Robot emotions generated and modulated by visual features of the environment

by Aaron S.W. Wong, Steven Nicklin, Kenny Hong, Stephan K. Chalup & Peter Walla

This is an author-prepared version of the article, reprinted from Proceedings of 2013 IEEE Symposium on Computational Intelligence for Creativity and Affective Computing (CICAC), p. 9-16.

http://dx.doi.org/10.1109/CICAC.2013.6595215

Robot Emotions Generated and Modulated by Visual Features of the Environment

Aaron S.W. Wong, Steven Nicklin,
Kenny Hong & Stephan K. Chalup
Newcastle Robotics Laboratory
School of Electrical Engineering and Computer Science
The University of Newcastle, Australia
Email: {aaron.wong, steven.nicklin, kenny.hong}@uon.edu.au
stephan.chalup@newcastle.edu.au

Peter Walla
School of Psychology
Centre for Translational Neuroscience
and Mental Health Research
The University of Newcastle, Australia
Email: peter.walla@newcastle.edu.au

***Abstract*—Emotions are generated and modulated by many factors in the ever-changing surrounding environment. A new and challenging task is to emulate emotional responses on a robot that are caused by visual stimuli, such that the robot's responses mirror that of the human user. This paper presents the initial stage of an affective system that has been trained on-line using reinforcement learning to generate and modulate emotions. The inputs of the system comprise a subset of emotionally relevant visual features extracted from the environment: colours, fractal dimension, and facial pareidolia. These inputs are mapped onto an output that expresses the associated emotion in terms of language. Pilot experiments demonstrate how a humanoid robot tries to learn through interaction with a human companion to express emotions associated with different environmental scenes in a (near) human-like manner.**

I. Introduction

Companion robots are popular in the world of science fiction; however, in the real world no robot has been sufficiently programmed to act as a useful companion robot. There are still many issues to be resolved before this can be achieved, including the basic skills of walking, localisation and object recognition, and more difficult functions such as language, memory and human-robot interaction.

Another issue to overcome if companion robots are to be successfully integrated into society is known as the uncanny valley hypothesis [1], which describes the discomfort felt by a person in the presence of an entity that seems almost but not completely human. The perception of aesthetics and associated expression of empathy plays an important role in natural communication and will have to become part of human-robot interaction on a sufficiently sophisticated level in order to ease any negative effects predicted by the uncanny valley hypothesis.

The manner in which a robot interacts with humans can persuade a person to like or dislike the robot. A robot that is demanding and abrupt may cause the user to dislike being commanded by an inanimate entity. However, an overly kind and pleasant manner may lead the user to not consider information delivered by the robot seriously. From these examples, it is important to understand that the circumstances surrounding the situation influence how the user will react.

Human emotions are shaped by cultural, genetic, environmental and internal factors [2]. Taking a broad approach to the context of the situation with respect to human-robot interactions, the surrounding environment has an effect on the way we humans emotionally interact with each other. This effect causes subtle differences in emotions, which amount to important features for a robot to exploit, especially when positive interactions with people are desired. The robot must first understand the surrounding environment in order to exploit these environmental features.

Extracting environmental features for use in human-robot interaction is a relatively new and challenging task. Research preceding the work presented in this paper consists mainly of studies focused on the direct extraction of emotion from the user. Most papers focus on the extraction of facial features as emotions for the communication of affective processing [3], while others concentrate on the extraction of sound or tone of voice as emotions to infer the affect of the user [4]. Some perform multi-modal extraction, where they examine a combination of these classes of features [5, 6]. Only recently, a clear and narrow definition of emotion and affective processing has been proposed, which facilitates emotion research in general [2] and also helps to develop meaningful affective communication via emotions between humans and robots. Not surprisingly, there is minimal research on the use of environmental features for affective communication in human-robot interactions.

The impact of environmental sensory stimuli, such as sight, sound, smell [2, 7], and touch, are known to have an effect on the emotions and behaviour of people [8]. In order for a robot to relate to humans, it must first be able to visualise and feel through the same modes of perception. Devising techniques suitable for processing visual input is an important aspect, as are those for obtaining features heard in environmental sound. Once these features have been obtained, it is important to organise them in a rational format, such that information of importance can be extracted and used appropriately in the correct context.

Having the ability to sense the environment can assist robots to become friendlier in the eyes of a user, as they can better relate to how humans feel, based on the surrounding

environment. Once the environmental features have been extracted, we use them to implement and train an affective system for a humanoid robot. In this paper we present the commencement of an experimental system that aims towards the development of a complete affective system generating emotions derived from features extracted from the surrounding environment.

The paper is organised as follows: In the next section an emotion model is introduced and some environmental features that trigger affective processing and generate human emotions are examined, followed by an outline of the related feature extraction algorithms. The proposed affective system is then described, with regard to the stages of input, training, and language module in the output. Pilot experiments using the system are then presented, followed by a description of the results obtained from each experiment. Finally, conclusions and plans for future work are discussed.

II. A Definition of Emotion and Features that Modulate Human Emotions

Every stimulus from the environment engages both cognitive and affective processing. Affective processing codes for pleasantness on a graded scale from unpleasant to pleasant. All externally triggered affective processing is also combined with affective information reflecting internal processes and together they can but not necessarily do generate emotions, which are the bodily consequences [2]. Exactly how and what kind of emotions are generated depends on factors that are genetic, cultural, and environmental in origin. One stimulus can generate different emotions in different people. Thus, personality shapes emotions. Visual features such as those listed in this paper readily stimulate users' affective processing and the generation of emotions [9]. Computer vision techniques allow for the extraction of low-level features from images, such as colour, texture, shape and spatial location of image elements. However, it is difficult to extract high level features automatically, such as the names of objects, scenes, behaviours and affective aspects (emotions) [9].

