

HUMAN GAIT ANALYSIS AND RECOGNITION USING SUPPORT VECTOR MACHINES

Deepjoy Das and Dr. Sarat Saharia

Department of Computer Engineering, Tezpur University, Assam
deepjoy2002@gmail.com
sarat@tezu.ernet.in

ABSTRACT

Human gait reveals feelings, intentions and identity which is perceived by most human beings. To understand this perceptual ability, Swedish psychologist Gunnar Johansson (1973), devised a technique known as PL (Point Light) animation of biological motion. In his work, the activity of a human is portrayed by the relative motions of a small number of markers positioned on the head and the joints of the body. This paper explores the basic concept of PL animation along with machine vision and machine learning techniques to analyze and classify gait patterns. Basically, frames of each video are background subtracted, the silhouette noise found were salt noise and noise connected in large blobs which are detected and removed based on morphological operations and area of connected components respectively. Image is then segmented and body points such as hand, knee, foot, neck, head, waist along with the speed, height, width, area of person are determined by an algorithm. We then fit sticks connecting pairs of points. The magnitude and direction of these stick along with other features forming a 24-dimensional feature vector for each frame of a video are classified using SVM Modelling using LIBSVM Toolkit. The maximum recognition accuracy found during testing by cross validation with parameters of LIBSVM was 93.5 %.

KEYWORDS

Human gait analysis, Recognition, body point identification, stick view, support vector machines.

1. INTRODUCTION

Biometric is the field of study that uses physiological or behavioural traits to identify a person. The study of human gait recognition is related to analysing human's distinctive way of walking and extract patterns used for recognition. Human gait can used in identification of actions, gender, mood, emotions and intensions of a person. Gait requires no subject contact compared to other biometric technologies like fingerprint, iris detection, face recognition which requires the subject to be in close proximity with the sensor, therefore gait can be far more superior compared to other biometric technologies for example Geisheimer et al. [1] used continuous wave radar developed to record human gait signatures and extracted various gait parameters using signal processing techniques like short time Fourier transform (STFT).

Accuracy of gait analysis and recognition depends on two factors 1) feature extraction techniques followed to extract features from the walking subject e.g., a) models can be a shape based model for example, Lee & Grimson [2] used shape based features by dividing the silhouette into 7 regions and by extracting centroid, aspect ratio of major and minor axis of the ellipse and the orientation of major axis of the ellipse from each of the region , b) motion based model for example, Johansson [3] devised the technique known as PL animation, by which an observer was able to recognize walking subject affixed with lights in the major joints of the body or c) a combination of shape & motion based model. Feature extraction techniques should closely model the information captured by human brain 2) Classification model chosen to classify the gait patterns e.g., supervised techniques like Artificial Neural Network, Support vector machines etc. or unsupervised techniques like clustering etc.

2. RELATED WORK

In 1973 Johansson [3] devised PL animation, according to him an observer can recognise a person familiar to him when walking with light affixed in the major joint of their body. Since then various studies have been carried out using PL animation, studies which verify identify [4], sex [5] [13], emotion [6], facial expression using PL face [7]. Some studies also concluded that increasing the number of PL point on the human body increases accuracy of the recognition [8].

PL animation was proven to be a strong motion analysis model as studies show that manipulating PL animation data or purposefully degrading the information of PL data by fluctuating dot contrast polarity, blurring dots and spatial-frequency filtering doesn't degrade visual recognition performance of the observer [9].

Inverted PL animation however has some interesting results, studies show that human action are not easily perceivable using inverted PL animation [10]. However, inverted PL animation has an application in Parkinson's disease patients as a walking aid [11].

Studies have also been carried out using the stick view obtained by connecting the dots of PL animation. Hodgins et al. [12] studied tested stick view model, polygonal models, and others to check observer recognition performance in each of the model and reported polygonal model performance was greater than stick view. Gender identification study carried using stick view by Yoo et al. [13] using body contour & joint angles as features with SVM classifier shows good result.

Apart from motion based model, human gait can be also represented by a shape based model. Shape based model can be divided into counter based model [14] and region based model [15] [16]. Contour based model only uses boundary shape information whereas region based methods divide the shape region into parts which are then used for shape representation. Feature extraction methods using Chain coding, Fourier & Wavelet descriptor, Autoregressive, B-spline falls under contour based category, whereas Geometric, Zernike, Legendre, Krawtchouk Moments, Convex Hull falls under region based method.

