Landmark-Based Sparse Color Representations for Color Transfer

Tzu-Wei Huang and Hwann-Tzong Chen
Department of Computer Science, National Tsing Hua University
30013 Hsinchu, Taiwan

Abstract

We present a novel image representation that characterizes a color image by an intensity image and a small number of color pixels. Our idea is based on solving an inverse problem of colorization: Given a color image, we seek to obtain an intensity image and a small subset of color pixels, which are called landmark pixels*, so that the input color image can be recovered faithfully using the intensity image and the color cues provided by the selected landmark pixels. We develop an algorithm to derive the landmark-based sparse color representations from color images, and use the representations in the applications of color transfer and color correction. The computational cost for these applications is low owing to the sparsity of the proposed representation. The landmark-based representation is also preferable to statistics-based representations,* e.g. *color histograms and Gaussian mixture models, when we need to reconstruct the color image from a given representation.*

1. Introduction

To design vision algorithms that fully take advantage of spatial and color characteristics of images is not straightforward. The effectiveness of the algorithms would heavily rely on the choice of the digitized representation for color images. With the most popular RGB model, a digital color image is stored as a stack of three matrices, each of which corresponds to one of the RGB channels. Although such a matrix-based color representation is quite convenient, for many vision or image-processing applications we may also need to obtain statistical information from all pixels, in addition to the RGB values of individual pixels.

Statistics-based image representations, such as color histograms, provide global information and have been widely used for solving various computer vision problems, *e.g.*, color constancy [13], image segmentation [2], visual tracking [3], [9], skin detection [6], and image retrieval [16]. Color histograms are easy to compute, stable in presence of partial occlusion, and robust in tolerating geometric transformations. However, a major disadvantage of color histograms is that they do not preserve the spatial information. To overcome this drawback, Huang *et al*. [4] present the *color correlogram*, which describes the global distribution of local spatial correlations of colors. Huang *et al*. show that the color correlogram outperforms traditional histogram methods on the task of content-based image retrieval. The color correlogram provides a distinctive representation and works well for indexing or detection tasks, but it is not suitable for color image processing since the correspondences between the color correlogram and the original image cannot be recovered after applying transformations to the coorelogram. Gaussian mixture models are also popular for modeling spatial-color information of images. Like the color correlogram, the computation of Gaussian mixture models is irreversible, *i.e*., given a spatial-color distribution represented as a Gaussian mixture model, we are not able to reconstruct the corresponding color image that yields the desired distribution.

We present a novel image representation that characterizes a color image by an intensity channel and a small number of color pixels. Our work is motivated by the image colorization algorithms of Levin *et al*. [7] and Yatziv and Sapiro [18]. Given a grayscale image and user-specified color scribbles on some pixels, their algorithms are able to colorize the input image under the constraints introduced by the scribbles. Our idea of color image representation is based on solving an inverse problem: Given a color image, we seek to obtain a grayscale image and a small subset of color pixels, which are called *landmark pixels*, so that the input color image can be recovered faithfully using the grayscale image and the color cues provided by the selected landmark pixels.

The proposed landmark-based color representation inherently integrates the spatial and chromatic information of color images. The landmark pixels themselves maintain the representative colors of an image, and the locations of landmark pixels combined with the underlying grayscale image describe the spatial correlations between nearby pixels. We will show that such a representation has great advantages when they are applied to color-related tasks like color transfer or color correction. With the proposed representation, we are able to adjust the color of landmark pixels, and then derive the color of the whole image from the modified landmark pixels and the grayscale image. Unlike statistics-based representations, landmark-based color representations can be used readily to reconstruct the corresponding color images.

In this paper we focus on applying the landmark-based representation to the problems of color transfer and color correction. Color transfer can be thought of as a general form of color correction that tries to adjust one image's color according to the color characteristics of another example image. Reinhard *et al*. [12] present a simple and efficient method for color transfer. Their idea is based on converting RGB color space into $l\alpha\beta$ color space [15], which minimizes the correlation between channels and can better accounts for color perception. The approach of Reinhard *et al*. assumes that each of the $l\alpha\beta$ channels is a Gaussian distribution, and the histogram matching process of transferring mean and variance is performed separately on each channel owing to the uncorrelated axes of $l\alpha\beta$ color space. This approach is considered global since all pixels are modeled by a single Gaussian, and Reinhard *et al*. propose to include manually selected *swatches* in both images to introduce local histogram matching. In contrast to global color transfer, Tai *et al*. [17] present an EM scheme to estimate the color distribution as a Gaussian mixture model. Their approach allows local color transfer between different color regions in images. However, the EM algorithm of [17] is time consuming; it takes several minutes to converge for an image of medium size.

