Incorporating Higher-Order Cues in Image Colorization

Tae Hoon Kim
thkim@diehard.snu.ac.kr
Kyoung Mu Lee
kyoungmu@snu.ac.kr
Sang Uk Lee
sanguk@ipl.snu.ac.kr

Dept. of EECS, ASRI
Seoul National University
Seoul, Korea

Colorization problem is to find the colors of all pixels $X = \{x_n\}_{n=1,\dots,|X|}$, given a grayscale image I with scribbles S with the desired colors. We work in the YUV color space where $Y = \{y_n\}_{n=1,\dots,|X|}$ is the monochromatic luminance channel, which we will refer to simply as intensity, while $U = \{u_n\}_{n=1,\dots,|X|}$ and $V = \{v_n\}_{n=1,\dots,|X|}$ are the chrominance channels, encoding the color. Our goal is to complete both the U and V channels, given $Y = I$. We deal with the only U channel in this paper, since the V channel can be treated in the same manner.

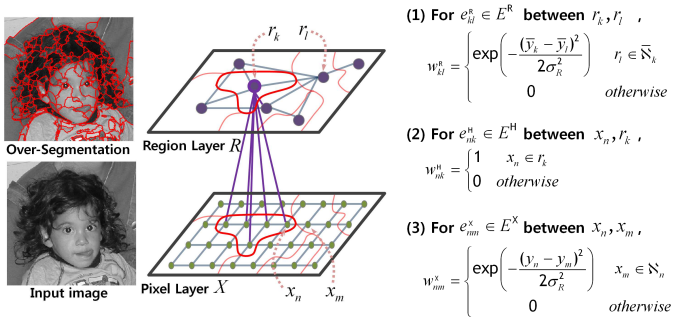
In this paper, we propose a new multi-layer graph model and an energy formulation that can incorporate higher-order cues for reliable colorization of natural images. In contrast to most existing energy functions [3] with unary and pairwise constraints, we address the problem of imposing a high-order constraint whereby pixels constituting each region tend to have similar colors to the representative color of the region they belong to. The representative colors of the regions that are generated by unsupervised image segmentation algorithms, act as higher-order cues. Unlike previous parametric models [2], they are automatically obtained by a non-parametric learning technique that estimates them from the resulting pixel colors in a recursive fashion. We formulate this problem in terms of two quadratic energy functions of pixel and region colors, that are supplementary to each other, in our proposed multi-layer graph model and estimate them by a simple optimization technique that minimizes both functions simultaneously.

Our proposed algorithm works as follows. We first design an undirected graph $G = (Q, E)$ where the nodes $Q = \{X, R\}$ consist of two types: pixels X and regions R, generated by an unsupervised segmentation algorithm such as Mean Shift [1], and the edges E are the links between two nodes as shown in Fig. 1(a). Each pixel $x_n \in X$ initially has an intensity $y_n \in Y$. For each region $r_k \in R$, we can generate its properties $\bar{y}_k$ as the mean intensity of the inner pixels $x_n \in r_k$: $\bar{y}_k = \frac{1}{|r_k|}\sum_{x_n \in r_k} y_n$. We then formulate both quadratic energy functions J^{X} and J^{R} for estimating the pixel colors $U = \{u_n\}_{n=1,\dots,|X|}$ and the region colors $\bar{U} = \{\bar{u}_k\}_{k=1,\dots,|R|}$, respectively, as follows.

$$\begin{aligned} J^{\mathrm{X}} &= \mathscr{E}^{\mathrm{X}}_{pairwise} + \lambda \mathscr{E}^{\mathrm{X}}_{unary} + \tau \mathscr{E}^{\mathrm{X}}_{region} \\ &= \sum_{x_n,x_m \in X} \tilde{w}^{\mathrm{X}}_{nm}(u_n - u_m)^2 + \lambda \sum_{x_n \in X^+} (u_n - u^+_n)^2 + \tau \sum_{x_n \in X} (u_n - \sum_{r_k \in R} \tilde{w}^{\mathrm{XR}}_{nk} \bar{u}_k)^2 \end{aligned} \tag{1}$$

$$\begin{aligned} J^{\mathrm{R}} &= \mathscr{E}^{\mathrm{R}}_{pairwise} + \lambda \mathscr{E}^{\mathrm{R}}_{unary} + \varepsilon \mathscr{E}^{\mathrm{R}}_{pixel} \\ &= \sum_{r_k,r_l \in R} \tilde{w}^{\mathrm{R}}_{kl}(\bar{u}_k - \bar{u}_l)^2 + \lambda \sum_{r_k \in R^+} (\bar{u}_k - \bar{u}^+_k)^2 + \varepsilon \sum_{r_k \in R} (\bar{u}_k - \sum_{x_n \in X} \tilde{w}^{\mathrm{RX}}_{kn} u_n)^2, \end{aligned} \tag{2}$$

