KEYWORD-BASED IMAGE COLOR RE-RENDERING WITH SEMANTIC SEGMENTATION

Fayez Lahoud[1], *Bin Jin*[1], *Maria V. Ortiz Segovia*[2] *and Sabine Süsstrunk*[1]

[1] School of Computer and Communication Sciences, EPFL, Switzerland
[2] Océ Print Logic Technologies, Creteil, France

ABSTRACT

Keyword-based image color re-rendering is a convenient way to enhance the color of images. Most methods only focus on the color characteristics of the image while ignoring the semantic meaning of different regions. We propose to incorporate semantic information into the color re-rendering pipeline through semantic segmentation. Using semantic segmentation masks, we first generate more accurate correlations between keywords and color characteristics than the state-of-the-art approach. Such correlations are then adopted for re-rendering the color of the input image, where the segmentation masks are used to indicate the regions for color re-rendering. Qualitative comparisons show that our method generates visually better results than the state-of-the-art approach. We further validate with a psychophysical experiment, where the participants prefer the results of our method.

Index Terms— keyword-based color re-rendering, semantic segmentation, statistical correlation, psychophysical experiment, image enhancement

1. INTRODUCTION

In most consumer cameras, the in-camera processing pipeline contains tone-mapping and color enhancement algorithms to render the captured image visually pleasing. However, the resulting images may still contain unnatural or unsatisfactory colors due to scene composition, mixed light sources, or other elements that confuse the automatic algorithms. Image color re-rendering, which aims at modifying the image colors for better visual appearance, is thus a popular image enhancement step especially for images shared on media platforms.

Keyword-based image color re-rendering is a convenient technique for that as users only need to specify a keyword to modify the colors. The state-of-the-art keyword-based color re-rendering algorithm [1] proposes to first learn the statistical correlations between keywords and color characteristics, and then modify the colors of the input image according to the learned correlations. This approach has the advantage of easy extensibility to many keywords with little human intervention. However, it ignores the semantic meaning of different regions in the image, which can result in unnatural artifacts in the re-rendered image. As shown in Fig. 1b, with the keyword *strawberry*, [1] re-renders the color of the strawberry to

(a) Input image (b) [1] (c) Ours

Fig. 1: Example color re-rendering result for the keyword *strawberry*.

be more vivid, but also modifies the other regions, like the yellow cake and the white background, which become more reddish.

We propose to incorporate semantic information into the color re-rendering pipeline through semantic segmentation. Following [1], we also first compute the statistical correlations between keywords and color characteristics. However, we first use semantic segmentation to locate the keyword-related regions and only compute the correlations using these regions. Color characteristics computed on the located regions rather than on the whole image [1] result in more accurate representations of the keyword, which leads to more accurate correlation measures. When applying the color re-rendering, we again use the semantic segmentation masks to indicate where to modify the colors, resulting in visually better results with fewer artifacts. For instance, in Fig.1c, our method enhances the colors of the strawberries while not affecting the non-strawberry regions. Both the qualitative comparisons and the psychophysical experiment validate that our method generates better color re-rendering results than the state-of-the-art method [1].

2. RELATED WORK

Various methods have been proposed to modify image colors [1–12]. [2, 3] propose different approaches to globally change the color of the input image according to a source image. [4, 5] extend the idea to local adjustment. Wang et al. [6] learn implicit color and tone adjustment rules from example images and apply those on the input image. Sample images are always required for these methods, which hinders their usage in real-world applications.

To remove the need of sample images, some other methods predefine a set of color re-rendering operations. [7, 8] use predefined color palettes or color themes to enhance an image. [9] detect several objects like human eyes and sky, and associate a color re-rendering function to each detected object. The lack of flexibility and generality limits the effectiveness of these methods.

Interactive methods like [10, 11, 13, 14] achieve effective color re-rendering results, while intensive user interventions are required in the process. On the contrary, keyword-based approaches require little human intervention while still being flexible and general. Wang et al. [12] first associate a color theme to each emotion keyword and apply color adjustment accordingly. Lindner et al. [1] generalize to more keywords by statistically analyzing the correlations between keywords and color characteristics. Our method extends [1] by integrating it with semantic segmentation, which leads to notably better color re-rendering results.

