Data-driven Affective Enhancement of Images

Abhishek Bhattacharya[*1], Christine Lisetti[2], Zhenyu Yang[3]

[1, 2, 3] School of Computing and Information Science, Florida International University
Miami, FL 33199, USA

[*1]abhat002@fiu.edu; [2]lisetti@cis.fiu.edu; [3]yangz@cis.fiu.edu

***Abstract*-With the current trend of digital camera revolution, it is increasingly important and desirable to generate high-impact photography that can have a lasting influence on the viewer. Recently, there is an increasing interest of computational/data-driven methodologies for more sophisticated controls such as aesthetics and emotions. Professionals usually rely on experience and intuition to choose appropriate color compositions which makes this process tedious and labor-intensive. Our goal in this paper is to learn implicit adjustment rules associated with the emotional properties of images from a set of training examples. Given a pair of image before and after adjustments, we discover the underlying mathematical relationships optimally connecting the color-contrast properties between them. As a first step we collect ground-truth data and construct a training database consisting of sample images and their different enhanced versions, and ask a set of human participants to mark their preferences. We employ machine learning techniques to understand the implicit relationship between an original image and its adjusted version, and derive statistical solutions to develop an enhancement function in a mathematical form. We conduct user studies to evaluate the effectiveness of our approach.**

Keywords- Image Enhancement; Affective Computing; Machine Learning

I. INTRODUCTION

With the increasing availability of digital cameras from mobile phones, hand-held devices, to high-end SLR and professional cameras, we can observe a revolution in the digital photography world. Digital cameras are getting increasingly familiar with decreasing cost and increasing functionalities. With this current trend, it is important and desirable to generate high-impact photography that can have a lasting impression on the viewer. This still remains a challenge for the typical user since they are not familiar with the sophisticated photographic techniques generally employed by the professionals. These specialized techniques that are required to produce impact photography are termed as correction, adjustment, enhancement; each of which may have distinctive characteristics in various context, but the general objective is to improve the image.

In a camera processing pipeline, the enhancement functionalities can be applied either real-time or post-capture. Real-time processing is performed at the time of capturing images (manual or automatic controls) with a wide variety of hardware and software techniques involving signal processing and filter functions. These types of functionalities (e.g. exposure control, image stabilization) are available with advanced devices such as high-end professional cameras but are generally not present in mobile phone cameras. The other type of processing is post-capture, which is our focus in this paper. Post-capture enhancement techniques are generally supported by automatic or manual editing controls in software systems such as Adobe Photoshop, GIMP, among others. Many post-processing techniques that improve the quality of images have been proposed where different correction techniques are used to improve the various properties such as contrast, brightness, color [1], [2], [3].

Recently, there is an increasing interest of computational/data-driven methodologies for more sophisticated controls such as aesthetics and emotions [4]. These techniques are attractive due to their computational nature which allows to automate the enhancement processing function through software control. Psychological studies confirm the strong influence of image and its properties such as color, contrast, textures on human perception and emotion [5], [6], [7]. Affect-based techniques are thus used to explore a variety of multimedia analysis applications such as retrieval [8], [9], presentation [10], summarization [11], browsing [12], deriving affective film semantics [13], visualization [14]. In this paper, we would like to address the following problem: *How do we modify the characteristics of an image (such as color, or contrast) to enhance its emotional impact on its viewers?*

The problem is challenging due to the ``semantic gap'' that exists between the image RGB channels and higher level emotional concepts such as *fear*, *happiness*, *anger*. Moreover, emotion is inherently subjective in nature which varies under different which makes it an exciting problem to explore. We know that an artist uses different color and their variations to invoke emotion in paintings. Similarly, a professional photographer carefully tunes the temperature and tint of existing colors in a photograph to convey specific emotional feeling. Professionals usually rely on experience and intuition to choose appropriate color compositions which makes this process tedious and labor-intensive. Thus, it is desirable to mathematically formulate emotional styles in an equation format, so that can then be applied to new images for invoking emotional appeal.

Our goal in this paper is to learn implicit adjustment rules associated with the emotional properties of images from a set of training examples. Given a pair of image before and after adjustments, we would like to discover the underlying mathematical relationships optimally connecting the color-contrast properties between them. Some of the key issues that arise in order to achieve our goal are: *(1) How to choose the set of image parameters which will be used for enhancing the emotional appeal of images? (2) How do we capture the notion of enhancement of images in a mathematical fashion to understand the implicit relationship between the original image and its adjusted version? (3) Given an arbitrary image, how to derive the*

enhancement operations that can be applied to the image to enhance its emotional appeal? To meet up these problems, as a first step we collect ground-truth data and construct a training database consisting of sample images and their different enhanced versions, and ask a group of human participants to mark their preferences. We implemented a simple and effective web-based user-interface for this purpose and sent invitations for participation to a group of participants. A web-based interface allows to reach out to a larger group of audience in a scalable fashion in comparison to a local desktop interface which is tied with the restriction of lab premises and time. To answer (1) in the above discussion, we selected a set of contrast and color properties which is found to have a high influence on the emotional impact of images [4]. We use machine learning techniques to understand the implicit relationship between an original image and its adjusted version to address item (2) above, and derive statistical solutions to develop an enhancement function in a mathematical form to address our third research question (3) above. We conduct user studies to evaluate the effectiveness of our approach.

The contributions of this paper are listed as follows: (a) per our knowledge, this is one of the first attempts to derive a computational framework for enhancing the emotional impact of images; (b) we collect ground-truth data from human participants which is a valuable resource for understanding the underlying relationship between them; (c) we employ a data-driven systematic framework to learn models from training data and derive generalized enhancement functions for arbitrary/unseen images using machine learning and statistical techniques; and (d) we derive an objective metric for evaluating the emotional enhancement of images using a color mood space model and use it to test our approach.

II. RELATED WORK

Image enhancement have been studied in literature from various contexts. Quality enhancement of images is one of the heavily studied areas and various proposals are found in the literature. The general objective of all these techniques are essentially to improve the quality of pictures by manipulating color, contrast, tonal, textural features of the degraded images [1], [2]. [3]. With the success of the above techniques, the need to build more sophisticated controls with complex characteristics emerged. Shape and orientation based adjustments for the various objects within the image were proposed by many researchers, which led to the development of interesting application areas. [15] presented a system for artistic perspective manipulation to produce a variety of effects such as changing the perspective composition of a scene which is not realizable with a camera using shape-preserving image warps. [16] proposed an easy-to-use enhancement technique for realistic reshaping of human bodies in a single image producing visually pleasing results with a variety of poses and shapes using a novel body-aware image warping technique. A data-driven enhancement system for facial attractiveness in frontal portraits is proposed in [17] in which a set of distances is learned from a variety of facial feature locations using a ground-truth training dataset.

Another line of research is followed by enhancing images using color and tonal properties of images to achieve various objectives. One general trend in this direction is to transfer the image characteristics of a source image to a target image by color harmonization [18], tonal and textural features [19]. Other proposals in this category of learning characteristic image properties from example images and then using this knowledge to enhance arbitrary or unseen images are: color harmonization [20], relighting effects by neutralizing the light colors using spatially varying while-balance [21], personal photo enhancement [22], color theme [23], color and tone style [24], collaborative personalization [25], [26]. [27] described an image restoration method using various operations such as white balance correction, exposure correction, and contrast enhancement by leveraging a large database of images gathered from the web. An effective de-colorization algorithm is proposed in [28] that preserves the appearance of the original color image by blending the luminance and the chrominance information in order to conserve the initial color disparity while enhancing the chromatic contrast. An efficient method for recovering reliable local sets of dense correspondences between two images with some shared content is described in [29] with applications in automatic example-based photograph enhancement such as adjusting the tonal characteristics of a source image to match a reference, transferring a known mask to a new image, and kernel estimation for image de-blurring.

Computational models for media aesthetics [4] and emotion [8] are slowly gaining importance in bridging the semantic gap [30] and to create new applications [12] due to close correlation with human visual perception. Though there are many image-editing software present on the market such as Adobe Photoshop and the open-source GIMP, these applications require manual control for enhancing images, which is not practical in large-scale. And although they are provided with automated controls such as the Retinex filter [1], these techniques do not consider affect or emotion while adjusting, which is our focus in this paper. We follow a data-driven approach in our framework since it is non-intrusive in nature (no human involvement after collection of training data) and more practically feasible due to its computational characteristics.

III. ENHANCEMENT FRAMEWORK

The high-level schematic framework of our image enhancement system is shown in Fig 1. The first step involves a training phase which essentially builds a database consisting of a set of images and their corresponding metadata which are rating values. We utilize the International Affective Picture System (IAPS) database [31], which is a standardized benchmark of emotion/affect ratings in images widely acknowledged by researchers in psychology-emotion domain and distributed for the purpose of research. The input database images $\boldsymbol{I^{in}}$ are fed into our enhancement processing framework to derive the corresponding output images $\boldsymbol{I^{out}}$ which are presented to human participants for rating metadata $\square$ using a web-based interface. The enhancement processing framework is a set of color-tonal image operations denoted by a vector of enhancement features $\boldsymbol{\varphi}$.

The training database records each participant observation $\square^{u}_{i}$ and their corresponding enhancement vector $\boldsymbol{\varphi}^{u}_{i}$ for each user $\boldsymbol{u}$ and image $\boldsymbol{I_i}$ in a table. Once the training database is prepared with sufficient samples, we analyze it to learn a statistical model that encodes the relationship between $\square$ and $\boldsymbol{\varphi}$. Specifically, we cluster the set of original images $\boldsymbol{I^{in}}$ in a number of different groups and then estimate the enhancement vectors of each group using a regression-based learning approach followed by an optimization solver. Once such a statistical model is learned, we can enhance a new or unobserved image by mapping the image to one of the trained clusters and applying the enhancement operator for that cluster.

