EXPRESSION OF EMOTIONS THROUGH BODY MOTION
A Novel Interface For Human-Robot Interaction

Nelson Gonçalves and João Sequeira
Institute of Systems and Robotics, Insituto Superior Técnico, Lisbon, Portugal

Keywords: Human-robot interaction, Emotions.

Abstract: An approach is presented for the expression of basic emotions through only the agent body pose and velocity. The approach is applied in human-robot interaction scenarios, where both humans and robots communicate only through their relative position and velocities. As a result, an interface for human-robot interaction is obtained, which does not require the use of haptic devices or explicit communication with humans, verbal for instance. The small set of emotions that can be conveyed enable humans and robots to anticipate the intentions of the opponent and adapt their behavior accordingly. The aproach is implemented using a webcam, simple vision processing algorithms and Hidden Markov models. The results of preliminary experiments are presented.

1 INTRODUCTION

The problem considered in this paper is the recognition of emotions in human-robot interaction (HRI) scenarios without explicit communication or the usage of haptic devices.

Such HRI problems can occur in many common applications. An example is that of a mobile robot advertising and selling products in a supermarket. Although it can move directly towards approach potential clients, this behavior may be considered too intrusive and unpleasant. Therefore, the robot must first estimate the interest of the clients without using haptic or voice interfaces. If a reasonable interest is perceived, the robot should then approach the clients.

Another application of interest is active surveillance, where mobile robots must intercept and identify intruders. These are not expected to cooperate and can even sabotage the robots. The intentions of the intruders must then be estimated at a safe distance and without explicit communication. In this application, the mobile robots can move aggressively, directly towards the intruders at high velocity. The purpose is to intimidate them and also to block potential exit pathways. In these applications, the use of traditional interface devices, such as voice or touch, is not efficient. The main reason is that in these applications, humans and robots keep some distance between them during most of the time. Another reason is that explicit communication, verbally for instance, may not be possible due to ambient background noise.

The proposed approach is to express and perceive emotions through the body pose and velocity. A friendly emotion can be expressed through a smooth path, executed at a low velocity. The antagonistic emotion of anger, may be expressed through sharp, discontinuous paths performed at a high velocity. An advantage of the proposed approach is an increase of the available bandwidth for human-robot communication, since the body motion is another possible communication channel. Another advantage is that agents can perceive the intentions of opponents at some distance and adapt their behaviors accordingly. This is relevant to robots in adversarial environments.

The remainder of this paper is as follows. A review of the literature is presented in Section 2. In Section 3 the nature of emotions and their forms of expression are discussed. A classifier for the recognition of emotions is presented in Section 4, which is evaluated in a set of preliminary experiments described in Section 5. Finally, in Section 6 the approach is discussed and future work is presented.

2 RELATED WORK

In the HRI problem considered, humans are not expected to explicitly communicate with robots or to use haptic interfaces. This is an uncommon scenario

Gonçalves N. and Sequeira J. (2009).
EXPRESSION OF EMOTIONS THROUGH BODY MOTION - A Novel Interface For Human-Robot Interaction.
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics - Robotics and Automation*, pages 465-470
DOI: 10.5220/0002220404650470

in HRI applications, (Fong et al., 2003; Goodrich and Schultz, 2007), where typical interfaces make use of voice, touch and human facial expressions. Nevertheless, the information conveyed through the body pose and velocity was considered in (Breazeal, 2003) and applied in practice in (Finke et al., 2005).

An early study on the expression of emotions in both humans and animals was conducted by Darwin in (Darwin, 1872). It is reported that the human body motion and stances, when expressing an emotion, are similar to when acting in accordance. For example, the body stances when expressing anger are almost identical to those when preparing for an actual attack. Also, it is well known that the state of mind has a strong influence on the motion of a person, (Nakamura et al., 2007). This is often exploited in computer animation to increase the realism of human characters, (Becheiraz and Thalmann, 1996; Neff and Fiume, 2006).

In neuro-psychological studies of human emotion, facial expressions typically receive much more attention than other forms of expression, (de Gelder et al., 2004). But in (Atkinson et al., 2004), it was found that emotions could be recognized from static and dynamic body stances. This was case also when human motion was represented using only a cloud of points. Finally, in (den Stock et al., 2008) the body motion was also found to bias the recognition of bimodal emotions from sound and vision cues.

3 EXPRESSION OF EMOTIONS

The nature of emotions is, to the best of the authors knowledge, an unsolved problem. Therefore, in this section an attempt is made to understand the nature emotions and how they can be perceived and expressed.

