Detecting affect from non-stylised body motions

Daniel Bernhardt and Peter Robinson

Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge, CB3 0FD, UK

Abstract. In this paper we present a novel framework for analysing non-stylised motion in order to detect implicitly communicated affect. Our approach makes use of a segmentation technique which can divide complex motions into a set of automatically derived motion primitives. The parsed motion is then analysed in terms of dynamic features which are shown to encode affective information. In order to adapt our algorithm to personal movement idiosyncrasies we developed a new approach for deriving unbiased motion features. We have evaluated our approach using a comprehensive database of affectively performed motions. The results show that removing personal movement bias can have a significant benefit for automated affect recognition from body motion. The resulting recognition rate is similar to that of humans who took part in a comparable psychological experiment.

1 Introduction

The human body has evolved not only to perform sophisticated tasks, but also to communicate affect and inter-personal attitudes. One possible distinction can be made between affect communicated through non-stylised and stylised motions. In a stylised motion the entirety of the movement encodes a particular emotion. Stylised motions normally originate from laboratory settings, where subjects are asked to freely act an emotion without any constraints. They also arise from stylised dance. This paper, however, concerns itself with the more subtle aspects of *non-stylised* motions. We will examine how affect is communicated by the manner, in which every-day actions, such as knocking or walking, are performed.

After reviewing relevant related work in the next section, we discuss the segmentation of a complex motion into a sequence of meaningful motion primitives in Sect. 3. Sect. 4 then describes a set of dynamic features which captures the segments' affective information. In order to disambiguate affective cues from personal movement idiosyncrasies we make use of a statistical normalisation procedure. Sect. 5 introduces the experimental validation results which are based on an existing motion database of acted every-day motions. A concluding discussion is given in Sect. 6.

2 Background

What makes a walk happy? How does an angry knock differ from a sad one? Indeed, are the visual cues from the body alone sufficient to judge a person's affect? Early research by Ekman suggested that people make greater use of the face than the body for judgements of emotion in others [1]. More recent results from psychology suggest, however, that emotional body language does constitute a significant source of affective information. In an experimental study Bull established that body positions and movements are consistently displayed and recognised during phases of interest/boredom and agreement/disagreement [2]. In a more recent study Pollick et al. examined the accuracy with which human observers could distinguish basic emotions from point-light arm movements [3]. They found that despite the impoverished nature of the displays, the recognition rates were significantly above chance level.

It is surprising that despite the apparent interest of psychologists in natural, non-stylised motions, research in affective computing has for a long time focused on stylised body movements. Only very recently, Kapoor et al. investigated *natural* behaviour in a computer-based tutoring environment. They demonstrated a correlation between frustration and various non-verbal cues including body posture [4]. Earlier, Camurri et al. had developed a vision-based library to analyse expressive body gestures based on both shape and dynamic cues. In recently reported results, they show correlations between affect expressed through stylised dance and dynamic measures such as quantity of motion and body contraction [5]. In a different study Kapur et al. showed how very simple statistical measures of stylised motions' dynamics can be used to distinguish between four basic emotions [6]. Our work builds on these latter dynamic approaches, gaining its main affective information from quantities such as limb velocity and acceleration.

However, the nature of non-stylised movements means that we will need to look more deeply into the structure of the motions involved. We need to understand which elements are governed by affect and which are confounded by other factors such as idiosyncrasies in personal movement (movement bias), gender or activity. This is a problem which animators have faced for a long time in order to create compelling and realistic motion sequences. As John Lasseter put it [7]:

> One character would not do a particular action the same way in two different emotional states. [...] No two characters would do the same action in the same way.

There has been a large body of recent work on the subject, which normally aims to provide better means for animators to control the expressions of their characters. Many of the approaches work on the notion of affective transforms applied to an underlying basic or neutral animation. In many cases these transforms change the spatial amplitude and speed of motions [8–10]. An interesting concept, as suggested by Rose et al., is the notion of motions being composed of a verb (a basic motion concept, such as walking) and an adverb which modifies the basic motion in various ways (e.g. happily, sadly, uphill or downhill) [11].

