A Neural Wake-Sleep Learning Architecture for Associating Robotic Facial Emotions

Chi-Yung Yau, Kevin Burn and Stefan Wermter

Abstract—**A novel wake-sleep learning architecture for processing a robot's facial expressions is introduced. According to neuroscience evidence, associative learning of emotional responses and facial expressions occurs in the brain in the amygdala. Here we propose an architecture inspired by how the amygdala receives information from other areas of the brain to discriminate it and generate innate responses. The architecture is composed of many individual Helmholtz machines using the wake-sleep learning algorithm for performing information transformation and recognition. The Helmholtz machine is used since its re-entrant connections support both supervised and unsupervised learning. Potentially it can explain some aspects of human learning of emotional concepts and experience. In this research, a robotic head's facial expression dataset is used. The objective of this learning architecture is to demonstrate the neural basis for the association of recognized facial expressions and linguistic emotion labels. It implies the understanding of emotions from observation and is further used to generate facial expressions. In contrast with other facial expression recognition research, this work concentrates more on emotional information processing and neural concept development, rather than a technical recognition task. This approach has a lot of potential to contribute towards neurally inspired emotional experience in robotic systems.**

I. Introduction

Most current interface robots and interactive agents have no capability of understanding the complexity of the human interaction, or demonstrating empathy or emotion. However, it is anticipated that future robots should be able to empathise with users and generate appropriate expressions to communicate to the user their emotive states during interaction [1].

In recent years, many researchers have investigated how emotional information can be processed and the emotional stimulus-response phenomenon in the human brain. These findings have the potential to help develop computational systems that can be installed in robots to make human-robot interaction more natural and satisfying [2]. There are several important brain areas and mechanisms involved in emotional information processing, including the amygdala, the insula, and the mirror system [17]. An overview of the various brain areas involved is shown in figure 1 and in this paper we concentrate on the amygdala region as an important area for emotion recognition. The aim of this research is to investigate the representation of robotic facial expression information, in order to facilitate the design of emotive computational architectures.

Chi-Yung Yau, Kevin Burn and Stefan Wermter are with the Hybrid Intelligent Systems research group, School of Computing and Technology, University of Sunderland, U.K. (e-mail: alan.yau@sunderland.ac.uk; kevin.burn@sunderland.ac.uk; stefan.wermter@sunderland.ac.uk). And this work has been funded partially by NEST-043374 coordinated by SW.

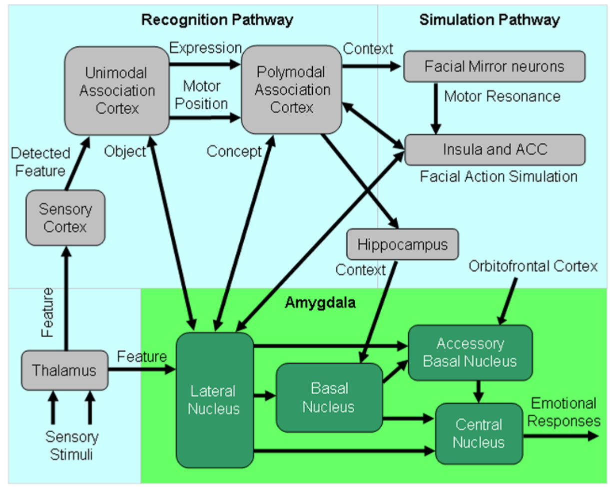

Figure 1. Diagram of the thalamus, the sensory cortex, the association cortex, mirror neurons, insula and amygdala for emotional information processing in the human brain.

The paper is structured as follows. Section II discusses the brain areas involved in emotion and facial expression information processing, and how we can recognize emotion from facial expressions. In section III there is a short introduction to the biologically-inspired neural network (Helmholtz machine) and its learning algorithm. Section IV presents the design of our architecture, whilst section V outlines recent experiments and results. Finally, conclusions and future work are summarized in section VI.

II. Neural Mechanisms Underlying Emotion

The human brain is a complex information processing system and emotions interfere at multiple levels of cognitive information processes. Referring to the H-Cogaff cognitive architecture schema [3, 4], varieties of emotions are not only considered as alarms to trigger responses in the reactive level but also interfere with different level of cognitive processes.

