A Segmentation Method of Catadioptric Images for Gait Recognition in Unconstrained Environment

Yohan Dupuis, Xavier Savatier, Jean-Yves Ertaud, Ghaleb Hoblos
Instrumentation, IT and Systems Department
IRSEEM, Research Institute for Embedded Systems
ROUEN, France
{firstname.lastname}@esigelec.fr

Abstract

Gait is an emerging biometric technology. It enables biometric at a distance. The first step in gait recognition is the silhouette extraction. However, most of the work involves indoor controlled environment or well-exposed outdoor scenes. Furthermore, they are all applied to perspective-like pictures. This paper addresses a method for silhouette extraction on catadioptric images in indoor and uncontrolled lighting environments. We introduce a new segmentation method based on the K-means clustering algorithm. This method is robust to the stroboscopic effect induced by the light source. We finally present a local method to obtain perspective-like pictures enabling further processing. Principal Component Analysis (PCA) is usually used for dimensionality reduction of datasets. Most of the time, the geometrically asset of the PCA is unused. In this work, we take advantage of this particular point to propose a local unwrapping technique of catadioptric pictures.

1. Introduction

Robotics and biometrics are nowadays present in our daily life. Iris Recognition is a [1]well-established field. As a morphological feature, it can be circumvent if no proof of life is taken into account. Behavioral biometrics is slowly emerging. The context of this work implies the use of a ground patrol robot. It aims at performing surveillance tasks and, as a result, being able to authenticate the persons presents in its surrounding environment.

The sensor used in this study is known as a catadioptric system. Catadioptric is the combination of catoptrics and dioptrics. These words respectively refer to the mirror and to the lens and as a consequence the camera as a whole. The association of a camera and a mirror are often designed to achieve a 360-degree field of view (FOV) in azimuth [1]. This FOV increases robot's performance for navigation and localization tasks [2]. Even if, the first aim of catadioptric sensor is navigation,

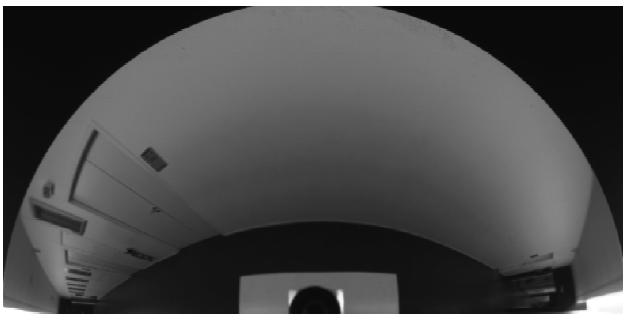

Figure 1. Background Model

we believe we can use it to perform authentication tasks as well. The main advantage of catadioptric sensors over perspective sensors is the ability of continuous tracking and gait analysis it offers. Indeed, unless the user leaves the scene, the sensor FOV increases the period where the gait analysis and recognition can be performed. Moreover, the most promising recognition techniques published so far uses the user profile [3] or views close to this situation [4]. Still, the chance that this situation happens with perspective camera is way lower than with 360-degree azimuth FOV in a random scene. As a consequence, after an exhaustive state-of-the art, we are convinced that catadioptric sensors can fulfill all the assumptions implied in existing gait recognition techniques.

The first step in a biometric system is feature extraction. Silhouette extraction for gait recognition has already been studied [3]. However, most of the work has been done for conventional cameras so far. Post-information extraction also requires standard perspective view of the scene. Furthermore, segmentation is achieved in indoor light-controlled environment [4]. When outdoor scenes are considered, they do not include shadows or other artifacts that might make the segmentation false [3]. In our case, we considered a low and uncontrolled illumination light indoor environment. The most advanced techniques are based on Gaussian Mixture Models (GMM) [5]. GMM is a pixel wise temporal

978-0-7695-4175-4/10 $26.00 © 2010 IEEE
DOI 10.1109/EST.2010.34

technique used to model the background, which works quiet well when the scene illumination changes to certain extend. We will show how and why this method fails in our case.

