Wang XH, Jia J, Liao HY *et al.* Affective image colorization. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 27(6): 1119–1128 Nov. 2012. DOI 10.1007/s11390-012-1290-4

Affective Image Colorization

Xiao-Hui Wang[1] (王晓慧), Jia Jia[1] (贾 珈), Han-Yu Liao[2] (廖捍宇), and Lian-Hong Cai[1] (蔡莲红)

[1] *Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*
[2] *Academy of Art & Design, Tsinghua University, Beijing 100084, China*

E-mail: wangxh09@mails.tsinghua.edu.cn; jjia@tsinghua.edu.cn; liaoliao1992@gmail.com; clh-dcs@tsinghua.edu.cn

Received September 5, 2012; revised September 17, 2012.

Abstract Colorization of gray-scale images has attracted many attentions for a long time. An important role of image color is the conveyer of emotions (through color themes). The colorization with an undesired color theme is less useful, even it is semantically correct. However this has been rarely considered. Automatic colorization respecting both the semantics and the emotions is undoubtedly a challenge. In this paper, we propose a complete system for affective image colorization. We only need the user to assist object segmentation along with text labels and an affective word. First, the text labels along with other object characters are jointly used to filter the internet images to give each object a set of semantically correct reference images. Second, we select a set of color themes according to the affective word based on art theories. With these themes, a generic algorithm is used to select the best reference for each object, balancing various requirements. Finally, we propose a hybrid texture synthesis approach for colorization. To the best of our knowledge, it is the first system which is able to efficiently colorize a gray-scale image semantically by an emotionally controllable fashion. Our experiments show the effectiveness of our system, especially the benefit compared with the previous Markov random field (MRF) based method.

Keywords image colorization, affective word, color theme

1 Introduction

Color can enhance the expressiveness of an image. A wonderful colorization not only gives a gray-scale image good visual sense, but also endows it with much richer semantic meaning. The traditional interaction-based colorization method needs users to manually specify scribbles and their colors[1]. To reduce this manual labor, some studies focus on example-based colorization which uses an existing color image for colorization[2-5]. However, these methods need a reliable reference image with both similar content and the same style for transfer. Sometimes choosing such a reference image is not an easy task. To avoid this problem, Chia *et al.* introduced a nice system to semantically colorize an image recently[6]. Using a semantic label, it automatically selects the most suitable references from the Internet. This approach provides a more friendly interface for non-experienced users, as it does not need to manually choose proper references.

The image color is the main conveyer of emotions through color themes (templates of colors), which is demonstrated by various psychological studies[7-9]. Images with the same content but different color themes may have totally different emotions. So a proper colorization can also give images much richer emotions. And the colorization with an undesired color theme is less useful, even if it is semantically correct. However, all the above methods do not consider the emotional aspect of colorization. Although semantically correct results may be produced by utilizing the internet images[6], these results could not be affective enough, especially when the user wants a precise control on the target emotion. Semantic and richly affective colorization results can much better express the artistic conception and greatly improve the visual quality.

This paper is the extension version of [10]. In this paper, we propose a novel framework to affectively colorize a gray-scale image with consideration of the semantics. The input is a gray-scale image and an affective word. Based on the art theories, we adopt the *image-scale* space to relate affective words and color themes. For each object in the input image, a set of the internet images is downloaded and filtered. And each object

Regular Paper

This work is supported by the National Basic Research 973 Program of China under Grant No. 2011CB302201, the National Natural Science Foundation of China under Grant Nos. 61003094, 60931160443. This work is also funded by Tsinghua National Laboratory for Information Science and Technology (TNList) Cross-Discipline Foundation of China, and supported by the Innovation Fund of Tsinghua-Tencent Joint Laboratory of China.

*The preliminary version of the paper was published in the Proceedings of the 2012 Computational Visual Media Conference.

is assigned a suitable reference using a generic algorithm. We also propose a patch match based approach for object level colorization. The contributions can be summarized as follows:

- *Affective Colorization System.* We develop an affective colorization system to automatically colorize a gray-scale image, which could meet the specific emotion. As far as we know, it is the first colorization system which considers both the semantics and the affective aspect.
- *Reference Selection Guided by the Affective Word.* We propose a generic algorithm based selection method to select the most suitable reference for each object of the gray-scale image. This method can efficiently select semantically correct references which also meet the desired emotion. We also design an optional manual tuning step for better personalizing the colorization.
- *Image Colorization Algorithm.* We propose a new hybrid texture synthesis based colorization method. Experiments show that compared with the Markov random field (MRF) based approach, the colorization can produce better meaningful results.

2 Related Work

We have reviewed the work related to our colorization framework, including colorization and color composition in the art theories.

2.1 Colorization

Colorization methods can be divided into two categories — interaction-based methods and example-based methods. This interaction-based colorization can trace its history back to the early 1970s, and was first used to colorize old films. In the early industrial work, after images were segmented into regions, the artists needed to manually specify a color for each region. Though the interaction-based method successfully dealt with a lot of old pictures and films, the labor-intensity of these methods seriously restricted their popularity. To reduce the labor requirement, Levin *et al.*[1] proposed a scribble-based method. The user just needs to give some colored scribbles on the gray-scale image, and colors are automatically propagated to the remain pixels based on local similarities. The process is similar to the edit propagation[11-12]. This method was later extended by Huang *et al.*[13] to reduce the color blurring effect. However, even with these improvements, users still need considerable interactions, and it is hard for non-professionals to select proper colors and draw approximate scribbles.