The idea of viewing body language as analogous to natural language is not a new one. Ray Birdwhistell argued that complex motions can be broken down into an ordered system of isolable elements which he called kinemes [12]. The notion of a universal set of kinemes or motion primitives is a compelling one as it gives structure to the otherwise vast complexity of human motion — a goal which the Facial Action Coding System [13] has achieved so successfully for the face. It is this segmentation into motion primitives which will help us to discard the structural information of non-stylised motions, leaving the essentially dynamic cues which we will use to distinguish different affects.

3 Motion Analysis

For this work we used a motion-captured database recorded at the Psychology Department, University of Glasgow [14]. It gave us access to a collection of knocking, throwing, lifting and walking motions performed by 30 individuals (15 male and 15 female) in neutral, happy, angry and sad affective styles. Most of our quoted results are based on the approximately 1200 knocking motions from the database.

The skeletal structure of the recorded bodies is represented by 15 joints, positioned relative to a world frame. In order to obtain a rotation- and scale-invariant representation, we transform the joint positions into a body-local coordinate system and normalise them with respect to body size. Let f stand for the dimension of time, measured in frames. We denote the time-varying signal of normalised joint positions as the matrix $\boldsymbol{\Psi}$. We can also represent the motion in terms of the joint rotations over time, $\boldsymbol{\Theta}$. A particular body configuration at frame f can be represented as a row vector, denoted as $\boldsymbol{\psi}_f$ or $\boldsymbol{\theta}_f$. The gth positional or rotational degree of freedom at frame f is written as $\psi_{f,g}$ or $\theta_{f,g}$ respectively. We will also make use of the dot notation (e.g. $\dot{\theta}_{f,g}$) to denote derivatives with respect to time. Finally, projecting a body configuration vector onto a subspace is written as the linear operator $P(\cdot)$, e.g. $P_{\mathrm{rh}}(\boldsymbol{\psi}_f)$ denotes the position of the right hand relative to the body-centred coordinate system at time frame f.

3.1 Motion Segmentation

The goal of motion segmentation is to parse high-dimensional body movements into a sequence of more basic primitives. In general, this is a hard problem which is of interest to researchers from many different areas, including gesture recognition and robotics. Our approach is based on the work by Fod et al. [15]. It makes use of an objective function $E(f)$ which is a measure for the overall motion energy (activation) at time frame f. In many ways this concept of energy is analogous to that employed in the segmentation of speech into phonemes or words [16]. Let $\dot{\theta}_{f,g}$ denote the angular speed of the gth rotational degree of freedom at time frame f. Then we can define the body's motion energy as a weighted sum of the rotational limb speeds.

$$E(f) = \sum_{k=1}^{n} w_k \dot{\theta}_{f,k}^2 \tag{1}$$

In essence, E will be large for periods of energetic motion and will remain small during periods of low motion energy. Fig. 1 shows E for repeated knocking.

Fig. 2 illustrates how the observed energy peaks coincide with actions such as arm raises or individual forward and backward movements during the knock. Local minima in E can be observed whenever the trajectory of the right arm changes direction. We can use these insights to segment a complex motion as follows.

1. Compute E for the whole motion sequence.
2. Threshold the signal at a threshold t. Mark all frames f for which $E(f) > t$.
3. Find all connected regions of marked frames and regard them as individual motion segments.
4. Extend the segments to the preceding and succeeding local minima of E.

