Image Colorization with an Affective Word

Xiaohui Wang, Jia Jia, Hanyu Liao, and Lianhong Cai

Key Laboratory of Pervasive Computing, Ministry of Education
Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Computer Science and Technology, Tsinghua University
`wangxh09@mails.tsinghua.edu.cn, jjia@tsinghua.edu.cn,`
`liaoliao1992@gmail.com, clh-dcs@tsinghua.edu.cn`

Abstract. An important role of image color is the conveyer of emotions (through color themes). The colorization is less useful with an undesired color theme, even semantically correct, which has been rarely considered previously. In this paper, we propose a complete system for the image colorization with an affective word. We only need users to assist object segmentation along with text labels and give 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 adopted to select the best reference for each object. Finally, we propose a hybrid texture synthesis approach to colorize each object. Our experiments show that the results of our system have both the correct semantics and the desired emotions.

Keywords: Image colorization, affective word, color theme

1 Introduction

Color can enhance the rich expressive force of an image. A wonderful colorization not only gives a grayscale image good visual sense, but also endows it with much richer semantic meaning. The interaction based colorization method need users to manually specify scribbles and their colors [1]. To reduce this manual labor, some works 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 the similar contents 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 a proper reference.

The image color is the main conveyer of emotions through the color themes (templates of colors), which is convinced by various psychological studies [7–9]. Images with the same contents and the different color themes may have totally different emotions. So a proper colorization can also give the 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.

In this paper, we propose a novel framework to affectively colorize a grayscale image with consideration of the semantics at the same time. As far as we know, it is the first colorization system which considers both the semantics and the affective aspect. The input is a grayscale image and an affective word. Based on the art theories, we adopt the *image-scale* space to build the relation between affective words and color themes. For each object in the input image, a set of the internet images is downloaded and filtered. We propose a selection method based on the generic algorithm to efficiently select the semantically suitable reference images which also meets the desired emotion. We also offer a patch match based approach for the object level colorization.

2   Related Work

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

2.1   Colorization

Colorization methods can be divided into interaction based methods and example based methods. Interaction based methods need users to give some colored scribbles on the grayscale image, and colors are automatically propagated to the remain pixels based on local similarities to complete the colorization process [1]. Since the users need considerable interactions, 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 grayscale image [10, 2]. Recent methods borrow the ideas from machine learning to predict the correspondences between the reference color image and the input grayscale image [5, 4]. All these methods require users to find reference images with similar contents and appropriate appearance, which is usually a difficult task.

With the rapid development of the Internet, data driven processing is attracting more attentions than before. To find the suitable reference images, Chia et al. [6] proposed to filter the internet images, which shared the similar framework as Sketch2Photo [11]. This method is a nice supplementary for the example based colorization, and can also be used as a preprocessing step of these methods. However, it ignores the emotional appearance aspect when filtering, which is the main consideration in this paper.

2.2   Color composition in the art theories

Color composition is the color distribution of an image and the key element for the artistic feeling [7]. Artists often use a set of colors called a *color theme* to represent the color composition. The most commonly used color themes are 3-color themes and

5-color themes, which are templates of three and five colors respectively. The rational study of color themes is a hot field in computer vision and graphics recently. Daniel Cohen-Or et al. applied the existing aesthetical color harmony models to harmonize images [12]. P. O'Donovan et al. studied the color compatibility of 5-color themes from large datasets [13]. However, the emotional aspect of color themes is often ignored.

Kobayashi systematicly studied the relationship between color themes and emotions based on the psychophysical investigations [8, 9], which has already been successfully used in graphic design. In [8], it maps 1,170 3-color themes to 180 affective words, such as romantic, elegant, etc. Furthermore, the relation between 490 5-color themes and 180 affective words is built in [9]. An affective space called *image-scale* is used in [8, 9] to quantitatively describe the emotions. The space has two dimensions including *warm-cool* and *hard-soft*. Fig.1 illustrates a few examples of color themes and the corresponding affective words in this space.
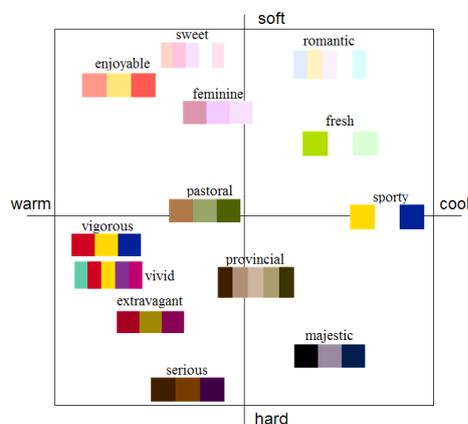

