Automatic Extraction of Human Gait Feature for Recognition

Sonia Das *, Sukadev Meher **
*(Dept. Of Electronics and Communication, National Institute of Technology, Rourkela, Odisha)
** (Dept. Of Electronics and Communication, National Institute of Technology, Rourkela, Odisha)

ABSTRACT

Human gait is a new biometric resource in visual surveillance system. It can recognize individual as the way they walk. In the walking process, the human body shows regular periodic variation, such as upper and lower limbs, knee point, thigh point, height, etc. which reflects the individual's unique movement pattern. However from a computational perspective, it is quite difficult to extract some feature points (knee, thigh, leg, and hip) because of occlusion of clothes, carrying bags. Height is one of the important features from the several gait features which is not influenced by the camera performance, distance and clothing style of the subject. This paper proposes DLT method of predicting height variation signal from the gait cycle of each subject. Height estimation has done using calibrated camera images. The variation of height signal is further analyzed using various transform: DHT, DFT, and DCT. Euclidian distance and MSE are computed on feature vectors to recognize individual.

Keywords **- Gait Recognition, silhouette Detection, Camera Calibration, Height Measurement.**

I. INTRODUCTION

In visual surveillance system human identification is a challenging job. Human identification uses many biometric resources, for instance fingerprint, palm print, face, Iris, and hand geometry. Each individual resource requires a bound fundamental interaction between a person and a system. But gait as a behavioral biometric resource, which is non-invasive and arguably non-concealable nature [1], [3]. The biometric features of face will not give satisfactory result when the distance between the camera and person is large. In such case gait features will give an estimable result. Human gait recognition works from the observation that an individual's walking style is unique and can be used for human identification. The extraction and analysis of human walking movements, or gait, has been an ongoing area of research since the advent of the still camera in 1896 [2].Gait as a biometric is strongly motivated by the need for an automated recognition system for visual surveillance and monitoring applications. In gait feature extraction, methods such as modeled based approach [4], silhouette based methods are commonly adopted [5]. Model based approaches to feature extraction, use prior knowledge of the object, which is being searched for, in the image scene. Models used are typically stick representations either surrounded by ribbons or blobs. But the disadvantage of implementing a model based approach to be that the computational costs, due to the complex matching and not implementing in real time system. Silhouette based method, recognizing people by gait intuitively depend on how the silhouette shape of an individual changes over time in an image sequence.

There are many properties of gait that might serve as recognition features. Early medical studies suggest that there are 24 different components to human gait and if all movements are considered, gait is unique [6]. However from a computational perspective, it is quite difficult to accurately extract some of the components such as angular displacements of thigh, leg and hip points using current computer vision system and some others are not consistent over time for the same person. So precise extraction of body parts and joint angles in real visual imagery is a very cumbersome task as non-rigid human motion encompasses a wide range of possible motion transformations due to the highly flexible structure of the human body and to self occlusion. Furthermore different types of clothing is also a crucial factor for the extraction of above components.

A new method is proposed here to extract some gait feature and perform some recognition task very efficiently. The gait feature extraction process and its identification process of the method are described in Section II and Section III respectively. In feature extraction process the binary silhouettes are extracted using a simple background subtraction. Direct Linear Transformation method (DLT) is employed to transform 3-D information to 2-D information which makes the system computationally efficient for real time implementation. This proposed method calculates the variation of height during walking. The measurement of height of a person is not affected by his clothing style as well as the distance from the camera. At any distance the height can measure, but for that Camera Calibration is essential. Previously Lee et al. [8] used Direct Linear Transformation (DLT) method to estimate the height of a stationary person. So he only used intrinsic parameter of the camera. In the proposed method we used both intrinsic as well as extrinsic parameters of the camera because extrinsic parameter gives both translational and rotational matrix. Then we used DLT method to find the height of a moving person for each frame. The height is varying for each movement of subject. As human walking style is periodic [11] so the accumulation of height will give a periodic signal. We extract the height variation signal of individual person. Section III gives a process of human identification in which different windowing techniques (Blackman window, Rectangular window) have applied on height signal to get finite samples from continuous signal. These samples are further used for spectral analysis. DCT (Discrete cosine Transformation), DFT (Discrete Fourier Transformation), and DHT (Discrete Heartily Transformation) techniques are applied on the samples. N- Harmonics are selected from the transformation coefficients. These coefficients are known as feature vectors which are stored in the database. Euclidian distance and MSE are calculated on these feature vectors. When feature vectors of same subject are compared, then a maximum value of MSE is selected, known as Self Recognition Threshold (SRT). Its value is different for different transformation techniques. It is used to identify individuals. Section IV gives simulation results and discussion and Section V gives conclusion.