Over a decade of research predicted that PL animation is the dominant model for motion analysis of humans and other living creatures. Our approach towards classifying gait pattern falls under motion based model viz. PL animation using Support Vector Machines. The study combines the ancient & intuitive idea of feature extraction using powerful classification model. Our throughout study concludes that motion based models are more powerful than shape based motion in terms of classification accuracy with less no. of features.

3. DATA COLLECTION

This section explains about the video capturing procedure followed, typically in terms of camera specification, camera placement, posture of walking subject and other specification. A camera fitted on a tripod was well adjusted so that the side view of walking subject can be captured. 10 videos of each walking subject were captured in 5 different locations (In each location 2 videos of each subject were captured). Different location have different light intensities, shadow fall, wind speed etc. We collected a video of 50 different walking subject. Videos are captured using Nikon S3000 which has Optical Sensor Resolution of 12 megapixels, frame rate of 25 frames/second and lens focal length 4.9-19.6mm/F3.2-5.9. The resolution of video used was 320×240 (QVGA Resolution).



Figure 1: Captured Frames during Data Collection

4. PREPROCESSING

After the data collection step, we go through a series of pre-processing step sequentially to get a perfect silhouette for feature extraction, the feature extracted is then modelled by SVM.

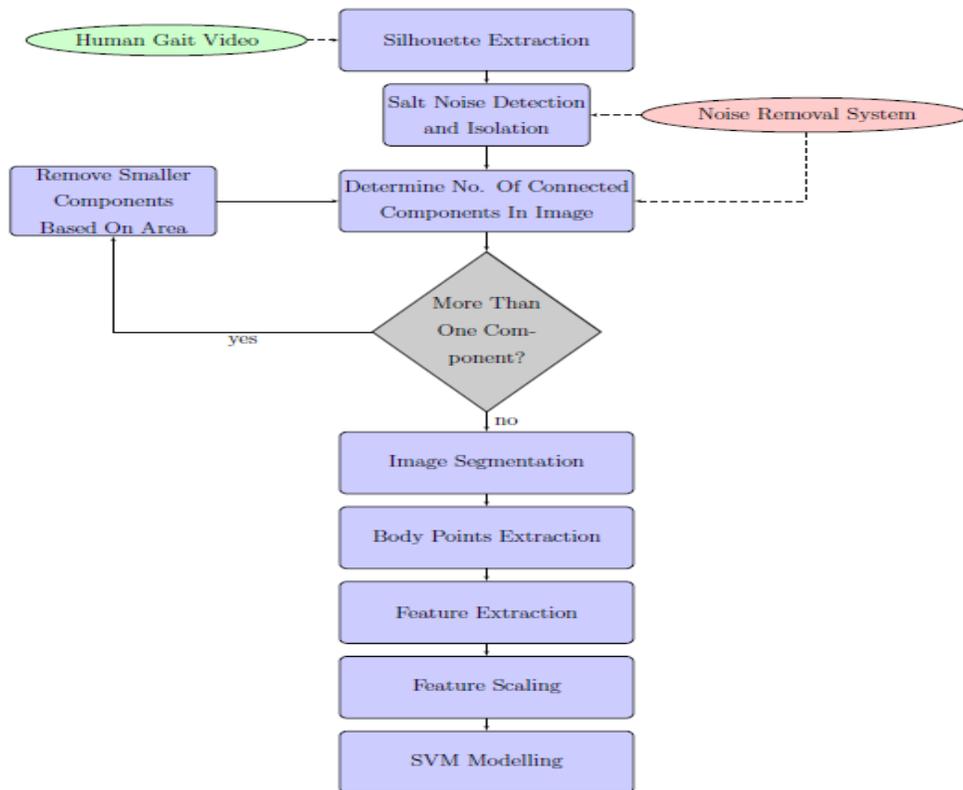


Figure 2: Flowchart of Preprocessing & Feature Extraction

4.1 Silhouette Extraction

Frame difference is the simplest form of background subtraction. The current frame is simply subtracted from the previous frame, and if the difference in pixel values for a given pixel is greater than that of threshold T_h then the pixel is considered part of the foreground [17].