Example-based colorization methods usually do not require user interactions. Using some color images, these methods automatically colorize a given gray-scale image. Such methods are closely related to color transfer[14-15] which aims at changing colors of existing images. Reinhard *et al.*[16] proposed a statistics based color transfer method, focusing on global color replacement. Although it does not directly handle colorization, the idea was soon extended to gray-scale images[2]. Recent methods borrow ideas from machine learning to predict correspondence between reference color images and the input gray-scale image. A local color transfer method via EM (expection maximization)-based probabilistic segmentation is proposed by Tai *et al.*[5]. Charpiat *et al.*[4] proposed a global optimization method to predict the possible color distribution for each individual pixel. All these methods require similarities between reference images and input images, but finding reference images with similar contents and appropriate appearance is usually a difficult task.

With the rapid development of the Internet, data driven processing is attracting more attentions than before[17-19]. To find suitable reference images, Chia *et al.*[6] proposed to filter internet images, sharing a similar framework as Sketch2Photo[20]. This method is a nice supplementary for the example-based colorization, and can also be used as a preprocessing step of these methods. While as mentioned in Section 1, the method ignores the emotional appearance aspect when filtering, which is the main consideration in this paper.

2.2 Color Composition in Art Theories

Color composition is the color distribution of an image and the key element for artistic feeling[7]. Artists often use a set of colors called *color theme* to represent the color composition. The most commonly used color themes are 3-color themes and 5-color themes. The rational study of color themes is a hot topic in computer vision and graphics recently[21]. Daniel Cohen-Or *et al.* applied the existing aesthetical color harmony models to harmonize images[22]. Peter O'Donovan *et al.* studied color compatibility of 5-color themes from large datasets[23]. However, the emotional aspect of color themes is often ignored. For example, pink tint often represents romance, and dark color is often associated with solemn feelings.

Kobayashi systematically studied the relationship between color themes and emotions based on the psychophysical investigations[8-9], which had already been successfully used in graphic design. In [8], Kobayashi mapped 1 170 3-color themes to 180 affective words, like romantic, elegant. Furthermore, the relation between 4 905-color themes and 180 affective words is built in [9]. An affective space, called *Color Image Scale*, is used in [8-9] to quantitatively describe the emotions.

The space has two dimensions — warm-cool and hard-soft. Fig.1 illustrates a few examples of color themes and the corresponding affective words in this space.

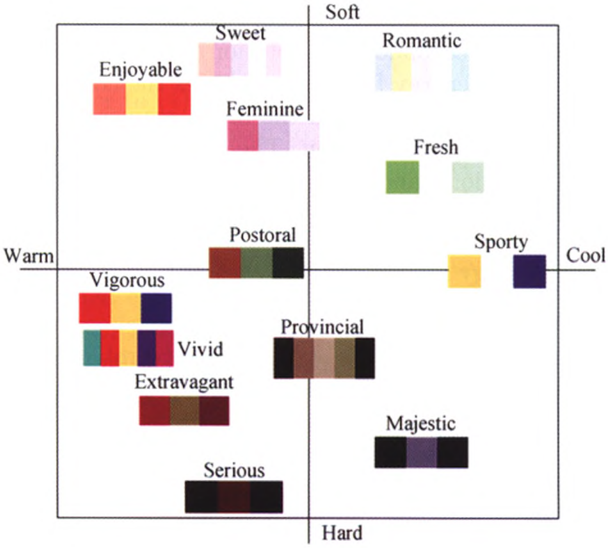

Fig.1. Examples of color themes and corresponding affective words in the image-scale space.

3 Overview

The overall pipeline of our system is illustrated in Fig.2. The input is a gray-scale image and an affective word to express the emotion, such as romantic, serious. The gray-scale image (Fig.2(a)) is first semi-automatically segmented into objects by a graphcut-based segmentation technique[24-26]. The user gives each object a text label by which we download and filter images from the Internet, and each object is given a set of candidate references as shown in Fig.2(b). According to the affective word, color themes are selected from a database built from on-line communities using the image-scale space as shown in Fig.2(c). To select the best reference for each object, we design a hybrid energy function to balance various requirements mainly including the similarity between the references and the input objects, as well as the conformity of the references with the candidate color themes. And a generic algorithm is adopted to optimize the energy (Fig.2(d)). Finally, we use a patch match based approach (a hybrid texture synthesis) for object-level colorization (Fig.2(e)).

