Part Based Recognition of Pedestrians Using Multiple Features and Random Forests

Gladis S. John
Curtin University of Technology
Bentley, Perth, WA - 6102
gladis.john@postgrad.curtin.edu.au

Geoff A.W. West
Curtin University of Technology
Bentley, Perth, WA - 6102
g.west@curtin.edu.au

Mihai Lazarescu
Curtin University of Technology
Bentley, Perth, WA - 6102
M.Lazarescu@curtin.edu.au

Abstract

This paper explores a discriminative part-based approach for recognising people in video. It uses many regions to model the background and foreground and a random forest for classification. The objective is to overcome the limitations of more holistic approaches that try to recognise people as a single region with the consequential need to segment each person as one representation. Attributes of each blob, their relationships and variation over video frames are argued to be useful features for discrimination. In this paper the attributes of each blob are considered as a first step in the recognition process. We evaluate our approach through a comparison of three state of the art classifiers: Bagging, Adaboost and a Multilayer Perceptron (MLP), with the Random Forest (RF) using 10 fold cross validation. A detailed statistical analysis shows that the random forest classifier is more accurate compared to the other methods in terms of discrimination between regions describing people and those of the background.

1. Introduction

Variations in human body dimensions, appearance, articulation, and environmental factors such as lighting conditions and camera viewpoint makes pedestrian classification challenging. Popular methods use background subtraction [28] for initial detection using assumptions that do not work under all circumstances. To overcome some of the issues, current methods focus on approaches such as bottom-up [22] and part based methods [21]. More recently, a discriminative approach to part detection has been proposed [7] using sample patches from images, categorizing them into feature vectors and providing discrete labels for classification [18]. The motivation of our work comes from the claim that better learning and good classification can only be obtained by including many different descriptors to account for shape variations and occlusions of the human body [10]. Shape analysis plays a vital role in pedestrian classification tasks and substantial work has been carried out in defining essential attributes of shape as feature vectors for classification [29]. Regions detected in images contain perceptually important information and are reasonably robust to noise thus making them powerful image features. Common region-based methods use moment descriptors to describe shape. Hu moments provide a description of shape that is invariant to translation, scale and rotation. Concise image features can also be extracted from the spectral domain and the effect of noise reduced by rejecting high frequencies. One such feature is the Fourier descriptor that has been popular in the computer vision field for many applications. Many machine learning techniques have been demonstrated to efficiently categorize such image features and better results can be achieved using combined methods. One such method is Random Forests that consists of multiple independently learning random decision trees that produce very low error rates in multi-class problems while maintaining high computational efficiency. Further, they are quite robust to labeling noise that is unavoidable in segmentation problems.

In our research, we are exploring a three-stage approach to classification in videos: (1) using individual regions in each image, (2) their relationships in each image, and (3) their relationships between adjacent images in the video. Initial observation of region growing over video shows good stability in region detection. Each stage will use machine learning to increasingly discriminate between the objects of interest (in this case pedestrians) and their background. The hypothesis is that combinations of regions is rich enough to separate pedestrians from other objects including the background. This paper concentrates on stage (1). In our approach, we use seeded region growing to extract regions and classify those regions as belonging to background or foreground using the Random Forest classifier. Features describing foreground and background regions are extracted using boundary-based Fourier descriptors and region based Hu moments from ground truth images. The labelled foreground and background regions are then classified using the random forest classifier. We evaluate the method on an extensive dataset that includes thousands of training samples under various conditions in an indoor environment. We compare the results with other well known classifiers: Bagging, Adaboost and the Multilayer Perceptron. Results show that random forests give better results than the other classifiers and show the advantages of using multiple features for pedestrian description. In this

978-0-7695-4271-3/10 $26.00 © 2010 IEEE
DOI 10.1109/DICTA.2010.68

paper, a brief overview of the literature will be discussed in the next section. Image modelling and feature extraction are described in section 3 followed by experimental results in section 4. Finally conclusions and future recommendation are in section 5.

