Global Color Consistency Correction for Large-Scale Images in 3-D Reconstruction

Yunmeng Li, Yinxuan Li, Jian Yao, Ye Gong, and Li Li

***Abstract*—Global color consistency correction for multiview images in three-dimensional (3-D) reconstruction is an important problem. The color differences between the images will affect the result of dense matching, thereby reducing the geometric accuracy of the mesh model. Moreover, it will also affect the result of texture mapping, causing color differences in the textured model. The color correction method based on global optimization is mainly used to solve this problem. And existing methods usually use sparse matching points as the color correspondences, but the correction results are not accurate enough as a result of the sparsity of the points. Besides, their efficiency of solving large-scale images globally is low. This article proposes a novel color correction method to eliminate the color differences between large-scale multiview images effectively. The core idea of our method is to group images by graph partition algorithm, and then perform intragroup correction and intergroup correction in sequence. First, for each pair of images, we calculate the reliable matching regions around the sparse points as the color correspondences according to the local homography principle. Compared with sparse matching points, our strategy can achieve more accurate color correction results. Next, for large-scale images, we partition them into many groups. For each group of images, the correction parameters are solved to eliminate the color differences of the images included in the group. Finally, we eliminate the color differences between groups by intergroup correction to achieve overall color consistency. Experimental results on typical datasets demonstrate that the proposed method is better than the current representative methods. The proposed method shows better color consistency in the extreme cases, and also exhibits higher computational efficiency on large-scale image sets.**

***Index Terms*—Color consistency, graph partition, local homography, multiple-view images.**

I. Introduction

THE image-based three-dimensional (3-D) reconstruction is a process of recovering real-world 3-D information from two-dimensional (2-D) images, and is widely used in many fields such as smart city, cultural relics restoration, virtual reality, and unmanned driving system. However, in the process of multiview image acquisition, especially for large-scale scenes, the acquisition conditions such as shooting time, weather, illumination, and camera parameters are inevitably different, resulting in color differences (CDs) in the images. If the original uncorrected images are used for 3-D reconstruction, it will first affect the result of dense matching, which partly depends on the similarity between pixels. And the result of matching will undoubtedly further affect the geometric accuracy of the reconstructed mesh model. Moreover, using the original images for subsequent texture mapping will also cause CDs between patches in the textured model. Geometric dislocation and texture inconsistency will greatly affect the visual quality of the final model. Therefore, it is essential to eliminate CDs between multiview images through color consistency correction technology.

Over the years, the color correction researches in image stitching have been continuously developed and improved [1]–[3]. The global optimization strategy is the mainstream method among them, and the color correspondences are usually calculated based on the overlapping area between the images. This type of methods cannot be directly applied to the color correction for multiview images in 3-D reconstruction. Since the shooting angles and positions of the images used for 3-D reconstruction are usually very different, it is difficult to precisely align these images. The current basic scheme is to calculate the color correspondences based on the sparse matching points of multiview images generated by structure from motion (SfM) [4]. Obviously, this strategy is not reliable enough compared with the strategy of using the overlapping area as the color correspondences. The characteristics of the feature points determine that their distribution is not uniform, and most of the feature points locate at the borders and corners. Thus, some color information of the images is ignored and the color correction results may be influenced. Some researchers propose to calculate color correspondences based on dense matching, but this strategy is very time-consuming [5]. In addition, the global optimization framework is not suitable for processing large-scale images, because it solves all correction parameters at the same time. When the number of images increases to a certain extent, the solution time will be greatly increased. In addition, the scale of the equation coefficient matrix also increases, and the solution may be interrupted due to out of memory error.

Considering the above two aspects, we propose a novel color consistency correction method for large-scale multiview images.

Manuscript received December 30, 2021; revised March 14, 2022; accepted April 8, 2022. Date of publication April 13, 2022; date of current version May 3, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42101440, in part by the Shenzhen Central Guiding the Local Science and Technology Development Program under Grant 2021Szvup100, and in part by the Open Research Fund Program of Key Laboratory of Digital Mapping and Land Information Application, Ministry of Natural Resources. *(Yunmeng Li and Yinxuan Li are co-first authors.) (Corresponding author: Li Li.)*

Yunmeng Li, Yinxuan Li, Ye Gong, and Li Li are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: yunmeng.li@whu.edu.cn; yinxuanli@whu.edu.cn; cvrs-gongye@whu.edu.cn; li.li@whu.edu.cn).

Jian Yao is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China, and also with the GongQing Institute of Science and Technology, Jiujiang 332020, Chinacountry> (e-mail: jian.yao@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3167264

In response to the problem of inaccurate color correspondences, we propose a novel strategy to obtain corresponding regions around the sparse matching points using local homography. In this way, not only are the statistical color correspondences more accurate, but the calculation time is within an acceptable range as well. For large-scale image color correction, we propose a two-step strategy based on graph partition. First, we construct a graph based on the image adjacency relationship, and use the graph partition algorithm to group the images. Then, we perform intragroup correction to eliminate CDs within each group of images, and perform intergroup correction to eliminate CDs in images of different groups. Because the number of images included in each group is far less than the total number of images, the computational time and memory requirement for intragroup correction are greatly reduced. In addition, each group of images can be corrected independently and parallel. The key contribution of this article can be summarized as follows.

1 Instead of using the sparse points as the color correspondences, we propose a novel strategy based on local homography to extract more accurate color correspondences. Our proposed method is more robust since color correspondence is the basis for color consistency correction.
2 To perform global color correction on large-scale images efficiently, we propose a two-step strategy based on graph partition. This strategy effectively reduces the computational burden of equation solving by grouping images and sequentially performing intragroup correction and intergroup correction.

Experimental results on various datasets show that our algorithm shows better efficiency and effectiveness than the current representative algorithms.

The rest of this article is organized as follows. In Section II, an overview of related work is provided. Section III introduces the proposed color correction algorithm in detail. The experiments are presented in Section IV. Finally, Section V concludes this article.

II. Related Work

We introduce image color correction algorithms from two perspectives: Color transfer between two images, and color consistency correction for multiple images.

A. Color Transfer

Color transfer refers to adjusting the tone of the source image to make it consistent with the target image. According to the mapping principle, color transfer approaches can be divided into two categories: Nonparametric and parametric.

Most of nonparametric methods use look-up tables to directly record the mapping of all color levels in the image. Usually, the look-up table is calculated based on the 2-D joint histogram of the known color correspondences between the two images. Pitié *et al.* [6] proposed an iterative approach based on the one-dimensional probability density function (PDF) transfer, which can accurately transfer the source PDF to target PDF and obtain the one-to-one color mapping table. Su *et al.* [7] proposed to decompose the source image into multiple base layers and detail layers, and then perform color transfer for each base layer using Pitié *et al.*'s method [6]. Finally, the base layers were combined with the boosted detail layers to generate the final corrected image. Hwang *et al.* [8] proposed a color mapping method based on moving least squares (MLS), and then combined MLS with the probabilistic modeling of the color transfer in the 3-D color space to better deal with misalignment and noise. Liu *et al.* [9] proposed to decompose the natural image into reflectance image and shading image through unsupervised deep learning. Palette-based recoloring is performed only on the reflectance image to avoid color overflow. However, the above methods cannot eliminate local CDs; therefore, some stronger local color transfer schemes are proposed. Wu *et al.* [10] proposed a content-based local color transfer algorithm, which integrates the spatial distribution of the target color style in the optimization process. Hwang *et al.* [11] improved their previous algorithm [8] by introducing bilateral weights to realize local color transfer that takes into account the spatial variation of colors. Niu *et al.* [12] first performed global coarse color correction using Pitié *et al.*'s method [13], and then performed local color correction to improve the local color consistency.

Different from the nonparametric methods, the parametric methods describe the mapping relationship between images as a global transformation, such as the linear model, the gamma model, and so on. Reinhard *et al.* [14] first proposed a global color transfer method based on mean and standard deviation. This method also converts the image from RGB color space to decorrelated $l\alpha\beta$ color space to process each color channel independently. Xiao and Ma [15] proposed a gradient preserving color transfer method, which takes the histogram of the target image and the gradient map of the source image as the reference information to solve the corrected image. Nguyen *et al.* [16] applied white-balance to remove color casts in the source and target images, and then performed Xiao *et al.*'s method [15] to match the overall brightness between the two images. Afterwards, a 3-D color gamut mapping was performed to ensure that the color of the corrected image is completely within the target color gamut. Similarly, the local color transfer scheme is also applied in the parameterized method. He *et al.* [17] constructed the guide image according to the semantic correspondences of the source image and the target image, and then iteratively optimized the local linear model from coarse to fine. Wu *et al.* [18] generated a weighted attention map using the saliency and gradient maps, and then used Reinhard *et al.*'s method [14] to correct the salient and nonsalient regions respectively.

Although many color transfer methods are difficult to extend to multiple images, it is clear that they also have made a great contribution to the development of color consistency correction for multiple images. In general, parametric methods are more suitable for extending to multiple images than nonparametric methods.

B. Color Consistency Correction

Color correction for multiple images, also known as color consistency correction, is a key technology in aerial image mosaicking, panoramic stitching, and other fields. There are

mainly two types of methods, namely methods based on path propagation and methods based on global optimization.