The following subsections will examine three visual features that are used in this paper: fractal dimension, colour, and the impact of facial expression patterns. However, it is important to note that 1.) more features exist, some of which are shown in other literature [9, 10], and 2.) personality, character and memory also have influence on emotions.

A. Fractal Dimension of Edge Patterns

Fractals are "rough or fragmented geometric shapes that can be subdivided into parts, each of which is (at least approximately) a reduced-size copy of the whole" [11]. Fractal dimension can be defined as a measure of complexity, of how the detail in the pattern changes with the scale. This value is similar to the generally understood terms, 'two dimensions' or 'three dimensions' (2D or 3D). However, the fractal dimension is not limited to these integer dimensions, and may exist in the spaces between them.

One way to think of the fractal dimension is to consider a shape such as a coastline on a map. This may have a fractal dimension greater than one, but less than two, suggesting that the shape is more complex than a simple line, but too sparse to be considered two-dimensional. This occurs by the shape of the coastline exhibiting some two-dimensional space-filling characteristics, while being created from a one-dimensional line [12].

Recently, applications of the fractal dimension have appeared in art to evaluate beauty. The fractal dimension value accounts for more than just the visual complexity, allowing it to better account for the variance in judgements of perceived beauty than that of visual complexity alone [13, 14].

Through the use of psychological experiments, the fractal dimensions (FDs) of images are related to the images' affective properties, and a higher FD makes people feel more messy [15]. The FD of a monotonous image is lowest, at around 1.37, while the FD of a messy image will be higher, for example around 1.91, and the FD of harmonious images will be medial at around 1.68.

In the field of the built environment and nature, experiments have been conducted [16] to ascertain the correlation between participants' interests, and scenes of various fractal dimension. The results obtained from experiments involving 220 participants [16, 17], indicate that there are three categories with respect to aesthetic preference for fractal dimension: 1.1-1.2 low preference (less), 1.3-1.5 high preference, and 1.6-1.9 low preference (too complex and challenging to comprehend). Although these values change from person to person, creative people tend to prefer greater complexity in scenes consisting of higher fractal dimensions [15].

B. Dominant Colours of the Environment

In the 1960s, Johannes Itten [18] summarised the theory of the effects of colour on emotions. In his theories, he postulated that "colour effects are in the eye of the beholder, and that the deepest truest secrets of colour effect are invisible to the eye, and are held by the heart alone". This implies that the emotional responses that result from visual perception of colour eludes the set rules or logic that could be placed in any conceptional formulation between individuals, making this problem a challenging task for any machine.

Itten stated that there are seven features of colour that have an effect on the way we feel: the contrast between hues, the contrast between light and dark, cold and warm, complementary colours, simultaneous colours, saturation of colours and extension of colours. All these features can be obtained by comparing different values of the hue, saturation and intensity (HSI) model of colour.

Recent psychological studies have been conducted that employ a number of Itten's features to determine a relationship between the colour and emotion space. In Mao et al. [19], histograms of hues were obtained from colours of an image and were categorised by the perception of monotonous, harmonious or 'mussy' (confused). In Wang et al. [20], an extended study was conducted to examine other emotions with regard to their valence and arousal. This work led to the generation of seven rules that relate to valence, and 17 rules that relate to arousal. These rules were then used to

predict the feelings of a person based on the features obtained from the HSI model.

The field of emotion semantic image retrieval [21] examines the converse of our problem. It is important to note that this field also studies the features of an image in relation to the emotion space. It can also be seen from [3, 20, 21] that colours and their related sub-categories are frequently used as features that affect emotions.

C. Facial Pareidolia Effects

The proposal by Chalup et al. [22, 23] suggests that pareidolia of abstract faces and facial expressions that appear in house designs can produce an emotional response from an observer. Facial pareidolia is the ability to "see" faces in random or vague stimuli. The use of faces and facial expressions as a stimulant for an emotional response is supported by a number of studies. These studies have identified regions of the brain that are dedicated to processing faces (the fusiform gyrus) [24] and processing of emotional expressions (the amygdala) [25].

These findings imply that we are 'switched on' to looking for faces whether we like it or not. Most importantly, they can be perceived beyond our awareness [26]. For a robot to behave like a human, it should be able to manifest this phenomenon and use this information as an added dimension to derive its emotional perception of its surrounding environment. Since a neural basis exists for processing faces and facial expressions, we hypothesise that the emotional response produced from abstract face-like patterns could generate emotive cues, similar to those generated by fractal dimensions and colour.

III. Implementation on a Humanoid Robot

The robotic platform used for this study is a humanoid robot, the Dynamic Anthropomorphic Robot with Intelligence - Open Platform (DARwIn-OP) [27] by ROBOTIS as seen in Figure 1. The robot stands 45 cm tall, weighs approximately 5 kg and is equipped with 20 degrees of freedom. The robot uses an Intel atom processor to enable autonomous operation using the NUPlatform software architecture [28]. The on-board processor performs all object recognition and cognitive functions required to evaluate the environment. The robot has a forward facing high definition camera, a set of stereo microphones in the head, and speakers in its chest so that it is able to communicate and interact with people. The camera produces images at 30Hz with a resolution of 640 by 480 pixels, in the image format of YUV422.

