# Affective Computing For Empathic Behaviour Change - Emotion detection via body language

Mathias Fuchs
University of Bern
Master of computer science
Matr. Nr.: 09-923-764
Email: fuchsmat@students.unibe.ch

***Abstract*—There are many reasons why it is crucial for a computer to be able to understand in which emotional state the user is in e.g. Lie detection, help people with aspergers to recognize emotions, machines that interact like humans etc. The research area of affective computing strives to achieve this feat. 55% of human communication is made up of non-verbal signals [?], [?]. This is why body language plays an important role to help detect the emotional state of a person. It is especially useful in situations where emotion detection through face or voice is failing. For example recognizing emotions from a greater distance, without being able to see the face or hear the person [?]. In this paper I give an overview over the body language recognition approaches done today and propose a model which detects the emotional state of a person based on the way they walk.**

## I. Introduction

For a long time people were convinced that human behaviour is "all nurture and no nature" [?]. However already Darwin [?] suggested that along with the facial expression, the human body movements and the gestures also represent the state of mind and the corresponding emotions of humans. We know today that body language plays a very important role to understand the affective state of a person [?], [?]. Surprisingly only 7% of human communication are made of words and 55% are made up of non-verbal communication [?], [?], [1].

The idea of this paper was to focus on micro expressions in body language. However there are no bodily micro expressions as in facial micro expressions. A micro expression is a "very fast facial movement lasting less than one-fifth of a second" [?]. Body language in comparison can be subconscious but it can also faked more easy than a facial micro expression. This is the reason that I focus on the emotion detection via body language in this paper.

There are various uses for emotion detection by body language. Some of those are detecting the *affective state* of a person, *lie detection*, the degree of *accessibility* towards another person etc. Indicators for interpreting body language can be *body position and distance* [?], [?], *body movement* [?], [1] and *hand form* [?], [?]. This list is not necessarily concluding, but those are the parts that this paper focusses on.

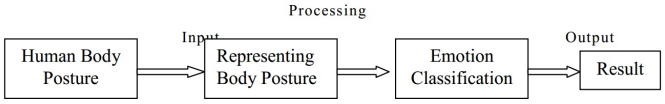


Fig. 1. The phases of a bodily emotion detection system [1]

In the field of affective computing however, it is not only a challenge to interpret the body language, but also to detect the position of the human posture and get useful data out of it. This leads to the two main challenges: *1. Detection of the body posture/ movement* and *2. How to interpret the posture representation* ( see Figure 1 [1]). However those two steps are combined to one in certain machine learning approaches [?], [?], [?].

In the following work I will first give an overview over the psychology behind emotion detection through body language. After that I present papers that provide approaches for both, the detection of the human posture as well as for interpreting the found results. Finally I propose a theoretical model of how to detect emotion by human gait in real-time.

## II. Emotions through body language

As already stated, 55% of our communication consists of non-verbal cues, like body language. The expression of emotions has been studied extensively [?], [?], [?], [?]. According to Eckman [?] there are 6 basic categories of emotions:

- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise

Those emotions seem to be universal and the same across different cultures [?]. The motivation for those 6 categories goes back to the theory, that the same facial muscles are used for the same emotions across different cultures [?]. Those emotions are however only universally the same when it comes to facial expressions, but when it comes to body language, especially to human gait, there can be great differences between the individual humans. This can lead to some problems which we will see in section VI. This categorical notion of emotions allows systems to more easily categorize new samples to one

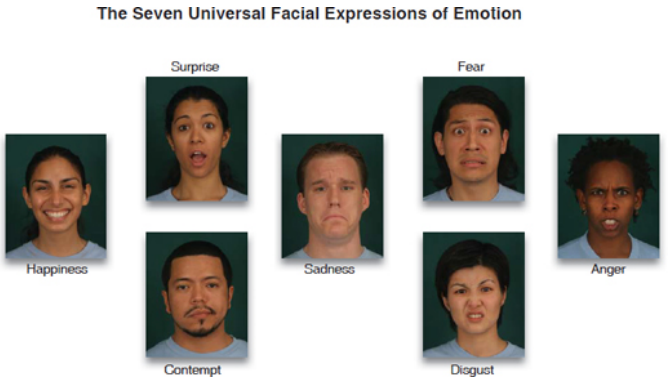


Fig. 2. Facial expression of the six basic emotions ( and contempt)