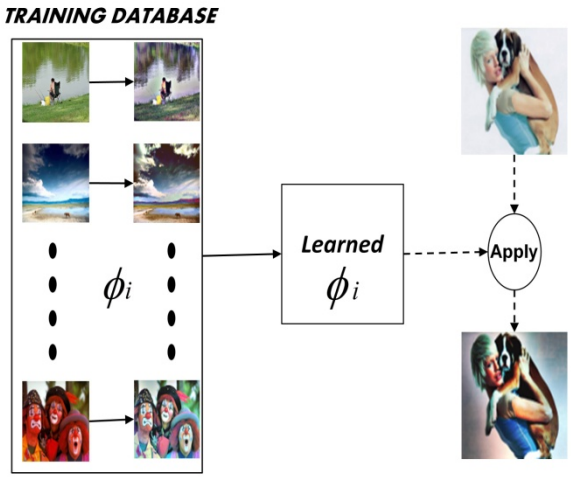


Fig. 1 Block diagram of the enhancement framework. φ is the enhancement vector which is a set of control features based on color and tonal properties

A. Image Database

As already mentioned, we utilize an affective image database known as International Affective Picture System (IAPS) as ground-truth for building our models. The IAPS currently includes around 1,200 images depicting a variety of human experience: joyful, sadness, fearful, angry, threatening, attractive, among others with a virtual world of pictures. The stimuli are standardized on the basis of ratings of pleasure and arousal in a 9-point scale for each dimension. Each image is rated by approximately 100 participants covering a diverse group based on gender, cross-culture, age, and many more. All the images are in color and have a resolution of 1024x768. Each trial is associated with a 5 seconds preparatory cue followed by a 6 seconds presentation of the to-be-rated image, and then 15 seconds for final ratings of pleasure and arousal. The affective space that is defined by the mean ratings of pleasure and arousal are illustrated in Fig 2, plotted in the 2-dimensional space. We manually selected a representative subset from the original set since it requires a lot of human resource to collect sufficient amount of samples for the total set of 1,000 images.

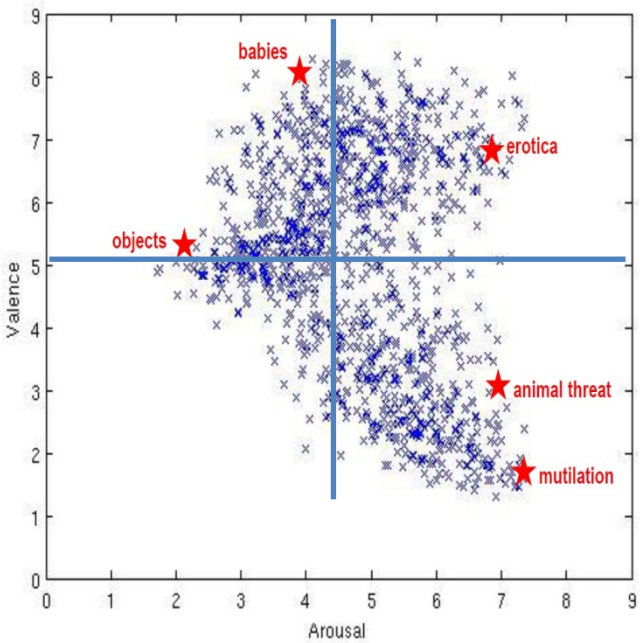


Fig. 2 Each image in the IAPS is placed in a 2-dimensional affective space on the basis of its mean valence and arousal rating on a 9-point scale with markers at some distinctive image categories to get an insight into the relationship between emotional semantics and the 2D space

B. User Interface

We designed a web-based user interface for collection of training data related to the affective enhancement of images, specifically the V^u_i, A^u_i columns in the database table (discussed later in Section V). Our user interface is kept simple and intuitive so that an ordinary user can easily enter their ratings as shown in Fig 3. Engaging professional photographers for manual image adjustment to build training data appears to be an attractive option, but it is more expensive with high resource consumption for even collecting a small number of samples. Thus, our target set of participants are university students without any prior knowledge of professional photography. The possible options for designing the rating procedure to capture emotional enhancement are manual or decision-based. In a manual-based procedure, the participant is provided with various controls to edit the image so that the subject can perform the enhancement based on his/her judgment. This process is geared for professional photographers (not easy for ordinary users to enhance emotional impact with edit controls) and local desktop setting, which does not fit our design choices. The decision-based procedure generally involves a binary yes/no or good/bad type of opinion which is very simple for the ordinary users but carry very less information.

We identified a simple procedure with more meaningful information than the decision-based process, using a 2D coordinate grid which is located at the right side of the user interface shown in Fig 3. The 2D coordinate system is calibrated with valence-arousal dimension on a 9-point scale which is similar to the ratings in IAPS dataset. The

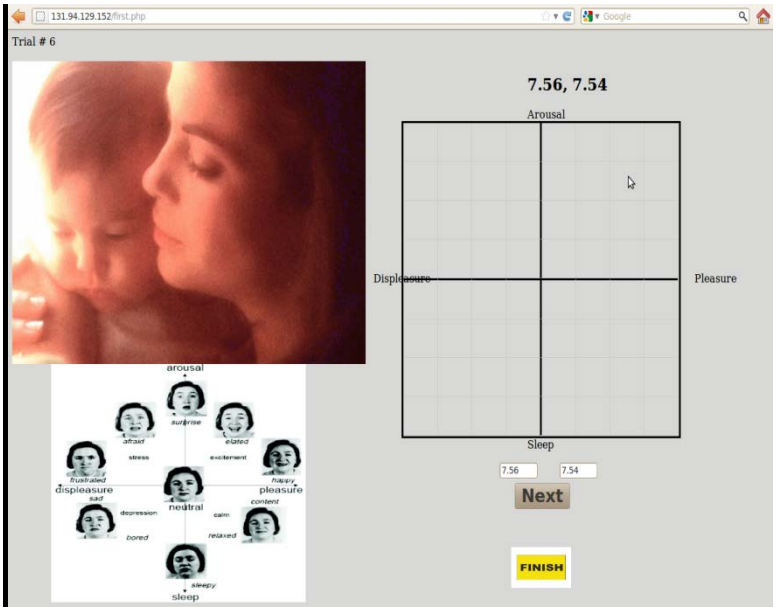


Fig. 3 The web-based user interface for collecting data of affective enhancement for the images displayed on the left of the panel

The 2D-grid in Fig 3 the (0,0) coordinate point represents the neutral state (i.e. no arousal and no pleasure), whereas in Fig 2 the (4.5, 4.5) coordinate point represents the neutral state. We employ the single stimulus continuous quality evaluation (SSCQE) methodology for this set-up [36]. The user just needs to mark a point in the 2D affective space with the help of a mouse which will indicate its rating for emotional impact with respect to the image displayed in the left side of the interface. This 2D rating procedure, based on the affective space with valence-arousal calibration, associates more meaningful information than a simple yes/no decision process and is also relatively simple for the non-expert user. We illustrate the relationship between image and the corresponding 2D affective spatial coordinates using examples from the IAPS dataset in the web interface before the start of the final rating procedure. This provides sufficient insight to the participant regarding the spatial position of the displayed images in the affective space. During the rating procedure we allow multiple clicks but only the coordinate valence-arousal values of the final click are submitted. As soon as the participant clicks on a position in the 2D space, the system displays the value of the clicked coordinate positions at the top of the grid as shown in Fig 3, which allows better judgment by the subject. A reference image is always kept in the bottom-left portion of the user interface as (cf Fig. 3) which provides the position of some higher-level emotional concepts such as ``fear'', ``anger'', ``sadness'' in the 2D affective space. The images on the top-left portion of the interface are various enhanced versions of the base training set from the IAPS dataset, and they are generated dynamically and displayed after each submit clicked by the participant. The various relevant information such as image metadata, enhancement parameters, and valence-arousal emotional ratings are stored in a training database for data analysis. We incorporated a set of twenty-five images from the IAPS database with a uniform distribution of Valence-Arousal rating values for the collection of training data samples.

IV. ENHANCEMENT CHANNEL

For designing the enhancement channel, we first need to identify the vital image parameters which are known to influence human affect. There are already ample evidence in the literature where it is found that images can influence human emotions to

a considerable extent [5]. Moreover, recently there are many proposals to indicate that different image properties such as color, contrast, saturation, or brightness contribute to invoke human emotions, and moreover these properties can be exploited to address retrieval [8], summarization [11], presentation [10].

After carefully studying the literature, we found evidence of various features that are related to emotion (such as color, texture, composition, content), and the two most widely used parameters are contrast and color [8]. Following our observation, we designed the enhancement channel based on the following parameters: power curve and S-curve shaping: for contrast adjustment; and color tint and temperature: for color adaptation. Though the framework can be extended with more effective parameters, we limited the number of parameters primarily to limit the complexity of our algorithm and keep its feasibility, since the computational space will increase exponentially with the number of parameters. Fig 4 shows the entire block of components that are incorporated in our enhancement channel. The initial preprocessing step involves the linearization of the nonlinear RGB space with gamma curves applied to the input image $\boldsymbol{I^{in}}$. This is followed by another preprocessing step of auto-enhancement of the images which is intended for normalizing the difference in image qualities and bringing them within comparable limits. Then the channel applies two contrast adjustment functions i.e., power curve and S-curve shaping, followed by the color operators i.e., temperature and tint adaptation. These are the core operations and the enhancement vector is derived from these transformations. Finally, a post-processing step of inverse linearization is performed to revert back to the nonlinear space to produce the final enhanced image $\boldsymbol{I^{out}}$.

