

CONSISTENT TONAL CORRECTION FOR MULTI-VIEW REMOTE SENSING IMAGE MOSAICKING

Menghan Xia, Jian Yao,* Li Li, Renping Xie, Yahui Liu

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, P.R. China
Email: jian.yao@whu.edu.cn Web: <http://cvsr.whu.edu.cn/>

Commission III, WG III/5

KEY WORDS: Color Transfer, Gain Compensation, Histogram Adjustment, Consistent Tonal Correction, Optimal Adjusting Order

ABSTRACT:

In this paper, we propose an effective approach for consistent tonal correction of multi-view images during mosaicking. Our method is specifically designed for mosaicking multi-view remote sensing images acquired under different conditions and/or presenting inconsistent tone. To avoid the correlation of three channels in original *RGB* images, we convert them to an orthogonal color space $l\alpha\beta$ in advance. First of all, the tones of sequential images are transferred from an example image reasonably via our improved color transfer algorithm. Secondly, the more refined adjustments take place in the luminance channel l and color channels α and β , independently. In the luminance channel, the global gain compensation is applied to minimize the luminance difference between pairs of images by the least square estimator. In the color channels, the specifically designed stepwise histogram adjustments make all the images consistent tone as a whole, including the initial correction transferring the color characteristics of the automatically selected reference subset to other images in an optimal order and the consistent correction readjusting each image by referring all their neighbors based on the overlaps. Thirdly, we creatively transfer the original structures to the previously corrected images by a local linear model, which can preserve the local structures of the original images. Finally, several groups of convincing experiments on both challenged synthetic and real data demonstrate the validity of our proposed approach.

1. INTRODUCTION

In the last decades, the high spatial resolution (HR) remote sensing images have been widely used for the applications in many aspects, such as environmental surveillance, resource management, and geographic mapping. For many applications, image mosaicking is a routine procedure which merges two or more images with overlapping areas into a single composite image as seamless as possible in both geometry and color tone. The basic steps of image mosaicking are comprised of geometric alignment, tone correction, seamline searching, and feathering or blending. All these steps are necessary for generating a consistently mosaicked image with least geometric and radiation deviations. However, compared to the other major steps, tone correction has received less attention. Although with a similar effect to tone correction, image blending alone can't obviously eliminate all the color difference between images under grave situations, as the example shown in Figure 1. Hence researches about the tonal correction algorithms are becoming more and more necessary recently.

In the remote sensing field, most of the works on solving the tonal difference for multi-view mosaicking are radiometric normalization (or gain compensation) (Canty et al., 2004, Canty and Nielsen, 2008), and others often take relatively simple treatments (Li et al., 2015). Although these algorithms work for some cases, it may fail to completely compensate for color difference between different views when the lighting conditions vary dramatically. In image processing and computer graphics communities, lots of color manipulation methods have been developed in recent years. These approaches can be divided into two categories : parametric and non-parametric at a high level (Xu and Mulligan, 2010).

Approaches based on the transformation model are parametric,

*Corresponding author

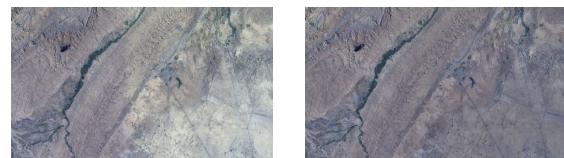


Figure 1. An example illustrating that blending can't obviously eliminate the tonal difference alone. The left mosaicked image was obtained directly by a multi-band blending technique. The right mosaicked image was generated by our proposed tonal correction followed by a multi-band blending technique.

which assume the color relation between images can be described by a transform: $\hat{\mathbf{I}} = f(\mathbf{I})$, where f stands for any transformation equation for a color vector \mathbf{I} . (Reinhard et al., n.d.) were pioneers in establishing the concept of color transfer, with an approach to modify the color distribution of the original image based on the global color statistics of an example image in the decorrelated color space $l\alpha\beta$. Their work has been widely used as the baseline approach by other approaches. To operate color in the *RGB* space directly, (Xiao and Ma, 2006) proposed an ellipsoid mapping scheme. (Ilie and Welch, 2005) proposed to use a general polynomial transform to correct the color vector in the *RGB* space. In order to get a more accurate mapping relation, (Hwang et al., 2014) corrected each pixel by an independent affine model, which is the solution of probabilistic moving least square based on point matching. (Tai et al., 2005) proposed a local color transfer scheme based on probabilistic image segmentation and region mapping using Gaussian mixture models (GMM) and the expectation-maximization (EM) algorithm. (Xiang et al., 2009) improved this work in the case that multiple source images are available for selection. By non-rigid dense correspondence, the algorithm proposed by (HaCohen et al., 2011) can handle shared content under non-rigid deformations and color variations, and compute a parametric color transfer model between pairs of

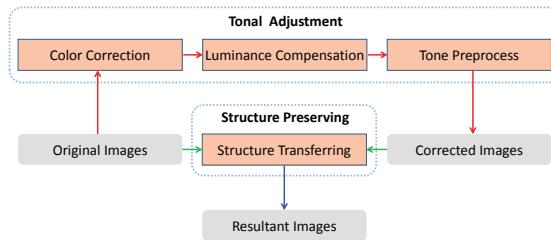


Figure 2. The flowchart of our tonal correction algorithm.

images.

Non-parametric approaches assume no particular model for the color mapping, and most of them use a look-up table to directly record the mapping of the full range of color levels. To simplify the look-up table, (Yoo et al., 2013) proposed to search the major colors in both original image and example image by clustering and build mapping relation between major colors through a defined similarity metric. Actually, the look-up table is often replaced by some kind of curve, like Gamma curve, S-curve and B-splines, etc. (Moulou et al., 2013) utilized the intensity values in the quantiles of histograms to depict the mapping relation, which are optimized globally as a convex problem. The algorithm developed by (HaCohen et al., 2013) can achieve the consistent appearance of connected photos by globally optimizing a quadratic cost function, which is based on the mapping curve as quadratic spline with 7 nodes. Furthermore, (Pitié et al., 2005, Pitié et al., 2007) proposed an N -dimensional probability density function (PDF) transfer approach to reduce the high-dimensional PDF matching problem to the one-dimensional PDF matching by Radon Transform.