The path propagation methods first select an image as the reference image, and then gradually transfer its color information to other images along the best propagation path. Pan *et al.* [19] proposed a network-based method to eliminate the CDs between images. This method determines the transferring path based on two factors: the time of transfer and the size of overlap. They also used a global to local strategy to improve the performance of color correction. Under the assumption that the error of each transfer step is equal, Chen *et al.* [20] determined the reference image and the transferring path using the shortest distance algorithm to reduce the normalized error. In essence, the path propagation method is the color transfer between two images. The accumulated errors may appear in the propagation process. In addition, there is no clear standard for the selection of reference image.

The global optimization methods, which solve the correction parameters of all images simultaneously, are the current mainstream methods. Such methods establish an independent color mapping model for each image, and convert the color consistency correction problem into a global energy optimization problem. The goal of the energy function is to minimize the total CDs between the images after color correction. The optimal solution of the energy function is the final parameters of the image mapping models. Brown and Lowe [1] first proposed to calculate the gain for multiple images by globally optimizing the energy function, and used the mean gray value of the overlapping area as the color correspondence. Xiong and Pulli [21] proposed to perform color correction in the decorrelated color space. Specifically, the linear model is still applied to the color channels, and the gamma model is applied to the luminance channel. Xia *et al.* [2] used a more complex piecewise quadratic spline curve as the mapping model to obtain more powerful color correction capability. To eliminate local CDs, Yu *et al.* [22] proposed a global to local optimization strategy. They first used least squares optimization to eliminate global CDs between images. A local correction strategy is further performed to eliminate the residual local CDs. In order to avoid meaningless global solution, Xie *et al.* [23] proposed a global optimization method with a supervised initial solution. The initial solution of each image is obtained by the path propagation method, and then the distance between the corrected color and the initial solution is served as a regular term to constrain the optimization process.

The color consistency correction in 3-D reconstruction is similar to that in image stitching. The most significant difference between them is the color correspondence extraction method. HaCohen *et al.* [5] applied the dense matching points between images as color correspondences, and then automatically propagated the changes made by the user to other images by minimizing a global energy cost function. Moulon *et al.* [24] proposed to use the virtual line descriptor filter to further extend the set of color correspondences generated by feature matching. Shen *et al.* [25] proposed a multiview color correction approach for 3-D modeling. After reconstructing the triangular mesh from images, this method accurately calculates overlapping regions by reprojecting the mesh to images. Park *et al.* [4] used scale-invariant feature matches as the color correspondences, and proposed a robust low-rank matrix decomposition method to robustly estimate the parameters of the gamma model. Yang *et al.* [26] also used the sparse matches as the color correspondences, but they solved the parameters by minimizing the variance of all observed color values for each sparse point. In addition, they also proposed a hierarchical correction strategy for large-scale images.

III. Method

Given a large-scale multiview images $\mathcal{I} = \{I_i\}_{i=1}^{N}$, where N is the number of images. We attempt to eliminate the CDs between the input images. Considering the low efficiency and memory consumption of global color consistency optimization for large-scale images, we propose a two-step strategy, including intragroup correction and intergroup correction. The first step of our approach is to group the images. We construct the graph based on the adjacency relationship. Let $G = \{V, E\}$ denote the graph, where V and E represent the images and the adjacency relationship, respectively. Specifically, we first use the SfM algorithm [27] to obtain sparse matching points, which indicate the adjacency relationship between the images. Generally speaking, the more matching points, the more content shared between the two images. Then we group the images by applying graph partition in the adjacency graph G.

After that, for each image pair, we need to acquire the accurate color correspondences, which are the basic measurement for color consistency optimization. Sparse matching points are often used as color correspondences [24], [26], but this way usually ignores color information of many regions other than feature points. However, more accurate color correspondence extraction methods such as dense matching require a lot of time. In our method, we propose a novel strategy that only computes matching pixels for regions near the sparse matching points using local homography. Compared with other methods, this strategy does not need to compute dense matching for all pixels; therefore, reliable corresponding pixels can be obtained efficiently. For convenience and robustness, the quantiles of the cumulative color distribution histogram constructed from corresponding pixels are extracted as the final color correspondences.

Based on the above image groups and color correspondences, we implement the two-step strategy to perform color consistency optimization for large-scale images. For each group of images, the global energy function [2] is constructed based on the color consistency and gradient preservation constraints, etc. The correction model parameters of the images included in each group can be solved by minimizing the energy function globally. After intragroup correction, the CDs of images belonging to the same group are eliminated, but there are still CDs between images of different groups. Therefore, intergroup correction is required. We regard each group of images as a virtual image. All the virtual images form a virtual image group. The similar correction can be performed on the virtual image group to achieve the

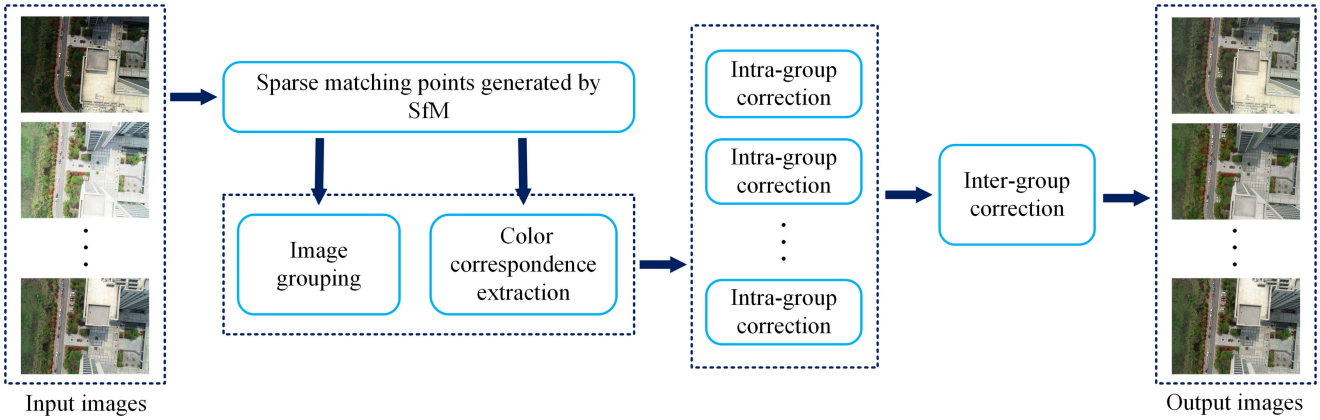

Fig. 1. Workflow of our proposed approach. For the input image set, sparse matching points are first obtained using SfM. Then, we construct the graph and cluster the images into many groups. In addition, for each pair of images, the color correspondences are constructed using the local homography. Next, the intragroup correction is performed to eliminate the color differences for images inside the group. Finally, we perform intergroup correction to eliminate the color differences between groups.

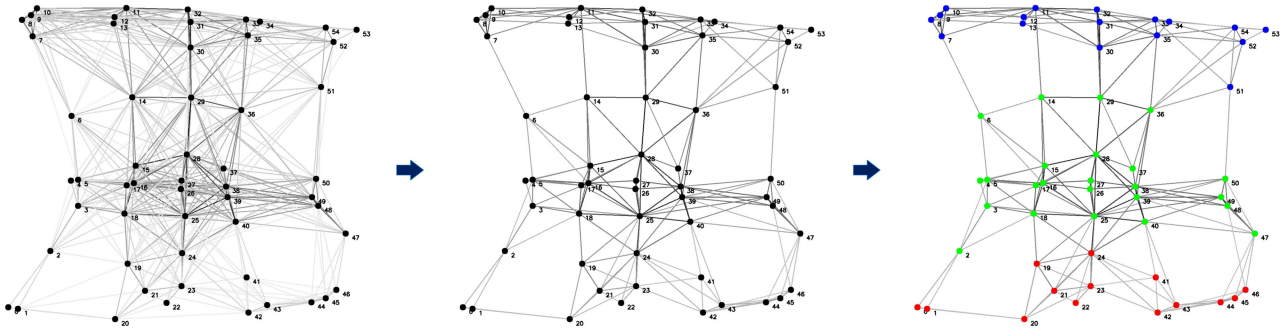

Fig. 2. Graph partition result of School dataset. (a) Initial adjacency graph of this dataset. (b) Simplified graph. (c) Graph partition result. In (a) and (b), the edge with darker color indicates the higher weight. In (c), the different colors of the nodes indicate different groups.

color consistency of the entire image set. The workflow of the proposed approach is shown in Fig. 1.

A. *Image Grouping*

According to the sparse matching points obtained by the SfM algorithm, we can easily know the adjacency relationship between the images and construct the adjacency graph $G = \{V, E\}$, where the vertices $V = \{v_i\}_{i=1}^{N}$ represent the images, and the edges $E = \{e_{ij}, i \neq j\}$ denote the adjacency relationship between images. The weight w_{ij} of e_{ij} is positively correlated with the number of matching points between the image I_i and I_j. $w_{ij} = 0$ means that there is no matching point between the two images. An example of initial graph is shown in Fig. 2(a).