Fig. 3. Bodily signs of the six basic emotions [**?**]

of those existing categories, instead of having to create new categories from scratch [**?**]. We can find clear signs of all those emotions in our faces (see Figure 2[1]). Recently more research has been done in the field of detecting emotions through body language, like body movement and body pose [**?**], [**?**], [**?**]. In Figure 3 we can see a representation of the body pose for the 6 basic emotions.

The research is not conclusive about how to interpret body language. In Figure 4 we can see a possible way on how to interpret certain bodily signs based on the body position and body movement [1].

| Emotion | Body Posture |
|---|---|
| Anger | Head backward, no chest backward, no abdominal twist, arms raised forwards and upwards, shoulders lifted. |
| Joy | Head backward, no chest forward, arms raised above shoulder and straight at the elbow, shoulders lifted. |
| Sadness | Head forward, chest forward, no abdominal twist, arms at the side of the trunk, collapsed posture. |
| Surprise | Head backward, chest backward, abdominal twist, arms raised with straight forearms. |
| Pride | Head backward or lightly tilt, expanded posture, hands on the hips or raised above the head. |
| Fear | Head backward, no abdominal twist, arms are raised forwards, shoulders forwards. |
| Disgust | Shoulders forwards, head downwards. |
| Boredom | Collapsed posture, head backwards not facing the interlocutor. |

Fig. 4. A table of bodily signs for different emotions [1]

The distance between two people can also show the attitude between them. If the distance is less than the local social norms permit, we can infer a negative attitude [**?**]. However this has to be treated carefully, because the context is important. For example, if two people that are in a romantic relationship with each other, then less distance is not a negative sign. It has also been found, that the meaning of distance depends on the expectancies of the subjects. If the subjects expect a positive interaction, then a close distance is a positive sign, and if they expect a negative interaction, then a close distance is a negative sign [**?**]. Examples of different approaches will be shown in section III.

In addition larger body movements like the gait of a person, can also tell us something about the emotional state. It has been found that the arm swing, the stire length and the heavy-footedness of a persons gait, can reveal their emotional state [**?**]. For example angry movements tend to be larger, faster and seem jerky, compared to normal movements [1]. Whereas fearful and sad movements tend to be less energetic, smaller and slower [1]. Examples of different approaches will be shown in section VI.

The form of the hand can also give information about the emotional state of a person. For example open palms could mean pleasure/ openness, closed hands towards the chest could mean a sense of pride and clenched fists could mean anger [**?**], [**?**]. Kipp and Martin (2009) et al. [**?**] even stated, that for right-handed people, the right hand was more used when experiencing anger and the left hand more, when their feelings were relaxed and positive. Examples of different approaches will be shown in section V.

In the following sections I summarize a few approaches that extract and/or interpret body language in different ways.

## III. Emotion detection through body position and movement

### *A. Lie Detection based on Facial Micro Expression, body Language and Speech Analysis*

A very interesting approach was done by Barathi [**?**].

*1) Body pose extraction:* To extract the body poses from videos the Limb Action Model Converter [**?**] has been used. This converter uses Microsoft Kinect as a base. The Limb Action Model extracts 10 limbs from a body posture: ”Spine to center shoulder, center shoulder to head, left/right shoulder to left/right elbow, left/right elbow to left/right wrist, left/right hip to left/right knee and left/right knee to left/right ankle” [**?**]. In Figure 5 we can see how the posture is represented after the extraction.