Such emotional information processing mechanisms can be found in a part of the brain called the amygdala, which is located within the medial temporal lobes (see figure 1). Although the amygdala appears to be intertwined at all stages of emotional cognitive processing, its role can be summarised as evaluating and regulating emotion valence, emotional learning, and modulating emotional memory and behaviours [5-10]. Previously, the amygdala was thought to activate for negative emotions only, but there is increasing evidence suggesting that it activates for both positive and negative emotions [11]. Also, looking inside the amygdala's internal structure, there are four main nuclei: Lateral nucleus (L), Basal nucleus (B), Accessory Basal nucleus (AB) and central nucleus (Ce). L receives stimuli from the sensory cortex via the thalamus and unimodal association cortex for recognized objects and concepts. They are sent to the B, which contains many connections to the hippocampus for emotional context formation. The AB connects to the frontal cortex for attention and emotional behaviour control, and the central nucleus connects to the basal ganglia and brainstem for generating reactive responses and regulating the behaviours via varying physiological arousal.

Two sides of the amygdala are involved in different roles and tasks. Most neuroscience scientists state that the amygdala exhibits different activations for both negative and positive basic facial expressions, but different activation behaviours between the left and right sides of the amygdala [12]. Also, the amygdala contains greater activation in explicit emotion identification tasks rather than in implicit age identification tasks [13].

There are three ways to recognize emotions from facial expressions: direct recognition from stimuli without any manipulation, recognition via associated knowledge, and recognition via facial actions simulation [14]. When raw facial expressions are projected to the amygdala directly via the thalamus, this is normally called the "low route" and is mainly for producing immediate responses. Alternatively, facial expressions can be recognized as different objects and be associated with different concepts via the "what" cortical visual pathways. The results are then sent to the amygdala for generating object-emotion and concept-emotion association responses [15].

In addition, the spatial movement and location of facial motors can be detected via the "where" cortical visual pathway. Such detections will activate the face-related mirror neurons, whose processing is called *motor resonance*, and which perform the facial action simulation. In this case, the simulated action representations are interpreted at the insula. This is called *neural mechanisms of empathy* or *motor theory of empathy* [16-18]. When the amygdala receives such action representations, the relevant emotional responses will be generated and fed back to the insula and anterior cingulated cortex (ACC) for eliciting bodily representation of emotions (called feelings) such as disgust and pain [19, 20]. Due to the presence of this mechanism of empathy, humans can observe others and evoke the activity of corresponding motor neurons; simulate and retrieve the goal and meaning of actions; and understand the deeper and ambiguous emotional meaning of expressions.

After reviewing the neural mechanisms of emotion and facial expressions, it is observed that emotional information processing in the brain involves numerous top-down and bottom-up connections between the cortex and amygdala. The emotional information processing of facial expressions should be not only performing recognition, but also the reconstruction/regeneration of expressions from high level concepts.

III. Helmholtz Machine And Wake-Sleep Algorithm

The reason for developing this neurocognitive architecture is to demonstrate how emotional information associates with specific emotion labels. Thus, a statistical artificial neural network called a Helmholtz Machine (HM) is used because it supports bottom-up and top-down connections that can potentially explain how we recognise the concepts of emotion and reconstruct emotional facial expressions.

A Helmholtz machine models cortical bottom-up and top-down pathways in our perceptual system [21-24]. It operates like a statistical engine performing density estimation, in order to transform the sensory feature input into a reduced internal representation. There are two processing models: a recognition model and a generative model. A recognition model is used to estimate a probability distribution from the inputs and represents them in the higher layer; indeed, it is performing the discrimination of inputs. A generation model is a reverse top-down model that is able to reconstruct the output, which is generated by the recognition model. During learning, each stochastic neuron at the higher/hidden layer (B) is the weighted sum of the input layer (A). Conversely, at the input layer is the weight sum of the hidden layer (see figure 2a). They are defined as follows:

$$\delta(\Pr(s_y^b)) = \delta(1/(1+\exp(-\sum_x s_x^a \cdot R_{xy}^{ab} + b_y^R))) \quad (1)$$

where δ is the stochastic function whose output is 1 with Pr(s) and 0 with 1-Pr(s); s^b is the probability state of a neuron in the hidden layer with index y; R^{ab} is the recognition weight; and b^R is the bias.

$$\delta(\Pr(s_x^a)) = \delta(1/(1+\exp(-\sum_y s_y^b \cdot G_{yx}^{ba} + b_x^G))) \quad (2)$$

where s^a is the probability state of a neuron at the input layer with index x, G^{ba} is the generative weight and b^G is the bias.