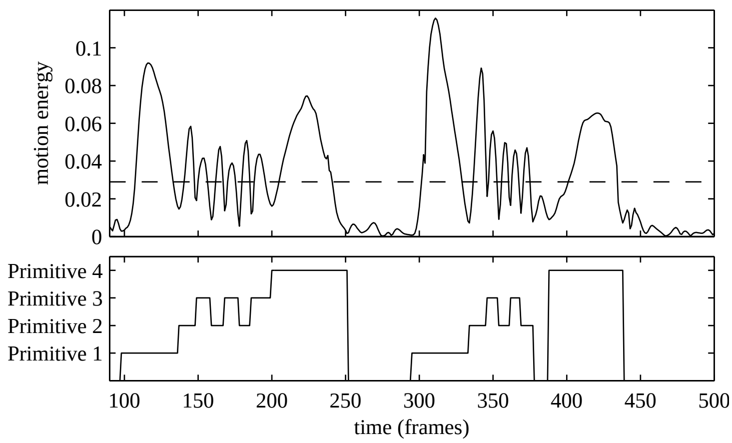

Fig. 1. Objective function $E(f)$ (top) with automatically calculated optimal segmentation threshold $t_{opt} = 0.029$ for part of a repeated knocking motion. The bottom shows the parse of the above motion sequence into four motion primitives and periods of no motion.

Obviously, our choice of t has a major impact on the nature of the segments. Fod et al. use empirically derived thresholds. If this method is to be used in a general framework, however, we need an automatic way of finding an optimal t. We propose the following solution. For every pair (E, t_n) we obtain a number of segments by thresholding E at t_n. Let $numseg_E(t_n) = s_n$ be the function which computes the number of segments s_n for any such pair. Fig. 3 shows $numseg$ for the motion in Fig. 1 and sampled at various thresholds. Our goal is to find a threshold which will exhibit all major motion segments (energy peaks) while filtering out small scale motions due to low-level signal noise. We note that noise is mainly registered during periods of low energy (e.g. between frames 250–300 and 450–500 in Fig. 1). Let t_0 be an empirical noise threshold. Then the optimal threshold t_{opt} is defined as the threshold which maximises the number of major motion segments.

$$t_{opt} = \underset{t}{\text{argmax}}\{numseg_E(t)\} \qquad \text{subject to } t_{opt} > t_0. \tag{2}$$

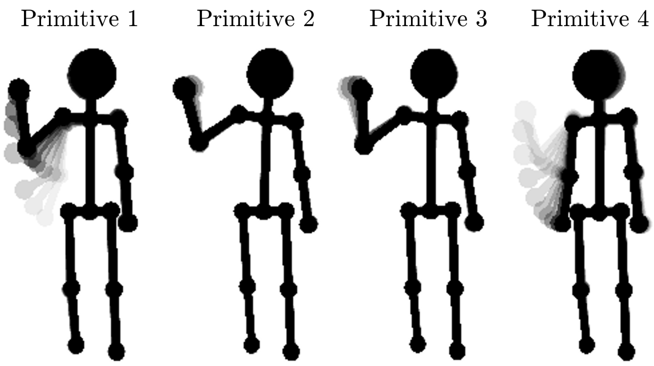

Fig. 2. Four phases of a knocking motion exhibiting distinct peaks of motion energy. Our algorithm detects each of the phases as a separate motion segment. Each segment is labelled with one of four automatically derived motion primitives. The primitives coincide with the semantically meaningful basic actions “Raise arm”, “Knock”, “Retract”, “Lower arm”.

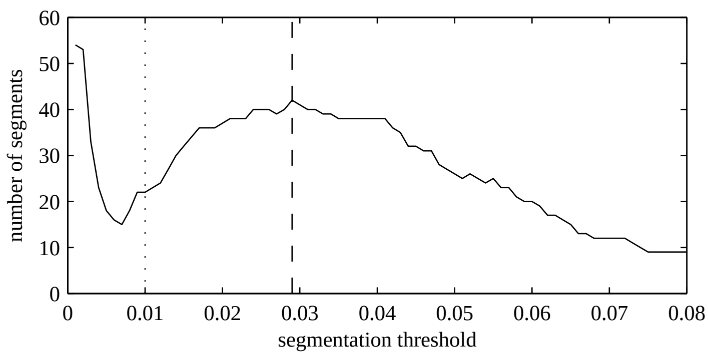

Fig. 3. $numseg_E(t)$ for a repeated knocking motion sampled for thresholds between 0.001 and 0.08. The diagram also shows t_{opt} (dashed) and t_0 (dotted).

