Example Based Colorization Using Optimization

Yipin Zhou*
Brown University

Abstract

In this paper, we present an example-based colorization method to colorize a gray image. Besides the gray target image, the user only needs to provide a reference color image which is semantically similar to the gray image. We first segment both the target image and reference image and find correspondences at the segmentation level between these two images. The use of segmentation level can not only speed up the colorization process, but also obtain higher probability to maintain spatial coherence while doing color transfer than using independent pixel directly. Then for corresponding segments we apply pixel-wise chromatic value transfer from reference color image to target image only to the pixels with high confidence. And we use an optimization method [Levin et al. 2004] to propagate those sparse colors to the entire image. The features we use to measure the pixel confidence enable our method work well in both random scene images and images with obvious foreground and background structure. Finally, experimental results and user study on a large set of images demonstrate that our colorization method is competitive with previous state-of-the-art methods.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Radiosity;

Keywords: colorization, segmentation, reference image, optimization

Links: DL PDF

1 Introduction

Image colorization, the process of adding color to grayscale images, can increase the visual appeal of the images. However, colorizing a image and make it perceptual meaningful is an under constrained problem because there are many colors can be assigned to a pixel with known intensity.

To reduce the ill-posedness, human interaction usually plays an important role in the colorization process. The interactive colorization methods [Levin et al. 2004] [Huang et al. 2005] require users to draw color scribbles on the target image, and an optimization method will be applied to propagate those colors to the entire image. Interactive methods rely on extensive manually work from users and also qualified results often require users have a good sense of choosing and matching suitable colors.

Another main class of techniques are example-based colorization methods [Welsh et al. 2002] [Irony et al. 2005] [Liu et al. 2008] [Chia et al. 2011] [Gupta et al. 2012] [Charpiat et al. 2008], which take a color reference image as the input and transfer color from the reference image to the target grayscale image. These methods can reduce the user effort, while require more consideration on how to transfer the color properly.

In this paper, we present a method combined the advantages of both interactive techniques and example-based techniques. We use an reference image as the color information source and only transfer the color from reference to the pixels of the target image with high confidence. By doing that, we have sparse color scribbles avoiding manual effort and then we propagate them to the entire image using [Levin et al. 2004]'s optimization-based method. Specifically, the features we use to measure the confidence includes luminance value and standard deviation which are used by [Levin et al. 2004], and SURF, Gabor features which are applied by [Gupta et al. 2012]. SURF is chosen for its discriminative attributes and efficiency compared with SIFT descriptor and Cabor is applied for its effective representation of texture, which are very helpful to select the color from the right place of the reference image. Besides, we also use high-level salient map as the last feature to enforce the spatial consistency.

We evaluate our method on a broad range of images compromising random scene images and images with obvious foreground and background spacial layout such as portrait. Then we compare our results with existing methods and apply a simple user study to demonstrate our method can yield visually meaningful and appealing images.

2 Related Work

Existing work on colorization can be broadly divided into two classes: interactive colorization methods and example-based colorization methods.

Interactive colorization [Levin et al. 2004] proposed a simple but still effective colorization algorithm that needs the users add color scribbles manually to the image as indications and propagate those color scribbles to the entire image automatically. The quality of the results highly depend on the user's effort and aesthetic taste. [Huang et al. 2005] improved the propagation method by reducing color blending at edges.