[1]https://hubpages.com/health/Facial-Expressions-Emotions-and-Feelings

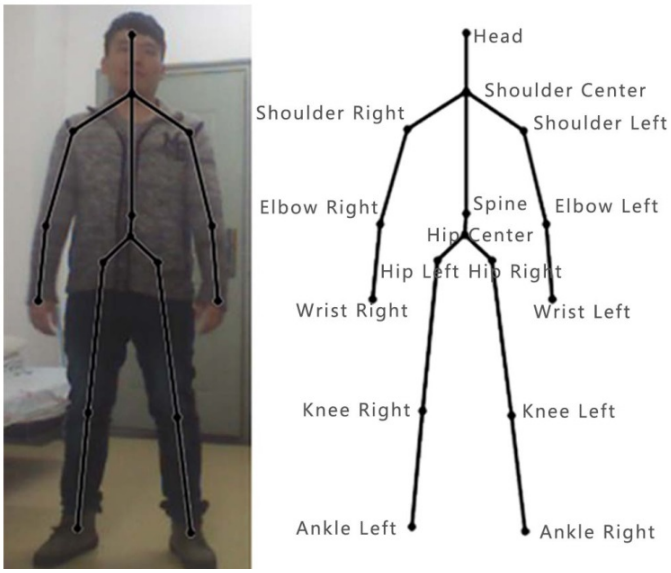


Fig. 5. Skeleton joints extracted with the Limb Angle Model [**?**]

*2) Interpretation:* As the paper focusses on lie detection, they used the following signs for lying [**?**], [**?**], [**?**]:

- Increasing hand to face/mouth gestures
- Nose touching: Because of an adrenaline rush, the capillaries open up, which causes the nose to itch
- Place the hand close to or over the mouth
- Small gestures like lip biting, hands rubbing, fidgeting
- Clenched fist, crossed arms

According to those specifications the system is trained with images of liars that were exhibiting those typical body language cues for lying and pictures of people who were not lying. All the images are subjected to the Limb Action Model Converter. Finally they clustered the converted pictures with k-means.

Sadly they did not provide an evaluation for the body language part of their method.

### *B. EDBL - Algorithm for Detection and Analysis of Emotion Using Body Language*

*1) Body pose extraction:* The EDBL approach by Singh et al. [**?**] relies on a pose estimation which is using postelets for humnan parsing [**?**]. Different postelets are extracted from images. A linear SVM classifier is trained for detecting the presence of each postelet. In the end we get a complete model of the human. See Figure 6 for a graphical representation of the process.

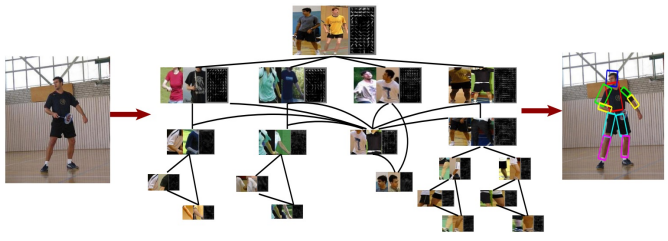

Fig. 6. Graphical illustration of the postelet model [**?**]

Based on the postelet representation a line graph (see Figure 7) of the extracted pose is created and used for the interpretation of the body posture.

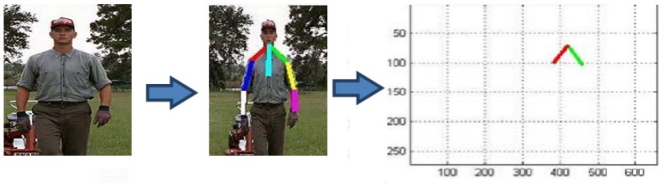


Fig. 7. EDBL stick pose representantion [**?**]

*2) Interpretation:* To figure out the emotional state of a person, the position of the shoulders is interpreted. There is a differentiation between three distinct types of poses: 1. normal/ calm, 2. confused, amazed or in doubt and 3. depressed/ not interested. Figure 8 shows images of the three different types.