4 Reference Selection

4.1 Object Filtering

Generally speaking, internet image search is unreliable, as the labels of some crawled images may be irrelevant to the image content. We do not solve this challenging problem, but adopt the shape filtering and the candidate selection step in Subsection 4.3 to get the most suitable reference. The shape information of the input object is crucial for shape filtering, but automatic image segmentation is sometimes not robust enough for certain images. So we rely on a semi-automatic approach to segment the input gray-scale image into multiple objects. The user roughly gives strokes to specify the objects, and a graphcut-based segmentation[24-26] is used for segmentation. As the segmentation is user assisted, it is generally suitable for any input gray-scale image. Then each object is manually labeled by a word for internet image search.

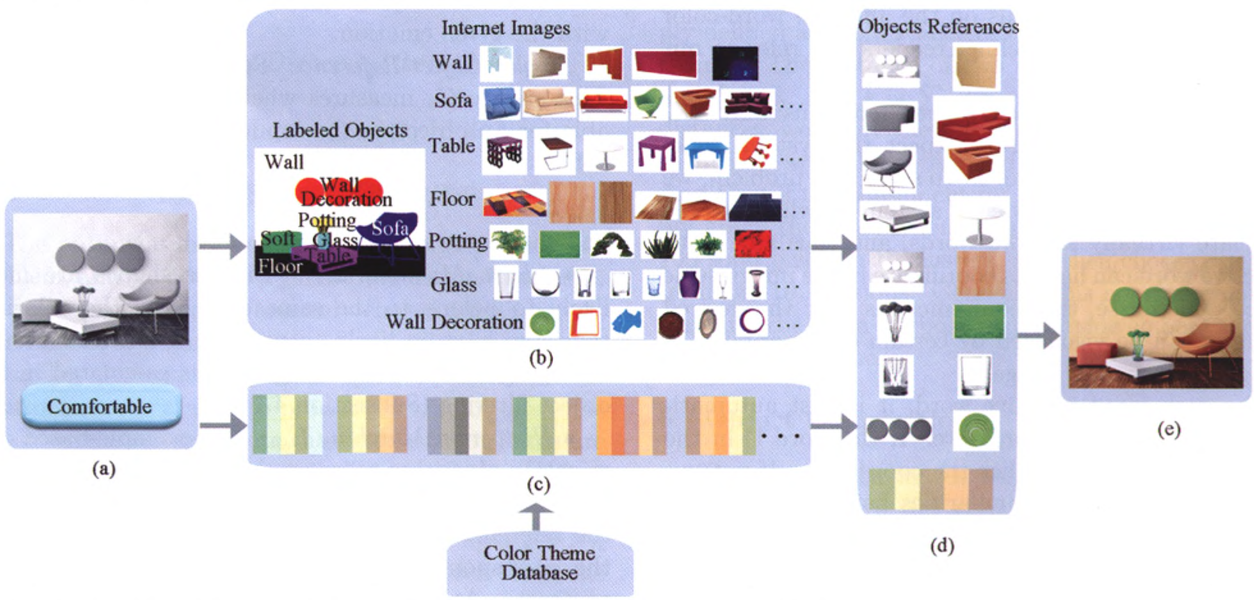

Fig.2. Pipeline[10]. (a) Input. (b) Object filtering. (c) Color theme selection. (d) Generic algorithm based reference selection. (e) Image colorization.

We first download a set of pictures (about 500~1 000) from the Internet such as Google Image Search and Flickr with each label. The salient region is automatically extracted for each downloaded image using the global contrast based salient region detection[27]. Then using the contour consistency filtering in [20], we select the images whose salient regions are similar to the outer contour of the gray-scale image object with shape context descriptors[28]. In addition, we allow for the shape deformation. The shape context matching cost and the affine registration are summed up to get the overall score. It is used to rank the extracted salient regions, and the top $50 \sim 100$ objects (salient regions in images, also called references or reference objects) are retained for further reference selection (Subsection 4.3).

4.2 Color Theme Selection

Color Theme Database Construction. We construct a color theme database with about 400 000 color themes. Each color theme t_i is labeled with a coordinate $a_i = (wc_i, hs_i)$ in the image-scale space, where wc_i and hs_i are the values of the warm-cool axis and the hard-soft axis respectively. We model the relationship between color themes and coordinates using the LASSO regression framework[23]. The training data is the mapping between 1 170 3-color themes and 180 affective words, and the relations between 4 905-color themes and these 180 words. All color themes and 180 words are labeled with the coordinates. The visual features of a color theme are colors in the HSV (hue, saturation, and value) color space and the color contrast. For more details on the affective word-color theme relationship modeling, readers can refer to our recent work[29].

Initially the image-scale space contains only 180 words. However, the users expect the system works for any input affective word. To solve this problem, we adopt the HowNet knowledge system[30] which is able to relate virtually any two words, and thus any input affective word can have its coordinate in the image-scale space. Of course, the requirement is that the desired emotion can be visualized by certain colors applied to the given gray-scale image.

For each input affective word, it is first automatically labeled with a coordinate. If the word is in the set of the 180 words, it has the coordinate. If not, we calculate the semantic similarities between it and the 180 words by the HowNet knowledge system[30]. The coordinate of the word is the weighted average value of M ($M = 5$) most similar words. N_T ($N_T = 100$ in our experiments) candidate color themes nearest to the coordinate are selected.