Example-based colorization [Welsh et al. 2002] introduced a colorization method based on swatches matching between reference image and target image. However, this method still requires user to manually mark the corresponding patches and maintains weak spatial consistency. To keep the spatial consistency, [Irony et al. 2005] proposed a colorization method which needs manually segmented regions of reference image as an additional input and automatically determine for the pixels of target image which reference segment it should learn its color from. [Charpiat et al. 2008] does color transfer by minimizing an energy function using the graph cut algorithm. While their method heavily depends on finding a suitable reference image. [Liu et al. 2008]decomposes the target and reference images into illumination and reflectance layers and does color transfer based on the reflectance. This method is robust to the illumination difference between target and reference images

*e-mail: yipin_zhou@brown.edu

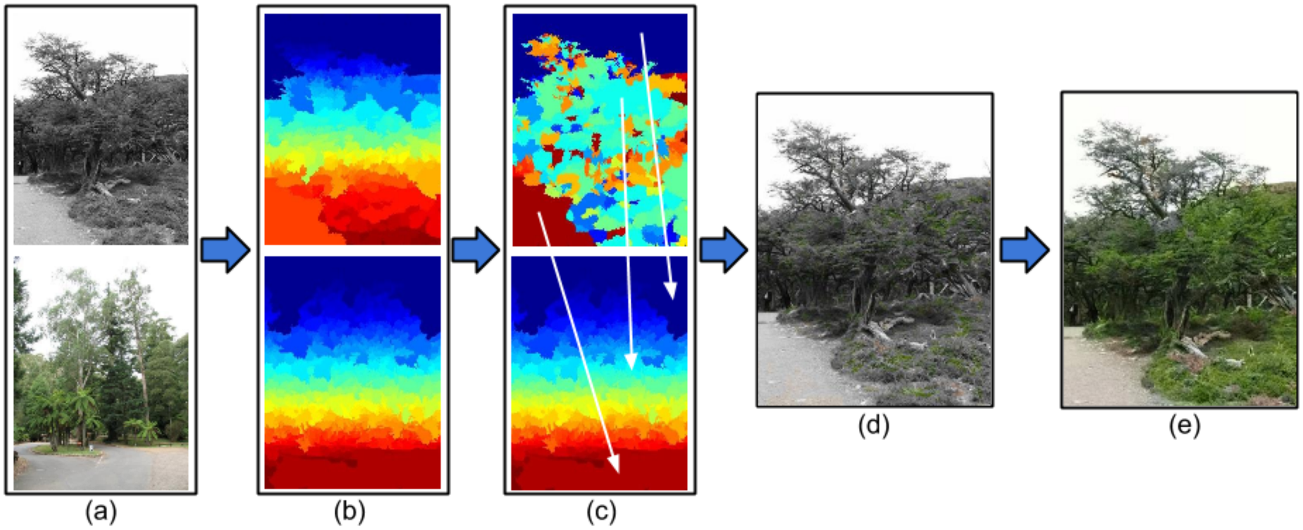

Figure 1: *Overview of our colorization method. (a) Input target gray image and reference color image. (b) Segmentation visualization of both target and reference images. (c) For each target segment, we find a corresponding reference segment, where the corresponding segments are visualized as the same color. And for each pixel in target segments, we find the optimal pixel from corresponding reference segments and (d) only transfer the color to those pixels with high confidence. (e) Colorization result after propagation.*

while it requires several reference images with similar viewpoints to insure a valid intrinsic image decomposition. [Chia et al. 2011] developed a method which obtains reference images from internet using a novel image filtering framework. To colorize a grayscale image, it requires the user segments the target image into foreground and background parts and provide semantic text label for each object. It transfers the color to foreground and background parts separately so this method works well for images with clear foreground/background structure. Recently, [Gupta et al. 2012]introduced a method which adopts a fast cascade feature matching scheme to find the correspondences between target and reference images and develops a image space voting framework to enforce the spatial coherence.

3 Overview

An overview diagram of our approach is shown in figure 1. To colorize a grayscale target image, the user needs to provide a reference color image which is semantically similar to the target image and it is better to also have the similar spatial layout. This is the only input required from the user. Then we segment both the target and reference images using Mean shift algorithm [Comaniciu and Meer 2002]. And compute features for each pixel. The features of one segment is the average of features of all the pixels within this segment. Based on features of segments, we find the segments correspondences between target and reference images. For each pixel in target segments we find the optimal pixel from the corresponding reference segment and transfer the color for those pixels with high confidence. Finally, we minimize an energy function to propagate those sparse colors to the entire image.