The wake-sleep learning algorithm is primarily used in the stochastic Helmholtz machines and its goal is not only to learn an economic representation to describe the observed inputs, but also to accurately reconstruct the inputs. There are two learning phases, 'wake' and 'sleep'. In the wake phase, a sampled input is fed into the recognition model for estimating an economic output (see eq.1), and then it is used to reconstruct the input via the generative model (see eq.2). By comparing with the original sampled input, the difference can be used to update the generative weight by the delta rule. They are defined as follows:

$$G_{yx}^{ba} = G_{yx}^{ba} + \eta(s_x^a - s_x^{aG})s_y^b \quad (3)$$

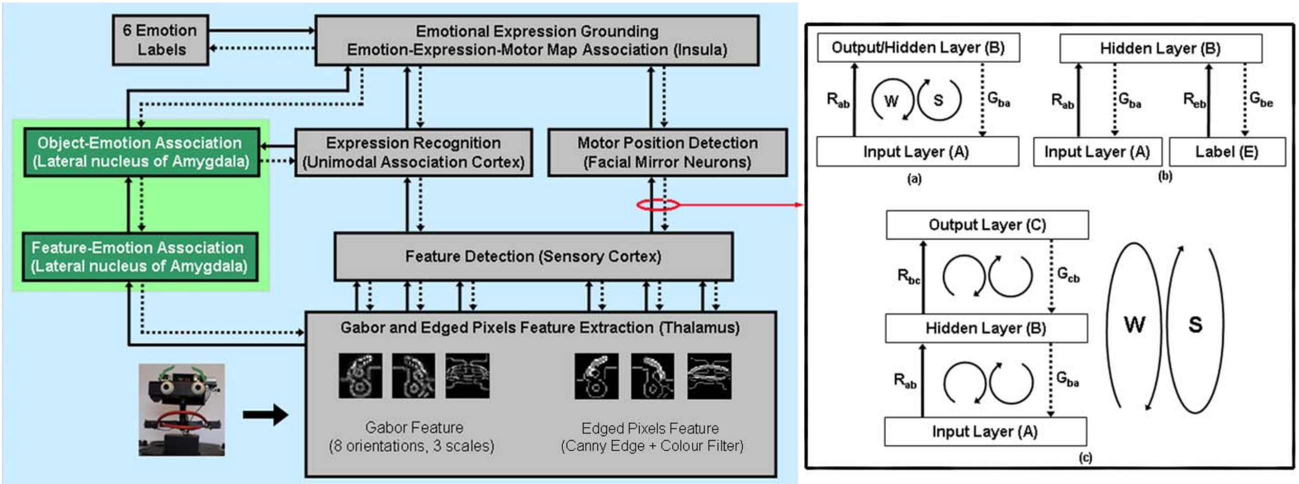

Figure 2. On the left hand is the overall architecture. On the right hand, there are three basic Helmholtz machines showing the way of wake(W)-sleep(S) phases learning : (a) is the single-layer model for demonstrating the bottom-up association and transformation of information; (b) is the model to link two single layer HMs side-by-side. They share the same hidden layer representing the association where either layer can be specified to represent the target label; (c) also contains two HM, but they are connected together hierarchically. The lower HM's output is connected to upper HM's input. This connection model is commonly used in recognition tasks.

$$b_x^G = b_x^G + \eta(s_x^a - s_x^{aG})(-1) \tag{4}$$

Thus, the generative weight's updating is driven by the recognition model in order to increase the chance of reconstructing an accurate input. In the sleep phase, a stochastic pattern or a supervised pattern is fed into the generative model and reversely performs the reconstruction. This phase is driven by the generative model and updates the recognition weight with the aim of improving the estimation. The equations are:

$$R_{xy}^{ab} = R_{xy}^{ab} + \eta(s_y^b - s_y^{bR})s_x^a \tag{5}$$

$$b_y^R = b_y^R + \eta(s_y^b - s_y^{bR})(-1) \tag{6}$$

In addition, Oja's decay term is added to limit the rate of growth of each weight:

$$R_{xy}^{ab} = R_{xy}^{ab} - \eta\varepsilon((s_y^b)^2 \cdot R_{xy}^{ab}) \tag{7}$$

$$G_{yx}^{ba} = G_{yx}^{ba} - \eta\varepsilon((s_x^a)^2 \cdot G_{yx}^{ba}) \tag{8}$$

The wake-sleep algorithm is applied not only in the Helmholtz machine, but also in the restricted Boltzmann machine (RBM) for performing hand written character recognition and face recognition [25, 26]. Six layer hierarchical networks have been used to auto-encode images into 30 dimensions and good results have been obtained that outperform principal component analysis (PCA). In fact RBM is very similar to HM, although RBM has symmetrical connections, whereas HM does not. Therefore, in this research the wake-sleep algorithm was chosen to construct the whole architecture, both for being biologically-inspired while still having a reasonable training and testing performance for networks.

IV. Architecture Design

A. Neuroscience-Inspired Design

For development purposes, firstly we assume the cortices are responsible for analyzing the facial expressions. The lateral nucleus of the amygdala is responsible for categorising facial expressions into different type of responses. The role of the central nucleus of the amygdala is to process actual emotional responses for regulating behaviours.