However, it is expensive to consider all matching relationships, because we need to calculate the color correspondences for each edge. And the scale of the coefficient matrix is proportional to the number of edges. In addition, the edges with a few matching points are unreliable. Thus, we tend to only consider the major edges in our method. Namely, we need to simplify the initial graph. To solve this problem, the intuitive method is to directly discard the edges with small weights. Another simple way is to keep only the first T edges with higher weights for each vertex. We combine these two methods. We first use a threshold to remove edges with a few matching points, and then set a parameter T to limit the number of edges for per vertex. For each edge e_{ij}, we calculate the normalized weight w_{ij} as follows:

$$w_{ij} = \begin{cases} 0, & \text{if } m_{ij} \in [0, M_1] \\ \dfrac{m_{ij} - M_1}{M_2 - M_1}, & \text{if } m_{ij} \in (M_1, M_2) \\ 1, & \text{if } m_{ij} \in [M_2, +\infty) \end{cases} \tag{1}$$

where m_{ij} is the number of matching points between I_i and I_j. For a dataset, the number of matching points for major image pairs is similar. Thus, to facilitate the graph partition method, we define two adaptive thresholds M_1 and M_2 to linearly stretch and normalize the weight. The calculation method of M_1 and M_2 is as follows. For each vertex v_i in the graph G, v_i usually has several edges. The number of matching points of each edge is different. For each vertex v_i, we first calculate the maximum

Fig. 3. Calculation process of the homography matching regions. (a) Original image pair. (b) Sparse matching points of two images. (c) Grid division of the first image. (d) For a valid grid that contains matching points, the local homography is calculated using the neighboring matching points. (e) Projection result of one grid. (f) Mask and (g) the final extracted matching regions.

number of matching points m_i as

$$m_i = \{\max(m_{ij}) | j \in [0, N), j \neq i, m_{ij} \neq 0)\}. \quad (2)$$

Then, we define a set $\mathcal{M} = \{m_i\}_{i=0}^{N}$ that represents the maximum number of matching points for all vertices. M_1 and M_2 are calculated as

$$M_1 = a_1 \times \min(\mathcal{M}), M_2 = a_2 \times \max(\mathcal{M}) \quad (3)$$

where a_1 and a_2 are two coefficients. $\min(\cdot)$ and $\max(\cdot)$ represent the minimum and maximum values of all elements included in $\mathcal{M}$, respectively. In general, we set $a_1 = 0.6$ and $a_2 = 0.8$.

After calculating the weights of all edges, we directly discard the edges with a weight equal to 0. However, there are still a lot of redundant edges in the graph. Thus, we further restrict that the maximum number of edges of each vertex is less than T. In general, we set $T = 5$. We find that $T = 5$ generates the similar color consistency result compared with no this restriction. The simplified graph is shown in Fig. 2(b).

Based on the simplified graph, we apply the graph partition algorithm to group images. The graph partition, also known as graph clustering, aims to divide a graph into many uniform parts. Graph clustering usually aims to maximize within-cluster association or minimize the cut between clusters. In this article, we directly apply the normalized cut method [28] to complete the task of graph partition. The normalized cut is to minimize the cut relative to the degree of a cluster instead of its size, which is formulated as follows:

$$NCuts(G) = \min \sum_{c=1}^{C} \frac{\sum_{I_i \in V_c, I_j \notin V_c} W_{ij}}{\text{degree}(V_c)} \quad (4)$$

where C is the number of clusters. V_c represents the cth cluster. $\text{degree}(V_c)$ is the degree of V_c, which is calculated as the sum of the degrees of vertices included in V_c. The result of graph partition is shown in Fig. 2(c).

B. Color Correspondence Extraction

The current multiview image color correction methods for 3-D reconstruction usually use sparse matching points as color correspondences, which is not reliable enough. Some researchers calculate the dense matching for each image pair to obtain a more accurate color correspondences, but it is time-consuming. We propose a method to balance these two factors, as shown in Fig. 3, which can effectively improve the accuracy of color correspondences and control the time consumption at a normal level.

The core idea of this method is to calculate the matching regions around the sparse matching points using local homography. For any pair of images (I_i, I_j), as shown in Fig. 3(a), we already obtained the sparse matching points in the stage of

graph construction. These sparse matching points are regarded as the guidance. First, we divide the first image I_i into $m \times n$ grids at the interval l_{grid}, as shown in Fig. 3(c). We will further check whether each grid includes the matching points. We only calculate the corresponding regions for the grids that include at least one matching point.

Inspired by the local image warping methods [29], we calculate the matching regions for each grid using local homography. We assume that the transformation between two local regions can be approximated by a local homography. For each grid in image I_i, to estimate the local homography matrix, we need to prepare enough matching points. The best choice is the one that uses the matching points included in the current grid. However, in most of cases, there are not enough matching points in the current grid. Thus, we search the matching points from the surrounding grids until the number of matching points reaches the predefined threshold. In our article, to ensure that the estimation of homograph matrix is robust, we set the number threshold as 8. Based on the matching points, the projection error functions are established and minimized to estimate the parameters of local homography matrix, as shown in Fig. 3(d). Then, we directly project the grid onto the second image I_j using the estimated homography matrix to find the corresponding matching regions, as shown in Fig. 3(e).

In Fig. 3(f) and (g), we visually show the final matching regions of two images. However, for the local regions located by the obvious objects such as buildings, the local homography only can roughly approximate the transformation between two images. Thus, there may be some outliers in matching regions. To suppress the influence of the outliers, instead of directly using the matching pixels as the color correspondences, we apply the statistic-based method to extract the color correspondences from the matching regions. We calculate the cumulative distribution functions (CDFs) for the matching regions of two images. Then, we take the same quantiles of two CDFs (color value pairs with the same frequency) as the color correspondences.

To illustrate the effectiveness of the proposed color correspondence extraction method, we compare it with the widely used strategy that applies the sparse matching points as the color correspondences. In Fig. 4, we present the color correction results offered by the homography matching regions and sparse matching points. From the first row of Fig. 4, we observe that the input image I_1 is dark and the input image I_3 is bright. When dealing with the darker image I_1, the correction result of the method using sparse points is obviously grayer, and our result is more vivid. For the brighter image I_3, our correction result has more natural color and moderate brightness.

In addition, to better illustrate why the method with the use of homography matching regions can generate the better results than that with the use of sparse matching points, we calculate the probability density function (PDFs) and CDFs of sparse matching points and homography matching points. In Fig. 5, we show the PDFs and CDFs of two pairs of images (I_1, I_2) and (I_3, I_4). It can be seen that when sparse matching points are used as color correspondences, the pixel value distribution is relatively concentrated, and the broken lines in the CDFs picture are also steep. After calculating the homography matching

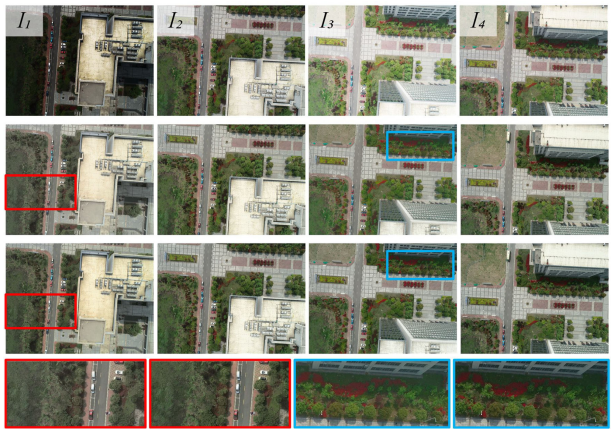

Fig. 4. (Row 1) Original images. (Rows 2 and 3) Correction results of the method using sparse matching points and our proposed method using homography matching pixels. (Row 4) Local enlarged details.

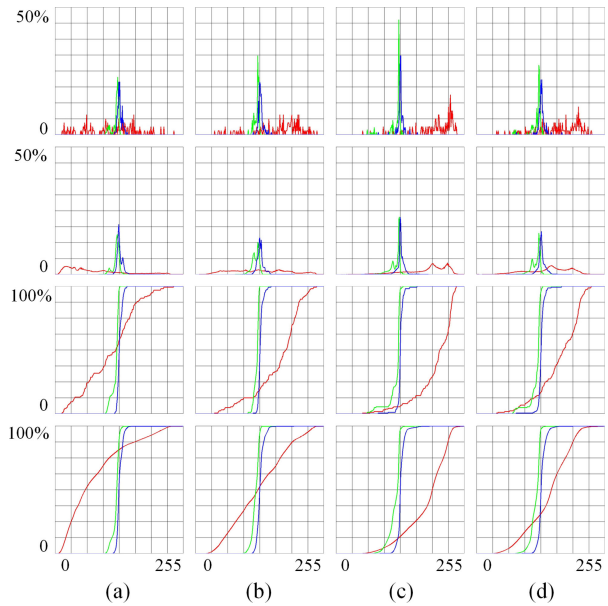

Fig. 5. PDFs and CDFs of two image pairs (I_1, I_2) and (I_3, I_4). (Rows 1 and 2) PDFs of sparse matching points and homography matching regions. (Rows 3 and 4) Corresponding CDFs. The red, green, and blue lines represent the Y, Cb, and Cr channels, respectively. (a)I_1. (b)I_2.(c) I_3.(d)I_4.

pixels, the range of pixel value distribution is wider, and the broken lines in the CDFs picture are also more gentle. This shows that using homography matching pixels as color correspondences can more accurately describe the color information of images.