Instead of modeling the color distribution of an image as a single Gaussian or a mixture of Gaussians, some approaches try to matching the color histograms directly. The N-dimensional pdf transfer algorithm proposed by Pitié *et al*. [10] achieves multi-dimensional histogram matching by projecting the color distributions onto various axes and performing 1D histogram matching on the projected marginal pdf iteratively. Morovic and Sun [8] present another way of direct color histogram matching based on the *earth mover's distance* (EMD) [14]. The major disadvantage of the approaches of [10] and [8] is that pixels of similar color are sometimes assigned to very dissimilar colors, resulting in significant grain artifacts or over-saturated outputs. To address this problem, Pitié *et al*. [11] further present a variational technique that manipulates the gradient field, such that the color correlations between neighboring pixels in the input image can be maintained. The additional process increases the complexity of the original N-dimensional pdf transfer algorithm, but is quite effective for reducing grain artifacts.

In this work we integrate EMD with the landmark-based representation for color transfer. Our approach derives the sets of landmark pixels from the input image and the example image, and uses EMD to transfer two sets of landmark pixels. Since we require that the sets of landmark pixels to be small, the computational cost of applying EMD is quite low. Furthermore, as mentioned previously, the landmark-based representation accounts for the spatial and color characteristics in an image, and thus our approach does not cause the problem of grain artifacts observed in [8] or [10].

It is also worth noting that if the two images for color transfer have a strong structural correspondence but a very weak color resemblance, our landmark-based representation still perform well while all of the aforementioned methods may not work without substantial modifications. For example, the color correction scenario presented in [5] is about transferring the color of a correctly-exposed but blurred image to an under-exposed image. The structures of such two images should be roughly the same, but the color distributions would widely differ. We apply the landmark-based representation to this type of color-correction problem, and demonstrate the advantage of our approach in handling both spatial and chromatic correspondences between images through the experimental results. We also extend our color transfer method to video, which require more stable transferring results among frames. Our approach, with the landmark-based color representation, produces stable color transfer outputs based on the correspondences between the example image and each input frame, as well as the correspondences between consecutive frames.

2. Landmark-Based Sparse Color Representations

The general problem of deriving a landmark-based color representation from a color image can be stated as follows: Given a color image $\mathbf{I}$, we wish to find an intensity image $\mathbf{Y}$, a small subset X of color landmark pixels of $\mathbf{I}$, and a colorization algorithm ψ taking $\mathbf{Y}$ and X as its input, so that the reconstruction error of the colorization output $\psi(\mathbf{Y}, X)$ with respect to $\mathbf{I}$ is minimized, subject to the constraint that the cardinality of the set X is not greater than c. In other words, we try to solve an inverse problem of colorization, which can be defined more formally by

$$\begin{aligned} \{\mathbf{Y}^*, X^*, \psi^*\} = \ & \underset{\mathbf{Y},X,\psi}{\arg\min} \ \left\|\mathbf{I} - \psi(\mathbf{Y}, X)\right\|_1 , \\ & \text{s.t. } |X| \leq c \, . \end{aligned} \tag{1}$$

The error is measured by the 'element-wise' L_1-norm of the difference between $\mathbf{I}$ and the colorization result $\psi(\mathbf{Y}, X)$, *i.e*., we compute the sum of the absolute differences over all RGB channels of all pixels. Obviously the optimization problem in (1) is ill-posed, and it is too much to ask for all the unknowns. In this work we make the problem tractable by using a specific colorization algorithm and assuming that the intensity image is fixed, and thus we can simply focus on solving for the landmark set X:

$$X^* = \arg\min_{X} \left\|\mathbf{I} - \hat{\psi}(\hat{\mathbf{Y}}, X)\right\|_1 \,, \qquad \text{s.t. } |X| \leq c \,. \tag{2}$$

