

Affective Gait Recognition and Baseline Evaluation from Real World Samples

Vili Kellokumpu, Markus Särkiniemi, Guoying Zhao

Center for Machine Vision and Signal Analysis
University of Oulu
{kello, misarkin, gyzhao}@ee.oulu.fi

Abstract. Over the years a lot of research efforts have been put into recognizing human emotions from facial expressions. However, in many scenarios access to suitable face data is difficult, and therefore there is a need for methodology that can be used when people are observed from a distance. A potential modality for this is human gait. Early attempts to recognize human emotion from gait have been limited to acted data. Furthermore, in these approaches the data has been captured in controlled settings. This paper presents the first experiments for automated affective gait recognition using non acted real world samples. A database of 96 subjects affected by positive or negative feedback is collected and two baseline methods are used to recognize the affective state of a person. The baseline results are promising and encourage further study in this domain.

1 Introduction

Human emotion recognition is one of the most active areas of computer vision research. It is driven by its importance in human-machine interfaces. It also has application potential in other domains such as social sciences, health care and security. Most of the research efforts so far have focused on facial expression recognition. This limits the applicability of emotion recognition techniques because suitable frontal face data of sufficient quality cannot be captured when subjects are observed from a distance.

It is evident that humans can observe a great deal of information from human movement. For example, recognizing friends based on the way they move even from a distance is very natural. In computer vision research this phenomena is used in gait recognition: human movement is used as biometric signature to identify people based on the way they walk. In gait recognition the aim is to discriminate between the walking styles of different people. However, human walking movement conveys even more subtle information, such as cues about feelings or mood of a person. It has been shown that human observers are able to identify emotions from gait [6]. Another study suggests that humans are able to recognize emotions using walking speed and posture [8].

Earlier attempts at automated emotion recognition from human body movement have been limited to analyzing acted data. In this work, our hypothesis is



Fig. 1. When subjects are observed from a distance and facial expressions are not available, analyzing gait offers additional cues for estimating the affective state of a person.

that we can automatically recognize the affective state of a person by observing their gait using real non acted data. This is in contrast to previous works that perform emotion analysis from subjects artificially behaving and posing their feelings. So far the availability of non acted data has been a limiting factor and in this work tackles this limitation by collecting data from tv-broadcast material.

The rest of the paper is organized as follows. Related work on gait is reviewed in Section 2. The data collection for our experiments is described in section 3. Section 4 describes the chosen baseline methods and section 5 presents the obtained experimental results. Finally, Section 6 concludes the paper.

2 Related Work

A lot of research on recognizing human emotion from gait has been done over the years. The work is typically restricted by data that is based on acted emotions. Montepare et al. showed that identifying emotions from gait is possible for human observers [6]. In the experiments ten subjects observed five walkers that were instructed to express four different emotions: sadness, anger, happiness and anger. The results suggest that gait characteristics such as the amount of arm swing, stride length, heavyfootedness, and walking speed differentiated the

emotions expressed by the walkers.

Crane and Gross [3] used autobiographical memory tasks to record video and motion capture data of people feeling anger, sadness, content, joy or neutral. In their data collection participants completed a self-report of felt emotions. The data was then viewed and classified by human observers. It was argued that emotions can be successfully elicited in the laboratory setting, emotions can be recognized in the body movements of others, and that body movements are affected by felt emotions.

Roether et al. [8] investigated the influence of posture and dynamic cues for the expression and recognition of emotions in gait. To identify emotion-specific features they motion-captured the neutral and emotionally expressive (anger, happiness, sadness, fear) gaits of 25 subjects. They found that in terms of kinematics, size and speed of movements were among the most important features driving the perception of emotional gaits.

Instead of gait, Atkinson et al. [1] used a database of people expressing emotions or performing simple actions with whole-body movement. The actors in the database were free to interpret and express the emotions as they saw fit, with only minimal guidance as to the sorts of situations in which people might experience those emotions. Human observers were then shown either a full-light or patch-light stimuli of the movement. It was then discovered that kinematics help distinguish emotions expressed by body gestures

It is evident that earlier work suffers from a low number of subjects. Also, the data is not typically recorded in an authentic manner, i.e. the subjects are asked to act as if feeling an instructed emotion.

Apart from emotion recognition, a lot of work has been devoted for human gait analysis in computer vision. Many methods have been proposed for gait recognition aiming at biometric recognition [11]. The methods and databases used for gait recognition traditionally rely on using silhouettes as input.