II. PROPOSED MODEL

Human height during his walking varies periodically and the variation of height depends on the style of walking of the subject. So this variation of signal gives a pattern which is used as a gait feature. We captured only side view videos of 10 different subjects in the indoor environment at different distances from the fixed camera. Each subject walked 4 to5 times at different time both day and night. Side view videos are taken because gait features are easily perceived.

The overview of the proposed model for height extraction is shown in Fig. 1. The algorithm consists of three main modules. The first module tracks the walking person and extracts the head and feet point from each frame. The detection of head and feet points consist of following steps-Input video frames extraction, silhouette detection, corner point's detection, head and feet point detection. The second module uses calibration process to get intrinsic and extrinsic camera parameters. Direct Linear Transformation is used in the third module to give approximate height of the subject in each frame.

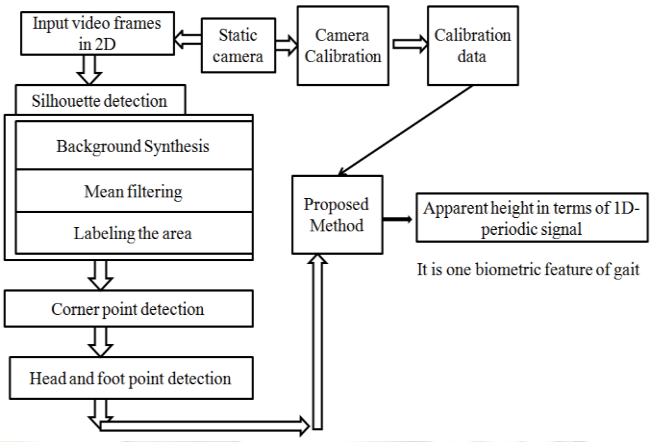


Fig. 1. System overview of gait feature extraction process.

2.1 Silhouette Detection Method

The silhouette of the subject detected in three- step operations.

Step-1: After the extraction of video frames from video sequences, background model is constructed using simple statistical method [12]. Frame Differencing Method is used to separate the foreground to the background. This method is used because we have taken static background video.

Step-2: Mean filtering is used to remove some back-ground noise after back ground subtraction.

*Step-3: E*ach connected element is labeled; area of each one is calculated ,and then compare the area of silhouette with other elements and treat other element as background.

Four examples of subjects with their background models and silhouettes are shown in Fig. 2.

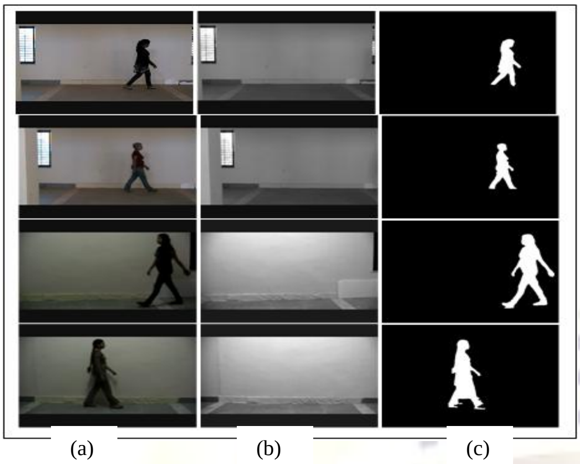


Fig. 2. Examples of moving silhouette extraction.(a)Original images, (b) Backgrounds Constructed(c) Extracted Silhouettes.