Finally, the sensor design imposes to consequently close the optic aperture. The hyperbolic mirror parameters used in this study impose a distance separating the camera and hyperbola foci of 73.143 mm. To the best of our knowledge, no manufacturers provide optics, which can give a convenient depth of field at such a small distance. Without highly reducing the aperture size, we would have to favor certain radius of the mirror. The remaining parts would be blurry. As a consequence, the catadioptric system would loose its interest.

The main drawback of such a narrow aperture is the diminution of intensity range in the scene. It results in decreasing the discriminative power of regular segmentation methods. Fig. 1 gives an overview of the study environment.

Previous works of image segmentation for gait recognition in catadioptric images already exists [6]. Besides, the unwrapping process is conducted on the entire picture after segmentation. In order to embed the algorithm achieves real time performances, one needs to remove non-required operations. As a consequence, we believe that it would be interesting to locally unwrap the ROI and not the entire catadioptric picture.

In the first part, we present our segmentation method and different statistical tools used to perform it. A comparison with popular methods is conducted. Secondly, we expose our rectification methods. We finally present our first results regarding feature extraction for gait recognition.

2. Silhouette Extraction

The first stage in silhouette extraction is the foreground segmentation. The statistical background model is built on the n first frames where no foreground objects are supposed to be present. The statistical model can be the mean, the median or a mixture of Gaussians distribution of the n frames. Then, the actual frame is compared to that model. The comparison can be the difference and then a threshold is applied to determine the foreground pixels. When the mixture of Gaussians is used, one can use the probability density function to obtain the probability that a pixel belongs to the background model. However, when the model is built on RGB picture, a model is established for each channel. The resulting problem is how to combine the similarity outcome. To overcome this problem, we adopted an original method. We considered the RGB components of each pixel as a vector. The best way to compare two vectors is their norms ratio and the angle between them.

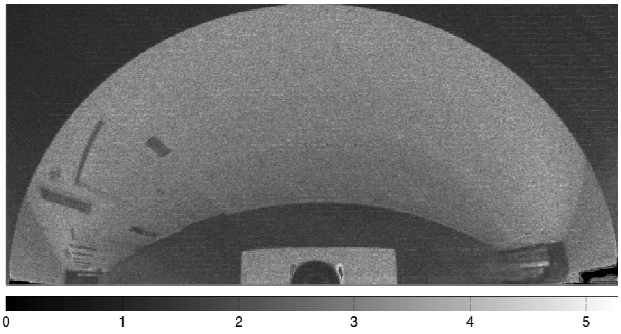

Figure 2. Standard Deviation over two seconds

The cosine similarity is a measure of similarity between two vectors of n dimensions [7]. Let's consider it in context to our work:

$$\cos(M_{i,j}, I_{i,j})_{i,j} = \frac{M_{i,j} \bullet I_{i,j}}{\left\|M_{i,j}\right\| \times \left\|I_{i,j}\right\|} \qquad (1)$$

where :

- $\cos(M_{i,j},I_{i,j})_{i,j}$ is the angular similarity measure at pixel coordinate (i , j).
- $M_{i,j}$ is a 1x3 vector containing the RGB information of the background model at pixel coordinate (i , j).
- $I_{i,j}$ is 1x3 vector containing the RGB information of the current frame at pixel coordinate (i , j).
- $M_{i,j} \bullet I_{i,j}$ represents the dot product.

The outcome is one when the vectors are collinear and zero when they are perpendicular and as a consequence different. This similarity measure ranges between zero and one. Nevertheless, a shadow is comparable to shorten the vector norm. The color components, as a result the vector direction in the RGB color space, remains unchanged. We also considered the length of the vectors:

$$R_{i,j} = \frac{\left\|I_{i,j}\right\|}{\left\|M_{i,j}\right\|} \qquad (2)$$

where $R_{i,j}$ is a ratio similarity measure at pixel coordinate (i , j).

