2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

Applying LCS to Affective Image Classification in Spatial-Frequency Domain

Po-Ming Lee, Tzu-Chien Hsiao

Abstract—**Affective image classification is a task aims on classifying images based on their affective characteristics of inducing human emotions. This study achieves the task by using Learning Classifier System (LCS) and spatial-frequency features. The model built by using LCS achieves Area Under Curve (AUC) = 0.91 and accuracy rate over 86%. The result of the LCS is compared with other traditional machine-learning algorithms (e.g., Radial-Basis Function Network (RBF Network)) that are normally used in classification tasks. The study presents user-independent results which indicate that the horizontal visual stimulations contribute more to the emotion elicitation than the vertical visual stimulation.**

I. Introduction

People experience emotion in their daily life by feeling happy, angry and various emotions induced by stimulus and events that are emotionally relevant. Because it is human nature to pursue happiness and avoid pain, the research finding related to human emotion can be easily transferred to diverse applications. For example, behavioral economics [1], media studies and advertisement [2, 3]. Some researches focused on the use of emotional relevant stimulus to attract the attention of subjects, and to make subjects remember more on the product presented [3]. In the area of image, print advertisement and the use of affective images for attracting the attention of subjects during web browsing were reported [2]. Guideline of extracting emotional relevant features in a web page is also available [4].

Due to the development of personal computer, software and World-Wide-Web (WWW), people nowadays generate huge amount of content (e.g., daily news, articles on variety topics and personal data) and upload them to the internet every day. To enable end users to explore the content on the internet, Google and Yahoo! such the search engine provider index these contents. Currently most of the web content indexing works are done based on text-based technologies. Although text-based indexing technologies are suitable for articles, the limitation of the text-based method is obvious when images are the indexing target. Traditionally, image search is done based on the file name of the target image, and the description (e.g., tags) of the target image. Despite image search based on content has been provided recently (e.g., Google Picture and Yahoo! Image Search), such the method still leaves rooms for improvement. Hence, our study aims to demonstrate a novel technique for indexing and classifying the affective characteristics of images. Equipped with this technology, a search engine would be able to lead end users to target images that may potentially improve their “feelings”.

To index the affective characteristics of images on the internet, an intuitive approach is to have a large number of people manually rate all the images and calculate descriptive statistics from the ratings. On the other hand, recent studies utilize color, texture, and composition information of images; also the application of content analysis, to achieve affective image classification [5, 6].

However, the features related to spatial-frequency domain that are proven to be useful for pattern recognition have not been explored yet. In addition, contributed by recent advances in methodology, the resolution in frequency analysis has been improved. Hence, this study achieves the affective image classification task by using the features related to spatial-frequency domain, and the Extended Classifier System for Function approximation task (XCSF)[7] (i.e. one of a latest version of Learning Classifier System (LCS)[8]). The dataset used for the classification task is collected from a human-subject experiment conducted in our laboratory.

The remainder of this paper is structured as follows: section II provides a review on theory and methodology revealed in the area of psychology and neuroscience; section III describes the human-subject experiment conducted; section IV presents the data analysis process; section V presents the XCSF; and lastly, obtained results and discussion are provided in section VI.

II. Literature Review

A. Emotion Theory

One of the difficulties in studying emotion is that how to define it. Although there is a tendency for researchers to intuitively define a set of discrete basic emotions (e.g., happy, surprising, sad, and angry [9]), recently the dimensional theory of emotion, in replacement of the traditional assumption of discrete emotions, has been proposed and demonstrated to be more suitable than the traditional manner of describing emotions in a number of studies [10, 11]. Dimensional theory defines emotions by a two dimensional

Po-Ming Lee is with the Institute of Computer Science and Engineering, National Chiao Tung University, Taiwan (R.O.C.). (E-mail: pmli@cs.nctu.edu.tw).

Tzu-Chien Hsiao is with the Institute of Biomedical Engineering; Department of Computer Science; and Biomedical Electronics Translational Research Center and Biomimetic Systems Research Center of National Chiao Tung University, Taiwan (R.O.C.). (Phone: +886-3-571-2121#54786; E-mail: labview@cs.nctu.edu.tw).

This work was fully supported by the Taiwan Ministry of Science and Technology under grant numbers NSC-102-2220-E-009-023 and NSC-102-2627-E-010-001. This work was also supported in part by the UST-UCSD International Center of Excellence in Advanced Bioengineering sponsored by the Taiwan Ministry of Science and Technology I-RiCE Program under grant number NSC-101-2911-I-009-101; and in part by "Aim for the Top University Plan" of the National Chiao Tung University and Ministry of Education, Taiwan, R.O.C.

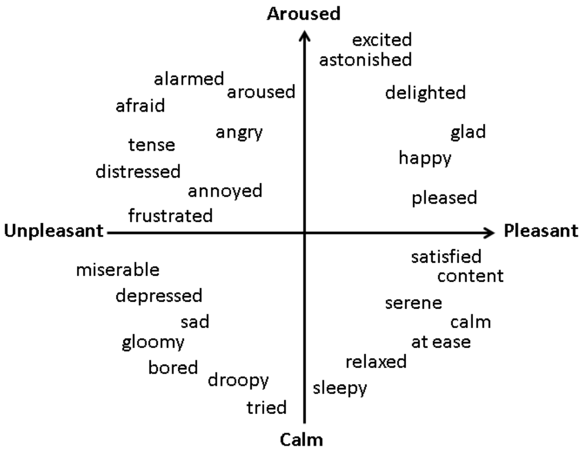

Fig. 1. Definition of emotions in a two-dimension affective space

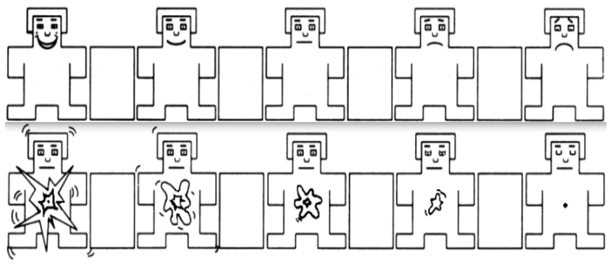

Fig. 2. The SAM used in this study, in which the upper row represents valence and the lower row represents arousal

