FUSION OF MULTI-ANGULAR AERIAL IMAGES BASED ON EPIPOLAR GEOMETRY AND MATRIX COMPLETION

Yanting Ma [*], *Dehong Liu* [†], *Hassan Mansour*[†],
Ulugbek S. Kamilov[†], *Yuichi Taguchi* [†], *Petros T. Boufounos*[†], *and Anthony Vetro*[†]

[*]Department of ECE, North Carolina State University, Raleigh, NC 27606
[†] Mitsubishi Electric Research Laboratories, Cambridge, MA 02139

ABSTRACT

We consider the problem of fusing multiple cloud-contaminated aerial images of a 3D scene to generate a cloud-free image, where the images are captured from multiple unknown view angles. In order to fuse these images, we propose an end-to-end framework incorporating epipolar geometry and low-rank matrix completion. In particular, we first warp the multi-angular images to single-angle ones based on the estimated fundamental matrices that relate the multi-angular images according to their projective relations to the 3D scene. Then we formulate the fusion process of the warpped images as a low-rank matrix completion problem where each column of the matrix corresponds to a vectorized image with missing entries corresponding to cloud or occluded areas. Results using DigitalGlobe high spatial resolution images demonstrate that our algorithm outperforms existing approaches.

Index Terms— image fusion, multi-angular, epipolar geometry, matrix completion.

1. INTRODUCTION

A common problem in aerial photography is that acquired optical aerial images are contaminated by clouds obscuring the view. This hinders the usability of the images in applications such as urban planning and environmental monitoring. Due to the high cost of aerial image acquisition, it is desirable to generate cloud-free images by fusing multiple cloud-contaminated images captured at different time or even from different view angles.

Various image fusion methods have been proposed in the literature to tackle the cloud removal problem as well as general image fusion problems [1–4]. For instance, Agarwala *et al.* [4] proposed a graph-cut based digital photomontage method, which uses graph cut to select the best image for each pixel while making the seams as invisible as possible. With the development of compressive sensing (CS) techniques, image fusion has recently been modeled as CS reconstruction [5], a matrix completion [6], or a robust principal component analysis (PCA) [7, 8] problem. Wang *et al.* [6] assumed that the scene has simple structure and that pixels at the same spatial location represent the same material along the temporal dimension. Under these assumptions, the matrix formed by concatenating the vectorized images has low rank, and robust matrix completion was utilized to recover the missing pixels. Aravkin *et al.* [8] developed a variational approach based on robust PCA, where cloud-contaminated images, as an example, are decomposed into two parts:

the sparse part which corresponds to clouds and the low-rank part which corresponds to cloud-free images. To improve the computational efficiency, Zhang *et al.* [9] developed an online robust tensor PCA method to process sequentially collected images for the cloud removal application.

The aforementioned fusion methods are generally valid for processing well-aligned images such as multi-temporal images viewed from the same angle or slightly different view angles but with low spatial resolution. However, they perform poorly on multi-angular high spatial resolution images mainly for the following two reasons. First, these methods ignore the parallax effect of 3D scenes. Images captured from different view angles are typically not well aligned due to the parallax effect, especially images of urban areas where buildings exhibit different elevations. For CS-based reconstruction methods that require critical alignment to use a low-rank prior, a slight mis-alignment of the images may lead to significantly blurred fusion results and therefore decrease the spatial resolution. Second, these methods are incapable of recovering occluded area since all images are assumed to be captured from the same view angle. To meet the requirement of modern high spatial resolution image applications, it is imperative to develop effective algorithms suitable for fusing multi-angular images of 3D scenes.