A. Colour

The hue, saturation, and intensity (HSI) colour space was used in previous papers [19, 20]. HSI is a more intuitive model when compared with the convoluted chromatic values of the YUV422 colour space. Additionally, the HSI colour space corresponds to Ittens' theory of effects of colour on emotion. For these reasons, HSI was the colour space of choice in our emotion system.

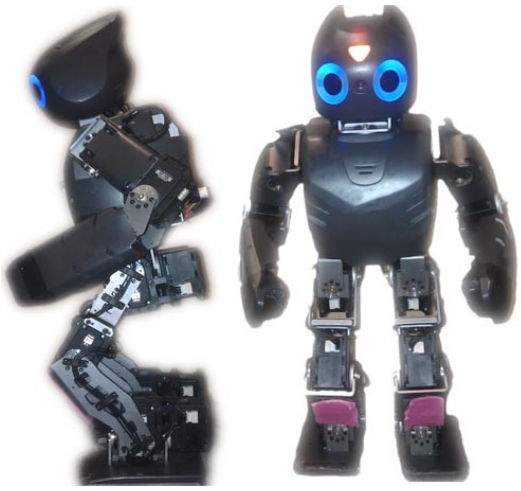

Fig. 1. The DARwIn-OP robot used in experiments. The robot's speaker was used to communicate the robot's feelings to the user.

Since the robot collects images in the original format of YUV422, a transformation was required to convert from the original colour space to HSI colour space [29–32].

Once the input colours have been converted to HSI, they can be used as a feature to better distinguish associated emotions, similarly to that used in [3, 20, 21]. As there are 640 by 480 pixels to process, only a random selection of 10,000 sample pixels are transformed for each image frame processed. The average of these transformed HSI samples is then found for each individual component: hue, saturation and intensity. These raw components are then used as part of the environmental emotion feature vector for further processing.

B. Box Counting Method for Fractal Dimension

In order to approximate the fractal dimension of a particular image, a number of box grid sizes are generated. These grids consist of squares with sides of length δ. The grids are used to divide the image into boxes. Each of these boxes is then evaluated to determine if they are filled or unfilled. A filled box contains some part of the image, while an unfilled box does not. The number of filled boxes for this particular box size is determined $N_\delta(F)$. Key steps in this process:

1) Grid set generation: Initial sizing is determined by the size of the original image. The size is chosen so that a minimum of 3 boxes can fit along the largest dimension of the image, while a minimum of 2 boxes fit along the shorter dimension.

From the largest grid size the remaining sizes in the set are generated iteratively by dividing the current grid size by a constant divisor. In our case, we used $\sqrt{2}$. Once the box size reaches the limit of detail in the image i.e. reduces to less than the size of a single pixel, the set is complete.

2) Result calculation: The box-counting fractal dimension is found by approximating the limit in Equation 1 [33–36].

To approximate this value the log of the number of filled boxes $N_\delta(F)$ is plotted against the negative log of the box sizes δ as shown in Equation 2. Linear regression is

performed using a least squares method to determine the linear fit to the data points. The box counting dimension estimate is then taken from the slope of this fitted line.

$$\dim_B F = \lim_{\delta \to \infty} \frac{\log N_\delta(F)}{-\log \delta} \tag{1}$$

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} -\log(\delta) \\ \log(N_\delta(F)) \end{bmatrix} \tag{2}$$

C. Machine Pareidolia System

To replicate the pareidolia effect of faces and produce an emotional response from facial expressions, several ν and SVDD one-class face detectors from a study of pareidolia of faces and expressions [37] were used. In addition, we have chosen several facial expression analysers from the same study, which uses pairwise adaptive C and ν-SVM classification [38]. In both scenarios [37, 38], they have evaluated their models with a training dataset consisting of human faces and facial expressions where a number of image preprocessing techniques and a number of resolutions were assessed.

To begin finding faces for our robot, a detection window of $n \times n$ pixels is scanned across an image at multiple scales and locations. The size of the detection window corresponds to the optimal resolution derived from cross validation. We gather a number of image scales by resizing the original image until it is no smaller than the size of the detection window. The resize down scale factor we use is 0.8 – this effectively performs a top down search where we start with the whole image and then examine it closer upon further iterations. At each scale we move our detection window of $n \times n$ pixels from top left to bottom right. Rather than examining every pixel location we define a step size to be roughly 10% of the width or height of the current image scale, whichever is smaller. For each face detected we classify its expression into one of seven universal facial expressions of emotion – happy, sad, surprised, fearful, angry, disgusted and contemptuous [39, 40]. The overall emotional response can then be defined by the expression that dominates among the faces detected.

IV. Machine Learning for a Robot with Environmentally Affected Emotions

The task is finding a mapping from feature space, the extracted visual features that affect human emotions, to the emotion space of the user(s). To form this mapping, we employ an artificial neural net that is trained using reinforcement learning in interaction with the user(s).

A. Reinforcement Learning

Reinforcement Learning [41], coupled with an artificial feed forward neural network, was used to learn on-line a mapping of observed visual features or current state s, to the output emotion a that reflects that of the current user.