Fig. 1. Examples of color themes and the corresponding affective words in the *image-scale* space.

3   Overview

The overall pipeline of our system is illustrated in Fig.2. The input is a grayscale image and an affective word which is used to express the emotion, such as romantic, intense, serious, etc. (a). The grayscale image is first semi-automatically segmented into objects by a graphcut based segmentation technique [14]. The user labels each object with a text label by which we download and filter a set of images from the Internet, and each object is given a set of candidate references (b). According to the affective word, a set of color themes are selected from a database built using on-line communities by means of the *image-scale* space (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 (d). Finally, we use a patch match based approach (a hybrid texture synthesize) for object-level colorization to get the final result (e).
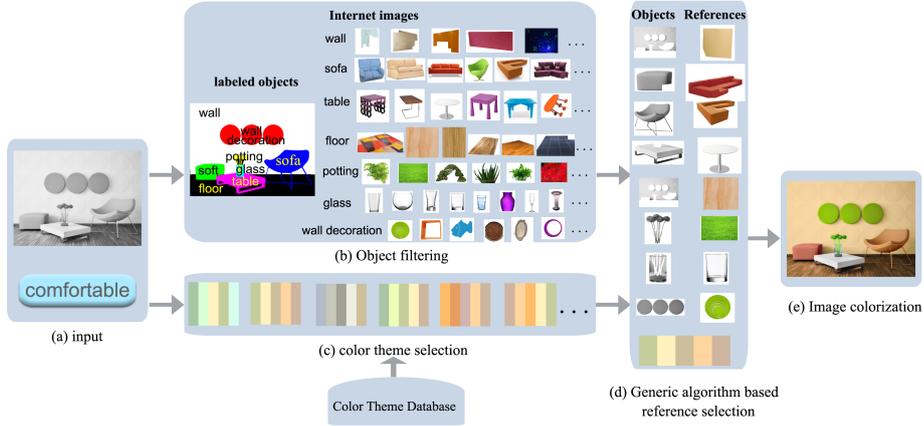

Fig. 2. Pipeline.

4   Reference image selection

4.1   Object filtering

To start our framework, we rely on a semi-automatic approach to segment the input grayscale image into multiple objects. The user roughly gives strokes to specify the objects, and a graphcut based segmentation [14, 15] is used to finally segment the image. To fully utilize the internet images, each object is manually labeled by a semantic text.

We first download a large set of pictures (about 500-1000) from the Internet such as Google Image Search and Flickr for each label. The salient region is automatically extracted from each downloaded image by applying the method called global contrast based salient region detection [16]. Then as same as the contour consistency filtering in [11], we select the images whose salient regions are similar to the outer contour of the grayscale image object with shape context descriptors [17]. 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. This score is used to rank the extracted salient regions, and the top $50 - 100$ objects are retained for further reference image selection (Section 4.3).

4.2   Color theme selection

Color theme database construction We construct a color theme database with 400,000 color themes. Each color theme t_i is labeled with an affective vector $a_i = (wc_i, hs_i)$ 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. For details on the affective word-color theme relationship modeling, readers can refer to our recent work [18].

For each input affective word, it is first automatically labeled with an affective vector. If the word is in the set of the 180 words [8], it has the affective vector. If not, we calculate the semantic similarities between it and the 180 words by the HowNet knowledge system [19], which is a bilingual general knowledge database describing the relations between concepts and attributes. The affective vector 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 affective vector are selected.