4 Colorization algorithm

4.1 Segmentation level correspondences

Before applying pixelwise color transfer, we first segment the images using mean shift and find correspondences between target and reference segments. The reasons why we use segmentation level are as following. First, applying segmentation correspondences will speed up the colorization process since for each pixel in target image we can find the optimal pixel from corresponding segment of reference image instead of searching whole image. The second reason is that finding segmentation correspondences has higher probability to keep spatial consistency compared with using pixel correspondences directly for regions tend to contain more spatial information than independent pixels. Third, segmentation correspondences allow us to select pixels with high confidence from every segments and we will have sparse colors on each spatial part. Instead, if we only use the pixelwise correspondences, for two images have one spatial part very close related, the pixels with high confidence will only belong to that part and poor result with monotonous color will be generated after propagation.

We use [EDISON] to perform Mean shift segmentation. To get suitable number and size of segment regions, we set the spatial bandwidth and range bandwidth both equal to 8 through experiments.

4.2 Features extraction

For each pixel in target gray image and reference color image, we compute 5 features based on luminance value, standard deviation, Gabor feature, SURF feature and high-level salient map. Each feature of one segment is the mean value of that feature of all pixels that belong to that segment. We compute each feature as follows:

Luminance value We use the CIELuv color space to transfer the color from reference pixels to target pixels, which make it easier to separate luminance and color components. We use luminance layer as the luminance value for each pixel.

Standard deviation We also need to consider the neighborhood statistics, so we compute standard deviation of the luminance values of each pixel neighborhood. For all the results in this paper, we use a neighborhood size of 5×5 pixels.

Gabor We apply Gabor filter [Manjunath and Ma 1996] to the image and compute a 40-dimensional feature for each pixel. Similar