2. Related Work

Classifying people in images using analytical descriptions is very challenging due to the non-rigid nature and variation in appearance of different people, the dynamic changes in natural scenes, illumination changes and the presence of other moving objects. It is argued [34] that an object shape model has more advantages compared to other methods like blob detection as it is less sensitive to noise and lower-level processing parameters. A model is required because occlusion and poor segmentation means only part of the person may be visible or detected and hence classification has to occur using a partial set of features [10]. Shape description is one of the key parts of image content description for image retrieval. Shape analysis plays a vital role in many computer vision problems such as recognition, matching and registration and is desired to be invariant to translation, scale and rotation [29]. Various features describing shape have been developed. A taxonomy of shape descriptors depending on different criteria is proposed by [20].

Fourier descriptors have been used to describe shape since the 1970s. Fourier descriptors have been used for plane closed curves [31] and for the identification of three dimensional objects. Elliptical descriptors have been proposed [11] to represent a shape by a set of ellipses. Elliptical descriptors have been used to perceptually group surfaces of revolution [26] and a novel approach for human silhouette recognition based on Fourier descriptors used 40 normalised descriptors and a nearest centroid classifier [25]. They achieved a recognition rate of 97% when tested on real images of humans. Multiscale Fourier descriptors have been used in shape classification [12], shape-based image retrieval [13] and defect image retrieval [14]. A system has been proposed for content-based image retrieval using Fourier descriptors on a logo database [6]. Acceptable results were produced for both classification and abstraction queries.

Analysing shape in the spectral domain instead of the spatial domain overcomes the common problem in digital images of noise. Additionally, the spectral features of a region are usually more concise than spatial features. However, contour-based shape descriptors exploit only boundary information, and may not be able to deal with disjoint shapes where good boundary information is not available, for example under occlusion [33]. Region based techniques use all the pixel information to describe shape and can be used when the information is only partially available. Common region-based methods use moment descriptors to describe shape. These include Hu moments, Legendre moments, Zernike moments and pseudo-Zernike moments. The Hu moments derived by [9] are invariant to rotation scaling and translation and have been widely used to describe geometric characteristics of objects in pattern recognition.

A robust video marking scheme based on Zernike and Hu moments has been proposed [19]. Extracting these descriptors is fast and hence have been used in robotic applications in the area of real time vision [27]. Standard moments have been used to identify three-dimensional objects from 2D images [23]. Recently a novel classification scheme has been proposed representing a modified distance transform as moment invariants using partial object information [17]. The method has the potential to handle changes in illumination, pose and inter-class and intra-class variations.

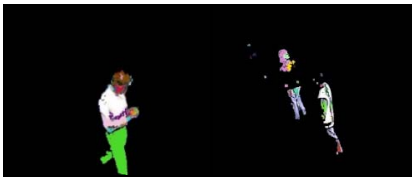
Many approaches compare pixel colors in a frame to a learned stationary background model for detecting motion blobs. This method works under constrained environments [35]. Under such circumstances, modeling the foreground *as well as the* background and using a machine learning approach should improve classification. Many machine learning processes have demonstrated good classification of image categories using image features [18] and ensembles of randomly created clustering trees. These trees are fast to train, test and robust to background clutter. Random forests were introduced by [4] and gained much interest in the computer vision field due to their simplicity, speed and performance [15]. Random forests have been built [5] based on image patches to automatically extract an object from video sequences. Images have been classified by the object categories they contain using random forests [3]. This demonstrated that using a random forest classifier significantly reduces training and testing costs compared to classifiers such as multi-way SVM without reducing performance.

A more recent approach to classification is to model the background using keypoints to extract the foreground [8]. Good results were obtained using Adaboost to classify the foreground. The work of Hough forests for object classification [7] is of significance as models of the background and foreground use patches producing good classification results. This paper differs from this approach in that it carries out discriminative region based learning of the foreground and background. Additionally we learn these regions under different conditions and occlusions. Our contribution is a technique that works with real world data with minimal assumptions and is computationally inexpensive.