4.3   Generic algorithm based reference image selection

Even with the above rough filtering, choosing the appropriate reference images is still a difficult problem. Generic algorithm is adopted to select the most approximate reference image for each object. Here we call a set of all these object-reference correspondences a solution. Formally, the problem can be formalized as follows: There is a set of objects $O = \{o_1, o_2, ...o_n\}$, and each object o_i has a set of candidate reference images $R_i = \{r_{i,1}, r_{i,2}, ..., r_{i,t_i}\}$. In order to achieve an optimized solution $f : o_i \rightarrow R_i, i = 1, 2, ...n$, we design a comprehensive energy G measuring the suitability of the solution, and the goal is to calculate $argmin_f G(f)$.

The energy function considers the consistency with both the original input grayscale image and the given emotion. To formalize the energy, we rewrite it as

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

where $G_s(s) = \sum_i E_s(o_i, r_i)$, $s = \{r_1, r_2, ..., r_n\}$ 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.

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

where E_{ss} is the shape context matching cost which is calculated in the object filtering (Section 4.1). E_{sh} is the histogram matching cost which is defined as the distance between the two histograms. E_{sc} measures the consistency between the color theme extracted from the reference image and the candidate color themes (refer to Equation 3). 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 (eg. orange, horse, etc.), θ_1 can be set to 1. On the other hand, for the objects such as sofa/furniture which are semantic insensitive, θ_1 can be set smaller, e.g. $[0.1, 0.2]$.

$G_a(s)$ is the affective suitability measurement which is the consistency between the color theme of the solution and the candidates, and is defined as

$$G_a(s) = \mathscr{D} \min_{i=1}^{M} D(theme_o, theme_i) \tag{3}$$

where $\mathscr{D}$ is a constant to normalize G_a to $[0, 1]$. $theme_i = \{c_1^i, c_2^i, ..., c_m^i\}$ is one of the M candidate themes where c_j^is are colors in the HSV color space and $theme_o$ is the theme extracted from the colorized image using K-means. $D(\cdot, \cdot)$ is the distance between two themes, which is defined as

$$D(theme_1, theme_2) = \min_{p \in P} \sum_{i=1}^{m} d(c_{p(i)}^1, c_i^2) \qquad (4)$$

where P is the set of permutations of $1, 2, ...m$ and $d(\cdot, \cdot)$ is the Euclidean distance.

5   Object colorization

We use a patch match based colorization method. In order to better understand our method, we first introduce some basic concepts. The example based colorization is essentially a correspondence finding problem. That is, given a grayscale image A and its reference color image B, our task is to find a function $U : A \to B$, and then each pixel 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 grayscale guided texture synthesis, and can also be thought of 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:

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

where i is the pixel iterating over image A. e_1 measures the consistency of 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 algorithms. While the first term considers the grayscale correspondence, the second term e_2 measures the quality of the synthesized colors which is also the main energy term in texture synthesis.

As in [20], 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 $U(i) = u_i$.

The optimization of E is essentially a labeling problem, which is NP-hard. Here we use an iterative process similar to PatchMatch [20], which is also commonly used in the 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 the minimal error:

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

Note that we increase the energy term e_2 gradually, as it is not reliable in the first few iterations. In our experiments, we use 6 iterations and let α be $0, 0, 0, 0.15, 0.2, 0.2$ respectively.

6   Experiments

We have implemented our system on a machine with two quad-core 2.26GHZ CPUs. We apply 10 iterations in the generic algorithm based reference selection, and the algorithm terminates within 1 minute. The colorization step takes less than half a minute for a 1024×768 image. We also manually setup a small knowledge base indicating some objects which do not have common shapes, such as the floor, the sky, etc. They are usually used as backgrounds. We do not extract the salient regions from them and use the full images as references.