4.3 Generic Algorithm Based Reference Selection

A generic algorithm is adopted to select the most suitable reference for each object (we call a set of object-reference correspondences a solution). In the generic algorithm, the energy function measures the consistency between the references and the input gray-scale image as well as the given emotion. The system can also cluster the solutions, and users can easily select one to satisfy their own tastes.

4.3.1 Energy Function

Even with the above rough filtering, choosing the appropriate references is still a difficult problem. The problem can be formalized as follows: There is a set of objects $O = \{o_1, o_2, \dots o_n\}$, and each object o_i has a set of candidate references $R_i = \{r_{i,1}, r_{i,2}, \dots, r_{i,t_i}\}$, where t_i is the number of o_i's candidate references. In order to achieve an optimized solution $f : o_i \to R_i$, $i = 1, 2, \dots, n$, we design a comprehensive energy G measuring the suitability of the solution, and the goal is to calculate $\arg\min_f G(s)$.

The energy function considers a set of carefully planned elements, mainly relating to two important factors: consistency with the input gray-scale image and the given emotion. To formalize the energy, we rewrite it as

$$G(s) = G_s(s) + G_a(s), \tag{1}$$

where $G_s(s) = \sum_i E_s(o_i, r_i)$, $s = \{r_1, r_2, \dots, r_n\}$ is a solution, E_s is an energy measuring the suitability of a single object-reference pair, and $G_a(s)$ measures the affective suitability of a solution, that is the consistency with the given emotion.

Single Object-Reference Suitability Measurement. This energy G_s measures whether a reference is suitable for an object in the input gray-scale image. We define

$$E_s = \theta_1 E_{ss} + \theta_2 E_{sh} + \theta_3 E_{sc} \tag{2}$$

and each issue is normalized to $[0, 1]$.

• *Shape Matching E_{ss}.* Similar shapes are more likely to belong to the same class, and the colorization step is more likely to produce a semantic result. The shape context matching error is calculated in the object filtering (Subsection 4.1).

• *Histogram Matching E_{sh}.* The example-based colorization algorithm relies on the matching between the object in the input image and the gray-scale version of the reference (reference gray-scale). It is hard to find the correspondence for two images with strong light contrast. A better histogram matching can help to generate more uniform correspondence. This energy is defined as the distance between the two histograms.

• *Color Consistency Matching* E_{sc}. This item measures the consistency between the reference and the candidate color themes (determined by the affective word). To measure this, we calculate the conformity between the extracted color theme of a reference and every color theme in the candidates, and select the best as its energy. Refer to (4) for the definition of this energy.

θ_1, θ_2 and θ_3 are used to balance the above factors. In our experiments, θ_2, θ_3 are usually set to 1. If the object is semantic sensitive (e.g. orange, horse), θ_1 can be set to 1. On the other hand, for the object like sofa/furniture which is semantic insensitive, θ_1 can be set smaller, e.g., $[0.1, 0.2]$.

Affective Suitability Measurement. This energy measures the affective suitability, which is the consistency between the color theme of a solution and the candidate color themes. For a solution $\{r_1, r_2, \ldots, r_n\}$, we resize each reference object to the same scale as their corresponding objects $o_1, o_2, \ldots, o_n$. The color theme of a reference object is extracted by K-means. We then compare the extracted theme with the candidate color themes. For two color themes $theme_1 = \{c_1^1, c_2^1, \ldots, c_m^1\}$ and $theme_2 = \{c_1^2, c_2^2, \ldots, c_m^2\}$, where c_i^1, c_i^2 are colors in the HSV color space. The distance between two themes is defined as

$$D(theme_1, theme_2) = \min_{p \in P} \sum_{i=1}^{m} d(c_{p(i)}^1, c_i^2), \tag{3}$$

where P is the set of permutations of $1, 2, \ldots m$, and d is the Euclidean distance. Suppose the extracted theme is $theme_o$ and the candidate themes are $theme_1, theme_2, \ldots, theme_M$, $G_a(s)$ is defined as

$$G_a(s) = \mathscr{D} \min_{i=1}^{M} D(theme_o, theme_i), \tag{4}$$

where $\mathscr{D}$ is a constant to normalize G_a to $[0, 1]$.

4.3.2 Generic Algorithm Based Optimization

The optimization of energy G is a highly nonlinear problem, and we develop a generic algorithm to solve it. We start from a set of randomly chosen solutions. At each iteration, we select a set of "parent" solutions with possibility determined by their energies. These parent solutions are then used to produce new solutions. The details are as follows. Note that the last four items are components of an iteration.

• *Initialization.* We randomly choose N_I solutions.

• *Selection of Parent Solutions.* We sample $2 \times N_R$ solutions (N_R pairs) with replacement. Each time, the possibility of a solution s is proportional to $e^{-\alpha G(s)}$, where $G(s)$ is the energy defined by (1) in Subsection 4.3.1, and α is usually set to 1.