affective space, of which the two dimensions are "valence" and "arousal". The valence represents that whether the emotion experienced is pleasant, whereas the arousal represents the amplitude of the emotion aroused. This study adopts the dimensional theory of emotion. The philosophy of the theory adopted is illustrated in Fig. 1. Furthermore, the dimensional theory of emotion also explains how human emotion is elicited and the roles of stimulus plays in emotion elicitation by relating the emotion theory to the motivational system of human. The motivational system guides human to behave in the tendency of "approach" or "avoidance" when presented with emotionally relevant stimulus (the "stimulus" can be an object, a scenario, or a type of circumstance)[12]. The reason of a stimulus to be emotionally relevant could be considered as a result of evolutionary process, that is, can be related to the need of survival. For example, the stimulus that stimulates positive emotions is found related to food and sex, whereas the stimulus that stimulates negative emotions was found related to danger and death. The umbrella term "emotionally relevant" can simply be understood as a capability to elicit certain emotions of a person (either positive or negative emotions) [13].

The dimensional theory of emotion has attracted substantial attention in the field of psychology since proposed, and is commonly adopted in latest studies [3, 10, 14]. On the other hand, brain scientists focused on biological proof. The pathway, the mechanisms of brain, autonomic nervous system, and organs, that are accounted for emotion responses have being revealed [15]. Other researchers reported the experimental results on the relationship between emotion and decision making [16], and also the relationship between emotion and memory [17].

B. Self-Assessment Manikin (SAM)

To assess the two dimensions of the affective space, the Self-Assessment Manikin (SAM), an affective rating system devised by Lang [18] was used to acquire the affective ratings. The SAM is a non-verbal pictorial assessment that is designed to assess valence and arousal directly by means of two sets of graphical manikins. The SAM has been extensively tested in conjunction with the International Affective Picture System (IAPS) and used in diverse theoretical studies and applications [3, 10, 14]. The SAM takes a very short time to complete (5 to 10 seconds). For using the SAM, there is little chance of confusion with terms as in verbal assessments. The SAM was also reported to be capable of indexing cross-cultural results [19] and the results obtained using a Semantic Differential scale (the verbal scale provided in [20]). The SAM that we used was identical to the 9-point rating scale version of SAM that was used in [21], in which the SAM ranges from a smiling, happy figure to a frowning, unhappy figure when representing the affective valence dimension. On the other hand, for the arousal dimension, the SAM ranges from an excited, wide-eyed figure to a relaxed, sleepy figure. The subjects were presented with 5 figures and 4 buttons (i.e. the interval between the figures). The subject could select the 5 figures comprising each scale, or the 4 buttons, which results in a 9-point rating scale for each dimension. The scores 1, 3, 5, 7, 9 were associated with the 1st, 2nd, 3rd, 4th, and 5th figure respectively; whereas the scores 2, 4, 6, and 8 were associated with the buttons between the 1st and 2nd, 2nd and 3rd, 3rd and 4th, 4th and 5th figure respectively. Ratings are scored such that 8 represents negative valence and low arousal; and 0 represents positive valence and high arousal.

C. Affective Image Classification

To predict emotions of subject elicit by image, Mikels et al. (2005) categorized images in IAPS into different categories, and identified the images that are especially excellent in eliciting emotions of subjects [22]. Wu et al. accomplished the affective image classification by using Semantic Differential features [23]. Subsequently, various features were reported for being beneficial for performing affective image classification task [24, 25]. On the other hand, Machajdik et al. and Zhang et al. pointed out the disadvantages of applying content analysis on affective classification task due to individual difference [5, 26]. In addition, the obtained accuracy rates from previous studies were relatively low. Furthermore, the use of definition in "discrete" emotions also caused the experiment hard to reproduce in countries other than the United States. Hence, the conclusion is remaining inconclusive in our point of view.

III. HUMAN SUBJECT EXPERIMENT

A. Experimental Procedure

To build an intelligent system that could predict the emotions of subjects elicit by image, a human subject experiment was conducted. The entire experiment conducted in this study complies the IAPS protocol of emotion inducement described in [27] to guarantee the effectiveness

