Extracting Silhouette-based Characteristics for Human Gait Analysis using One Camera

Trong-Nguyen Nguyen, Huu-Hung Huynh
DATIC, University of Science and Technology
Danang, Vietnam
{ntnguyen, hhhung}@dut.udn.vn

Jean Meunier
DIRO, University of Montreal
Montreal, Canada
meunier@iro.umontreal.ca

ABSTRACT

With the strong development of computer vision, health care and in-home monitoring systems are widely applied. Gait analysis is one of the main problems, which needs to be solved in such systems. Most of the recent researches implemented on 3D information of each walking person extracted from stereo cameras or devices with sensors, thus it leads to an increase in the computational cost and price. Therefore, we propose an approach for performing gait analysis using only one normal camera. This paper presents how characteristics are extracted from the walking person's silhouette for gait analysis, in detail, detecting abnormal gaits. Experiments are performed with normal gaits and three different types of anomaly, which consist of hunched back, left-right asymmetry, and sudden motion variation. The obtained results show that there is no case of omission or false detection, and our solution can be integrated into real-time systems.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Motion, Video analysis*; I.4.9 [**Image Processing and Computer Vision**]: Applications

General Terms

Theory

Keywords

silhouette, motion, cycle, likelihood, approximated ellipse

1. INTRODUCTION

Human fall and bone-related injuries are the highest risks for elderly. Many solutions were proposed to provide timely assistance. Some recent studies used wearable devices with sensors such as infrared sensors, acceleration sensors and support buttons. However, old people often forget to take them on. Another disadvantage is that if a person is unconscious after the fall, support buttons will become useless. In order to overcome these limitations, vision-based systems are used to monitor and provide warning automatically without human intervention. Such systems are suitable for old people, which live alone.

Many approaches have been proposed for detecting human fall and gait analysis, in which used data are captured via calibrated cameras or devices with sensors such as Microsoft Kinect. The price and computational cost become an issue of concern. Meanwhile, some fall detection approaches were presented using only one normal camera, but it is difficult to generalize to the problem of gait analysis. Therefore, implementing the gait analysis with only one camera is necessary. We found that a sequence of walking person's silhouettes contains the gait and motion information. In this paper, we present gait characteristics, which are extracted from human silhouettes determined using one camera, and how abnormal gaits can be detected based on these features.

2. RELATED WORK

Most researches performed the gait analysis by estimating parameters such as stride length, step length, cadence and speed. In [5], the study was implemented on two groups of people, which are patient with and without a history of falls. The criteria for separating these two groups is a variety of parameters determined using body-worn kinematic sensors. Although the fall risk can be detected ideally via such devices, it is inconvenient for old people to wear these equipments for a long time. In addition, the batteries have to be charged every several hours, so that it interrupts the monitoring.

Recently, markerless motion capture systems have been focused because of its convenience. Most of such approaches used the information of footfalls on the ground in order to estimate related parameters. In detail, the acquisition process was implemented via a set of calibrated cameras. Some other methods examined gait based on depth information obtained from a camera with sensors (e.g. Microsoft Kinect). This device played an important role in clinical settings, in which it was used as a screening tool for representing the asymmetry of participants' gait [11]. In [4], signals similar to wearable gyroscopes were generated by applying a machine learning method on the Kinect. In the same vein, Stone et al. [16] demonstrated the potential of a depth sensor for estimating basic gait parameters at home. However, such systems still have several disadvantages for in-home monitoring because

SoICT '14, December 04 - 05 2014, Hanoi, Viet Nam
Copyright 2014 ACM 978-1-4503-2930-9/14/12 ...$15.00
http://dx.doi.org/10.1145/2676585.2676612

of the limited distance range (0.8m - 4m) and the field of view ($57^\circ \times 43^\circ$ degrees) of the Kinect.

In [15], Stone et al. also estimated gait parameters based on the silhouettes using a set of two calibrated web cameras, which are mounted high in the environment. The acquisition process captures multiple views of the same scene at five frames per second, in which the image resolution is 640×480 pixels. The background subtraction is performed on the raw images captured by the two cameras based on information of texture and color characteristics. In order to obtain a 3D intersection, the extracted silhouettes are projected into a discretized volume space by combining intrinsic and extrinsic calibration parameters. Such 3D intersections are then used to determine the location of footfalls in a walking sequence. The experimental results show that the physical parameters can be estimated efficiently, and the gait measurement (which consist of average left/right stride length and stride time, along with walking speed, for each short walking sequence) then obtains good accuracy. However, the complexity (synchronization and calibration are required) and computational cost may reduce the performance of real-time systems, especially for the 3D reconstruction process. In addition, the cost of equipment may also be an issue.