Although the above mentioned approaches have solved some key problems arising in color transfer effectively and made practical effects, they are only feasible for two or several images, because of the processing complexity or the specificity for some situations. In this paper, we aim to propose an approach which is designed for the tonal correction of a remote sensing image sequence containing dozens or even hundreds of images. Firstly, all the input images are preprocessed by a local transformation according to the statistical parameters of an example image in each channel in the $l\alpha\beta$ space, independently. Then, the global gain compensation is applied to minimize the luminance difference among all images. Besides, the specifically designed stepwise histogram adjustments are combined to eliminate the color differences between pairs of images. The initial adjustment searches a consistent reference subset and transfers the tone of the reference subset to other images in an optimal order, which promotes the tonal consistency of all the images effectively. The joint adjustment based on overlaps with neighboring images will uniformize the color difference further. Finally, we creatively transfer the local structures of the input images to the previously-corrected ones by a local linear model.

The remainder of this paper is organized as follows. The proposed consistent tonal correction for multi-view remote sensing image mosaicking is detailedly described in Section 2. Experimental results on both challenged synthetic and real data are presented in Section 3 followed by the conclusions drawn in Section 4.

2. OUR APPROACH

Color spaces with correlations between different channels, such as the *RGB* space, always complicate the color manipulation. For these spaces, color channels must be processed in tandem

if we want to change the appearance of a pixel in a coherent way. Hence, we convert the input images into an orthogonal color space $l\alpha\beta$ without correlations before tonal correction (Reinhard and Pouli, 2011).

As the flowchart illustrated in Figure 2, the original images will be processed in two stages: tonal adjustment and structure preserving in our tonal correction algorithm. First of all, since the image sequence may take on very different tones, we have to conduct a preprocessing to attenuate the obvious tonal deviations. Then, the luminance compensation and the color correction are performed independently, and the appearances of images will change obviously in this stage, which determines the final tone as a whole. Finally, each original image will be guided by the correspondingly corrected image with the local linear model, like the guided filter (He et al., 2013), to transfer its local structures.

2.1 Tonal Preprocessing

The tone of an image is some kind of global intensity characteristic, and the mean value and standard deviation of the image intensities are the most basic indicators of the image tone. In the computer vision field, (Reinhard et al., n.d.) proposed a simple but effective algorithm which borrows one image's color characteristics from another. For simplicity, they manipulate an image in each channel of the $l\alpha\beta$ color space individually. Taking the single channel image for example, given an image \mathbf{I} and its reference image \mathbf{I}_r , the intensity value of a point \mathbf{p} in \mathbf{I} is denoted as $\mathbf{I}(\mathbf{p})$ and its corrected value can be calculated as:

$$\hat{\mathbf{I}}(\mathbf{p}) = \frac{\sigma(\mathbf{I}_r)}{\sigma(\mathcal{N}_{n \times n}(\mathbf{p}))} (\mathbf{I}(\mathbf{p}) - \mu(\mathcal{N}_{n \times n}(\mathbf{p}))) + \mu(\mathbf{I}_r), \quad (1)$$

where $\mu(\bullet)$ and $\sigma(\bullet)$ denote the mean value and standard deviation of the intensity values in a pixel set, respectively, and $\mathcal{N}_{n \times n}(\mathbf{p})$ denotes the $n \times n$ neighborhood of \mathbf{p} in \mathbf{I} .

After this transformation, the resultant image has the standard deviation and mean value that conform to the reference image appropriately. Although this algorithm does well in most situations, the quality of the resultant image depends on the similarity in content components of these two images. For example, if the original image contains lots of buildings and the reference image includes more green vegetation, then the buildings of the corrected image will be covered with a mist of green. Unfortunately, containing multiple kinds of objects with diverse tones is the often case of remote sensing images, and hence it is hard to select a suitable reference image having similar components with all input images. To effectively apply this algorithm on remote sensing images, we make some improvements described as follows.

In fact, the image tone is determined by two basic factors, luminance and color. Compared to the color, the luminance is much less sensitive to the components in images, which can be transferred via Eq. (1) successfully in most of situations. However, the color has to be processed more carefully. We propose to modify the equation (1) into a weighted form, whose processing effects can be presented in different strengths according to the weight setting. The weight value should be set proportionally to the similarity between the original image and the reference image. The improved transferring equation is defined as:

$$\hat{\mathbf{I}}(\mathbf{p}) = \frac{\sigma(\mathbf{I}_r)}{\lambda\sigma(\mathcal{N}_{n \times n}(\mathbf{p})) + (1-\lambda)\sigma(\mathbf{I}_r)} (\mathbf{I}(\mathbf{p}) - \mu(\mathcal{N}_{n \times n}(\mathbf{p}))) + \lambda\mu(\mathbf{I}_r) + (1-\lambda)\mu(\mathcal{N}_{n \times n}(\mathbf{p})), \quad (2)$$

where $\lambda \in [0, 1]$ determines the degree of transformation effects with respect to the reference image. While $\lambda = 1$, the above equation will degenerate to the form of Eq. (1).

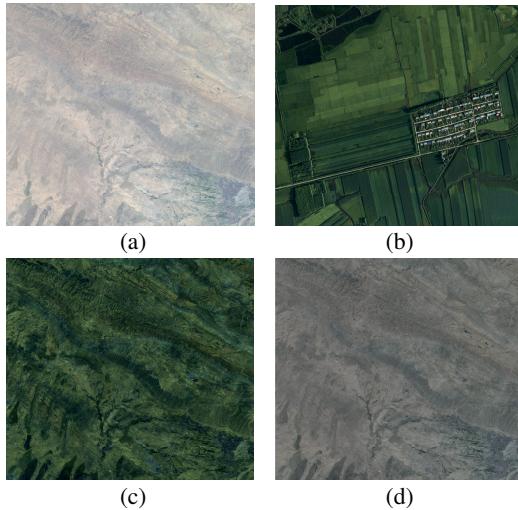


Figure 3. A comparative illustration of our improved color transferring equation: (a) the original image with a little over exposure; (b) the reference image; (c) the resulting image via Eq. (1); (d) the resulting image via Eq. (2) with the default weight combination.