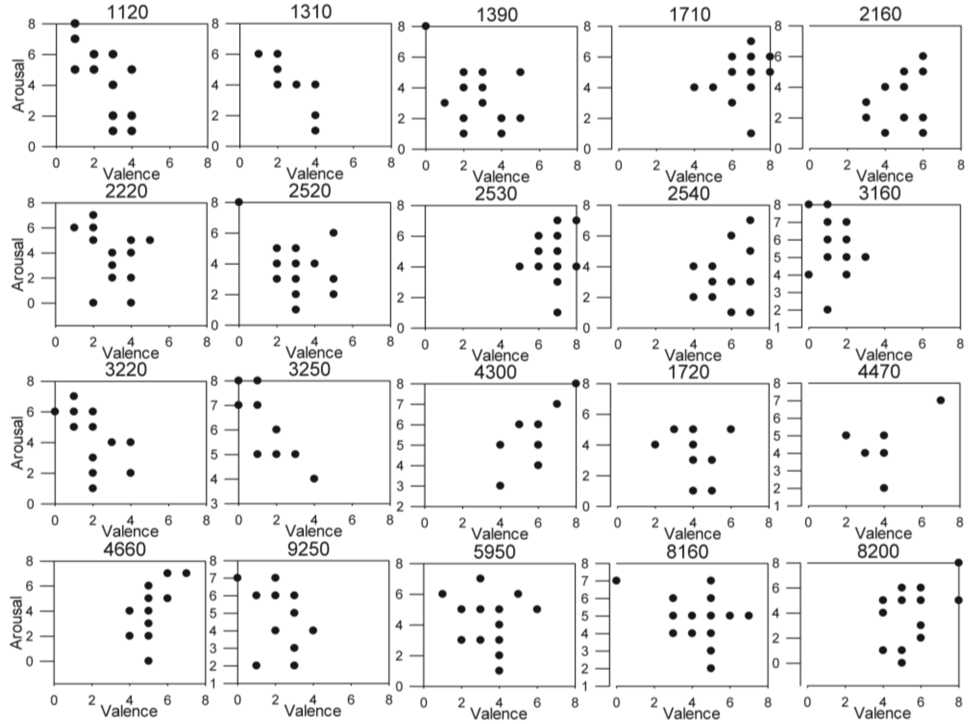

Fig. 3. The distribution of the emotion induced by each image

of the emotion induction procedure, and the clarity of the experimental design for reproduction. During the experiment, the subjects were requested to look at a screen which sequentially presents images and to correspondingly rate these images presented, by using computer-based SAM (through the use of mouse). The duration of the experiment was 10 minutes for each subject. Each trial (i.e. presentation of an image) started by presenting an image and displayed it for 6 seconds, then presented the SAM on the screen for the subject to manually rate the affective characteristics (i.e. self-report the induced emotion) of the presented image. The SAM was followed by a 15 seconds delay to ensure the emotional status of subject return to baseline before the start of next trial and a reasonable length to keep the subjects involved in the experiment.

B. *Affective Images*

This study utilizes 20 images selected from IAPS [28] database developed and distributed by the NIMH Center for Emotion and Attention (CSEA) at the University of Florida. The IAPS is developed to provide a set of normative emotional stimuli for experimental investigations of emotion and attention and can be easily obtained through e-mail application. The IAPS database contains various affective images proved (by using descriptive statistics) to be capable of inducing diverse emotions in the affective space [11]. The images selected from the IAPS database comply the IAPS image set selection protocol described in [28], which includes the constraint about the number of images used in a single experiment and the distribution of the emotions induced by the images selected. The image ids of the used images are as follows: 1120, 1310, 1390, 1710, 1720, 2160, 2220, 2520, 2530, 2540, 3160, 3220, 3250, 4300, 4460, 4470, 4660, 4750, 5950, 8160, 8200, and 9250. These images can be found in the IAPS database [28] using the ids listed above. The order of the image presentation was randomized to eliminate the effects due to the presentation sequence.

C. *Environment Setting*

The images were presented using a general PC with 32-inch (81.28 centimeters) monitor. The subjects were sat in a comfortable bed at a distance of approximately 1.5 meters away from the monitor in an EMI shielding room (Acoustic Inc. US) in which eliminates most of noise interferences and electrical noises. The CO_2 concentration of the environment was monitored during the entire experiment to guarantee reasonable CO_2 concentration (500 ppm ~ 1,300 ppm) to keep subjects sustain their attention during the experiment.

D. *Subject Selection*

There were 16 university students participated in the study (15 subjects is the typical sample size required in the field of affective image classification studies [23, 26]), ranging in age between 20 and 28 (mean = 23.44, standard deviation = 2.19; 10 men, 6 women). All subjects reported they were healthy, with no history of brain injury, cardiovascular problems, had normal or corrected-to-normal vision, and normal range of finger movement. The experiment and the manner of using data obtained from human subject were approved (Protocol No: 100-014-E) by the Institution Review Board (IRB) of the National Taiwan University Hospital Hsin-chu Branch.

IV. EXTRACTION OF SPATIAL-FREQUENCY FEATURES

A. *Collected Data Set*

The collected dataset contains 20 images (1024x768 JPEG) used in the experiment and the affective ratings of these images, that were rated by 16 subjects through the SAM. The experiment totally acquired 318 rows of raw data (images, and the affective ratings of the images, 20 rows for each subject). Two rows were excluded due to machine mal functioning. Figure 3 presents the distribution of the ratings of all images. Figure 3 shows that most the subjects were aroused with either unpleasant feelings or pleasant feelings by the displayed images. No obvious skew was found in the distribution of valence and arousal (histogram was examined but not shown).

B. Preprocessing Using Hilbert-Huang Transform (HHT)

Spatial-frequency analysis on image is one of the well-known techniques used in the field of image processing and computer vision [29, 30]. The information in frequency domain was found abundant by physiologists [31]. It was found that various spatial-frequencies can lead to distinct characteristics of visual stimulations. Moreover, the orientation of visual stimulation can cause different efficacies in stimulations of cortical receptors [32, 33]. Traditionally, Fast Fourier Transform (FFT) is used to transform an image into frequency domain. However, due to the assumption of that series of target data should be at least piecewise stationary, the FFT-based techniques (e.g., spectrogram), is not suitable for modeling local phenomena or when higher resolution is required. Hence, recently Hilbert-Huang Transform (HHT) was proposed to obtain higher frequency resolution toward Instantaneous Frequency (IF) [34]. Later, the use of such concept in spatial-frequency analysis was also reported [35]. The HHT is a two-phase transformation, which firstly apply an Empirical Mode Decomposition (EMD) on the target data series to extract Intrinsic Mode Functions (IMFs). Secondly, Hilbert Transform (HT) is applied to each IMF to obtain required frequency domain information (i.e., IF). The EMD is a shifting process that can be used to extract IMFs from a data series X(s). The IMF is defined as a monocomponent by satisfying the criterias as following: 1) has the number of zero crossings and extrema one difference at most, 2) symmetric with respect to the local mean, and 3) the X(s) should has at least two extrema. After the procedure of EMD, n IMFs, namely, IMF1, IMF2, IMF3, …, IMFn, and the residuals(r_n), denoted as $X(s) = \sum_{j=1}^{n} C_j + r_n$ provided in Formula (1), are extracted from X(s). The residuals (r_n) is a data series which is the remainder series of target data series after the EMD shifting process removes all the IMFs from the original target data series.