In this study, the artificial neural network consists only of one layer of weights, i.e. with no hidden layer of units. Inputs to this network are the environmental features, which consist of colours, fractal dimension, and facial pareidolia, that are extracted and encapsulated in a vector, the current state s. The number of the input neurons is equivalent to the dimension of the input vector. The output of the neural network consists of eight neurons, where each neuron represents an emotion class. The weights of the neurons were adjusted using reinforcement learning.

The reinforcement learning strategy employed for this task was on-policy temporal difference (TD) learning [42], more commonly known as the TD(λ) algorithm with eligibility traces $e_t(s)$. In the following experiments, the trace decay parameter λ was 0, and the discount rate γ was set to 0.9.

The robot learns based on user interaction by reward input $r \in [+1, -1]$, where the weights of the neural network are updated using a function that consists of the current state s, observed reward r, next state s', and the step size α. $V(s)$ is the value obtained from the neural network using s. This function is updated as follows,

$$V(s) \leftarrow V(s) + \alpha e_t(s)\left[r + \gamma V(s') - V(s)\right] \tag{3}$$

For $\lambda = 0$, the eligibility trace $e_t(s)$ is simple:

$$e_t(s) = \begin{cases} \gamma\lambda e_{t-1}(s) + 1 = 1 & \text{if } s_t = s_{t-1} \\ \gamma\lambda e_{t-1}(s) = 0 & \text{otherwise} \end{cases} \tag{4}$$

The learning process described by equations 3 and 4 adjusts the weights based upon the on-line reward input received from the user.

B. Language Output and Spoken Interface

The robot communicates to the user through spoken words. Once an emotion vector has been obtained using the on-policy affective system through visual features of the environment, the words are then read out aloud to the user as part of a sentence. These spoken words consist of a combination of the following emotional words:

SAD	ANGRY
SURPRISED	FEAR
DISGUSTED	CONTEMPT
HAPPY	NEUTRAL

The words selected are dependent on the values of the output emotion vector, containing the likelihood of each of the eight possible emotions as predicted by the system. By having eight outputs, this allows the robot to expand on the interaction and words spoken to the user. An example of the spoken output is as follows.

I feel very ..... and a little .....

The user can then choose to respond to the spoken words signifying the particular emotions. If these emotional words do not represent the current feeling that the user is experiencing, he or she can select "no". Otherwise the user can select "yes", if he or she agrees with the robot's emotion. These responses correlate to the reward of the system and will trigger an update of $V(s)$. These responses are currently programmed as button inputs at the back of the robot but could as well be communicated to the robot via language.

V. Results

A. Fractal Dimension

Under laboratory conditions with consistent lighting and static scenes, parameters of the fractal dimension algorithm were adjusted to examine the type of input the robot could use for learning emotion. Here we were searching for parameters in the fractal dimension algorithm that had high variance to different scenes. In this situation we compared grey-scale images and Canny edge images. A selection of these images can be seen in Figure 2.

It was found that within the set scene, using a grey-scale image the robot obtained on average a fractal dimension of 1.70 with ± 0.10 variance. While using an edge image, the robot obtained on average 1.32 with ± 0.15 variance. It can also be seen that grey-scale images were affected by noise, resulting in a higher fractal dimension. Even though the results show similar variance, it is worthwhile to note that the laboratory is a relatively calm environment, and the average fractal dimension should correspond to a lower number [16, 17]. In stating this, the edge image was used for external experiments outside the laboratory.

Fractal dimensions were processed in different environments using the robot. The fractal dimension of a scene can vary with the motion state i.e. walking or stationary. As a robot manoeuvres through an environment, its motion of walking and panning to view the scene can impair and modify the result of fractal dimension output. This reduces the number of edges, due to the blurring of the image leaving only the dominant edges. This is similar to increasing the threshold on the edge detection algorithm, or blurring the image in its pre-processing stages before calculating the fractal dimension. The results leave an impression of the dominant edges such as the skyline [43].

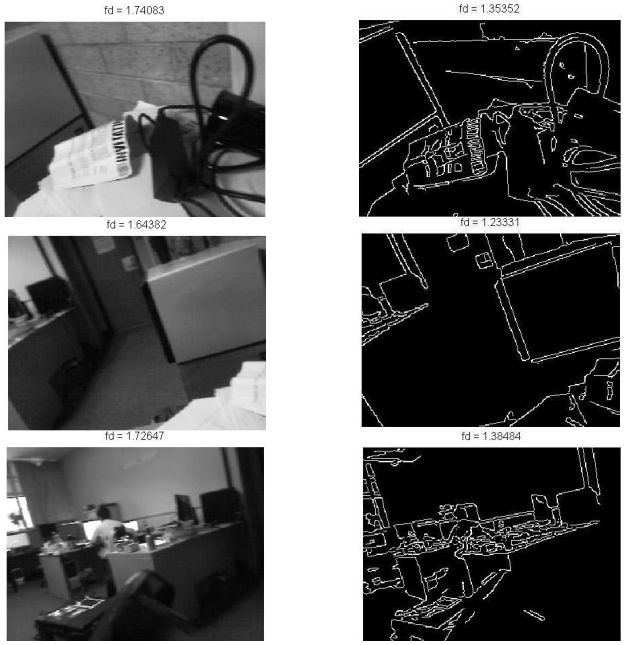

Fig. 2. Grey-scale Versus Canny Edge: Grey-scale images were affected by noise, resulting in a higher fractal dimension number. Even though the results show similar variance, it is worthwhile to note that the laboratory environment is a relatively calm environment, and the average fractal dimension should correspond to a lower number [16, 17].