3. METHOD

Similar to [1], our algorithm consists of two steps. We first analyze the correlations between keywords and color characteristics, and later modify the color of the input image according to the learned correlations and the input keyword. Semantic segmentation is used in both steps to learn accurate correlations and to locate the keyword-related regions for local color re-rendering.

3.1. Semantic segmentation

Semantic segmentation algorithms segment an image into regions that are related to certain keywords. In this work, we use the weakly supervised semantic segmentation algorithm from [15] for its easy extensibility to any keyword. Fig. 2 shows an example of the segmentation mask from [15].

(a) Input image (b) Segmentation mask

Fig. 2: Semantic segmentation result for the keyword *banana*.

3.2. Keyword-color correlation

Following [1], we statistically analyze the correlations between keywords and color characteristics[1] based on a large image/tags database. Such correlations indicate which color characteristic is mostly related to an input keyword, hence indicating how to re-render the image colors according to the keyword.

Given a large database of image/tags pairs $\{I_i, A_i\}$, where I_i represents the ith image and A_i is the set of the corresponding tags, the goal is to measure the correlation between a keyword k and a color characteristic j. For each image in the database, Lindner et al. [1] propose to first compute its color characteristic vector as:

$$C_i^j = F_j(I_i) \tag{1}$$

where F_j represents the function for computing the characteristic j, and C_i^j is the color characteristic vector for image I_i. A keyword k splits the database into two distinct subsets, $\mathcal{I}_k = \{I_i | k \in A_i\}$ and $\mathcal{I}_{\overline{k}} = \{I_i | k \notin A_i\}$, which leads to two separate color characteristic sets, $\mathcal{C}_k^j = \{C_i^j | I_i \in \mathcal{I}_k\}$ and $\mathcal{C}_{\overline{k}}^j = \{C_i^j | I_i \in \mathcal{I}_{\overline{k}}\}$.

$\mathcal{C}_k^j$ is the set of color characteristic vectors of j for images that contain the tag k, and $\mathcal{C}_{\overline{k}}^j$ is the corresponding set for images that do not carry the tag k. Therefore, the difference between $\mathcal{C}_k^j$ and $\mathcal{C}_{\overline{k}}^j$ implies the correlation between k and j. However, since C_i^j is computed on the whole image, which also encloses the regions that do not correspond to k, the computed correlation is not as accurate due to the noise introduced from those non-related regions. For example, for the image in Fig. 2a with keyword *banana*, the C_i^j computed on the whole image not only describes the color of the banana but also the sky and the grass, which introduces noise to the $\mathcal{C}_k^j$ set.

We propose to use semantic segmentation to improve the accuracy of the keyword-color correlations. Assume M_i^k is the semantic segmentation mask of image I_i for keyword k, like Fig. 2b. The new color characteristic vector $\hat{C}_i^j$ for image I_i is computed as:

$$\hat{C}_i^j = F_j(M_i^k \cdot I_i) \tag{2}$$

Because of M_k^j, $\hat{C}_i^j$ mainly encodes the color information for the regions that correspond to the keyword k with significantly less noise from the unrelated regions. A new characteristic set for $\mathcal{I}_k$ is build as $\hat{\mathcal{C}}_k^j = \{\hat{C}_i^j | I_i \in \mathcal{I}_k\}$, which better represents the color of keyword k than $\mathcal{C}_k^j$ used in [1]. Hence, the dissimilarity between $\hat{\mathcal{C}}_k^j$ and $\mathcal{C}_{\overline{k}}^j$ is a more accurate indication of the correlation between k and j. We do not apply semantic segmentation on $\mathcal{I}_{\overline{k}}$ because images in $\mathcal{I}_{\overline{k}}$ do not carry keyword k.