A. Linearization

Linearization is generally adopted to minimize the variation between device hardware since different cameras may have been calibrated with different intrinsic properties. We use the simple power curve for the process of linearization as follows:

$$c = c^{\gamma} \tag{1}$$

where c is the normalized values of the R,G,B channel and □ = 2.2 The inverse linearization process or reverting back to the nonlinear space which is the last component block of the enhancement channel in Fig 4 is the same operation but with γ = 1/2.2 = 0.45.

B. Auto-Correction

We initiate an auto-correction step to factor out the differences of qualities among the various images in the IAPS database. This allows to correct the degraded images if present in the database to a more acceptable level upon which our enhancement operator will function. We follow a generic methodology consisting of white balance and contrast stretch which are the two most simple operations for correcting an image.

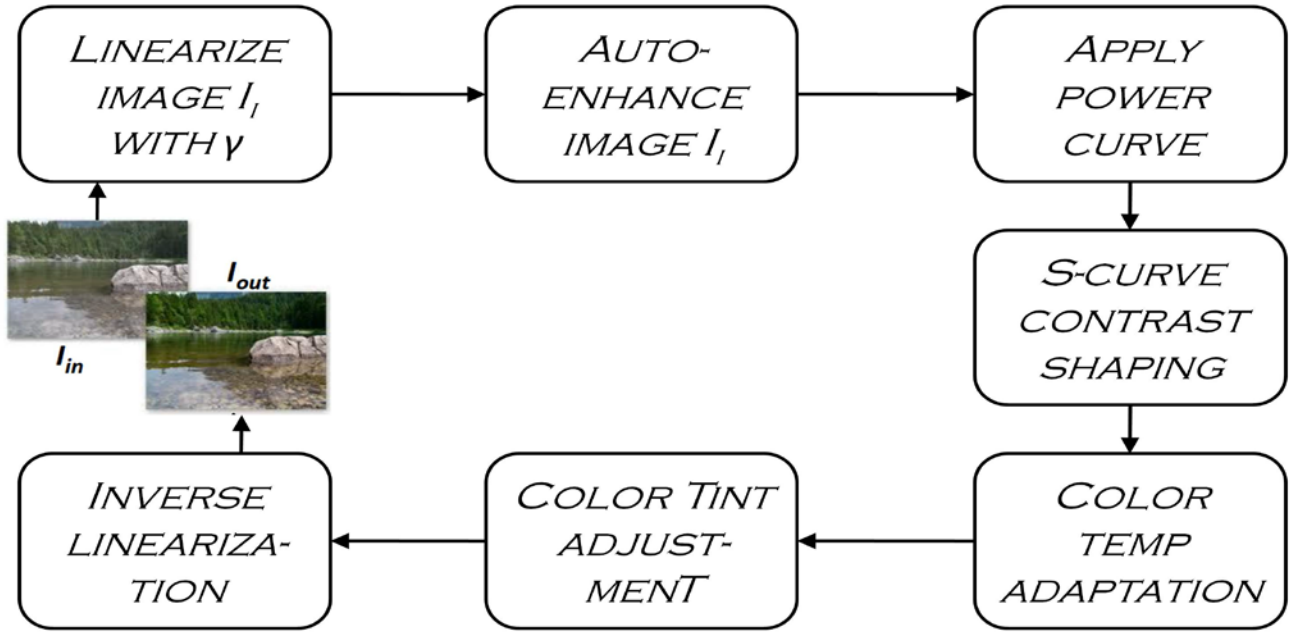


Fig. 4 Exploded view of the entire enhancement framework with the various sub-components and the end-to-end sequential operator chain followed by an image from the point of input to the final output.

White balance is the process of removing unrealistic color casts from images which are mostly generated from different acquisition condition with varying illuminations. The human visual system is generally able to render the perceived colors of objects almost independent of illuminations by a phenomenon known as Color Constancy, which is mimicked in the digital

cameras by a process of Automatic White Balancing to suppress unrealistic colors. From a computational perspective, white balance is a two-step procedure: the illuminant is estimated, and the image colors are then corrected on the basis of the estimate. For estimating the illuminant, we follow the Gray World algorithm which assumes the average surface color in a scene is gray and the shift from gray of the measured averages on the three channels corresponds to the color of the illuminant. We find the average values of the RGB color components and use their average to find an overall gray value for the image. Each color component is then scaled by a factor ratio of the gray value to the appropriate average of each color component.

Contrast stretch is the process of evenly distributing the pixel intensities within a defined span which is essentially used to stretch the compact region of the pixel values in the image histogram. We fix the span at 0.5% on the darker side and 1% on the brighter side followed by linear stretching of the original values to the new range with a shift and scale operation applied to all the color bands.

C. Contrast Shaping

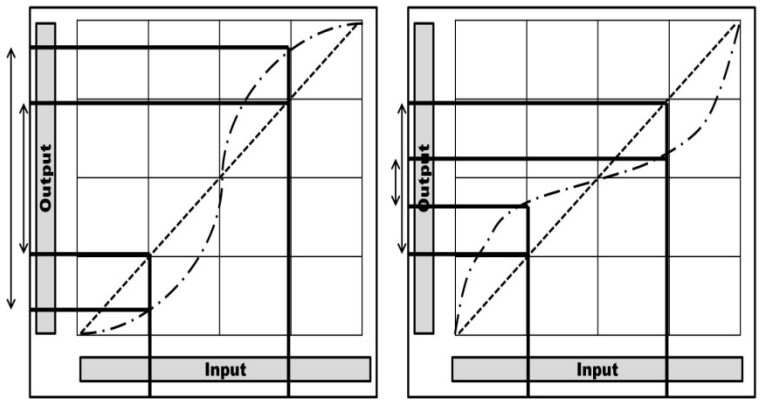
We apply two transformations for contrast adjustment i.e., power curve and S-curve as shown in Fig 4 as follows:

Power curve is similar to the gamma curve that was employed in the linearization process, but its value is not fixed at 2.2 as before, and is rather adjusted at different levels which allows to learn the affective preference:

$$i_{out} = i_{in}^{\gamma} \tag{2}$$

where $i_{in} i_{in}$ and i_{out} are the normalized input and output intensities of each pixel.

S-curve is the most common tool for adjusting contrast and is found in almost all commercial products such as Adobe Photoshop. The contrast of the tonal curve can be increased or decreased to varying degrees by applying different shaping parameters for the S-curve as shown in Fig 5.



(a) (b)

Fig. 5 (a) shows increase of contrast and (b) shows decrease of contrast by varying the shaping parameters of S-curve

The formula for S-curve shaping is as follows:

$$i_{out} = \begin{cases} q - q\left(1 - \dfrac{i_{in}}{q}\right)^{\alpha} & if\ i_{in} \le q \\ q + (1-q)\left(\dfrac{i_{in}}{1-q}\right)^{\alpha} & otherwise \end{cases} \tag{3}$$

D. Colour Temperature

Color temperature has been found to have a high influence on emotion whereby the ``warmth-ness" or ``cool-ness" of an image can indicate the nature of affect generated by the picture [5]. Many professional photographers adjust color temperature in pictures using commercial software such as Adobe Photoshop to generate desired effects of emotion. Warmer temperatures give the images a sense of warmth and coziness, while cooler temperatures can make images seem cold and harsh [5]. The color temperature of an image can be determined by comparing its chromaticity values with that of an ideal black-body radiator. The scale of color temperature varies from 1,900 Kelvin to almost around 10,000 Kelvin and the standard is calibrated at 6,500 Kelvin which signifies daylight settings. Color temperatures over 5,000 Kelvin are known as cool colors (bluish white) while lower color temperatures around 2,500~3,000 Kelvin are known as warm colors (yellowish white to red) as illustrated in Fig 6.

Fig.6 Block of image showing the variation of the different color temperatures on a single image. The color temperature generally varies in a blue-yellow axis. The middle image is the original one and moving towards the left increases the color temperature (yellow end) and towards the right decreases the color temperature (blue end).

The computation algorithm estimates the correlated color temperature, T_c by way of interpolation from lookup tables and charts using the popular Robertson's method. The first step involves converting the color space from RGB to CIE XYZ, and then the chromaticity values (u, v) are obtained using the CIE 1931 coordinate system as follows:

$$u = \frac{4x}{-2x+12y+3} \quad v = \frac{6y}{-2x+12y+3} \tag{4}$$

Then a search is initiated through the set of standard isotherms from the lookup table to find the two tightest adjacent lines between which the test chromaticity i.e., (u, v) lies. Then, the correlated color temperature or T_c can be calculated as follows:

$$\frac{1}{T_c} = \frac{1}{T_i} + \frac{d_i}{d_i - d_{i+1}}\left(\frac{1}{T_{i+1}} - \frac{1}{T_i}\right) \tag{5}$$

The distance between the test point (u_T,v_T) and the ith i^{th} isotherm (u_i,v_i) is given by:

$$d_i = \frac{(v_T - v_i) - m_i(u_T - u_i)}{\sqrt{1+m_i^2}} \tag{6}$$

where m_i is the slope of the ith isotherm computed by $m_i=-1/l_i$ and l_i is the slope of the locus at (u_i,v_i).