We obtain two similarity features between the actual frame and the background statistical model: angular measures and ratio measures. Still, the problem remains the same: how to combine these measures. We know that two clusters of pixels exist: the background cluster and

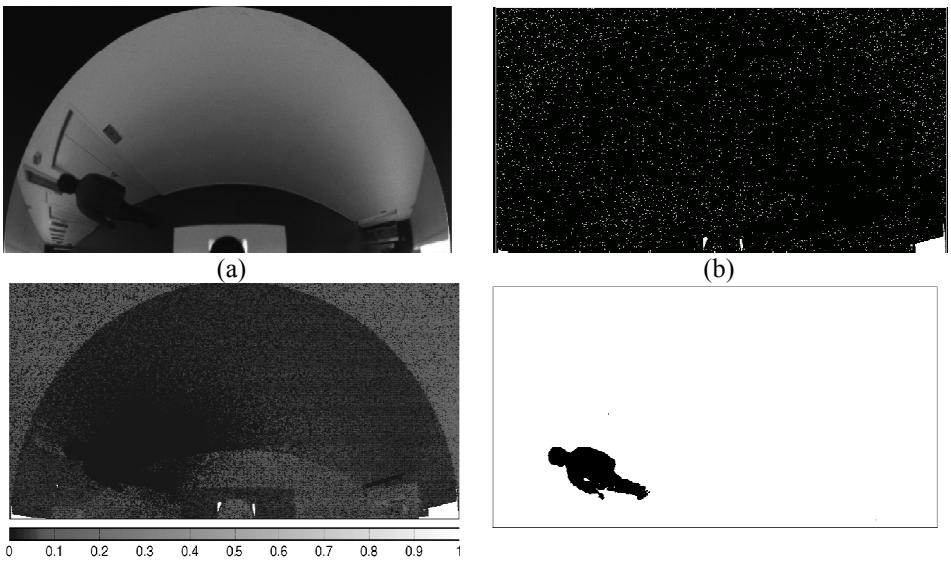

Figure 3. (a) Input picture (b) Direct Background substraction (c) Probability that a pixel belong to the background (d) Our method. Foreground pixels are in black.

the foreground cluster. Moreover, the pixel class is a priori unknown.

Machine learning algorithms are well suited for this kind of application. Unsupervised learning, such as K-means clustering, is one the existing tools appropriate for our problem [7]. The K stands for the number of expected clusters in the pixel distribution. We already know that two clusters exist. The algorithm statistically assigns a class to each pixels in the two dimensional space defined by the similarity measures. Each cluster has a center of mass: a centroid. The algorithm iteratively minimizes the within-cluster sum of squares from the centroids. The centroids are moved up to a threshold displacement. Moreover, we use intra-frame context to fasten the centroid convergence. For the first frame, the centroids are randomly initialized. On the remaining pictures, the centroid position at *t-1* is used to initialize the values of the centroid at *t*. As the camera frame rate is 15 fps, the scene may not have changed too much. This approach enables to reduce from a hundred to two the average number of iterations required to maximize intra-class distance. As the K-means algorithm is sensitive to scale disparity between the different dimensions, we also normalize ratio similarity to fit between [0:1]. K-means algorithm has already been used for image processing [5]. It has notably been used for color clustering [8] as well. To the best of our knowledge, it is the first time that it is used to actually segment foreground element from the rest of the picture.

In our experiment, we simulated an electronic surveillance camera situated in a hallway. As mentioned earlier, our pictures suffer from a fluttering effect caused by the light sources used in the experiment. As a result the pixels vectors will vary a lot during the model construction. Figure 2 shows the average standard deviation during the construction of our model, i.e. the first two seconds of the experiment. We can notice that a major part of the image suffers from a large standard deviation. In order to construct the background model, we used three statistical measures: mean, median and standard deviation. The mean is too sensitive to outliers to be used as a simple model. The median vector is described as the number separating the higher half of a population from the lower half. As a consequence, that statistical measure is more appropriate to solve our purpose. The mean and standard deviation were used to construct the probability density function (PDF) as expressed in the following equation:

$$N(\mu_{i,j,k}, \sigma^2_{i,j,k}) = \frac{1}{\sigma\sqrt{2\Pi}} \times e^{\left(\frac{-(I_{i,j,k}-\mu_{i,j,k})^2}{2\sigma^2_{i,j,k}}\right.} \quad (3)$$

Where :

- $\mu_{i,j,k}$ is the mean computed with the model images at pixel coordinate (i , j) for the channel k.
- $\sigma^2_{i,j,k}$ is the standard deviation computed with the model images at pixel coordinate (i , j) for the channel k.