• *Reproduction of New Solutions.* For a pair of solutions $r_1, r_2, \ldots r_n$ and $r'_1, r'_2, \ldots r'_n$, we generate a new solution $r''_1, r''_2, \ldots r''_n$, where $r''_i = r_i$ or $r''_i = r'_i$ with possibility $p = \frac{\mathcal{P}}{\mathcal{P}+\mathcal{P}'}$ and $p' = 1 - p$ respectively. $\mathcal{P} = e^{-\beta G_s(r_i)}, \mathcal{P}' = e^{-\beta G_s(r'_i)}$, where β is usually set to 1. The new solutions and the original N_I solutions form a set of $N_I + N_R$ solutions.

• *Variation.* We randomly choose 10% of the $N_I + N_R$ solutions to undergo mutations. For each of these solutions, we randomly reassign 1/3 of its objects' references.

• *Generation of a New Group.* We sort the energies of the $N_I + N_R$ solutions and cut off N_R solutions with the biggest energies to keep the size.

In our experiments, we simply set $N_I = 5\,000$ and $N_R = 1\,000$. In cases when the total number of solutions is small, N_I and N_R can be smaller. N_S (100 in our experiments) best solutions are finally selected when the algorithm finishes.

4.3.3 Manual Selection Tuning

An emotion can be represented by kinds of color themes, and different users prefer different ones. Choosing the solution with the lowest energy from previous results usually works, and this step gives users flexibility to control the final results for personalization requirements. The generic algorithm based optimization gives a set of candidates, each of which ensures reliable selection. Suppose the N_S solutions are $t_1, t_2, \ldots t_{N_S}$. We realize that exploring in a large solution set is a boring task, especially when most of them are similar. For two references r, r', their distance $d(r, r')$ is simply defined as the distance of their corresponding color themes. The distance of two solutions $\{r_1, r_2, \ldots r_n\}$, and $\{r'_1, r'_2, \ldots r'_n\}$ is $\sum_i d(r_i, r'_i)$. Using the pairwise distances, we apply the spectral clustering[31] to divide them into several clusters and display cluster centers. And then if the chosen cluster is large enough, we simply apply the same scheme. In practice, we cluster the 100 solutions into 10 clusters, and then choose directly from the 10 solutions. Generally this is an alternative step and the main purpose is to provide elaborate control of the pipeline.

5 Object Colorization

We use a patch match based colorization method which can improve the visual quality in our experiments. The patch match based correspondence finding algorithm is shown to be a success in image editing[32], and has been integrated into some commercial softwares such as Photoshop CS5. Because of its controllability and randomized characteristics, patch match based methods are proved to be more sufficient than the traditional MRF-based approaches[6].

In order to better understand our colorization method, we first introduce some basic concepts. The example-based colorization is essentially a correspondence finding problem. That is, given a gray-scale image A and its reference color image B, our task is to find a function $U : A \rightarrow B$, and then each point in A can take the corresponding color in B. For convenience, here we use the term "image" to refer to previous "object". Though an "object" has an irregular shape rather than a rectangle, the coloring step shares nearly the same operation flow.

Our colorization can be considered as a process of gray-scale guided texture synthesis, and can also be thought as a hybrid of image correspondence finding and classical texture synthesis. To measure the quality of a colorization, we conceptually minimize the following energy function:

$$E = \sum_{i \in A} (e_1(i, U(i)) + e_2(i, U(i))), \tag{5}$$

where i is the pixel iterating over image A, and e_1 measures the consistency between two local patches with centers i and $U(i)$ respectively, noting that B needs to be grayed for the gray comparison. This term is mainly used in the image correspondence finding algorithm. While the first term considers the gray-scale correspondence, the second term e_2 measures the quality of the synthesized colors, which is also the main energy term in the texture synthesis.

5.1 Initialization

As in [32], we simply use a randomized initialization. For each point i in image A, we randomly select a point u_i in image B, that is $U(i) = u_i$.

5.2 Iterative Correspondence Refinement

The optimization of E is essentially a labeling problem, and is thus NP-hard. Here we use an iterative process similar to PatchMatch, which is also commonly used in texture synthesis algorithms. In each iteration, we consider to update each point i in image A in the scan-line order. We consider the correspondence set S from point i, point i's neighbors and points chosen randomly, and select the one with minimal error:

$$U(i) = \min_{u \in S} e_1(i, u) + \alpha e_2(i, u). \tag{6}$$

Note that we increase e_2 gradually, as it is not reliable in the first few iterations. In our experiments, we use six iterations and let α be $0, 0, 0, 0.15, 0.2, 0.2$ respectively.

6 Experiments

6.1 Performance and Results

We have implemented our method on a machine with two quad-core 2.26 GHZ CPUs. We apply 10 iterations in the generic algorithm based reference selection. Besides the possible user interactions in the segmentation/labeling/manual tuning step, and the image downloading step which heavily depends on the internet connections, the whole system takes less than 1.5 minutes for colorizing a $1\,000 \times 750$ image. We also manually setup a small knowledge base indicating some objects which do not have common shapes, such as floor, sky. For them, we do not extract the saliency regions and use the full images as references, especially the background.

We have validated the proposed method with various input examples. Fig.3 shows the main results. The first column shows four input gray-scale images as well as the segmentations and the text labels. We achieve two different colorization results through optionally giving two affective words. The color theme (the second column) is chosen from the color theme database according to the affective word, which is closest to the final composition. The colorization results (the third column) not only well conform to the given affective word, but also are semantic because the selected objects usually belong to the same class.