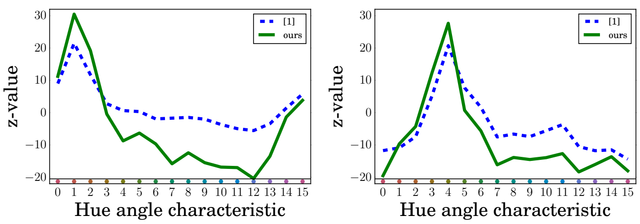
As in [1], we apply the Mann-Whitney-Wilcoxon (MWW) ranksum test [16] to measure the dissimilarity between $\hat{\mathcal{C}}_k^j$ and $\mathcal{C}_{\overline{k}}^j$. From the MWW ranksum test, we compute a $z_{k,j}$ value to represent the correlation between k and j:

$$z_{k,j} = \frac{T - \mu_T}{\sigma_T} \tag{3}$$

[1]Such as hue angle, chrome and CIEL histograms. Refer to [1] for all the color characteristics

where T, μ_T, σ_T are the ranksum, the expected mean and the expected variance according to the MWW test.

The magnitude of z reflects the strength of the correlation, while its sign shows the direction of the correlation. In Fig. 3, we show two typical z values examples computed by our method and [1]. For keyword *strawberry*, the z values computed using our method show a stronger peak around the red hues with smaller values for the other colors than that from [1]. A similar trend is observed for the keyword *sunflower*, demonstrating that our method calculates more accurate keyword-color correlations than [1], as strawberry is only strongly correlated with the red color and sunflower is normally yellow.



(a) Strawberry (b) Sunflower

Fig. 3: z values between the hue angle characteristic and keywords *strawberry* and *sunflower*.

3.3. Local color re-rendering

Keyword-based image color re-rendering takes an input image I_i and a keyword k, and modifies the colors of I_i to be visually more appealing according to the keyword k. To determine how to modify the colors, we refer to $\{z_{k,j} | j \in \mathcal{J}\}$, where $\mathcal{J}$ represents the set of all color characteristics, and choose to enhance the color characteristic j that has the largest $z_{k,j}$ value. For instance, if j is the hue angle histogram, then we modify the hue channel of I_i in LCH color space and convert it back to RGB.

The difference $\delta^j_{I_i,k}$ between the I_i's color characteristic vector $\hat{C}^j_i$ and those in $\hat{\mathcal{C}}^j_k$ is an indication of how to modify the color characteristic j.

$$\delta^j_{I_i,k} = \begin{cases} max[0, Q_{75\%}(\hat{\mathcal{C}}^j_k) - \hat{C}^j_i] & if\ z_{k,j} \geq 0 \\ max[0, \hat{C}^j_i - Q_{25\%}(\hat{\mathcal{C}}^j_k)] & if\ z_{k,j} < 0 \end{cases} \quad (4)$$

where $Q_p(.)$ is a set's p^{th} quantile value. Conceptually, $z_{k,j}$ indicates whether k and j are correlated, $\delta^j_{I_i,k}$ (rewrite as δ for simplicity) indicates how much the color characteristic of I_i, namely $\hat{C}^j_i$, is different from the majority in $\hat{\mathcal{C}}^j_k$. Hence $\delta z_{k,j}$ together implies whether and how to enhance the colors of I_i according to the keyword k.

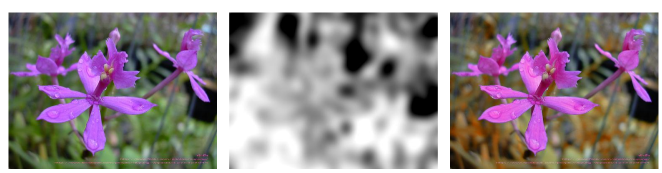
The re-rendering operation is defined as a nonlinear mapping function the same as in [1], with its derivatives m computed as:

$$m = \begin{cases} max[1/m_{max}, 1/(1 + S\delta z_{k,j})] & if\ \delta z_{k,j} \geq 0 \\ min[m_{max}, 1 + S|\delta z_{k,j}|] & if\ \delta z_{k,j} < 0 \end{cases} \quad (5)$$

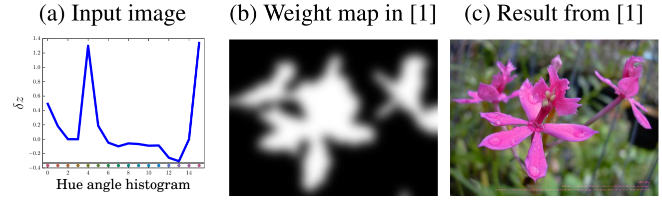
where S is a constant that controls the strength of the non-linearity. The derivatives are clipped to $[1/m_{max}, m_{max}]$ to reduce extreme enhancement artifacts. The mapping function is obtained by integration over the derivatives. Conceptually, such a mapping function map the color characteristic $\hat{C}^j_i$ to the majority values if the correlation is positive and suppress $\hat{C}^j_i$ if the correlation is negative.