To address these problems, we propose an end-to-end image fusion framework for multi-angular aerial images incorporating epipolar geometry and matrix completion techniques. In particular, we warp all images of different view angles to a target view angle based on our estimated epipolar geometry. Then we formulate the image fusion process as a low-rank matrix completion problem, where each column of the matrix corresponds to a vectorized warped image and missing entries correspond to cloud or occluded areas to be recovered. We examine our method on a collection of multi-angular high spatial resolution images without any prior knowledge of view angles or elevation information of the scene. Our main contribution of this paper is to combine 3D epipolar geometry and low-rank matrix completion such that images of a 3D scene captured from different view angles can be fused even for misaligned images. Besides the cloud removal application of aerial images, our method can also be extended to other image enhancement applications of fusing non-aerial multi-angular images.

2. IMAGE FUSION FRAMEWORK

2.1. Image Acquisition

Our dataset consists of a total of $N(N \geq 3)$ multi-angular aerial images $\{\mathbf{I}_n\}_{n=1,...,N}$ of a common urban area. Considering the influence of clouds and different sun light directions, we model the

acquired image $\mathbf{I}_n$ as

$$\mathbf{I}_n = \mathbf{M}_{\Omega_n} \odot (\mathbf{\Psi}(\theta_n)\mathbf{V}) + \mathbf{M}_{\Omega_n^C} \odot \mathbf{C}_n + \mathbf{M}_{\Omega_n^S} \odot \mathbf{S}_n, \quad (1)$$

where $\mathbf{M}_{\Omega_n}$, $\mathbf{M}_{\Omega_n^C}$, and $\mathbf{M}_{\Omega_n^S}$ are binary masks of corresponding visible scene area Ω_n, cloud area Ω_n^C, and shadow area Ω_n^S in the nth image respectively; $\odot$ represents element-wised product; $\mathbf{V}$ represents the 3D scene of interest; $\mathbf{\Psi}$ accounts for the 6 degree-of-freedom camera using pin-hole projection model parameterized by θ_n; $\mathbf{C}_n$ and $\mathbf{S}_n$ stand for clouds and shadows respectively. The area union $\Omega = \Omega_n \cup \Omega_n^C \cup \Omega_n^S$ covers the whole image area.

The goal is to recover the cloud- or shadow-contaminated areas in a pre-defined target image $\mathbf{I}_m$ with $m \in \{1, 2, ..., N\}$, using the observed image set $\{\mathbf{I}_n\}_{n=1,...,N}$ such that the reconstructed image $\hat{\mathbf{I}}_m$, which ideally is $\mathbf{\Psi}(\theta_m)\mathbf{V}$, can be interpreted for further applications. In the special case when the camera parameters θ_n are constant values for all $n = 1, ..., N$, meaning that all images are captured from the same view angle, we have well-aligned images since $\mathbf{\Psi}(\theta_n)\mathbf{V}$ are identical, and therefore the fusion problem is much simpler to solve than our current one. The main challenge of our problem is that the camera angles vary for different images and their specific values are unknown a priori. It may significantly degrade the fusion performance by simply treating them as constant, as to be shown in Fig. 2 (d). Fortunately, we can make use of epipolar geometry to estimate the correspondence of any pair of the multi-angular images and align them using point transfer.

Our framework of image fusion contains three main steps, which are cloud and shadow detection, image warping, and image fusion. Since the first step can be performed using some straightforward methods based on the intensity and variance of small patches [10], we only address the later two steps in the following sections.

2.2. Image Warping using Epipolar Geometry

2.2.1. Basics of Epipolar Geometry

In some applications, an affine camera model, which assumes camera at infinity, is sufficient for processing images with weak perspective. In this scenario, one can estimate a 2D homography from at least four correspondence points in the visible areas, and then the entire image can be warped using the same homography [11]. However, the affine camera model is unable to capture the parallax effect of 3D objects. To process images captured from multiple view angles, it is necessary to consider the epipolar geometry using the projective camera model [12], where the point transfer involves the epipolar geometry among at least three images.

