate the performance of our method by experimenting with the popular CMU MoBo database [18]. We use the LBP-TOP methodology described in the previous section and verify the performance of the new multiresolution uniform patterns on gait recognition.

3.1 CMU Database

To show the suitability of the dynamic texture based approach to gait recognition, we experimented with the CMU database. The reason for choosing this database is that we can easily compare our performance to others described in the literature [9, 11, 19, 20, 21]. Furthermore, this database was chosen because it contains examples of people walking with different speeds. As we are using dynamic textures, we want to see how this change in dynamics affects our performance.

The CMU Moba Database contains 25 different people (23 male, 2 female) walking on a treadmill. Each subject performs walking under four different conditions: walking slow, walking slow holding a ball, walking fast and walking on an inclined surface. Each sequence is recorded at 30 frames per second and is 11 seconds long in duration.

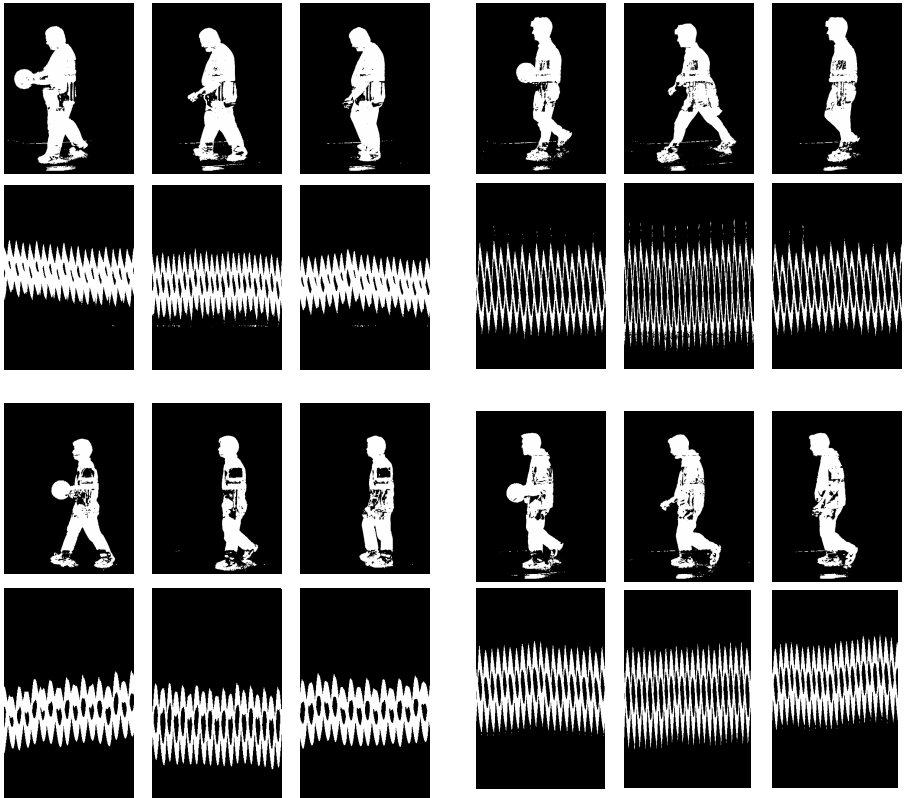


Fig. 6. Illustration of the data in the CMU Moby database under the three different conditions: walking with a ball, walking fast and walking slow. The first silhouette of each sequence is shown together with a single xt slice from whole the duration of the sequence. Notice the change in dynamics and appearance between the conditions.

The database is ideal for testing our method as it contains both appearance (walking and holding a ball) and motion (walking slow vs. fast) changes. The incline walking is excluded from our experiments. Figure 6 gives examples of the database.

3.2 Results

The experiments were run on all the combinations of walking slow (S), walking slow holding a ball (B) and walking fast (F). The test was also performed separately and jointly for appearance and motion features. Table 1 summarizes our results using two LBP kernels (LBP-TOP with $R_x=R_y=2$, $R_t=5$, $P_x=P_y=P_t=8$, and with $R_x=R_y=4$, $R_t=6$, $P_x=P_y=P_t=8$) with the traditional and the proposed multiresolution approaches.