6.2 Evaluation on the Generic Algorithm Based Reference Selection

To investigate the effects of the generic algorithm in Subsection 4.3.2, we apply the method described in Subsection 4.3.3. Fig.4 shows such an example. The input gray-scale image, the affective word, and the labeled objects are the same as those in Fig.2. We cluster the resulting 100 solutions into 10 clusters, and the 10 centers as well as the references for each object are shown row by row. These solution clusters are quite different and the user can easily choose his/her favourite one. The user can also choose a solution in a finer level within a cluster. Fig.5 shows color themes for solutions in the first cluster.

6.3 Comparison with Other Colorization System

Compared with [6], given an affective word, we can achieve the results meeting the desired emotions, which are in line with the individuals' various emotional expectations. For example in Fig.6, the input is from Fig.9 in [6]. With our system, it is much easier to get two more results with distinct emotions, while [6] does

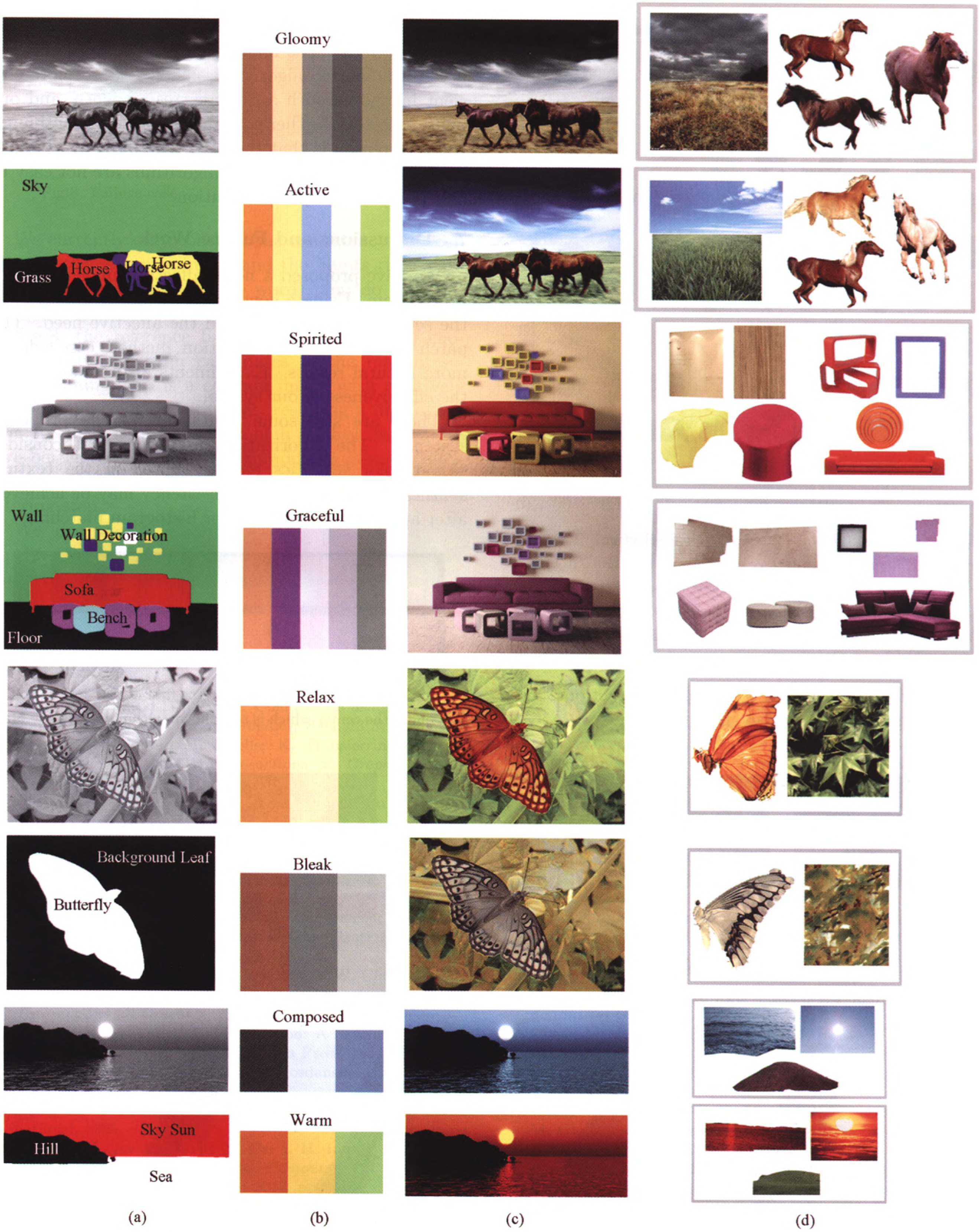

Fig.3. Colorization results. (a) Input images and labels. (b) Affective words and color themes. (c) Colorization results. (d) Reference images.