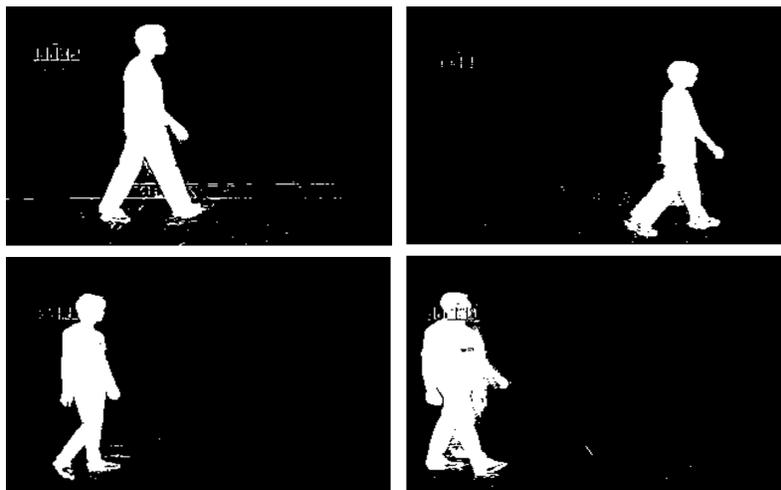


Figure 3: Background Subtraction

$$|frame_i - frame_{i-1}| > T_h$$

The estimated background is just the previous frame and it is very sensitive to the threshold T_h .

The threshold T_h has been determined by hit and trial method.

4.2 Noise Detection and Removal Techniques

The gait representation technique used in this paper may be partially biased by noisy silhouette. To overcome this, in this work the noisy silhouette are processed by the silhouette noise removal algorithms which efficiently removes two types of noise, which are mentioned as follows:

1. Intensity errors spread randomly across the silhouette (also called salt Noise) are removed by using a morphological opening operation on the silhouette with the square structuring element of size 6×6 (The morphological open operation is an erosion followed by a dilation, using the same structuring element for both operations).
2. Errors connected in large blobs in silhouette (greater than the size of structuring elements) are removed by finding the number of connected components in the binary silhouette and then filtering the component based on the area of the components. The low valued areas are filtered out and the largest area of connected component (which is the person in the silhouette) remains in the resulting silhouette.



Figure 4: Original Grayscale Frames Processed after Noise removal Techniques, the algorithm correctly identifies the largest area after filtering out the rest

5. FEATURE EXTRACTION

We identify nine major points of human body, sticks are fitted into the frames using the points. The magnitude of the sticks and angle between the sticks, along with height of the body and velocity of the centroid constitutes a feature vector for a single frames of a single class. A total of 24 features were used.

5.1 Estimation of Body Points

1. For each frame the centroid is calculated (the centroid is always inside the walking person image).
2. Keeping the vertical axis distance from the centroid fixed, the max and min point is searched in vertical direction.
3. Using max and min point the total height (H) of the person is estimated in each frame.
4. For a body height H, an estimate of head is 0% of H, neck 13% of H, Hand is between 35-50% of H, waist is 47% of H, knee is 71.5% and foot is between 93-98% of H. (all the points are distance from the horizontal axis).

The image is segmented further to search other body point. We now search in the horizontal direction keeping the fixed range 35-50% H in vertical direction to find the front and rear hand point (in term of distance from vertical axis) respectively. Similarly, the front and rear coordinates points of the knee and foot are identified.

Body parts were points are fitted are as follows: Head, Neck, Front Hand, Rear Hand, Centroid, Front Knee, Rear Knee, Front foot, Rear foot.

5.2 Computation of major angles between body parts

Apart from this we determine 9 angle values which are 1. Angle between the head and the body [some person walk with their head bend towards front/back] 2. Angle between the front hand and the body [Measures the amount of front arm swing] 3. Angle between the rear arm and the body [Measures the amount of rear arm swing] 4. Angle between front thigh and vertical axis [Measures the amount of front knee swing] 5. Angle between rear thigh and vertical axis [Measures the amount of rear knee swing] 6. Angle between front leg shin and vertical axis 7. Angle between rear leg shin and vertical axis 8. Angle between front leg foot and horizontal axis 9). Angle between rear leg foot and horizontal.