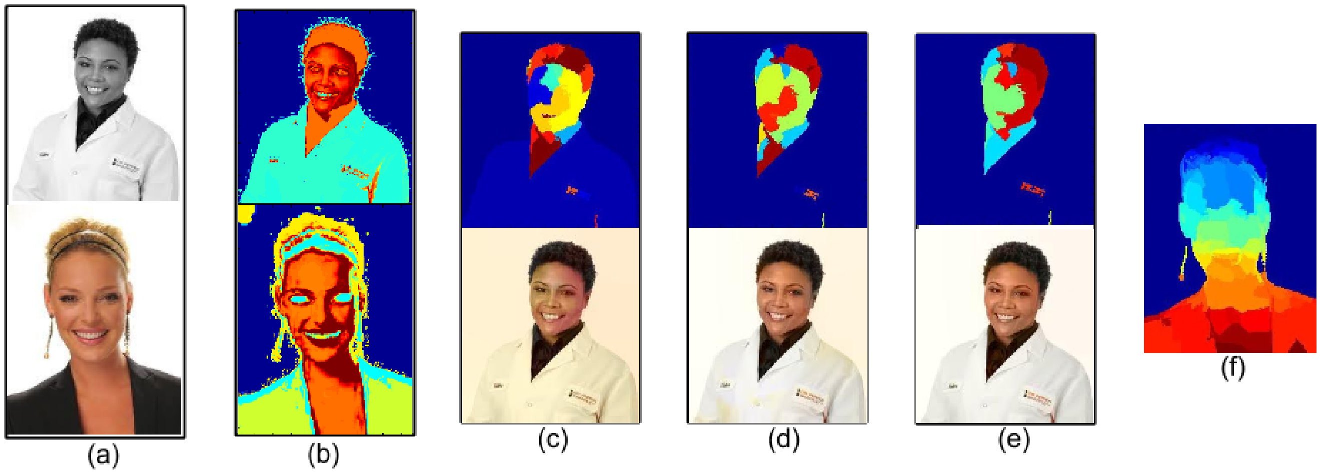

Figure 2: *Effects of 5 features to spacial consistency and colorization results. (a) Input target gray and reference color images. (b) Salient maps of target and reference images. (c) Colorization using luminance and standard deviation features. Above is the segments correspondences with reference image(f), corresponding segments own the same color. As we can see, only using these two features yields a poor spatial consistency(the blue segment on target woman's face means the it's color will be transferred from hair part of the reference woman). Below is the colorization result. (d) Colorization using luminance, standard deviation, Gabor and SURF features. As we can see, the spatial consistency is obviously improved. (e) Colorization using above mentioned 4 features and Salient map. Both spatial consistency and colorization results are further improved. (f) Segmentation of reference image using mean shift.*

to the work in [Gupta et al. 2012], we set 8 orientations (0 from $\frac{7}{8}\pi$) and five exponential scales exp(i$\times\pi$) (i = 0,1,2,3,4,5).

SURF descriptor We extract a 128-dimensional extended SURF (Speed Up Robust Features) descriptors [Bay et al. 2008] at each pixel.

Salient map Based on the intuition that if two images have similar semantic content and spatial layout, the human brain and visual system tend to pay similar attention to the corresponding regions between two images, namely, regions with relatively high salient value in one image have higher probability to correspond to regions in the other image with relatively high salient value, we use salient map to further enforce spatial coherence. In this paper, we apply [Liu et al. 2011]'s method to compute normalized salient map, which incorporate the high-level concept of salient object into the process of visual attention computation and has a good indication for where a user's attention is while perusing images.

In segments and pixelwise correspondences, for each segment(pixel) in target image, the corresponding segment(pixel) in the reference image is the one with the least distance to the target segment(pixel). The distance is defined as:

$$\begin{aligned} D(A,B) = w_1E_1(A,B) + w_2E_2(A,B) + w_3E_3(A,B) \\ + w_4E_4(A,B) + w_5E_5(A,B) \end{aligned} \quad (1)$$

A,B represent segment(pixel) A in target image and segment(pixel) B in reference image. And we denote E_1, E_2, E_3, E_4, E_5 as the Euclidean distance between the luminance, standard deviation, Gabor, SURF and salient map features and w as their weights. In this paper, we set w_1,w_2,w_4,w_5 to be 0.3, 0.1, 0.2, 0.2 and 0.2 respectively. Figure 2 shows how do those features help to maintain spatial consistency. For each segment in target image we choose the segment in reference image with the least distance and for each pixel in target segment we select an optimal pixel in corresponding reference segment with the least distance. For each target segment, we consider 15% pixels with least distance to their optimal pixels from reference image as the high confidence pixels and only transfer color to UV chrominance channels of those pixels.

4.3 Optimization

Since we have the sparse colors on the target image, we would like to propagate the colors to the entire image using [Levin et al. 2004], an optimization-based interpolation method based on the principle that neighboring pixels with similar luminance(intensity) should have similar color.

This interpolation method works in YUV color space, where Y is the luminance channel and U, V are color channels. In image I, we convert the constraint that two neighboring pixels r,s should have similar colors if their luminance values are similar to equation in least-square sense. The goal of this step is to minimize the equation:

$$J(C) = \sum_{r \in I}(C(r) - \sum_{s \in N(r)} wC(s))^2 \quad (2)$$

where N(r) is the set of neighboring pixels of pixel r, C(r) represents color of U or V channel of r and w is a weighting function, large then two pixels have similar luminance, and small when two luminance values are different.

$$w = \exp(\frac{-(Y(r) - Y(s))^2}{2\sigma_r^2}) \quad (3)$$

Y(r), Y(s) are luminance value of r and s, and σ_r represents the variance of the luminance in a window around r. One can refer [Levin et al. 2004] for further details. And figure 3 shows several groups of propagation results.