The procedure of EMD, different from the Fourier and Wavelet Decomposition, is fully data-driven. By being adaptive and unsupervised, the EMD improves the efficiency of signal decomposition and can be applied to the non-linear and non-stationary signal (details on the procedure of the EMD please refer to [34]). After the EMD, the HT is then applied to each IMF

$$Y_j(s) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{C_j(\tau)}{s-\tau} d\tau \qquad (1)$$

Each IMFj can be represented by the conjugate pair of $Y_j(s)$ and $C_j(s)$, hence can be represented by an analytical signal Z(s) = $C_j(s) + iY_j(s) = a(s)e^{i\theta(s)}$, in which the amplitude $a_j(s) = \sqrt{C_j(s)^2 + Y_j(s)^2}$ and phase $\theta_j(s) = \arctan(Y_j(s)/C_j(s))$. Based on the definition stated above, the IF_j can be derived by applying a derivative on $\theta_j(s)$ (i.e. $\omega_j = d\theta_j(s)/ds$). Then, an analytical representation of X(s), can be derived $X(s) = \sum_{j=1}^{n} a_j(s)[i \int \omega_j(s)\, ds]$.

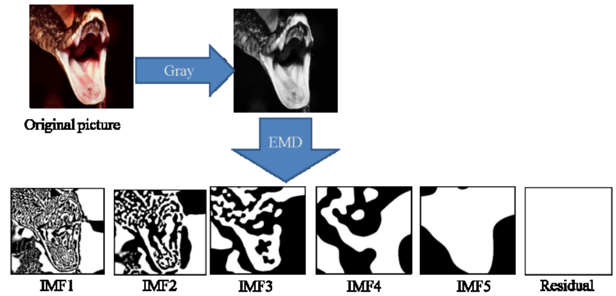

Fig. 4. The illustration of the data processing in our study, the application of 2D EMD on IAPS picture 1120.

Originally, EMD was proposed to decompose one-dimensional data. To construct two-dimensional HHT, the concept of EMD was extended to 2D in our study based on the concept listed as follows: 1) identify the extrema (maxima and minima) of the image by sliding a 3-by-3 grid; 2) generate two smooth 2D surfaces to fit the found maxima and minima; 3) compute the local mean by averaging two surfaces; and 4) the equation of applying 2D EMD then can be rewrite from Formula (1), to $f(x,y) = \sum_{j=1}^{n} C_j(x,y) + r_n(x,y)$.

The data processing done in our study is illustrated in Fig. 4. The original image (1024x768 resolution) was first down-sampled to 128x128 resolution. The color setting was changed from RGB color into gray color. Second, the 2D EMD is applied to the image. For extracting IFs from IMFs, this study applies the concept of partial HT by applying 1D HT to each orientation (i.e. each row and each column) and unit, in order to extract spatial-frequency features that account for different orientations of visual stimulations [32]. The IF analysis method used in our study was inspired by the work in [36] which provides a show case on estimating the changes of IF data series. This study mainly adopts three indexes as follows: 1) $F\_Q\_IMF_j$ represents frequency value in the 1st quarter of the histogram area of IF_x; 2) $A\_I\_IMF_j$ represents the ratio between the 1st and the 2nd halves of the histogram area of IF_j; 3) $M\_I\_IMF_j$ represents the ratio between the maxima found in the 1st and 2nd halves of the histogram area of IF_j. This study applies totally 12 features listed below as follows:

1. The vertical side (the direction of applying 1D HT) $F\_Q\_IMF_1$, the horizontal side (the direction of applying 1D HT) $F\_Q\_IMF_1$, the vertical side $A\_I\_IMF_1$, the horizontal side $A\_I\_IMF_1$, the vertical side $M\_I\_IMF_1$, the horizontal side $M\_I\_IMF_1$;
2. the vertical side $F\_Q\_IMF_2$, the horizontal side $F\_Q\_IMF_2$, the vertical side $A\_I\_IMF_2$, the horizontal side $A\_I\_IMF_2$, the vertical side $M\_I\_IMF_2$, the horizontal side $M\_I\_IMF_2$.

We totally acquired 318 rows of the feature vector from the collected data set. The method that we use to build the prediction model is introduced in the following section.

V. LEARNING CLASSIFIER SYSTEM

The study applies XCSF to the affective image classification task to cope with any possible non-linear characteristics contained in the target dataset.

A. XCSF

The XCSF is an extension of the Extended Classifier System (XCS), a machine learning system based on Michigan-Style Classifier Systems (CSs). Since the 1990s, the XCS has attracted considerable attention in the field of LCSs because of its theoretical advances and its applicability in practice [37, 38]. The XCS is a rule-based online learning algorithm which can extract knowledge from a previous unknown dataset in an iterative manner. The XCS can also be regarded as a system that manages a set of classifiers that are represented in the traditional production system form of "IF state THEN action". By integrating the Genetic Algorithm (GA) component (also named rule discovery component in the XCS), the set of classifiers evolves occasionally and searches for a set of classifiers that yields the maximal generality and accuracy. The original XCS was designed to perform on datasets with discrete inputs and a discrete output.

B. XCSF

In 2002, XCSF, as a version of XCS used for function approximation was proposed [39]. The XCSF allows both real value inputs and real value outputs. In addition, the version of XCSF implemented in [40] allows multiple outputs The input accepts real value by using rotating hyperrectangle and rotating hyperellipsoid for condition representation [7, 41]. On the other hand, instead of selecting a discrete value as output according to fitness-weighted prediction value, the classifiers in the XCSF directly map the desire output using the prediction value produced by the linear approximation (i.e. $h(\vec{x}) = \vec{\omega}\,\vec{x}$ in which $\vec{x}$ represents the input vector and $\vec{\omega}$ represents weight vector). Each classifier in the XCSF updates its weight vector using Recursive Least Squares (RLS) method [7]. For performing the RLS, each classifier manage by XCSF updates its weight vector using

$$\vec{\omega} \leftarrow \vec{\omega} + \vec{k}\left[y_t - \left(\overrightarrow{x^*} - \overrightarrow{m^*}\right)^T \vec{\omega}\right] \tag{5}$$

where y_t represents target output, and $\vec{k}$ represents the gain vector computed by

$$\vec{k} = \frac{V^T\left(\overrightarrow{x^*} - \overrightarrow{m^*}\right)}{\lambda + \left(\overrightarrow{x^*} - \overrightarrow{m^*}\right)^T V^T \left(\overrightarrow{x^*} - \overrightarrow{m^*}\right)}\,. \tag{6}$$

