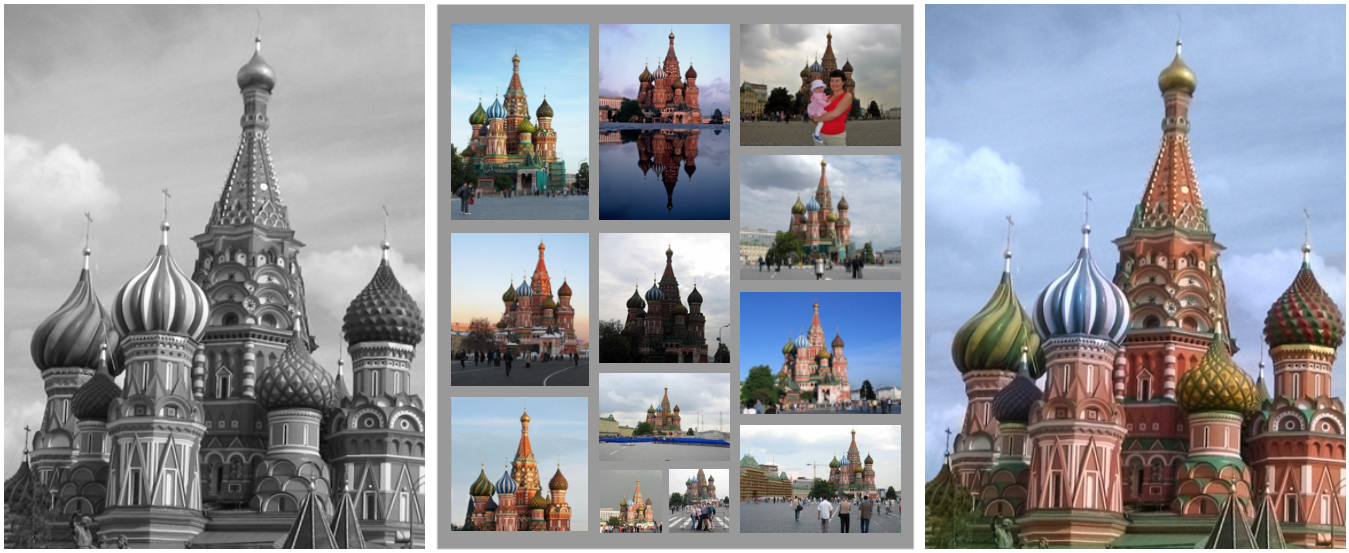
Intrinsic Colorization

Xiaopei Liu[1] Liang Wan[1,3] Yingge Qu[1] Tien-Tsin Wong[1] Stephen Lin[2] Chi-Sing Leung[3] Pheng-Ann Heng[1]
[1]The Chinese University of Hong Kong [2]Microsoft Research Asia [3]City University of Hong Kong



(a) Target grayscale image (b) Reference images from Internet (c) Colorized result

Figure 1: *Colorization of St. Basil's Cathedral: To colorize the target image in (a), we utilize the references searched from the Internet in (b). After recovering the intrinsic color of the target, we colorize to obtain the result in (c) without the influence of illumination and dynamic objects in the references.*

Abstract

In this paper, we present an example-based colorization technique robust to illumination differences between grayscale target and color reference images. To achieve this goal, our method performs color transfer in an illumination-independent domain that is relatively free of shadows and highlights. It first recovers an illumination-independent *intrinsic reflectance image* of the target scene from multiple color references obtained by web search. The reference images from the web search may be taken from different vantage points, under different illumination conditions, and with different cameras. Grayscale versions of these reference images are then used in decomposing the grayscale target image into its intrinsic reflectance and illumination components. We transfer color from the color reflectance image to the grayscale reflectance image, and obtain the final result by relighting with the illumination component of the target image. We demonstrate via several examples that our method generates results with excellent color consistency.

Keywords: colorization, intrinsic images

1 Introduction

The need to colorize classical black-and-white (grayscale) movies and photographs has driven several works in recent years. A state-of-the-art technique performs this task by propagating user-provided color scribbles on the grayscale target image to the rest of the image using a color optimization process [Levin et al. 2004]. Scribbling, however, can be tedious for images with complex details, and requires some skill to obtain natural-looking results. In [Irony et al. 2005], a method is presented for automatically generating such scribbles from an example image provided by the user.

Although the use of an example image can save considerable labor, the quality of the result depends heavily on the choice of example. In particular, significant colorization errors may arise when the illumination condition of a reference image differs from that of the target. These errors arise from the basic assumption in previous colorization works that similarities (or differences) in grayscale intensities indicate similarities (or differences) in colors. Intensity disparities due to differences in shadows and highlights between a reference and a target image can therefore mislead colorization algorithms. In this paper, we propose a technique called "intrinsic colorization" to address this problem of illumination inconsistency between target and reference images.

The key idea of our method is to reduce the effects of illumination in the target and reference images prior to color transfer, and then reintroduce the illumination of the target image in the final result. To compute an illumination-independent reference image, we perform intrinsic image decomposition, which separates an image into a reflectance (albedo) component and an illumination (shading) component. Intrinsic image estimation is an ill-posed problem, since at each scene point there are two unknowns (reflectance and illumination) for each image measurement. So in a scheme similar

to [Weiss 2001], we utilize multiple images of the target scene to obtain a more reliable decomposition. Using grayscale versions of the reference images together with the grayscale target image, we also compute the intrinsic images of the target photograph. Colors from the color reference reflectance image are then transferred to the grayscale target reflectance image at pixels with high confidence in the reference decomposition result. These transferred reflectance colors are used as scribbles that are propagated using the method in [Levin et al. 2004], and then the illumination image of the target photograph is factored back in to generate the colorization result.