In the pioneer work by Darwin, (Darwin, 1872), and William James, (James, 1884), is argued that at least some emotions are a form of instinctive reaction to stimuli received from the environment. The reason is the similarity in some expressions among humans from very different cultures. Also, it is not plausible that a conscientious process it at the origin of emotions in animals. Nevertheless, since these initial contributions many other definitions of emotion have been unsuccessfully proposed, (Scherer, 2005).

Although the question of what is an emotion is yet unanswered, it is more relevant for the HRI problem to answer questions related to the causes of emotions. The answer to these questions is stated in terms of the causation categories by Aristotle, (Russell, 2004). If these are known, then suitable models can be build and used for perceiving and expressing emotions. The first question to be posed is: "why do humans express emotions ?". A possible answer is given in terms of the final causation category:

Assumption 1 (Manifestation of Emotions). *The final cause of an emotion is the change in the agent state, perceptible to external observers.*

The final causation category is identified with the concepts of purpose and ultimate goals. Thus, in this paper it is assumed that the purpose of an emotion is to be announced to others, through a change in the agent state. It is clear that other answers are possible if other causes are identified. The answer could be given in terms of specific hormones or physiological mechanisms, such as in (Scherer, 2005). These answers belong, respectively, to the material and efficient causation categories.

The final causation category is used because an important design guideline is obtained. That the emotions an agent can express do not form part of the state. In order to understand this argument, consider the case where emotions form part of the agent state. Then the sequence of emotions being expressed is uniquely determined by the state past history and dynamics. As a result, the agent state changes could be known in advance and there would be no need to express them. Therefore, in this paper emotions are considered not part of the agent state, but instead part of the agent actions. The difference between emotions and other actions is that the former cannot be applied to the environment. As the result of this discussion, a definition for emotions is obtained.

Definition 1 (Agent Emotion). *An emotion is an action executed by the agent on his state, producing an externally perceptible state change.*

This definition is useful for the design of HRI interfaces. For an application example, consider the case where facial expressions are used to express emotions. Let the state of the agent, human or robot, be the configuration of the mouth and the eyebrows. An emotion is then the act of displaying a particular configuration of the mouth and eyebrows. Similarly, emotions can be perceived by identifying the respective state configurations.

The previous definition does not provide clues on how the state is altered through an emotion. Thus, the next question is: "why is an emotion expressed through some state changes and not others ?". The answer is given using the efficient causation category, which is related to the concepts of method or function. A possible answer is then that the forms of emotion expression in humans are function of evolutionary pressures. An immediate consequence is that un-

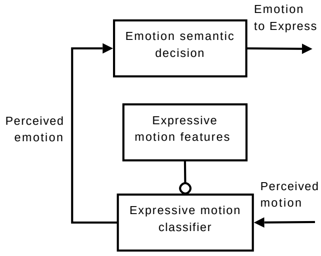

Figure 1: Architecture for emotion perception and decision.

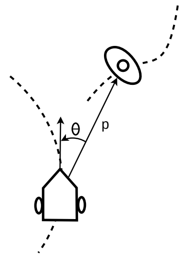

Figure 2: Typical HRI situation.

der different environments, different forms of expression would emerge. Another consequence is that a learning algorithm could be employed to determine the human forms of expression. But in general, it would require that humans and robots interact during an excessively long period of time. Therefore, it is more practical to mimic, where possible, the forms of expression in humans and domestic animals.

4 PERCEPTION OF EMOTIONS

In the remainder of the paper and when clear from context, humans and robots are both referred to as agents. It is assumed that agents move in a 2D plane. Furthermore, robots are assumed not to possess any anthropomorphic features.

The proposed architecture for the perception of emotions is presented in Figure 1. The motion of humans is perceived and classified using features of interest defined *a priori* by the system designer. Thus the recognition problem can be formulated without knowledge on the semantics of emotions, since from Definition 1, only state changes must be perceived. After the type of emotion is perceived, the robot must decide which emotion to display. In this step is required knowledge of the context and also the meaning of emotions. The solution is presented in Sub-Section 4.2, where the notion of empathy is used.

4.1 Expressive Motion Classifier

The design of the classifier of emotions from the human body motion is formulated as time series classification problem. The human body is approximated by the geometric center and the features of interest are the human pose and velocity relative to the robot. This choice is based on the expression of emotions by human actors described in (Atkinson et al., 2004) and the social distances presented in (Becheiraz and Thalmann, 1996; Pacchierotti et al., 2006). An accurate estimation of the features values is not required. The reason is that the basic emotions, such as fear and disgust, are fundamental to survival and are expressed with clear state changes.