E. Colour Tint

Color tint can be viewed as orthogonal to color temperature whereby the control changes along the green-magenta axis which are at two complementary positions in the color wheel as illustrated in Fig 7. To adjust color tint, we utilize the hue channel by converting to the HSV color space. Now, we need to design an adjustment function which can change the tint of the hue channel by varying degrees. The strategy involves shifting the hue of each pixel towards the desired tint i.e., along the green-magenta axis. Instead of a simple linear shift of the hues which may produce unwanted color artifacts, we apply a technique adapted from [20] using Gaussian function whereby the overall tone of the image is not greatly affected. Let the hue of pixel p be denoted by h_p and the tint hue of the sector associated with green-magenta axis to be h_t, then we can derive the value of shifted hue $h_p^{'}$ as follow:

$$h_p^{'} = h_t + \frac{w}{2}\left(1 - G_\sigma\left(\left\|h_p - h_t\right\|\right)\right) \tag{7}$$

Fig. 7 Block of image showing the variation of the different color tints on a single image. The color tint generally varies in a green-magenta axis. The middle image is the original one and moving towards the left increases the color tint (green end) and towards the right decreases the color tint (magenta end).

where w is the arc-width of the green-magenta sector and G_σ is the normalized Gaussian function (such that $G_\sigma(x) \in (0,1]$ with mean 0 and standard deviation σ; the hue distance $\|\bullet\| \|\cdot\|$ is the arc-length distance on the hue wheel measured in radians. The width of the Gaussian σ is a user-defined parameter that may vary between zero and w, where larger values of σ creates hue concentrations near the sector center and smaller values lead to concentrations near the sector boundaries. In our implementation, we use the mid-point $\sigma=w/2$. Thus, the adjustment parameter for color tint or T is the value of w which can be varied to achieve different effects of tint.

Thus the final enhancement vector can be found to be $\phi = \langle \gamma, q, \alpha, K, T \rangle$ where γ is from power curve, q, α from S-curve, K from color temperature, and T from color tint.

V. LEARNING FRAMEWORK

In this section we describe our learning framework which basically involves deducing a best set of enhancement vector (with respect to its match with the user's preference) which is later used for applying to unseen/test images. Until now, we have observed the enhancement feature vector covering various adjustments of an image, which is utilized to gather training data through the web-based interface. The input to the learning framework is a table of training data where each row is denoted as follows:

$\langle I_i, \phi_i^u, V_i^{gt}, A_i^{gt}, V_i^u, A_i^u \rangle$ for user u and training/seen image Ii as described before, IAPS ground-truth affective values (V_i^{gt}, A_i^{gt}) and user-rated affective values of (V_i^u, A_i^u) ,and the expected output of the best set of enhancement vector is $\phi' = \langle \gamma', q', \alpha', K', T' \rangle$ for test/unseen image It.

One point worth noting is that we use the original rating values from the IAPS database, and augment them with new user-rated values of their corresponding image, after applying enhancement operators. This is required in our methodology to detect the amount/pattern of user-rating with respect to the enhancement operations.

In order to solve this objective, we need to resolve the following problems: (1) how to find an enhancement vector XXXX for an unseen/test image? (given the fact that the training samples which are labelled with seen/training images Ii and the new/test images Ii are distinct subsets);

(2) How to define the degree of enhancement in each training sample or specifically, how to define the relationship strength between IAPS ground-truth (V_i^{gt}, A_i^{gt}) and user-rated values of (V_i^u, A_i^u) for each image I_i since, per our design interface, each sample of the training database has varying degrees of enhancement?

(3) How to learn the relationship between the enhancement vector E and the valence-arousal based emotional values (V,A) ?

(4) How to compute the best set of enhancement vector φ' from the training data samples?

We take a modular approach to address each of the above concerns as follows: k-means clustering for (1), a mapping function for (2), regression-based learning for (3), and an optimization solver for (4) as shown in Fig 8.

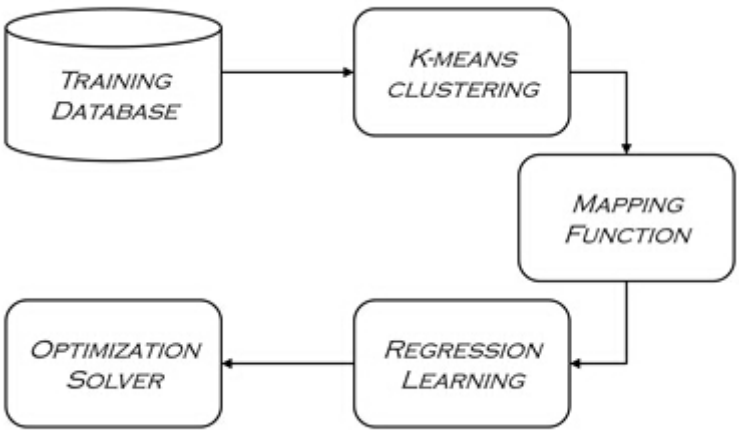


Fig. 8 Block diagram of the learning framework where the objective is to derive a best set of enhancement vector from a training database of samples

A. Clustering

This is the first component in the learning framework where the samples from the training database are fed to a K-means clustering algorithm for processing. The clustering is performed on the ground-truth values from the IAPS dataset $\left(V_i^{gt}, A_i^{gt}\right)$. The rationale behind the adoption of clustering is as follows: IAPS images are tagged with user-rated valence-arousal ratings in a 2-dimensional scale, but we wish to divide them into separate classes (in line with discrete theory where the emotions are classified as happy, anger, sad, etc.) and then derive an enhancement vector for each of those classes. This observation follows from previous research where it has been shown that each discrete class of emotion possesses distinctive color/contrast/tonal characteristics of image which are exploited for the affect-based retrieval of images [9]. Thus, our hypothesis is to enhance the emotional characteristics of images from each class with a distinctive function, so that the emotional semantics (in the form of color/contrast properties) can be embedded in the images effectively. Another way is to explore the enhancement in the valence-arousal dimensions, whereby we can learn enhancement functions for valence and arousal separately. In such a case, we only need to omit this component of clustering and the rest of the framework can be used to attain the objective. Enhancing an image in terms of valence and arousal is an interesting direction which has not been explored before, and we intend to explore this area in future.

The K-means clustering is a popular unsupervised learning algorithm where the data points are grouped into a user-defined (K) number of clusters.

Given a set of training examples from the database $\{x^{(1)}, x^{(2)},...,x^{(m)}\}$ where each $x^{(j)}$ is a valence-arousal tuple $\left(V_{i,(j)}^{gt}, A_{i,(j)}^{gt}\right)$ and $j \rightarrow 1$ *to* m is the number of samples in the database, the algorithm starts by initializing cluster centroids $\mu_1,\mu_2,...,\mu_K$ randomly. Then at each iteration step until convergence, it assigns a training example $x^{(j)}$ to the closest cluster centroid μ_K by minimizing the distance between them, followed by moving each cluster centroid μ_K to the mean of the member points assigned to it as follows:

$$\forall i: \; c^{(k)} := \arg\min_k \left\| x^{(i)} - \mu_k \right\|^2 \tag{8}$$

$$\forall k: \; \mu_k := \frac{\sum_{i=1}^{m} 1\left\{c^{(i)} = k\right\} x^{(i)}}{\sum_{i=1}^{m} 1\left\{c^{(i)} = k\right\}} \tag{9}$$

where $c^{(j)}$ is the cluster membership of sample j, $1\{.\}$is a membership function between sample i and cluster j evaluating 1 if present in cluster or zero otherwise, and $\|.\|$ is a sum of squared distance function between training sample $x^{(i)}$ and the assigned cluster centroid μ_k. After k-means clustering, each sample in the training database will be: $\left\langle I_i, \varphi_i^u, V_i^{gt}, A_i^{gt}, V_i^u, A_i^u, c_i^k \right\rangle$where the new field c_i^k denotes the cluster membership of image I_i in cluster k. In our framework, we assume $K=4$, which can intuitively correspond to major affect categories (i.e., happy, sad, anger, fear) [8]. Fig 9 shows the plot corresponding to the deduced clusters on a subset of 167 images from the IAPS dataset.

B. Mapping Function

Now, we try to derive a metric that can quantify the degree of enhancement (D) of each sample from the training database. Formally, the amount of change from $VA_{(j)}^{gt} = \left\langle V_{(j)}^{gt}, A_{(j)}^{gt} \right\rangle$ to $VA_{(j)}^{u} = \left\langle V_{(j)}^{u}, A_{(j)}^{u} \right\rangle$ for sample j of ground-truth rated V-A values in IAPS database with enhancement rated V-A values in training database. One naive way is to use a metric of $D_{(j)} = \left\| VA_{(j)}^{u} - VA_{(j)}^{gt} \right\|$ here $\|.\|$ is an Euclidean distance function, but this is not a good choice since it does not consider the direction between two data points. In reference to the ground-truth point $V_{(j)}^{gt}$ as the origin in the affective space, $V_{(j)}^{u}$ can be placed in four quadrants i.e., first quadrant $[V^+,A^+]$ when $V_{(j)}^{u} > V_{(j)}^{gt} \,\&\&\, A_{(j)}^{u} > A_{(j)}^{gt}$, second quadrant $[V^+,A^-]$ when $V_{(j)}^{u} > V_{(j)}^{gt} \,\&\&\, A_{(j)}^{u} < A_{(j)}^{gt}$, and likewise for the others. Here, we would like to take another hypothesis as follows: the degree of enhancement D should also consider the quadrant relationship between $V_{(j)}^{gt}$ to $V_{(j)}^{u}$ unlike the naive approach discussed before. The rationale behind this hypothesis is intuitive: enhancement to $[V^+,A^+]$ should have higher relevance than $[V^+,A^-]/[V^-,A^+]$, which should be higher than $[V^-,A^-]$, since both the valence and arousal values are increased in the first case, either valence/arousal is increased and the other one decreased in the second case, and both the valence and arousal are decreased in the last case. So, the enhancement degree D should consider three levels of relevance based on the directional aspect of the comparison, apart from the positional aspect defined by $\|.\|$ as discussed above.