5 Results

To evaluate our colorization method, we compare our result images with existing state-of-the-art colorization methods using the

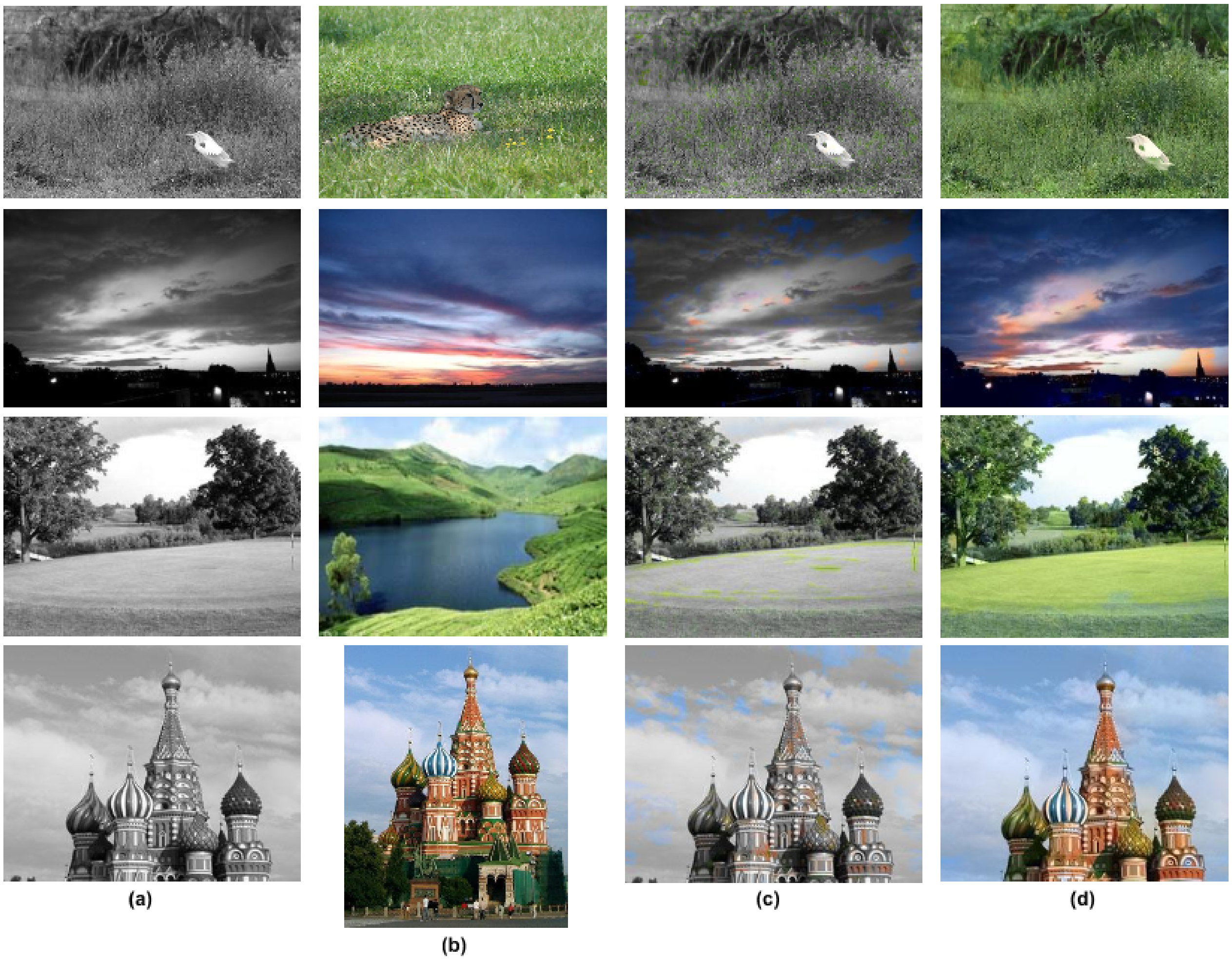

Figure 3: *(a) Input target gray image. (b) Reference color image. (c) Applying color transfer to pixels with high confidence. (d) Propagation results*