B. Machine Pareidolia

To determine the appropriate face detectors and facial expressions from [37], we had several options in the pre-processing stage that required consideration. Included image pre-processing techniques considered were grey-scale, histogram equalised grey-scale; and their respective Sobel and Canny edges. Our initial chosen face detector was trained on images of size 30 by 30 pixel resolution. However, when tested on the robot, false positives were produced. Our reasoning was that this outcome was due to inconsistencies in the robot's camera settings and environmental lighting conditions when compared to that of the training data.

However, we have gathered the set of what we considered as true and false positive samples from our initial 30 by 30 pixel resolution grey-scale face detector and ran them against all possible one-class face detectors (i.e. the number of image pre-processing techniques and the number of resolutions evaluated in [37]) to isolate the face detector which best fit our purpose. Effectively, our face detector for the robot was determined using a two stage process. The first stage was a chosen 30 by 30 pixel resolution grey-scale face detector to obtain our set of true and false positives and use these samples to determine a favourable face detector. To further compensate for what we considered as false positives, we made use of the Canny edge operator where we defined a threshold of a minimum number of edges to become a positive face. If this threshold was reached the image was considered as a possible true positive.

While testing on the robot, it was noted that the grey-scale performed significantly faster, when compared to the Sobel classifier. Additionally, the speed of obtaining facial pareidolia results was hampered by the resolution of each possible face searched, and the number of faces it was required to search through. To increase the speed of processing the image resolution was reduced to 320 by 240 pixels, with the scaling set to a factor of 0.8, and pareidolia scan step size was set at 10% of the image size.

The current facial pareidolia sub-component consists of input regions of interest of 30 by 30 pixels. The format of these inputs is grey-scale $\in [0, 255]$, without histogram equalisation. The face detector module consists of a singular value decomposition (SVDD) SVM, which employed a radial basis function (rbf) kernel, whose ν and γ parameters were set to 0.05 and $1.22e^{-4}$ respectively.

In Figures 3 and 4 shows faces that are extracted from the surrounding environment as seen by the robot; the angry face is a vague face, while the detected surprised face is one that has been created from shadows. In Figure 4, for visual comparison, images of average faces used for the testing phase of the machine pareidolia system are shown against the detected faces. In our set-up, the machine pareidolia system had a strong bias to angry faces (approximately 50%

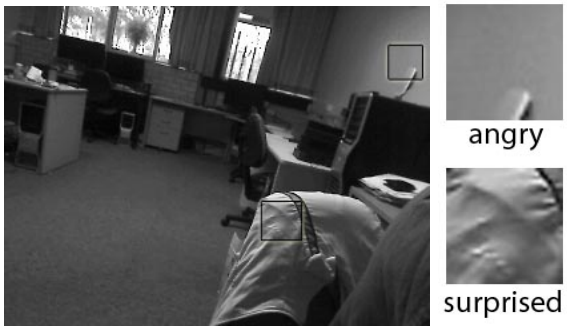

Fig. 3. Facial Pareidolia: Left image shows the image collected from the system, boxes show regions of interest detected as faces. Top Right, enlarged detected angry face. Bottom Right, an enlarged surprised face is detected.

Fig. 4. Visual Comparison of Faces: The left box contains an example of the average surprise face used in the testing data [37] and an extracted image output from the facial pareidolia sub-component of the affective system. Similarly, in the right box, an example of the happy expression.

of faces processed on the robot), as they consist of straight contours located near the mouth area, which correlates to straight shading in the lower region of interest.

C. Learning for an affective system using visual features of the environment

The robot was first tested under laboratory conditions. For every reinforcement learning problem, it is essential that parameters are tuned such that they match the problem. In this case, for an on-policy affective system trained on a real robot using visual features of the environment, we required a system with a fast convergence rate. In several pilot tests we determined a suitable step-size parameter α in Equation 3 in order to obtain acceptable results in a short time.

During experimentation under laboratory conditions, it was found that large values of α never recovered into a positive reward. This became unstable and eventually led to outputs of random emotions being expressed by the robot. Additionally, α values that were too small correlated to a slow rate of learning, and would eventually converge to the desired emotion after an extended period of time. This scenario was undesirable, since the user would feel distracted and disheartened with the robot. This would cause a comparatively greater effect on the emotional state of the user, as the user would be frustrated with the robot, loosing concentration on the effect of the surrounding environment.

The results, as seen in Figure 5, show an increase in the maximum reward parameter over a period of 100 - 200 iterations (approximately two seconds between each iteration). The curve shows the characteristics of a converging learning algorithm, where the graph plateaus into a region of stability. Here, we define stability to be the state where no further training was required (no positive or negative reward)

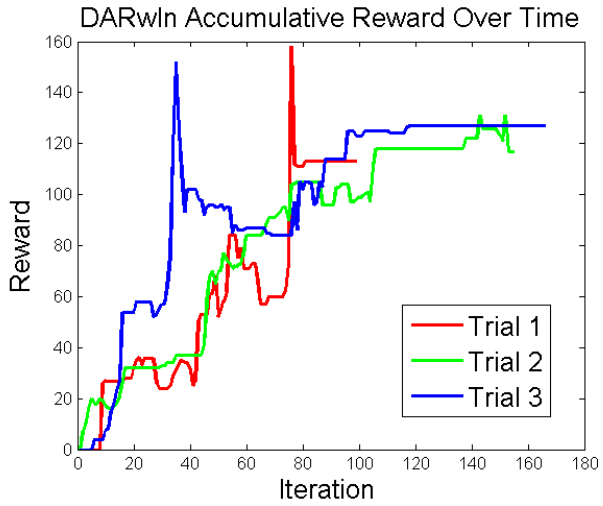

Fig. 5. Training the affective system in a laboratory environment. System training was terminated when the rate of learning had stagnated or a time limit of 5 minutes was reached.

and simultaneously have the ability to replicate the user's affective status.