First the slope of the normal shoulder position $\theta$ is calculated. An initialisation pose is needed in order to do this. Then it is compared to other images of the same person and if the slope is bigger than the slope of $\theta$ then it is a sign of doubt, confusion or being amazed. If the slope is smaller than the slope of $\theta$ then it is a sign of a depressed, not-interested or lazy emotional state.

This approach also offered no evaluation to validate it.


Fig. 8. Meaning of shoulder positions. (a) shows a person in normal or calm position, (b) shows a person in confused or amazed state and (c) shows a person in a depressed or not interested pose. [**?**]

## IV. Affect Detection from Body Language during Social HRI

*1) Body pose extraction:* This approach of McColl et al [**?**] also uses Microsoft Kinect as a base to extract the body pose. It creates an ellipsoid model as seen in Figure 9. It works by extracting body poses with Kinect. Then the extracted body poses are observed to see if a static pose is displayed. Once a static pose is identified the segmented body parts are fitted with ellipsoids. The static body poses are further explained in subsubsection IV-2.

Fig. 9. An ellipsoid model od the human body built on the basis of Microsoft Kinect. [**?**]

*2) Interpretation:* To figure out the emotional state of the subject, static body poses are used. The poses are based on the Nonverbal Interaction States Analysis of the Davis Nonverbal States Scale [**?**]. Body angle, trunk lean and arm position are evaluated. The resulting metrics are the following:

- **Three different body angles:** *Toward(T)*: $0° - 3°$ angle from the robot, *Neutral(N)*: $3° - 15°$ angle from the robot, *Away(A)*: $> 15°$ from the robot.
- **Trunk lean:***Upright*: The shoulders are over the hips, *Forward/ backward lean*: The shoulders are closer/ farther away than the hips in relation to the robot, *Right/ left lean*: The right/ left shoulder is tilted past the right/ left hip.
- **Arm positions:** *T*: The arms are closer to the robot than the upper trunk, *A*: The arms are farther from the robot than the trunk, *N*: else.

According to those metrics, 4 accessibility levels have been defined (see Figure 10). Level I-IV, where I is least accessible and IV is the most accessible state. The arm orientation is used for a finer scaling of the accessibility levels.

| Trunk Orientation | Accessibility Level | Arm Orientation | Finer-Scaling |
|---|---|---|---|
| Upper/Lower trunk: T/N or N/T combined with upright or forward leans, T/T with all possible leans | IV | T | 12 |
| | | N | 11 |
| | | A | 10 |
| Upper/Lower trunk: T/N or N/T except positions that involve upright or forward leans | III | T | 9 |
| | | N | 8 |
| | | A | 7 |
| Upper/Lower trunk: N/N, A/N, N/A, T/A, A/T with all possible leans | II | T | 9 |
| | | N | 8 |
| | | A | 7 |
| Upper/Lower trunk: A/A with all possible leans | I | T | 3 |
| | | N | 2 |
| | | A | 1 |

Fig. 10. Different accessibility levels defined to determine the level of accessibility of the person against the system [**?**]

This approach classified the accessibility correctly with an accuracy of 88%. This a solid result. However the approach is very limited in its application and needs a highly specific set-up.

### *A. Recognizing Emotions Expressed by Body Pose: a Biologically Inspired Neural Model*

This approach by Schindler et al. [**?**] aims to categorize images into the 6 basic emotions [**?**] *angry, disgusted, fearful, happy, sad, surprised and neutral*. The Idea is to model the visual pathway of recognition of a human. This means the process in which the visual cortex extracts cues from the visual input to determine an emotion [**?**]. The model is inspired by [**?**], [**?**]. This approach is interesting, because it combines various commonly used and efficient techniques like MAX pooling, PCA and SVM in a hierarchical way that follows the organization of the human visual cortex.

The visual cortex consists of the following areas: *Primary visual cortex (V1), secondary visual cortex or prestriate cortex (V2), visual area (V3), visual area (V4) and middle temporal visual area (V5)*[2]. The model is built according to those areas (see Figure 11).