C. Global Optimization

Global optimization, that is, obtains the color mapping models of all images. Based on the extracted color correspondences, we establish and minimize the global energy function to solve the optimal parameters of color mapping models. However, for large-scale images, solving all image correction parameters at the same time is not only time-consuming, but also easy

to fall into an unsolvable situation due to memory limitation. Therefore, we propose an efficient two-step strategy to indirectly achieve the effect of global optimization through intragroup and intergroup correction.

Let $\mathcal{I} = \{I_i\}_{i=1}^{N}$ denote the original images. After performing image grouping, all images are divided into many groups. Let $\mathcal{V} = \{V_c\}_{c=1}^{C}$ denote the groups. In the intragroup correction stage, we process each group of images independently. For each group V_c, global color consistency optimization is performed to obtain the color mapping models for images included in V_c. The CDs between the images included in V_c can be effectively eliminated. Let $\mathcal{F} = \{F_i\}_{i=1}^{N}$ denote the intragroup color mapping models of all images. In the intergroup correction stage, we treat each group of images as a virtual image. Let $\mathcal{J} = \{J_c\}_{c=1}^{C}$ denote all virtual images. The global color consistency optimization is performed again to eliminate the CDs between virtual images. Let $\mathcal{F}' = \{F_c'\}_{c=1}^{C}$ denote the intergroup color mapping models. For intragroup and intergroup color consistency optimization, we apply the same color correction method presented in our previous work [2] to effectively eliminate the CDs. The general definition of its global energy function is as follows:

$$E = \sum_{w_{ij}\neq 0} w_{ij} E_{\text{data}}(f_i, f_j) + \lambda \sum_{i=1}^{N} E_{\text{regulation}}(f_i)$$
$$\text{subject to}: C_{\text{rigid}}(f_i), \forall i \in [1, N] \tag{5}$$

where w_{ij} is the weight of edge e_{ij}, and λ represents the balance coefficient. f_i and f_j represent the correction models of the images I_i and I_j. $E_{\text{data}}(f_i, f_j)$ represents the data term to constrain the CDs between the corrected adjacent images to be as small as possible. The second part $E_{\text{regulation}}(f_i)$ enforces certain constraints on the color correction softly, including regularizing the parameters and stretching dynamic range. Finally, the increasing monotonicity and mapping domain are guaranteed through the rigid constraint term $C_{\text{rigid}}(f_i)$. For more details, refer to [2].

After solving the parameters of the mapping models, we can correct the images in turn. Generally speaking, for each pixel p in I_i included in group V_c, its correction process is formulated as

$$\widetilde{I}_i(p) = F_c'(F_i(I_i(p))) \tag{6}$$

where $I_i(p)$ and $\widetilde{I}_i(p)$ denote the intensity values of pixel p in the original image I_i and corrected image $\widetilde{I}_i$, respectively.

IV. Experiments

The experiments consist of three parts. In the first part, in order to determine the graph simplification parameters involved in Section III-A, experiments are conducted on several datasets. In the second part, to illustrate the effectiveness of our proposed color correction method, we compare the proposed method with several existing methods [4], [24], [25]. Because most of color correction methods cannot process large-scale datasets, we apply several small-scale datasets to conduct the comparative experiments. In the third part, to illustrate that the proposed method can efficiently and effectively process large-scale images, we evaluate the performance of the proposed method on a large-scale dataset.

A. Datasets and Evaluation Indicators

We evaluate the proposed color correction approach on various types of datasets. The details of each dataset are shown in Table I. In order to increase the challenge of color consistency optimization, we randomly adjust the original colors of the images included in School [26], YongZhou, and Cathedral [30] datasets. These three synthetic datasets are used for comparative experiments. Four Internet datasets, namely ShangQing, ZiXiao, Gendarmenmarkt [31], and NotreDame [32], are used to evaluate the correction effect of the proposed method on natural datasets. Three reconstruction datasets PiazzaDante, Navona [33], and Toledo [34] are used to evaluate the performance of our proposed method on the actual textured model. The last dataset WHU-MVS [35] is used to evaluate the performance of our proposed method on large-scale images. Furthermore, in the parameter determination experiments, two datasets are involved, School and Navona.

In addition, in order to quantitatively evaluate the performance of color correction methods, we apply the peak signal-to-noise ratio (PSNR) and Structural SIMilarity (SSIM) to measure the similarity of color and structure. The calculation formula of PSNR is as follows:

$$\text{PSNR} = 10 \times \log_{10} \frac{\text{MAX}^2}{\text{MSE}} \tag{7}$$

where MAX is the possible maximum pixel value of the image, and MSE is the mean square error of the two images. When calculating SSIM, a sliding window is generally used to divide the image into S blocks, and then the structural similarity of the corresponding block is calculated. Finally, the average value is used as the structural similarity measure of the two images. The calculation formulas are as follows:

$$\text{SSIM} = \frac{1}{S} \sum_{s=1}^{S} \text{SSIM}(x_s, y_s) \tag{8}$$

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{9}$$

where μ_x and μ_y are the mean values of the block x and y, σ_x and σ_y are the variances, and σ_{xy} is the covariance of x and y. c_1 and c_2 are two constants. Higher PSNR and SSIM scores mean that the corrected image is more similar to the original image, retaining more information of the original data. But these two metrics cannot evaluate the color consistency between images. Therefore, the scores of these two indicators cannot fully reflect the quality of the color correction results. For this reason, we additionally calculated another reference metric, the CD. It reflects the color consistency of the dataset by calculating the average of the differences between the color correspondences of adjacent images after color correction. The calculation formula

TABLE I
DETAILS OF DATASETS USED IN THE EXPERIMENTS

	School	YongZhou	Cathedral	ShangQing	ZiXiao	Gendarmenmarkt	NotreDame	PiazzaDante	Navona	Toledo	WHU-MVS
Number	55	20	34	16	17	17	15	39	92	87	1776
Width	4864	3680	1936	1343	4032	4000	1600	1144	4000	4000	5376
Height	3648	2456	1296	899	3024	3000	1200	856	3000	3000	5376
Platform	UAV	UAV	Handheld	Handheld	Handheld	Handheld	Handheld	Handheld	Handheld	UAV	UAV
Synthetic	YES	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO

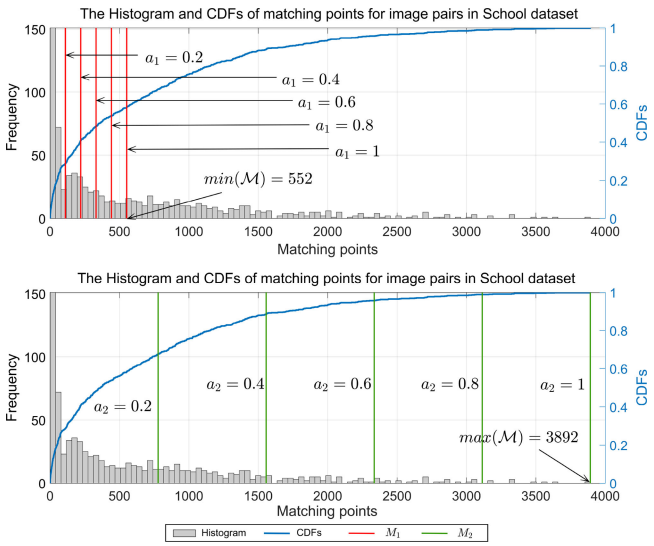

Fig. 6. Histogram and CDFs of matching points for image pairs in School datasets. The upper and lower figures show the positions of M_1 and M_2 within the matching point range when a_1 and a_2 take different values, respectively.

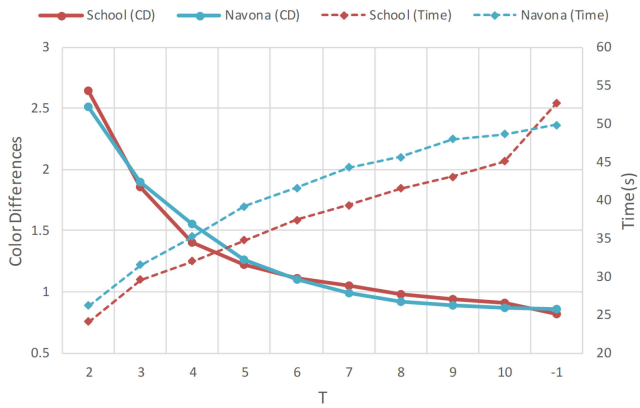

Fig. 7. Relationship between T and CD/time on two datasets ($T = -1$ means that there is no limit to the number of connected edges per vertex).

is as follows:

$$\mathrm{CD} = \frac{1}{N_{\mathrm{edges}} \cdot K} \sum_{w_{ij} \neq 0} \sum_{k=1}^{K} \|\hat{v}_{i,k} - \hat{v}_{j,k}\|_2 \qquad (10)$$

where N_{edges} is the number of edges whose weight is not zero, $(\hat{v}_{i,k}, \hat{v}_{j,k})$ represents the kth color correspondence values of the image pair after image correction, and K is the number of color correspondences. The smaller the CD, the better the color consistency between the images.