Secondly, we assume humans have the innate capability to analyse facial expressions via direct recognition and facial motor position detection. The lateral nucleus of the amygdala should contain the feature-emotion association, the object-emotion association, and also some higher conceptual and contextual level association. Due to the complexity of development, some brain structures and their functions are omitted in the architecture, for example the mirror system for action simulation and recognition, and the hippocampus and basal nucleus of the amygdala for emotional memory formation and retrieval.

Thirdly, depending upon the nature of the brain structure, we can specify whether the networks are supervised for acquired feature detection, where external information is required, or unsupervised for innate feature detection and categorisation. It is intended to let the architecture learn user's expressions online in the future. The detailed configurations of the networks chosen for this work are shown in table 1, which will be discussed further in section V.

Fourthly, a symbolic approach is applied [27-29] where different linguistic emotion labels are grounded into internal cognitive representations, including the perception of facial features, recognition of basic facial expressions, and simulation of facial actions. In the following sections, the explanation of each part is discussed.

Table 1. The configurations of the Helmholtz machines

Network	Input Layer (A)	Output Layer (B)	Stage 1	Stage 2	Stage 3
Left Eyebrow: Pixel feature (lebcf_net)	(30*30)=900	16	unsupervised	supervised (exp_net, pos_net)	n/a
Left Eyebrow: Gabor feature (lebgf_net)	(30*30*24)=21600	16	unsupervised	supervised (exp_net, pos_net)	n/a
Right Eyebrow: Pixel feature (rebcf_net)	(30*30)=900	16	unsupervised	supervised (exp_net, pos_net)	n/a
Right Eyebrow: Gabor feature (rebgf_net)	(30*30*24)=21600	16	unsupervised	supervised (exp_net, pos_net)	n/a
Lips: Pixel feature (lipcf_net)	(60*53)=3180	25	unsupervised	supervised (exp_net, pos_net)	n/a
Lips: Gabor feature (lipgf_net)	(60*53*24)=76320	25	unsupervised	supervised (exp_net, pos_net)	n/a
Expressions Recognition (exp_net)	114	64	n/a	unsupervised	n/a
Motor Position Detection (pos_net)	114	15	n/a	supervised (predefined label)	n/a
Feature-Emotion Association: Hidden layer (emo_hdd)	124500	64	supervised (emo_net)	n/a	n/a
Feature-Emotion Association: output layer (emo_net)	64	4	supervised (predefined label)	n/a	n/a
Object-Emotion Association (objemo_net)	(4+64)=68	x (undetermined)	n/a	n/a	unsupervised
Emotion-Expression-Motor Grounding (emoexp_net)	(64+15+x)	6	n/a	n/a	supervised (ext. label)

* "predefined label" means that it is innate and no leaning is required.
* "ext. label" means that it comes from language cortex and external input is required.

B. Feature Extraction

A feature layer represents the features after pre-processing, including the edge-colour and Gabor features. A robot facial expression image data set is used in our research (see section V for details of the data set) where each image is decomposed into three regions: left eyebrow, right eyebrow and lips. Currently, a Gabor filtering toolbox [30] is used for extracting 8 orientations and 3 scaled Gabor features directly. Therefore, there are in total six subsets of features, including 4980 pixel features (left eye brow: 900 + right eye brow: 900 + lips: 3180) and 119520 Gabor features (21600 + 21600 + 76320).

C. Feature Detection

There are six single-layer HMs for six feature subsets. The role of each HM is to detect and extract hidden features from lower-level features. The single-layer HM can be configured to estimate the components from the features and form an economic binary representation. Six feature detection networks perform individually to produce a total of 114 hidden features to represent six regions of pixel and Gabor features.

D. Expression Recognition

Expression recognition is a process that models the "what" stream of visual processing. After feature detection (dimension reduction), the features of a facial expression are further recognized as different expression cognitive symbols. The single layer HM is used as an association memory that memorizes the expressions in the network weights. The size of this HM is equal to the amount of single facial expressions that humans can memorize. Currently, unsupervised learning is applied to transfer the detected features into 64 predefined expression categories. Unsupervised learning is a machine learning approach that does not require any target outputs during learning and purely performs a categorisation task based on observing the difference between inputs and the number of categories specified.

E. Facial Motor Position Detection

This section is not the same as expression recognition, but is inspired by the mirror system that allows humans to recognize facial motor actions via motor resonance. It is thought that this capability is innate and may be well organized in the brain from birth. In our architecture, a one-layer HM with supervised learning is used. The representation of this layer is a motor map, the same as the one used for controlling facial actions. In this architecture, 15 binary units are specified, which represent 3 motor positions for each side of eyebrow and 9 for the lips. In terms of information, this HM transfers the detected features into motor positions.