In general, we tend to refer the luminance information of the reference image greatly, while adjust color characteristics slightly during tonal preprocessing. In our experiments, we set the default weight combination as $(0.8, 0.1, 0.1)$ whose components are for l , α , and β channel respectively. Figure 3 illustrates an comparative experimental result for our improved color transfer strategy. From Figure 3, we observe that the reference image contains main objects as green vegetation while the test image contains bare land. Applying the traditional color transfer via Eq. (1) directly covers the image with green tone, as shown in Figure 3(c). However, applying our improved color transfer via Eq. (2) with the default weight settings enhanced the contrast of the test image, and meanwhile maintained its original color characteristics.

To alleviate noises, we compute the mean value and standard deviation on the Gaussian filtered images. Besides, to increase efficiency, we only count the mean values and standard deviations at some grid points on the image based on their local neighborhood and the statistic parameters of other pixels on the image are bilinearly interpolated from those of grid points.

2.2 Global Luminance Compensation

After tonal preprocessing with respective to some reference image, all the input images usually present the luminance at the similar level. However, the slight difference between some pairs of images still exists because of lacking joint adjustment which takes the difference between pairs of images into account globally. Since the tonal preprocessing has eliminated most of the luminance deviation, we choose a simple gain model as the luminance refinement model.

As for a pair of adjacent images \mathbf{I}_i and \mathbf{I}_j , we define \mathbf{I}_{ij} as the subset of \mathbf{I}_i corresponding to the common pixels in the overlap with \mathbf{I}_j , and so does \mathbf{I}_{ji} . Obviously, \mathbf{I}_{ij} and \mathbf{I}_{ji} completely share the same points in the mosaicked image framework. The difference between \mathbf{I}_i and \mathbf{I}_j in the overlap after the adjustment can be written as:

$$e_{ij} = \sum_{\mathbf{p} \in \mathbf{I}_{ij}} ((a_i \mathbf{I}_i(\mathbf{p}) - a_j \mathbf{I}_j(\mathbf{q}))^2 + \alpha((a_i - 1)^2 + (a_j - 1)^2)), \quad (3)$$

where \mathbf{p} and \mathbf{q} denote a pair of corresponding points from overlapping regions \mathbf{I}_{ij} and \mathbf{I}_{ji} , respectively, a_i and a_j stand for the gain coefficients of \mathbf{I}_i and \mathbf{I}_j , respectively, and α is a fixed weight for preserving the average variance of the images after adjustment. In practice, we tend to evaluate the difference between two overlapping images after adjustment based on the mean values $\bar{\mathbf{I}}_{ij}$ and $\bar{\mathbf{I}}_{ji}$ in the overlapping region, and thus the above equation (3) is modified as:

$$e_{ij} = (a_i \bar{\mathbf{I}}_{ij} - a_j \bar{\mathbf{I}}_{ji})^2 + \alpha((a_i - 1)^2 + (a_j - 1)^2). \quad (4)$$

The above modification greatly simplifies the computation and improves some robustness to other interference factors, such as small geometric alignment bias between overlapping images. Considering the joint adjustment for all the N images, the total difference can be written as:

$$E = \sum_{\substack{i=1 \dots N, j=1 \dots N \\ \mathbf{I}_i \cap \mathbf{I}_j \neq \emptyset}} e_{ij}, \quad (5)$$

where $\mathbf{I}_i \cap \mathbf{I}_j \neq \emptyset$ means that there exists at least an overlapping region between the two images \mathbf{I}_i and \mathbf{I}_j .

We aim to find such a set of coefficients $\mathcal{A} = \{a_i\}_{i=1}^N$, which minimizes the total difference between each pair of images in overlaps. It can be achieved by solving Eq. (5) through the linear least square algorithm.

The combination of tonal preprocessing and global luminance compensation is a kind of processing scheme from coarse to fine, which improves the luminance consistency of all the input images gradually.

2.3 Color Correction Based on Histogram Mapping

As mentioned in Section 2.1, the preprocessing corrects the color characteristics just in a slight way. We observe that the color characteristics are sensitive to the positions and categories of objects in an image, which can not be operated in a global or uniform form. Therefore, each image should be corrected in an individual way according to its characteristics to eliminate the existing color deviations. Histogram mapping between overlapping regions is a good choice to solve this problem. To correct the color of each image to a consistent tone, we propose a stepwise processing strategy composed of initial correction and consistent correction.

2.3.1 Initial Correction by Referring Optimal Subset Actually, each image may have several overlapping neighbors with distinctive tones, and it's hard to determine which image should be adjusted first and which image should be selected as the reference one. Hence, we have to find the major tone of all the preprocessed images as the guidance, namely to search a subset of images with consistent color as reference ones, whose characteristics of histograms would be propagated to other images by several steps of histogram mapping in an optimal order.

Distance of Histograms: To judge the color difference between two adjacent images, we compute the distance of histograms as a quantitative metric. Firstly, we count the cumulative distribution histograms (CDFs) of the overlapping regions from an image pair $(\mathbf{I}_i, \mathbf{I}_j)$, and extract the intensity values $\mathcal{V}_i = \{v_k^i\}_{k=1}^K$ and $\mathcal{V}_j = \{v_k^j\}_{k=1}^K$ at K selected discrete points with equal interval probabilities $\{p_k\}_{k=1}^K$, as illustrated in Figure 4(a), which does not consider the bottom and up 0.5% regions at the probability range $[0, 1]$. Then, the corresponding sample set at selected discrete points in the CDF curves are fitted by the quadratic B-spline in the coordinate system $ov_i v_j$, as shown in Figure 4(b).