Let us consider an example where three images of a 3D scene, which are the ith, jth, and mth ($i, j, m \in \{1, 2, ..., N\}$, and $i \neq j \neq m$) of the N images respectively, are captured from three different view angles. Let $\mathbf{x} = (x_1, x_2, x_3)^T$, $\mathbf{x}' = (x_1', x_2', x_3')^T$, and $\mathbf{x}'' = (x_1'', x_2'', x_3'')^T$ be the homogeneous coordinates of the three images, respectively. According to the epipolar geometry [12], if $\mathbf{x}$ and $\mathbf{x}'$ are correspondence points, denoted by $\mathbf{x} \leftrightarrow \mathbf{x}'$, there exists a fundamental matrix of the ith and the jth images $\mathbf{F}_{ji} \in \mathbb{R}^{3 \times 3}$ of rank 2 satisfying

$$\mathbf{x}'^T \mathbf{F}_{ji} \mathbf{x} = 0, \quad (2)$$

for correspondence point set $\{\mathbf{x} \leftrightarrow \mathbf{x}'\}$. Similarly, we have

$$\begin{cases} \mathbf{x}''^T \mathbf{F}_{mi} \mathbf{x} = 0, & \text{for } \{\mathbf{x} \leftrightarrow \mathbf{x}''\}, \\ \mathbf{x}''^T \mathbf{F}_{mj} \mathbf{x}' = 0, & \text{for } \{\mathbf{x}' \leftrightarrow \mathbf{x}''\}. \end{cases}$$

Let $\mathbf{l} = \mathbf{F}_{mi}\mathbf{x}$ and $\mathbf{l}' = \mathbf{F}_{mj}\mathbf{x}'$ be epipolar lines corresponding to points $\mathbf{x}$ and $\mathbf{x}'$, respectively. Since $\mathbf{x}''\mathbf{l} = 0$ and $\mathbf{x}''\mathbf{l}' = 0$, or $\mathbf{x}''$

lies in the both lines $\mathbf{l}$ and $\mathbf{l}'$, $\mathbf{x}''$ can be computed as

$$\mathbf{x}'' = \mathbf{l} \times \mathbf{l}' = (\mathbf{F}_{mi}\mathbf{x}) \times (\mathbf{F}_{mj}\mathbf{x}'). \quad (3)$$

Note that our goal is to recover missing $\mathbf{x}''$ in the cloud- or shadow-contaminated areas in a target image $\mathbf{I}_m$. Thus the correspondences $\mathbf{x} \leftrightarrow \mathbf{x}''$ and $\mathbf{x}' \leftrightarrow \mathbf{x}''$ are not available. Nevertheless, the point transfer formula (3) enables us to obtain $\mathbf{x}''$ without requiring the correspondences $\mathbf{x} \leftrightarrow \mathbf{x}''$ or $\mathbf{x}' \leftrightarrow \mathbf{x}''$. To this end, we first estimate the fundamental matrices $\mathbf{F}_{ji}$ for all image pairs $\{i, j\}$ from the cloud-free areas, and then find dense correspondence points in the visible areas by applying a modified SIFT-flow.

2.2.2. Fundamental Matrix Estimation

To estimate the fundamental matrices $\mathbf{F}_{ji}$ for all pairs $\{i, j\}$, we acquire the correspondence points by matching the SIFT features in Euclidean distance at the key points detected by the SIFT detector [13]. Specifically, let Ω_i^{SIFT} and Ω_j^{SIFT} be the coordinate sets of the key points in the ith and jth image, respectively. Then the correspondence points $\{\mathbf{x} \leftrightarrow \mathbf{x}'\}$ are defined to be all pairs from $\mathbf{x} \in \Omega_i^{\text{SIFT}}$ and $\mathbf{x}' \in \Omega_j^{\text{SIFT}}$ that satisfy the following condition

$$\|s_i(\mathbf{x}) - s_j(\mathbf{x}')\|_2 \leq \mathcal{C} \cdot \min_{\tilde{\mathbf{x}}' \in \Omega_j^{\text{SIFT}} \setminus \mathbf{x}'} \|s_i(\mathbf{x}) - s_j(\tilde{\mathbf{x}}')\|_2, \quad (4)$$

where $s_i(\mathbf{x})$ represents the SIFT feature vector at $\mathbf{x}$ in the ith image, $\Omega_j^{\text{SIFT}} \setminus \mathbf{x}'$ denotes the set Ω_j^{SIFT} excluding the singleton $\mathbf{x}'$, and $\mathcal{C}$ is a constant whose value $0 < \mathcal{C} < 1$, typically $\mathcal{C} = 2/3$.