B. Parameter Determination

In Section III-A, we set three parameters, a_1, a_2, T, to simplify the graph. In this section, we conducted experiments to explore reasonable values for these parameters. First, we estimate a_1 and a_2 by taking statistics on the distribution of matching points for all edges in the graph. Fig. 6 shows the positions of M_1 and M_2 within the range of matching points when a_1 and a_2 take different values. The value of a_1 ranges from 0 to 1, and its main function is to exclude edges with few matching points. From the histogram distribution in Fig. 6, when the frequency is high, the number of matching points is small, which is less robust. Therefore, it is more appropriate for a_1 to be between 0.4 and 0.8, then the value of M_1 is between 220.8 and 441.6. The value range of a_2 is between a_1 and 1, mainly to prevent the weight value of the edge from being too concentrated due to the existence of extreme values. It can be seen from the CDFs curve in Fig. 6 that when a_2 is about 0.8, the CDFs value is close to 1. This shows that there are very few edges with more than this number of matching points. In general, we set $a_1 = 0.6$ and $a_2 = 0.8$.

Then, we conducted experiments on two datasets, School and Navona, to test the effect of different T on the correction results, as shown in Fig. 7. Here, the metric CD is used to indicate the correction effect of different T. It can be seen that when T is too small, there are few edges between image nodes, and the CD of the correction result is large. When T gradually increases, the CD of the correction results decreases and tends to be stable, and the corresponding processing time also increases gradually. In general, we set $T = 5$ to strike a balance between processing time and correction effect. We found that $T = 5$ produced similar color consistency results compared to $T = -1$.

C. Comparative Experiments

C. Qualitative Evaluation: To illustrate the superiority of our method, we compared the proposed color correction method with the other three methods [4], [24], [25] on three synthetic datasets, namely School, YongZhou, and Cathedral datasets. We named the selected three comparative methods as Moulon *et al.*'s method [24], Park *et al.*'s method [4], and Shen *et al.*'s method [25], respectively. Moulon *et al.*'s method has been integrated into openMVG [36]. Thus, we directly install openMVG to show the performance of this method. The *MATLAB* code

Fig. 8. Color correction results on School dataset. Limited by the space, only part of the images are displayed. (Row 1) Input images. (Rows 2–5) Results of Moulon *et al.*'s method [24], Park *et al.*'s method [4], Shen *et al.*'s method [25], and our proposed method.

Fig. 9. Color correction results on YongZhou dataset. Limited by the space, only part of the images are displayed. (Row 1) Input images. (Rows 2–5) Results of Moulon *et al.*'s method [24], Park *et al.*'s method [4], Shen *et al.*'s method [25], and our proposed method.

of Park *et al.*'s method is provided by the authors. Because the source code of the method presented in [25] is not provided, we reimplement Shen *et al.*'s method based on sparse matching points instead of reprojection.

Fig. 8 presents the experimental results of the four color correction methods on the School dataset. There are large exposure inconsistencies between the images included in the School dataset. It can be seen that Moulon *et al.*'s method [24] offers the worst results. The corrected images are hazy due to the loss of image gradient. In addition, the CDs between corrected images are still obvious. This method applies the linear model to approximate the CDs, and their color correction ability is relatively weak. Park *et al.* [4] used matching patches instead of sparse matching points as the color correspondences, so on most images in the School dataset, their results are similar to our results in terms of color consistency and image quality. However, their approach is not robust enough. When the brightness changes drastically, abnormal correction results may occur, such as the penultimate image. The results of Shen *et al.*'s method [25] are similar with the results generated by Moulon *et al.*'s method in color consistency. The contrast of the images corrected by Shen *et al.*'s method is better than the images corrected by Moulon *et al.*'s method. This is because Shen *et al.*'s method has linearly stretched the images to increase the contrast after the color correction. As shown in the last row of Fig. 8, the results of our proposed approach are the best.

Fig. 9 presents the experimental results of the four color correction methods on the YongZhou dataset. The images included in the YongZhou dataset have serious inconsistencies in terms of exposure and color. The results of Moulon *et al.*'s method [24]

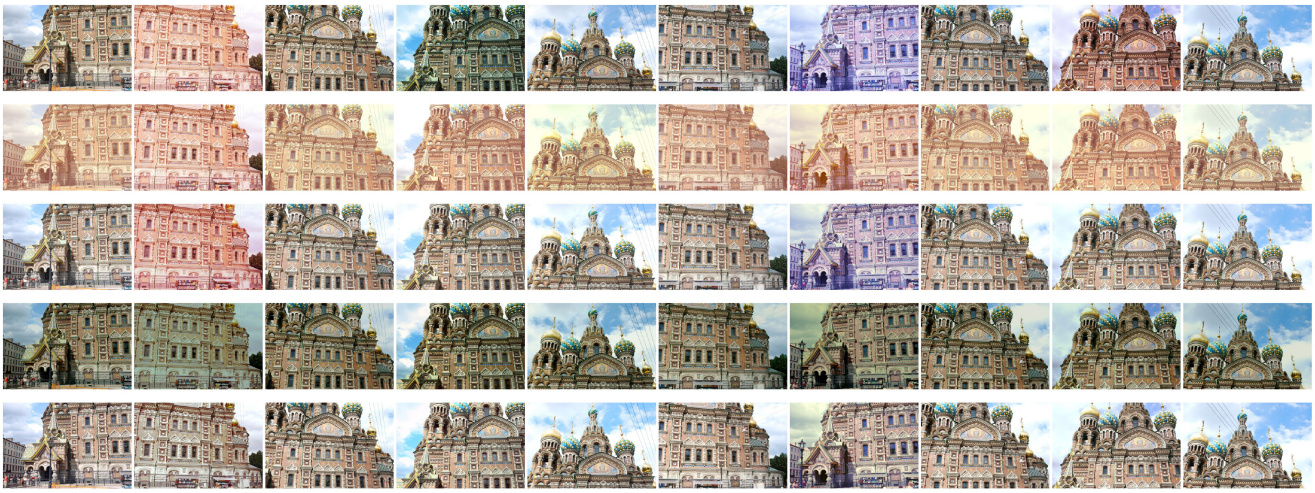

Fig. 10. Color correction results on Cathedral dataset. Limited by the space, only part of the images are displayed. (Row 1) Input images. (Rows 2–5) Results of Moulon *et al.*'s method [24], Park *et al.*'s method [4], Shen *et al.*'s method [25], and our proposed method.

TABLE II
QUANTITATIVE EVALUATION RESULTS OF FOUR METHODS ON THREE SYNTHETIC DATASETS

	School			YongZhou			Cathedral		
	PSNR	SSIM	CD	PSNR	SSIM	CD	PSNR	SSIM	CD
Moulon *et al.*'s method [24]	17.98692	0.86549	3.96771	**26.21981**	0.88125	2.17786	16.32965	0.86732	2.63666
Park *et al.*'s method [4]	**23.18811**	**0.92620**	1.98963	23.96554	**0.96092**	3.57422	26.26926	0.97533	3.21175
Shen *et al.*'s method [25]	19.44912	0.89320	2.85018	16.85603	0.86871	2.13007	15.48504	0.89058	2.06527
Our approach	22.58832	0.91946	**1.19950**	23.61740	0.93116	**0.90196**	**26.52322**	**0.97552**	**1.25453**

The bold entities means that this method has the best result of the comparison methods.

and Shen *et al.*'s method [25] have poor color consistency. In addition, the results of Moulon *et al.*'s method have serious loss of detail. The results of Shen *et al.*'s method have color overflow, such as the areas around the roads in the last few pictures. Except for the third image, the visual quality of the rest of images corrected by Park *et al.*'s method [4] is good. In this dataset, our method also offers the best correction results.

In the Cathedral dataset, there are strong color inconsistencies between the images, as shown in the first row of Fig. 10. From Fig. 10, it can be seen that the results of Moulon *et al.*'s method still have large residual CDs, and the tones of corrected images are unnatural. Park *et al.*'s method fails to correct the images with drastic tone differences. The tones of the second and seventh corrected images of Park *et al.*'s method are still different with other images. The results of Shen *et al.*'s method also have the problem of unnatural image tones. For this challenging dataset, our method still can generate the corrected images with invisible CDs and high visual quality; even the tone differences of input images are very large.

In addition, we tested the four methods on four natural Internet datasets, namely, ShangQing, ZiXiao, Gendarmenmarkt, and NotreDame datasets. Due to space limitations, only the results of our method are presented, as shown in Fig. 11. The quantitative evaluation results of the four methods on the Internet datasets are given in the next section. Since Internet pictures are taken by handheld cameras with different shooting angles and times, there are large CDs between original images. However, the proposed method performs well on these datasets. We successfully eliminate the CDs between the original images, especially for the areas of buildings.