Fig. 1(c) shows our colorization result using the reference images shown in (b). Note that our method can correctly colorize the image without interference from illumination effects such as shadow and specular highlight in the target grayscale image. A unique advantage of our technique is its ability to colorize old photographs taken many years ago, even though the photographs are noise-contaminated and of poor quality.

The wide availability of online images provides a base for our method to obtain multiple reference images of a target scene. With several color references registered to the target grayscale image, our system can compute intrinsic images more reliably. Although differences in vantage points between reference images and the target image may allow only partial registration, we present a technique to jointly use all of the available information for intrinsic image decomposition. The use of web resources makes the proposed system especially convenient for automating the colorization of well-known monuments and rigid structures/buildings. We demonstrate its effectiveness via multiple examples.

2 Previous Work

Existing work on colorization can be roughly divided into scribble-based colorization and example-based colorization.

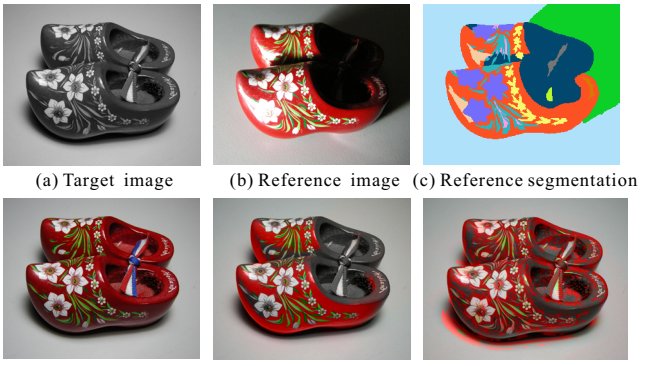
Scribble-based colorization This class of techniques performs colorization based on scribbles placed by users onto the target grayscale image. The classical work, called color optimization, was proposed by Levin et al. [2004]. It optimizes the color of all image pixels with the scribbles as constraints. Later, Huang et al. [2005] prevented the color from bleeding over object boundaries by using adaptive edge detection. Yatziv et al. [2006] proposed a faster scribble-based color optimization technique by chrominance blending. However, these methods all require intensive user intervention, especially when the image contains complex structures or is full of textures. To avoid the burden of scribbling over images with complex textures, Qu et al. [2006] and Luan et al. [2007] both employed texture continuity to colorize pattern-intensive manga and natural images, respectively.

Example-based colorization Another class of techniques automates the colorization process by providing an example image. It does not rely on the user's skill or experience to choose suitable colors for a convincing colorization. Inspired by image analogies [Hertzmann et al. 2001] and the color transfer technique of [Reinhard et al. 2001], Welsh et al. [2002] proposed a pixel-based approach to colorize an image by matching swatches between the target grayscale and color reference images. This work was later improved by Irony et al. [2005]. Its basic idea is that regions with similar textures are colorized with similar colors. A segmented reference image is needed for the purpose of texture matching. The texture matching technique used by Irony et al. was later improved by Schnitman et al. [2006].

Our work can be classified as example-based colorization. Instead of making use of a single reference image, our method utilizes multiple reference images from the Internet for colorization in order to avoid problems arising from illumination differences. Approaches based on Internet images have recently been used for different applications ([Snavely et al. 2006], [Hays and Efros 2007], [Lalonde et al. 2007]). We regard Irony's method [2005] as the state-of-the-art technique in example-based colorization and compare our work with theirs in this paper. Note that existing techniques do not account for the illumination, which may severely degrade the quality of colorization.

3 Illumination Inconsistency

In example-based colorization, the goal is to give objects in the grayscale image the same intrinsic colors of corresponding objects in the reference image, while maintaining its own original illumination. This is a challenging task, since what we observe in images is not the intrinsic reflectance colors, but the combined appearance of reflectance and illumination, including shadows and highlights. Transfer of *image colors* that include these illumination effects instead of *reflectance colors* can therefore lead to erroneous colorization results when the illumination is inconsistent between the reference and target images. We illustrate this problem in Fig. 2 where two different methods are used to transfer color from the reference image in (b) to the target image in (a). Note that the illumination of (a) and (b) are substantially different even though the views are exactly the same. In (e), we directly transfer the colors from (b) to (a) in YUV color space by copying U and V channels. Even if the reference and target images have the same pose, the inconsistent illumination conditions introduce unnatural, false or missing colors compared to the ground-truth color image in (d). In (f), we apply Irony's method with reference to the initial segmentation (c) which is manually prepared based on the reference image (b). Although the regions of missing colors are smaller than that of direct color transfer in (e), the inconsistent illumination causes false texture matches and hence misleads the colorization into an erroneous result.



(a) Target image (b) Reference image (c) Reference segmentation

(d) Ground-truth image (e) Direct color transfer (f) Result of Irony's method

Figure 2: *Discrepancies due to illumination inconsistency. Even when the target image (a) and the reference image (b) are perfectly aligned, inconsistent illumination causes severe errors in the colorization result of (e) and (f), where direct color transfer and Irony's method are applied, respectively. The ground-truth color image in (d) is converted to grayscale to obtain the target image in (a).*

For accurate color transfer, the effect of illumination must first be removed. One way to obtain the illumination-independent reflectance colors is to measure the bidirectional reflectance distribution function (BRDF) of points in the scene, which is clearly impractical for colorization applications. In this paper, we propose to extract the intrinsic reflectance component of the scene from a set of reference images using a decomposition method based on [Weiss