3.2 Motion Primitives

Ideally, we would like to group the extracted segments into semantically meaningful clusters representing primitive motions. One approach to define such primitives would be to use a comprehensive list as devised by Bull or Birdwhistell to transcribe their psychological or anthropological observations [2, 12]. Due to their generality, however, these sets are large. Many of the listed primitives are irrelevant for any particular context. Indeed, context often governs the *affective and social meaning* of movements [12]. We therefore adopt a more context-dependent approach to the definition of motion primitives. It is based on the clustering of a set of example motions which are representative for a certain context. For our current scenario the context is very specific (knocking) and therefore the number of motion primitives is rather small. In more complex scenarios such as "everyday activities" or "interpersonal conversations" we would expect to require a larger set of primitives to represent all observed movements well.

Consider the database of affective knocking motions described above. After segmenting the movements, we need to find a representation for the segments which allows us to compare and cluster them. We therefore consider the *joint angles* of the motions and time-normalise them. This is done by resampling each segment at 25 equally spaced intervals. We also subtract the segments' means in order to capture the *relative* motion rather than the absolute body configurations. Next, we wish to group the segments into semantically distinct categories. We hypothesised that the knocking motions can be divided into four basic phases: lift arm, repeatedly knock and retract, lower arm. We therefore used a simple k-means clustering algorithm with $k = 4$. In a completely unsupervised scenario without any prior knowledge of the number of motion primitives, we would choose a clustering technique which automatically determines an optimal number of clusters such as hierarchical or Markov clustering. The following steps summarise our algorithm to compute a set of motion primitives from a set of example motions:

1. Segment the set of motions as described in Sect. 3.1.
2. Time-normalise all segments. Subtract sample means.
3. Cluster the normalised segments.
4. The clusters (or cluster centroids) represent the motion primitives.

The four derived motion primitives for the set of knocking motions are visualised in Fig. 4. The four degrees of freedom of the right arm are represented as four separate curves.

Having defined our primitives, we can now *parse* a new motion by following steps 1 and 2 as outlined above and replacing steps 3 and 4 by an assignment to the closest cluster centroid (most similar primitive). Fig. 1 illustrates how a repeated knocking motion (energy curve shown on top) has been parsed into a sequence of primitives (bottom). The motion is parsed in a semantically meaningful fashion. Fig. 2 shows that primitives 1 and 4 correspond to the larger scale motions of raising and lowering the right arm while primitives 2 and 3 capture the smaller scale knocking motions. We will now turn to the analysis of the dynamic and affective parameters of the segmented motions.