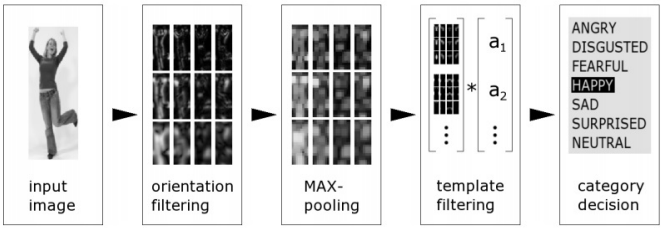


Fig. 11. "Illustration of the neural model. From the raw image on the retina, local orientation is extracted (area V1), pooled over spatial neighborhoods (V2/V4), and filtered with learned complex features (V4/I).T The filter response serves as input into a discriminative classifier (IT). Parameters were chosen for illustration purposes and are different from the actual implementation." [**?**]

The recall of the system achieved 82% whereas the recall for human testers was 87%. This means that on average the model only miss-classified 5% more than the human testers. Interesting to see is the direct comparison against the human for the single emotions that were classified (see Figure 12). Anger and disgust are the most difficult to classify, for the machine as well as for the humans. However the classification gap between the machine and the human is the biggest for anger, disgust and fear.

This approach is quite interesting, because it performs fairly well, even though the training set only consists of 50 actors that created a total of 696 images.

[2]https://en.wikipedia.org/wiki/Visual_cortex

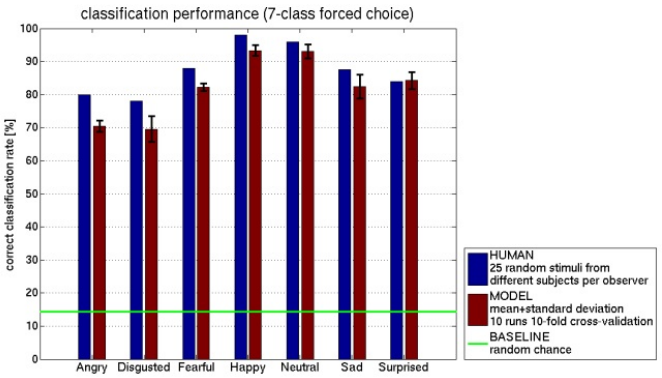


Fig. 12. The classification performance of the model in comparison to the human testers [**?**]

### *B. Real Time Multi-Person 2D Pose Estimation using Part Affinity Fields*

The approach of Cao et al. [**?**] inputs a given image into a two-branch convolutional neural network. First a feed forward network simultaneously predicts a set of confidence maps of the body part locations (see Figure 13 (b)) and a set of vector fields of part affinities, which encode the degree of association between the different parts (see Figure 13 (c)). After that the confidence maps and the affinity fields are parsed by greedy inference (see Figure 13 (d)) and they output the 2D key points for all the people in the image.

Like this we could also analyse the shoulder position like it was done in [**?**] and other movement based emotion detection. However it would not be possible to do angle-based evaluations like in McCol et al's approach [**?**].

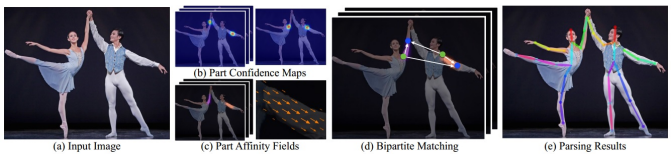


Fig. 13. OpenPose process visualized [**?**]

## V. Emotion detection through hand form

### *A. A system for person-independent hand posture recognition against complex backgrounds*

*1) Hand pose extraction:* Most approaches in this field, require the hand to be in front of a static background or the hand to be the only skin coloured item in the picture, in order for it to be recognized [**?**], [**?**], [**?**], [**?**]. The approach of Triesch and Malsburg [**?**], proposes a solution using elastic graph matching (EGM) with multiple feature types. EGM is "a neurally inspired object recognition architecture" [**?**]. In EGM the views of objects are visualized as a labelled 2-d graph. The nodes of the graph contain a local image description and the edges are labelled with a distance vector which represents the distance between the nodes.