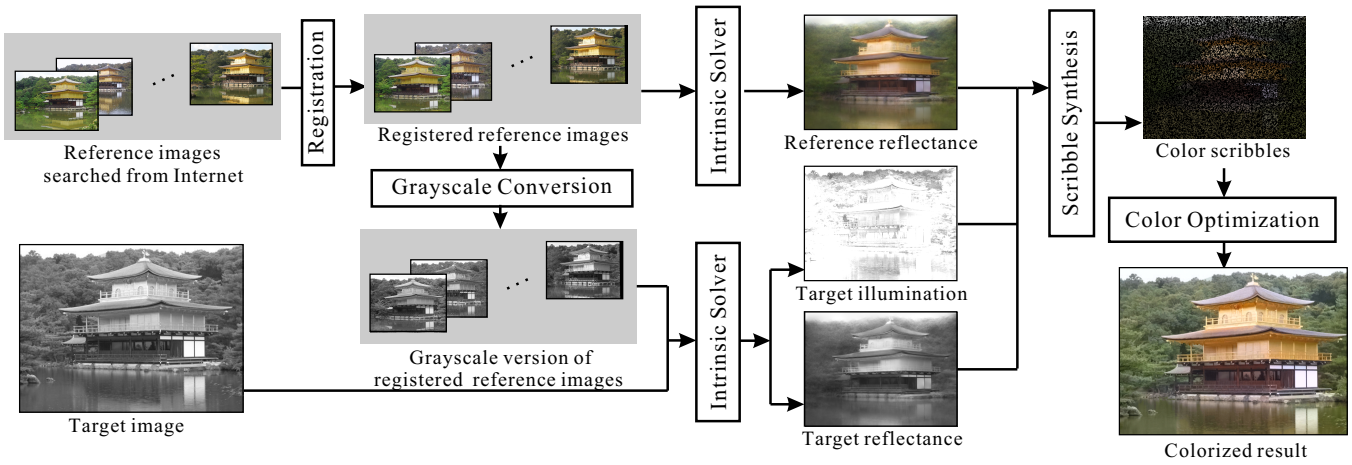


Figure 3: *An overview of intrinsic colorization.*

2001]. This view-dependent intrinsic reflectance image is a mid-level description to represent the underlying surface reflectance, and it is invariant to illumination changes. Extracting intrinsic images is practical for colorization purposes since only a few reference images are sufficient for a reasonable decomposition. We acquire these images from the vast collection of images available on the web. Since we colorize the grayscale target with intrinsic reflectance colors recovered from reference images, we refer to our method as *intrinsic colorization*.

4 Algorithm

The Internet serves as a huge distributed image database that contains multiple images of a given scene captured at different times and by different persons. It is an invaluable source of reference images from which to determine intrinsic reflectance colors for colorization of grayscale images or old photographs. Given a target grayscale image, our system searches for corresponding references from the Internet by attempting to register online images to the target image. Those images that can be registered at least in part are used as references for determining intrinsic colors.

An overview of our system is shown in Fig. 3. From N registered reference images, we solve for a single reference reflectance image of the target scene and N reference illumination images. Similarly, from a set containing the target grayscale image and the N reference images converted to grayscale, we can compute the target reflectance image and the target illumination image from the grayscale target.

Conceptually, colors from the reference reflectance image can then be transferred to the target reflectance image and combined with the target illumination image to obtain the final result. However, since this method may easily suffer from color bleeding (as explained later), we use an alternative but equivalent approach. The color reference reflectance image is first combined with the target illumination image to generate color scribbles which are then used to transfer color to the target image via color optimization. Details of these algorithmic components are presented in the remainder of this section.

4.1 Registration of Reference Images

Given images from the Internet or an image database, we first identify the set of reference images that can be used. Theoretically, using a larger number of reference images with different illuminations will lead to more statistically accurate computation of intrinsic images. We utilize SIFT [Lowe 2004], the state-of-the-art feature matching algorithm, to detect and match corresponding features in grayscale between each reference and the target images. These corresponding features are then used to register each candidate reference image to the target image. The registration is done using RANSAC [Fischler and Bolles 1987] to estimate an eight-parameter homographic projection matrix P from the target to the reference, which in turn is used to compute the re-projection error for all the corresponding features in the candidate reference image. The re-projection error for any feature point can be computed as

$$e(x') = \|x - P^{-1}(x')\| \tag{1}$$

where x and x' are the positions of matched points in the target and reference images, respectively. If most of the feature points have re-projection errors below a predefined threshold ε, we include this image into the reference image set.

Registration can then be done in two ways depending on the quality of feature matching. If most of the feature points in the reference image have re-projection errors below a small threshold ε_0, we perform global alignment using the estimated projection matrix P to transform the entire image. Otherwise, if there are many feature points with relatively large re-projection error, we perform triangle-based warping. We first triangulate the target image based on the feature points using Delaunay triangulation [Shewchuk 1996]. The reference images are then automatically triangulated since a registerable feature point in a reference image corresponds to another feature point in the target image. Folding generally does not occur in the triangulation of a reference image, since the target and reference images have consistent spatial relationship. Although folding is possible when the distance between two connected feature points is less than the matching error, this is generally avoided in our application because of the constraint on re-projection errors. Thus, with this texture-mapped mesh, we can warp the reference image by transforming the feature points to their matches (from SIFT) in the target image.