The channel wise PDF are then combined to obtain the probability of a pixel to be a member of the background as follows:

$$p(I_{i,j} \mid M_{i,j}) = \sum_{k=1}^{3} \frac{1}{3} N(\mu_{i,j,k}, \sigma^2_{i,j,k}) \quad (4)$$

$p(I_{i,j} \mid M_{i,j})$ is close to one when the pixel belongs to the background and close to zero when it is a part of the foreground.

Once the model was established, it was asked to a person to enter the scene. The pictures taken when the person entered were compared to the model established with the statistical tools mentioned earlier. Figure 3 shows the segmentation outcome for three popular methods: (b) direct background substraction, (c) GMM (d) our method. The first thing we can notice is that the simple background substraction (Fig. 3(b)) doesn't work at all. Fig. 3(c) shows an interesting aspect of our approach. As one can see black pixels has a probability close to zero so as to belong to the background. As a consequence, they have to be classified as foreground pixels. The person silhouette falls in this category. However, a complete blob of black pixels also exists on the wall and on the floor. They result from a shadow created by the person. An analysis of the ratio similarity confirms a diminution of the ratio while no changes of the angular measure were observed. This artifact is not present on the image resulting from our method. Furthermore, we can easily recognize the silhouette of the person. We can also observe that we effectively discriminated background and foreground pixel in the most challenging part of the scene: the floor. This aspect is particularly important when we recall that this paper propose a preprocessing technique for gait recognition. It's worth saying that this segmentation technique doesn't require any additional morphological operations. This characteristic is pretty important when we know that the final algorithm has to be embedded on a robot. As a partial conclusion, we can assert that we successfully segment foreground objects from a quite challenging environment. Still, regular gait analysis requires a person standing as one can expect when someone walks. The main way of tackling this question is to globally unwrap the catadioptric using the estimated model parameters. Nevertheless, locally unwrapping the catadioptric image would speed up the process.

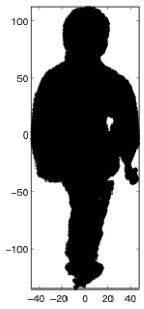

Figure 4. Projection onto the eigenspace

3. Silhouette Rectification

As we can notice in Fig. 3, the silhouette is tilted. Traditional image processing implies that the person silhouette is standing in a vertical manner. The picture unwrapping is a straightforward method. However, the process is still computationally expensive. Moreover, complete unwrapping is required when one wants to perform operation requiring perspective-like image. We already performed the segmentation on the catadioptric image itself. We consequently know where our Region of Interest (ROI) is located. In order to reduce the computational cost, we tried to work directly on the ROI. A statistical tool, the Principal Component Analysis allows projecting the data onto a new space. The PCA allows a structural analysis of the data. It is often used for dimensionality reduction. To the best of our knowledge, it hasn't been used for rectification of catadioptric images. The PCA is actually based on the SVD of the covariance matrix. The first principal component is actually the eigenvector associated with the largest eigen value of the covariance matrix. The principal components accounts for as much of the variability in the data as possible. The first component captures the maximum variability. The succeeding components take into account the remaining variability. Once the principal components are defined, the cluster is projected onto the new orthogonal base. A closer look at the foreground cluster of points may suggest that the person height may correspond to the first principal component. Indeed, a person skeleton is spatially more spread over its height than its width. In statistical words, we will have more variance along the height axis than the width axis. The second principal component would be the width of the person.

We filtered the Fig. 3(b) with a 3x3 median mask. We then applied the PCA to the foreground pixels of Fig. 3(d). The outcome is presented on Figure 4.