(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig.4. Generic algorithm based reference selection. The input gray-scale image, the affective word, and the labeled objects (shown in the first row) are the same as those in Fig.2. (a) Color themes. (b)~(i) are references of objects. (b) Sofa. (c) Sofa. (d) Table. (e) Wall. (f) Floor. (g) Wall decoration. (h) Potting. (i) Glass.

not directly support such convenient interactions. The various results in Fig.3 also show that.

We also give a comparison between the proposed colorization and the MRF-based colorization in [6], as shown in Fig.7. The main reason is that patch match based randomized algorithm is shown to be more efficient in finding such a quasi-correspondence, and our hybrid strategy further guarantees the color continuity. Another reason is that the MRF energy is highly non-linear, and the optimization algorithms are not easy to find the globally optimal solution.

7 Discussions and Future Work

We have proposed a novel affective image colorization system. Using a generic algorithm, it unifies both the semantic requirement and the affective need. The patch match based colorization method can achieve more natural results. Experiments also demonstrated the effectiveness of our system.

There are still some limitations of the proposed system. The colorization step does not consider global color statistics recently noted in the texture synthesis[33]. Better color consistency may be achieved after applying the global color histogram matching.

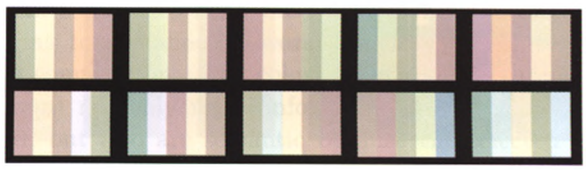

Fig.5. Color themes for solutions in the first cluster in Fig.5. The first one is the center which is also shown in the top of in Fig.5(a).

(a) (b) (c)

Fig.6. Comparison with [6]. (a) Input image. (b) Result of [6]. (c) Our results with different affective words.

(a) (b) (c) (d)

Fig.7. Comparison between our patch match based approach and the MRF-based approach. (a) Input gray-scale image. (b) Reference. (c) MRF-based colorization[6]. (d) Our result.

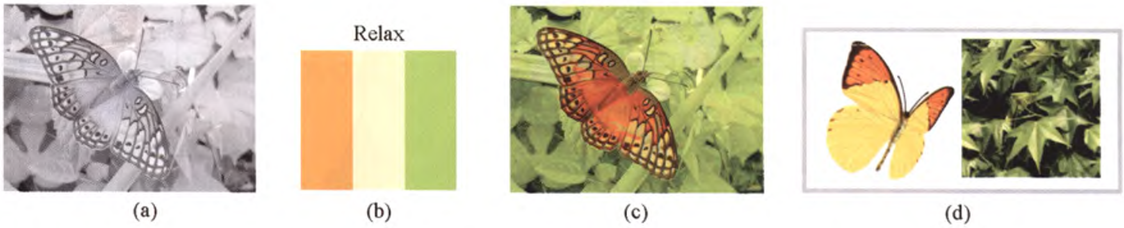

Fig.8. Failure case. (a) Input image. (b) Affective word and color theme. (c) Colorization result. (d) Reference images.

We currently only consider the consistency between the selected reference image and the inputs (the gray-scale image and the affective word), and do not consider the quality of the reference images. As shown in Fig.8, the reference butterfly image Fig.8(d) chosen by the user from candidates in the manual selection tuning step is a cartoon image with relatively monotonous color distributions, resulting a dull image (Fig.8(c)). It is obvious that high quality images can give better results. We plan to add the image quality measurement[34] as an important factor in the object filtering step.

We also plan to improve the various steps in our pipeline for better user experience. For example, we can investigate more image segmentation methods such as [35], and more sophisticated color-to-gray methods such as [36].

References

[1] Levin A, Lischinski D, Weiss Y. Colorization using optimization. *ACM Transactions on Graphics*, 2004, 23(3): 689-694.

[2] Welsh T, Ashikhmin M, Mueller K. Transferring color to greyscale images. *ACM Transactions on Graphics*, 2002, 21(3): 277-280.

[3] Ironi R, Cohen-Or D, Lischinski D. Colorization by example. In *Proc. the 16th Eurographics Workshop on Rendering Techniques*, June 29-July 1, 2005, pp.201-210.

[4] Charpiat G, Hofmann M, Schölkopf B. Automatic image colorization via multimodal predictions. In *Proc. the 10th ECCV*, October 2008, pp.126-139.

[5] Tai Y W, Jia J Y, Tang C K. Local color transfer via probabilistic segmentation by expectation-maximization. In *Proc. CVPR 2005*, June 2005, Vol.1, pp.747-754.

[6] Chia A Y S, Zhuo S J, Gupta R K, Tai Y W, Cho S Y, Tan P, Lin S. Semantic colorization with internet images. *ACM Transactions on Graphics*, 2011, 30(6), Article No.156.

[7] Arnheim R. Art and Visual Perception: A Psychology of the Creative Eye. University of California Press, 1954.

[8] Kobayashi S. Color Image Scale. Kosdansha International, 1992.

[9] Kobayashi S. Art of Color Combinations. Kosdansha International, 1995.