The λ (usually $0 \leq \lambda \leq 1$) used in Formula 6 and 7 represents the forget rate of RLS. The lower the value of λ is the higher the forget rate. The value of λ is set to 1.0 for having an infinite memory (mostly used in time invariant problems). The matrix V hold by each classifier updates recursively using

$$V^T = \lambda^{-1}\left[I - \vec{k}\left(\overrightarrow{x^*} - \overrightarrow{m^*}\right)^T\right] V^T\,. \tag{7}$$

The fitness value used for the GA in the XCSF is the relative classifier accuracy calculates from system error [42]. For further detail, sufficient information about XCS can be found in Butz's algorithmic description of XCS [42], and also the recent advances in XCSF [7, 39, 41]. To summarize, the XCSF can be understood as a manager which manages a set of classifiers. Each of the classifiers maps from a subspace in the feature space to the landscape-function output using a linear-fitting method. The XCSF used in this study is modified from the Java implementation version of XCSF contributed by Stalph and Butz (2009) [40]. For parameters setting, $\alpha = 1.0$; $\beta = 0.1$; $\delta = 0.1$; $\lambda = 1.0$; $\theta_{GA} = 50$; $\varepsilon_0 = 0.5$; $\delta_{rls} = 1000$; $\theta_{del} = 20$; $\chi = 1.0$; $\mu = 1.0$; $\theta_{sub} = 20$; the GA subsumption was turned on. Although the maximal population size N was set to 6,400~10,000 to maximize the performance of XCSF, the number of classifiers quickly converged to 5,400 during the model training. To examine the performance of the system, ε_0 was set to various values. However, it appears relatively small effect on the learning performance in regard to the learning speed and system error. During the model training, the XCSF was sequentially presented with 20,000 instances randomly selected from the training dataset.

C. Affective Image Classification

Models were built to predict the emotion ratings rated by subjects in terms of valence and arousal through SAM. The prediction of valence and arousal can be real number herein according to the definition of valence and arousal in the dimensional theory of emotion [10]. Besides the XCSF, this study also applies several well-known machine-learning techniques for comparison purpose. Zero-R is a majority voting learning scheme that predicts the majority class in any data set. In a classification task, the Zero-R classifies an instance into the majority class, whereas in a prediction task, the Zero-R predicts the mean value of all the instances. Thus, the. The performance of the Zero-R can be considered as a baseline performance of the classification class, which should be beaten by any algorithm that learns decision boundaries from the data set without over-fitting. One-layer method such as Linear Regression (LR) [43] and multi-layered method with transfer function such as Radial-Basis-Function (RBF) Network [44] were used in our study. Leave-One-Out-Cross-Validation (LOOCV) which leaves one sample out at each time as a testing set and the remaining samples as a training set was used for model building.

VI. RESULTS AND DISCUSSION

The performance evaluation based on Mean Absolute Error (MAE) and the standard deviation (represents by SD) of the MAEs achieved by the methods used is provided in Table 1. The MAE was calculated as following:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|V_i - VP_i| \text{ or } \frac{1}{N}\sum_{i=1}^{N}|A_i - AP_i|$$

TABLE I
THE PERFORMANCE ACHIEVED BY BENCHMARK CLASSIFIERS AND XCSF

Prediction Results		Affective Dimension	
		Valence	Arousal
Method	Statistics		
ZeroR	MAE	1.617	1.491
	SD of MAE	1.110	1.065
LinearReg	MAE	1.453	1.427
	SD of MAE	1.076	1.083
RBFNet	MAE	0.950	1.471
	SD of MAE	0.747	1.021
XCSF	MAE	0.950	1.461
	SD of MAE	0.755	1.011

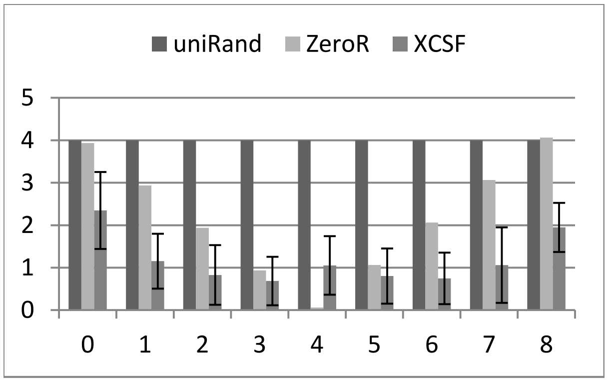

Fig. 5. The MAE of XCSF on each valence value. The standard deviation of MAE made by XCSF is nearly 0.58~0.91. Based on the SAM ratings, the maximal MAE is 8 and minimal MAE is 0.

in which N represents sample size; V_i and A_i represents the values of the valence and arousal corresponds to the i-th sample; and VP_i and AP_i represents the system prediction on the values of valence and arousal corresponds to the i-th sample. The MAEs are used here to evaluate the performance of a built model in predicting valence and arousal. The ZeroR represents the Zero-R classifier, LinearReg represents the LR model, and RBFNet represents the RBF network. The number of nodes (clusters) of the RBF network was set to 200 based on the result of the examination on the performance changes caused by the number of nodes.