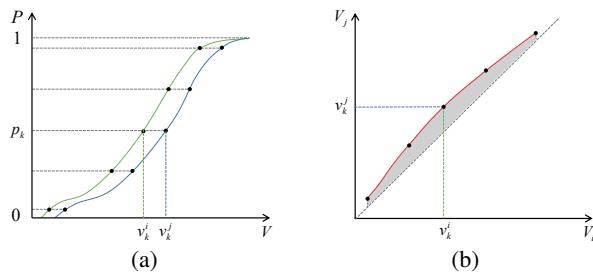


Figure 4. An illustration of measuring the distance of histograms: (a) the two CDF curves of the overlapping regions from an image pair $(\mathbf{I}_i, \mathbf{I}_j)$ with discrete nodes at equal interval probabilities; (b) the fitted curve by the quadratic B-spline fitting for histogram mapping, forming a shadow region with the clino-diagonal line (*base line*).

The distance of two histograms from the overlapping regions from the image pair \mathbf{I}_i and \mathbf{I}_j is measured as $d_H(\mathbf{I}_i, \mathbf{I}_j) = A_s/(p_K - p_1)$, where A_s denotes the area of the shadow region formed by the fitted curve and the clino-diagonal line as shown in Figure 4(b).

Correction Order: So far, we can construct a graph of images in which each image is regarded as a node and the neighboring relationship with some overlap is represented as an arc. The cost of an arc linking images \mathbf{I}_i and \mathbf{I}_j is defined as:

$$c(\mathbf{I}_i, \mathbf{I}_j) = \frac{1}{3}(d_H^\alpha(\mathbf{I}_i, \mathbf{I}_j) + d_H^\beta(\mathbf{I}_i, \mathbf{I}_j) + c_{base}), \quad (6)$$

where $d_H^\alpha(\mathbf{I}_i, \mathbf{I}_j)$ and $d_H^\beta(\mathbf{I}_i, \mathbf{I}_j)$ denotes the distances of histograms from overlapping regions \mathbf{I}_{ij} and \mathbf{I}_{ji} of \mathbf{I}_i and \mathbf{I}_j in the α and β channels, respectively. The base amount c_{base} , which was set as 1 in this paper, aims to avoid to find a least-cost path across too many nodes in some special cases. According to their costs under the predefined threshold combination or not, the arcs are classified into *consistent arcs* and *inconsistent ones*. In our experiments, the threshold combination is set as (0.03, 0.005) for α and β channel respectively, which can always give the ideal results according to lots of testing results. The nodes of the maximum connected sub graph with consistent arcs are selected as the reference set of images. Other images will find the optimal referring paths by the shortest path algorithm from the root node, i.e., the reference set of images. At last, the order of histogram mapping can be achieved through visiting each node from the root nodes by a BFS (breadth-first search) and each image will be corrected based on the histogram mapping with its parent node. Figure 5 illuminates the procedure of finding the optimal correction order.

2.3.2 Consistent Correction by Cross Reference The correction by referring the maximum-consistent subset transforms all images into a basically consistent color in a strong way. However, the color differences between pairs of images may distribute at diverse levels, because of no considering the neighboring relationship during initial correction. The appearances of adjacent images would be more consistent, once the deviations between all pairs of images were put in a similar level. So, we propose to further adjust images jointly with taking all neighbors into account. For the sake of efficiency, we readjust each image according to the weighted mapping curve which combines the histogram mapping curves related to its neighbors. The weight of each neighbor is set proportionally to the area of its corresponding overlap region. What's more, the weighted mapping curve won't change largely compared to the initial one due to the fact that the component curves shall be close to the

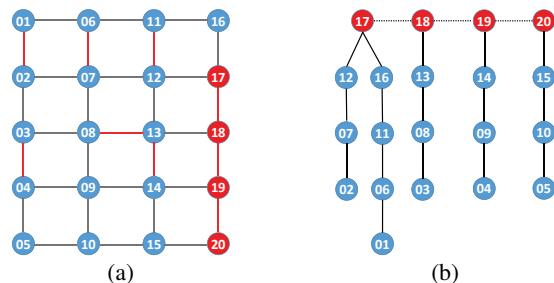


Figure 5. An illustration of finding the optimal correction order: (a) the graph of 20 images in 4 strips with 5 images each, whose arcs are labeled in red (consistent arcs) or in blue (inconsistent arcs). The nodes of the reference subset are labeled in red as the root ones. (b) the spanning tree generated by the shortest path algorithm, which is applied with the costs between reference images set as 0. The optimal correction order is the result of the breadth-first search from root nodes.

base line more or less after adjustment in the initial correction period. Notice that, if the initial correction fails to put images in a similarly consistent color, the joint adjustment in this way would be a disaster for the mapping curve of an image. That is to say, this step had better to be omitted in some extreme cases, such as that the observation area consists of very different objects like water area and bare land.

2.4 Structure Transferring

To achieve a consistent tone, all the input images have undergone a series of processes including color transferring in a weighted form, global luminance compensation and histogram mapping. Although the tone of each input image is corrected to our ideal status, the local structures in the image may have been weaken or even damaged sometimes, which often happens when operating the pixels of an image in a global form. In order to preserve the structure information of the original images in the finally corrected ones, we propose an efficient approach which transfers local structures in original input images to the corrected ones by a local linear model, which combines the merits of both parametric and non-parametric approaches.

Inspired by the idea of the guided filter (He et al., 2013), we process each original image by a local linear model under the guidance of its corresponding corrected image via tonal adjustment. A series of local linear transformations are applied on each original image to take on the appearance of the corresponding corrected one. Meanwhile, this operation can transfer the structure features of the original images to the corrected ones. Given an input image \mathbf{I} and its corresponding corrected one $\hat{\mathbf{I}}$, we take a pixel point \mathbf{p} as an example, whose error function of our guidance model can be written as:

$$e(\mathbf{p}) = \sum_{\mathbf{p}_l \in \mathcal{N}_{m \times m}(\mathbf{p})} (a\mathbf{I}(\mathbf{p}_l) + b - \hat{\mathbf{I}}(\mathbf{p}_l))^2, \quad (7)$$

where $\mathcal{N}_{m \times m}(\mathbf{p})$ denotes the $m \times m$ neighborhood of \mathbf{p} , and a and b are the two parameters of the guidance model for the pixel point \mathbf{p} . This above equation can be solved easily by the linear least square method. To avoid the effect of the smooth filter, we use the parameters estimated for local transformation directly, instead of averaging the parameters within some neighborhood as used in (He et al., 2013). Note that a and b are computed using all the pixels in the $m \times m$ neighborhood of \mathbf{p} . In this case, a smoothness-like constraint across pixels is naturally enforced in flat regions, while in regions with strong structures, such as edges, the contrast is reasonably maintained (Shan et al., 2010).