3. PROPOSED FEATURES

In order to compute parameters of motion, the walking person silhouette needs to be extracted. There are many methods developed to detect moving objects, in which background subtraction techniques [12] are appropriate for integrating into real-time systems. A tracking algorithm (e.g. particle filter [19]) can also be performed to improve accuracy. A sequence of obtained silhouettes will be used for calculating motion features.

At each frame, a sequence of characteristics is calculated from the corresponding silhouette. The obtained values are used to represent the posture of the walking person at this time.

3.1 Parameters of Approximated Ellipse

In many studies of gait analysis, estimating an approximated ellipse of the walking person and using its geometrical parameters are often implemented. In this approach, we determine the 95% confidence ellipse using the method presented in [13] that is briefly summarized below.

Consider a set of N pixels, the process estimates the ellipse which best fits the pixel locations with the smallest ellipse that encloses 95% of these points. The primary steps need to be performed are:

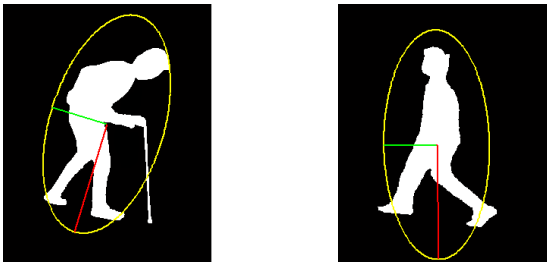
- Compute the average coordinates (x_m, y_m)
- Determine the covariance matrix Cov:

$$c_{xx} = \frac{\sum(x_i - x_m)^2}{N-1} \tag{1}$$

$$c_{yy} = \frac{\sum(y_i - y_m)^2}{N-1} \tag{2}$$

$$c_{xy} = c_{yx} = \frac{\sum(x_i - x_m)(y_i - y_m)}{N-1} \tag{3}$$

$$Cov = \begin{bmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{bmatrix} \tag{4}$$



(a) (b)

Figure 1: Approximated ellipse of human silhouette

- Calculate the eigenvalues (λ_1, λ_2) and unit eigenvectors (v_{11}, v_{21}) and (v_{12}, v_{22}) of Cov, in which $\lambda_1 > \lambda_2$.
- Finally, estimate parameters of approximated ellipse:
  - Center coordinates (x_m, y_m)
  - Semi-major axis has direction (v_{11}, v_{21}) and length $\sqrt{5.99\lambda_1}$
  - Semi-minor axis has direction (v_{12}, v_{22}) and length $\sqrt{5.99\lambda_2}$

The value 5.99 corresponds to the 95% confidence threshold of the χ^2 distribution with 2 degrees of freedom (for the 2 coordinates).

3.1.1 *Inclination Angle*

This information is helpful in detecting sudden incidents. For instance, the inclination angle of the approximated ellipse, which is the angle between horizontal and major axes, is very sensitive to differentiate fall from walk postures (e.g. [14]). Thus, body imbalance can be detected based on this characteristic. The range of this angle is $[0, \pi)$. This value in radians is determined using the following expression:

$$\theta = sin^{-1}(v_{11}) + \frac{\pi}{2} \tag{5}$$

3.1.2 *Axial Ratio*

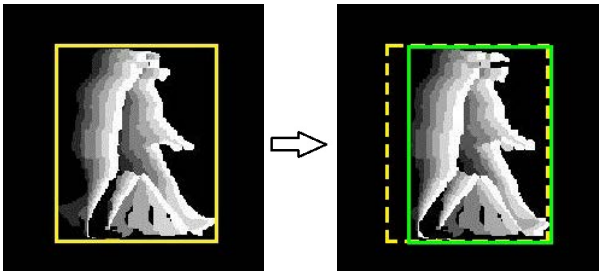
In many researches on activity detection and recognition, the aspect ratio of object's bounding box is commonly used. This feature is obtained by creating a rectangle enclosing the walking person, and the ratio of width to height is then determined. However, the aspect ratio of a walking person does not contain much information about the human posture as axial ratio of the approximated ellipse [2]. Another advantage of the axial ratio is that this characteristic is less affected by random noise in an image than the aspect ratio. Thus, in our study, the axial ratio is used instead of information estimated from the bounding box.