The spikes visible on the plot represent instances of random value generation, in an attempt to escape local maxima. On completion of training, the robot had successfully mimicked the emotion of the user for the different scenes provided within the laboratory, as seen from the convergence of reward in Figure 5.

As the affective system on the robot was trained, the weights from the input features (colour coded in Figure 6) to the eight output units/emotions (displayed on the vertical axis in Figure 6) were adjusted over time. Small random values were used for the weights of the neural network on initialisation. These weights quickly evolved through the reinforcement learning process.

The fractal dimension has dominance in the neural network, because it can be calculated for every sample image and only consists of one parameter unlike colour or pareidolia. Features originating from pareidolia were not as dominant in the weights of the neural network. The features of pareidolia do not appear as often due to restrictions and complex pre-processing. Weights relating to colour lie between fractal dimension and features from pareidolia, with hue weights dominantly appearing within the features of colour.

Once the robot had been initially trained, it was taken to a number of different venues situated around the University of Newcastle. At these venues the robot was able to collect different images of the different scenes, which were processed on-line. For every image processed, the robot would vocalise its emotions to the user, and the user would respond with a yes or no. During this time, the robot would use these inputs to further update its affective system, in an attempt to gain a closer match to the user's emotional state. Some examples of the environments, and data extracted from the robot perspective can be seen in Figure 7.

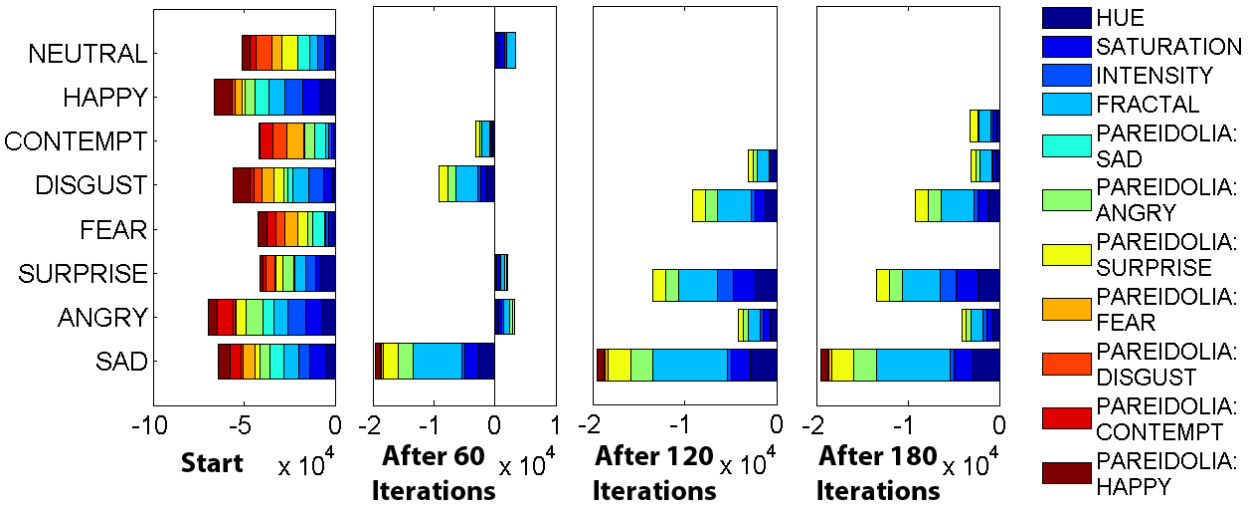

Fig. 6. Cumulative weights of the neural network evolving over time. Initialisation of the neural network weights are random (left), they are then updated during the learning process over time. The vertical axis represents the eight output units. The colour coded boxes represent the weights from the ten inputs. In the horizontal direction four snapshots are displayed (start, after 60 iterations, after 120 iterations, and at the end after 180 iterations). Fractal dimension has a large weighting value in most raw affective responses. Pareidolia features have a weak weighting in the network, as it does not appear as often during training compared to the features of fractal dimension or colour.