Some errors may exist in registration, depending on the selected value of ε_0, but a certain amount of registration error (i.e., up to about 10 pixels) can be tolerated by our colorization technique since these errors are not noticeable in color channels. Registration may be degraded by noise in either the target or the reference images. For the cases, such as an old target photograph or a low-quality reference image, the state-of-the-art image denoising algorithm (e.g. [Portilla et al. 2002]) may be applied as a preprocessing step. If denoising does not lead to satisfactory registration, the matching results may be manually adjusted. Manual adjustment

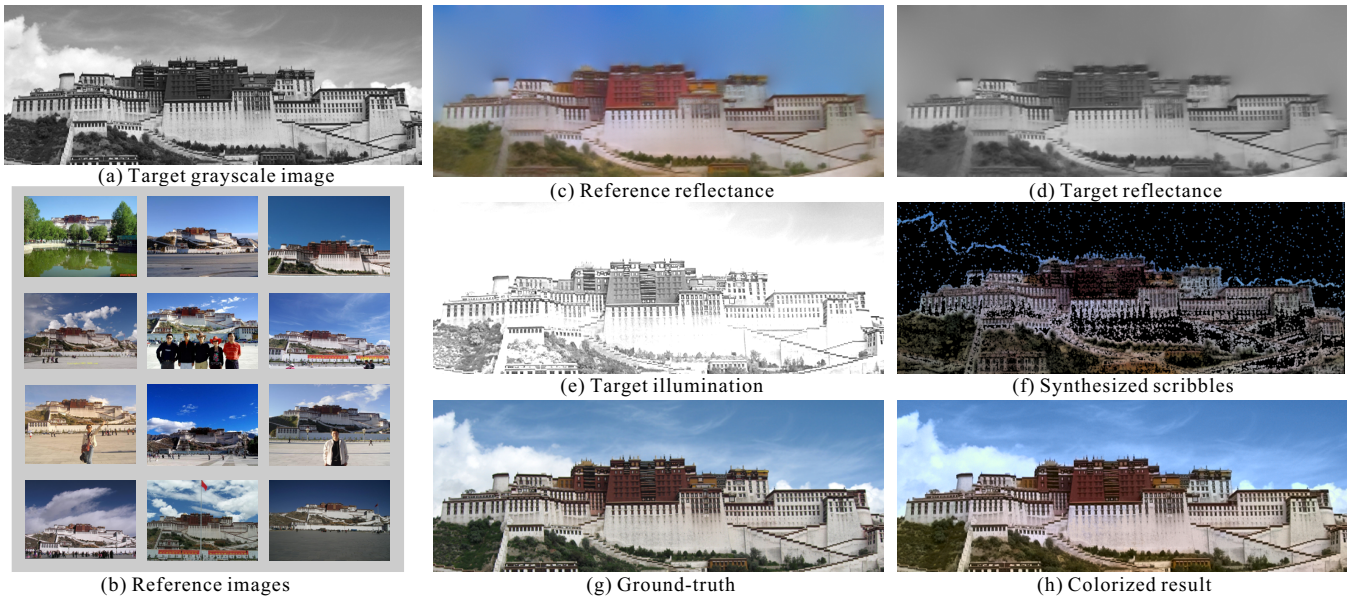


Figure 4: *Intrinsic colorization in action. (a) is the target grayscale image. (b) is a set of references acquired from the Internet. By registering all references and solving for intrinsic images with partial registration, we obtain a reference reflectance image (c), a target reflectance image (d), and a target illumination image (e). Color scribbles are automatically synthesized from (c) and (e), as shown in (f). With color optimization, the result in (h) is obtained. The ground-truth image is shown in (g) for comparison.*

may also be used to deal with false matches that result from significant illumination difference between the target and reference images. After registration, reference images with different resolutions are automatically scaled to match the resolution of the target image. Here, bilinear interpolation is used for scaling.

Once the reference images are registered, each reference may only partially cover the target image. Hence, we associate to each reference image a binary mask M_i to identify pixels that can be registered to the target image. This will be used for the following intrinsic image decomposition.

4.2 Recovering Intrinsic Components

As proposed by Barrow and Tenenbaum [1978], intrinsic images are a view-dependent, mid-level description of a scene. An image is usually decomposed into a reflectance image and an illumination image. Although not making explicit all the physical causes of image features, intrinsic images are very useful for supporting a range of visual inferences. Recovering the intrinsic image from one single image remains a difficult problem since it is highly ill-posed. Although certain progress has been made in single-image decomposition (e.g., Land et al. [1971], Freeman et al. [1998], Tappen et al. [2005]), significant improvement in accuracy and robustness, and simplicity in implementation, can be obtained when multiple images are taken into account. Weiss [2001] solved the decomposition problem for a set of N images with the same view but under different illumination conditions. With the assumption of constant scene reflectance, this method outputs one reflectance image and N illumination images. In our work, we adopt a modified version of this technique that handles images with partial registration.

Deriving Intrinsic Images We first describe the basic approach for intrinsic image decomposition in [Weiss 2001]. Consider N images captured from the same view but under different illumination conditions. All N images are assumed to share the same reflectance. We denote the intrinsic reflectance image by R and each intrinsic illumination image by L_k. Then each reference image can be expressed by the following relationship:

$$I_k(p) = R(p)L_k(p). \tag{2}$$

To linearize the equation, we take the logarithm on both sides of this equation:

$$i_k(p) = r(p) + l_k(p) \tag{3}$$

where logarithm counterparts are denoted in lowercase. To each logarithmic image i_k, we then apply derivative filters $f_x(p)$ and $f_y(p)$ along the x and y directions and get the corresponding filtered output images:

$$o_x(p,k) = f_x(p) \star i_k(p) \tag{4}$$

$$o_y(p,k) = f_y(p) \star i_k(p) \tag{5}$$

Based on the property that most natural images have sparse derivative filter output, we obtain a maximum likelihood estimate of the filtered reflectance derivative images by taking the median values of the filtered output images:

$$\hat{r}_x(p) = \text{median}_k\{o_x(p,k)|k = 1..N\} \tag{6}$$

$$\hat{r}_y(p) = \text{median}_k\{o_y(p,k)|k = 1..N\} \tag{7}$$

To obtain r, we solve the following Poisson equation with zero boundary condition:

$$\triangle r(p) = \nabla \cdot [\hat{r}_x(p), \hat{r}_y(p)]. \tag{8}$$

Substituting r into Eq. 3 gives the logarithmic illumination l_k for each reference image. Finally, inverse logarithm (exponential) is computed to obtain the reflectance image R and the N corresponding illumination images L_k.