The MAE of a prediction model which predicts at random on the value of valence and arousal is 4.0. Hence, the MAE 1.453±1.076 achieved by the LR seems to be fair. The MAE achieved by the RBF network is 0.949±0.747, which further shows a reduction of the error by 35%. This result indicates the existence of the non-linearity characteristic of the dataset collected. The MAE achieved by the XCSF was 0.950±0.755. The equivalence in the performance of RBF network and XCSF indicates the capability of XCSF on mapping non-linear functions. The mechanism of the XCSF in model building by managing a set of linear classifiers seems to be comparable to the multi-layered based method with non-linear transfer function. To further examine the performance of the XCSF, the MAEs of the XCSF on each valence and arousal value are also provided in Fig. 5 and Fig. 6. To compare the MAE achieved by the XCSF, the MAEs that achieved by uniRand, a classifier that makes predictions in a uniformly random manner are also included in these figures.

The performance of the XCSF in predicting each valence value is illustrated in Fig. 5 in which x-axis represents the valence value, y-axis represents the MAE value. The MAEs that the classifiers achieved on each valence are represented by three bars. The right most bar represents the MAE achieved by the XCSF. The MAEs of uniRand and ZeroR are represented by the first and the second bar.

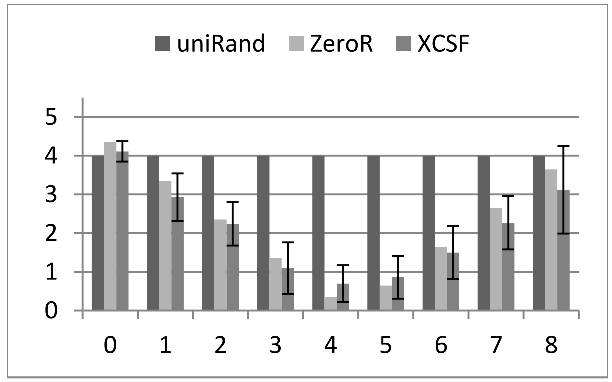

Fig. 6. The MAE of XCSF on each arousal value. The standard deviation of MAE made by XCSF is nearly 0.26~0.69. Based on the SAM ratings, the maximal MAE is 8 and minimal MAE is 0.

The MAE achieved by the XCSF is smaller than the MAE achieved by uniRand and ZeroR at most ratings (i.e. the value of valence and arousal). The MAE of XCSF is only larger than the MAE of ZeroR at the ratings near the mean values. A skew on the value of MAEs is observed for the largest and lowest valence values (i.e. 0~1 and 8). This is possibly due to the sample size of these ratings, since the numbers of samples of valence equals to 0, 1, and 8 are smaller. This finding suggests that insufficient sample size of a class (e.g., valence = 8) may lead to bad performance of XCSF in predicting the corresponding output value. However, the MAE of XCSF at valence 4.0 was not the lowest, which indicates that the larger sample size only guarantees the efficacy of XCSF in function approximation instead of eliminating all exist errors. In our observation, approximately 30 samples (which is nearly 10% of the number in our collected data set) is sufficient for the XCSF to build a model to predict a valence value in our collected dataset. On the other hand, this phenomenon happened in valence 0 and 8 could also be explained by a psychological approach. That is, some of the subjects reported that they tended to rate the values in the middle of the scale rather than those values that represent extreme emotional experiences. This may cause the non-linear characteristics of the distances between the levels of valence and arousal. Further clarification is required for this issue; an appropriate transformation may be applied to the data to improve the result.

Similar results can be found in Fig. 6, in which for the ZeroR, the prediction was set to 3.937. The MAE achieved by the XCSF in each level of arousal substantially outperformed the MAE of uniRand and ZeroR. However, the Fig. 6 shows the increase of the MAEs achieved by the XCSF at each level of arousal. This could be explained by that most subjects reported that during the experiment, they confused the definition of "being aroused" with "the tendency of valence". The reaction of the subjects is possibly caused by the cultural difference, but similar results were not highlighted previously in the research community that applies the IAPS and SAM.

TABLE II
THE RESULTS OF ROC CURVE PRODUCED BY XCSF AND RBF NETWORK

Prediction Target		V < 4	V > 4	A < 4	A > 4
Method	Estimation				
XCSF	AUC	0.913	0.914	0.594	0.573
	Accuracy	84.30%	86.80%	53.10%	56.00%
RBFnet	AUC	0.682	0.673	0.495	0.505
	Accuracy	70.50%	65.20%	50.10%	63.30%

V: Valence, A: Arousal.

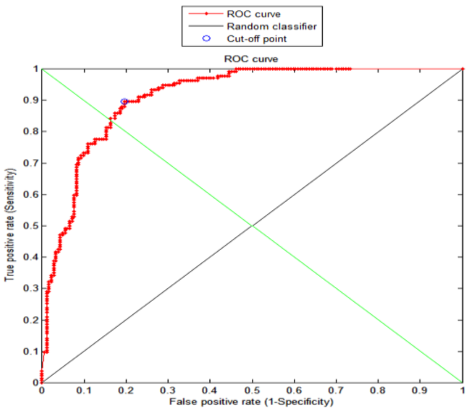

Fig. 7. The ROC curve made by XCSF on prediction of the event when valence of the image is rated smaller than 4 (neutral).

Gender information was not used for model building, but the within-subject analysis was conducted. However, we found that without the applying content analysis to the image, the effect of individual difference is relatively small.

To further examine the performance of the XCSF in this task, the ROC curve of the XCSF in predicting the value of valence of an image being rated smaller or larger than 4 (i.e. 4 is the value of valence that represents a neutral state) are provided in Fig. 7 and Fig. 8. The result shows that the AUC achieved by the XCSF is significantly (p-value < .000) larger than the random classifier (the Area Under Curve (AUC) of predicting valence < 4 is 0.913, in which the AUC of predicting valence > 4 is 0.914). However, the AUC achieved by the XCSF in predicting the value of arousal being smaller or larger than 4 is relative small (p-value < .005). The AUC of predicting arousal < 4 is 0.594 and the AUC of predicting arousal > 4 is 0.573. The accuracy rates achieved by the XCSF on the cut-off point are: 84.3% (for predicting valence < 4), 86.8% (valence > 4), and 53.1% (arousal < 4), and 56.0% (arousal > 4). In addition, the ROC curve of RBF network was also examined because the MAE made by RBF network was favorable in comparison with the MAE made by XCSF. The results are provided in Table 2.