5.3 Total features vector

The relative distances between the points were measured with respect to other points as a human observer always looks at the relative distance with other points. Thus the Euclidean distance between pairs of different point were used as a feature.

We use the following set of features after finding Euclidean distance:

1. Euclidean distance of head point from the neck point.
2. Horizontal distance of front hand from the centroid point.
3. Vertical distance of front hand from the centroid point.
4. Horizontal distance of rear hand from the centroid point.
5. Vertical distance of rear hand from the centroid point.
6. Horizontal distance of front knee from the centroid point.
7. Vertical distance of front knee from the centroid point.
8. Horizontal distance of rear knee from the centroid point.
9. Vertical distance of rear knee from the centroid point.
10. Horizontal distance of front foot from the front knee point.
11. Vertical distance of front foot from the front knee point.
12. Horizontal distance of rear foot from the rear knee point.
13. Vertical distance of rear foot from the rear knee point.



Figure 5: Vectors corresponding to the extracted point are fitted to image. Magnitude and the direction of these vector will be used for classification

Along with these 13 Euclidean distances, 9 angle values (mentioned in sub-section 4.2) along with height and centroid velocity were used to form a 24 dimensional feature vector for each frame of a video. As SVM works on fixed sized vectors, we consider only the first 100 frames of a video, we have 100×24 feature matrix for each video. Since, 50 different persons gait data are present and for each person we have 10 different samples with different parameters changed for example. Therefore, we form 50000×24 feature matrix for SVM modelling using LIBSVM. (i.e., 50 video of different person with 10 different videos of each person with 100 frames of each video vs. 24 features each frame)

5.4 Feature Scaling

The goal is to use the features in LIBSVM (SVM Toolkit), therefore we scale the features in range [0, 1].

$$M'(i, j) = \frac{M(i, j) - \min(i)}{\max(i) - \min(i)} \quad \forall 1 \leq i \leq 50000, 1 \leq j \leq 24$$

Where M and M' represents the unscaled & scaled feature matrix of size 50000×24 respectively. M' has values in range 0 to 1.

6. RESULTS

For SVM modelling LIBSVM [18] toolkit is used. We first divide the feature matrix into training and testing matrix size each of size 25000×24 (5 video of each person is used for training and testing respectively). We then prepare the class label matrix for each of training and testing matrix set. We take the testing matrix and perform cross validation with different c (cost) & g (gamma) values using the default kernel provided by LIBSVM [18] (i.e., radial basis function) to check the performance at unique parameter values. The unique parameter c (cost) & g (gamma) values which yields the highest performance is then fixed and the system is trained to generate a final train model. The system is then tested using training matrix with the train model plugged into the LIBSVM system. The highest accuracy achieved by our method was 93.5%. Confusion matrix for the first 10 classes are shown below in figure 6.

		ACTUAL CLASS LABELS											
		1	2	3	4	5	6	7	8	9	10
	1	23891	0	9	2	0	0	4	0	0	3
	2	1	24816	0	0	0	1	0	2	0	1
	3	0	0	24021	0	0	3	0	0	0	0
PREDICTED	4	0	4	0	23194	0	0	0	0	13	0
CLASS	5	3	0	2	0	22982	0	0	0	8	0
LABELS	6	1	0	0	4	0	23456	0	3	0	0
	7	0	1	0	0	0	0	24781	0	0	0
	8	0	2	0	1	2	0	7	22495	0	0
	9	0	0	0	0	0	0	0	0	24698	0
	10	0	0	0	0	0	0	1	0	0	23003

Figure 6: Testing Confusion Matrix of SVM Testing for first 10 Class

7. CONCLUSION

The result using the 24-dimensional feature vector has better recognition. The 24 features were selected so precisely that these features were the most discriminant than any other features. We can conclude that the motion based model is much stronger representation of human gait than shape based models as motion based feature distinctly recognizes a person with less no. of features than shape based model. For examples, Lee and Grimson [2] used 41 to 57 features by which they get a recognition rate of 98%.