It can be seen that results using just appearance are better than using just motion. Interestingly, the proposed multiresolution method works better on appearance, whereas the original one is better on the temporal planes. When all planes are used together, the new proposed multiresolution uniform LBP performs slightly better.

Table 1. The recognition results on the CMU database. Tests were performed using appearance and motion features separately and jointly. The use of uniform patterns is marked with u2.

Appearance = XY		Hist	B/S	S/B	B/F	F/B	S/F	F/S
Original Multiresolution	u2	228	75 %	79 %	63 %	79 %	76 %	76 %
	-	1020	79 %	79 %	67 %	79 %	76 %	80 %
New Multiresolution	u2	964	71 %	83 %	75 %	79 %	88 %	84 %
	-	262136	79 %	83 %	79 %	83 %	84 %	84 %
Motion = XY+XT								
Original Multiresolution	u2	456	63 %	83 %	54 %	58 %	68 %	56 %
	-	2040	88 %	88 %	33 %	42 %	48 %	40 %
New Multiresolution	u2	1928	67 %	75 %	50 %	54 %	56 %	56 %
	-	524272	96 %	88 %	33 %	42 %	56 %	44 %
Appearance 50% + motion 50%								
Original Multiresolution	u2	684	75 %	83 %	71 %	79 %	80 %	80 %
	-	3060	83 %	88 %	63 %	63 %	76 %	64 %
New Multiresolution	u2	2892	79 %	88 %	75 %	79 %	80 %	76 %
	-	786408	88 %	92 %	67 %	79 %	80 %	72 %
Appearance 65% + motion 35%								
XY = new multiresolution	u2	1420	75 %	83 %	75 %	83 %	88 %	88 %
XT+YT = original multiresolution	u2							

Table 2. Results reported in the literature for the CMU database

	S/B	B/S	F/B	B/F	S/F	F/S
CMU [19]	92 %	-	-	-	76 %	-
UMD [20]	48 %	68 %	48 %	48 %	80 %	84 %
MIT [21]	50 %	-	-	-	64 %	-
SSP [9]	-	-	-	-	54 %	32 %
SVB frieze [11]	77 %	89 %	61 %	73 %	82 %	80 %
LBP-TOP	75 %	83 %	75 %	83 %	88 %	88 %

Another interesting thing to note is that when uniform patterns are not used, the new sampling method is always better. This indicates that the more detailed structure information is important and kept by the proposed method. The downside is of course the huge histogram size, but with the proposed multiresolution uniform coding the histogram can be shortened to a reasonable size. The new coding of multiresolution LBP and uniform patterns are definitely worth further investigating.

It can also be seen that using just appearance gives better result when appearance and motion are given equal weight. The best overall result is obtained when the new multiresolution sampling is used without the uniform patterns, though this comes with a cost of a very long feature vector (262136 bins). We then gave appearance more weight (65%) and observed that by taking a combination of the new multiresolution uniform method on xy plane and the original multiresolution method on xt and yt planes, we can match the result of using just appearance with a feature vector that is only 0.5% in length (1420 bins). Also, the performance is better than using the new multiresolution uniform method on xy alone. It can be deduced that using both the

appearance and motion is important, though it seems appearance should be given more weight. We are currently working on boosting methods to further improve the results and learn the importance of different features.

The test results are encouraging as in all test combinations exists variation in either shape or motion, or both. The best recognition results are achieved on the F/S and S/F scenarios. In these cases the appearance of the person is roughly the same, though there is variation in the pace of walking. In the other scenarios the recognition result is slightly lower due to the change in appearance. Overall, our result is very good. Table 2 shows that our results are comparable to the state of the art methods reported in the literature. It can be seen that compared to others, we achieve a very good result in the most difficult scenarios F/B and B/F where there is variation in both appearance and dynamics.

4 Conclusions and Future Work

We have proposed a novel dynamic texture based method for human gait recognition. We extract LBP-TOP features in spatiotemporal space and use them to describe human walking movements. The method is computationally simple. We show excellent results on a popular benchmark database.

We have also proposed a new way of constructing multiresolution uniform LBP and we show slightly better results on gait recognition compared to the traditional multiresolution analysis. The method has potential to increase the performance of various LBP based applications. Our current work focuses on more exact formulation of the method and extensive analysis on real image data.