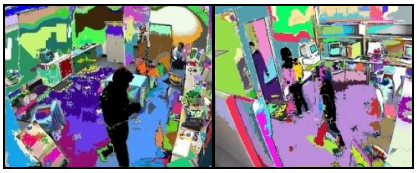
3. Feature Extraction and Image Modelling

We treat each video frame as consisting of a set of background and foreground regions generated by region growing. The first task is to label or ground truth these regions for classifier training. An efficient adaptive density background subtraction technique [32] was used to extract blobs where each blob represents a human. This process is carefully monitored and parameters chosen to give 100% person detection with a consequent false alarm rate such that the resulting ground truth is an acceptable representation. The result is a binary image of people masks overlaid on a background mask. Regions from the region growing process can then be assigned to be human or background regions. This process reduces the tediousness of manually ground truthing many thousands of frames and regions. However one problem with background subtraction methods is that they only extract moving objects as foreground and stationary people could be included in the background regions thereby producing false positives. However, by using a large number of images for training, there are significantly more examples of correct ground truth and only a few incorrect ones. Each video frame was then segmented to extract the foreground and background regions using automatically seeded region growing [1]. The regions corresponding to the people and background were then extracted from these frames by comparing them to the corresponding foreground and background masks. In cases where regions overlap with the foreground or background masks, only regions overlapping the foreground or background mask is considered. A frame and its corresponding foreground and background regions for two examples are shown in figure 1.

(a)



(b)



(c)

Figure 1: (a) Original frames (b) Foreground regions (c) Background regions.

Features describing foreground and background regions were then extracted using region-based Hu moments and boundary-based Fourier descriptors. The pixel coordinates of the contour for each region in the foreground and background were then extracted one and interpolated to get a power of two number of points for Fourier feature extraction. The elliptical Fourier descriptors and Hu moments used are summarised in the following sections.

3.1 Hu Moments

Hu moments are obtained by combining different normalised central moments that represent different aspects of the image that are invariant to scale, rotation and reflection [9].

The central moments are given by:

$$\mu_{p,q} = \sum_{i=0}^{n} I(x,y)(x - x_{\text{avg}})^p (y - y_{\text{avg}})^q$$

The normalised moments are given by:

$$\eta_{p,q} = \frac{\mu_{p,q}}{m_{00}^{(p+q)/2+1}}$$

The seven Hu moments used as features are given by:

$$h_1 = \eta_{20} + \eta_{02}$$

$$h_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$h_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$h_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$h_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)$$

$$h_6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$h_7 = (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)$$

3.2 Elliptical Fourier Analysis (EFA)

An ellipse from a region contour is parameterised by centre location, length of the minor and major axes, and orientation of the major axis. The tilt angle of the ellipse (r) is determined from the major and minor axis values. Elliptical Fourier descriptors suggested by [11] to represent a shape by a set of ellipses, approximate the closed contour as a sum of the Fourier harmonics. Each ellipse has a major and minor axis invariant to translation and rotation. They used n harmonics to describe a closed contour with k points and each harmonic has four Fourier coefficients a_n, b_n, c_n and d_n.

The coefficients are defined by the equations:

$$a_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{\Delta x_i}{\Delta t_i}\left[\cos\frac{2n\pi t_i}{T} - \cos\frac{2n\pi t_{i-1}}{T}\right],$$

$$b_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{\Delta x_i}{\Delta t_i}\left[\sin\frac{2n\pi t_i}{T} - \sin\frac{2n\pi t_{i-1}}{T}\right],$$

$$c_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{\Delta y_i}{\Delta t_i}\left[\cos\frac{2n\pi t_i}{T} - \cos\frac{2n\pi t_{i-1}}{T}\right],$$

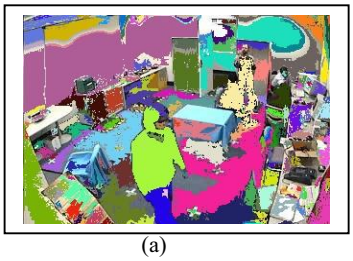
$$d_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{\Delta y_i}{\Delta t_i}\left[\sin\frac{2n\pi t_i}{T} - \sin\frac{2n\pi t_{i-1}}{T}\right].$$

The coefficients a_i and b_i geometrically represent the x-axis projection of the semi major and semi minor axis of the i^{th} harmonic and coefficients c_i and d_i that of y-axis. [16] proved that the complete information of the shape can be described using n Fourier harmonics using $4n$ independent features. It has been shown that there are advantages of this representation for geometric interpretation [30].