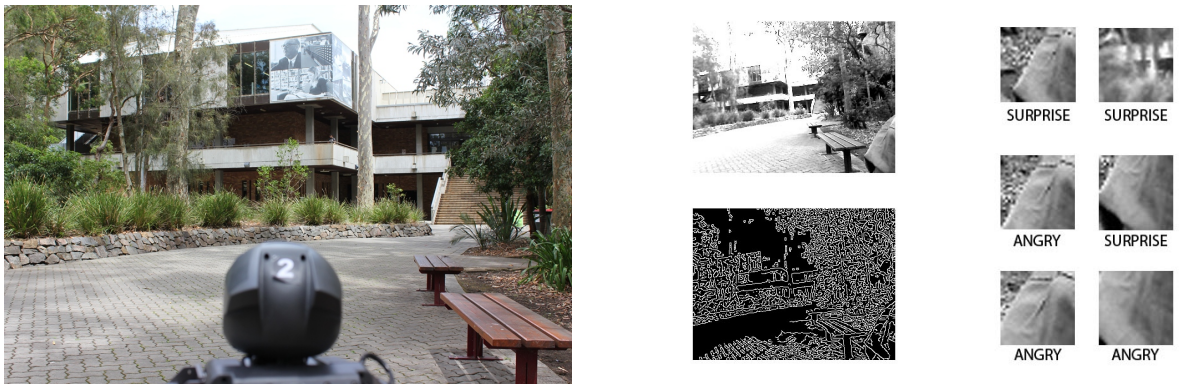

Fig. 7. DARwIn-OP with the partially trained affective system, taken outside to Callaghan Campus at the University of Newcastle for further training. Left: we see the DARwIn-OP and the surrounding environment, from the perspective of the user, Right: an image collected from the DARwIn-OP affective system, which contain grey-scale detected face images as output from the environmental pareidolia module, with their extracted labelled dominant facial expressions. The edge image is the image used to determine the fractal dimension, (Here we obtained a FD of 1.67).

VI. Discussion and Conclusion

This paper presents the commencement of a system that aims towards the development of a complete robot affective system using visual features from the environment in a step to bridge the discomfort gap before integrating robots into everyday society. The environmental emotion detection system implemented on a humanoid robot attempts to adapt to the user through on-line reinforcement learning. The on-line learning process allows the robot to adapt closely to the user's unique emotional status, which has resulted from combinations of cultural, genetic and environmental factors. The environmental emotion detection system allows the robot to predict the user's feelings, to help increase the user's level of comfort in interacting with the robot in various situations.

The inputs of the environmental emotion detection system presented here consist of features obtained from visual perception of the environment that have a "proven" effect on emotions [3, 10, 15, 19–21, 23]. There exist a number of other features known to have an effect on the emotional status of a person, such as the gaze of nearby pedestrians, acoustics, and general lines and shapes in the environment. In our future work, we aim at addressing more of these features.

VII. Acknowledgements

This project was supported by ARC DP1092679 "Modelling and predicting patterns of pedestrian movement: using robotics and machine learning to improve the design of urban space". The authors are grateful to all members of the Newcastle Robotics Laboratory who contributed to the NUbot software system that is part of RoboCup research. Overview of author contributions to the paper: Corresponding author, robot behaviour, system integration, and colour analysis (ASWW); fractal dimension (SN); pareidolia (KH); project supervision (SKC); emotions (PW). The authors are grateful to language editor R. Linich for checking the manuscript.

REFERENCES

[1] M. Mori, "Bukimi no tani [the uncanny valley]," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[2] P. Walla and J. Panksepp, "Neuroimaging helps to clarify brain affective processing without necessarily clarifying emotions," in *Novel Frontiers of Advanced Neuroimaging*, K. N. Fountas, Ed. InTech, 2013, ch. 6, pp. 93–118. [Online]. Available: http://dx.doi.org/10.5772/51761

[3] O. da Pos and P. Green-Armytage, "Facial expressions, colours and basic emotions," *Colour: Design & Creativity*, vol. 1, no. 1, pp. 1–20, 2007.

[4] M. E. Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572 – 587, 2011.

[5] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor, "Emotion recognition in human-computer interaction," *Signal Processing Magazine, IEEE*, vol. 18, no. 1, pp. 32 –80, 2001.

[6] H.-J. Go, K.-C. Kwak, D.-J. Lee, and M.-G. Chun, "Emotion recognition from the facial image and speech signal," in *SICE 2003 Annual Conference*, vol. 3, Aug. 2003, pp. 2890 –2895.

[7] P. Walla, "Olfaction and its dynamic influence on word and face processing: Cross-modal integration," *Progress in Neurobiology*, vol. 84, pp. 192–209, 2008.

[8] P. Kotler, "Atmospherics as a marketing tool," *Journal of Retailing*, vol. 4, no. 4, pp. 48–64, 1973-1974.

[9] S. Wang and X. Wang, "Emotion semantics image retrieval: An brief overview," in *Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, J. Tao, T. Tan, and R. Picard, Eds. Springer Berlin / Heidelberg, 2005, vol. 3784, pp. 490–497.

[10] S. K. Chalup and M. J. Ostwald, "Anthropocentric biocybernetic approaches to architectural analysis: New methods for investigating the built environment," in *Built Environment: Design Management and Applications*, P. S. Geller, Ed. Nova Scientific, 2010, pp. 121–145.

[11] B. B. Mandelbrot, *The Fractal Geometry of Nature*. Freeman, New York, 1977.

[12] D. Harte, *Multifractals : theory and applications / David Harte*. Boca Raton : Chapman & Hall/CRC, 2001.

[13] A. Forsythe, M. Nadal, N. Sheehy, C. Cela-Conde, and M. Sawey, "Predicting beauty: fractal dimension and visual complexity in art," *British Journal of Psychology*, vol. 102, no. 1, pp. 49–70, 2011.

[14] Y. Joye, "Some reflections on the relevance of fractals for art therapy," *The Arts in Psychotherapy*, vol. 33, no. 2, pp. 143 – 147, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S019745560500105X

[15] R. P. Taylor, "Reduction of physiological stress using fractal art and architecture," *Leonardo*, vol. 39, no. 3, pp. 245–251, June 2006.