To further identify the extracted knowledge, the prediction models built by the LR are provided in Formula 2 and 3.

$$\text{Valence} = 3.2893 * \text{F\_Q\_IMF1\_col} + 0.3651 * \text{F\_Q\_IMF1\_row} + 2.6606 * \text{A\_I\_IMF1\_row} + 0.4394 * \text{F\_Q\_IMF2\_row} - 0.2629 * \text{A\_I\_IMF2\_col} + 2.335 * \text{A\_I\_IMF2\_row} - 1.0522 \quad (2)$$

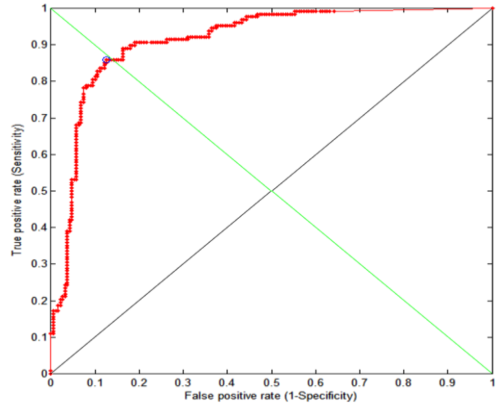

Fig. 8. The ROC curve made by XCSF on prediction of the event when valence of the image is rated larger than 4 (neutral)

$$\text{Arousal} = -2.4297 * \text{F\_Q\_IMF1\_col} - 0.9253 * \text{F\_Q\_IMF1\_row} + 0.1896 * \text{A\_I\_IMF1\_col} - 1.2495 * \text{A\_I\_IMF1\_row} + 0.4578 * \text{M\_I\_IMF1\_col} - 0.3721 * \text{A\_I\_IMF2\_col} + 0.1047 * \text{M\_I\_IMF2\_col} + 7.0513 \quad (3)$$

For building models by the LR, Akaike criterion was used for model selection and the M5's method was used for attribute selection, in which all the co-linear attributes were excluded. The equations show that F_Q_IMF1_col, F_Q_IMF1_row, and A_I_IMF2_row were the main factors that affect the affective ratings during the experiment. The F_Q_IMF1_col, F_Q_IMF1_row, and A_I_IMF2_row show positive relationship to the rating of valence. These results indicate that the stimulations from horizontal side are more effective than the stimulations from vertical side. The horizontal side of the image may contain abundant information. Conversely, the affective characteristics of an image in regard to making people feel aroused, is negative correlated with F_Q_IMF1_col and A_I_IMF1_row. In addition, the offset of the equation 3 is +7.0513. These results indicate that the effects of activation in motivational system due to a visual stimulus are influenced by the asymmetric of cortical receptors responsible for distinctive directions of the spatial-frequency visual stimulations.

VII. CONCLUSION

On indexing affective characteristics of the images on the internet, the related applications which may use these images to improve the quality of life (e.g., to ease people, or to excite people), become possible. In the near future, the results achieved by this study should be replicated using images obtain from sources other than IAPS database, to validate the experimental results in regard to generality. Further examination on the mechanisms and pathway between the affective information contained in spatial-frequency domain and the cortical receptors in human eye, is also suggested.

REFERENCES

[1] W. J. Kuo, T. Sjostrom, Y. P. Chen, Y. H. Wang, and C. Y. Huang, "Intuition and Deliberation: Two Systems for Strategizing in the Brain," *Science*, vol. 324, pp. 519-522, Apr 2009.

[2] R. M. M. I. Chowdhury, G. D. Olsen, and J. W. Pracejus, "Affective Responses to Images In Print Advertising: Affect Integration in a