To determine where to apply the mapping function, Lindner et al. [1] build a weight map for the input image and the re-rendering is then weighted by the weight map. The weight map is built as the z value between each pixel's color characteristic value and the input keyword. However, since no semantic information is considered when building the weight map, it often mis-identifies the keyword-related regions. As illustrated in Fig.4b, the background green leaves are also captured in the weight map.

We propose to again use semantic segmentation in this step, where we can better locate the keyword-related regions than the weight map in [1]. As shown in Fig. 4e, our segmentation mask is significantly more accurate in locating the orchid. Given a segmentation mask, we first smooth it with a Gaussian filter with $\sigma = 0.02\sqrt{h^2 + w^2}$, where h and w are the image's height and width. This is because the binary mask may introduce artifacts near the edges. The color re-rendering is then weighted according to the smoothed segmentation mask, which results in visually better results than [1]. As illustrated in Fig. 4c and 4f, [1] changes the color of the orchid as well as the green background due to the errors in the weight map. Our method enhances the color of the orchid while leaving the background unchanged. Fig. 4d shows the corresponding $\delta z_{k,j}$ values of our method, which clearly indicates that the orchids in the input image need to be enhanced to be more purple and red.



(a) Input image (b) Weight map in [1] (c) Result from [1]



(d) δz values (e) Our mask (f) Our result

Fig. 4: Local color re-rendering example for *orchid*.

4. EXPERIMENTS

We use the MIR Flickr dataset [17] for computing the keyword-color correlations, which contains one million images, each with multiple annotated tags. The same as in [1],

(a) Banana (b) Strawberry (c) Tulip (d) Ferrari (e) Sunflower (f) Cheese

Fig. 5: Qualitative comparison between the results of [1] and our method.

the parameter S in Eq. 5 is set to 2. m_{max} is set to 5 as a compromise to allow visible changes while lowering the extreme artifacts.

4.1. Qualitative Results

We show qualitative comparison between our method and [1] in Fig. 5. Clearly our method generates visually more appealing results than [1] with much less artifacts. For instance, in Fig. 5a, Lindner et al. [1] modifies the color of the banana to be unnatural cyan while ours correctly re-renders the color of the banana to yellow. This can be attributed to our more accurate keyword-color correlations by using semantic segmentation to filter out non-related regions. Similar observations can be made for Fig 5b and 5c. The colors of strawberries and tulips in our results are clearly more vivid and appealing than those from [1] for the same reason. In addition, in Fig 5d and 5e, our method enhances the colors of the Ferrari and sunflower to be more appealing while not affecting the background, since our segmentation masks accurately locate these objects. Lindner et al. [1] also modifies the colors of the backgrounds due to the errors of the weight maps, resulting in unnatural artifacts. Fig. 5f is a failure case of both our methods as our semantic segmentation algorithm segments part of the table as cheese. Such errors can be improved with the development of better semantic segmentation algorithms. More examples can be found in the supplementary material.

4.2. Psychophysical Experiment

We further validate our color re-rendering method by a psychophysical experiment on a crowd-sourcing website[2]. For this experiment, we choose 50 images covering different keywords. For each image, the participant is shown the original image, and the two re-rendered results from [1] and our method, and is asked to choose the more appealing one among the two re-rendered results. In total 50 users participate in the experiment with each of them labeling all 50 images.

#preference / keywords (#images)	[1]	Ours
strawberry(12)	4	8
banana(11)	3	8
desert(13)	4	9
sunflower(6)	1	5
tulip(5)	2	3
orchid(3)	1	2
all(50)	15	35

Table 1: Results of the psychophysical experiment.