F. Feature-Emotion Association

Here the lateral nucleus amygdala deals with the categorization of emotional stimuli into a predefined number of groups. This is a reactive level association process that associates features with emotional activation when the feature is directly fed into the amygdala from the thalamus (see figure 1). The output is then interpreted as different individual symbolic activations. It is assumed only four types of activation can appear at the lateral nucleus of the amygdala

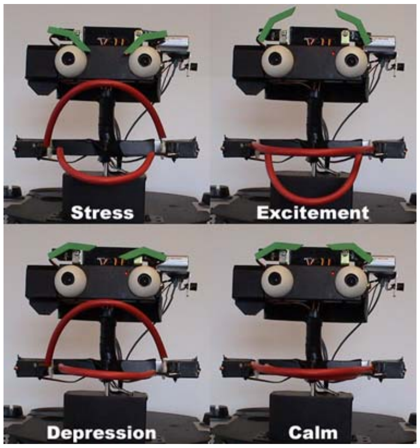

Figure 3. Example of robotic expressions with four assigned feature-level emotion labels.

i.e. positive-high arousal, positive-low arousal, negative-high arousal and negative-low arousal. A two-layer HM (refer to figure 2c) with supervised learning is used rather than just a one-layer HM, since the additional hidden layer improves the association.

G. *Object-Emotion Association*

This is similar to the previous feature-emotion association, but is configured for categorizing the stimuli into different pre-defined object activations. It is planned to use a single layer HM and unsupervised learning.

H. *Emotion-Expression-Motor Map Grounding*

This is the output layer of the whole architecture that associates all the emotional labels with the activation of the amygdala, recognised facial expressions and detected motor positions together, to form higher level concepts. Supervised learning will be used to ground all internal states to linguistic emotion labels.

V. Experiments and Result

To evaluate the architecture, a robot expression image dataset has been produced. There are 144 robotic expression images, including all combinations of the motor positions of eyebrows and lips, with manually assigned feature emotion labels (see figure 3). At feature level, we have grouped them into four emotion categories: excitement, stress, depression and calm, with the meaning positive-high arousal, negative-high arousal, negative-low arousal and positive-low arousal respectively.

The development work of the architecture is divided into three stages (see table 1). In the first stage, all feature manipulations from the brain structures that connect to the thalamus are included, such that the sensory cortex is represented by six feature detection networks and the lateral nucleus of amygdala is represented by feature-emotion association network. In the second stage, the expression recognition network and motor position detection network will be trained. The remaining networks, including object-emotion association and emotional expression grounding, will be implemented at the third stage. In this paper we focus on reporting the results of the first stage as follows.

In order to allow the HM to learn more efficiently, three approaches have been implemented. Firstly, the network performed the learning more often in the wake phase than the sleep phase. Since HM is generative-based learning, to let the wake phase learn more we can increase the chance of correctly adjusting the generative weights toward the input pattern via recognition models. Secondly, the lower level network was set to learn in unsupervised mode at stage one and changed to supervised mode to fine-tune the lower-level network. Thirdly, two different learning rates - one for the wake phase and one for the sleep phase - are applied. The wake phase is playing the role of M-step [31], so the increase of its learning rate can maximise the probability density (the likelihood) of the observed inputs. Alternatively, the sleep phase learning rate is playing the role of E-Step. The decrease in the learning rate can improve the estimation of the probability density of parameters, as if observing the inputs.

To evaluate the architecture, both the estimation of factors from input and the reconstruction of input from those factors

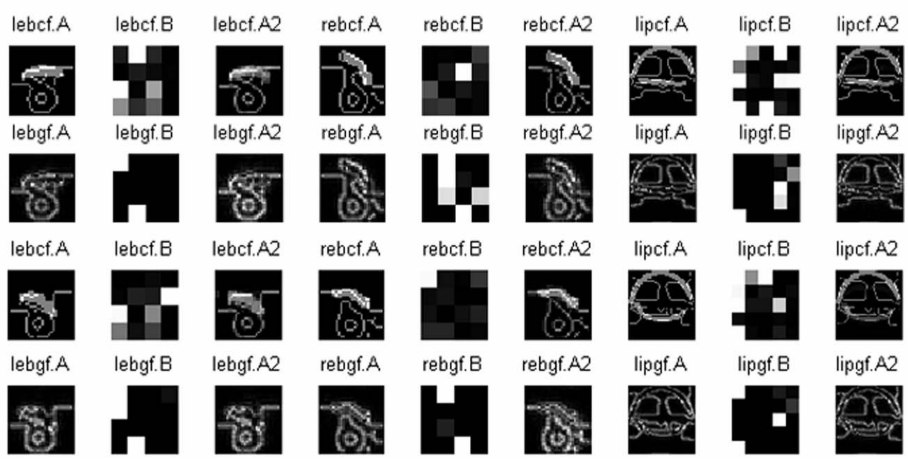

Figure 4. Test results of the feature detector recorded at stage one. The first row is the pixel feature image and the second row is the Gabor feature image. The first column is the input, the second column is the output representing the probability distribution, and the third column is the reconstructed image and so on.