Several recently proposed scribble-based colorization algorithms, *e.g.* [7] and [18], are quite effective in colorizing an intensity image with the hints given by the user. The approach of Levin *et al.* [7] assumes a locally linear relation between color and intensity. The colorization results can be obtained by solving sparse linear systems with user supplied constraints via scribbles. Another colorization algorithm presented by Yatziv and Sapiro [18] computes the geodesic distances between pixels, and determines the color of each pixel by blending the chrominances of observed pixels according to the geodesic distances. Here we choose to use the algorithm of Levin *et al.* as the colorization algorithm $\hat{\psi}$ for our problem, although the algorithm of [18] should be equally applicable.

To obtain a suitable intensity image, we use the standard function for converting RGB color space to YUV color space. The intensity image comprising the luminance values is determined by $\hat{\mathbf{Y}} = 0.2989\,\mathbf{R} + 0.5870\,\mathbf{G} + 0.1140\,\mathbf{B}$.

2.1. Selecting Landmark Pixels

In this section, we present an iterative algorithm for selecting a small set X of landmark pixels. Each landmark pixel in X is associated with its RGB values and the position. Since we assume that $\hat{\mathbf{Y}}$ and $\hat{\psi}$ in (2) are known, given any subset of pixels in the input color image $\mathbf{I}$, we may use this subset as color scribbles and perform the algorithm $\hat{\psi}$ to get a colorization result. If we just arbitrarily choose a subset of pixels as the scribbles, the colorization results will not look like the input image $\mathbf{I}$. We describe below an iterative algorithm that can determine X for the color image $\mathbf{I}$ and is able to achieve low reconstruction error.

We build a Gaussian pyramid $\mathbf{I}_0, \mathbf{I}_1, \ldots, \mathbf{I}_d$, where $\mathbf{I}_0$ is the input color image of full resolution and $\mathbf{I}_d$ is the coarsest level in the pyramid. We also build a Gaussian pyramid $\hat{\mathbf{Y}}_0, \hat{\mathbf{Y}}_1, \ldots, \hat{\mathbf{Y}}_d$ from the intensity image $\hat{\mathbf{Y}}$.

At each level k, we compute the residue image $\mathbf{E}_k$ by

$$\mathbf{E}_k = \left[\mathbf{I}_k - \hat{\psi}(\hat{\mathbf{Y}}_k, X_k)\right]_{\mathsf{Abs\_RGB}} \,, \tag{3}$$

where the operation $[\cdot]_{\mathsf{Abs\_RGB}}$ is to compute, at every pixel, the sum of absolute values over RGB channels. Therefore, each pixel of $\mathbf{E}_k$ keeps a residual value that represents its reconstruction error at the current level k. Some examples of $\mathbf{E}_k$ are shown in Fig. 1.

The residue image $\mathbf{E}_k$ is then divided into blocks. Each block is of size $h \times h$ pixels. We perform a simple step of kernel density estimation by carrying out kernel smoothing on $\mathbf{E}_k$ and obtain $\bar{\mathbf{E}}_k$. From each block of $\bar{\mathbf{E}}_k$ we pick