Intrinsic Images with Partial Registration In [Weiss 2001], all N images are assumed to be of the same view. Our application differs in that each reference image may be registered to only part of the target image. To solve this partial registration problem, we propose a labeling scheme. With the N mask images obtained in our registration process, we form a label set $L(p)$ for each pixel p in the target image that identifies all reference images whose partial registration includes p. The size of label set $L(p)$ may vary among

different pixels, and is determined by the coverage of different mask images. With these label sets, we modify Eqs. 6 and 7 to

$$\hat{r}_x(p) = \text{median}_k\{o_x(p,k)|k \in L(p)\} \quad (9)$$

$$\hat{r}_y(p) = \text{median}_k\{o_y(p,k)|k \in L(p)\} \quad (10)$$

and then evaluate with the same zero-boundary condition.

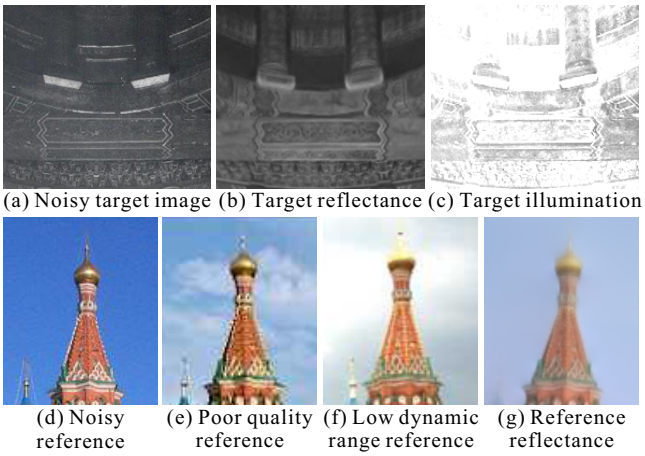
Illumination Separation Before Colorization Our core contribution is to reduce much of the illumination influence before colorization, so that our colorization is not degraded by differences in illumination conditions. To achieve this, we first obtain the illumination-independent color reference reflectance R_r and the illumination-independent grayscale target reflectance T_r.

To compute the reference reflectance image R_r, we take the N registered color references as input and perform our intrinsic image computation with partial registration. This returns a single reference reflectance image R_r, and N corresponding reference illumination images which are not used in our application. Fig. 4(c) shows an example of the reference reflectance image computed from the reference images in Fig. 4(b). The target reflectance image T_r is computed in the same manner except using grayscale version of the reference images together with the grayscale target image. Decomposition then yields $N + 1$ illumination images. Of these, only the one that corresponds to the target image is retained, and we denote this grayscale target illumination image as T_l. Examples of T_r and T_l are shown in Fig. 4(d) and (e), respectively. By transferring colors from R_r to T_r, we colorize the target image *without illumination*. Recombining the colorized target reflectance with the illumination component of the target grayscale illumination image T_l, we finally obtain the colorized target image *with illumination*.

Note that solving the intrinsic components avoids not only the influence of illumination, but also the influence of dynamic occluders in the reference images, since these occluders are essentially "dynamic illumination" that can be removed in the maximum likelihood estimate in Eqs. 9 and 10. This property is demonstrated in Fig. 1, where people standing in front of St. Basil's Cathedral in the reference images do not affect our colorization result. The same property also reduces the influence of noise in both reference and target images, since the same amount of noise does not in general persistently appear at the same location in different reference and target images. We examine this issue in Fig. 5, where (b) and (g) show close-ups of the recovered target reflectance image of Qinian Palace and the reference reflectance image of St. Basil's Cathedral in Fig. 9 (a) and Fig. 1, respectively. Although the target or some of the reference images are noisy, both of the recovered intrinsic reflectance images are smooth. However, noise may still exist in the target illumination if the target image is noisy, as demonstrated in Fig. 5(c). Therefore, our colorized image retains the same amount of noise as the target image, as shown in Fig. 9 (a). Similarly, the same amount of degradation is retained in the colorized result, if the target image is of poor quality or low dynamic range. On the other hand, if a small proportion of the reference images are noise-contaminated, and of poor quality or low dynamic range, these degradations can be regarded as "dynamic illumination" and will not influence the results, as demonstrated in Fig. 5(d) to (g).

4.3 Colorization

With the three intrinsic images (R_r, T_r, and T_l) in hand, we can perform intrinsic colorization. Assuming white illumination in the target image, this could be done by transferring color from the reference reflectance R_r to the target reflectance T_r and then multiplying by the target illumination T_l. Equivalently, we could simply multiply R_r and T_l to obtain the final image. However, errors



(a) Noisy target image (b) Target reflectance (c) Target illumination

(d) Noisy reference (e) Poor quality reference (f) Low dynamic range reference (g) Reference reflectance

Figure 5: *Influence of image degradations on intrinsic image solver. The top row shows the effect when noise exists in the target image (a). The recovered reflectance (b) is smooth, while the noise is present in the illumination image (c), which will lead to a colorization result with the same noise level. The bottom row shows the influence of various image degradations in the references. Even though the references include images with noise (d), compression artifacts (e) and of low dynamic range (f), the intrinsic image solver can still return a nice reflectance image (g).*

in registering the reference images can lead to a blurry color reflectance image R_r, as exemplified in Fig. 4(c). This blur in the reflectance image can seriously degrade colorization if it is not carefully accounted for, as demonstrated in Fig. 6(a) for colorization by direct multiplication of R_r and T_l. Severe color bleeding artifacts are noticeable in the rectangular regions.

Therefore, we do not trust the colors of all pixels in the reference reflectance R_r. Instead, we allow color transfer only for pixels considered to be less affected by misalignment errors. This limited transfer produces color scribbles that are propagated by color optimization [Levin et al. 2004].