Simultaneous Presentation Context," *Journal of Advertising,* vol. 37, pp. 7-18, 2008.
[3] C. Chang, "The Impacts of Emotion Elicited By Print Political Advertising on Candidate Evaluation," *Media Psychology,* vol. 3, pp. 91-118, 2001.
[4] K. Kyung-Sun, "Effects of emotion control and task on Web searching behavior," *Information Processing & Management,* vol. 44, pp. 373-385, 2008.
[5] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," presented at the Proceedings of the international conference on Multimedia, Firenze, Italy, 2010.
[6] P.-M. Lee, Y. Teng, and T.-C. Hsiao, "XCSF for prediction on emotion induced by image based on dimensional theory of emotion," in *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion*, 2012, pp. 375-382.
[7] M. V. Butz, P. L. Lanzi, and S. W. Wilson, "Function Approximation With XCS: Hyperellipsoidal Conditions, Recursive Least Squares, and Compaction," *Evolutionary Computation, IEEE Transactions on,* vol. 12, pp. 355-376, 2008.
[8] J. H. Holland and J. S. Reitman, "Cognitive systems based on adaptive algorithms," *SIGART Bull.,* pp. 49-49, 1977.
[9] A. Ortony and T. Turner, "What's basic about basic emotions," *Psychological review,* 1990.
[10] [M. M. Bradley, "Emotional memory: a dimensional analysis," in *Emotions: Essays on emotion theory*, S. H. M. v. Goozen, N. E. v. d. Poll, and J. A. Sergeant, Eds., ed Hillsdale, NJ: Lawrence Erlbaum, 1994, pp. 97-134.
[11] M. M. Bradley and P. J. Lang, "Emotion and motivation," in *Handbook of Psychophysiology*, J. T. Cacioppo, L. G. Tassinary, and G. Berntson, Eds., 3 ed New York, NY: Cambridge University Press, 2007, pp. 581-607.
[12] P. J. Lang, "The motivational organization of emotion: Affect-reflex connections," in *Emotions: Essays on emotion theory*, ed Hillsdale, NJ: Lawrence Erlbaum, 1994, pp. 61-93.
[13] P. J. Lang, "The Emotion Probe - Studies of Motivation and Attention," *American Psychologist,* vol. 50, pp. 372-385, May 1995.
[14] P. D. Bolls, A. Lang, and R. F. Potter, "The Effects of Message Valence and Listener Arousal on Attention, Memory, and Facial Muscular Responses to Radio Advertisements," *Communication Research,* vol. 28, pp. 627-651, 2001.
[15] D. Antonio R, "Emotion in the perspective of an integrated nervous system," *Brain Research Reviews,* vol. 26, pp. 83-86, 1998.
[16] A. Bechara, "The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage," *Brain and cognition,* vol. 55, pp. 30-40, 2004.
[17] K. S. LaBar and R. Cabeza, "Cognitive neuroscience of emotional memory," *Nat Rev Neurosci,* vol. 7, pp. 54-64, 2006.
[18] P. J. Lang, "Behavioral treatment and bio-behavioral assessment: Computer applications," in *Technology in Mental Health Care Delivery Systems*, J. Sidowski, J. Johnson, and T. Williams, Eds., ed Norwood, NJ: Ablex Pub. Corp., 1980, pp. 119-137.
[19] J. D. Morris, "Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response," *Journal of advertising research,* vol. 35, pp. 63-68, 1995.
[20] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*: the MIT Press, 1974.
[21] M. M. Bradley and P. J. Lang, "The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual," *University of Florida, Gainesville, FL, Tech. Rep. B-3,* 2007.
[22] J. Mikels, B. Fredrickson, G. Larkin, C. Lindberg, S. Maglio, and P. Reuter-Lorenz, "Emotional category data on images from the international affective picture system," *Behavior Research Methods,* vol. 37, pp. 626-630, 2005.
[23] Q. Wu, C. Zhou, and C. Wang, "Content-Based Affective Image Classification and Retrieval Using Support Vector Machines," in *Affective Computing and Intelligent Interaction*. vol. 3784, J. Tao, T. Tan, and R. Picard, Eds., ed: Springer Berlin / Heidelberg, 2005, pp. 239-247.
[24] D. Joshi, R. Datta, E. Fedorovskaya, L. Quang-Tuan, J. Z. Wang, L. Jia, and L. Jiebo, "Aesthetics and Emotions in Images," *Signal Processing Magazine, IEEE,* vol. 28, pp. 94-115, 2011.
[25] N. Liu, E. Dellandréa, B. Tellez, and L. Chen, "Associating Textual Features with Visual Ones to Improve Affective Image Classification," in *Affective Computing and Intelligent Interaction*. vol. 6974, S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds., ed: Springer Berlin / Heidelberg, 2011, pp. 195-204.
[26] H. Zhang, E. Augilius, T. Honkela, J. Laaksonen, H. Gamper, and H. Alene, "Analyzing Emotional Semantics of Abstract Art Using Low-Level Image Features," in *Advances in Intelligent Data Analysis X*. vol. 7014, J. Gama, E. Bradley, and J. Hollmén, Eds., ed: Springer Berlin / Heidelberg, 2011, pp. 413-423.
[27] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," University of Florida, Gainesville, FL2008.
[28] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International Affective Picture System (IAPS)," The Center for Research in Psychophysiology, University of Florida, Gainesville, FL1999.
[29] S. Li and B. Yang, "Multifocus image fusion using region segmentation and spatial frequency," *Image and Vision Computing,* vol. 26, pp. 971-979, 2008.
[30] H. Leonard, D. Annaz, A. Karmiloff-Smith, and M. Johnson, "Brief Report: Developing Spatial Frequency Biases for Face Recognition in Autism and Williams Syndrome," *Journal of Autism and Developmental Disorders,* vol. 41, pp. 968-973, 2011.
[31] D. John G, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research,* vol. 20, pp. 847-856, 1980.
[32] M. A. Webster and R. L. De Valois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," *J. Opt. Soc. Am. A,* vol. 2, pp. 1124-1132, 1985.
[33] K. Kobayashi, H. Fushiki, M. Asai, and Y. Watanabe, "Head and body sway in response to vertical visual stimulation," *Acta Oto-laryngologica,* vol. 125, pp. 858-862, 2005.
[34] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. Lond. A,* vol. 454, pp. 903-995, 1998.
[35] P. C. Tay, "AM-FM Image Analysis Using the Hilbert Huang Transform," in *Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on*, 2008, pp. 13-16.
[36] P. Caseiro, R. Fonseca-Pinto, and A. Andrade, "Screening of obstructive sleep apnea using Hilbert–Huang decomposition of oronasal airway pressure recordings," *Medical Engineering & Physics,* vol. 32, pp. 561-568, 2010.
[37] A. Quirin, J. Korczak, M. V. Butz, and D. E. Goldberg, "Analysis and evaluation of learning classifier systems applied to hyperspectral image classification," in *Intelligent Systems Design and Applications, 2005. ISDA '05. Proceedings. 5th International Conference on*, 2005, pp. 280-285.
[38] B. Ravichandran, A. Gandhe, and R. E. Smith, "XCS for robust automatic target recognition," presented at the Proceedings of the 2005 conference on Genetic and evolutionary computation, Washington DC, USA, 2005.
[39] S. W. Wilson, "Classifiers that approximate functions," *Natural Computing,* vol. 1, pp. 211-234, 2002.
[40] P. O. Stalph and M. V. Butz, "Documentation of JavaXCSF," Retrieved from University of Wurzburg, Cognitive Bodyspaces: Learning and Behavior website2009.
[41] S. W. Wilson, "Get Real! XCS with Continuous-Valued Inputs," *Learning Classifier Systems,* vol. 1813, pp. 209-219, 2000.
[42] M. Butz and S. Wilson, "An Algorithmic Description of XCS," in *Advances in Learning Classifier Systems*. vol. 1996, P. Luca Lanzi, W. Stolzmann, and S. Wilson, Eds., ed: Springer Berlin / Heidelberg, 2001, pp. 267-274.
[43] A. Kaw and E. Kalu, *Numerical Methods with Applications*: autarkaw, 2010.
[44] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.,* vol. 11, pp. 10-18, 2009.