The axial ratio is computed based on eigenvalues determined in previous processing steps as follows:

$$R_{axial} = \sqrt{\frac{\lambda_2}{\lambda_1}} \tag{6}$$

3.2 Motion Variation

Anomaly in walking can also be detected based on the sudden change of velocity, which can be seen for neuropathic



(a) frame t^{th} (b) frame $(t+1)^{th}$

Figure 2: Two consecutive MHIs and corresponding bounding boxes

and hemiplegic patients [17], or left-right asymmetry or unbalance walk, for instance. Information about movement can be obtained via motion analysis. Thus we expect that estimating motion-related parameter(s) could detect abnormal gaits. In many studies, in order to detect and represent motion in a sequence of images, optical flow [3] is commonly used. However, this technique is not appropriate to be used in our approach because of its computational cost. In other words, optical flow is not well-suited to be integrated into a real-time system. In addition, some errors may be generated in case of large movement such as fall. In this approach, we suggest to collapse the motion information into an image using motion history image (MHI) [1], which does not require large memory and time for storing. In this image, the intensity of a specific pixel is a function of the motion history at this location, where brighter value corresponds more recent movements of a person during walking or other motions. In other words, the width of accumulated silhouettes represents recent motion of the person in a period of time.

Consider a sequence of silhouette images extracted from background subtraction techniques, motion regions $D(x, y, t)$ can be determined by applying frame differencing [6] on these silhouettes. Each pixel at time t (x, y, t) of the MHI function H_τ represents the temporal history of motion at that location, occurring during a fixed duration τ , with $\tau \in [1, n]$ for a sequence of n consecutive frames.

The MHI function H_τ is defined as follows:

$$H_\tau(x,y,t) = \begin{cases} \tau & if D(x,y,t) = 1 \\ max(0, H_\tau(x,y,t-1) - 1) & if D(x,y,t) \neq 1 \end{cases} \quad (7)$$

In this paper, we represent the velocity variation by the change of widths in two consecutive MHIs, which is obtained by a simple subtraction. In order to avoid the effect of distance from the walking person to the camera, we use a relative value, which is a ratio, instead of using the number of pixels directly. Thus, the obtained value (may be positive or negative) is divided by the height h, since the variation of height values is much smaller than the width. This feature is determined using following equation:

$$V_t = \frac{w_t - w_{t-1}}{h_t} \quad (8)$$

where h and w are the height and width of the accumulated silhouettes bounding box in the MHI, respectively. An illustration is shown in Fig. 2.

3.3 Signed Distance between Legs

For each normal walk cycle, the distance between the legs usually satisfies certain rules. Therefore, anomaly in walking (abnormal gait) can be assessed if these rules are not obeyed. According to [7] and anthropometric data, the legs region can be approximated based on the silhouette bounding box by cropping the bottom area at $h_l = 33\%$ of the full height h. In the cropped region, we consider each row in turn and compute the length (in pixels) of the gap between the legs. Two cases can occur (see Fig. 3): (1) there is a gap between the legs (e.g. at heel strike) or (2) the two legs are close together and there is no gap between the legs (e.g. at mid-stance). These cases are now further presented below. We consider only the cropped legs region in order to compute the distance.

3.3.1 Gap Exists

This case occurs when it has a significant gap between the legs, that is when the number of rows with gap is greater than 20% legs region height. The distance value is then calculated as the average of n largest gap values (to reduce noise). Based on our experiments, we have chosen $n = 5$ but n should not exceed 20% of the total number of lines where a gap exists in order to avoid insignificant values. A negative sign is assigned to the obtained result to discriminate it from the next case (no gap) below.

$$D = -\frac{1}{n}\sum_{i=1}^{n} g_i \quad (9)$$

where g_i is the i^{th} largest gap.

3.3.2 No Gap

In this situation, the distance is estimated by calculating the average number of foreground pixels per row, corresponding to the division of the number of legs pixels to the height of the legs region, which is a positive value.

$$D = \frac{n_{fp}}{h_l} \quad (10)$$

where n_{fp} and h_l are the number of foreground pixels and the height of the legs region ($h_l = 33\%$ of the silhouette height h), respectively.