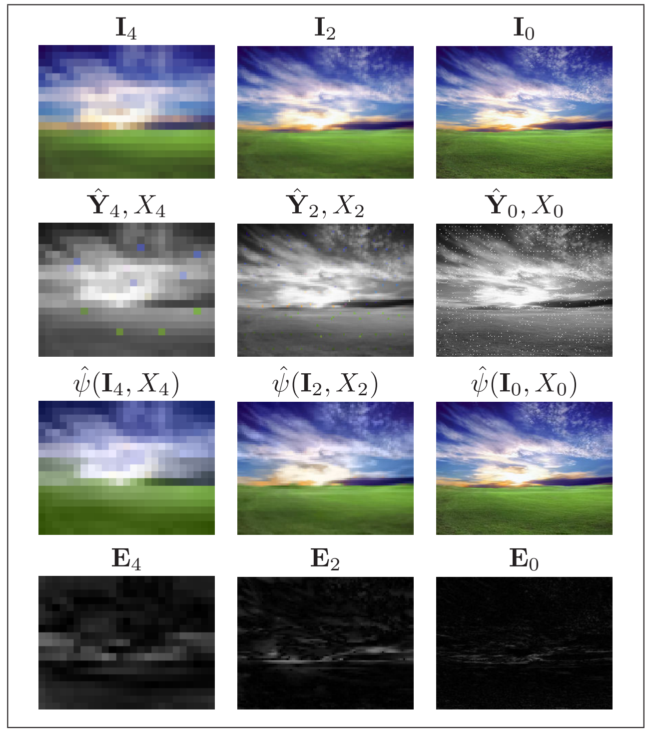

Figure 1. Deriving a landmark-based representation. Note that we use a coarse-to-fine scheme, and hence $\mathbf{E}_0$ is the residue image for the last iteration. (The intensity values of the landmark pixels in '$\hat{\mathbf{Y}}_0, X_0$' are enhanced for better visualization.)

the pixel that has the maximum value, and add it into the new candidate set of landmark pixels if its residual value is larger than threshold θ_k. The pixels which are chosen into the candidate set would approximately correspond to the local maxima of the residual surface. We compute the union of the new candidate set and the landmark set X_k of the current level, and denote it by X'_k. When moving up to the next finer level, the positions of landmark pixels in X'_k must be upsampled accordingly: At level $k-1$, we have the landmark set $X_{k-1} = [X'_k]_{\uparrow}$, where $[\cdot]_{\uparrow}$ denotes the operation of upsampling in spatial domain. Hence the landmark set X_k at level k for running the colorization algorithm can be derived from the previous level via appropriate upsampling. However, in the beginning of the selection process, the initial landmark set X_d of the coarsest level needs to be determined directly: We perform k-means on $\mathbf{I}_d$ and choose the centers as the initial landmark pixels.

The whole process of selecting landmark pixels is performed iteratively, from the coarsest level ($k = d$) to the level of the original resolution ($k = 0$). At each iteration we use the aforementioned method to expand the set of landmark pixels. Finally, at the finest level, we are able to obtain the set of landmark pixels for the input image $\mathbf{I}$ as $X = X'_0$. Fig. 1 shows an example of deriving the landmark-based representation from a color image. As can be seen in the intermediate results, the reconstruction error decreases during the process of selecting landmark pixels.

2.2. Deciding the Block Size w.r.t. the Sparsity Constraint

One of our goal is to find a sparse color representation for an input color image. We introduce the sparsity constraint $|X| \leq c$ in the optimization problem (2) to favor a smaller set of landmark pixels. The constraint can be satisfied through our landmark selection process if we choose a suitable block size (parametrized by h). Consider an input image of size $M \times N$ pixels. Assume that we build a Gaussian pyramid with a scale factor of 2, and the coarsest level in the pyramid is d. Based on our selection criterion, at level k we will add at most $\lfloor M/(2^k h) \rfloor \times \lfloor N/(2^k h) \rfloor$ landmark pixels, since the worse case is to pick one pixel from every block. Consequently, the total number of selected landmark pixels is bounded by

$$\begin{aligned} |X| &= \sum_{k=0}^{d} \left\lfloor \frac{M}{2^k h} \right\rfloor \times \left\lfloor \frac{N}{2^k h} \right\rfloor \\ &\leq \frac{MN}{h^2} \sum_{k=0}^{d} 4^{-k} = \frac{4MN}{3h^2} \Big(1 - \frac{1}{4^{d+1}}\Big) \end{aligned} \tag{4}$$

Because we wish to ensure $|X| \leq c$, we have the following condition:

$$h \geq \sqrt{\frac{4MN}{3c}\Big(1 - \frac{1}{4^{d+1}}\Big)}\,. \tag{5}$$

Given some value c, we may choose a suitable block size based on the condition, and the constraint $|X| \leq c$ will necessarily be satisfied.

3. Transferring Landmark Pixels

Our specific interest in this paper is to apply the landmark-based representation to color transfer and color correction. Different applications will require different settings, but they can all be modeled in the same approach with the proposed representation. One of the key issue in our approach is how to transfer one set of landmark pixels to another. We achieve this goal by using the earth mover's distance (EMD). Details about the analysis and the computation of EMD can be found in [14]. In the following we describe the idea of using EMD to decide the mapping of landmark pixels for color transfer.