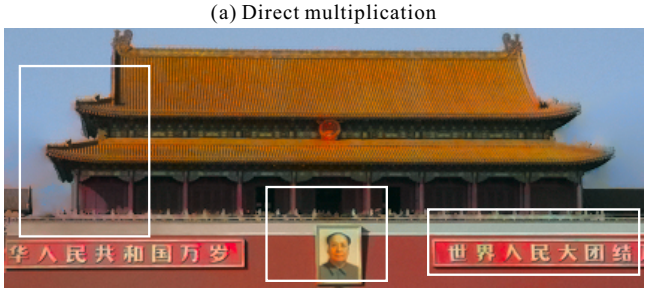
Scribbles could potentially be transferred from R_r to T_r prior to color optimization, and relit with T_l. However, T_r also suffers from blur due to registration errors, so color optimization on T_r also results in color bleeding artifacts, as shown in Fig. 6(b). To address this issue, we instead transfer colors from R_rT_l to T_rT_l, where T_rT_l is equivalent to the original target image, which is unaffected by misalignment blur. The pre-multiplication of the target illumination T_l can significantly reduce color bleeding artifacts in the color optimization. Furthermore, it directly yields the final colorized target image under the target illumination condition. The correctness of this pre-multiplication approach relies on the assumption that the illumination in the reference image is *white* light, as the target illumination T_l is equally applied (R:G:B = 1:1:1) to all three color channels.

Generating Scribbles The color scribbles are automatically generated from the product of reference reflectance R_r and target illumination T_l:

$$C(p) = R_r(p)T_l(p). \quad (11)$$

There may be color error among scattered pixels in the image due to intrinsic image computation on misaligned pixels. To minimize this error influence, we first over-segment the target grayscale image into segments of similar intensities using the mean-shift algorithm [Comaniciu and Meer 2002]. Then for each segment, pixels are uniformly sampled from $C(p)$, with the number of pixels in

(a) Direct multiplication



(b) Optimization on target reflectance

Figure 6: *Color-bleeding artifacts. In (a), colorization is performed by multiplying R_r and T_l. Inconsistent colors appear in the rectangular regions because of misalignment of reference images. In (b), colorization is performed with automatically computed scribbles and color optimization on T_r. Since T_r also suffers from misalignment problems, this result also exhibits color bleeding and inconsistency artifacts. The reference images from Fig. 8(a) are used to compute the two results.*

proportion to the segment size. This sampling is assumed to consist mostly of pixels that are not influenced by misalignment. Next, we compute a weighted average color within each segment:

$$\bar{C}_i = \frac{\sum_{p \in S_i} \omega(p) C(p)}{\sum_{p \in S_i} \omega(p)} \tag{12}$$

where S_i is the i-th segment and the $\omega(p)$ is a Gaussian weight computed by

$$\omega(p) = e^{-k(C(p)-\mu_s)^2/2\sigma_s^2} \tag{13}$$

with μ_s and σ_s denoting the mean and the variance of the segment colors, respectively. We empirically set $k = 1$. Pixel with color $C(p)$ close to $\bar{C}_i$, i.e.,

$$\frac{|C(p) - \bar{C}_i|}{\bar{C}_i} < \epsilon, \tag{14}$$

is considered to be the representative of the segment and less affected by misalignment error. This color is then used as the scribble color. The value of ϵ is set to 0.1 in our implementation. The scribbles are generated in YUV color space and Fig. 4 (f) shows an example of color scribbles computed with this approach.

Color Optimization Having obtained the synthesized color scribbles, we apply color optimization [Levin et al. 2004] to colorize the target grayscale image. This is achieved in YUV color space by minimizing the following quadratic objective function for both U and V channels respectively, with constraints on scribble colors:

$$J(C) = \sum_r (C(r) - \sum_{s \in N(r)} \omega_{rs} C(s))^2 \tag{15}$$

where

$$\omega_{rs} = g e^{-(Y(r)-Y(s))^2/2\sigma_r^2} \tag{16}$$

Examples	Reference No.	Target Image Res.	Time
Fig. 1	44	412×600	3m15s
Fig. 4	46	800×314	4m50s
Fig. 8(a)	34	649×408	2m35s
Fig. 8(b)	21	700×419	2m30s
Fig. 9(a)	8	338×600	1m04s
Fig. 9(b)	12	406×375	2m05s

Table 1: *Timing statistics.*

and $N(r)$ is the set of 4-connected neighbors of pixel r and Y is the target grayscale image. The normalization factor g is set so that the weights sum to one. In Fig. 4(h), we demonstrate a colorization result with this color optimization scheme. We also compare our colorization result to the ground-truth image as shown in Fig. 4 (g).