Finally, in order to avoid the effect of distance between camera and walker, the value D is normalized as follows:

$$D_{ratio} = \frac{D}{h} \quad (11)$$

where h is the height of walking person silhouette.

3.4 Walk Cycle Representation

By combining the four characteristics described above, $(R_{axial}, \theta, V_t, D_{ratio})$, we obtain the feature vector of the frame which is being processed. In our approach, the detection process is performed on each walk cycle. Therefore, we separate the sequence of walking frames into cycles.

We define that a cycle begins at the time when the distance between the legs, D in Eq. 10, changes from negative to positive and ends at the frame just before the next cycle. In other words, each cycle is determined based on the sequence of signs of the distance feature, in which both ends are located by pairs of opposite signs ("−", "+") as illustrated in Fig. 3.

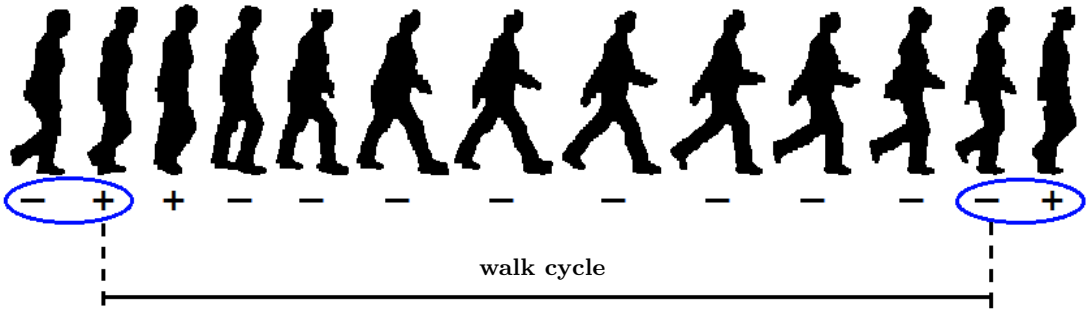


Figure 3: A walk cycle defined by two consecutive pairs of opposite signs ("−", "+"), in which the beginning and ending correspond to the plus sign of the first pair and the minus sign of the second pair, respectively.

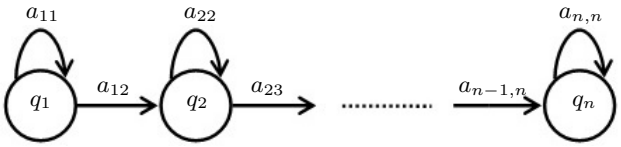


Figure 4: The HMM architecture with n states q_i, a_{ij} are the state transition probabilities from state i^{th} to state j^{th}. Emission probabilities of the codewords (observations) for each state are not shown.

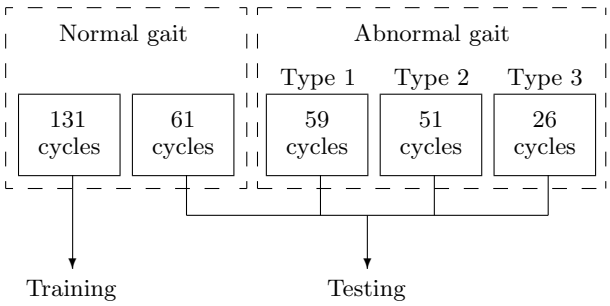


Figure 5: Separating data for training and testing

4. EXPERIMENTS

4.1 Building Model of Walk Cycle

In order to inspect the level of generalization of proposed characteristics, we performed modeling the normal gait in each walk cycle using a set of training samples; anomaly is then assessed based on the likelihood of normality with a threshold which is defined automatically during the training process. The vector $(R_{axial}, \theta, V_t, D_{ratio})$, which is calculated from each frame, is first converted into a codeword using a clustering technique. In other words, these vectors are grouped and each group corresponds to a codeword. Each sequence of such codewords, corresponding to each walk cycle, is considered as a feature vector. These vectors are then used for modeling the normal gait. In our experiments, the k-means technique was used to create the codeword because this is simple in calculation. We implemented the modeling process using Hidden Markov Model (HMM) due to the fact that this method is effective in describing the change of motion parameters in a sequence of frames. In detail, HMM is capable to represent the transition of different gait postures during walking. Our HMM architecture is shown in Fig. 4, in which the number of observations (not shown in the figure) is equal to the value k mentioned above.