3.1. The Earth Mover's Distance for Landmark Matching

The EMD algorithm is to solve the transportation problem of moving a given amount of goods from suppliers to consumers. In our case the problem can be formulated as matching the landmark pixels from one image to another. Let $X = \{\mathbf{x}_i\}$ and $Z = \{\mathbf{z}_j\}$ be two sets of landmark pixels extracted from two images. Each landmark pixel $\mathbf{x}_i$ (or $\mathbf{z}_j$) is represented by a five-dimensional vector comprising three components for RGB values and two components for pixel position. We define the cost (ground distance) of matching $\mathbf{x}_i$ to $\mathbf{z}_j$ as the Mahalanobis distance $\sqrt{(\mathbf{x}_i - \mathbf{z}_j)^T \mathbf{\Sigma} (\mathbf{x}_i - \mathbf{z}_j)}$, where the diagonal matrix $\mathbf{\Sigma}$ gives biases on different dimensions. Suppose that X is selected from the source image and Z from the target image. Since we want every landmark pixel in X can be matched to one landmark pixel in Z, we require that the number of target landmark pixels must be greater than the number of source landmark pixels. This requirement can be achieved alternatively by associating each landmark pixel with a weight w. Specifically, we have $S_X = \{(\mathbf{x}_i, w_{\mathbf{x}_i})\}$ and $S_Z = \{(\mathbf{z}_j, w_{\mathbf{z}_j})\}$, where $w_{\mathbf{x}_i} = 1$ and $w_{\mathbf{z}_j} = \max\{\lceil |X|/|Z| \rceil, 1\}$. The EMD algorithm can be used to solve the matching problems under this setting, and is able to produce the minimum-cost mapping from X to Z.

3.2. Image Color Transfer

By image color transfer we mean that, given an input image and a reference image, we wish to adjust the input image's color based on the color characteristics of the reference image. We consider the input image to be the source and the reference image to be the target. We extract the landmark pixels from each image, and use EMD to solve for the matching from the source to the target. The landmark pixels of the input image are then modified according to the matching result: The color components of each landmark pixel in the input image are replaced by the color of the corresponding landmark pixel in the reference image, while the position components remain unaltered. We then use the modified landmark pixels and the intensity image of the input image to perform colorization $\hat{\psi}$, and produce the resulting color image.

3.3. Color Correction with Paired Images

In this application we have two images of the same scene as the input. One of them is under-exposed and the other is well-exposed but blurred due to camera shake. We apply standard feature-based image registration technique (SIFT and RANSAC [1]) to do the alignment. Although the under-exposed image does not have much color information, it contains most details and can provide useful intensity information. We compute the intensity image of the under-exposed image and apply contrast enhancement to it. We denote the resulting intensity image by $\hat{\mathbf{Y}}_u$. On the other hand, since the blurred image contains the correct color characteristics, we may extract the landmark pixels from the blurred image to get color cues. The position components of landmark pixels need to be adjusted according to image registration, and the aligned landmark pixels are denoted by X_b. Color correction is then achieved by performing the colorization algorithm to get the output $\hat{\psi}(\hat{\mathbf{Y}}_u, X_b)$.

3.4. Video Color Transfer

The task of video color transfer is to adjust the color of a sequence of image frames by referring to the color characteristics of a single reference image. Instead of independently applying image color transfer to each image frame, we propose to perform color transfer via propagating landmark pixels. We compute the mapping from the first frame to the reference image as in image color transfer. For the second frame, we compute the mapping of its landmark pixels to the first frame. Likewise, we can determine the mapping from every subsequent frame to its preceding frame, and follow the mapping to trace back to the first frame. In this way every frame is able to match its landmark pixels to the reference image through the first frame. We then perform the colorization algorithm to get the output. The propagation scheme is helpful for producing stable results of color transfer.