2.2 Head and Feet Point Detection

Corner points detected after detection of silhouette, using Plessey corner detector. This operator considers a local window in the image and determines the average change of intensity resulting from shifting the window by a small amount in various directions. This operation repeated for each pixel position which assigned an interest value equal to the minimum change produced by these shifts. Points of interest are the local maximum of the interest values, since corners exhibit a large intensity variation in every direction. Once the corner points detected then select the top max point and bottom min point of the silhouette. These points are called as head and feet point respectively. The whole process is shown in Fig. 3.

Fig. 3. The process of getting Head and Feet points.

2.3 Camera Calibration

The goal of Camera Calibration is to determine the parameters of the camera. There are two types of camera parameters such as intrinsic and extrinsic. Intrinsic or internal camera parameters are those, which describe the projection of objects onto the camera image. They establish the relationship between the points in the camera reference frame and the pixel coordinates of the points on the images got from the camera. Extrinsic or external camera parameters describe the location and the orientation of the camera. They establish the relationship between the camera reference frame (coordinate system) and the world reference frame. Determine these parameters basically mean that we find the transformation (translation and rotation) which reconciles the 3-D coordinate system of the camera and that of the real world. We concerned about both intrinsic as well as extrinsic parameters because although we have used static camera but our subjects are moving. So we bound to take extrinsic parameters to get translational and rotational transformations.

In this paper, we have used Camera Calibration Toolbox developed by Jean-Yves Bouguet [10]. A set of 9 monochrome test images was taken from the Cannon LEGRIA FS305 camera. Test images featured a planar checkerboard grid differently oriented in the each image (shown on Fig. 4). Camera calibration Tool box features an algorithm that uses the extracted corner points of the checkerboard pattern to compute a projective transformation between the image points of the n different images.

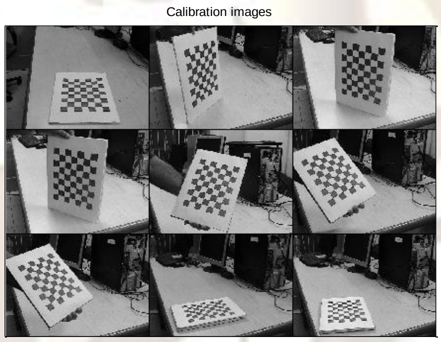


Fig. 4. Images used in calibration process

Afterwards, the camera intrinsic and extrinsic parameters are recovered using a closed-form solution.

The camera matrix is formed using both the parameters [7]. The matrix converts 3D image to 2D image using (1):

$$\begin{pmatrix} u \\ v \end{pmatrix} = K\begin{bmatrix} R & t \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (1)$$

K=intrinsic matrix = $\begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix}$

$f_x \ f_y$ are the focal length in x and y direction.

C_x, C_y are the principal point in x and y direction.

X, Y, Z are the object coordinates in the real world.

[R t] = joint rotation –translation matrix.

u, v are the image coordinates.

It is the combination of 3×3 rotation matrix and 1×3 Translation matrix.

It is the combination of 3×3 rotation matrix and 1×3 Translation matrix.

R= $\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$ t = $\begin{bmatrix} x_t & y_t & z_t \end{bmatrix}$

As in 2-D case, a homogeneous transformation matrix is defined, for 3-D case, a 3×4 matrix is obtained that performs the rotation given by R (α, β, □), followed by a translation given by $x_{t,}\ y_{t,}\ z_t$.

$$T = \begin{bmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha s\beta c\gamma + s\alpha s\gamma & x_t \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma & y_t \\ -s\beta & c\beta s\gamma & c\beta c\gamma & z_t \end{bmatrix}$$

Where $c\alpha = cos(\alpha)$, $c\beta = \cos(\beta)$, $c\gamma = \cos(\gamma)$,

$s\alpha = \sin(\alpha)$, $s\beta = \sin(\beta)$, $s\gamma = \sin(\gamma)$

α=yaw angle, β = pitch angle, □ = Roll angle.