In order to minimize the influence of outlier correspondence points in estimating the fundamental matrix, we consider the well-known random sample consensus (RANSAC) [14] method to fit (2) using the set of initial correspondence points such that a robust estimation of $\mathbf{F}_{ji}$ can be achieved.

2.2.3. Dense Correspondence Points Estimation and Transfer

Note that we have estimated the fundamental matrix for each image pair. In order to perform point transfer using (3), we need to find dense correspondence points $\{\mathbf{x} \leftrightarrow \mathbf{x}'\}$. SIFT-flow is a widely used algorithm that finds the displacement of the pixels in two images by solving an optimization problem [15]. The SIFT-flow algorithm, however, can not be applied to search correspondence points directly in our application for the following reason. Although SIFT-flow can warp one image such that it is visually similar to another image based on SIFT features and the smoothness of images, namely, neighboring pixels should have similar displacement, it does not guarantee that the estimated correspondence points represent a common point in the 3D scene. To solve this issue, we propose a modified objective function with an epipolar geometry constraint to improve the displacement estimation. Let $\mathbf{w}_{ji}^{\mathbf{x}} = (w_1^{\mathbf{x}}, w_2^{\mathbf{x}}, 0)$ denote the displacement in the jth image of the pixel $\mathbf{x} = (x_1/x_3, x_2/x_3, 1)$ in the ith image. Then the set of displacements $\{\mathbf{w}_{ji}^{\mathbf{x}}\}$ for every coordinate in the jth image is obtained by minimizing the following objective function

$$\{\mathbf{w}_{ji}^{\mathbf{x}}\} = \arg\min_{\{\mathbf{w}^{\mathbf{x}}\}} \sum_{\mathbf{x}} \|s_i(\mathbf{x}) - s_j(\mathbf{x} + \mathbf{w}^{\mathbf{x}})\|_1 + \gamma \sum_{\mathbf{x}} \|\mathbf{w}^{\mathbf{x}}\|_2^2$$
$$+ \sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{E}} \min(\alpha|w_1^{\mathbf{x}} - w_1^{\mathbf{y}}|, d) + \min(\alpha|w_2^{\mathbf{x}} - w_2^{\mathbf{y}}|, d)$$
$$+ \sum_{\mathbf{x}} \frac{\beta((\mathbf{x} + \mathbf{w}^{\mathbf{x}})^T \mathbf{F}_{ji}\mathbf{x})^2}{\|\mathbf{F}_{ji}\mathbf{x}\|_2^2 + \|\mathbf{F}_{ji}^T(\mathbf{x} + \mathbf{w}^{\mathbf{x}})\|_2^2}, \quad (5)$$

where α, d, γ, and β are non-negative tuning parameters, and $\mathcal{E}$ denotes a collection of all 4-connected neighborhoods in the image. The first term in the right-hand side of equation (5) measures the difference in SIFT feature, the second term penalizes large movements, the third term promotes similar displacement of neighboring pixels, and the last term penalizes the Sampson distance [12], which is the first-order approximation to the geometric error. Note that the first three terms are identical to the original SIFT-flow [15], whereas the last term is our modification. Following the original SIFT-flow algorithm, we use the accelerated belief propagation algorithm [16] to minimize the new objective function in (5).

Furthermore, considering that the occlusion effect may lead to incorrect correspondence points, we verify correspondence points with two flows for each pair of images. One is from the ith image to the jth image, and the other is from the jth to the ith. A pair of correspondence points is verified if both points return to themselves according to the two flows, i.e., $\mathbf{w}_{ji}^{\mathbf{x}} = -\mathbf{w}_{ij}^{(\mathbf{x}+\mathbf{w}_{ji}^{\mathbf{x}})}$. Otherwise, it is treated as an occlusion point.