A typical situation for HRI through motion is depicted in Figure 2. The mobile robot is represented by the polygonal shape and both agents are moving at different linear velocities. The robot is able to measure the relative pose of the human, at a constant rate Δ^{-1}. Since the robot is also moving, it will perceive an apparent motion of the human. This effect must be corrected to prevent erroneous classifications.

Consider a static frame $\{w_k\}$, which is coincident with the robot body frame at time t_k. Let $p(t_k)$ be the position of the human measured by the robot and assume that the human is static. Then, for a small enough interval Δ, at t_{k+1} the measured value should be

$$\hat{p}(t_{k+1}) = R(\omega(t_k)\Delta)(p(t_k) + v_r(t_k)\Delta) \qquad (1)$$

where $v_r(.)$ is the robot linear velocity, $\omega(.)$ the angular velocity about the robot frame origin and $R(.)$ is the rotation matrix from frame $\{w_k\}$ to the robot frame at time t_{k+1}. Let $p(t_{k+1})$ be the actual measured value by the robot at time t_{k+1}. If the human is not stationary, then the predicted and measured values are not equal and their difference is due to the human velocity

$$v_h(t_k) = (\hat{p}(t_{k+1}) - p(t_{k+1}))\Delta^{-1} \qquad (2)$$

The vector of observed motion features is then

$$f_k = (\|p(t_k)\|, \theta(t_k), \|v_h(t_{k-1})\|) \qquad (3)$$

where $\theta(t_k) = atan(p(t_k))$. The first two features model static properties of the expression of emotions, which are linked to focus of the agent on the observer. The relative velocity of the human is related to the intensity of the emotion.

The block diagram of the emotion classifier is presented in Figure 3. The input is an array of feature vectors, $m_{[k,k+n]} = (f_k, f_{k+1}, \ldots, f_{k+n})$. Through vector quantization, each feature vector f_k is replaced by a symbol s_k. Then the probability of each Hidden Markov Model e_i generating the sequence of symbols, $s_{[k,k+n]}$, is computed with the forward algorithm. The

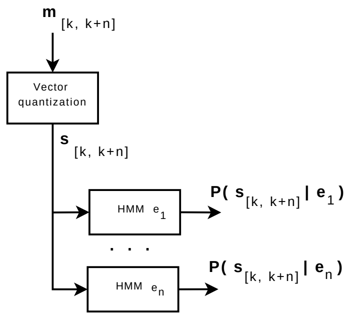

Figure 3: Emotion classifier from motion.

output of the classifier is a vector with the normalized value of these probabilities, $P(s_{[k,k+n]}|e_i)$.

A similar approach was used in (Takeda et al., 2007) with good results in estimating the next dance step for a robotic dance partner.

4.2 Emotion Semantic Decision

The expression of emotions in humans is closely linked to similar instinctive behaviors, (Darwin, 1872; James, 1884; Scherer, 2005). Therefore, it is reasonable to use a reactive approach to determine the emotion to express. The proposed solution is to make use of the concept of empathy. The emotion expressed by the robot, e^*, is the one assigned the highest probability by the classifier

$$e^* : \max_i \{P(s_{[k,k+n]}|e_i)\} \tag{4}$$

This is a straightforward solution and does not require knowledge of the emotion semantics or context. A similar method is used in (Takeda et al., 2007), where the dance step is selected based on the ratio between the two highest probabilities. If it is above some threshold the step associated to the highest probability is executed, otherwise the robot wheels are stopped. This method is not suitable for expressing emotions through motion because stopping can be perceived as an emotion, fear for instance. It is also not robust to classification errors and does not enable robots to take the initiative. The latter is an important aspect in general HRI problems. Since most humans are not familiar with robots, they may not expect an autonomous behaviors from these machines. The original solution can be improved by minimizing a decision cost

$$e*(\gamma) : \min_j \left\{ \sum_i c_{ij}(\gamma) P(s_{[k,k+n]}|e_i) P(e_i;\gamma) \right\} \tag{5}$$

where $c_{ij}(.)$ is the cost of expressing emotion e_i instead of emotion e_j and $P(e_i;\gamma)$ is the *a priori* probability of observing emotion e_i. The discrete parameter γ is used to define the context of the mobile robot application. For instance, in surveillance applications it is reasonable to expect humans to behave aggressively. The probability of observing anger is then greater than that of happiness and the robot should also prefer also to express hostility over happiness.