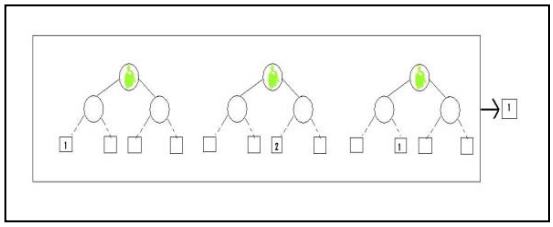
The elliptical parameters such as the major axis, minor axis, magnitude of the real and imaginary axes, the angles between the real and imaginary axes and the angle of rotation were then derived. Thus each region has fourteen Hu and Fourier feature values and together with the labels constitutes the fifteen element feature vector for training.

4. Random Forests

Random forests [4] is a classifier consisting of an ensemble of trees trained with random features. To model the background and foreground regions we generate 10 random binary trees each constructed with different samples from the training data. Each tree node contains the attribute that splits the data to be classified. Each leaf contains the distribution estimate of the classes based on the training data. Each tree votes for one of the classes and the instance is classified with the maximum class vote. An example showing a small part of a random forest classifying an image into region trees is shown in Figure 2. The upper body region of the person just below the middle in Figure 2(a) is classified by three trees and the majority vote is for label '1'.



(a)



(b)

Figure 2: (a) Region grown image
(b) An example region classification

The advantages of Random Forests are:

- They are fast and robust.
- They are computationally inexpensive for training on high dimensional datasets without significant overfitting [7].
- They are good at handling unbalanced datasets.
- They can easily handle multiclass problems and easy to train [15].

5. Experimental Results and Discussion

We experimented on a dataset consisting of 46 videos with an average length of 350 frames, each illustrating a different indoor environment with people moving around under varying lighting conditions. We studied a total of 188,226 foreground and background blobs for training and testing. The feature extraction was carried out using OpenCV and the Visual Studio 2005 development environment on an Intel Core 2 Duo CPU E6850 @ 3.00 GHz, 2.99 GHz, with 1.96 GB of RAM with Windows XP. The evaluation of the classification was carried using the Random Forest, Bagging, Adaboost and Multilayer Perceptron implemented in the Weka 3.6 machine learning tool with the default values of the parameters.

The four classifiers: Random Forest, Bagging, M1Adaboost and Multilayer Perceptron (MLP) were generated using 10 fold cross validation and a detailed analysis of the accuracy was carried out using True positive rate, False positive rate, Precision, Recall, the F-Measure, the area under the ROC, and the confusion matrix. We also compute the kappa statistic, mean absolute error and root mean squared (RMS) error.

Table 1: Detailed accuracy of classifiers

Classifier	Time in secs	Accuracy %	KS	Mean absolute error	RMSerror
Training					
Random Forest	5.83	93.95	0.88	0.09	0.22
MLP	72.45	69.92	0.39	0.38	0.44
Bagging	5.69	93.56	0.87	0.10	0.22
Adaboost	1.56	83.63	0.67	0.34	0.38
Testing					
Random Forest	5.66	99.34	0.99	0.02	0.083
MLP	74.52	67.64	0.38	0.39	0.43
Bagging	5.64	98.49	0.97	0.05	0.12
Adaboost	1.5	91.11	0.82	0.33	0.37

Table 2. Evaluation of various classifiers

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	Class
Random Forest	0.925	0.047	0.947	0.925	0.936	0.979	1
	0.953	0.075	0.933	0.953	0.943	0.979	2
MLP	0.635	0.242	0.707	0.635	0.669	0.763	1
	0.758	0.365	0.694	0.758	0.725	0.763	2
Bagging	0.901	0.033	0.962	0.901	0.93	0.978	1
	0.967	0.099	0.914	0.967	0.94	0.978	2
AdaBoost	0.713	0.051	0.928	0.713	0.807	0.864	1
	0.949	0.287	0.783	0.949	0.858	0.864	2

The results of all the classifiers are summarised in Table 1. The number of trees for the Random Forest classifier was set to ten following some preliminary evaluations. Despite the large volume of data the Random Forest performs the best with an accuracy of 93.95% on the training set and 99% on the test set. The higher accuracy of the test set is due to its statistical similarity with training set. The four classifiers could classify the foreground and background with the Random Forest classifier with a high accuracy of 99% with the test set. The bagging method also performs equally well. Adaboost is not so good, by computationally inexpensive with a trade off in accuracy. The higher values of the kappa statistic and the lower values of root mean square also demonstrate the good classification performance of the Random Forest