Given all correspondence points $\{\mathbf{x} \leftrightarrow \mathbf{x}'\}$ estimated by epipolar geometry regularized SIFT-flow (5), we warp all images except the target one, using the point transfer formula (3) to align them. More specifically, let $\mathbf{I}_i(\mathbf{x})$ and $\mathbf{I}_j(\mathbf{x}')$ denote the intensity of the ith image at $\mathbf{x}$ and that of the jth image at $\mathbf{x}'$, respectively. A point $\mathbf{x}''$ in the warped image satisfying $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \mathbf{x}''$ can be computed using (3) given the fundamental matrices $\mathbf{F}_{mi}$ and $\mathbf{F}_{mj}$. Therefore, we have warped images $\widehat{\mathbf{I}}_{ij}(\mathbf{x}'') = \mathbf{I}_i(\mathbf{x})$ and $\widehat{\mathbf{I}}_{ji}(\mathbf{x}'') = \mathbf{I}_j(\mathbf{x}')$, which are well aligned with the target image after warping, generated from a pair of multi-angular images $\mathbf{I}_i$ and $\mathbf{I}_j$, respectively.

2.3. Image Fusion using Matrix Completion

After warping all pairs of images chosen from the $N-1$ images (excluding the target image), we have obtained a total of $2 \times \binom{N-1}{2}$ well aligned images, but with missing pixels due to cloud contamination or occlusion. To improve the fusion performance, we only select a subset of the warped images that are potentially of high quality in terms of the correlation coefficient between a warped image and the dominant eigen image $\widehat{\mathbf{I}}_0$ of all warped images. A warped image is selected if its mutual correlation coefficient with $\widehat{\mathbf{I}}_0$ is higher than a preset threshold. A matrix $\mathbf{X}$ is then formed by concatenating vectors of the selected images $\{\widehat{\mathbf{I}}_{ij}\}$ as well as the target image $\mathbf{I}_m$.

Low-rank matrix completion estimates the missing entries of a matrix under the assumption that the matrix to be recovered has low rank. Since direct rank minimization is computationally intractable, convex [17] or nonconvex [18] relaxation is usually used to reformulate the problem. Here we consider the convex relaxation, i.e., a nuclear norm regularized least squares problem to perform matrix completion

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2^2 + \mu \|\mathbf{X}\|_*,$$

where $\mathcal{A}$ is a linear operator, $\mathbf{b}$ is a vector of observations, and $\mu > 0$ is a tuning parameter. In the image fusion problem, the linear operator $\mathcal{A}$ is a selection operator, which is defined as $\mathcal{A}(\mathbf{X}) = \text{vec}(\mathbf{X}_{\{\Omega_n\}})$, where $\{\Omega_n\}$ defines all visible area as indicated in (1). To solve the problem, we employ the accelerated proximal gradient (APG) algorithm [19] for efficient and fast solutions. The APG algorithm, which is an extension of Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [20] to the matrix setting, can be

specified as follows

$$\mathbf{Y}^k = \mathbf{X}^k + \frac{t^{k-1} - 1}{t^k}(\mathbf{X}^k - \mathbf{X}^{k-1}) \tag{6}$$

$$\mathbf{G}^k = \mathbf{Y}^k - (\tau)^{-1}\mathcal{A}^*(\mathcal{A}(\mathbf{Y}^k) - \mathbf{b}) \tag{7}$$

$$\mathbf{X}^{k+1} = \text{prox}_{\mu/\tau \|\cdot\|_*}(\mathbf{G}^k) \tag{8}$$

$$t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}, \tag{9}$$

where $k \geq 1$ is the iteration index, the initialization follows $t^0 = 1$ and $\mathbf{X}^0 = \mathbf{X}^1 = \mathbf{0}$, $\tau \geq L_\mathcal{A}$ (the Liptchitz constant of the operator $\mathcal{A}$), and $\text{prox}_{\mu/\tau \|\cdot\|_*}$ is a proximal operator defined as follows.