4.2 Estimating the Normality Threshold

After finishing the training process, the values of the log-likelihood corresponding to the training normal patterns are computed. A normality threshold is then estimated based on the mean and standard deviation of obtained log-likelihood values, as the following equation:

$$thresh = \mu + C\sqrt{\frac{\sum_{i=1}^{n}(\mathcal{L}_i - \mu)^2}{n}} \qquad (12)$$

where $thresh$ is the threshold which needs to be determined, C is a constant, $\mathcal{L}$ is the log-likelihood value, μ and n are the average and number of training patterns, respectively. Note that μ and C are negative because the log-likelihood is less than zero.

If the log-likelihood of a walk cycle is lower than this threshold, we assume that an anomaly has occurred. In practical systems, more robustness could be added to this method by requiring a minimum number of abnormalities through a specified period of time to confirm the anomaly. This step was not necessary in this study.

4.3 Experimental Results

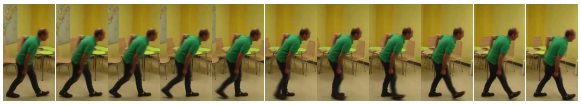
There are two sets of images used in our experiments: normal and abnormal gaits. Data in the first category, silhouette images, were collected from the CASIA gait database [18] (see Fig. 3), in which 56 sequences of silhouette images from 16 different people were used, corresponding to 192 cycles of normal gait. This dataset was separated into training and testing sets, in which the number of cycles is 131 and 61, respectively (see Fig. 5). The abnormal gaits in the other category were performed by three volunteers in a realistic environment at the University of Montreal (see Fig. 6), with the number of cycles corresponding to each subject in Table 1. The foreground extraction stage is performed using a combination of the Running Gaussian Average [12] and Frame Difference algorithms [6]. Our testing system was implemented in Matlab (MathWorks, USA) using the HMM toolbox [10].

The absolute values (for better visual illustration) of log-probabilities when testing normal and abnormal gaits with different values of parameter k and various numbers of HMM

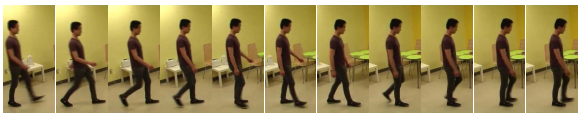
Table 1: Number of abnormal gait cycles performed by each subject

	Hunched back	Left-right asymmetry	Sudden motion variation
Subject 1	24	21	08
Subject 2	17	19	08
Subject 3	18	11	10

(a) left-right asymmetry



(b) hunched back



(c) sudden motion variation

Figure 6: Three abnormal gaits performed for testing

states is shown in Fig. 7. In our experiments, the constant C was set to -3.5.

We can see that normal and abnormal gaits were separated very clearly. In experiments, the average processing time was about 78 milliseconds per frame. Thus, our solution is expected to be appropriate for real-time systems, where the processing speed could be several times faster than our implementation in Matlab.

5. CONCLUSION AND DISCUSSION

In this paper, four silhouette-based characteristics are presented to support solving the gait analysis problem using only one camera. In experiments, they are used to detect three types of abnormal gait, which consist of sudden motion variation, left-right asymmetry and hunched back. At first, the movement is estimated by creating the motion history image. Four mentioned features are then determined to represent the posture of walking person in each captured frame. Finally, the k-means clustering technique is combined with HMM for modeling normal gaits. A normality threshold, which is defined during the training process, is used to assess anomalies. Experimental results show that proposed features provide high accuracy in gait analysis, and they can be integrated into real-time systems. The main contribution of our study is that the proposed features can represent the posture and motion of a walking person at a specific time using a low-cost device instead of a set of complicated calibrated cameras or sensors. We expect that different abnormal gaits can be modeled according to the described processing. The limitation of the proposed approach is that the walk cycle segmentation may not give desired accuracy when the movement in baseline sequences is either towards or away from the camera.

As further work, some other characteristics would be studied and assessed. In order to choose best features, the importance of each one can be rated using principal component analysis method [8]. Besides, our experiments will be developed for recognizing more specific types of abnormal gait, such as hemiplegic, diplegic, choreiform, and parkinsonian gaits [17]. In addition, detecting normal gaits with different machine learning algorithms, such as dynamic time warping [9], will be implemented and compared to choose the most appropriate model for each abnormal gait problem.