C. Quantitative Evaluation: In addition to the qualitative evaluation, we also performed the quantitative evaluation to convincingly illustrate the effectiveness of our approach. The quantitative evaluation results of all methods are reported in Table II. Because we use the synthetic datasets (School, YongZhou, and Cathedral) to conduct the comparative experiment, the PSNR and SSIM values can be calculated by comparing the corrected images with the original images. For each dataset, the average values of PSNR and SSIM of all images are used as the final results. For the CD metric, as described in the formula, it is the average of the CDs of all image pairs. It can be seen that the scores of Moulon *et al.*'s method [24] and Shen *et al.*'s method [25] have relatively poor scores. This conclusion is consistent with the previous qualitative comparison results. Park *et al.*'s method [4] and our proposed method offer better PSNR and SSIM scores. It indicates that the corrected images have a similar color and structure with the original images. However, Park *et al.*'s method has higher CD scores. It indicates that the CD between the correction results is large. And on these three datasets, the CD scores of the correction results of our method are the best. Combined with the qualitative results, it can be concluded that our method is more robust than Park *et al.*'s

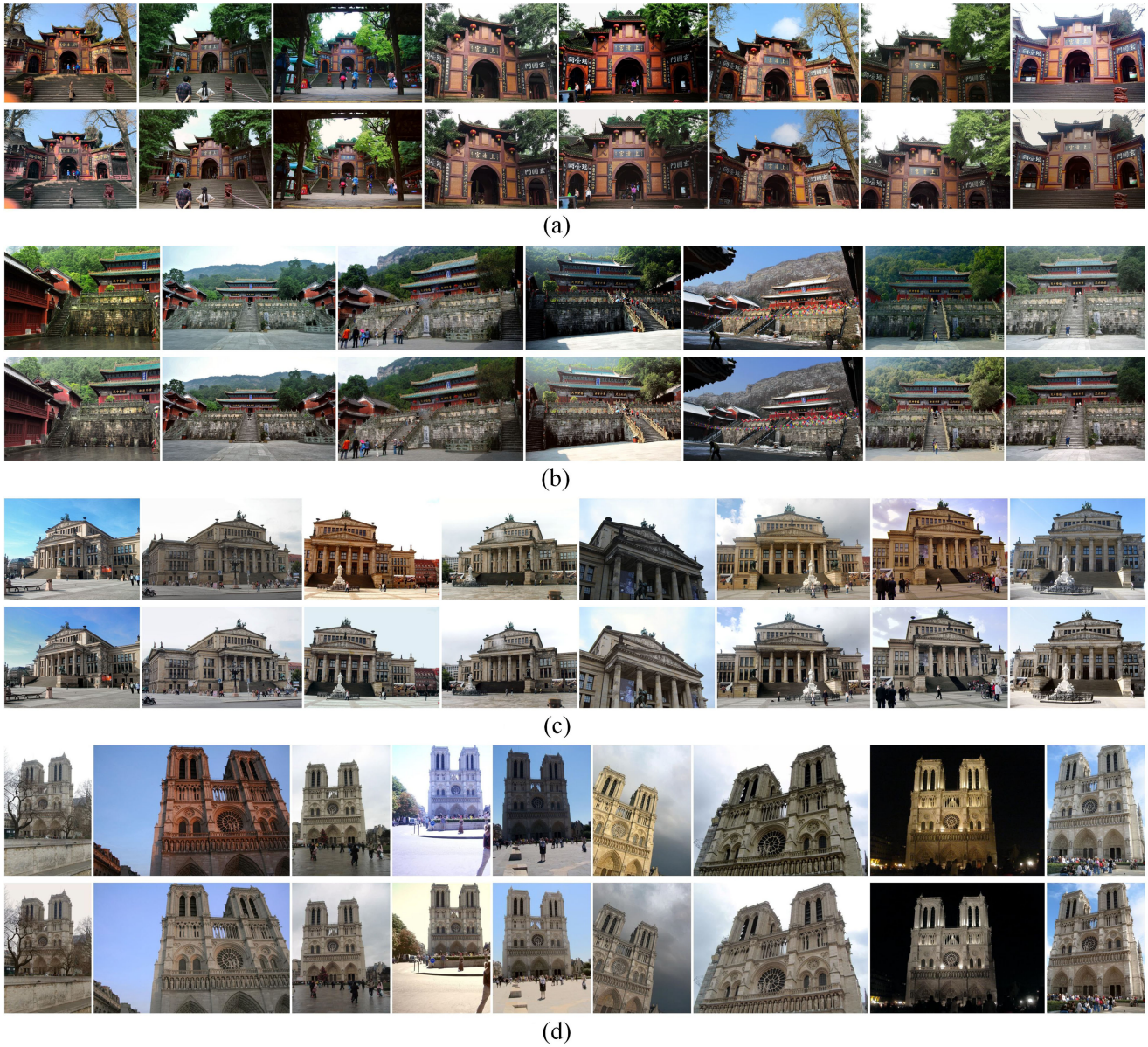

Fig. 11. Color correction results of four Internet datasets. The first rows of (a)–(d) show the original images, and the second rows of (a)–(d) show our color correction results. Limited by the length and space of manuscript, only part of the images in each dataset are displayed.

TABLE III
QUANTITATIVE EVALUATION RESULTS OF FOUR METHODS ON FOUR NATURAL INTERNET DATASETS

	ShangQing			ZiXiao			Gendarmenmarkt			NotreDame		
	PSNR	SSIM	CD	PSNR	SSIM	CD	PSNR	SSIM	CD	PSNR	SSIM	CD
Moulon *et al.*'s method [24]	20.743	0.851	6.692	22.340	0.895	4.965	21.352	0.919	6.078	17.666	0.837	5.216
Park *et al.*'s method [4]	**24.047**	**0.931**	5.128	**22.802**	**0.918**	4.114	**25.000**	**0.970**	2.444	**22.826**	**0.934**	2.874
Shen *et al.*'s method [25]	20.742	0.885	4.077	20.200	0.888	3.558	17.586	0.889	2.092	16.583	0.866	2.697
Our approach	23.804	0.915	**3.093**	22.410	0.908	**2.393**	23.396	0.948	**1.638**	21.705	0.920	**1.799**

The bold entities means that this method has the best result of the comparison methods.

method, and our method can achieve good color consistency even in the extreme cases.

We also conducted quantitative evaluations on four natural Internet datasets, as shown in Table III. As can be seen from the table, for the PSNR and SSIM metrics, Park *et al.*'s method [4] has the highest scores, and our method is the second. Different from synthetic datasets, in natural datasets, PSNR and SSIM reflect the similarity between the corrected images and the input