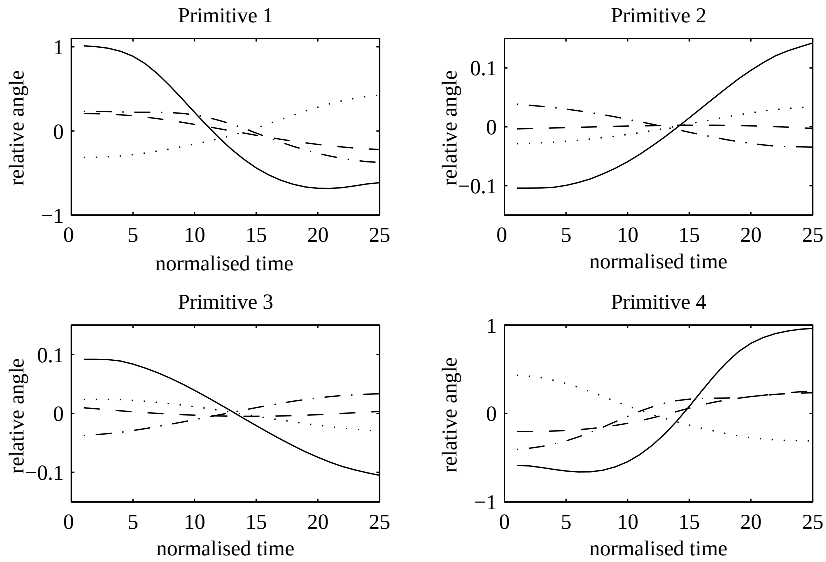

Fig. 4. Four motion primitives derived by k-means clustering. Dashed/dotted curves visualise rotations in the right shoulder, solid line visualises rotation in the right elbow. All other degrees of freedom are omitted as they exhibit no significant motion.

4 Affect Recognition

Angry movements in the analysed database tend to look energetic and forceful while sad knocks appear relatively slow and slack. Similar observations are true for the other classes of motions such as throwing and walking. This role of dynamic movement qualities such as velocity and acceleration in affect recognition has been stressed by several authors [3, 5, 6]. Never before, however, has the analysis of dynamics been attempted at the level of motion primitives. We propose this solution as a more flexible and well-founded alternative to the use of fixed or sliding windows as used in other recent works [5, 6].

We are using four statistical measures as features for affect recognition. They are computed over a whole motion segment such as an arm raise. For the analysed knocking motions only the right arm exhibits significant movement. Therefore all dynamic features are currently based on the right arm. Consider a motion segment $\boldsymbol{\Psi}$ of length n. Remembering that $P_{\mathrm{rh}}(\boldsymbol{\psi}_f)$ denotes the position of the right hand, we can define the features as follows.

1. Maximum distance of hand from body $\quad d_h = \max_{f=1}^{n} ||P_{\mathrm{rh}}(\boldsymbol{\psi}_f)||$
2. Average hand speed $\quad \bar{s}_h = \frac{1}{n}\sum_{f=1}^{n} ||P_{\mathrm{rh}}(\dot{\boldsymbol{\psi}}_f)||$
3. Average hand acceleration $\quad \bar{a}_h = \frac{1}{n}\sum_{f=1}^{n} ||P_{\mathrm{rh}}(\ddot{\boldsymbol{\psi}}_f)||$
4. Average hand jerk $\quad \bar{j}_h = \frac{1}{n}\sum_{f=1}^{n} ||P_{\mathrm{rh}}(\dddot{\boldsymbol{\psi}}_f)||$

We can also compute analogous features d_e, $\bar{s}_e$, $\bar{a}_e$, $\bar{j}_e$ based on the elbow motion. For any person p and motion segment m this gives us the feature vector $\boldsymbol{\phi_{p,m}} = (d_h, \bar{s}_h, \bar{a}_h, \bar{j}_h, d_e, \bar{s}_e, \bar{a}_e, \bar{j}_e)$.

4.1 The problem of individual movement bias

According to Lasseter's view quoted in Sect. 2, the two major factors affecting observable motion qualities are emotion and individual style. In Sect. 5 we shall show that our data does indeed reveal some global correlation between the above features $\boldsymbol{\phi_{p,m}}$ and the different emotion classes. Fig. 5(a), however, shows that the between-class variability of the two very different emotion classes sad and angry is smaller than we would hope. The hand speed distribution for sad knocks (black) overlaps heavily with that of angry knocks (white). In order to be separable through a pattern recognition approach, the two distributions should show a large between-class variability while exhibiting a small within-class variability. This exemplifies the problem of individual movement bias. Different people tend to display the same emotion in very different ways, thus impeding classification.

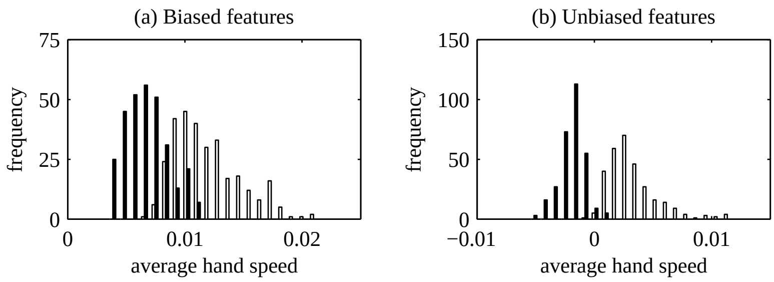

Fig. 5. Biased and unbiased feature distributions for sad knocks (black) and angry knocks (white).