As the used LBP-TOP features can also be calculated from the frame data without segmentation, we are also currently investigating gait recognition from image data rather than silhouettes. This extension makes the method more suitable for many applications.

Acknowledgements

This work was supported by European Regional Development Fund, the Finnish Funding Agency for Technology and Innovation, the Graduate School in Electronics, Telecommunication and Automation (GETA), and the Academy of Finland.

References

- [1] Nixon, M.S., Tan, T., Chellappa, R.: Human Identification Based on Gait. Springer, Heidelberg (2005)
- [2] Boyd, J., Little, J.: Biometric Gait Recognition. In: Tistarelli, M., Bigun, J., Grosso, E. (eds.) *Advanced Studies in Biometrics*. LNCS, vol. 3161, pp. 19–42. Springer, Heidelberg (2005)
- [3] BenAbdelkader, C., Cutler, R., Davis, L.: Motion based recognition of people in eigengait space. In: *Proc. FG*, pp. 267 – 272 (2002)
- [4] Veres, G.V., Gordon, L., Carter, J.N., Nixon, M.S.: What image information is important in silhouette based gait recognition? In: *Proc. CVPR*, vol. 2, pp. 776–782 (2004)

- [5] Liu, Z., Sarkar, S.: Improved Gait Recognition by Dynamics Normalization. In: PAMI, vol. 2, pp. 863–876 (2006)
- [6] Zhao, G., Cui, L., Li, H.: Gait Recognition Using Fractal Scale. PAA 10(3), 235–246 (2007)
- [7] Kellokumpu, V., Zhao, G., Pietikäinen, M.: Human activity recognition using a dynamic texture based method. In: Proc. BMVC, 10 p. (2008)
- [8] Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R.: Actions as Space-Time Shapes. In: Proc ICCV, vol. 2, pp. 1395–1402 (2005)
- [9] BenAbdelkaber, C., Cutler, R.G., Davis, L.S.: Gait Recognition Using Image Self-Similarity. EURASIP Journal on Applied Signal Processing 24(4), 572–585 (2004)
- [10] Liu, Y., Collins, R., Tsin, Y.: Gait Sequence Analysis Using Frieze Patterns. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2351, pp. 657–671. Springer, Heidelberg (2002)
- [11] Lee, S., Liu, X., Collins, R.: Shape Variation-Based Frieze Pattern for Robust Gait Recognition. In: Proc CVPR 2007, 8 p. (2007)
- [12] Niyogi, S.A., Adelson, E.H.: Analyzing and Recognizing Walking Figures in XYT. In: Proc. CVPR, pp. 469–474 (1994)
- [13] Huang, G., Wang, Y.: Human Gait Recognition Based on X-T Plane Energy Images. In: Proc. ICWAPR 2007, pp. 1134–1138 (2007)
- [14] Kobayashi, T., Otsu, N.: Action and Simultaneous Multiple-Person Identification Using Cubic Higher-order Auto-Correlation. In: Proc. ICPR, vol. 4, pp. 741–744 (2004)
- [15] Zhao, G., Pietikäinen, M.: Dynamic Texture Recognition Using Local Binary Patterns With an Application to Facial Expressions. PAMI 29(6), 915–928 (2007)
- [16] Zhao, G., Pietikäinen, M., Hadid, A.: Local Spatiotemporal Descriptors for Visual Recognition of Spoken Phrases. In: Proc. HCM 2007, pp. 57–65 (2007)
- [17] Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. PAMI 24(7), 971–987 (2002)
- [18] Gross, R., Shi, J.: The CMU Motion of body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University (2001)
- [19] Collins, R., Gross, R., Shi, J.: Silhouette Based Human Identification from Body Shape and Gait. In: International Conference on Face and Gesture, pp. 366–371 (2002)
- [20] Veerarghavan, A., Chowdhury, A.R., Chellappa, R.: Role of Shape and Kinematics in Human Movement Analysis. In: Proc. CVPR, vol. 1, pp. 730–737 (2004)
- [21] Lee, L., Grimson, W.: Gait Analysis for Recognition and Classification. In: International Conference Automatic Face and Gesture, pp. 155–162 (2002)