where $\tilde{w}^{\mathrm{X}}_{nm} = \frac{w^{\mathrm{X}}_{nm}}{\sum_{n'=1}^{|X|} w^{\mathrm{X}}_{nn'}}$, $\tilde{w}^{\mathrm{XR}}_{nk} = \frac{w^{\mathrm{H}}_{nk}}{\sum_{k'=1}^{|R|}, w^{\mathrm{H}}_{nk'}}$, $\tilde{w}^{\mathrm{R}}_{kl} = \frac{w^{\mathrm{R}}_{kl}}{\sum_{k'=1}^{|R|} w^{\mathrm{R}}_{kk'}}$ and $\tilde{w}^{\mathrm{RX}}_{kn} = \frac{w^{\mathrm{H}}_{nk}}{\sum_{n'=1}^{|X|} w^{\mathrm{H}}_{n'k}}$ in Fig. 1(b) are defined as the normalized weights. In a subset X^+ of X, each node $x_n \in X^+$ represents the pixel with user-given color $u^+_n \in U^+$. We similarly select a subset R^+ of R, where each node $r_k \in R^+$ is a region containing at least one pre-defined pixel, and it has the initial color $\bar{u}^+_k \in \bar{U}^+$, defined as the mean color of inner pre-defined pixels. Three weighting parameters λ, τ, and ε are initially fixed for our all experiments. The common energy model is characterized by the energy function only defined in unary and pairwise terms. Note that in our work, we propose to use the additional higher-order term $\mathscr{E}^{\mathrm{X}}_{region}$ in (1) with a constraint whereby a pixel color should be similar to its corresponding region color. Unlike the hard color consistency constraints that were used in other conventional region-based methods, the color consistency between its inner pixels is partly enforced with a weight τ. In (2), the third term $\mathscr{E}^{\mathrm{R}}_{pixel}$ is also proposed as another estimated unary constraint whereby a region color should be similar to the mean color of inner pixels. This term has the effect of refining the region colors from more informative pixel colors, when there is less color information from user-inputs.



(a) (b)

Figure 1: Illustration of a proposed graph. A node $Q = \{X, R\}$ denotes a pixel $x_{n,m} \in X$ (green circle) or a region $r_{k,l} \in R$ (violet circle). The boundaries of the regions, generated by unsupervised segmentation algorithm [1], are drawn in red color overlaid on the image in (a). (a) shows an example of the edges between a region and its corresponding pixels with violet lines. In (b), the weights $W = \{W^{\mathrm{X}}, W^{\mathrm{R}}, W^{\mathrm{H}}\}$ are assigned to $E = \{E^{\mathrm{X}}, E^{\mathrm{R}}, E^{\mathrm{H}}\}$ and classified by different criteria.

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

Figure 2: Colorization results on the natural images. (a) Input images with color scribbles. (b) Regions generated by an unsupervised image segmentation algorithm [1]. Colorizations by (c) Yatziv et al. [4]; (d) Levin et al. [3]; (e) Our algorithm.

Since two quadratic energy functions J^{X} and J^{R} in (1) and (2) are convex and supplementary to each other, we simply minimize them together by differentiating their matrix forms $\mathbf{J}^{\mathrm{X}}$ and $\mathbf{J}^{\mathrm{R}}$ with respect to two vectors $\vec{u} = [u_n]_{|X|\times 1}$ and $\vec{\bar{u}} = [\bar{u}_k]_{|R|\times 1}$, respectively, and set to zero. We finally have the colors of all pixels X and regions R for the colorization problem by a sparse matrix inversion technique.

Fig. 2 shows our final colorization results on natural images, compared with those of [3] and [4]. In Fig. 2(c), the results by [4] are largely sensitive to the size and position of each scribble. Fig. 2(d) shows that the method in [3] produces the over-smoothed results with color artifacts visible especially near edges away from the scribbles. In contrast, our approach provides superior performance with much higher-quality colorization results as in Fig. 2(e). These comparisons clearly confirm the robustness and accuracy of our algorithm.

[1] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[2] P. Kohli, L. Ladický, and P.H.S. Torr. Robust higher order potentials for enforcing label consistency. *Int. J. Computer Vision*, 82(3):302–324, 2009.

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

[4] L. Yatziv and G. Sapiro. Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing*, 15 (5):1120–1129, 2006.