Our approach to this problem is a normalisation procedure based on the following intuition. It seems a reasonable assumption that a person's motion idiosyncrasies influence his or her movements in a consistent fashion — after all we expect them to be governed by gender, physical build and other constant factors. Even dynamic factors such as mood might be changing slowly enough to be assumed temporarily constant. We therefore propose to model individual motion bias as an additive constant signature $\bar{\boldsymbol{\phi}}_{\boldsymbol{p}}$ which influences the motion features introduced above. We obtain an estimate of the unbiased motion features $\hat{\boldsymbol{\phi}}_{\boldsymbol{p,m}}$ by subtracting the personal bias.

$$\hat{\boldsymbol{\phi}}_{\boldsymbol{p,m}} = \boldsymbol{\phi_{p,m}} - \bar{\boldsymbol{\phi}}_{\boldsymbol{p}} \qquad \text{for any motion } m \tag{3}$$

An important problem is how to estimate $\bar{\boldsymbol{\phi}}_{\boldsymbol{p}}$. If we do not "know" a person, i.e. have no history of his or her movements, we may need to take an a priori guess, maybe conditioned on gender or other cues. However, if we have a history, we can compute $\bar{\boldsymbol{\phi}}_{\boldsymbol{p}}$ from all the observed motions. In our case, we take an average over all the knocking motions in the database in order to learn about a person's motion bias. Note that this operation does not tell us anything about affect-specific factors as all motions are treated equally and different affects are represented at equal frequencies in the database. Fig. 5(b) illustrates how this normalisation improves the between-class variability for the two shown classes considerably. Sect. 5 gives a more rigorous account of the improvements achieved when taking movement bias into consideration.

4.2 Machine Learning

We can use the biased or unbiased motion features to train a classifier which distinguishes the four emotions neutral, happy, angry and sad. We decided to use support vector machines (SVMs) with a polynomial kernel as they tend to exhibit good generalisation performance. The suitability of SVMs for this domain was demonstrated by Kapur et al. [6]. In order to solve the general problem of recognising the affect of a motion sequence, we train a family of binary SVMs $M^z_{x,y}$. The classifier $M^z_{x,y}$ aims to find the maximum margin between affect classes x and y for motion primitives of type z. Once these binary classifiers have been trained, we can classify a new motion as follows:

1. Segment motion into a list of primitives as described in Sect. 3.
2. Let the first segment in the list be of primitive type z. Apply all pairwise SVMs $M^z_{x,y}$. Classify the segment according to a majority vote.
3. Remove segment from the list and repeat from step 2 until list is empty.
4. Classify whole motion by majority vote of individual segment classifications.

5 Experimental Results

With the conducted experiments we aimed to answer three questions:

1. What recognition rate can be achieved with our approach?
2. How does movement bias (see Sect. 4.1) affect the recognition performance?
3. How do our results compare to related results found in the literature?

We used the knocking motions from our database to run Leave-One-Subject-Out cross-validation (LOSO-CV) tests. Overall, we used approximately 1200 motion samples with an equal proportion for each of the considered emotions neutral, happy, angry and sad. For each iteration of the cross-validation the system was therefore trained on around 1160 samples and validated on 40 samples. In different tests we found that the system does considerably better if we add some of the remaining 40 samples to the training set or perform a subject-independent cross-validation. In contrast to those tests, the figures we quote here are representative for the generalisation performance of our system for an unknown person.