T is the homogeneous transformation matrix

X, Y, Z are the object coordinate in the real world.

2.4 Direct Linear Transformations

A point under the feet of the subject is chosen as the image coordinate (u1, v1), and the coordinate of a point over the head of the subject (u2, v2) is arbitrarily chosen. Z2 can be obtained by solving the four equations in (2)

$$c_{11}X_1 + c_{12}Y_1 + c_{13}Z_1 + c_{14} - \left(u_1c_{31}X_1 + u_1c_{32}Y_1 + u_1c_{32}Z_1\right) = u_1$$

$$c_{21}X_1 + c_{22}Y_1 + c_{23}Z_1 + c_{24} - \left(v_1c_{31}X_1 + v_1c_{32}Y_1 + v_1c_{32}Z_1\right) = v_1 \qquad (2)$$

$$c_{11}X_1 + c_{12}Y_1 + c_{13}Z_2 + c_{14} - \left(u_2c_{31}X_1 + u_2c_{32}Y_1 + u_2c_{33}Z_2\right) = u_2$$

$$c_{21}X_1 + c_{22}Y_1 + c_{23}Z_2 + c_{24} - \left(v_2c_{31}X_1 + v_2c_{32}Y_1 + v_2c_{33}Z_2\right) = v_2$$

Assuming that the image coordinate under the feet Z1 is zero and the x– y coordinate of the points over the head and the feet are equal. In case of height only Z-coordinate is different. The difference of Z1 and Z2 will give the height of a subject.

2.5 Height signal Detection

Fig. 5 presents the extracted height model superimposed over the original sequence of images for subject in dataset.

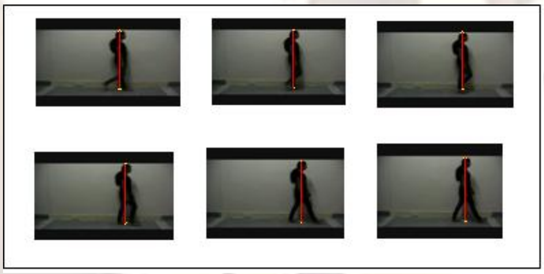


Fig. 5. Walking sequence with extracted height model for subject. Frames run from left to right.

In the swing phase (one feet is in the ground and other feet toe-off) the vertical segment between the head and feet and the height is maximum. In stance phase (when two feet contact in the ground and apart) the vertical segment that extends from the top of the head to the point halfway between the two feet. At that time height is the minimum. Fig. 6 shows variation of height pattern.

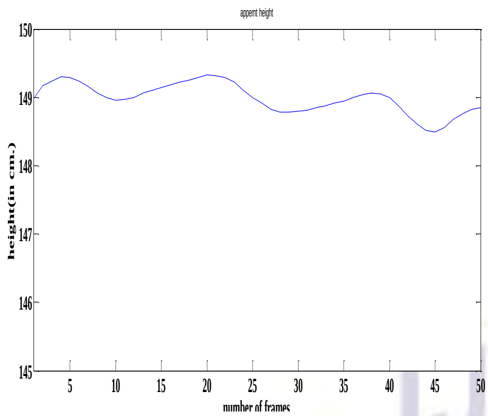


Fig. 6. Height changing pattern extracted with the height model for the sequence in Fig. 5.