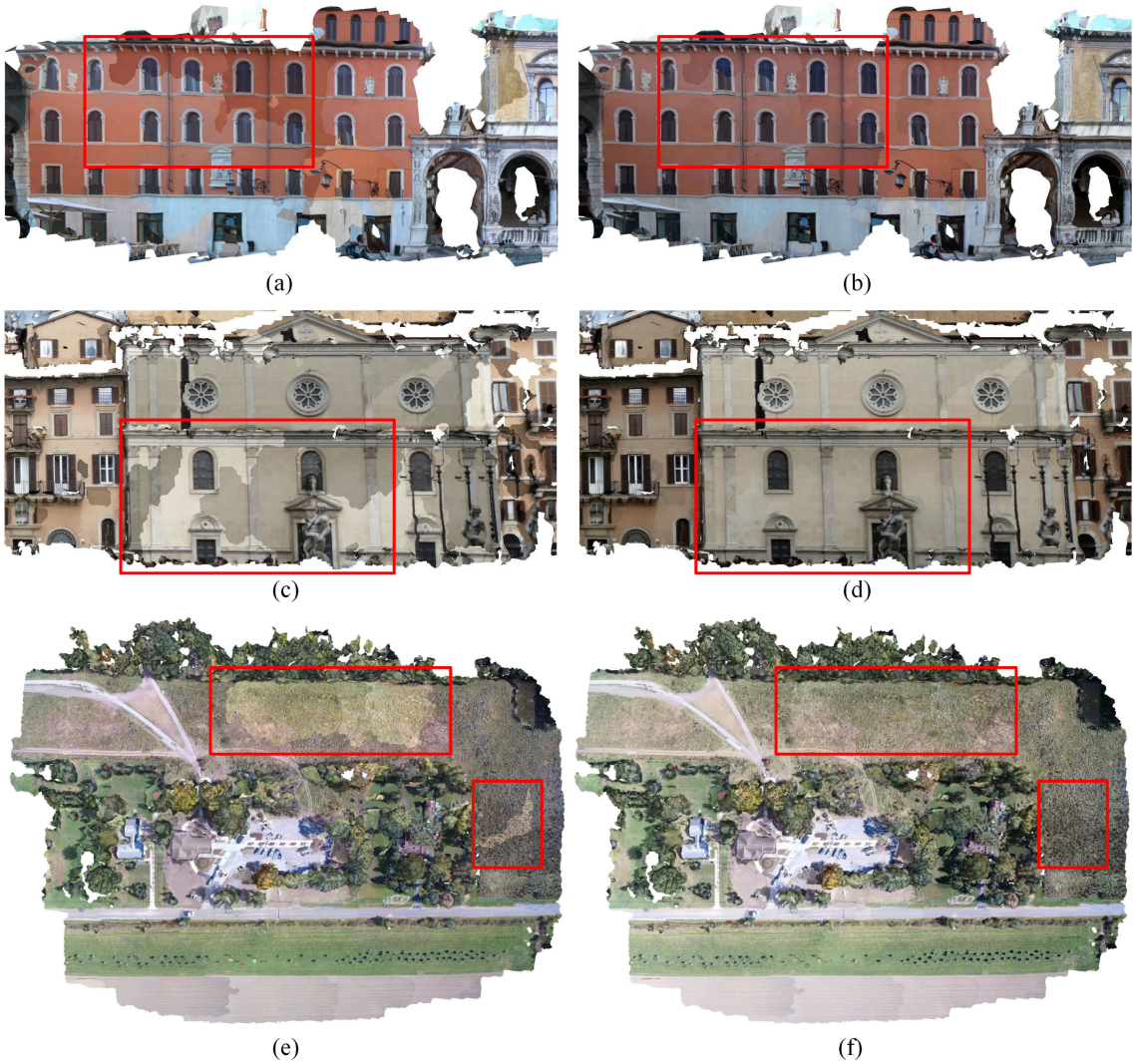

Fig. 12. Performance on textured models. (a), (c), and (e) Results without color correction on PiazzaDante, Navona, and Toledo. (b), (d), and (f) Results with color correction on PiazzaDante, Navona, and Toledo.

images. A high score simply means that the corrected image is similar to the input image. However, the input images have obvious CDs and brightness differences, and are not suitable as the reference image. Therefore, the PSNR and SSIM scores are only used as auxiliary information here. And for the CD metric, our method gets the best scores, followed by Shen *et al.*'s method [25]. This also supports our previous conclusion that our method achieves better color consistency results.

C. Texture Mapping: In the previous experiments, we tested our method on several synthetic and natural datasets. We illustrated that our method can generate the images with consistent tone and high quality. Then, we attempted to illustrate that our method can improve the visual quality of the textured 3-D models. We tested our method on PiazzaDante, Navona, and Toledo datasets. First, we reconstructed the 3-D models for these datasets using multiview environment [37]. Then, the textured 3-D models were generated using the texture mapping method presented in [38]. We applied the original images and the corrected images generated by our method to complete the texture mapping, respectively. The textured models with the use of original and corrected images are presented in Fig. 12. It can be seen that there are large CDs in the textured models when the original images are applied for texture mapping, as shown in the Fig. 12(a), (c), and (e). The visible seams would significantly degrade the visual quality of the textured 3-D models. With the use of the corrected images generated by our method, the visible seams disappear as our expect, as shown in Fig. 12(b), (d), and (f).

D. Large-Scale Dataset

A series of previous experiments on small datasets proved that the color correction ability of our proposed method is better. Then, we attempted to illustrate that our method can process large-scale datasets efficiently. In order to evaluate the effectiveness of our proposed two-step strategy, we conducted experiments on the large-scale dataset WHU-MVS. The dataset has 1776 images, the image size is 5376×5376, and the number

TABLE IV
EQUATION SOLVING TIME COMPARISONS OF THREE METHODS ON DIFFERENT NUMBERS OF IMAGES IN THE WHU-MVS DATASET (UNIT: SECOND)

Number of images	27	60	110	160	222	444	888	1110	1332	1554	1776
Moulon *et al.*'s method [24]	1.00	21.00	255.00	1587.00	6305.00	N/A	N/A	N/A	N/A	N/A	N/A
Shen *et al.*'s method [25]	0.04	0.21	1.09	2.91	8.08	53.62	391.01	769.41	1303.01	2047.49	Failed
Our approach without grouping	0.03	0.14	0.66	1.99	5.13	43.14	339.84	671.74	1121.66	1823.18	Failed
Our approach with grouping	0.03	0.14	0.66	1.99	1.37	3.08	6.26	7.08	8.72	10.64	11.14

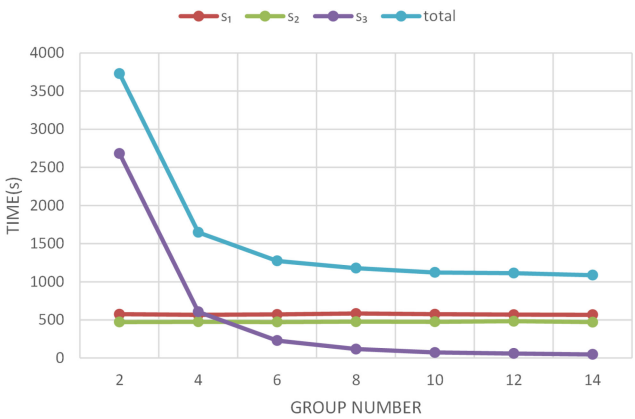

Fig. 13. Relationship between group number and time consumption on WHU-MVS dataset (s_1, s_2, and s_3 represent image I/O, color correspondence extraction, and correction parameter solving, respectively).

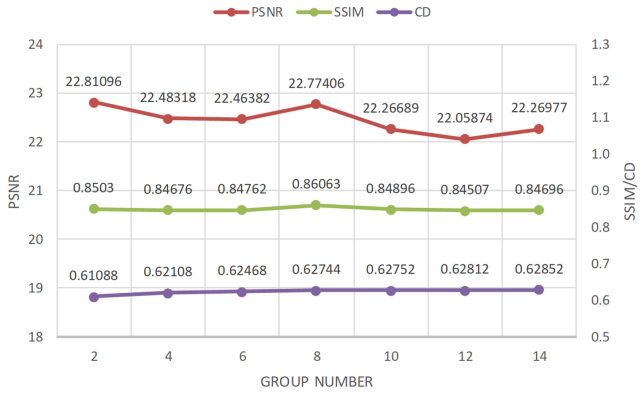

Fig. 14. Relationship between group number and PSNR/SSIM/CD on WHU-MVS dataset.

of image pairs is 5040. The experimental device is a desktop computer with 16 GB RAM and an Intel i7-8700 CPU running at 3.2 GHz. In order to facilitate the experiments, all images in WHU-MVS are resampled to 1344×1344. The relationship between the algorithm time consumption and the number of groups is presented in Fig. 13.

The main computation of our proposed algorithm includes three parts, namely image I/O, color corresponding calculation, and correction parameters solving. We named the time spent in the three parts as s_1, s_2, and s_3, respectively. s_1 is mainly related to the image size and the number of images. And s_2 is positively correlated with the number of image pairs. The proposed two-step strategy has nothing to do with these two parts, but for the convenience of comparison and understanding, they are shown together in Fig. 13.

The third part s_3 is the time to construct and solve the energy equation. We optimize this part through the proposed two-step strategy. By controlling the number of image groups, we can analyze the relationship between the time consumption and the number of groups. When the number of groups is 1, a *out of memory* error occurs, so the number of groups in the experiment starts from 2. From Fig. 13, it can be seen that as the number of groups increases, s_3 decreases sharply. When the number of groups is 2, there are about 888 images in each group, and the total solution time exceeds about 2500 s. When divided into 14 groups, there are about 126 images in each group, and the total solution time is only about 30 s. This proves that the proposed two-step strategy can effectively improve the computational efficiency of the correction algorithm on large-scale datasets. In addition, there is no parallel optimization in the above experiment. If parallel optimization is performed, s_1, s_2, and s_3 will all benefit.

Finally, we attempted to illustrate that the two-step strategy will not have a significant impact on the correction results. Therefore, the PSNR, SSIM, and CD scores were calculated for each set of correction results. The relationship between group number and PSNR/SSIM/CD is presented in Fig. 14. It can be seen that as the number of groups increases, PSNR and SSIM show a slight downward trend, while CD shows a slightly increasing trend. This shows that when the number of groups is different, the correction results are similar. In summary, our proposed method can not only greatly improve the efficiency of the algorithm, but also ensure the color correction effect. This meets the needs of color consistency correction for large-scale images.

On the large-scale dataset WHU-MVS, we compared the equation solving time of three methods by controlling the number of images involved in the operation, as shown in Table IV. We did not compare with Park *et al.*'s method [4], because this method is implemented in Matlab, so the efficiency is very low. Moulon *et al.*'s method [24] is implemented on the linux platform. It can be seen from the table that its solution efficiency is lower than the latter two methods. When the number of images is small, such as 110 images, it takes 255 s for Moulon *et al.*'s method [24] to solve the model parameters, but only 1.09 s for Shen *et al.*'s method [25]. The solution efficiency of our method without grouping is similar to Shen *et al.*'s method [25]. However, when the number of images is large, the time to solve the equations of Moulon *et al.*'s method [24] and Shen *et al.*'s method [25] increases greatly, and may even be impossible

to solve. After applying a two-step strategy based on image grouping, it can be seen that our method takes a significant reduction in the time to solve the equations. In our method with grouping, image grouping is applied when the number of images is greater than 200. Because grouping is not necessary when the number of images is small, we simply set the number of groups equal to the number of images divided by 111. The above time comparison experiments demonstrate that our method can handle color correction on large-scale datasets more efficiently.