are monitored. Firstly, the performance of HMs in categorising non-binary input patterns for feature detection is monitored. A good result of six feature detection networks has been recorded after applying the first approach to let the wake phase learn more. This is shown in figure 4. The factors of non-binary input features are estimated and represented by fewer binary units at the output layer. In addition, their reconstructions are clear.

In figure 5, feature-emotion association network results are generated from one sample input. Although a two-layer HM was designed, the implementation used two single-layer HMs “emo_hdd” and “emo_net” instead. The facial feature is fed into the input layer (“emo_hdd.A”) and then transferred to the output layer (“emo_hdd.B”). The output of “emo_hdd” was connected to the input layer of “emo_net” (“emo_net.A”) and the result obtained at output layer (“emo_net.B”).

In figure 6, the feature detection results after applying all three HM optimization approaches are shown. The results are presented region-by-region in order to clearly indicate how HMs discriminate input features and represent them by units.

VI. Discussion

The main contribution of this paper is to model the amygdala’s processing of facial expression information and link this to external emotion labels, in order to develop emotion concepts. A simplified robotic facial expression dataset has been used.

For feature detection, the results show that the input data is successfully represented by few unique units, which can be used to reconstruct the low-level input feature precisely. Some noise (or “non solid” units) can be found among the detected binary units shown in figure 4. These are the units that have not fully learnt or are overlapping with other features. However, this can be improved at a later stage by applying different wake-sleep learning rates and the result is shown in figure 6. By observing these results, we can primarily conclude that the HMs should able to detect and estimate factors from both pixel features and large-dimensionality Gabor features by an unsupervised learning approach.

For feature-emotion association, the result shown in figure 5 is the representation of the type of activation responses at the lateral nucleus of the amygdala. At a reactive level, we assume this association is an innate capability and performed rapidly; therefore, as previously discussed, four possible categories to the inputs were assigned. A reasonable and explainable result has been collected and although the reconstructed image is less clear than the original, the factors of low level features are successfully estimated. The reason is that we only group them into four possible emotion categories, so overlapping occurs and is shown at the reconstruction. In addition, in contrast with six feature detection networks, the hidden layer of the emotion network is trying to perform the categorisation of inputs all-in-one, without dividing them into regions (left and right eyebrows and lips). The total combination of inputs is much larger than

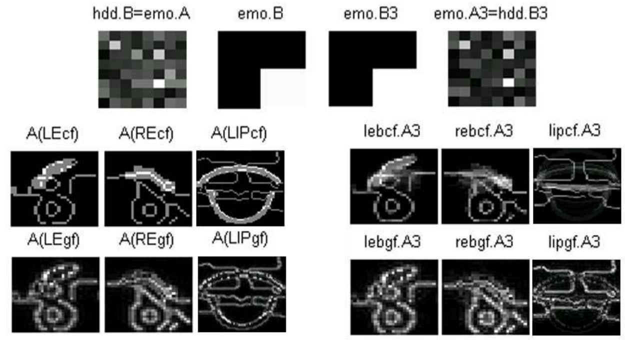

Figure 5. The information flow from bottom network “emo_hdd” to higher network “emo_net”. This result is collected at iteration=100000.