III. HUMAN RECOGNITION

The process of human feature identification is shown in Fig. 6. In the proposed model 1-D height signal is generated by combining the height of each frame of the silhouette. Different windowing techniques such as Blackman window and Rectangular window are applied to get finite samples from this continuous signal. Windowing techniques are applied to avoid leakage outside the finite interval. DFT (Discrete Fourier Transformation), DHT (Discrete Heartily Transformation), DCT (Discrete cosine Transformation), are used on the samples to extract the feature. DFT is a powerful tool for analyzing and measuring the continuous signal. It produces the average frequency content of the signal over the entire time that the signal was acquired. It maps a length-N complex sequence to a length-N complex spectrum. Discrete Heartily Transformation is a real-valued transform closely related to DFT of a real-valued sequence. It directly maps a real-valued sequence to a real-valued spectrum while preserving some of the useful properties of DFT [9]. Discrete cosine Transformation is a real and orthogonal transformation technique. It expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. In DCT most of the signal components are stored in the lower harmonic components. These transformations are applied on the apparent height signals. Then N-harmonics from the transformation coefficients are selected. These are the feature vectors stored in the database. If any subject comes then extract his features by using the above process, then compare these features with the database features by using Euclidian distance and MSE computation. If MSE value is greater than SRT then different subjects, otherwise same subject.

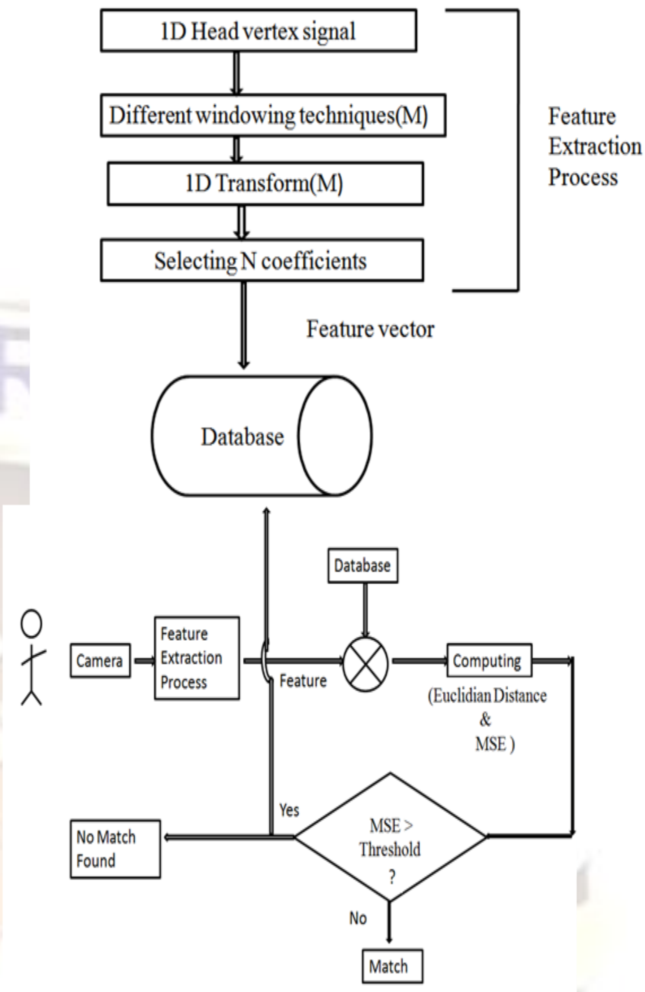


Fig. 6. The process of feature identification.

IV. SIMULATION RESULTS AND DISCUSSION

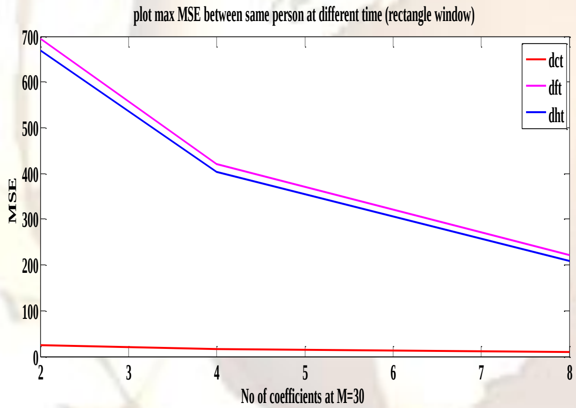
A video camera is placed with a plane normal to the subject’s path in indoor environment with controlled illumination. Subjects are asked to walk in front of a plain, static background. A static video camera captured video sequences at a distance of 500 cm, 250 cm, and 200cm. The length of each Video is 3-4 sec. and the frame size is 720×480 pixel resolution. This paper used 50 consecutive frames from each video. This is a very important step as the total result depends on the quality of the gait captured.