[10] Wang X H, Jia J, Liao H Y, Cai L H. Image colorization with an affective word. In *Proc. Computational Visual Media Conference 2012*, November 2012, pp.51-58.

[11] Xu K, Li Y, Ju T, Hu S M, Liu T Q. Efficient affinity-based edit propagation using K-D tree. *ACM Transactions on Graphics*, 2009, 28(5), Article No.118.

[12] Li Y, Ju T, Hu S M. Instant propagation of sparse edits on images and videos. *Computer Graphics Forum*, 2010, 29(7): 2049-2054.

[13] Huang Y C, Tung Y S, Chen J C, Wang S W, Wu J L. An adaptive edge detection based colorization algorithm and its applications. In *Proc. the 13th MULTIMEDIA*, November 2005, pp.351-354.

[14] Huang H, Zang Y, Li C F. Example-based painting guided by color features. *The Visual Computer*, 2010, 26(6/8): 933-942.

[15] Xiao X Z, Ma L Z. Gradient-preserving color transfer. *Computer Graphics Forum*, 2009, 28(6/8): 1879-1886.

[16] Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. *IEEE Computer Graphics and Applications*, 2001, 21(5): 34-41.

[17] Chen T, Tan P, Ma L Q, Cheng M M, Shamir A, Hu S M. PoseShop: Human image database construction and personalized content synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 2012, http://doi. ieeecomputersociety.org/10.1109/TVCG.2012.148, Sept. 2012.

[18] Huang H, Zhang L, Zhang H C. Arcimboldo-like collage using internet images. *ACM Transactions on Graphics*, 2011, 30(6), Article No.155.

[19] Liu H, Zhang L, Huang H. Web-image driven best views of 3D shapes. *The Visual Computer*, 2012, 28(3): 279-287.

[20] Chen T, Cheng M M, Tan P, Shamir A, Hu S M. Sketch2Photo: Internet image montage. *ACM Transactions on Graphics*, 2009, 28(5), Article No.124.

[21] Csurka G, Skaff S, Marchesotti L, Saunders C. Building look & feel concept models from color combinations: With applications in image dessification, retrieval, and color transfer. *The Visual Computer*, 2011, 27(12): 1039-1053.

[22] Cohen-Or D, Sorkine O, Gal R, Leyvand T, Xu Y Q. Color harmonization. *ACM Transactions on Graphics*, 2006, 25(3): 624-630.

[23] O'Donovan P, Agarwala A, Hertzmann A. Color compatibility from large datasets. *ACM Transactions on Graphics*, 2011, 30(4), Article No.63.

[24] Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 2006, 70(2): 109-131.

[25] Rother C, Kolmogorov V, Blake A. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 2004, 23(3): 309-314.

[26] Cheng M M, Zhang F L, Mitra N J, Huang X L, Hu S M. Repfinder: Finding approximately repeated scene elements for image editing. *ACM Transactions on Graphics*, 2010, 29(4), Article, No. 83.

[27] Cheng M M, Zhang G X, Mitra N J, Huang X L, Hu S M. Global contrast based salient region detection. In *Proc. CVPR 2011*, June 2011, pp.409-416.

[28] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(4): 509-522.

[29] Wang X H, Jia J, Cai L H. Affective image adjustment with a single word. To appear in *The Visual Computer*.

[30] Dong Z D, Dong Q. HowNet and the Computation of Meaning. World Scientific, 2006.

[31] Luxburg U. A tutorial on spectral clustering. *Statistics and Computing*, 2007, 17(4): 395-416.
[32] Barnes C, Shechtman E, Finkelstein A, Goldman D B. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 2009, 28(3), Article No.24.
[33] Chen J T, Wang B. Solid texture synthesis using position histogram matching. In *Proc. the 11th CAD/Graphics*, August 2009, pp.150-153.
[34] Sheikh H R, Bovik A C. Image information and visual quality. *IEEE Transactions on Image Processing*, 2006, 15(2): 430-444.
[35] Zhong F, Qin X Y, Peng Q S. Robust image segmentation against complex color distribution. *The Visual Computer*, 2011, 27(6-8): 707-716.
[36] Wu J L, Shen X Y, Liu L G. Interactive two-scale color-to-gray. *The Visual Computer*, 2012, 28(6-8): 723-731.

Xiao-Hui Wang is a Ph.D. candidate at the Department of Computer Science and Technology, Tsinghua University, China. She received her B.S. degree in computer science and technology in 2009 from Beijing Jiaotong University, China. Her research interests include image processing and pattern recognition.

Jia Jia is now an assistant professor at the Department of Computer Science and Technology, Tsinghua University, China. She received her Ph.D. degree in computer science and technology in 2008 from Tsinghua University, China. Her research interests are affective computing and computational speech perception.

Han-Yu Liao is a junior student of the Academy of Art & Design, Tsinghua University, China. She learns interaction design and information design.

Lian-Hong Cai is now a professor at the Department of Computer Science and Technology, Tsinghua University. She directs the Human-Computer Speech Interaction Laboratory. Her major research interests include human computer speech interactive, speech synthesis and text-to-speech (TTS), speech corpus, and multimedia technology.