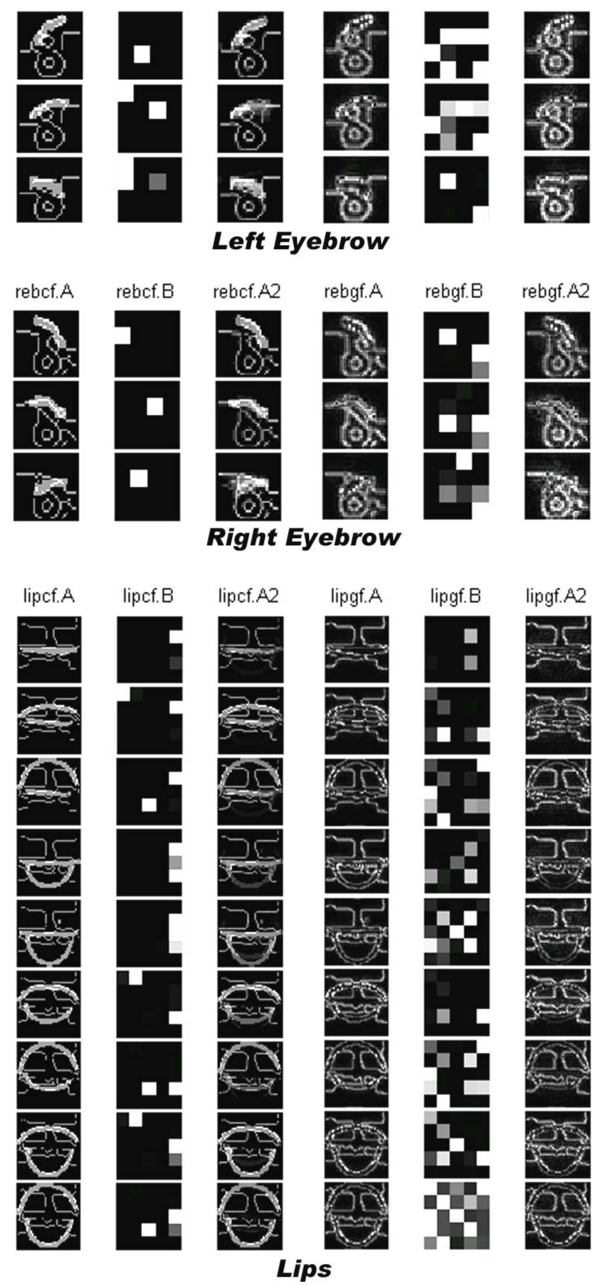

Figure 6. The complete feature detection results after applying all HM optimization approaches.

the feature detection network. Considering HM is a statistical neural network which uses a stochastic learning approach, it may be that it is sampling all the combinations that require the longest times to process, so that the learning and optimising of HMs can become very difficult.

Another difficulty of the implementation is the estimation of the initialization parameters for the HMs. In particular, in this scenario the low level pixel and Gabor features are not binary features and the difference between data can be small, especially for the Gabor features. Also, non-binary inputs make the HM more difficult to initialise and optimise, even when a weight decay term is added. It may be that the problem can be solved by replacing the two separated connection weights with symmetric connection weights, similar to some of the latest approaches using restricted Boltzmann machines with wake-sleep algorithms [25, 26].

Future work should include: the continuous implementation of the complete architecture; the integration of a basic expression behaviour controller to control the action of a robot head; further detailed analysis of the architecture, such as the performance of the retrieval of face images and motor position map based on higher level concepts (i.e. "happy smiling face" can retrieve happy emotion, smiling face and its motor position).

In conclusion, the paper has summarised how the brain processes emotional information and the basic theory of the Helmholtz machine. A design and development description of the architecture has been provided, including the assumptions made and how the system has been biologically-inspired at the architectural level. Primary experimental results have demonstrated that the Helmholtz machines work in both supervised and unsupervised learning modes. Based on the bottom-up recognition results and top-down reconstructed results, it can explain how humans recognise facial expressions and generate feature level emotional activation. The research will be extended to associate those internal recognition results with linguistic emotional labels, in order to develop specific emotional experiences.

References

[1] R.W. Picard and J. Klein, "Computers that Recognise and Respond to User Emotion: Theoretical and Practical Implications," *Interacting with Computers*, vol.14, issue 2, pp. 141-169, 2002.

[2] M.A. Arbib and J.M. Fellous, "Emotions: from brain to robot," *Journal of TRENDS in Cognitive Sciences*, vol.8, no.12, pp.554-561, 2004.

[3] A. Sloman, "Beyond Shallow Models of Emotion," *Cognitive Processing*, vol.12, no.1, pp.117-198, 2001.

[4] A. Sloman and R.L. Chrisley, "More things than are dreamt of in your biologically inspired robots," *Cognitive Systems Research*, vol.6, pp.145-174, 2005.

[5] E.A. Phelps, "Emotion and Cognition: Insights from Studies of the Human Amygdala," *Annual Review of Psychology*, vol.57, pp.27-53, 2006.

[6] R. Adolphs, "Neural systems for recognizing emotion," *Journal of Current Opinion in Neurobiology*, vol.12, pp.169-177.

[7] E.T. Rolls, "Vision, emotion and memory: from neurophysiology to computation," *International Congress Series*, vol.1250, pp.547-573, 2003.

[8] J. Zhu and P. Thagard, "Emotion and action," *Philosophical Psychology*, vol.15, no.1, pp.19-36, 2002.