We show the results of the experiment in Table 1. For 35 out of the 50 images, our color re-rendering results are preferred over the counterpart from [1], which means a 70% preference rate of our results compared to 30% of [1]. Moreover, our method is more favored on all keywords, further proving that our method is independent of the keyword.

5. CONCLUSION

We propose to integrate semantic segmentation into the keyword-based image color re-rendering pipeline. The semantic segmentation is first employed to improve the calculation of the keyword-color correlations, where the segmentation masks help to remove the influence of the non-keyword-related regions and lead to more accurate correlation measures. We also use semantic segmentation to locate the keyword-related regions in the input image, and re-render their colors according to the computed correlations. By incorporating semantic segmentation, our keyword-based color re-rendering method generates notably better results than the state-of-the-art approach [1], demonstrated by both the qualitative comparison and the psychophysical experiment.

[2] www.clickworker.com

6. REFERENCES

[1] Albrecht Lindner, Appu Shaji, Nicolas Bonnier, and Sabine Süsstrunk, "Joint statistical analysis of images and keywords with applications in semantic image enhancement," *ACM International Conference on Multimedia*, pp. 489–498, 2012.

[2] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[3] Xuezhong Xiao and Lizhuang Ma, "Color transfer in correlated color space," *ACM International Conference on Virtual Reality Continuum and its Applications*, pp. 305–309, 2006.

[4] Yu-Wing Tai, Jiaya Jia, and Chi-Keung Tang, "Local color transfer via probabilistic segmentation by expectation-maximization," *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 747–754, 2005.

[5] Fuzhang Wu, Weiming Dong, Yan Kong, Xing Mei, Jean-Claude Paul, and Xiaopeng Zhang, "Content-based colour transfer," *Computer Graphics Forum*, vol. 32, no. 1, pp. 190–203, 2013.

[6] Baoyuan Wang, Yizhou Yu, and Ying-Qing Xu, "Example-based image color and tone style enhancement," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 64, 2011.

[7] Baoyuan Wang, Yizhou Yu, Tien-Tsin Wong, Chun Chen, and Ying-Qing Xu, "Data-driven image color theme enhancement," *ACM Transactions on Graphics*, vol. 29, no. 6, pp. 146, 2010.

[8] Huiwen Chang, Ohad Fried, Yiming Liu, Stephen DiVerdi, and Adam Finkelstein, "Palette-based photo recoloring," *ACM Transactions on Graphics*, vol. 34, no. 4, pp. 139, 2015.

[9] Gianluigi Ciocca, Claudio Cusano, Francesca Gasparini, and Raimondo Schettini, "Content aware image enhancement," *Congress of the Italian Association for Artificial Intelligence*, pp. 686–697, 2007.

[10] Alla Maslennikova and Vladimir Vezhnevets, "Interactive local color transfer between images," *Graphicon Conference*, 2007.

[11] Gierad P Laput, Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar, "Pixeltone: a multimodal interface for image editing," *SIGCHI Conference on Human Factors in Computing Systems*, pp. 2185–2194, 2013.

[12] Xiaohui Wang, Jia Jia, and Lianhong Cai, "Affective image adjustment with a single word," *The Visual Computer*, vol. 29, no. 11, pp. 1121–1133, 2013.

[13] Li Xu, Qiong Yan, and Jiaya Jia, "A sparse control model for image and video editing," *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 197, 2013.

[14] Xiaowu Chen, Dongqing Zou, Qinping Zhao, and Ping Tan, "Manifold preserving edit propagation," *ACM Transactions on Graphics*, vol. 31, no. 6, pp. 132, 2012.

[15] Bin Jin, Maria V Ortiz Segovia, and Sabine Süsstrunk, "Webly supervised semantic segmentation," *IEEE International Conference on Computer Vision and Pattern Recognition*, 2017.

[16] Frank Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[17] Mark J. Huiskes and Michael S. Lew, "The MIR flickr retrieval evaluation," *ACM International Conference on Multimedia Information Retrieval*, pp. 39–43, 2008.