*2) Interpretation:* Various graphs are created from training images and then images of new hands are matched to the most similar existing image. In Figure 14 we can see an example of a graph, which is matched to a new hand.

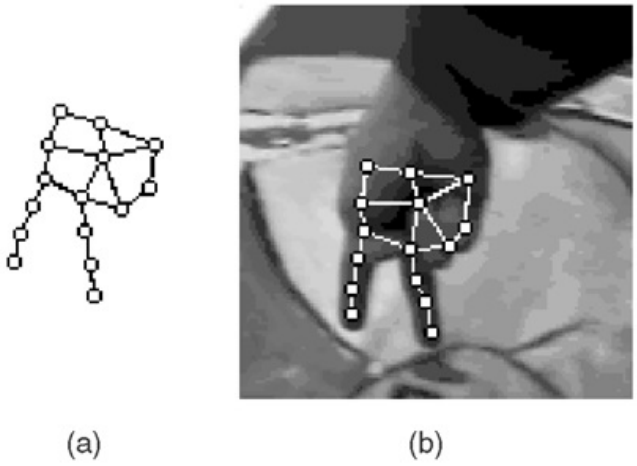


Fig. 14. Picture of an elastic graph of a hand, and a new hand that matches it [**?**]

The reported accuracy of this approach is 92.9% against simple backgrounds and 85.8% accuracy against complex backgrounds. With this approach we can easily recognize if a hand is for example clenched to a fist ( i.e. anger/aggression) and it is not limited, because it also works against complex backgrounds.

## VI. Emotion detection through human gait

### *A. Emotion Detection from Natural Walking*

*1) Gait information extraction:* Although this approach of Cui et al. [**?**] is not 100% fitting to this topic, it is interesting and worth to mention. To extract the data from the gait, they attached two smartphones (Samsung I9100G) to one wrist and one ankle separately [**?**]. The phone's accelerometer was used to gather data of the walk. After preprocessing it, various features are extracted using Principal Component Analysis (PCA).

To build the training set, the 59 participants were asked to walk naturally back and forth on an area of 6m x 0.8m for two minutes [**?**]. After that they reported their current emotional state from 1-10. This was done twice. On the first round, 1 was no anger and 10 was angry, whereas in the second round, 1 was not happy and 10 was very happy. After that the participants watched an emotional film clip for emotional priming[3] (for anger). After this the participants were asked to walk again for one minute, and report the anger score after the walk and recall the anger score right after the film. After at least three hours, the same experiment was repeated with a happy film.

[3]https://en.wikipedia.org/wiki/Priming_(psychology)

*2) Interpretation:* The classification approach was done in different ways. The most successful one was using SVM. It correctly classified the emotion of the walk with 90%.

This approach shows, that we can detect the emotional state of a person based on his/her gait, however different people walk very individually. This is why it is necessary to get a baseline walk from each person. Hence it is very hard/ impossible to instantly infer information from a person's walk, when seeing the person for the first time (more to person-dependent and person-independent gait in section VII).

### *B. Emotion recognition using Kinect motion capture data of human gaits*

*1) Gait information extraction:* The approach of Li et al. [**?**] uses a set up with two Kinect cameras that were placed oppositely at the two ends of a 6m x 1m footpath. The emotional states of the participants were recorded in the same way as in the experiment described under subsection VI-A. The data reported by the two Kinect cameras is processed independently. Kinect outputs a stick figure with the different body joints, (see Figure 15) of the recorded person, marked. The recorded data contains the 3-dimension position of the 14 joints. This results in a 42 dimension vector. Because each record consists of T frames, the data of one record is represented by a T * 42 matrix. Fourier transformation is used for further feature extraction.