$$\text{prox}_{\lambda \|\cdot\|_*}(\mathbf{G}) = \mathbf{U}\mathbf{D}_\lambda \mathbf{V}^\mathbf{T},$$

where the singular value decomposition of $\mathbf{G}$ is $\mathbf{G} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ and the entries of $\mathbf{D}_\lambda$ are given by

$$\mathbf{D}_\lambda(p, q) = \begin{cases} \mathbf{D}(p, q) - \lambda, & \text{if } \mathbf{D}(p, q) > \lambda \\ 0, & \text{otherwise} \end{cases}.$$

The adjoint $\mathcal{A}^*$ puts the elements of the vector back into the matrix form and fills the missing entries in $\{\Omega - \Omega_n\}$, the complement of visible area $\{\Omega_n\}$, with zeros. Using the APG algorithm (6) – (9), the estimated cloud-free target image $\widehat{\mathbf{I}}_m$ is obtained by extracting the corresponding column from the reconstructed matrix $\mathbf{X}^{k+1}$ computed in (8) after the iteration index k exceeds some predefined maximum iteration number K, and converting the column vector back to the image form.

3. NUMERICAL RESULTS

Our dataset is a collection of $N = 13$ DigitalGlobe multi-angular high spatial resolution images of a common urban scene, where each is of size 14178×14178. The images are of the same 3D scene of interest according to the ground objects but captured from different angles so they are not well aligned to each other. Three regions with clouds, each of pixel size 1024×1024, are cropped from the original 13 images for testing. In our experiments, the intensity of the images are normalized to be in $[0, 1]$. We set parameters in (5) as $\alpha = 30$, $d = 300$, $\gamma = 0.01$, and $\beta = 10$, and those in (8) as $\mu = 20$, $\tau = 1$, and maximum iteration number $K = 100$. Since most existing algorithms assume that the images are well-aligned, they either fail or produce blurry images when directly applied to our dataset due to the inappropriate assumption. We compare our method with graph-cut based digital photomontage [4], robust PCA [7], and online tensor robust PCA algorithm [9] to demonstrate the improvement of our proposed method.