The projection onto the eigenspace successfully straightened the silhouette. The first component corresponds to the person height and the second component to the person width as expected. The algorithm also shows promising results for smaller ROI.

In this part, we expose the methods we combine in order to obtain a robust segmentation of a person in an

unconstrained environment. We also introduce a brand new technique used to locally unwrapping the catadioptric images once the ROI is determined.

4. First Results

As mentioned in the title, this paper aims at presenting a preprocessing framework for gait analysis on catadioptric images. As a result, we would also like to present our first results in this context. Gait is based on body unbalance. It entails a body asymmetry originating from the arm and leg swing. We characterized that asymmetry by measuring the ratio of the mass of pixels with values inferior to zero along the horizontal axis over the mass of pixels superior to zero. The Fig. 5(a) presents the ratio evolution when a subject was asked to walk straight from the left hand side to the right hand side of the hallway. The dashed curve is actually the polynomial approximation of the plain curve. It reflects the trend of the plain curve. Even if it requires further investigation, the trend seems to reflect the person pose variation as well as arms movements. It explains why we have peaks followed by valleys and not the opposite. We are really interested in the variation of the ratio. As a result, we analyzed the residual signal (Fig. 5(b)). We were interested in extracting the gait frequency from the silhouette.

We logically performed a Fourier transform of the signal in order to obtain its frequency signature. The experiment was performed on two subjects so far (Fig. 5(c) and Fig. 5(d)). The maximum was obtain at f=1.4 Hz for the subject A and f=1.875Hz for subject B. The accuracy of these results has to be taken within its context. As a matter of fact, the frame rate is 15 fps. As a result, the sampling rate of our study will be 15 Hz. The ratio recording was performed during 5 seconds. The resulting spectral resolution is 0.33 Hz. A visual analysis of the footage gives a approximated frequency of 1.6 Hz for subject A and 2 Hz for subject B. These second results show the constituency of the first ones. In fact, the pixel-based technique gives a result within the spectral resolution. We could expect that this analysis gives better results with a smaller spectral resolution. In order to achieve this performance, we need to increase the length of the user's footage.

5. Conclusions

Gait analysis is an emerging biometrics as a recognition technique at a distance. However, the first techniques developed assume good lighting conditions. Most of the time these constrained can't be achieved. Our case of study is a good example where regular segmentation technique can't give good results and moreover it would need further steps to be added. It would result in a heavy processing. This would not be a