One of the simplest yet effective features for gait recognition is the average silhouette proposed by Liu and Sarkar [5]. In this approach silhouettes are aligned and averaged into a single image template without considering the dynamics between frames. On the contrary, Kellokumpu et al [4] treated the human gait pattern as a dynamic texture. The method calculates Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) and thus utilizes both appearance and dynamics of human movements for gait recognition. The average silhouette and dynamic texture based methods are selected as baseline methods in this work and are briefly introduced in Section 4.

3 Data from Real Life Samples

As discussed above, earlier work suffers from a low number of subjects. Furthermore, the subjects are asked to act as if feeling an instructed emotion when data is captured. To overcome these limitations, we collect affective gait data from 96 subjects after they have received positive or negative feedback. Table 1 compares

Table 1. Comparison of emotional databases in literature. Symbols: X = No, Y = Yes.

	Subjects	Emotions	Sensors / Markers	Acted
Atkinson [1]	32	5	X	Y
Pollick et al. [7]	2	19	Y	Y
Montepare et al. [6]	5	4	X	Y
Wallbott [10]	16	14	X	Y
Crane & Gross [3]	42	5	Y	Y
Roether et al. [8]	25	5	Y	Y
Our work	96	2	X	X

the databases used for affective gait analysis. The data used in our experiments is made publicly available

The bases for the data collection are various TV-shows where candidates audition or present their talent in order to compete for attractive prizes. After an audition, the judges of the show determine if a candidate will advance in the competition or if he/she is eliminated. The starting point of the data capture is the moment when the result of a performance is published and the candidate leaves the stage affected by the positive or negative verdict.

The data is collected from different shows and different seasons. While the clips are selected so that the subject is captured from the side view, the capturing conditions vary between shows and seasons. Furthermore, the collected samples include variation also within the videos as they are captured with cameras mounted on moving studio cranes.

Altogether videos of 96 subjects are collected. The data contains an equal number of samples (48) of both positive and negative class. The cuts in the broadcast data are typically rather fast and capturing long gait sequences is challenging. Therefore, the number of half cycles, i.e. steps, in the collected samples varies from one to four.

Since most of the methods and databases used for biometric gait recognition rely on silhouettes, we also adopt this approach. The TV broadcast data that is used for the data collection is captured with moving cameras and it may also contain dynamic lighting variations. This makes automatic extraction of silhouettes difficult. Instead, we use user assisted graph cut algorithm [2] to do the segmentation. The resulting silhouettes are realistic and comparable to what a background subtraction algorithm should produce from a video captured with a static camera.

Figure 2 illustrates the user input for an example frame and gives a few examples of the silhouette quality. It is evident that the data contains a lot of natural variation in the subjects body shape, size, clothing etc.

One limitation of this approach to data collection is that the ground truth for the affective state can only be categorized either as positive or negative instead of a more detailed categorization. As discussed above, in the acted scenarios a



Fig. 2. Illustration of a the user annotation for segmentation and example silhouettes from the database

more detailed ground truth label may be set. Also, the data collection does not allow to estimate the intensity of the felt reaction.

4 Methods

Since gait recognition methods are capable of capturing small differences in human movement to recognize people from the way they walk, we adopt a similar approach in this work. We use existing features originally proposed for gait recognition and apply them with a Support Vector Machine to learn a classifier for affective gait recognition.

In the preprocessing stage, the silhouette sequence is first normalized and stacked into a 3D volume. In feature extraction either average silhouette or LBP-TOP features are computed. Finally the sequence is classified based on the features into a positive or negative class.

4.1 Preprocessing

We treat gait as a spatiotemporal volume. In order to build the volume, silhouettes are scaled and aligned. First, the silhouette size is normalized by scaling it into a fixed height. The height is set to 128 pixels similar to [5, 9]. Then, vertical alignment is performed by aligning the silhouettes based on their centroids. Finally the silhouettes are stacked into a 3D volume.