5 EXPERIMENTAL RESULTS

A set of experiments where conducted to evaluate the emotion classifier, with results presented in this section. The interface was implemented in C++ language using a standard, of-the-shelve webcam mounted on top of a Pioneer P3-AT robot. The purpose is to evaluate the classifier from the robot world perspective view. The experiments where performed with a human wearing a bright, green colored vest to facilitate the detection of color blobs. The webcam was calibrated to measure the distance and angle of the human under the assumption that the height of the hip is constant. The blob detection is affected by high frequency noise because the vest surface is wrinkled and is not perceived with an uniform color. A median filter was applied the values of the blob centroid, to remove some of the noise. The maximum sampling rate was approximately 6 samples per second, much slower than the rate in (Takeda et al., 2007) for instance.

The parameters for the vector quantization procedure where determined by hand, based on the social distances discussed in (Becheiraz and Thalmann, 1996; Pacchierotti et al., 2006). The values of the relative position and angle are quantized in $\{1.0, 2.0, 3.5, 4.5, 5.5\}[m]$ and $\{-40^o, 0^o, 40^o\}$, respectively. The norm of the relative velocity is quantized in $\{0.5, 1.5, 2.0, 3.0, 4.5\}[m/s]$, where $0.5\,m/s$ roughly corresponds to the human being stopped. With respect to the use of clustering algorithms, this approach does not require a large amount of data to be properly trained. Also, the quantization values determined by the algorithms may not reflect the social distances used by humans.

The emotions considered in the experiments where: (i) anger, (ii) fear, (iii) friendliness and (iv) apathy, which can be understood as the agent not expressing any emotion. Their expression was exemplified by a human in front of the robot. The human ran towards the robot to express anger, while friendliness was expressed with a normal pace. In order to express fear, the human moved toward the robot but at halfway stopped and moved away. The expression of apathy was exemplified with the human moving parallel to the webcam image plane or away from

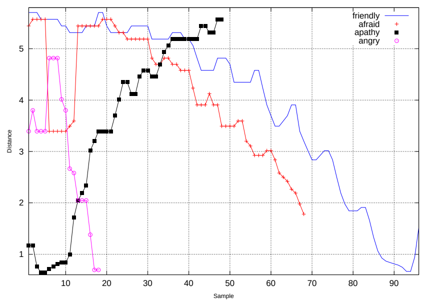

Figure 4: Example of feature $\|p(t_k)\|$ for each emotion.

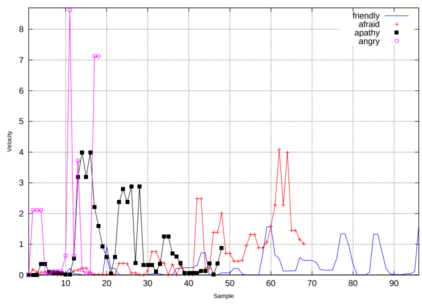

Figure 5: Example of feature $\|v_h(t_{k-1})\|$ for each emotion.

the robot. A total of twenty videos where recorded, with five examples for each emotion. The features $\|p(t_k)\|$, $\|v_h(t_{k-1})\|$ and $\theta(t_k)$ are plotted in Figures 4 to 6, taken from an example of each emotion. In these figures is visible that the distance and angle features produced distinctive sequences for each of the emotions emotion. For example, friendliness and fear produce similar distance sequences but clearly distinct sequences of angles. The estimation of the values for the velocity feature is sensitive to the noise and detection failures of the vision system. Another source of error is the height of the hip which as small variations during the motion. Although some patterns are visible in Figure 5 for each emotion, the sequences of values are very irregular and with also abnormally high values.

The HMM of each emotion was trained using the quantized sequence of features from all the videos. Each HMM is composed with five states and a left-right transition structure. After the training phase, the emotion classifier was evaluated using all of the emotion examples. The prior probabilities of each emotion are equal, $P(e_i;\gamma) = 0.25$, and the elements of the decision cost matrix are all unitary. In Table 1 are summarized the classification for each set of videos. The numbers between parenthesis in the first column represent the total number of sequences, $s_{[k,k+n]}$ per set of videos. The numbers in the other columns represent the number of times the corresponding emotion

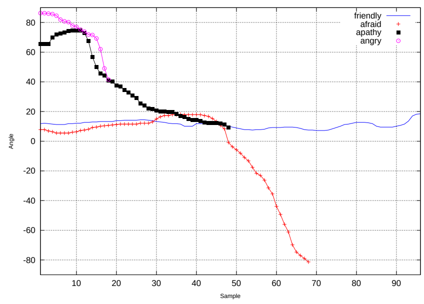

Figure 6: Example of feature $\theta(t_k)$ for each emotion.

was perceived. From Table 1, the emotions of friendliness and fear had the highest number of correct classifications but not anger and apathy. The main reason is that these emotions have sequences with much smaller dimensions, see Figure 4 for instance. Nevertheless, the data in Table 1, is useful to determine the values for c_{ij} and $P(e_i;\gamma)$, which reduce the classification errors. These can also, to some degree, be handled by the selection and expression of emotions in the robot. For instance, the first action of the robot can be to stop and observe the human in order to reduce classification errors. This is a common trait in the expression of fear in humans, for example.