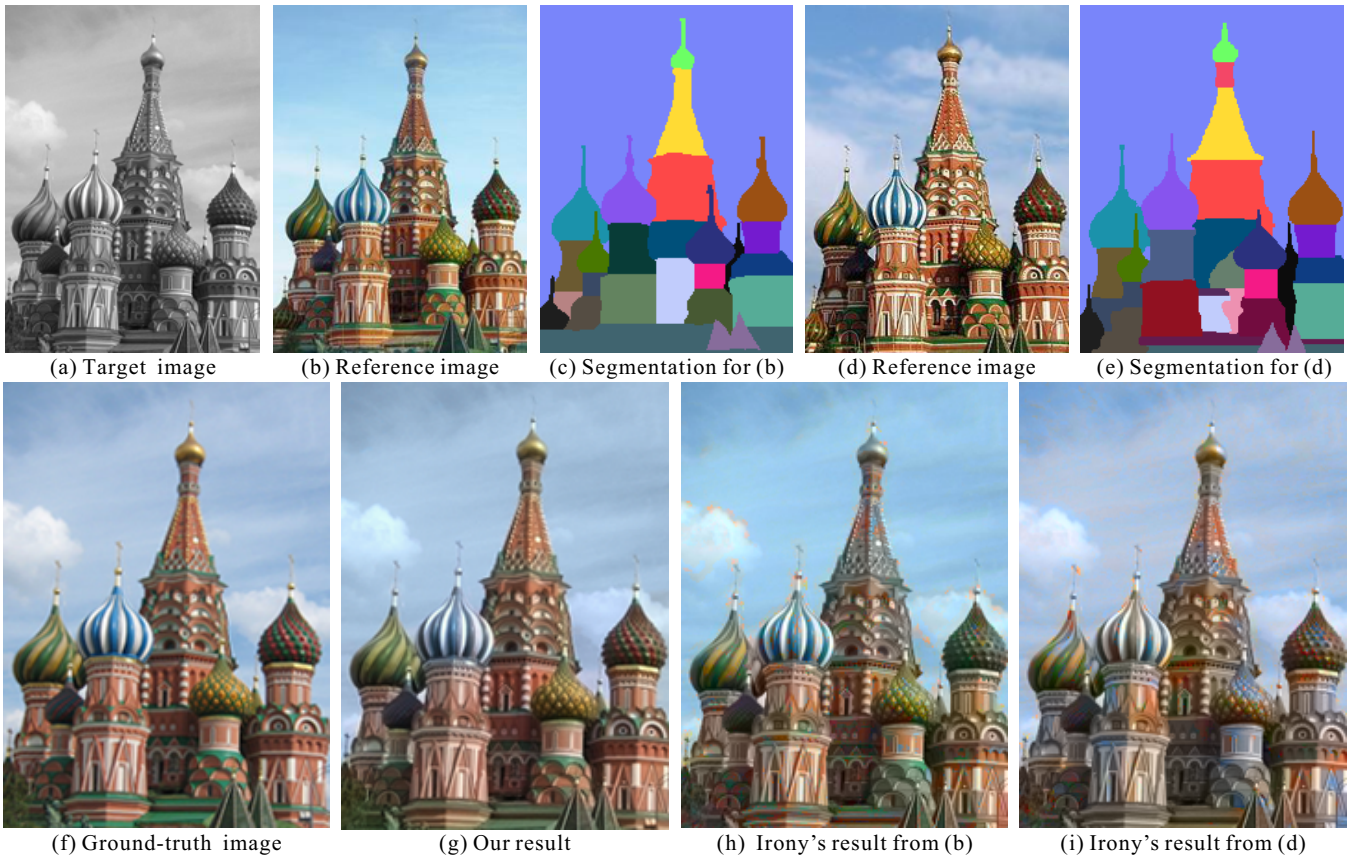
5 Results

Figs. 1, 4, 8 and 9 show multiple colorization results by our method. All references are searched from the Internet and may differ in illumination conditions and/or in the presence of foreground occluders. By decomposing the reflectance images as well as the target grayscale image into reflectance and illumination images, the intrinsic color can be used to reliably colorize the target image.

A unique and direct application of our method is to colorize old black-and-white photographs taken many years ago and of extremely low quality. The "Qinian Palace" and "Temple of Heaven" target images in Fig. 9 are both old photographs captured in the 1950s. These images are severely contaminated by noise. Nevertheless, our method can still obtain reasonable colorization results. Here, we assume the color-to-gray process during the acquisition of these old photographs is consistent to our grayscale conversion of reference images. We also assume that the objects in the target images have maintained their colors over the years.

To demonstrate the stability of our technique on images acquired in different illumination conditions, we compare our colorization result of St. Basil's Cathedral to that of the example-based colorization by Irony et al. [2005] in Fig. 7. We applied Irony's method twice with two different reference images, one with illumination similar to the target image (see (b)) and the other with illumination significantly different from the target image (see (d)). Fig. 7 (h) and (i) show the results using the reference images in (b) and (d), respectively. The one with more similar illumination gives better result, but still contains noticeable false color. On the other hand, our method gives a reasonable and consistent colorization result in (g) by determining the intrinsic colors. The comparison suggests that the illumination difference in reference and target images can substantially affect the quality of colorization.

For most of the examples in our paper, the colorization can be done automatically, depending on the quality of the computed image registration. Automatic image registration, however, may fail in some cases, so users may be needed to adjust the matching results. This is the only user interaction that is required in our colorization process. Table 1 shows the total processing time for colorizing each of the examples on a computer equipped with an Intel Core(TM)2 6400 @ 2.13GHz CPU and 3GB system memory. It depends on the size of the target image as well as the number and the size of reference images. The time for user interaction varies for different examples, which depends on the number of reference images failed in registration. For the example of Temple of Heaven, which has the maximum number of reference images failed in registration, it requires around 5 minutes to adjust the matching results.



(a) Target image (b) Reference image (c) Segmentation for (b) (d) Reference image (e) Segmentation for (d)

(f) Ground-truth image (g) Our result (h) Irony's result from (b) (i) Irony's result from (d)

Figure 7: *Comparison of our method to that of Irony et al. [2005]. (a) is the target grayscale image. (b) is the first reference image for the Irony method. It has a similar illumination as the target. (c) is the corresponding segmentation for (b). (d) is the second reference image for the Irony method. This one has an illumination substantially different from the target, as seen from the highlights and shadows. (e) is the segmentation for (d). The results of the Irony method given (b) & (c) and (d) & (e) are shown in (h) and (i), respectively. Greater differences in illumination lead to more serious colorization error. (g) is our result. (f) is the ground-truth image.*

Limitations As mentioned previously, intrinsic colorization assumes the illumination in all images be white light. For the intrinsic image decomposition to be effective, we also require sufficient reference images be obtained from web search in order to provide enough registerable images with different illuminations. Because of this, our method is intended to colorize well-represented scenes that can be well aligned with web images, and may not properly colorize personal items. In addition, our method is restricted to colorizing static scenes using images of the same scene viewed from similar directions. Content in the target image that is inconsistent to the reference images, or objects in the old photographs that are long gone, may not be colorized properly. Significant view changes among reference images may cause occlusions that impair the registration process. The scene in a target image also may not be fully covered by the set of reference images. Image inpainting techniques could potentially be used to fill in the colors of those uncovered regions.