Details of the accuracy of all the four classifiers together with their true positive rate, false positive rate, precision, recall, F-measure, and area under ROC are summarised in Table 2. The class value for foreground is denoted by 1 and background by 2. Although there is variation in the results across classifiers and measurement, generally the Random Forest and Bagging classifiers produce the best results overall. A confusion matrix for each of the classifiers is presented in Table 3

Table 3: Confusion Matrices

Classifier		a	b	<-classified a=1;b=2
Random Forest	a	8327	677	1
	b	462	9360	2
MLP	a	5714	3290	1
	b	2373	7449	2
Bagging	a	8113	891	1
	b	321	9501	2
AdaBoost	a	6423	2581	1
	b	500	9322	2

From the confusion matrices it is evident that the Random Forest classifier produces good discrimination between regions of foreground and background giving similar results to Bagging. Adaboost and the MLP were significantly worse. The area under the ROC curve is best for the Random Forest, with the highest true positive rate and precision making it the best classifier for large volumes of data while at the same time being computationally inexpensive.

6. Conclusion

This paper has explored the usefulness of region gowing and region features for discriminating between foreground (in this case people) and background. It recognises that each person or object will be represented by a number of regions, each subject to variation across frames and object. Regions are described using a number of methods and a feature vector of 14 attributes used. Machine learning using the Random Forest technique is used for discrimination and compared to three other popular classifiers. The Random Forest method performs well and shows the power of describing humans and background using a number of quite variable regions. Future work will explore other features from the large numbers that have been proposed in the literature and extend the methodology to consider the relationships and coherence between regions describing humans within each video frame and across video frames.

7. References

[1] Adams,R., & Bischof,L. (1994). Seeded region growing. IEEE Trns. On PAMI, 16(6), 641-647.

[2] Ballaro, B., F. Isgro, et al. (2004). "Silhouette encoding and synthesis using elliptic Fourier descriptors, and applications to videoconferencing." Journal of Visual Languages and Computing: 391-408.

[3] Bosch, A., A. Zisserman, et al. (October, 2007). Image Classification using Random Forests and Ferns. IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil.

[4] Breiman, L. (2001). "Random Forests." Machine Learning 45(1): 5 - 32.

[5] Chen, H. T., T. L. Liu, et al. (2006). Segmenting Highly Articulated Video Objects with Weak-Prior Random Forests. ECCV, Springer-Verlag Berlin Heidelberg.

[6] Folker, A. and H. Samet (2002). Content-based Image Retrieval Using fourier Descriptors on a Logo Database. 16th Int.Conf. on Pattern Recognition, Quebec City, Canada.

[7] Gall, J. and V. Lempitsky (2009). "Class-Specific Hough Forests for Object Detection." IEEE Trans on Pattern Analysis and Machine Intelligence: 1022 - 1029.

[8] Hamdoun, O. and F. Moutarde (2009). PhD forum: Keypoints-based background model and foreground pedestrians extraction for future smart cameras Third ACM/IEEE International Conference on Distributed Smart Cameras, 2009, Paris, France

[9] Hu, M. K. (1962). "Visual pattern recognition by moment invariants." IEEE Trans. Inform. Theory 8: 179-187.

[10] John, G. S., G. West, et al. (June,2009). Measures for the evaluation of segmentation methods used in people tracking. IEEE Int. Conf on Multimedia and Expo, New York,USA.

[11] Kuhl, F. P. and C. R. Giardina (1982). "Elliptic Fourier features of a closed contour." Computer Graphics and Image Processing 18: 236-258.

[12] Kunttu, I., L. Lepisto, et al. (2003). Multiscale Fourier Descriptor for Shape Classification. 12th International Conference on Image Analysis and Processing.

[13] Kunttu, I., L. Lepisto, et al. (2004). Multiscale Fourier Descriptor for Shape-Based Image Retrieval. 17th International Conference on Pattern Recognition (ICPR'04).