We first validate our fusion framework on synthetic cloud-contaminated images, in which case we have access to the ground truth. In particular, we select a cloud-free image (Fig. 1(a)) as the target image, which serves as the ground-truth from a set of multi-angle images with examples shown in Fig. 1(a)–(b). The synthetic cloud-contaminated image used to test our algorithm is generated by artificially replacing an area with a white cloud-like patch (Fig. 1(c)). The reconstruction result of online tensor robust PCA is shown in Fig. 1(d), which is blurry due to misalignment. We notice that in Fig. 1(d) the objects that are close to the ground are less blurry than those with higher altitude. This is because the objects close to the ground are roughly aligned, whereas there is a significant movement of the building tops, as can be seen by comparing Figs. 1(a) and 1(e).
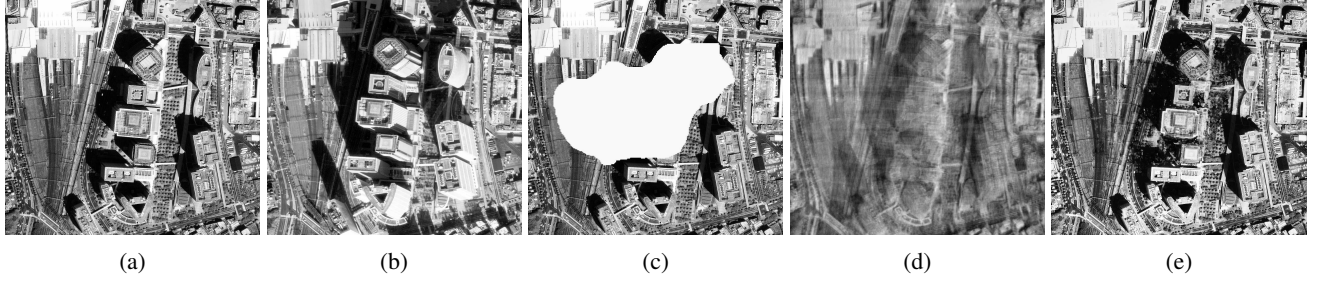
Fig. 1: Simulation results for synthetic cloud-contaminated images: (a), (b) example images captured from different view angles; (c) synthetic cloud-covered target image using (a) as groundtruth; (d) fused result by online tensor robust PCA; (e) fused image by our algorithm.
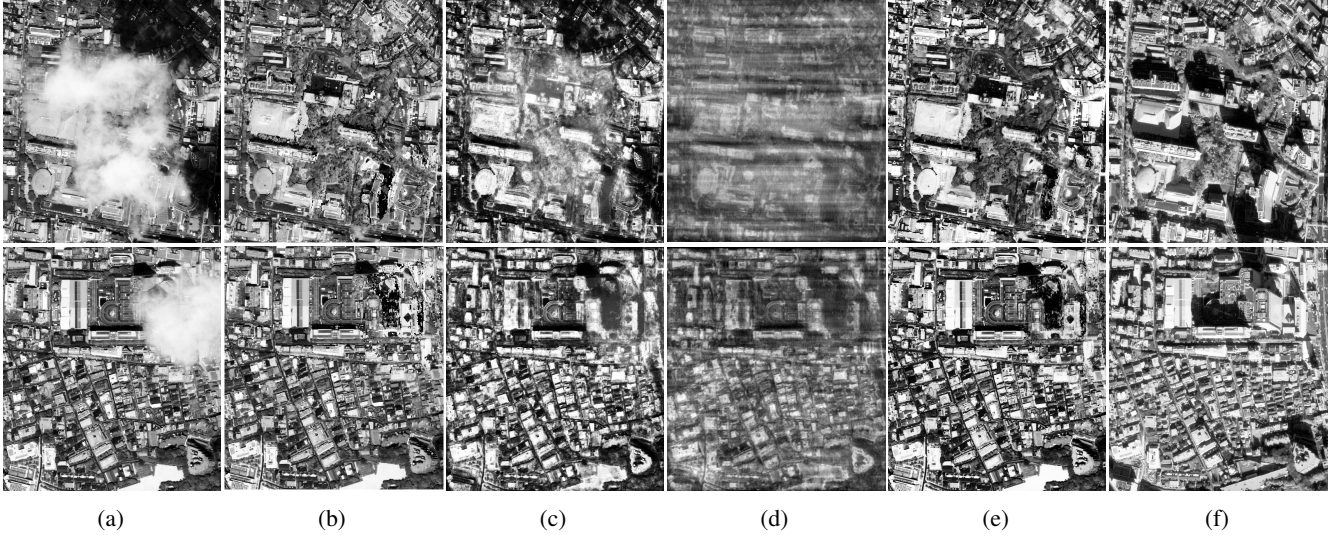

Fig. 2: Simulation results for real cloud contaminated images. Each row corresponds a set of testing images. Columns from the left to the right show examples of (a) the original cloud-contaminated images; (b)–(d) reconstructed results by graph cut photomontage, robust PCA, and online tensor robust PCA respectively; (e) reconstructed results by our algorithm; and (f) one of the images that are used for image fusion.

Our result shown in Fig. 1(e) has correctly located all the buildings at different elevations, comparing to the ground truth shown in Fig.1(a), and produced a visually plausible sharp image.