test cases of [Gupta et al. 2012] and the colorization results of other methods are from Gupta's paper. Figure 4 compares our method against [Gupta et al. 2012] and [Charpiat et al. 2008]'s colorization algorithm, where the first group of results have the reference image different with target image but with similar semantic content and similar spatial layout, while the reference images of second group have exactly the same foreground object with that of the target image but the viewpoints are slightly different. Figure 5 shows colorization results, with comparisons to existing state-of-the-art methods: [Welsh et al. 2002] [Irony et al. 2005] [Charpiat et al. 2008] [Gupta et al. 2012]. As the result shows, though we use [Gupta et al. 2012]'s test cases, our colorization results can outperform other methods while be competitive with Gupta's results.

5.1 User study

Finally, we perform a simple user study to further evaluate our colorization method. We engage 10 volunteers and show them a set of test images one by one to tell whether it is an artificial colored image or a real image. Each subject is given 5 to 10 seconds for every image to make their decision. Our test set includes 10 artificial images and 10 real images in random sequence. The result of user study is shown in table 1. Averagely, there are 64% of artificial colored images that are considered as real, while interestingly

Table 1: *Fake as real column represents the probability of the artificial images that are considered as real. Real as real column shows the probability of the real images that are considered as real.*

Index	Fake as real	Real as real
Subject 1	60%	70%
Subject 2	80%	70%
Subject 3	70%	80%
Subject 4	90%	100%
Subject 5	70%	60%
Subject 6	30%	70%
Subject 7	50%	50%
Subject 8	70%	60%
Subject 9	60%	60%
Subject 10	60%	70%
Total	64%	69%

only 69% real images are thought as real, which because the subjects suggest themselves that there must be some fake images during the whole testing process and also, to some degree, demonstrates they have high requirements for the real images. Another interesting thing is that when asked how do subjects discriminate

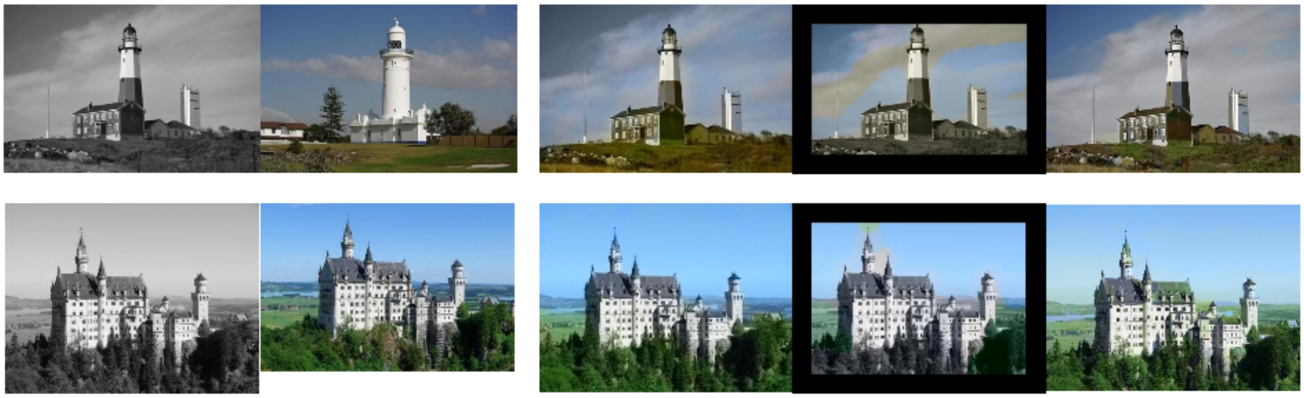

Figure 4: *Comparison with other state-of-the-art methods.*

Figure 5: *Comparison with other state-of-the-art methods.*

whether an image is artificial colored or not, most of them agree that they pay more attention to whether the color assortment and hue of whole image is natural, instead of concentrating on incorrect colorization in tiny places. This intuition can be an important guide for the improvement of our colorization algorithm in the future.