[14] Kunttu, I., L. Lepisto, et al. (2006). "Multiscale Fourier descriptors for defect image retrieval." Pattern Recognition Letters 27(2): 123-132.

[15] Lepetit, V. and P. Fua (September, 2006). "Keypoint Recognition using Randomized Trees." IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9): 1465- 1479.

[16] Lin, C. S. and C. L. Hwang (1986). "New Forms of Shape Invariants From Elliptic Fourier Descriptors." Pattern Recognition 20(5): 535-545.

[17] Minhas, R., A. A. Mohammed, et al. (2009). A Generic Moment Invariants Based Supervised Learning Framework for Classification Using Partial Object Information,Canadian Conference on Computer and Robot Vision, British Combia.

[18] Moosman, F., E. Nowak, et al. (2008). "Randomized Clustering Forests for Image Classification." IEEE Trans on Pattern Analysis and Machine Intelligence 30(9): 1632-1646.

[19] Paraskevi, K. T., S. N. Klimis, et al. (2005). HumanVideo Object Watermarking Based on HU Moments. IEEE workshop on Signal Processing Systems, Athens, Greece.

[20] Pavlidis, T. (1978). "A Review of Algorithms for Shape Analysis." Computer Graphics and Image Processing 7: 243-258.

[21] Quattoni, A., Collins, M., & Darrell, T. (2005). Conditional random fields for object recognition. *Advances in Neural Information Processing Systems 17*, 1097-1104.

[22] Ramanan, D., & Forsyth, D. A. (June 2003). *Finding and Tracking People From the Bottom Up.* Paper presented at the Computer Vision and Pattern Recognition (CVPR), Madison, WI. Segmenting highly articulated video objects with weak prior random forests

[23] Reeves, A. P., R. J. Prokop, et al. (1988). "Three-Dimensional Shape Analysis Using Moments and Fourier Descriptors." IEEE Trans on Pattern Analysis and Machine Intelligence 6: 937-943.

[24] Richard, C. W. and H. Hemani (1974). "IdentlficaUon of 3-dimensional objects using Fourier descriptors of the boundary curve." IEEE Trans. Systems, Man and Cybernetics(4): 371-378.

[25] Rocio Diaz de, L. (2000). Human Silhouette Recognition with Fourier Descriptors. 15th Intl. Conf on Pattern Recognition.

[26] Rosin, P. and G. West (1992). "Detection and verification of surfaces of revolution by perceptual grouping." Pattern Recognition Letters 13: 453-461.

[27] Sanz, P. J., R. Marin, et al. (2005). "Including Efficient Object Recognition Capabilities in Online Robots:from a statistical to a Neurl-Network Classifier." IEEE Trans on Systems, Man and Cybernetics, Part C 35: 87-96.

[28] Stauffer, C. and W. Grimson (1999). Adaptive background mixer models for real time tracking. IEEE Int. Conf. on Computer Vision and Pattern Rec.

[29] Veeraraghavan, A., A. K. Roy-Chowdhury, et al. (2005). "Matching Shape Sequences in Videos with Applications in Human Movement Analysis." IEEE Trans on Pattern Analysis and Machine Intelligence 27(12).

[30] Yip, R. K. K. and P. K. S. Tam (1994). "Application of Elliptic Fourier Descriptors to Symmetry Detection Under Parallel Projection." IEEE Trans on Pattern Analysis and Machine Intelligence 16(3): 277-286.

[31] Zahn, C. T. and R. Z. Roskies (1972). "Fourier descriptors for plane closed curves." IEEE Trans. Computers: 269-281.

[32] Zivkovic, Z., & Van Der Heijden, F. (2005). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*.

[33] Zhang, D. and G. Lu (2002). "Shape-based image retrieval using generic Fourier descriptor." Signal Processing: Image Communication: 825-848.

[34] Zhao, T. and R. Nevatia (2004). "Tracking Multiple Humans in Complex Situations." IEEE Trans on Pattern Analysis and Machine Intelligence 26(9): 1208-1221.

[35] Zhao, T., R. Nevatia, et al. (2008). "Segmentation and Tracking of Multiple Humans in Crowded Environments." IEEE Trans on Pattern Analysis and Machine Intelligence 30(7): 1198-1211.