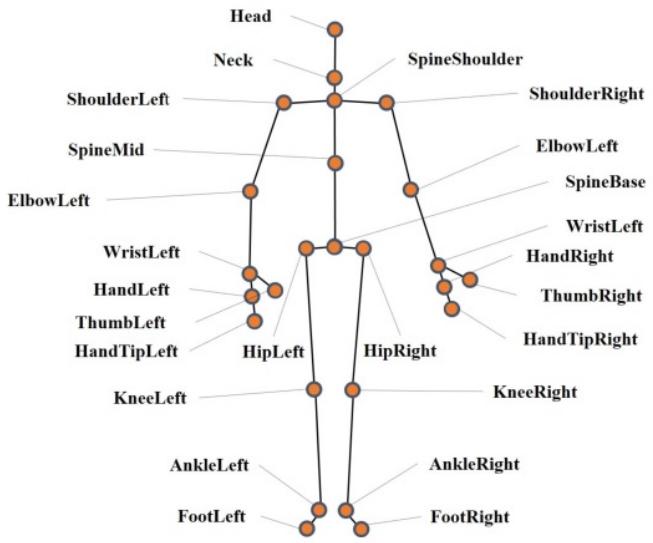


Fig. 15. Stick figure extracted by kinect [**?**]

*2) Interpretation:* The data is classified in various ways using NaiveBayes, RandomForests, LibSVM and SMO [**?**]. The most successful approach is using NaiveBayes. The reported accuracy is 80.5% for recognizing angry and neutral states with the Kinect1 camera, using NaiveBayes. 75% accuracy is achieved with the Kinect2 camera. For recognizing happy or neutral, the achieved accuracy is 79.6% with Kinect1 and 61.8% with Kinect2. However the distinction between angry and happy only succeeds with an accuracy of 52%.

This approach shows that it is also possible to recognize emotions from human gait, without using any sensors. However it makes a great difference whether the person is recorded from the front or the back.

## VII. Proposed model for emotion detection through human gait

In this section I review the possibilities for recognizing emotions through human gait in a real-time scenario. There are various problems and limitations which are discussed in subsection VII-A. Finally I propose a concept for a system based on the possibilities and limitations in subsection VII-B.

### *A. Possibilities and limitations*

The problem with many emotion recognition algorithms is, that they often need a neutral initialization pose [**?**], [**?**] because humans and especially human gaits are very individual. This makes a real-time system very hard, without having data of a specific person beforehand. However there are various gait databases available [**?**], [**?**], which could be used for training with machine learning algorithms. This still has to be treated with great care because 1. The databases only consist of acted emotions (currently no spontaneous gait database exists that is publicly acessible) and 2. As already mentioned, due to the high individuality of gaits [**?**], [**?**]. In [**?**], the inter-individual recognition of emotions through gait was only around chance level.

However this leads to the next problem, the expression of the affective state in gait is often also associated with the velocity [**?**], [**?**], [**?**] and other measures like pressure etc., in addition to the movement itself. This is part of the problem that gait data is often characterized by a high dimensionality, temporal dependency, high variability and nonlinearities [**?**], which makes a real-time classification difficult (performance wise). This is why it is important to preprocess the gathered data and limit the feature space and do a good feature extraction. PCA, KPCA, LDA and GDA are commonly applied for feature selection [**?**].

| Feature | NN | Naive Bayes | SVM |
|---|---|---|---|
| PCA-FT-PCA | 43 | 41 | 57 |
| Velocity, Cadence, Stride Length | 52 | 45 | 45 |
| Significant Subsection | **63** | 49 | 47 |
| Sig. Subsection + PCA (15PC) | 58 | 52 | **62** |
| Sig. Subsection + KPCA (15PC) | 36 | **60** | **60** |
| Sig. Subsection + LDA | **63** | 55 | **62** |
| All Joint Angles | 56 | 45 | 25 |
| All Joint Angles + PCA (30PC) | 50 | 50 | **69** |
| All Joint Angles + KPCA (23PC) | 28 | 58 | 25 |
| All Joint Angles + LDA | 52 | 53 | 53 |