V. Conclusion

In this article, we propose a color consistency correction approach for large-scale images in image-based 3-D modeling. The proposed approach obtains more accurate color correspondences through local homography matching, so the final color correction results are more reliable. In addition, this approach proposes a two-step strategy based on graph partition, which effectively reduces the computational burden when correcting large-scale images. Experiments show that our approach can achieve good color consistency results on various datasets, and outperforms the existing color correction methods. Although the matching regions applied in our method can get more color information than sparse matching points, it still unavoidably ignores some low-frequency information of image. In the future work, we will pay attention to the problem of color consistency correction in weak texture areas.

References

[1] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, 2007.

[2] M. Xia, J. Yao, R. Xie, M. Zhang, and J. Xiao, "Color consistency correction based on remapping optimization for image stitching," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 2977–2984.

[3] L. Li, Y. Li, M. Xia, Y. Li, J. Yao, and B. Wang, "Grid model-based global color correction for multiple image mosaicking," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 11, pp. 2006–2010, Nov. 2021.

[4] J. Park, Y. W. Tai, S. N. Sinha, and I. So Kweon, "Efficient and robust color consistency for community photo collections," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 430–438.

[5] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Optimizing color consistency in photo collections," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 1–10, 2013.

[6] F. Pitie, A. C. Kokaram, and R. Dahyot, "N-dimensional probability density function transfer and its application to color transfer," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1434–1439.

[7] Z. Su, D. Deng, X. Yang, and X. Luo, "Color transfer based on multiscale gradient-aware decomposition and color distribution mapping," in *Proc. ACM Int. Conf. Multimedia*, 2012, pp. 753–756.

[8] Y. Hwang, J. Y. Lee, I. So Kweon, and S. Joo Kim, "Color transfer using probabilistic moving least squares," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3342–3349.

[9] X. Liu, L. Zhu, S. Xu, and S. Du, "Palette-based recoloring of natural images under different illumination," in *Proc. Int. Conf. Comput. Commun. Syst.*, 2021, pp. 347–351.

[10] F. Wu, W. Dong, Y. Kong, X. Mei, J. C. Paul, and X. Zhang, "Content-based colour transfer," *Comput. Graph. Forum*, vol. 32, no. 1, pp. 190–203, 2013.

[11] Y. Hwang, J. Y. Lee, I. S. Kweon, and S. J. Kim, "Probabilistic moving least squares with spatial constraints for nonlinear color transfer between images," *Comput. Vis. Image Understanding*, vol. 180, pp. 1–12, 2019.

[12] Y. Niu, X. Zheng, T. Zhao, and J. Chen, "Visually consistent color correction for stereoscopic images and videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 697–710, Mar. 2020.

[13] F. Pitié, A. C. Kokaram, and R. Dahyot, "Automated colour grading using colour distribution transfer," *Comput. Vis. Image Understanding*, vol. 107, no. 1/2, pp. 123–137, 2007.

[14] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, Jul./Aug. 2001.

[15] X. Xiao and L. Ma, "Gradient-preserving color transfer," *Comput. Graph. Forum*, vol. 28, no. 7, pp. 1879–1886, 2009.

[16] R. M. Nguyen, S. J. Kim, and M. S. Brown, "Illuminant aware gamut-based color transfer," *Comput. Graph. Forum*, vol. 33, no. 7, pp. 319–328, 2014.

[17] M. He, J. Liao, D. Chen, L. Yuan, and P. V. Sander, "Progressive color transfer with dense semantic correspondences," *ACM Trans. Graph.*, vol. 38, no. 2, pp. 1–18, 2019.

[18] Z. Wu and R. Xue, "Color transfer with salient features mapping via attention maps between images," *IEEE Access*, vol. 8, pp. 104884–104892, 2020.

[19] J. Pan, M. Wang, D. Li, and J. Li, "A network-based radiometric equalization approach for digital aerial orthoimages," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 401–405, Apr. 2010.

[20] C. Chen, Z. Chen, M. Li, Y. Liu, L. Cheng, and Y. Ren, "Parallel relative radiometric normalisation for remote sensing image mosaics," *Comput. Geosci.*, vol. 73, pp. 28–36, 2014.

[21] Y. Xiong and K. Pulli, "Color matching of image sequences with combined gamma and linear corrections," in *Proc. Int. Conf. ACM Multimedia*, 2010, pp. 261–270.

[22] L. Yu, Y. Zhang, M. Sun, X. Zhou, and C. Liu, "An auto-adapting global-to-local color balancing method for optical imagery mosaic," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 1–19, 2017.

[23] R. Xie, M. Xia, J. Yao, and L. Li, "Guided color consistency optimization for image mosaicking," *ISPRS J. Photogramm. Remote Sens.*, vol. 135, pp. 43–59, 2018.

[24] P. Moulon, B. Duisit, and P. Monasse, "Global multiple-view color consistency," in *Proc. Conf. Vis. Media Prod.*, 2013.

[25] T. Shen, J. Wang, T. Fang, S. Zhu, and L. Quan, "Color correction for image-based modeling in the large," in *Proc. Asian Conf. Comput. Vis.*, 2016, pp. 392–407.

[26] J. Yang, L. Liu, J. Xu, Y. Wang, and F. Deng, "Efficient global color correction for large-scale multiple-view images in three-dimensional reconstruction," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 209–220, 2021.

[27] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4104–4113.

[28] I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1944–1957, Nov. 2007.

[29] J. Zaragoza, T. J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2339–2346.

[30] O. Enqvist, F. Kahl, and C. Olsson, "Non-sequential structure from motion," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2011, pp. 264–271.

[31] W. Kyle and S. Noah, "Robust global translations with 1DSfM," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 61–75.

[32] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," in *Proc. ACM SIGGRAPH*, 2006, pp. 835–846.

[33] M. Farenzena, A. Fusiello, and R. Gherardi, "Structure-and-motion pipeline on a hierarchical cluster tree," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2009, pp. 1489–1496.

[34] OpenDroneMap, "Example datasets for OpenDroneMap," 2021. [Online]. https://github.com/OpenDroneMap/ODMdata#odmdata

[35] J. Liu and S. Ji, "A novel recurrent encoder–decoder structure for large-scale multi-view stereo reconstruction from an open aerial dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6050–6059.

[36] P. Moulon, P. Monasse, R. Perrot, and R. Marlet, "OpenMVG: Open multiple view geometry," in *Proc. Int. Workshop Reproducible Res. Pattern Recognit.*, 2016, pp. 60–74.

[37] S. Fuhrmann, F. Langguth, and M. Goesele, "MVE-A multi-view reconstruction environment," in *Proc. Eurographics Workshop Graph. Cultural Heritage*, 2014, pp. 11–18.

[38] M. Waechter, N. Moehrle, and M. Goesele, "Let there be color! large-scale texturing of 3D reconstructions," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 836–850.

Yunmeng Li received the B.E. degree in surveying and mapping engineering from Northeastern University, Shenyang, China, in 2019. She is currently working toward the M.E. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China.

Her research interests include color consistency correction for multiple images, image mosaicking, and image processing.

Yinxuan Li received the B.E. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2014. She is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University.

Her research interests include color correction for multiple images, texture mapping for 3-D models, image-based 3-D reconstruction, 2-D image processing and analysis, and machine learning.

Jian Yao received the B.Sc. degree in automation from Xiamen University, Xiamen, China, in 1997, the M.Sc. degree in computer science from Wuhan University, Wuhan, China, and the Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, Hong Kong, in 2006.

He was a Research Assistant with Shenzhen R&D Centre of City University of Hong Kong, from 2001 to 2002; a Postdoctoral Fellow with Computer Vision Group, IDIAP Research Institute, Martigny, Switzerland, from 2006 to 2008; a Research Grantholder with the Institute for the Protection and Security of the Citizen, European Commission Joint Research Centre (JRC), Ispra, Italy, from 2009 to 2011; and a Professor with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences, China, from 2011 to 2012. Since April 2012, he has been a Hubei "Chutian Scholar" Distinguished Professor with the School of Remote Sensing and Information Engineering, Wuhan University, where he is currently the Director of Computer Vision and Remote Sensing (CVRS) Lab. He has authored or coauthored more than 100 papers in international journals and proceedings of major conferences and holds more than 30 patents. His research interests include computer vision, image processing, deep learning, LiDAR data processing, robotics, SLAM, etc.

Ye Gong received the B.E. degree in automation from the Huazhong University of Science and Technology and the M.E. degree in detection technology and automation equipment from the Huazhong Institute of Photoelectronic Technology, China, in 2015 and 2018, respectively. He is currently working toward the doctorate degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China.

His research interests include structure from motion, simultaneously location and mapping, and image processing.

Li Li received the B.E., M.S., and Ph.D. degrees in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2013, 2016, and 2019, respectively.

He is currently an Associate Researcher with the School of Remote Sensing and Information Engineering, Wuhan University. His research interests include image mosaicking, texture mapping, LiDAR data processing, and machine learning.