Now, we can easily derive D using simple scaling and shifting techniques as follows: Assuming, we want the value of D to be distributed in [0,1], a, b, c are the starting point for each quadrant mapping such that $0 \le a < b < c < 1$, the diagonal end-point 2d coordinate of each quadrant opposite to origin be Z^q (where Z^I for $[V^+,A^+]$ is (9.0,9.0), Z^{II} is (0.0,9.0), and likewise since we already know that the IAPS affective 2d space is calibrated in a 9-point scale):

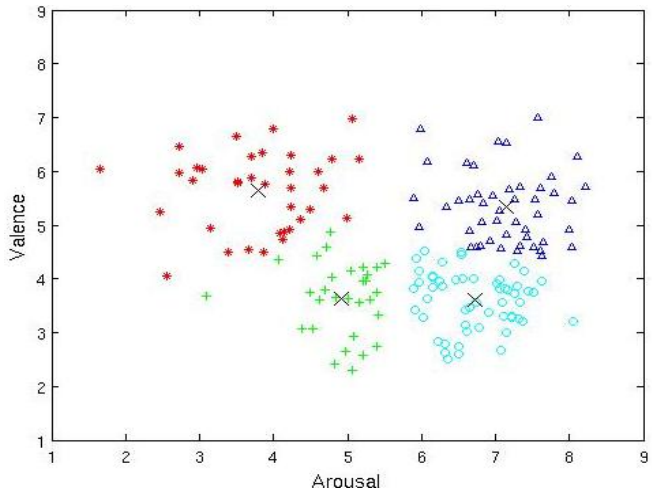


Fig. 9 The clustered images from the IAPS database in the affective space and the respective centroids marked with a X

$$R_{(j)} = \frac{\left\| VA^{u}_{(j)} - VA^{gt}_{(j)} \right\|}{\left\| Z^{q}_{(j)} - VA^{gt}_{(j)} \right\|} \tag{10}$$

$$D_{(j)} = s + \frac{R_{(j)}}{e - s} \tag{11}$$

where [$s=a$, $e=b$] is the segment interval for enhancement quadrant [V^-,A^-], [$s=b$, $e=c$] for [V^+,A^-]/[V^-,A^+], and [$s=c$, $e=1.0$] for [V^+,A^+] with respect to each training sample j as shown in Fig: 10.

In our framework, we heuristically assume $a=0.0$, $b=1.0$, $c=3.0$ which classifies varying degree of relevance for different enhancement quadrants. Thus, each training sample can be denoted as $\langle I_i, \varphi^u_i, D, c^k_i \rangle$ (some fields are dropped since they are not required for further processing) where we quantified the degree of enhancement by D. Design choices different from our hypothesis can also be easily integrated in the framework with small modifications of the mapping function.

One such choice of importance could be based on voting, where the number of samples in each quadrant is calculated for each image, and then the importance is computed in proportion to the majority share. If there are more number of training samples for image I_i with enhancement to [V^+,A^-] (majority of user's prefer enhancement of I_i by increasing valence and decreasing arousal), then the highest importance is given for this quadrant mapping, unlike our hypothesis where we assume a fixed bias.

C. Regression Learning

The goal of this component is to learn a relationship function between the enhancement vector $\varphi^{(k)} = \langle \gamma^{(k)}, q^{(k)}, \alpha^{(k)}, K^{(k)}, T^{(k)} \rangle$ for each cluster k. First, we divide the training samples based on its cluster membership $c_{(j)}^{k}$, and then we learn a regression function for each cluster. The rationale for choosing a regression learning approach is due to its suitability for deriving a function of input variables to correlate with the output variable, see Fig 10. Regression is a supervised learning algorithm which initially starts with a representation of function/hypothesis:

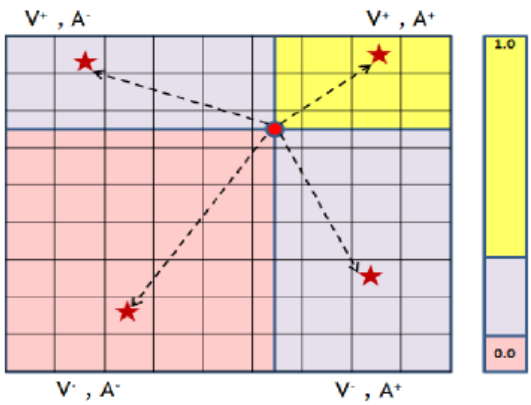


Fig.10 The pictorial depiction of the mapping function using scaling and shifting techniques from the 2D affective space on the left to the 0.0-1.0 linear scale on the right to capture the degree of affective enhancement

$$h(\phi) = \sum_{i=0}^{n} \theta^i \phi^i = \theta^T \phi \tag{12}$$

where θ^i are the regression coefficients parameterizing the space of linear function mapping from φ, $n=5$ in our case where $\varphi=<\gamma, q, \alpha, K, T>$ and $x_0=1$ which makes θ_0 the intercept term. We compute the coefficient vector θ^T using a gradient descent approach to solve the least mean squares optimization problem of minimizing the error term as per the following update rule:

$$\forall i \rightarrow 1 \; to \; 5 : \theta_i := \theta_i + \alpha \sum_{j=1}^{m} \left[D_{(j)} - h\left(\phi_{(j)}\right) \right] \phi^i_{(j)} \tag{13}$$

where α is the learning rate, m is the number of training samples, and the above process is repeated till convergence. After termination, a set of five regression coefficients and an intercept term θ^T are derived, and the process is repeated for each cluster $c^{(k)}$ to obtain a set-of-set of regression parameters $\{\theta^T\}^{(k)}$.

D. Optimization Solver

Now, we proceed to the last component of the learning framework which involves an optimization problem of deriving a set of enhancement vector φ^k for each cluster k by maximization of $D^{(k)}$. This problem involves maximizing the function learned in Eq: 12 so that we can derive a best set of enhancement vector φ' (our initial objective when we started with the design of the learning framework as discussed before) which statistically provides the maximum degree of enhancement (or highest value of $D^{(k)}$). We formulate the problem as a methodology for linear programming (LP) as follows:

$$\max mize : \theta^T \phi + \theta_0 \tag{14}$$

$$subject \;\; to : 0 \prec \phi \prec h \tag{15}$$

$$\phi \succ \phi_{(j),\max} \tag{16}$$

where φ is the optimization vector, θ is a vector of known coefficients, h is the upper bound of φ variables which is 1 in our case since we have them normalized, and $\prec, \succ$ denotes element-wise inequality. $\varphi_{(j),max}$ is one set of $\varphi_{(j)}$ values for a training sample j with a maximum value of $D_{(j)}$ selected from the training database since we want to derive a best set of vector φ' which is explicitly greater than the current maximum $\varphi_{(j),max}$ from the database. We use the CVX [32] optimization solver which exploits the widely known interior-points method to solve the linear problem by moving through the interior of the feasible region, unlike traversing the edges between vertices on a polyhedral set typically performed in the simplex algorithm. We repeat the process to derive the best set of enhancement vector $\varphi_{(k)}'$ for each cluster. Thus, we are now ready with a separate function of enhancement variables $\left\langle \gamma'_{(i)}, k'_{(i)}, \alpha'_{(i)}, K'_{(i)}, T'_{(i)} \right\rangle$ for each cluster i which can be applied to any unseen/test images for affective enhancement.

E. Applying Enhancement

We are interested in predicting enhancement preference for test image I_{test} for arbitrary/unseen source. The key observation lies in the fact that, once we know the cluster membership of the new image, then we just select the corresponding enhancement vector $\left\langle \gamma'_c, k'_c, \alpha'_c, K'_c, T'_c \right\rangle$ to find the individual parameters, and then adjust the contrast and color based on the values accordingly. The adjustment procedure is described in Sec: IV with fixed values of the derived best set of enhancement vector. Thus, the problem really reduces down to establishing the cluster membership of the new image I_{test}. Images with V-A ratings in a 9-point scale from IAPS database or some other sources can easily be integrated in our framework. We find the closest cluster centroid μ_k and assign the image to that cluster, thereby applying the enhancement operator from the corresponding cluster.

The scenario is a little bit complicated for unrated new images where we need to somehow identify the cluster membership of the image. One solution is to employ a classifier such as SVM to learn a model of image features and its associated class of cluster membership from the training database samples (training sample are grouped into various clusters in Sec: V.A). A single multi-class classifier or multiple binary-class classifiers can be designed for this purpose. For multiple binary-class classifiers, an aggregation technique is required to derive a single prediction of cluster membership. The selection of image features is a critical issue and is still an open problem, but recent research provides considerable success with high classifier accuracy for affective retrieval of images [9]. Some of the possible image features are: saturation/brightness contrast, hue statistics, colorfulness, Itten contrasts, wavelet features, level of detail, dynamics, among others [8]. The classifier predicts the cluster membership and the rest of the procedure follows what we discussed before. We can also employ regression to learn the valence and arousal ratings with image features as input variables from training database samples. The learned model predicts the V-A values for unseen/new images, which can be mapped to a cluster based on the distance to the centroids.

VI. EXPERIMENTAL EVALUATION

Evaluating the quality of affective enhancement is difficult, since there is no universal agreement on the perceptual metric.

Its nature is more subjective and so developing objective evaluation models are challenging.

A. User Study

We carried out a user study to evaluate the effectiveness of our enhancement framework with respect to a set of testing images. A simple web-based interface is built for this purpose as shown in Fig 11. The interface facilitates pair-wise comparison between the original and the adjusted images. The user feedback mechanism is provided with three options: left (indicates the original invokes more emotion than the adjusted i.e., negative case for our method), right (adjusted image is more emotional than the original i.e., success of our method), and neutral (within comparable limits i.e., enhancement does not produce significant gain). We prepared a written set of instructions for the participants to familiarize themselves with the task, interface, and the images following ITU-R500 recommendation [36]. Five initial random trials are provided without recording the results for stabilizing the participant's opinion [36]. We employ the double-stimulus categorical scaling methodology for the subjective evaluation in our user study as indicated in [36]. The participants are presented with a series of image pairs in random order for rating [36]. Each user session typically lasts for 30 minutes with 30-sec interval between each stimulus. For the grading scale, we selected adjectival categorical judgment method with three levels corresponding to +ve, -ve, and neutral analogous to the ITU-R scales[36]. The users rating are stored in a database for later processing. We sent the web link for the user study to a mixed participant group with a few users who already provided feedback for building the training ground-truth data and some new users. This will provide an evaluation which can test the generalization of our approach by allowing some new participants in the user study, thereby avoiding any preference bias with a fixed set of user group (those already provided feedback for training ground-truth data collection). Twenty-five participants contributed to the user study voluntarily through campus-wide ads. The participants were all proficient with English language (13: native, 7: advanced, 5: intermediate). Among them, 14 were male and 11 were female. All participants were between the age of 18-48 years (Mean=33.72, Standard Deviation=4.39). Every participant reported normal or corrected to normal vision. We selected a testing set of 50 images from the IAPS database for the user evaluation study. The participant group is entirely from university students of various levels (undergrad to graduate) and no one is an expert of photography as per our knowledge. We explicitly mentioned to the participants to judge the results purely from an emotional aspect and avoid any bias due to quality, since our main target is to enhance the affective features rather than the general perceptual quality of the image. We provided some introductory notes in the web-interface before the study starts, about some examples and insights of judging the emotional aspect of an image similar to the procedure done in Sec: III.B.