[16] R. P. Taylor, B. Spehar, J. A. Wise, C. W. G. Clifford, B. R. Newell, C. M. Hagerhall, T. Purcell, and T. P. Martin, "Perceptual and physiological responses to the visual complexity of fractal patterns," *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 9, no. 1, pp. 89 – 114, 2005.

[17] B. Spehar, C. W. Clifford, B. R. Newell, and R. P. Taylor, "Universal aesthetic of fractals," *Computers & Graphics*, vol. 27, no. 5, pp. 813 – 820, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0097849303001547

[18] J. Itten, *The art of color : the subjective experience and objective rationale of color*. John Wiley, New York, 1973.

[19] X. Mao, B. Chen, and I. Muta, "Affective property of image and fractal dimension," *Chaos, Solitons & Fractals*, vol. 15, no. 5, pp. 905 – 910, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960077902002096

[20] S. Wang, R. Ding, Y. Hu, and H. Wang, "Analysis of relationships between color and emotion by classification based on associations," in *Computer Science and Software Engineering, 2008 International Conference on*, vol. 1, Dec. 2008, pp. 269 –272.

[21] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proceedings of the International Conference on Multimedia*. New York, NY, USA: ACM, 2010, pp. 83–92. [Online]. Available: http://doi.acm.org/10.1145/1873951.1873965

[22] S. K. Chalup, K. Hong, and M. J. Ostwald, "A face-house paradigm for architectural scene analysis," in *CSTST 2008: Proc. of The Fifth International Conference on Soft Computing as Transdisciplinary Science and Technology*, R. Chbeir, Y. Badr, A. Abraham, D. Laurent, and F. Ferri, Eds. ACM, 2008.

[23] S. K. Chalup, K. Hong, and M. Ostwald, "Simulating pareidolia of faces for architectural image analysis," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 2, pp. 262–278, 2010.

[24] N. Kanwisher and G. Yovel, "The fusiform face area: a cortical region specialized for the perception of faces," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1476, pp. 2109–2128, 2006.

[25] P. Vuilleumier, "Neural representations of faces in human visual cortex: the roles of attention, emotion, and viewpoint," in *Object Recognition, Attention and Action*, N. Osaka, I. Rentschler, and I. Biederman, Eds. Springer, Tokyo Berlin Heidelberg New York, 2007, pp. 109–128.

[26] M. Finkbeiner and R. Palermo, "The role of spatial attention in nonconscious processing: A comparison of face and nonface stimuli." *Psychological Science*, vol. 20, pp. 42–51, 2009.

[27] I. Ha, Y. Tamura, H. Asama, J. Han, and D. Hong, "Development of open humanoid platform darwin-op," in *SICE Annual Conference (SICE), 2011 Proceedings of*, 2011, pp. 2178 –2181.

[28] J. Kulk and J. S. Welsh, "A nuplatform for software on articulated mobile robots," in *Leveraging Applications of Formal Methods, Verification, and Validation*, ser. Communications in Computer and Information Science, R. Hähnle, J. Knoop, T. Margaria, D. Schreiner, and B. Steffen, Eds. Springer Berlin Heidelberg, 2012, pp. 31–45.

[29] A. R. Smith, "Color gamut transform pairs," *SIGGRAPH Comput. Graph.*, vol. 12, no. 3, pp. 12–19, 1978. [Online]. Available: http://doi.acm.org/10.1145/965139.807361

[30] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall, 2008.

[31] A. Hanbury, "A 3d-polar coordinate colour representation well adapted to image analysis," in *Image Analysis*, ser. Lecture Notes in Computer Science, J. Bigun and T. Gustavsson, Eds. Springer Berlin Heidelberg, 2003, vol. 2749, pp. 804–811.

[32] ——, "Constructing cylindrical coordinate colour spaces," *Pattern Recogn. Lett.*, vol. 29, no. 4, pp. 494–500, Mar. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2007.11.002

[33] L. S. Liebovitch and T. Toth, "A fast algorithm to determine fractal dimensions by box counting," *Physics Letters A*, vol. 141, no. 8–9, pp. 386 – 390, 1989. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0375960189908542

[34] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science*, 1st ed. New York: Springer-Verlag, 1992.

[35] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed. West Sussex: John Wiley & Sons Ltd., 2003.

[36] N. Sarkar and B. Chaudhuri, "An efficient differential box-counting approach to compute fractal dimension of image," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 24, no. 1, pp. 115 –120, 1994.

[37] K. Hong, S. K. Chalup, and R. A. R. King, "Scene perception using pareidolia of faces and expressions of emotion," manuscript accepted for publication, IEEE Symposium on Computational Intelligence for Creativity and Affective Computing, CICAC 2013 (accepted).

[38] ——, "An experimental evaluation of pairwise adaptive support vector machines," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, June 2012, pp. 1 –8.

[39] P. Ekman and W. Friesen, "Constants across cultures in the face and emotion." *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[40] P. Ekman and W. V. Friesen, "A new pan-cultural facial expression of emotion," *Motivation and Emotion*, vol. 10, no. 2, pp. 159–168, 1986.

[41] D. Aberdeen, O. Buffet, F. P. Selmi-Dei, X. Zhang, and T. Lopes., "libpgrl," Software, 2007. [Online]. Available: http://code.google.com/p/libpgrl/

[42] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, Cambridge, 1998.

[43] S. K. Chalup, N. Henderson, M. Ostwald, and L. S. Wiklendt, "A computational approach to fractal analysis of a cityscape's skyline," *Architectural Science Review*, vol. 52, pp. 126–134, 2009.