To improve the performance of the algorithm in terms of running time, this algorithm can be tried with anomaly detection algorithm. In which a priority can be assigned to the each features (e.g., the most discriminant feature among all the person gets the top priority). When a videos is fed into the recognition system, we take some top priority features and try to classify it to a classes of existing videos in database. If an anomaly is detected (i.e., if the frames of video cannot be classified in any existing classes) and maximum no. of samples (say x no. of samples) are not classified within a single class, then we add the lower priority feature and classify again. The procedure would yield better running time of the algorithm as its breaking the features into distinct set for classification and computing each feature & testing consumes significant amount of time.

REFERENCES

- [1] Geisheimer, Jonathon L., William S. Marshall, and Eugene Greneker. "A continuous-wave (CW) radar for gait analysis." *Signals, Systems and Computers*, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on. Vol. 1. IEEE, 2001.
- [2] Lee, Lily, and W. Eric L. Grimson. "Gait analysis for recognition and classification." *Automatic Face and Gesture Recognition*, 2002. Proceedings. Fifth IEEE International Conference on. IEEE, 2002.
- [3] Johansson, Gunnar. "Visual motion perception." *Scientific American* (1975).
- [4] Cutting, James E., and Lynn T. Kozlowski. "Recognizing friends by their walk: Gait perception without familiarity cues." *Bulletin of the psychonomic society* 9.5 (1977): 353-356.
- [5] Kozlowski, Lynn T., and James E. Cutting. "Recognizing the gender of walkers from point-lights mounted on ankles: Some second thoughts." *Attention, Perception, & Psychophysics* 23.5 (1978): 459-459.
- [6] Clarke, Tanya J., et al. "The perception of emotion from body movement in point-light displays of interpersonal dialogue." *Perception-London* 34.10 (2005): 1171-1180.
- [7] Bassili, John N. "Facial motion in the perception of faces and of emotional expression." *Journal of experimental psychology: human perception and performance* 4.3 (1978): 373.

- [8] Blake, Randolph, and Maggie Shiffrar. "Perception of human motion." *Annu. Rev. Psychol.* 58 (2007): 47-73.
- [9] Ahlström, Vicki, Randolph Blake, and Ulf Ahlström. "Perception of biological motion." *Perception* (1997).
- [10] Sumi, Shigemasa. "Upside-down presentation of the Johansson moving light-spot pattern." *Perception* 13.3 (1984): 283-286.
- [11] Dunne, J. W., G. J. Hankey, and R. H. Edis. "Parkinsonism: upturned walking stick as an aid to locomotion." *Archives of physical medicine and rehabilitation* 68.6 (1987): 380-381.
- [12] Hodgins, Jessica K., James F. O'Brien, and Jack Tumblin. "Perception of human motion with different geometric models." *Visualization and Computer Graphics, IEEE Transactions on* 4.4 (1998): 307-316.
- [13] Yoo, Jang-Hee, Doosung Hwang, and Mark S. Nixon. "Gender classification in human gait using support vector machine." *Advanced concepts for intelligent vision systems*. Springer Berlin Heidelberg, 2005.
- [14] Lu, Jiwen, and Erhu Zhang. "Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion." *Pattern Recognition Letters* 28.16 (2007): 2401-2411.
- [15] Little, James, and Jeffrey Boyd. "Recognizing people by their gait: the shape of motion." *Videre: Journal of Computer Vision Research* 1.2 (1998): 1-32.
- [16] Pranjit Das, Deepjoy Das and Dr. Sarat Saharia "Gait Analysis and recognition for human identification", National Seminar on Advances in Electronics and Allied Science & Technology, (NaSAEAST- 2013)
- [17] Das, Deepjoy, and Dr Saharia. "Implementation And Performance Evaluation Of Background Subtraction Algorithms." *arXiv preprint arXiv:1405.1815* (2014).
- [18] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." *ACM Transactions on Intelligent Systems and Technology (TIST)* 2.3 (2011): 27.

AUTHORS

Deepjoy Das

Has done Bachelor of Technology from Tezpur University and was held as Assistant Project Engineer in Indian Institute of Technology for one year. His research interest includes image processing with supervised and unsupervised learning and handwriting recognition.



Dr. Sarat Saharia

He is currently working as Associate Professor in the department of computer science and engineering at Tezpur University, Assam, India. His research interest includes pattern recognition and image processing.