6 Conclusion

In this paper, we present a novel method to colorize a grayscale image by extracting intrinsic reflectance colors from multiple references. These references are all obtained directly from web search. With previous example-based methods, consistent colorization results are difficult to obtain when illumination conditions between the target grayscale image and its references are different. By reducing the influence of illumination with intrinsic image decomposition, reliable colorization results can be generated.

Our current method assumes that a sufficient number of registerable reference images of the target scene can be acquired from the Internet. In cases when no exact match is available, color information from similar scenes may potentially be used in computing intrinsic reflectance images. Relevant scenes may be identified through texture analysis. Alternatively, single-image methods [Tappen et al. 2005] for intrinsic image decomposition may also be employed to determine reflectance colors of the target scene, when reference images are very scarce.

Acknowledgements

We would like to thank Daniel Cohen-Or and Dani Lischinski for their advice on this project. Thanks to all reviewers for their constructive comments and guidance in shaping this paper. We also would like to thank Google and Baidu for providing the many reference images. This project was supported by the Research Grants Council of the Hong Kong Special Administrative Region, under the RGC Earmarked Grants (Project No. CUHK 417107), and the research grants from City University of Hong Kong (Project No. 7002230).

References

BARROW, H., AND TENENBAUM., J. 1978. Recovering intrinsic scene characteristics from images. *Computer Vision Systems*, 3–26.

Tiananmen

7 out of 34 reference images from Internet

Ground-truth

Colorized Tiananmen

(a) Colorization of Tiananmen

5 out of 21 Reference images from Internet

Saint Michel

Ground-truth

Colorized Saint Michel

(b) Colorization of Saint Michel

Figure 8: *Colorization results for real photographs of various scenes. (a) Tiananmen gate. (b) Saint Michel.*

COMANICIU, D., AND MEER, P. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 5, 603–619.

FISCHLER, M. A., AND BOLLES, R. C. 1987. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Readings in computer vision: issues, problems, principles, and paradigms*, 726–740.

FREEMAN, W. T., AND VIOLA, P. A. 1998. Bayesian model of surface perception. In *NIPS '97: Proceedings of the 1997 conference on Advances in neural information processing systems 10*, MIT Press, Cambridge, MA, USA, 787–793.

HAYS, J., AND EFROS, A. A. 2007. Scene completion using millions of photographs. *ACM Trans. Graph. 26*, 3, 4.

HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. *ACM Trans. Graph.*, 327–340.

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 *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, ACM, New York, NY, USA, 351–354.

IRONY, R., COHEN-OR, D., AND LISCHINSKI, D. 2005. Colorization by example. In *Rendering Techniques 2005*, IEEE Computer Society Press, 201–210.

LALONDE, J.-F., HOIEM, D., EFROS, A. A., ROTHER, C., WINN, J., AND CRIMINISI, A. 2007. Photo clip art. *ACM*

5 out of 8 references from Internet

Qinian Palace

Colorized Qinian Palace

(a) Colorization of Qinian Palace

11 out of 12 references from Internet

Temple of Heaven

Colorized Temple of Heaven

(b) Colorization of Temple of Heaven

Figure 9: *Colorization results for old photographs. (a) Qinian Palace. (b) Temple of Heaven. Both were acquired in the 1950s.*

Trans. Graph. 26, 3, 3.

LAND, E. H., AND MCCANN, J. J. 1971. Lightness and retinex theory. *Journal of the Optical Society of America (1917-1983) 61* (Jan.), 1.

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

LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2, 91–110.

LUAN, Q., WEN, F., COHEN-OR, D., LIANG, L., XU, Y.-Q., AND SHUM, H.-Y. 2007. Natural image colorization. In *Rendering Techniques 2007 (Proceedings Eurographics Symposium on Rendering)*, 309–320.

PORTILLA, J., STRELA, V., WAINWRIGHT, M. J., AND SIMONCELLI, E. P. 2002. Image denoising using gaussian scale mixtures in the wavelet domain. *IEEE Transactions on Image Processing 12*, 1338–1351.

QU, Y., WONG, T.-T., AND HENG, P.-A. 2006. Manga colorization. *ACM Trans. Graph. 25*, 3, 1214–1220.

REINHARD, E., ASHIKHMIN, M., GOOCH, B., AND SHIRLEY, P. 2001. Color transfer between images. *IEEE Computer Graphics and Applications 21*, 5, 34–41.

SCHNITMAN, Y., CASPI, Y., COHEN-OR, D., AND LISCHINSKI, D. 2006. Inducing semantic segmentation from an example. In *Asian Conference on Computer Vision*, 373–384.

SHEWCHUK, J. R. 1996. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. *Applied Computational Geometry: Towards Geometric Engineering 1148* (May), 203–222. From the First ACM Workshop on Applied Computational Geometry.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3d. *ACM Trans. Graph. 25*, 3, 835–846.

TAPPEN, M. F., FREEMAN, W. T., AND ADELSON, E. H. 2005. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 9, 1459–1472.

WEISS, Y. 2001. Deriving intrinsic images from image sequences. In *Eighth IEEE International Conference on Computer Vision*, vol. 2, 68–75.

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

YATZIV, L., AND SAPIRO, G. 2006. Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing 15*, 5, 1120–1129.