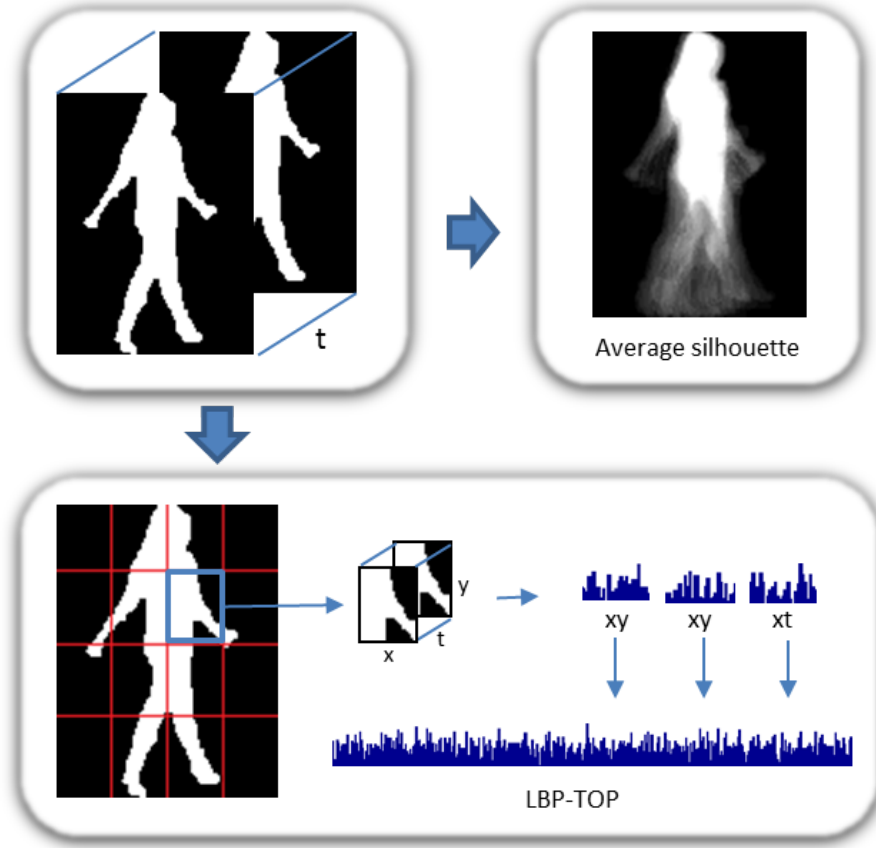


Fig. 3. Illustration of preprocessing and feature calculation for average silhouettes and LBP-TOP

4.2 Features

Based on the spatiotemporal volume, two feature representations are calculated: average silhouette [5] and LBP-TOP [4]. The reason for choosing these features is twofold. Firstly, the average silhouette is extensively used as a baseline method in gait recognition. Secondly, average silhouette only captures appearance information whereas LBP-TOP representation captures motion information in addition to appearance information, which is important according to the previous studies [6] and [8].

Average silhouette is computed simply as an average along time axis for each pixel. The dynamic texture features are extracted by calculating the LBP texture features on three orthogonal planes of the volume: xy , xt and yt . Figure 3 illustrates the feature calculation for both feature types from the spatiotemporal volume.

Table 2. Classification results for affective gait recognition with average silhouettes and LBP-TOP.

Method (feature + classifier)	performance
Average silhouette + NN	59%
LBP-TOP + NN	64%
Average silhouette + rbf SVM	72%
LBP-TOP + rbf SVM	78%

We follow the implementation details as presented in [4, 5]. The parameters we used for LBP-TOP on all planes are as follows: radius is set to one and the number of sampling points is eight. Also, uniform patterns are used. It should be noted that larger values for radiuses were used in [4], but since our data volume size is much smaller, small radius is more practical. Furthermore we divide the volume into 4x4 blocks to preserve more spatial structure.

4.3 Classification

Since the aim of gait recognition is to differentiate people from each other, it means that matching sequences directly based on the gait features will result in matching sequences to people moving similarly and not to people with similar feelings. For affective gait recognition this means that such simple matching should fail to work. Therefore we need to learn the differences in the feature representation that indicate differences in the affective state.

For classification we consider two methods. Firstly nearest neighbor (NN) classification is considered as a sanity check. Euclidean distance is used as a distance measure for the average silhouettes. For LBP-TOP histogram intersection is used as a similarity measure. Secondly, in order learn the important cues that enable emotion recognition, we use Support Vector Machines to learn the difference between positive and negative classes.

5 Experiments

We used the leave one out cross validation to test the average silhouette and LBP-TOP methods for affective gait recognition on the collected data. Thus one sample is used for testing while the others are used for training. This is repeated for all samples. The results are summarized in Table 2.

Figure 4 shows the SVM classification results with various parameters (powers of 2). It can be seen that fairly good results can be achieved with a range of parameter combinations.

Nearest neighbor based classification typically used in gait recognition leads to matching people to similar looking people. A biometric cue should be invariant to emotion, and this is a likely reason why the nearest neighbor classification basically fails to generalize for emotion recognition and achieves only slightly above random performance. In contrast to this, SVM seems to be able to do

better in learning the differences caused by the affective state of the subject from the training data.