For the purpose of comparison, we are not aware of any related affective enhancement tool, and thus we compared with a generic baseline i.e., GIMP auto-enhance function. GIMP's auto-enhance tool is geared for enhancing the perceptual quality of images and applies the same enhancement function to each image without considering the emotional aspect. This is unlike our approach of applying different functions for different emotional categories. Each enhanced image stimuli (i.e., image on the right side in Fig 11) is alternately rotated between our approach and GIMP auto-enhance in the web-interface during user feedback collection.

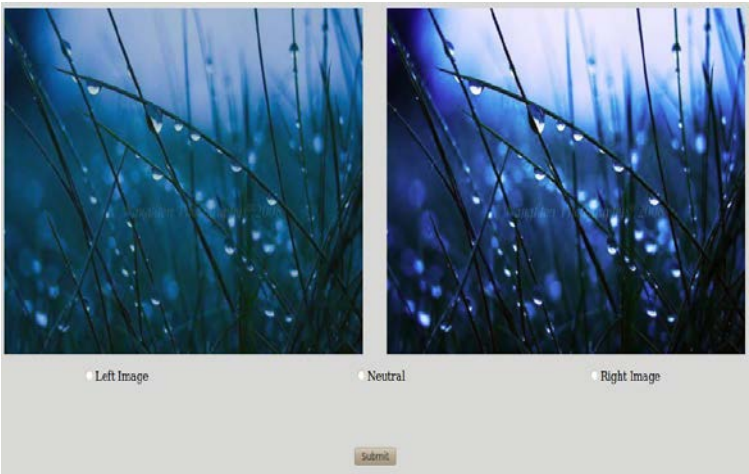


Fig.11 The web-based interface for user study which allows the subject to indicate its preference of affective enhancement between the original (left) and the adjusted (right) images.

The results are plotted in Fig 12. The plot depicts the various percentage ratios with respect to positive, negative, and neutral ratings as collected from user feedback for GIMP auto-enhance tool and our affective enhancement mechanism. The positive label indicate the users preference for the enhanced version of the pair-wise comparison (right button option in Fig 11), negative label indicate users preferred the original version (left button option in Fig 11), and neutral indicates no strong preference or rejection for the enhancement (middle button option in Fig 11). The plot also analyses the variation of performance for different clusters (considers the case for the number of clusters $K=4$.) and the final plot on the right end corresponds to the total combined results.

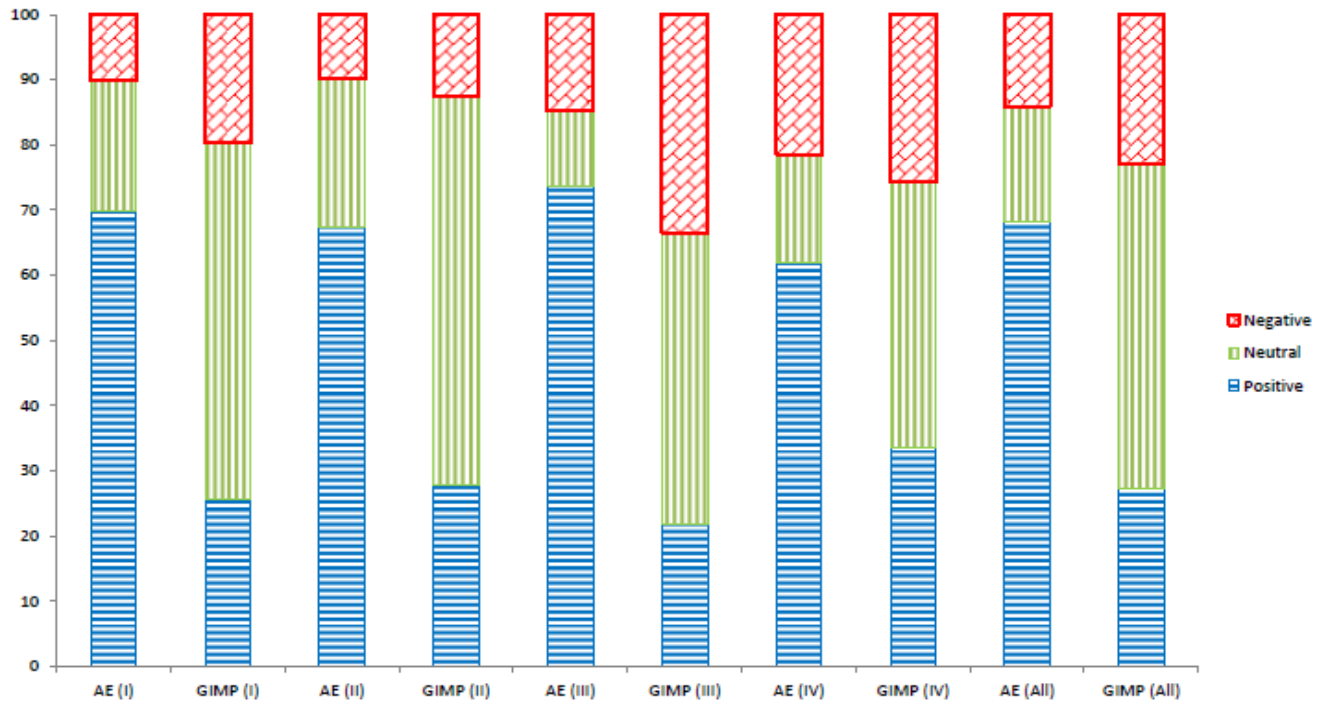


Fig. 12 The X-axis labels are associated with the two comparative methodologies i.e., AE for our Affective Enhancement and GIMP auto-enhance tool. The symbol within the parentheses indicates the cluster membership and the 'All' label considers the combined results. The Y-axis labels are the percentage ratio collected from user feedback for three different cases i.e., Positive, Neutral, and Negative

Some of the observations from the plot are as follows:

(1) The average case results for our affective enhancement are 68.15%-17.73%-14.13% for positive neutral-negative respectively, which is considerably higher than GIMP auto-enhance enhance with 27.23%-49.9%-22.88%. This can be justified by the fact that, GIMP auto-enhance is not designed for emotional adjustments, and only considers quality adaptation;

(2) Comparing the results cluster-wise, there is not much significant variation for our method with Cluster: 3 having a slight gain over the others with no apparent reason. This shows that our method performs relatively satisfactorily across the different clusters. GIMP has some variation among the cluster results, but generally it can be observed that a high bias is for the neutral case wherein the user did not have any strong preference or rejection for the emotional enhancement.

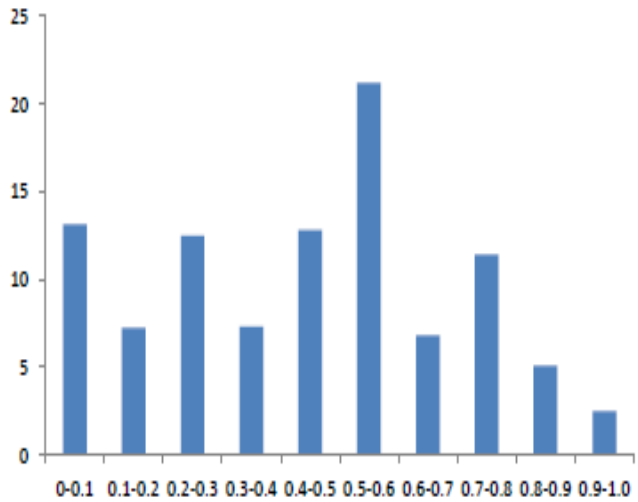


Fig. 13 Histogram plot depicting the distribution of enhancement degree using the objective function as discussed in Sec: VI.B. The X-axis shows the specified intervals objective metric values computed from Eq: 20 and the Y-axis defines the % ratio of the number of testing image samples within the respective histogram bins

B. Objective Evaluation

Deriving an objective metric for affective enhancement of images is a hard problem due to the inherent subjectivity of human emotion. In this section, we design an objective metric to quantify the degree of emotional enhancement in images strictly form image properties and without any ground-truth data. We exploit the color mood space conversion introduced in [33], [34], [35] to achieve our objective. The color mood space gives specific formulae to calculate mood scales from color appearance attributes such as luminance, hue, and chroma which makes it an exciting candidate for objective evaluation of affective enhancement. The mood space is a three coordinates axes of the space called as activity, weight, and heat. The methodology provides empirical formulations of the transformation from the CIELAB color space to the proposed color mood space.