The confusion matrices for LOSO-CV using biased and unbiased features are shown in Table 1. Note that angry and sad knocks are classified more reliably than neutral and happy ones. The most significant factor which negatively affects recognition rates (sensitivity) is the confusion between neutral and happy knocks. In answer to question 2 above, we find that using unbiased features improves the overall recognition rate considerably from 50% to 81%. Our informal observations from Sect. 4.1 have hence been confirmed.

Table 1. Confusion matrices for LOSO-CV using biased features (left) and unbiased features (right). The classification procedure distinguished between four emotions: neutral (neu), happy (hap), angry (ang) and sad. All average and affect-specific sensitivities are above chance level (0.25).

Truth	classified as			
	neu	hap	ang	sad
neu	**0.379**	0.229	0.127	0.265
hap	0.281	**0.411**	0.179	0.129
ang	0.176	0.203	**0.591**	0.030
sad	0.214	0.139	0.023	**0.624**
	average sensitivity: **0.501**			

classified as			
neu	hap	ang	sad
0.742	0.199	0.007	0.052
0.278	**0.653**	0.056	0.013
0.013	0.063	**0.924**	0.000
0.066	0.010	0.000	**0.924**
average sensitivity: **0.811**			

We can obtain a measure for the more objective recognition efficiency η if we normalise the achieved sensitivity by the sensitivity expected by chance (sometimes referred to as generality [17]).

$$\eta = \frac{\text{achieved sensitivity}}{\text{sensitivity expected by chance}} \tag{4}$$

In our case we would expect a classifier which assigns one of the four affect classes at random to achieve a sensitivity of 25%. Therefore the efficiencies of our classifiers for biased and unbiased features are $\eta_b = 2.0$ and $\eta_{ub} = 3.24$ respectively. We can use these measures to compare our results to those of related experiments in the next section.

6 Discussion and Future Work

For our discussion we consider the results of two other related experiments. We were using part of a database which was created by Pollick et al. for psychological work. In one particular study they examined how accurately *human subjects* could classify affect from knocking motions displayed as point-light or full video stimuli [3]. The only major difference from our experimental setup was their forced choice between five rather than our four emotional states (afraid being the additional class). They report that humans achieved a recognition rate of 59% for point-light and 71% for full video stimuli. These figures illustrate that even humans are far from perfect at classifying affect from non-stylised body motions. We can calculate the efficiency $\hat{\eta}$ achieved by humans as defined in Eq. 4. For point-light and video displays humans exhibit efficiencies of $\hat{\eta}_{pl} = 2.95$ and $\hat{\eta}_v = 3.55$ respectively.

One of the major contributions of our work derives from the fact that classifying affect from non-stylised motions is harder than from stylised ones. This is demonstrated by the experiments performed by Kapur et al. [6]. They recorded stylised emotions and compared the accuracy of various machine learning techniques as well as human performance. For the task of distinguishing four basic emotions from point-light displays, humans achieved a recognition rate of 93% ($\eta = 3.72$). This is considerably higher than human performance reported by Pollick et al. for non-stylised movements ($\hat{\eta}_{pl} = 2.95$). For SVMs the recognition rate was lower at 83.6% ($\eta = 3.34$). These results are summarised in Table 2.

We have shown that using unbiased dynamic features based on motion primitives boosts the recognition rate considerably. Our computational approach exhibits a better efficiency than humans for classifying affect in non-stylised movements from point-light displays. The performance of our approach is also comparable to that of Kapur et al. This is significant since their stylised motion data contained solely affective information. For our non-stylised motions, on the other hand, only certain subtle aspects communicate affect while most of the motion signal is governed by the independent semantic meaning of the motion.