Also an interesting observation from the results is that it seems that the dynamic texture based description performs better than the average silhouette. It seems that the dynamics of gait are a good cue for analyzing the affective state of a person. This is in line with the observations of Montepare [6] and Roether [8].

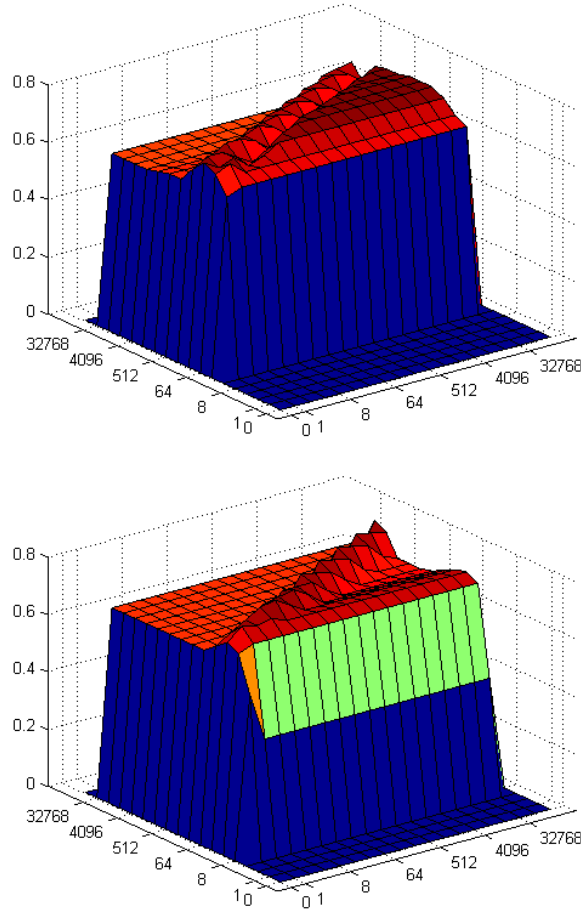


Fig. 4. (top) average silhouette, (bottom) LBP-TOP. Performance plots with various parameter combinations for rbf SVM

6 Conclusions

This work describes the first video based experiments for recognizing affect from human gait patterns based on real world data. Furthermore we provide the community with the first affective video gait dataset captured in real world conditions.

Although the experimental results suggest that it is indeed possible to recognize affect from human gait, the problem is far from solved. One important aspect to consider as future work is the evaluation of the level of the observed affect and not just the classification to positive and negative. Also, for practical applications, it would be useful to study the problem using the image data directly instead of silhouettes.

Acknowledgement. This work was sponsored by the Academy of Finland, Infotech Oulu and Nokia Visiting Professor grant.

References

1. Atkinson, A. P., Tunstall, M. L., Dittrich W. H.: Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, vol. 104, pp. 59-72, 2007.
2. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9), pp. 1124-1137, 2004
3. Crane, E., Gross, M.: Motion capture and emotion: affect detection in whole body movement. *Affective Computing and Intelligent Interaction*, vol. 4738, pp. 95-101, 2007.
4. Kellokumpu, V., Zhao, G., Li, S., Pietikäinen, M.: Dynamic texture based gait recognition. *International Conference on Biometrics, ICB*, 2009
5. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: averaged silhouette *Proceedings of the 17th International Conference on Pattern Recognition*, vol.4, pp. 211-214, 2004.
6. Montepare, J.M., Goldstein, S.B., Clausen, A.: The identification of emotions from gait information *Journal of Nonverbal Behavior*, 11(1), 33-42, 1987.
7. Pollick, F. E., Paterson, H. M., Bruderlin, A., Sanford, A. J.: Perceiving affect from arm movement *Cognition*, vol. 82, pp. B51-B61, 2001.
8. Roether, C. L., Omlor, L., Christensen, A., Giese, M. A.: Critical features for the perception of emotion from gait *Journal of Vision*, vol. 9, pp.1-32, 2009.
9. S. Sarkar, Phillips, P. J., Liu, Z., Robledo, L., Grother, P., Bowyer, K. W.: The Human ID Gait Challenge Problem: Data Sets, Performance, and Analysis *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2), pp. 162-177, 2005.
10. Wallbott, H. G.: Bodily expression of emotion *European Journal of Social Psychology*, vol. 28, pp. 879-896, 1998.
11. Wang, J., She M., Nahavandi, S., Kouzani, A.: A review of vision-based gait recognition methods for human identification *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 320-327, 2010