Automatic colorization of grayscale images using multiple images on the Web

Yuji Morimoto* Yuichi Taguchi Takeshi Naemura
The University of Tokyo The University of Tokyo The University of Tokyo

1 Introduction

Colorization is the process of adding color to monochrome images and video. It is used to increase the visual appeal of images such as old black and white photos, classic movies, and scientific visualizations. Since colorizing grayscale images involves assigning three-dimensional (RGB) pixel values to an image whose elements are characterized by one feature (luminance) only, the colorization problem does not have a unique solution. Hence, human interaction is typically required in the colorization process. Although existing colorization methods attempt to minimize the amount of user intervention, they require users to manually select a similar image to the target image or input a set of color seeds for different regions of the target image. In this paper, we present an entirely automatic colorization method using multiple images collected from the Web. The method generates various and natural colorized images from an input monochrome image by using the information of the scene structure.

2 Our Method

Our colorization method is entirely automatic and outputs multiple natural colorized images by using one million images downloaded from Flickr (http://www.flickr.com/). To colorize an input grayscale image \( I_m \) (Figure 1), the method first selects reference color images (source images) from the image database. It then generates multiple color images by colorizing \( I_m \) using each source image.

We select images that have similar scene structure with \( I_m \) as the source images, because images that have similar structure are likely to have a similar color. To find the source images from one million images, we used gist scene descriptor, which is a feature vector to describe the global scene in lower dimension. The gist scene descriptor aggregates oriented edge responses at multiscales into very coarse spatial bins. We used a gist scene descriptor built from 6 oriented edge responses at 5 scales aggregated to a 4 x 4 spatial resolution and converted one million images to 480 dimensional vectors [Hays and Efros 2007]. We calculated the SSD between the gist of \( I_m \) and every gist of the one million images and selected the most similar 100 images. From these images, we used the 20 images that have the most similar aspect ratio to \( I_m \) as the source images (Figure 2).

Next, we colorize \( I_m \) using these source images. For this purpose, we modified Welsh et al.’s method [2002], which colorizes a grayscale image by transferring color from a source color image. In order to transfer chromaticity values from a source image to \( I_m \), each pixel in \( I_m \) must be matched to a pixel in the source image. The comparison is based on the luminance value and the standard deviation of the luminance in a pixel neighborhood with size of 5 x 5 pixels. Since our method selects a source image that is similar to \( I_m \) not only in color trend but also in scene structure, we can also use the structure information for colorization. For example, since two images have similar structure, a color at the top left of the image is likely to appear in the same part of the other image. We embedded this information in the pixel matching algorithm by penalizing the matching score when the location between the two pixels is separated. Once the best matching pixel is found, chromaticity values are transferred to the target pixel. To transfer only chromaticity values, images are converted to the \( l\alpha\beta \) color space. The \( l\alpha\beta \) color space consists of an achromatic luminance channel \( l \) and two chromatic channels \( \alpha \) and \( \beta \). This color space minimizes the correlation between the three coordinate axes of the color space. By transferring \( \alpha \) and \( \beta \), we can transfer only chromaticity values without altering the luminance.

As a result, our method produces multiple colorized images (Figures 3 and 4), from which users can choose the one that they like.

3 Conclusions and Future Work

We presented a novel colorization method using one million images collected from the Web. Our method is entirely automatic and produces multiple natural colorized images by exploiting the similarity of the scene structure. Currently, the limitation of our method is that it sometimes produces unnatural colorized images, caused by source images that are structurally similar but semantically different. To solve this problem, we are considering combining the obtained color images to generate a more robust result.

References


*E-mail: morimoto@nae-lab.org