Figure 6. An illustration of our proposed structure transferring algorithm: (Left) the original image suffering from underexposure; (Middle) the adjusted result by exposure compensation via *Photoshop*; (Right) the structure-transferring image, which not only presents the similar appearance with the adjusted image but also preserves the local structures of the original image.

To demonstrate the effects of our algorithm, we take a poorly-exposed image as the original input image and its adjusted result through editing manually in *Photoshop* as the corrected image, and the structure-transferring image is showed in Figure 6.

3. EXPERIMENTAL RESULTS

To sufficiently evaluate the performance of our proposed method, we tested it on a set of synthetic images and two groups of real multi-temporal image data captured by Chinese satellite ZY-3.

3.1 Synthetic Data

In this experiment, a set of synthetic image data comprised of 5×5 images was clipped from a wide-view high resolution remote sensing image with the size of 3652×3601 pixels, in which there exists a 20% overlap between two adjacent images, as shown in Figure 7(a). All these clipped images were adjusted individually in diverse ways in both aspects of luminance and color via *Photoshop*, as illustrated in Figure 7(b). During our experiment, we selected the 23-th clipped image as the reference image for tonal preprocessing, which was followed by the luminance, color correction and local structure transferring. From simply mosaicked images by a simple superposition as shown in Figure 7(b), we observe that the tonal preprocessing greatly alleviates the luminance differences of images on the whole as depicted in Figure 7(c) while our proposed color correction method further minimizes the color deviations of images as shown in Figure 7(d), which are more easily identified at enlarged representative regions as shown in Figures 7(e)-(f).

Table 1 shows the numerical comparative analysis at different stages of our proposed method, in which the number denotes the average of distances of histograms between all pairs of images. From Table 1, it's easy to observe that the tonal preprocessing greatly minimizes the luminance differences of images but slightly decreases the color deviations. The proposed stepwise color corrections eliminated the color deviations in α and β channels gradually. The initial correction by referring to the consistent reference subset eliminates the major color deviations between image pairs while the joint adjustment is responsible for reallocating the deviations to be well-distributed among all the pairs of images.

3.2 Real Multi-Temporal Data

Firstly we tested our method on a set of multi-temporal images with three strips consisting of 21 images with the down-sampling

Stages	l	α	β
No-Operation	0.3550	0.0464	0.0072
Tonal Preprocessing	0.0459	0.0414	0.0066
Initial Correction	0.0422	0.0099	0.0017
Joint Correction	0.0422	0.0069	0.0015

Table 1. Average tonal differences in l , α and β channels after different processing stages for synthetic image data.

size of 2294×1904 captured by Chinese satellite ZY-3. These three strips, comprised of 01 ~ 07, 08 ~ 14, and 15 ~ 21 images, respectively, as shown in Figure 8(a), were captured at different times, 2012.06.26 – 15 : 34, 2013.06.15 – 15 : 42, and 2012.06.16 – 15 : 37, respectively. From Figure 8(a), it is easy to observe the inconsistent tone between images of both intra-strip and inter-strip. By applying our proposed tonal correction based on the automatically selected maximum reference subset {12, 13, 14, 19, 20}, a consistent tone was achieved among all corrected images, as shown in Figure 8(b) where the final tone on the whole is close to the primitive tone of the reference image subset more or less. Via our proposed tonal correction, it is difficult to identify the borders for almost of all the pairs of images in Figure 8(b) while their corresponding borders before correction in Figure 8(a) are easily observed. For a more detailed visual comparison, four enlarged representative regions are illustrated in Figures 8(d)-(g). Besides, several image pairs with a little residual tonal deviations, such as {01,08}, {13,20} and {14,21}, are also not a trouble, which can be smoothed by any feathering or blending algorithm easily. All in all, our method just aims at tonal correction, which should be followed by the subsequent operations like seamline searching and image fusion in the full procedure of image mosaicking. To demonstrate the compatibility of our approach applied in the whole image mosaicking procedure, we fused all the corrected images by a multi-band blending algorithm with the open source software *Enblend*¹, which generated the finally mosaicked image as shown in Figure 8(c).

However, in the corrected result as shown in Figure 8(b), there are still some flaws on some image regions, for example, the images 05 and 19, whose brightnesses distribute unevenly. It is a problem introduced by tonal preprocessing in which the transformation parameters of pixels are bilinearly interpolated from some fixed grid points. In such a strategy, pixels, located in the margin regions between a bright region and a dark one, are prone to be brightened overly, because its transformation parameters are influenced by the dark region which shall be brightened through the transformation. As it should be, such situation can be alleviated by setting grid points more compactly, which can also be combined with setting a smaller weight to weaken the effect of transformation. A group of examples are reported in Figure 9, from which we observe that the over-brightened margin region around the dark region is smaller with the more compact grid points setting.

The numerical comparative analysis at different stages is given in Table 2. Different from those in Table 1, we observe from Table 2 that the joint correction operation slightly enlarged the color deviation in the α channel. As we can know from the operation as described in Section 2.3.2, the joint adjustment is not necessary to decrease the average difference between images, especially when the neighbors of some image have obviously different tonal characteristics, whose function is to uniformize the deviations between pairs of images contributing a relatively uniform tone as a whole.