Given a color $c=<L^*,a^*,b^*>$, its corresponding point $e=<a,w,h>$, the color mood space is a nonlinear function of c defined by the following equations:

$$a = -2.1 + 0.06\left[\left(L^* - 50\right)^2 + \left(a^* - 3\right)^2 + \left(\frac{b^* - 17}{1.4}\right)^2\right]^{\frac{1}{2}} \tag{17}$$

$$w = -1.8 + 0.04(100 - L^*) + 0.45 * \cos(h - 100^0) \tag{18}$$

$$h = -0.5 + 0.02(C^*)^{1.07} * \cos(h - 50^0) \tag{19}$$

where L^* is CIELAB lightness, C^* is CIELAB chroma, h is CIELAB hue angle, and a^*, b^* are CIELAB coordinates. On generalization of the additive relationship [35], we compute the color mood of an image by averaging the color of every pixel.

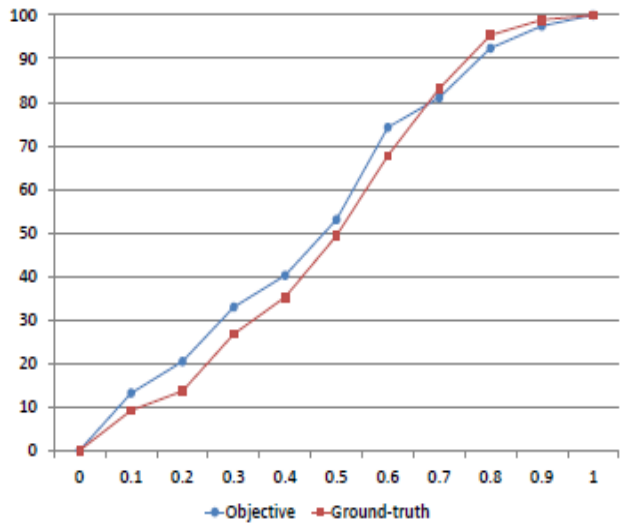


Fig. 14 CDF plot of the distribution of enhancement degree between the Ground-truth (methodology proposed in Sec: V.B for calculating the enhancement between the IAPS dataset ground-truth and the user collected response) and the objective technique described in Sec: VI.B.

Given, e computed from original image I_i, and e' computed from enhanced image I_i' using the above transformation function, we quantify the degree of enhancement as follows:

$$D_{(i)} = \left\| e' - e \right\| \tag{20}$$

where ||.|| is the Euclidean distance and at the end we normalize the values to (0,1). Fig 13 plots the distribution of affective enhancement using Eq: 20 on the same testing sample set employed in Sec: VI.A. We cannot directly compare the objective metric with the subjective results from Sec: VI.A due to the difference in scales. To find the degree of correlation between the subjective ratings and the objective values, we analyzed the test samples and found the following. By classifying the objective scale as: 0-0.1 (negative), 0.1-0.3 (neutral), and 0.3-1.0 (positive), the percentage ratios computed are: 13.7%, 19.2%, and 67.1% respectively, which correlates well with the subjective percentage ratios (9.15%, 12.75%, and 78.1% respectively) from the "AE-All" column in Fig 12. The Pearson's correlation index between the subjective and objective results for the test samples is found to be 0.8125, which indicates that the objective metric of Eq: 20 is a good indicator of affective enhancement in images.

Next, we compare the correlation between the computed objective metric to the method used in Sec: V.B for calculating the enhancement between the IAPS dataset ground-truth and the user collected response from our web-interface for samples in the training database. The CDF plot is shown in Fig 14. It can be observed from the plot that both the techniques have a reasonable correlation between them, based on the computed metric of affective enhancement. We can also observe that, based on the

analysis discussed, we can say that a user-rating with higher valence and higher arousal [V^+, A^+] (as described in Sec: V.B from the user feedback for training database) translates to a high value of $D_{(i)}$, one higher and the other lower [V^+, A^-]/[V^-, A^+] corresponds to moderate values of $D_{(i)}$, and lower valence-arousal [V^-, A^-], corresponds to very low values of $D_{(i)}$. So, our design in Sec: V.B for classifying the enhancement degree based on its directional orientation of $VA^u_{(j)} = \langle V^u_{(j)}, A^u_{(j)} \rangle$ with respect to the original image IAPS ratings $VA^{gt}_{(j)} = \langle V^{gt}_{(j)}, A^{gt}_{(j)} \rangle$ is a fair choice to make.

C. Analysis by Examples

In this section, we observe several image examples with their corresponding enhanced versions as produced by our affective enhancement function for the four different clusters. Clusters: I, II, III, IV are shown in Fig's 15, 16, 17, 18, respectively. It can be observed that the same enhancement function is applied for all images in the same cluster, thereby the enhancement color and contrast ratios are different across different clusters but remain the same within the cluster. The top row of two image pairs are rated as “positive” by a majority of users, bottom-left pair as “neutral”, and bottom-right pair as “negative”.

VII. CONCLUSION

We presented an affect-based image enhancement methodology which can adjust an input image based on the emotional characteristics of the underlying relationship. The first step involves a training phase which essentially builds a database comprising of a set of images and their corresponding metadata which are rating values. The input database images are fed into our enhancement processing framework to derive the corresponding output images which are presented to human participants for rating metadata using a web-based interface. The enhancement processing framework is a set of color-tonal image operations denoted by a vector of enhancement features. Once the training database is prepared with sufficient samples, we analyze it to learn a statistical model that encodes the relationship between the user-provided ratings and the enhancement vector. Specifically, we cluster the set of original images in a number of different groups and then estimate the enhancement vectors of each group using a regression-based learning approach followed by an optimization solver. Once such a statistical model is learned, we can enhance a new or unobserved image by mapping the image to one of the trained clusters and applying the enhancement operator for that cluster. As per our knowledge, this is one of the first attempts to derive a computational framework for enhancing the emotional impact of images. We collected ground-truth data for affective enhancement from human participants which is a valuable resource for understanding the underlying relationship between them. We employed a data-driven systematic framework to learn models from training data and derive generalized enhancement functions for arbitrary/unseen images using machine learning and statistical techniques. We derived an objective metric for evaluating the emotional enhancement of images using a color mood space model and used it to test our approach.

Fig. 15 A block of image pairs enhanced with our proposed affective methodology. For each pair, the left image is the original one and the right image is enhanced one. The results are taken from user study of test samples as discussed in Sec: VI.A for Cluster: I. The image pairs on the top row are rated by users as ``positive'' samples, image pair on the bottom-left is marked as ``neutral'', and image pair on the bottom-right is marked as ``negative''.

Fig.16 A block of image pairs enhanced with our proposed affective methodology. For each pair, the left image is the original one and the right image is enhanced one. The results are taken from user study of test samples as discussed in Sec: VI.A for Cluster: II. The image pairs on the top row are rated by users as ``positive'' samples, image pair on the bottom-left is marked as ``neutral'', and image pair on the bottom-right is marked as ``negative''

Fig.17 A block of image pairs enhanced with our proposed affective methodology. For each pair, the left image is the original one and the right image is enhanced one. The results are taken from user study of test samples as discussed in Sec: VI.A for Cluster: III. The image pairs on the top row are rated by users as ``positive'' samples, image pair on the bottom-left is marked as ``neutral'', and image pair on the bottom-right is marked as ``negative''

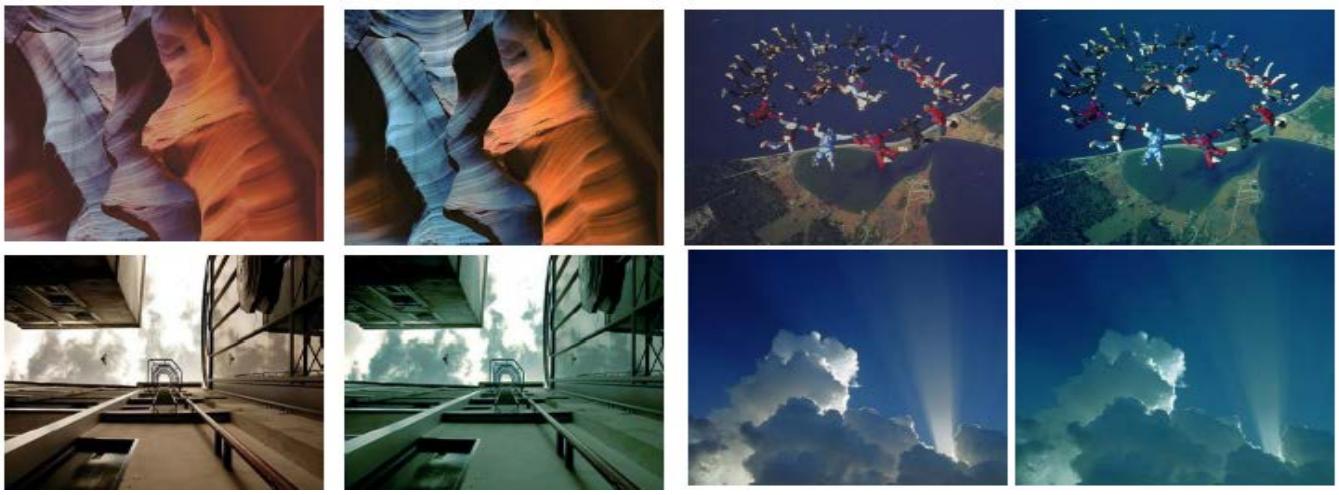


Fig.18 A block of image pairs enhanced with our proposed affective methodology. For each pair, the left image is the original one and the right image is enhanced one. The results are taken from user study of test samples as discussed in Sec: VI.A for Cluster: IV. The image pairs on the top row are rated by users as ``positive'' samples, image pair on the bottom-left is marked as ``neutral'', and image pair on the bottom-right is marked as ``negative''

REFERENCES

[1] Z. ur Rahman, D. J. Jobson, and G. A. Woodell, “Retinex processing for automatic image enhancement,” Human Vision and Electronic Imaging VII, vol. 4662, no. 1, pp. 390–401, 2002. [Online]. Available: http://link.aip.org/link/?PSI/4662/390/1

[2] A. Majumder and S. Irani, “Perception-based contrast enhancement of images,” ACM Trans. Appl. Percept.,vol. 4, no. 3, Nov. 2007. [Online]. Available: http://doi.acm.org/10.1145/1278387.1278391

[3] J. Mukherjee and S. Mitra, “Enhancement of color images by scaling the dct coefficients,” Image Processing, IEEE Transactions on, vol. 17, no. 10, pp. 1783 –1794, oct. 2008.