The method is tested on walking style of 10 different subjects and each subject is advised to walk minimum 4-5 times. First we compared the features of walking style of each subject with himself and others. Three transformation techniques applied and their coefficients compared to get MSE. In the table 1 we put some experimental results of MSE of same subjects and different subjects to get threshold values for each transformation using two different windowing techniques such as Blackman and rectangular window.

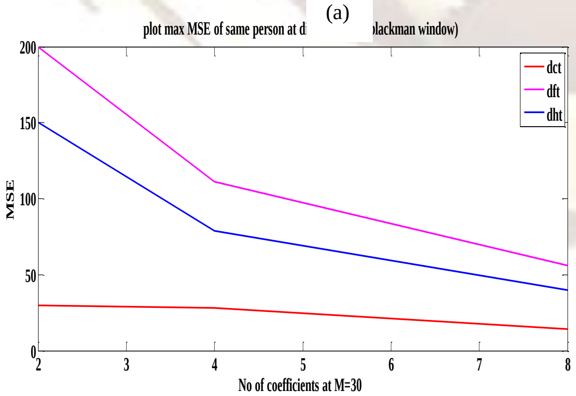
TABLE 1. The Results of MSE when same subjects and different subjects are compared.

Method	Blackman Window Technique					
	$P1\sim P2$	$P1\sim P2'$	$P1\sim P3$	$P1\sim P1'$	$P1\sim P1''$	$P1\sim P1'''$
DCT	57.042	40.51	27.69	4.641	1.8354	1.7046
DFT	1361.2	966.8	659.1	111.3	41.82	37.88
DHT	1308.7	995.6	592.7	79.6	32.56	24.38
	Rectangular Window Technique					
DCT	158.34	92.15	56.7	15.85	8.76	5.78
DFT	6594	4681	2731	421.1	182.4	113.6
DHT	6562	2358	1358	208.5	85.7	54.99

Here $P_{1,} P_{2,} P_3$ are the coefficients of different subjects. $P_1^{'}, P_1^{''}, P_1^{'''}$ are the coefficients of same subject's P_1 at different time of walking. We used Blackman window technique and Rectangular window technique to get M number of samples from the 1D signal. Then DCT, DFT, DHT transformation techniques applied on them. Then N no. harmonics coefficients are taken from M no. of sample coefficients which makes a N×1 feature vector. A tradeoff between computational burden and for better result we selected first 4 coefficients among 30 samples. In Fig. 7 threshold values for different transformation techniques are plotted using Rectangular and Blackman window.



(a)



(b)

Fig. 7. The above graphs show the threshold values by using different transformation techniques. Mean Square Error in y axis and no of coefficients in x-axis: (a, b) Compare MSE for same subject walking style using Blackman and Rectangular window techniques respectively.

We carried out these threshold values for recognition of other 10 subjects. These subjects also walked 4-5 times at different time. Table II compares recognition rate of different transformation methods used in the proposed model. A comparison of different of average recognition rate of subjects for different transformations is shown in Fig. 8.

$$Recogination\ Rate = \frac{Number\ of\ Subjects\ Correctly\ recognized}{Total\ number\ of\ Subjects\ tests} \times 100$$

TABLE II. Recognition rate of different transformation techniques.