¹Enblend is available at <http://enblend.sourceforge.net/>

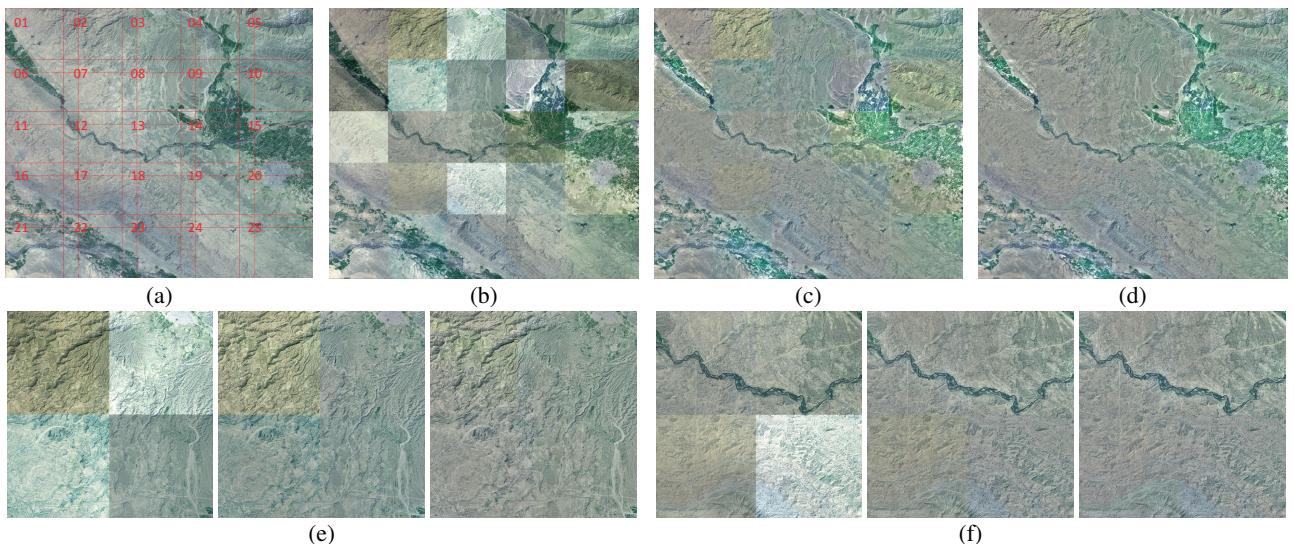


Figure 7. The experimental results on the synthetic image data: (a) the original image with red border lines of clipped images with a 20% overlap between two adjacent ones; (b) the manually adjusted images individually via *Photoshop* displayed as a single simply mosaicked image; (c) the tone-preprocessed result, in which all clipped images present at a similar luminance but the color deviations between adjacent images do not disappear; (d) the finally corrected result, which presents a consistent tone on the whole with the reference subset {11,12,13,16,18,19,21,23,24,25}; (e)-(f) two enlarged representative regions shown with the manually adjusted results, the tone-preprocessed ones, and the finally corrected ones from left to right.

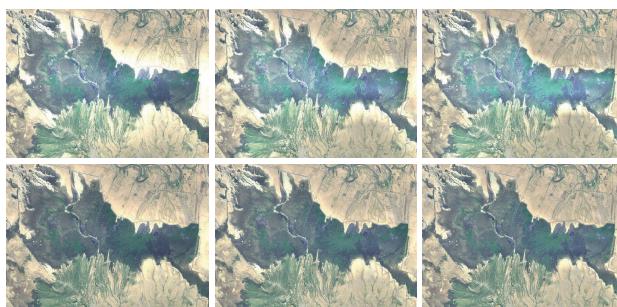


Figure 9. The corrected results of an example region on the image 19 in Figure 8(a) when setting grid points in different densities during tonal processing: 8×8, 16×16, and 24×24, respectively, from left to right. The weight values for the l channel are set as 1.0 and 0.5 for the images in the top and bottom rows, respectively.

Stages	l	α	β
No-Operation	0.1183	0.0442	0.0053
Tonal Preprocessing	0.0491	0.0403	0.0049
Initial Correction	0.0401	0.0138	0.0017
Joint Correction	0.0401	0.0145	0.0015

Table 2. Average tonal differences in l , α and β channels after different processing stages for real multi-temporal images shown in Figure 8(a).

To sufficiently evaluate our method, we tested it on another group of four strips of multi-temporal images consisting of 16 images as shown in Figure 10(a), which were also captured by Chinese satellite ZY-3 at different times. The visual comparison and the corresponding numerical analysis are provided in Figure 10 and Table 3, respectively. The similar performance on this group of image data can be observed, except for the region on the image 04 of the mosaicked image shown in Figure 10(c), which contains a large area of water and presents a little strange tone after the color correction. This newly emerged problem tells that our color

Stages	l	α	β
No-Operation	0.1975	0.0715	0.0067
Tonal Preprocessing	0.0631	0.0641	0.0059
Initial Correction	0.0526	0.0100	0.0017
Joint Correction	0.0526	0.0089	0.0013

Table 3. Average tonal differences in l , α and β channels after different processing stages for real multi-temporal images shown in Figure 10(a).

correction algorithm is infeasible for those images whose tone can't be sufficiently represented by that of its overlap region with its neighbors. That's to say, our method can achieve very good correction effects in the premise that each image basically has an uniform tonal characteristic within itself.

4. CONCLUSION

In this paper, we proposed a full set of processes for consistent tonal correction of multi-view remote sensing images for mosaicking. The utilized tonal preprocessing can decrease the deviations between images integrally by the color transferring equation in a weighted form with an example image. Then, the global luminance compensation in the l channel and the stepwise color correction in the α and β channels further minimize the luminance difference and the color deviations, respectively, with a consistent tone on the whole. Finally, the evaluation on several groups of representative synthetic and real images demonstrates that the proposed method is effective and feasible. Of course, there exist also some aspects to be possibly improved. One is the strategy of selecting the reference image subset, which won't select two or multiple subsets disconnected with each other, though they are in a very consistent tone. This improvement may decrease the times of histogram mapping from images to the reference subset, which may improve the quality of correction greatly. Next, the consistent correction should be implemented by a more robust algorithm to deal with some extreme cases. All of these problems will be deeply investigated in our future work.