6. ACKNOWLEDGMENTS

This work is supported by DATIC - Danang University of Science and Technology, Natural Sciences and Engineering Research Council of Canada (NSERC) and the project "Visually impaired people assistance using multimodal technologies" funded by Vlaamse Interuniversitaire Raad (VLIR) in the framework of VLIR's Own Initiatives' Program 2012 under the grant number ZEIN2012RIP19. Portions of the research in this paper use the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences.

7. REFERENCES

[1] M. A. R. Ahad. *Motion History Images for Action Recognition and Understanding.* Springer Briefs in Computer Science. Springer, 2013.

[2] T. Brinkhoff, H.-P. Kriegel, and R. Schneider. Comparison of approximations of complex objects used for approximation-based query processing in spatial database systems. In *Data Engineering, 1993. Proceedings. Ninth International Conference on*, pages 40–49, Apr 1993.

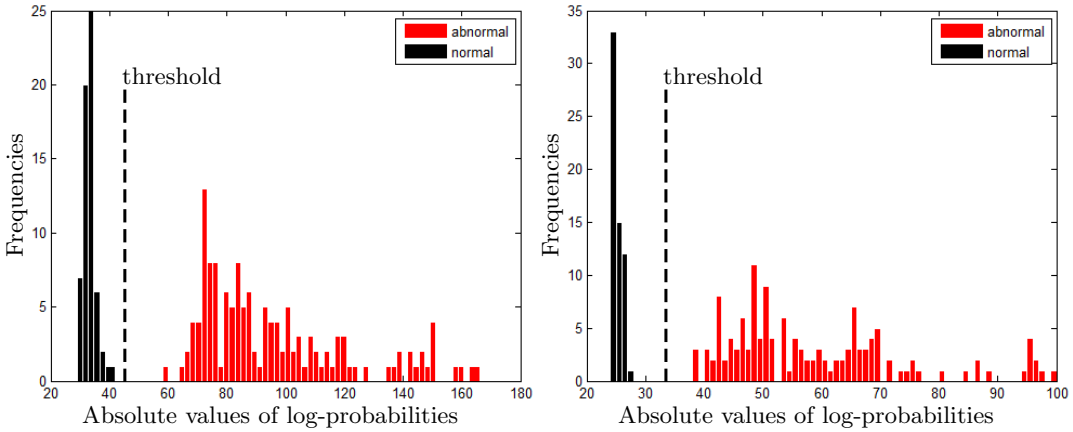
[3] D. J. Fleet and Y. Weiss. Optical flow estimation. *Mathematical models for Computer Vision: The Handbook*, 2005.

[4] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster. Full body gait analysis with kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1964–1967, Aug 2012.

[5] B. Greene, A. Donovan, R. Romero-Ortuno, L. Cogan, C. N. Scanaill, and R. Kenny. Quantitative falls risk assessment using the timed up and go test. *Biomedical Engineering, IEEE Transactions on*, 57(12):2918–2926, Dec 2010.

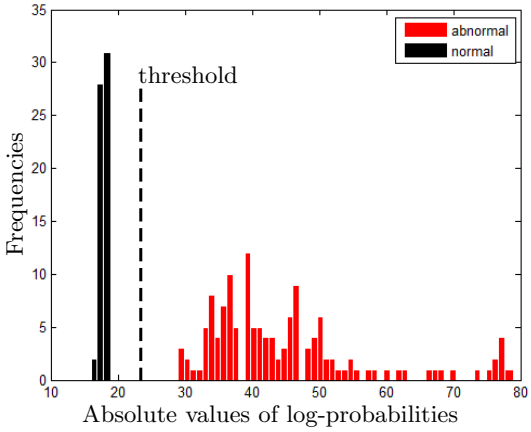
[6] R. Jain and H.-H. Nagel. On the analysis of accumulative difference pictures from image sequences of real world scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-1(2):206–214, April 1979.

[7] F. Jean, R. Bergevin, and A. Albu. Body tracking in human walk from monocular video sequences. In *Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*, pages 144–151, May 2005.