4. Experimental Results

Fig. 2 shows two experimental results of color transfer. We compare our approach with the methods of Reinhard *et al*. [12] and Pitié *et al*. [10]. The color transfer results produced by [12] look distorted because the color distributions in the reference images are non-Gaussian. The results generated by [10] contain grain artifacts, due to the mapping from highly peaked 1D distributions to flatter ones. Our approach produces more natural results; the results are closer to what we would expect if we are asked to assign a new color to each pixel in the input image according to the reference image. Furthermore, the results present correct contrast and do not contain grain artifacts, since the image intensities are unchanged. The size of input images and reference images are about 640×480 pixels. The computation time of the method of [12] is less than 3 seconds for each result. The method of [10] takes 106 seconds and 98 seconds, while ours takes less than 50 seconds for each result. All the experiments are done in Matlab on a Core 2 2GHz PC. We require that $|X| \leq 1000$, so each image has at most 1000 landmark pixels. In Fig. 3a we present two examples of color correction with blurred and under-exposed images. Our algorithm is able to produce visually pleasing results with correct color and good sharpness. In Fig. 3b we show an example of video color transfer. The results are quite stable owing to the propagation scheme with landmark pixels.

5. Conclusion

The landmark-based color representation presented in this paper is a new type of image representation for color processing. It effectively encodes chromatic and spatial information of a color image in a form that the original image can be recovered from the selected landmark pixels and the underlying intensity image. We have shown that this representation is useful for the applications of color transfer and color correction. The ability of recovering the color image from the landmark-based representation also makes our approach preferable to statistics-based representations. We are trying to apply the approach to other tasks including co-segmentation and de-noising, and to extend such kind of reversible representation to other image cues besides color.

Acknowledgments. This research is supported in part by NSC grants 96-2221-E-007-132-MY2 and 98-2221-E-007-083-MY3.

References

[1] M. Brown and D. G. Lowe. Recognising panoramas. In *ICCV*, pages 1218–1227, 2003.

[2] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *ICCV*, pages 1197–1203, 1999.

[3] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR (2)*, pages 142–149, 2000.

[4] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih. Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3):245–268, 1999.

[5] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum. Bayesian correction of image intensity with spatial consideration. In *ECCV (3)*, pages 342–354, 2004.

[6] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.

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

[8] J. Morovic and P.-L. Sun. Accurate 3d image colour histogram transformation. *Pattern Recognition Letters*, 24(11):1725–1735, 2003.

[9] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *ECCV (1)*, pages 661–675, 2002.

[10] F. Pitié, A. C. Kokaram, and R. Dahyot. N-dimensional probablility density function transfer and its application to colour transfer. In *ICCV*, pages 1434–1439, 2005.

[11] F. Pitié, A. C. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1-2):123–137, 2007.

[12] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001.

[13] C. R. Rosenberg, M. Hebert, and S. Thrun. Color constancy using kl-divergence. In *ICCV*, pages 239–246, 2001.

[14] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

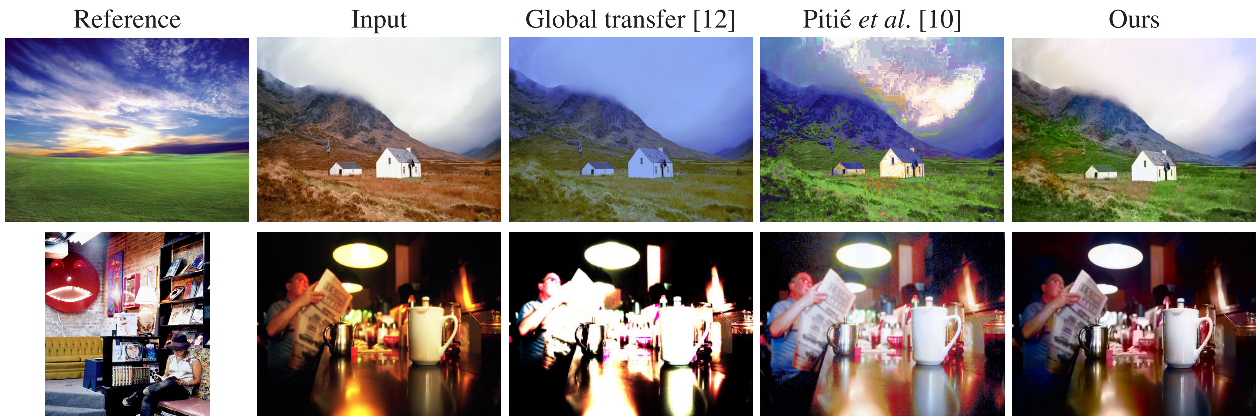

Figure 2. Experimental results of color transfer.