We have validated the system with various input examples. Fig. 3 shows the main results. The first column shows the input grayscale images as well as the segmentations and the corresponding labels. We achieve two different colorization results through optionally giving two affective words. The second column gives the words and the corresponding color themes. The third column demonstrates the colorization results and the last column shows the selected references. The colorization results well conform to the given affective words. And they are generally semantic as the selected objects usually belong to the same class.
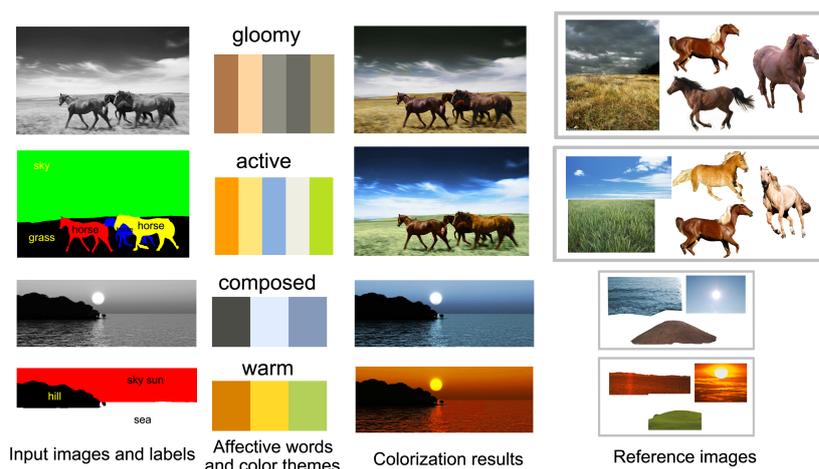
Fig. 3. Colorization results.

7   Conclusions

We have proposed a novel system of the image colorization with an affective word. Using a generic algorithm, it unifies both the affective requirement and the semantics. It gives users a convenient and flexible interface to perfectly colorize a grayscale image. Experiments also convinced the effectiveness of our system.

References

1. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. ACM Transactions on Graphics **23** (August 2004) 689
2. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. ACM Transactions on Graphics **21**(3) (2002) 277–280
3. Ironi, R., Cohen-Or, D., Lischinski, D.: Colorization by example. In: Rendering Techniques. (2005) 201–210
4. Charpiat, G., Hofmann, M., Schölkopf, B.: Automatic image colorization via multimodal predictions. Computer Vision–ECCV (2008) 126–139
5. Tai, Y.W., Jia, J., Tang, C.K.: Local color transfer via probabilistic segmentation by expectation-maximization. Computer Vision and Pattern Recognition **1** (2005) 747–754
6. Chia, A., Zhuo, S., Gupta, R., Tai, Y., Cho, S., Tan, P., Lin, S.: Semantic colorization with internet images. ACM Transactions on Graphics **30**(6) (2011) 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 (1991)
9. Kobayashi, S.: Art of Color Combinations. Kosdansha International (1995)
10. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Computer Graphics and Applications **21**(5) (October 2001) 34–41
11. Chen, T., Cheng, M.M., Tan, P., Shamir, A., Hu, S.M.: Sketch2photo: internet image montage. ACM Transactions on Graphics **28**(5) (December 2009) 124:1–124:10
12. Cohen-Or, D., Sorkine, O., Gal, R., Leyvand, T., Xu, Y.: Color harmonization. ACM Transactions on Graphics **25**(3) (July 2006) 624
13. O'Donovan, P., Agarwala, A., Hertzmann, A.: Color compatibility from large datasets. ACM Transactions on Graphics **30**(4) (July 2011) 63:1–63:12
14. Rother, C., Kolmogorov, V., Blake, A.: "grabcut": interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics **23**(3) (August 2004) 309–314
15. Cheng, M.M., Zhang, F.L., Mitra, N.J., Huang, X., Hu, S.M.: Repfinder: finding approximately repeated scene elements for image editing. ACM Transactions on Graphics **29** (July 2010) 83:1–83:8
16. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, IEEE (June 2011) 409–416
17. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(4) (2002) 509–522
18. Anonymous: Affective image adjustment with a single word. Submitted to The Visual Computer
19. Dong, Z., Dong, Q.: HowNet and the Computation of Meaning. World Scientific (2006)
20. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics **28**(3) (July 2009) 24:1–24:11