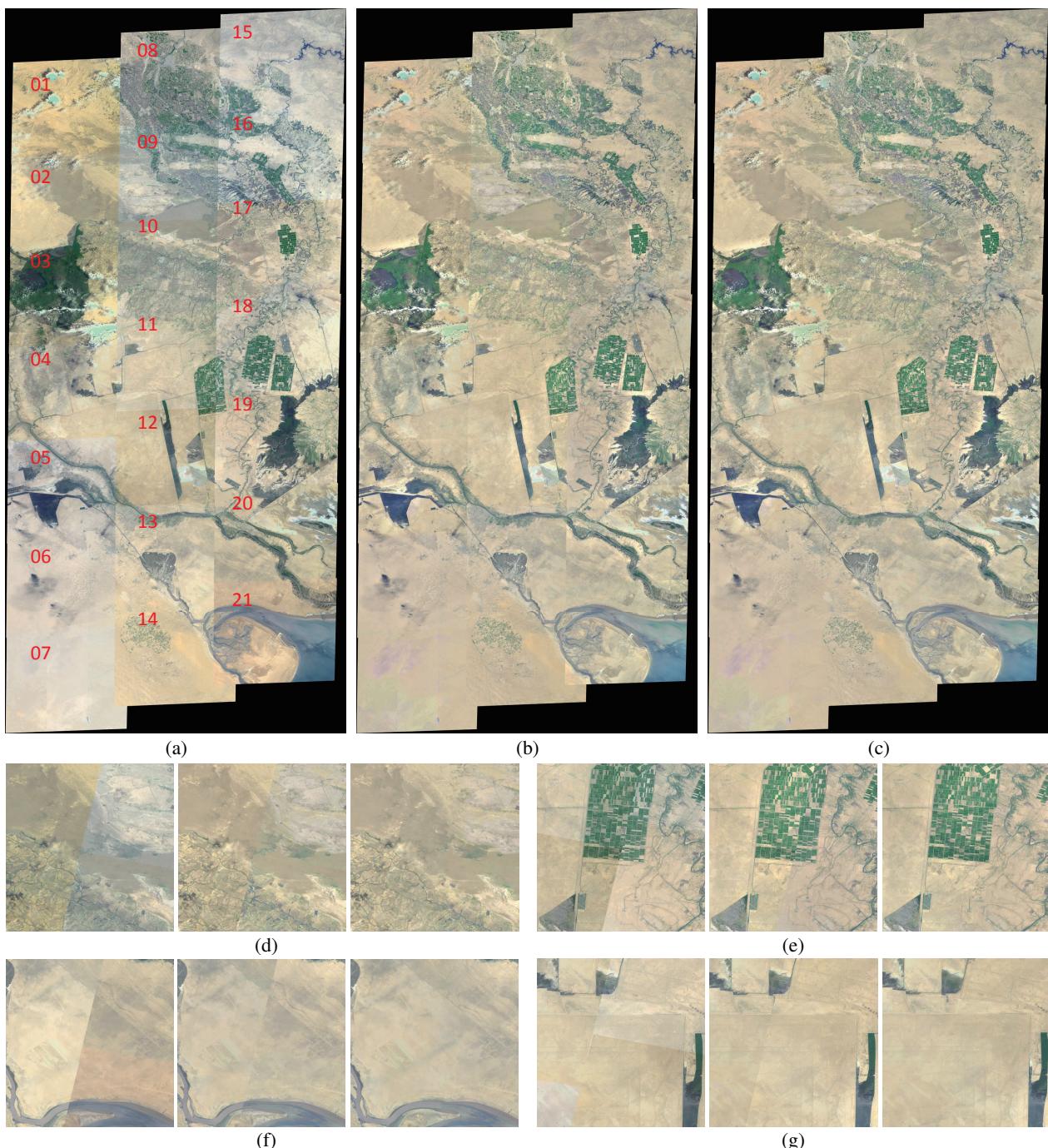


Figure 8. The experimental results on three strips of multi-temporal images: (a) the mosaicked image by a simple superposition from 21 input images; (b) the finally corrected results based on the maximum reference subset {12, 13, 14, 19, 20}; (c) the mosaicked image by applying a multi-band blending on the finally corrected images in (b) with the open source software *Enblend*; (d)-(g) four enlarged representative regions clipped from the mosaicked images before and after consistent tonal correction and those after applying multi-band blending on the corrected result, as shown in (a), (b) and (c), respectively.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Project No. 41571436), the Hubei Province Science and Technology Support Program, China (Project No. 2015BAA027), the National Basic Research Program of China (Project No. 2012CB719904), and the National Natural Science Foundation of China under Grant 91438203.

REFERENCES

- Canty, M. J. and Nielsen, A. A., 2008. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively reweighted MAD transformation. *Remote Sensing of Environment* 112, pp. 1025–1036.
- Canty, M. J., Nielsen, A. A. and Schmidt, M., 2004. Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sensing of Environment* 91, pp. 441–451.
- HaCohen, Y., Shechtman, E., Goldman, D. B. and Lischinski, D., 2011. Non-rigid dense correspondence with applications for image