6 Conclusion

In this paper, we present a colorization method to bring a target gray image into life by transferring color properly from a reference image with semantically similarity. We extract features from each pixel and build segmentation level correspondences between segments of target and reference images. Then for each pixel in a target segment we find optimal pixel with the least feature distance from the corresponding reference segment and we only transfer values of UV color channels for pixels with relatively high confidence. Finally, we apply an optimization based interpolation method to propagate sparse colors to the entire image. We generate our colorization results based on a broad range of images and compare the our results with results of existing state-of-the-art method to demonstrate that our method is competitive. We also develop a simple user study which shows that our colorization results are pretty convincing even compared with real images.

In the future, we would like to further explore features with improved discriminative potential that can better build correspondences between target and reference images and measure the confidence of pixels to yield more accurate color transfer. Besides, we are also willing to develop a image filtering framework which can automatically find suitable reference image based on the semantic and spatial layout information of target image from internet.

References

BAY, H., ESS, A., TUYTELAARS, T., AND VAN GOOL, L. 2008. Speeded-up robust features (surf). *Comput. Vis. Image Underst. 110*, 3 (June), 346–359.

CHARPIAT, G., HOFMANN, M., AND SCHLKOPF, B., 2008. Automatic image colorization via multimodal predictions.

CHIA, A. Y.-S., ZHUO, S., GUPTA, R. K., TAI, Y.-W., CHO, S.-Y., TAN, P., AND LIN, S. 2011. Semantic colorization with internet images. In *Proceedings of the 2011 SIGGRAPH Asia Conference*, ACM, New York, NY, USA, SA '11, 156:1–156:8.

COMANICIU, D., AND MEER, P. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 24*, 5 (May), 603–619.

EDISON. Code for the edge detection and image segmentation system.

GUPTA, R. K., CHIA, A. Y.-S., RAJAN, D., NG, E. S., AND ZHIYONG, H. 2012. Image colorization using similar images. In *Proceedings of the 20th ACM international conference on Multimedia*, ACM, New York, NY, USA, MM '12, 369–378.

HUANG, Y.-C., TUNG, Y.-S., CHEN, J.-C., WANG, S.-W., AND WU, J.-L. 2005. An adaptive edge detection based colorization algorithm and its applications. In *Proceedings of the 13th annual ACM international conference on Multimedia*, ACM, New York, NY, USA, MULTIMEDIA '05, 351–354.

IRONY, R., COHEN-OR, D., AND LISCHINSKI, D. 2005. Colorization by example. In *Proceedings of the Sixteenth Eurographics conference on Rendering Techniques*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, EGSR'05, 201–210.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2004. Colorization using optimization. *ACM Trans. Graph. 23*, 3 (Aug.), 689–694.

LIU, X., WAN, L., QU, Y., WONG, T.-T., LIN, S., LEUNG, C.-S., AND HENG, P.-A. 2008. Intrinsic colorization. *ACM Transactions on Graphics (SIGGRAPH Asia 2008 issue 27*, 5 (December), 152:1–152:9.

LIU, T., YUAN, Z., SUN, J., WANG, J., ZHENG, N., TANG, X., AND SHUM, H.-Y. 2011. Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell. 33*, 2 (Feb.), 353–367.

MANJUNATH, B. S., AND MA, W. Y. 1996. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell. 18*, 8 (Aug.), 837–842.

WELSH, T., ASHIKHMIN, M., AND MUELLER, K. 2002. Transferring color to greyscale images. *ACM Trans. Graph. 21*, 3 (July), 277–280.