6 CONCLUSIONS

An HRI interface for the expression of emotions through body motion was presented. The approach was implemented in practice with a standard vision system. Despite the simplicity of the implementation, acceptable results where obtained. In addition, the bottlenecks of the system performance where identified. Thus, given more efficient feature estimation methods, it is reasonable to assert the feasibility of the proposed HRI interface. Since human body motion is emotionally charged, (Atkinson et al., 2004), no prior training in robotics or specially designed hardware is required in this HRI interface. Thus, it can be used in HRI applications where humans are un-skilled in mobile robotics.

The interface is valuable also to other mobile robot applications. As argued by António Damásio, in (Damasio, 2006) and elsewhere, emotions are fundamental to successful decision making in humans. Thus the ability to express them without the need for additional hardware is by itself a feature of interest. Since any movement of the agent can be perceived as an emotion, knowledge of the application context is required for disambiguation purposes. The discrete parameter γ was introduced to account for the appli-

Table 1: Emotion classifier results with $c_{ij} = 1.0$ and $P(e_i;\gamma) = 0.25$.

Video Set \ Emotion	*Friendly*	*Afraid*	*Apathy*	*Anger*
Friendly (23)	10	4	5	4
Afraid (31)	9	11	8	3
Apathy (11)	3	3	1	4
Anger (13)	3	3	7	1

cation context.

The vision system low frame rate and the detection failures had a negative impact on the system performance. Therefore, future work is aimed at increasing the frame rate and the robustness of the human detection. For instance, through better the use of hardware and a detection algorithms, such as a face detector. Also, the approach must be evaluated using groups of humans with different backgrounds in mobile robotics.

ACKNOWLEDGEMENTS

This work was supported by European Project FP6-2005-IST-6-045062-URUS, and Fundação para a Ciência e a Tecnologia (ISR/IST pluriannual funding) through the POSConhecimento Program that includes FEDER funds. Nelson Gonçalves is working under grant SFRH/BD/23804/2005, from Fundação para a Ciência e a Tecnologia.

REFERENCES

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33:717–746.

Becheiraz, P. and Thalmann, D. (1996). A model of nonverbal communication and interpersonal relationship between virtual actors. In *Computer Animation '96. Proceedings*, pages 58–67.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1-2):119–155.

Damasio, A. (2006). *Descartes' Error*. VINTAGE RAND.

Darwin, C. (1872). *Expression of the Emotions in Man and Animals, The*. Oxford University Press Inc, 3 sub edition.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., and Hadjikhani, N. (2004). Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proc Natl Acad Sci U S A*, 101(47):16701–16706.

den Stock, J. V., Grezes, J., and de Gelder, B. d. (2008). Human and animal sounds influence recognition of body language. *Brain Research*, 1242:185–190.

Finke, M., Koay, K. L., Dautenhahn, K., Nehaniv, C. L., Walters, M. L., and Saunders, J. (2005). Hey, i'm over here - how can a robot attract people's attention? In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 7–12.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4):143–166.

Goodrich, M. A. and Schultz, A. C. (2007). Human-robot interaction: a survey. *Found. Trends Hum.-Comput. Interact.*, 1(3):203–275.

James, W. (1884). What is an emotion ? *Mind*, 9:188–205.

Nakamura, T., Kiyono, K., Yoshiuchi, K., Nakahara, R., Struzik, Z. R., and Yamamoto, Y. (2007). Universal scaling law in human behavioral organization. *Physical Review Letters*, 99(13):138103.

Neff, M. and Fiume, E. (2006). Methods for exploring expressive stance. *Graphical Models*, 68(2):133–157.

Pacchierotti, E., Christensen, H. I., and Jensfelt, P. (2006). Evaluation of passing distance for social robots. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 315–320.

Russell, B. (2004). *History of Western Philosophy (Routledge Classics)*. Routledge, 2 edition.

Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729.

Takeda, T., Hirata, Y., and Kosuge, K. (2007). Dance step estimation method based on hmm for dance partner robot. *Industrial Electronics, IEEE Transactions on*, 54(2):699–706.