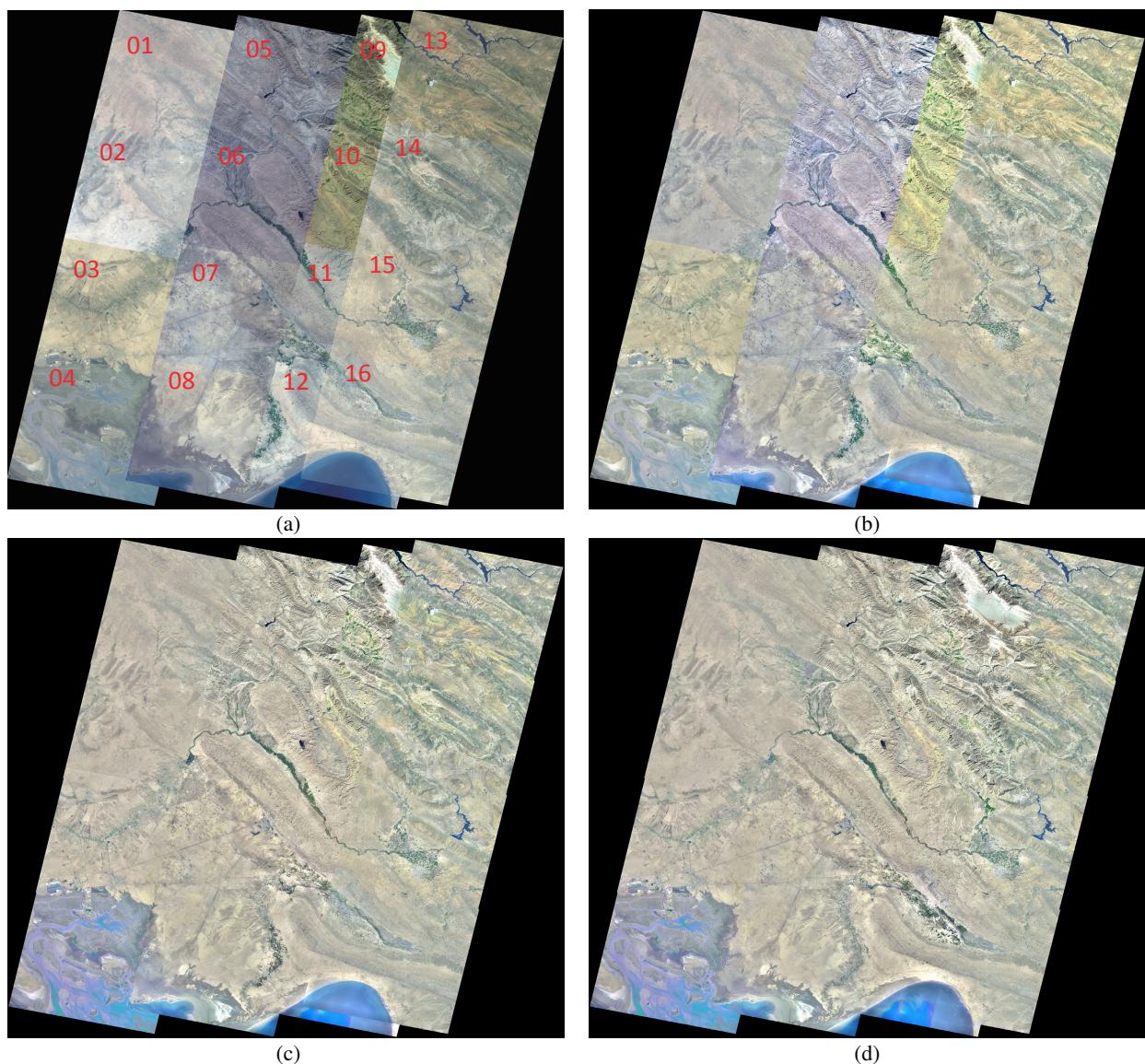


Figure 10. The experimental results on four strips of multi-temporal images: (a) the mosaicked image by a simple superposition from 16 input images; (b) the transferred results via tonal preprocessing, with a basically consistent luminance; (c) the finally corrected results based on the maximum reference subset {14, 15, 16}; (d) the mosaicked image by applying a multi-band blending on the finally corrected images in (c) with the open source software *Enblend*.

enhancement. *ACM Transactions on Graphics (TOG)* 30(4), pp. 70:1–70:10.

HaCohen, Y., Shechtman, E., Goldman, D. B. and Lischinski, D., 2013. Optimizing color consistency in photo collections. *ACM Transactions on Graphics* 32(4), pp. 38:1–38:10.

He, K., Sun, J. and Tang, X., 2013. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(6), pp. 1397–1409.

Hwang, Y., Lee, J.-Y., Kweon, I. S. and Kim, S. J., 2014. Color transfer using probabilistic moving least squares. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3342–3349.

Ilie, A. and Welch, G., 2005. Ensuring color consistency across multiple cameras. In: *IEEE International Conference on Computer Vision (ICCV)*.

Li, X., Hui, N., Shen, H., Fu, Y. and Zhang, L., 2015. A robust mosaicking procedure for high spatial resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing* 109, pp. 108–125.

Moulou, P., Duisit, B. and Monasse, P., 2013. Global multiple-view color consistency. In: *International Conference on Visual Media Production (CVMP)*.

Pitié, F., Kokaram, A. C. and Dahyot, R., 2005. N-dimensional probability density function transfer and its application to color transfer. In: *IEEE International Conference on Computer Vision (ICCV)*.

Pitié, F., Kokaram, A. C. and Dahyot, R., 2007. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding* 107(1C2), pp. 123–137.

Reinhard, E. and Pouli, T., 2011. Colour spaces for colour transfer. Lecture Notes in Computer Science, Vol. 6626, Springer Berlin Heidelberg.

Reinhard, E., Ashikhmin, M., Gooch, B. and Shirley, P., n.d. Color transfer between images.

Shan, Q., Jia, J. and Brown, M. S., 2010. Globally optimized linear windowed tone mapping. *IEEE Transactions on Visualization and Computer Graphics* 16(4), pp. 663–675.

Tai, Y.-W., Jia, J. and Tang, C.-K., 2005. Local color transfer via probabilistic segmentation by expectation-maximization. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Xiang, Y., Zou, B. and Li, H., 2009. Selective color transfer with multi-source images. *Pattern Recognition Letters* 30(7), pp. 682–689.

Xiao, X. and Ma, L., 2006. Color transfer in correlated color space. In: *ACM International Conference on Virtual Reality Continuum and Its Applications*.

Xu, W. and Mulligan, J., 2010. Performance evaluation of color correction approaches for automatic multi-view image and video stitching. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yoo, J.-D., Park, M.-K. and Cho, J.-H., 2013. Local color transfer between images using dominant colors. *Journal of Electronic Imaging* 22(3), pp. 033003.