Table 2. Comparison of results from our and related experiments.

experiment	Kapur et al. [6]		Pollick et al. [3]		Our results (Sect. 5)	
motions	**stylised**		**non-stylised**		**non-stylised**	
classifier	human	SVM	human		SVM	
features	biased		pt.-light	video	biased	unbiased
# emotions	4	4	5	5	4	4
sensitivity	93%	84%	59%	71%	50%	81%
efficiency	**3.72**	**3.34**	**2.95**	**3.55**	**2.0**	**3.24**

We are currently working on overcoming various limitations of our approach. In the version presented here, the algorithm only considers the right arm for extracting affect-related dynamic features. Incorporating features from other body parts will help us to analyse motions such as walking, which are not primarily based on arm movements. Furthermore, the torso and head can be expected to hold valuable cues even for heavily arm-based actions [9]. One challenge which needs to be addressed in this direction is the generalisation of motion segments to multiple body parts. One could either compute segments for each body part individually or attempt to capture more subtle relationships between limbs by defining segments in terms of the simultaneous motion of multiple body parts.

Another limitation which might have to be addressed in the future is the assumption that motion primitives are clearly separable by local motion energy minima. For very smooth and cyclic motions, for example, it might be possible to exhibit the basic cycle periods of a motion as segments. Furthermore, it might not be sufficient to employ a single segmentation threshold t_{opt} if we analyse a series of movements which are very different in terms of exhibited motion energy. This problem might be solved by recomputing t_{opt} either regularly or, alternatively, whenever we detect a significant change in motion energy.

Finally, we are investigating better ways to estimate the personal movement signature $\bar{\boldsymbol{\phi}}_{\boldsymbol{p}}$. Ultimately, we would like to be able to predict $\bar{\boldsymbol{\phi}}_{\boldsymbol{p}}$ from as few and unconstrained example motions as possible.

References

1. Ekman, P.: Differential communication of affect by head and body cues. Journal of Personality and Social Psychology (1964) 726–735
2. Bull, P.E.: Posture and Gesture. Volume 16. Pergamon Press (1987)
3. Pollick, F.E., Paterson, H.M., Bruderlin, A., Sanford, A.J.: Perceiving affect from arm movement. Cognition **82** (2001) 51–61
4. Kapoor, A., Burleson, W., Picard, R.W.: Automatic prediction of frustration. International Journal of Human-Computer Studies (2007)
5. Camurri, A., Lagerlöf, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. International Journal of Human-Computer Studies **59**(1-2) (2003) 213–225
6. Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., Driessen, P.F.: Gesture-based affective computing on motion capture data. In: Affective Computing and Intelligent Interaction. (2005) 1–7
7. Lasseter, J.: Principles of traditional animation applied to 3D computer animation. In: SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques. (1987) 35–44
8. Amaya, K., Bruderlin, A., Calvert, T.: Emotion from motion. In: Graphics Interface '96, Canadian Human-Computer Communications Society (1996) 222–229
9. Zhao, L., Badler, N.I., Costa, M.: Interpreting movement manner. In: CA '00: Proceedings of the Computer Animation. (2000) 98–103
10. Polichroniadis, T.P.: High Level Control of Virtual Actors. PhD thesis, Computer Laboratory, University of Cambridge (2001)
11. Rose, C., Cohen, M.F., Bodenheimer, B.: Verbs and adverbs: Multidimensional motion interpolation. In: IEEE Computer Graphics and Applications. Volume 18. (1998) 32–40
12. Birdwhistell, R.L.: Kinesics and Context: Essays on Body Motion Communication. University of Pennsylvania Press (1970)
13. Ekman, P., Friesen, W.V.: Facial action coding system. Consulting Psychologists Press (1978)
14. Ma, Y., Paterson, H.M., Pollick, F.E.: A motion capture library for the study of identity, gender, and emotion perception from biological motion. Behavior Research Methods **38** (2006) 134–141
15. Fod, A., Matarić, M.J., Jenkins, O.C.: Automated derivation of primitives for movement classification. Autononmous Robots **12**(1) (2002) 39–54
16. Wang, D., Lu, L., Zhang, H.J.: Speech segmentation without speech recognition. In: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing. (2003) 468–471
17. Huijsmans, N., Sebe, N.: Extended performance graphs for cluster retrieval. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition. Volume 1. (2001) 26–32