We also perform experiments on practical captured cloud-contaminated images and compare our framework with several other existing fusion methods. Considering that all the other methods require aligned images, we share our intermediate aligned images using epipolar geometry with all the methods for further fusion process. The results for two sets of images are shown in two rows of Fig. 2, respectively. Column (a) shows examples of the original cloud-contaminated images; columns (b)–(d) show the best results we can get using graph-cut based digital photomontage, robust PCA, and online tensor robust PCA methods respectively, and column (e) presents the results using our proposed algorithm. Since we do not have the ground-truth, instead we plot one of the images that are used for image fusion in column (f) of Fig. 2 as a reference for visual comparison. Note that the fusion results using existing methods either exhibit artifacts or blurs in the cloud-contaminated and shadow areas, while our framework convincingly outperforms these methods in overall visual quality. According to the reference image viewed from a different angle, our fused images preserve very good fidelity and resolution of the 3D scene. When we check the details of our

fused images, we observe that the building tops are well recovered, while the side surfaces of the buildings and the areas that are close to the high buildings are less satisfactory due to occlusions. This is however to be improved in our future work.

As regarding to the computational time, it takes about 15 seconds to detect clouds for all images, 28 minutes for each pair of images to finish image warping, and around 4 minutes for the final matrix completion-based image fusion to obtain a 1024×1024 cloud-free image. All experiments were performed on a desktop with a 3.6GHz Intel Xeon CPU and 64G RAM using Matlab R2010b.

4. CONCLUSION

We proposed an end-to-end framework for recovering the cloud-contaminated areas in a collection of multi-angular aerial images. Our framework first warps images viewed from different angles to the same view angle of the target image using epipolar geometry regularized SIFT-flow followed by point transfer, then fuses the missing pixels of contaminated areas or occluded areas using a low-rank matrix completion method. Numerical results on cloud contaminated images demonstrated that our algorithm significantly improves the performance in fusing multi-angular high spatial resolution images.

5. REFERENCES

[1] A. Maalouf, P. Carre, B. Augereau, and C. Fernandez-Maloigne, "A bandelet-based inpainting technique for clouds removal from remotely sensed images," *IEEE transactions on geoscience and remote sensing*, vol. 47, no. 7, pp. 2363–2371, 2009.

[2] L. Lorenzi, F. Melgani, and G. Mercier, "Inpainting strategies for reconstruction of missing data in vhr images," *IEEE Geoscience and remote sensing letters*, vol. 8, no. 5, pp. 914–918, 2011.

[3] C.-H. Lin, P.-H. Tsai, K.-H. Lai, and J.-Y. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE transactions on geoscience and remote sensing*, vol. 51, no. 1, pp. 232–241, 2013.

[4] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *ACM Transactions on Graphics (TOG)*, ACM, 2004.

[5] L. Lorenzi, F. Melgani, and G. Mercier, "Missing-area reconstruction in multispectral images under a compressive sensing perspective," *IEEE transactions on geoscience and remote sensing*, vol. 51, no. 7, pp. 3998–4008, 2013.

[6] J. Wang, P. A. Olsen, A. R. Conn, and A. C. Lozano, "Removing clouds and recovering ground observations in satellite image sequences via temporally contiguous robust matrix completion," *arXiv preprint arXiv:1604.03915*, 2016.

[7] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[8] A. Aravkin, S. Becker, V. Cevher, and P. Olsen, "A variational approach to stable principal component pursuit," *UAI Proceedings 2014*, 2014.

[9] Z. Zhang, D. Liu, S. Aeron, and A. Vetro, "An online tensor robust PCA algorithm for sequential 2D data," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2434–2438, 2016.

[10] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Computer vision, graphics, and image processing*, vol. 29, no. 1, pp. 100–132, 1985.

[11] H. Mansour and A. Vetro, "Video background subtraction using semi-supervised robust matrix completion," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6528–6532, 2014.

[12] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[15] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, pp. 978–994, May 2011.

[16] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International journal of computer vision*, vol. 70, no. 1, pp. 41–54, 2006.

[17] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[18] R. Chartrand, "Nonconvex splitting for regularized low-rank+ sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5810–5819, 2012.

[19] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 15, pp. 615–640, 2010.

[20] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.