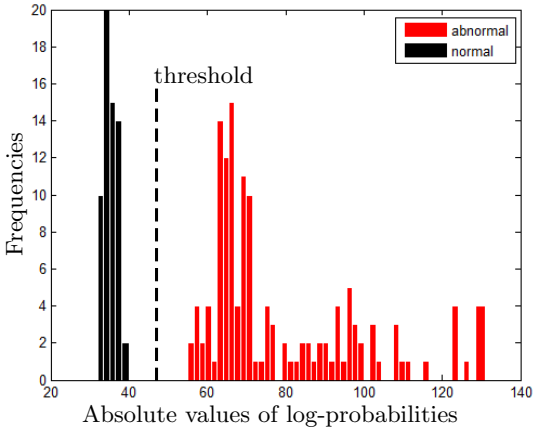


(a) 5 states and 32 observations (k=32)

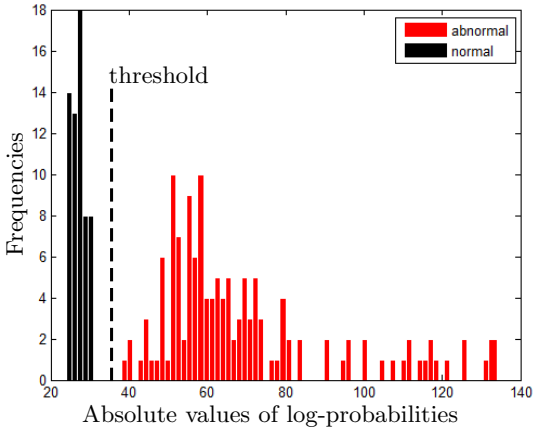
(b) 5 states and 16 observations (k=16)



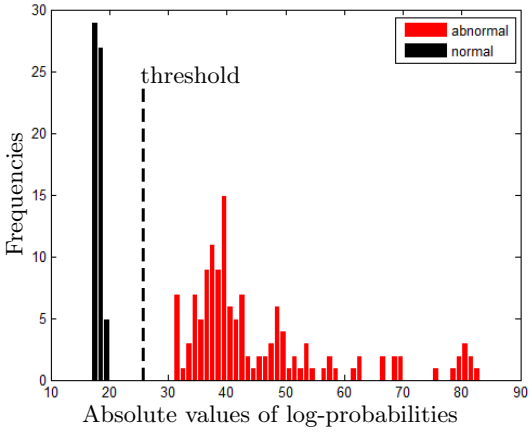
(c) 5 states and 8 observations (k=8)



(d) 4 states and 32 observations (k=32)



(e) 4 states and 16 observations (k=16)



(f) 4 states and 8 observations (k=8)

Figure 7: Absolute values of log-probabilities, $|\mathcal{L}|$, calculated from HMM output for normal and abnormal gaits. Black and red color correspond to normal and abnormal gaits, respectively. There are 6 graphs, corresponding to the testing results of 6 HMMs (with different numbers of states and observations). The threshold is defined in the training process. In all experiments, the constant C was -3.5 and there was no case of false detection (false positive) or omission (false negative).

[8] I. T. Jolliffe. Principal Component Analysis and Factor Analysis. In *Principal Component Analysis*, Springer Series in Statistics, chapter 7, pages 150–166. Springer New York, New York, 2002.

[9] M. Muller. *Information retrieval for music and motion.* Springer, 2007.

[10] K. Murphy. Hidden markov model toolbox for matlab. http://www.cs.ubc.ca/ murphyk/Software/HMM/hmm.html, 2005. Accessed: 2014-10-14.

[11] H. A. Nguyen, E. Auvinet, M. Mignotte, J. de Guise, and J. Meunier. Analyzing gait pathologies using a depth camera. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4835–4838, Aug 2012.

[12] M. Piccardi. Background subtraction techniques: a review. In *SMC (4)*, pages 3099–3104. IEEE, 2004.

[13] W. C. Rose. Mathematics and signal processing for biomechanics. udel.edu/biology/rosewc/kaap686/reserve/cop/center of pressure.html, 2010. Accessed: 2014-10-14.

[14] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Fall detection from human shape and motion history using video surveillance. In *Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on*, volume 2, pages 875–880, May 2007.

[15] E. Stone and M. Skubic. Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing. In *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pages 6491–6494, Aug 2011.

[16] E. Stone and M. Skubic. Unobtrusive, continuous, in-home gait measurement using the microsoft kinect. *Biomedical Engineering, IEEE Transactions on*, 60(10):2925–2932, Oct 2013.

[17] University of Utah and University of Nebraska. Neurologic exam. http://library.med.utah.edu/neurologicexam, 2013. Accessed: 2014-10-14.

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

[19] T. Yang, P. G. Mehta, and S. P. Meyn. Feedback particle filter. *IEEE Trans. Automat. Contr.*, 58(10):2465–2480, 2013.