[4] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Wang, J. Li, and J. Luo, “Aesthetics and emotions in images,” Signal Processing Magazine, IEEE, vol. 28, no. 5, pp. 94 –115, sept. 2011.

[5] A. Valdez, P.and Mehrabian, “Effects of color on emotions,” Journal of Experimental Psychology: General, 123(4), pp. 394–409, 1994.

[6] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” Systems, Man and Cybernetics, IEEE Transactions on, vol. 8, no. 6, pp. 460 –473, june 1978.

[7] E. Phelps, S. Ling, and M. Carrasco, “Emotion facilitates perception and potentiates the perceptual benefits of attention.” Psychological Science,17 (4), pp. 292–299, 2006.

[8] J. Machajdik and A. Hanbury, “Affective image classification using features inspired by psychology and art theory,” in Proceedings of the international conference on Multimedia, ser. MM’10. New York, NY, USA: ACM, 2010, pp. 83–92. [Online]. Available: http://doi.acm.org/10.1145/1873951.1873965

[9] J. Lee and E. Park, “Fuzzy similarity-based emotional classification of color images,” Multimedia, IEEE Transactions on, vol. 13, no. 5, pp. 1031 –1039, oct. 2011.

[10] X. Xiang and M. S. Kankanhalli, “Affect-based adaptive presentation of home videos,” in Proceedings of the 19th ACM international conference on Multimedia, ser. MM ’11. New York, NY, USA: ACM, 2011, pp. 553–562.[Online].Available:http://doi.acm.org/10.1145/2072298.2072370

[11] H. Joho, J. M. Jose, R. Valenti, and N. Sebe, “Exploiting facial expressions for affective video summarisation,” in Proceedings of the ACM International Conference on Image and Video Retrieval, ser. CIVR ’09. New York, NY, USA: ACM, 2009, pp. 31:1–31:8. [Online]. Available: http://doi.acm.org/10.1145/1646396.1646435

[12] S. Zhang, Q. Tian, Q. Huang, W. Gao, and S. Li, “Utilizing affective analysis for efficient movie browsing,” in Image Processing (ICIP), 2009 16th IEEE International Conference on, nov. 2009, pp.1853 –1856.

[13] A. Hanjalic and L.-Q. Xu, “Affective video content representation and modeling,” Multimedia, IEEE Transactions on, vol. 7, no. 1, pp.143 – 154, feb. 2005.

[14] S. Zhang, Q. Huang, S. Jiang, W. Gao, and Q. Tian, “Affective visualization and retrieval for music video,” Multimedia, IEEE Transactions on, vol. 12, no. 6, pp. 510 –522, oct. 2010.

[15] R. Carroll, A. Agarwala, and M. Agrawala, “Image warps for artistic perspective manipulation,” ACM Trans. Graph., vol. 29, no. 4, pp. 127:1–127:9, Jul. 2010. [Online]. Available: http://doi.acm.org/10.1145/1778765.1778864

[16] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, “Parametric reshaping of human bodies in images,” ACM Trans. Graph., vol. 29, pp. 126:1–126:10, July 2010. [Online]. Available:http://doi.acm.org/10.1145/1778765.1778863

[17] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, “Datadriven enhancement of facial attractiveness,” ACM Trans. Graph., vol. 27, no. 3, pp. 38:1–38:9, Aug. 2008. [Online].Available:http://doi.acm.org/10.1145/1360612.1360637

[18] K. Sunkavalli, M. K. Johnson, W. Matusik, and H. Pfister, “Multi-scale image harmonization,” in ACM SIGGRAPH 2010 papers, ser. SIGGRAPH ’10. New York, NY, USA: ACM, 2010, pp. 125:1–125:10. [Online].Availablehttp://doi.acm.org/10.1145/1833349.1778862

[19] S. Bae, S. Paris, and F. Durand, “Two-scale tone management for photographic look,” in ACM SIGGRAPH 2006 Papers, ser. SIGGRAPH ’06. New York, NY, USA: ACM, 2006, pp. 637–645. [Online]. Available: http://doi.acm.org/10.1145/1179352.1141935

[20] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu, “Color harmonization,” ACM Trans. Graph., vol. 25, no. 3, pp. 624–630, Jul. 2006. [Online]. Available: http://doi.acm.org/10.1145/1141911.1141933

[21] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, “Light mixture estimation for spatially varying white balance,” ACM Trans. Graph., vol. 27, no. 3, pp. 70:1–70:7, Aug. 2008.[Online].Available:http://doi.acm.org/10.1145/1360612.1360669

[22] N. Joshi, W. Matusik, E. H. Adelson, and D. J. Kriegman, “Personal photo enhancement using example images,” ACM Trans. Graph., vol. 29, no. 2, pp. 12:1–12:15, Apr. 2010.[Online].Available:http://doi.acm.org/10.1145/1731047.1731050

[23] B. Wang, Y. Yu, T.-T. Wong, C. Chen, and Y.-Q. Xu, “Datadriven image color theme enhancement,” ACM Trans. Graph.,vol. 29, no. 6, pp. 146:1–146:10, Dec. 2010.[Online].Available:http://doi.acm.org/10.1145/1882261.1866172

[24] B. Wang, Y. Yu, and Y.-Q. Xu, “Example-based image color and tone style enhancement,” ACM Trans. Graph., vol. 30, no. 4, pp. 64:1–64:12, Jul. 2011. [Online]. Available: http://doi.acm.org/10.1145/2010324.1964959

[25] S. B. Kang, A. Kapoor, and D. Lischinski, “Personalization of image enhancement,” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, june 2010, pp. 1799 –1806.

[26] J. Caicedo, A. Kapoor, and S. B. Kang, “Collaborative personalization of image enhancement,” in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, june 2011, pp. 249–256.

[27] K. Dale, M. Johnson, K. Sunkavalli, W. Matusik, and H. Pfister, “Image restoration using online photo collections,” in Computer Vision, 2009 IEEE 12th International Conference on, 29 2009-oct. 2 2009, pp. 2217 –2224.

[28] C. O. Ancuti, C. Ancuti, and P. Bekaert, “Enhancing by saliency-guided decolorization,” in Proceedings of the 2011 IEEE Conference

on Computer Vision and Pattern Recognition, ser. CVPR '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 257–264. [Online]. Available:http://dx.doi.org/10.1109/CVPR.2011.5995414

[29] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," ACM Trans. Graph., vol. 30, no. 4, pp. 70:1–70:10, Jul. 2011. [Online].Available:http://doi.acm.org/10.1145/2010324.1964965

[30] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in Proceedings of the 9th European conference on Computer Vision -Volume Part III, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag,2006, pp. 288–301.

[31] 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, FL, Tech.Rep., 2008.

[32] M. Grant and S. Boyd, "Matlab software for disciplined convex programming." http://stanford.edu/ boyd/cvx/, 2008.

[33] L.-C. Ou, M. R. Luo, A. Woodcock, and A. Wright, "A study of colour emotion and colour preference. part i: Colour emotions for single colours," Color Research Applications,vol. 29, no. 3, pp. 232–240, 2004. [Online]. Available:http://dx.doi.org/10.1002/col.20010

[34] L.-C. Ou, M. R. Luo, A. Woodcock, and A. Wright, "A study of colour emotion and colour preference. part ii: Colour emotions for two-colour combinations," Color Research Application, vol. 29, no. 4, pp. 292–298, 2004. [Online]. Available:http://doi.wiley.com/10.1002/col.20024

[35] L.-C. Ou, M. R. Luo, A. Woodcock, and A. Wright, "A study of colour emotion and colour preference. part iii: Colour preference modeling," Color Research Applications, vol. 29, no. 5, pp. 381–389, 2004. [Online]. Available: http://dx.doi.org/10.1002/col.20047

[36] ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures.

Abhishek Bhattacharya has completed a PhD in Computer Science from Florida International University (FIU) with a dissertation titled "Affect-based Modeling and it Application in Multimedia Analysis Problems". His scientific research interests are: media analytics with affective and social signal based modeling using machine learning and statistical techniques, P2P networks, content distribution/streaming, distributed systems, data management, processing, mining and analytics. Specific details can be found on his academic page located at http://users.cis.fiu.edu/~abhat002/ where all his conference and journal articles are listed. He also had 5 years of industry experience prior to joining the FIU PhD program, and is currently a Senior Staff Engineer at IBM Systems and Technology Group, Bangalore, India.

Christine Lisetti is a faculty at Florida International University in the School of Computing and Information Sciences, where she directs the Affective Social Computing Laboratory (see http://ascl.cis.fiu.edu/). She is one of the founders of Affective Computing, a new field of computer science which lies at the intersection of Artificial Intelligence (AI) and Human-Computer Interaction (HCI) in Computer Science, of emotion and personality theories in Psychology, and of social interaction in Communication. She has been a member of the founding Editorial Board of the IEEE Transactions on Affective Computing since its launch in 2010. Her research interests include building intelligent systems that can recognize and adapt to people's emotions and personality, and understand their affective experiences and perceptions.

Zhenyu Yang have research interests include networking and distributed systems with a focus on communication and multimedia systems involving quality-of-service management, overlay and peer-to-peer networks, content delivery network, wireless networks, 3D collaborative environments, ubiquitous computing, human-computer interface/interaction, and development of multimedia applications particularly to promote collaborative and interdisciplinary research (e.g., art, education and medical informatics).