Fig. 16. Average accuracy of inter-individual classification approaches with various pre-processing and classification methods [**?**]

| Feature | NN | Naive Bayes | SVM |
|---|---|---|---|
| PCA-FT-PCA | 70 | 70 | 78 |
| Velocity, Cadence, Stride Length | 85 | 83 | 76 |
| Significant Subsection | 87 | **93** | **89** |
| Sig. Subsection + PCA (15PC) | **91** | 85 | **89** |
| Sig. Subsection + KPCA (15PC) | 87 | 75 | 88 |
| Sig. Subsection + LDA | **93** | **93** | **93** |
| All Joint Angles | **91** | **93** | 79 |
| All Joint Angles + PCA (15PC) | **92** | **92** | **95** |
| All Joint Angles + KPCA (15PC) | 88 | 47 | 25 |
| All Joint Angles + LDA | 47 | 45 | 47 |

Fig. 17. Average accuracy of person-dependent classification approaches with various pre-processing and classification methods [**?**]

Figure 17 and Figure 16 highlight the importance of appropriate preprocessing, feature selection and picking a proper classifier as well as the importance of person-dependent (Figure 17) and inter-individual classification (Figure 16).

A positive aspect is that the emotion detection can be reduced to a classification problem (classify into one of the 6 emotion types [**?**] described in **??**. The most successful approaches for emotion detection in body movement in general (not gait specific) seem to be support vector machines [**?**], [**?**] and artificial neural networks [**?**], [**?**]. However there is not much work found for emotion detection through human gait using artificial neural networks [**?**]. This might be due to the fact, that there is not enough training data available.

### *B. Concept*

Based on those limitations the following approach could be useful:

- Create/ find a large enough gait database to train the model initially.
- Extract the pose using openPose [**?**], as it is working in real-time and without any special equipment. Any camera can do it.
- Do the feature extraction according to Figure 16. A good choice would be to use Sig. Subsection + LDA [**?**].
- Use a nearest node (NN) classifier [**?**].
- Do a continuous recording of the individual's gait, together with an evaluation about how they feel in this moment. This could be easily doable in a workplace environment for example by having a *mood* button right after entering the building. This step is important because if the system can compare person-dependent data it will greatly improve the accuracy.

A question which I can not answer is, if the classification can work in real-time, or if it takes too long. All the research that I found did not do it in real-time. Only the openPose pose extraction algorithm works in real time [**?**].

An artificial neural networks approach or a combination could also work, but the research is very thin on this topic [**?**], [**?**].

## VIII. Conclusion

One thing that we always have to keep in mind is that emotion recognition is not an easy task, even for the human himself [**?**], [**?**], [**?**]. The approaches in emotion detection through body language are dominated by a range of machine-learning algorithms. The research clearly shows that it is possible to detect human emotions from body language in a fairly accurate way.

There are many different approaches that tackle specific problems in this domain e.g. emotion detection through body position, body movement, gait, hand form etc. Most of them perform fairly well in their specific domain. However there are very few holistic approaches that do not require a specific set-up.

Sometimes various emotions are harder to distinguish than others. For example distinguishing anger and a neutral state is easier than distinguishing anger and happy [**?**] (see subsection VI-B) and some emotions share similar features like for example happiness and surprise [**?**]. Overall it seems harder to distinguish between negative emotions, by body positioning and body movement [**?**].

A lot of those approaches require initialisation poses [**?**], [**?**], [**?**], [**?**] which prevents the system from being usable instantly and in real-time. This is not surprising, because humans are distinct individuals and their neutral body poses/ gaits/ movements can be very different from each other.

The whole research is very machine-learning dominated. Approaches using artificial neural networks become increasingly more popular [**?**] as they might bring great advances in this field. Finally I came to the conclusion that it is not really important to have metrics on how to interpret the different emotions, but instead we can use human-created datasets to train the network and being able to learn how to recognize the emotion.

## References

[1] S. Singh, N. Sethi, and V. Sharma, "Significance of bodily movement for detection and analysis of emotions: A."