[9] J.E. LeDoux, "Emotion Circuits in the Brain," *Annual Review of Neuroscience*, vol.23, pp.155-184, 2000.

[10] G. Richter-Levin, "The Amygdala, the Hippocampus, and Emotional Modulation of Memory," *The Neuroscientist*, vol.10, no.1, pp.31-39, 2004.

[11] M.A. Williams, F.McGlone, D.F. Abbott and J.B. Mattingley, "Differential amygdala response to happy and fearful facial expressions depend on selective attention," *NeuroImage*, vol.24, pp.417-425, 2005.

[12] D.A. Fitzgerald, M. Angstadt, L.M. Jelsone, P.J. Nathan and K.L. Phan, "Beyond threat: Amygdala reactivity across multiple expressions of facial affect," *NeuroImage*, vol.30, pp.1441-1448, 2006.

[13] U. Habel, C. Windischberger, B. Derntl, S. Robinson, I. Kryspin-Exner, R.C. Gur and E. Moser, "Amygdala activation and facial expressions: explicit emotion discrimination versus implicit emotion processing," *Neuropsychologia*, vol.45, pp.2369-77, 2007.

[14] R. Adolphs, "Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms," *Behavioral and Cognitive Neuroscience Reviews*, vol.1, no.1, pp.21-61, 2002.

[15] C.I. Hooker, L.T. Germine, R.T. Knight and M.D'Esposito, "Amygdala Response to Facial Expressions Reflects Emotional Learning," *The Journal of Neuroscience*, vol.26, no.35, pp.8915-8922, 2006.

[16] L. Carr, M. Lacoboni, M.C. Dubeau, J.C. Mazziotta and G.L. Lenzi, "Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas," *Proceedings of the National Academy of Sciences of the United States of America*, vol.100, no.9, pp.5497-5502, 2003.

[17] K.R. Leslie, S.H. Johnson-Frey and S.T. Grafton, "Functional imaging of face and hand imitation: towards a motor theory of empathy," *NeuroImage*, vol.21, pp.601-607, 2004.

[18] V. Gallese, C. Keysers and G. Rizzolatti, "A unifying view of the basis of social cognition," *Journal of TRENDS in Cognitive Sciences*, vol.8, no.9, pp.396-403, 2004.

[19] B. Wicker, C. Keysers, J. Plailly, J.P. Royet, V. Gallese and G. Rizzolatti, "Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust," *Neuron*, vol.40, pp.655-664, 2003.

[20] P.L. Jackson, E. Brunet, A.N. Meltzoff and J. Decety, "Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain," *Neuropsychologia*, vol.44, pp.752-761, 2006.

[21] P. Dayan, G.E. Hinton, R.M. Neal and R.S. Zemel, "The Helmholtz Machine," *Neural Computation*, vol. 7, pp.889-904, 1995.

[22] G.E. Hinton, P.Dayan, B.J. Frey, R.M. Neal, "The wake-sleep algorithm for unsupervised neural networks," *Science*, vol. 268, no.5214, pp.1158-1161, 1995.

[23] P. Dayan and G.E. Hinton, "Varieties of Helmholtz machine," *Neural Networks*, vol.9, pp.1385-1403, 1996.

[24] K.G. Kirby, *A tutorial on Helmholtz Machines*, Department of Computer Science, Northern Kentucky University, 2006.

[25] G.E. Hinton, *To Recognize Shapes, First Learn to Generate Images*, Department of Computer Science, University of Toronto & Cannadian Institute for Advanced Research, 2006

[26] G.E. Hinton and R.R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol.313, 2006.

[27] S. Harnad, "The Symbol Grounding Problem," *Encyclopedia of Cognitive Science*, Nature Publishing Group, 2003.

[28] A. Cangelosi, "Approaches to Grounding Symbols in Perceptual and Sensorimotor Categories," In: H. Cohen and C. Lefebvre (Eds), *Handbook of Categorization in Cognitive Science*, Elsevier, 2005.

[29] E. Chinellato, A. Morales, E. Cervera and A.P. del Pobil, "Symbol grounding through robotic manipulation in cognitive systems, *Robotics and Autonomous Systems*, vol.55, pp.851-859, 2007.

[30] J.Ilonen, J.K. Kamarainen and H. Kalviainen, *Efficient Computation of Gabor Features*, Department of Information Technology, Lappeenranta University of Technology, Research report, ISSN 0783-8069, vol. 100, 2005.

[31] R.M. Neal and P. Dayan, *Factor Analysis Using Delta-Rule Wake-Sleep Learning*, Department of Statistics, University of Toronto, Technical Report, no. 9607, 1996.