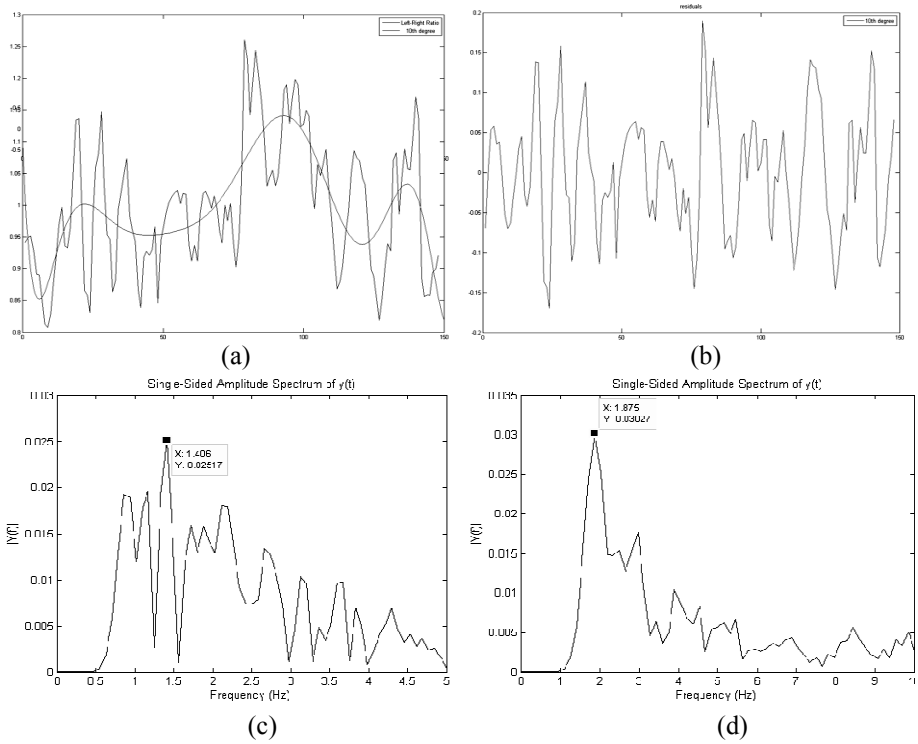

Figure 5. (a) Ratio evolution (plain line) and polynomial approximation (dashed line) (b) Residual after trend removal (c) FFT of the residual for subject A (d) FFT of the residual for subject B.

real time processing and as a consequence would be limited to desktop environment. In this paper, we introduced two new image processing techniques. The first one, dedicated to pure image segmentation outperforms GMM in indoor unconstrained environments. Its major assets are the high reduction of false positive foreground pixels. Only few post filtering is required to obtain good results. Moreover, it's also robust to low color variation between the model and the actual image. In our environment, it successfully enables to extract the legs from the floor. Secondly, we propose a new unwrapping technique for ROI in a gait analysis context. Based on the PCA of the ROI, it doesn't require unwrapping the entire image required before. This can significantly reduce the computational time required.

Our approach suffers from some limitations. Indeed, the PCA is a linear approach. The catadioptric silhouette will suffer from physical unbalance. The chest of the person, in our configuration, will represent a larger amount of pixels than the legs. The PCA will be drifted by this phenomenon. In order to overcome this problem, we plan to use a weighted-PCA approach. The weights will depend on the mirror geometry. We then plan to implement the algorithm with OpenCV library in order to really measure its real-time potential. We finally intend to start implementing feature extraction of gait characteristics using the silhouette extracted with the technique we introduced.

Acknowledgment

This work is part of the NOmad Biometric Authentication (NOBA) project funded by ERDF under the Interreg IVA program (Ref. No. 4051) in collaboration with University of Kent.

References

[1] Nayar, S.K., "Catadioptric Omnidirectional Camera." Washington : IEEE Computer Society. IEEE Conference on Computer Vision and Pattern Recognition. pp. 482-488.

[2] Boutteau, R. , Savatier X., Ertaud JY. and Mazari B., *Mobile Robots Navigation.* ROUEN : In-teh, 2010. pp. 1-24. 978-953-307-076-6.

[3] Bouchrika, I., *Gait Analysis and Recognition for Automated Visual Surveillance.* University of Southampton. Southanpton : s.n., 2008. PhD Thesis.

[4] Bouchrika, I. and Goffredo, M. and Carter, J.N. and Nixon, M.S., "Covariate Analysis for View-point Independent Gait Recognition." Sassari, Italy : s.n., 2009. 3rd IAPR/IEEE International Conference on Biometrics. pp. 990-999.

[5] Farnoosh, R., Gholamhossein Y. and Behnam Z., "Image segmentation using Gaussian mixture models." IUST International Journal of Engineering Science, 2008, Issue 1-2, Vol. 19, pp. 8-13.

[6] Makihara, Y.S. and Sagawa, R. and Mukaigawa, Y. and Echigo, T. and Yagi, Y.S., "Gait Identification Considering Body Tilt by Walking Direction Changes." Electronic Letters on Computer Vision and Image Analysis, July 2009, Issue 1, Vol. 8, pp. 15-26.

[7] Tan, P.N. and Steinbach, M. and Kumar, V., *Introduction to Data Mining.* Boston : Addison Wesley, 2006.

[8] Ilea, Dana E. and Whelan, Paul F., "Color image segmentation using a spatial K-means clustering algorithm." Dublin : s.n., 2006. 10th International Machine Vision and Image Processing Conference. pp. 146-153.