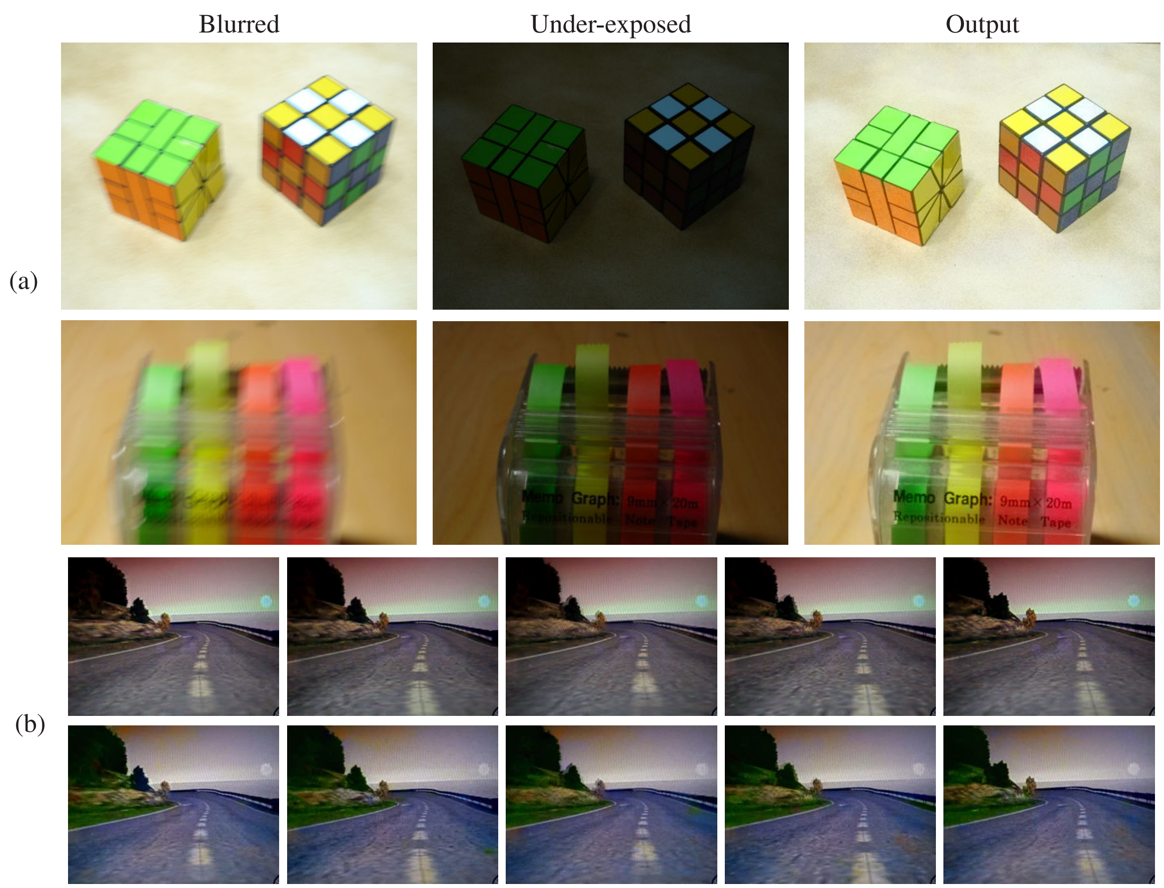

Figure 3. (a) Two examples of color correction using a pair of blurred and under-exposed images. (b) Video color transfer. The reference image for transferring color to the image sequence is the top-left image in Fig. 2.

[15] D. L. Ruderman, T. W. Cronin, and C. chin Chiao. Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A*, 15:2036–2045, 1998.

[16] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[17] Y.-W. Tai, J. Jia, and C.-K. Tang. Local color transfer via probabilistic segmentation by expectation-maximization. In *CVPR (1)*, pages 747–754, 2005.

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