Methods	Window Techniques	Recognition Rate (%)
DFT	Blackman window	46
DCT		56
DHT		53
DFT	Rectangular window	57
DCT		55
DHT		60

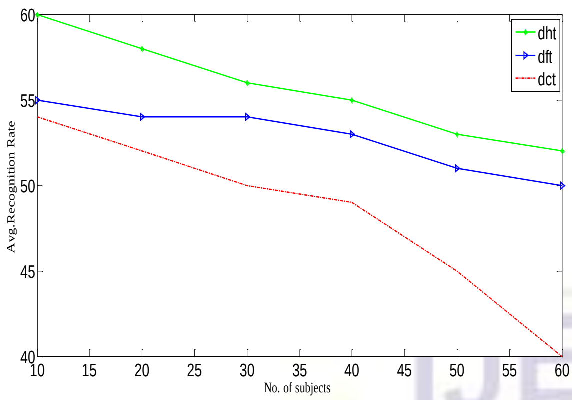


Fig. 8. Compare average recognition rate between different transformations using rectangular window.

V. CONCLUSION

We extract one feature of gait known as height for human recognition. Height variation of a person during walking is radically distinctive from person to person. Our experimental result shows analysis of height signals using different windowing techniques and different transformation techniques. These techniques give six different threshold values. The SRT value of DCT is 10, DHT is 100, and DFT is 200 in case of Blackman window. In case of Rectangular window SRT value of DCT, DHT, and DFT is 30, 670, and 700 respectively. By using these threshold values we compared recognition performance of all three techniques. Rectangular windowing technique using DHT gives better performance than others.

References

[1] W. Hu, T.Tan, L.Wang, S.Maybank, A survey on visual surveillance of object motion and behaviors, *IEEE Trans. System, Man, and Cybernetics-Part C, 34(3)*, 2004, 45-90.

[2] P. C. Cattin, D. Zlatnik, and R. Borer, Biometric system using human gait, In *Mechatronics and Machine Vision in Practice*, 2001.

[3] K. Rohr, Towards model-based recognition of human movements in Image sequences, *CVGIP: Image Understanding, 59(1)*, 1994, 94–115.

[4] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, *IEEE Trans. Pattern Analysis, and Machine Intelligence*, 25(12), 2003, 1505-1518.

[5] R. Collins, R. Gross, and J. Shi, Silhouette-Based Human Identification from Body Shape and Gait, *Proc. Int. Conf. Automatic Face and Gesture Recognition, Washington, Dc*, 2002, 366-371.

[6] A. Kale, A. Rajagopalan, N. Cuntoor, and V. Kruger, Gait-Based Recognition of Humans using Continuous HMMs, *in Proc. Int. Conf. Automatic Face and Gesture Recognition, Washington, DC*, 2002, 336-341.

[7] Zhang, A Flexible Camera Calibration by Viewing a Plane from Unknown Orientations, *In proceedings of ICCV'99, Corfu, Greece*, September 1999, 666-673.

[8] J. Lee, E.D. Lee, H.O. Tark, J.W. Hwang, D.Y. Yoon, Efficient height measurement method of surveillance camera image, *Forensic Sci Int. 177* , 2008, 17 – 23.

[9] H. Sorensen, D. Jones, C. Burrus, M. Heideman, On computing the Discrete Hartley Transform, *IEEE Trans. Acoustics., Speech and Signal processing, 33(4)* , 1985, 2131-1238.

[10] Fetic, A., D. Juric and D. Osmankovic, The procedure of a camera calibration using Camera Calibration Toolbox for MATLAB, In *MIPRO, 2012 Proceedings of the 35th International Convention. IEEE*, 2012, 1752-1757.

[11] C. BenAbdelkader, R. Culter, and L. Davis, Person identification using automatic height and stride estimation, *in Proc. Int. Conf. PatternRecognition, vol. 4, Québec, Canada*, 2002, 377–380.

T. Horprasert, D. Harwood, L.S. Davis, A statistical approach for real-time robust background subtraction and shadow detection, *IEEE ICCV'99 Frame-Rate Workshop, Corfu, Greece*, 1999.