

AN INTEGRATIVE OBJECT-BASED IMAGE ANALYSIS WORKFLOW FOR UAV IMAGES

Huai Yu ^{a, b}, Tianheng Yan ^a, Wen Yang ^{a, b, *}, Hong Zheng ^{a, b}

^a School of Electronic Information, Wuhan University, Wuhan, China

^b Wuhan University DJI Joint Laboratory of Unmanned Aerial Vehicles, Wuhan University, Wuhan, China
(yuhuai, yanth_eis, yangwen, zh)@whu.edu.cn

Commission VI, WG VI/4

KEY WORDS: UAV Images, Image Stitching, Binary Partition Tree, OBIA, Hierarchical Segmentation

ABSTRACT:

In this work, we propose an integrative framework to process UAV images. The overall process can be viewed as a pipeline consisting of the geometric and radiometric corrections, subsequent panoramic mosaicking and hierarchical image segmentation for later Object Based Image Analysis (OBIA). More precisely, we first introduce an efficient image stitching algorithm after the geometric calibration and radiometric correction, which employs a fast feature extraction and matching by combining the local difference binary descriptor and the local sensitive hashing. We then use a Binary Partition Tree (BPT) representation for the large mosaicked panoramic image, which starts by the definition of an initial partition obtained by an over-segmentation algorithm, i.e., the simple linear iterative clustering (SLIC). Finally, we build an object-based hierarchical structure by fully considering the spectral and spatial information of the super-pixels and their topological relationships. Moreover, an optimal segmentation is obtained by filtering the complex hierarchies into simpler ones according to some criterions, such as the uniform homogeneity and semantic consistency. Experimental results on processing the post-seismic UAV images of the 2013 Ya'an earthquake demonstrate the effectiveness and efficiency of our proposed method.

1. INTRODUCTION

1.1 Motivation and Objective

Nowadays, Unmanned Aerial Vehicles (UAVs)-based imaging systems have been applied in many remote sensing applications, such as agriculture and forestry, natural disasters and environmental issues. UAVs have several advantages over traditional remote sensing platforms, such as high flexibility and lower cost in collecting image data, higher speed and more safety. More importantly, UAVs are able to operate rather close to the object, which leads to images acquired with very high resolution (cm to dm pixel size). The high-resolution images provide sufficient details for identification and extraction of object parameters. However, A Unmanned Aerial System (UAS)-based image acquisition commonly results in hundreds of very high resolution, small footprint images, which pose great challenges for subsequent applications. The large overlap between neighbour images makes it inefficient to analyse large scale area which consists of hundreds of images. A simple solution to this problem is increasing the UAV flight altitude thus a single image can cover larger area and the total amount of image pieces can be decreased. However, it is inapplicable since civilian UAV can only fly at limited altitude. An alternative is the stitching of small images with limited view. The stitching output is a large scale panorama with very high resolution (VHR). Thus the segmentation and classification of large scale VHR panorama is the important tasks for processing UAV image. However, the interpretation of very large scale images remains a great challenge for the big data volume and semantic complexity.

The focus of our work is an entire per-processing chain for understanding UAV images, which includes several problems:

--The large number of image pieces with abundant overlap requests an efficient algorithm to integrate these images into a panorama. UAV images are characterized with more abundant information of the edge and texture, thus the conventional stitching methods are difficult to achieve practical requirements due to the low efficiency, insufficient memory and sparse point clouds in images. In addition, the accuracy of registration plays a crucial role in generating the panoramic view.

--The multi-scale VHR image representation. A VHR image is a unification of multi-scale objects, with large-scale objects at coarse level, e.g. forest, street, residential area, and small objects at fine level, e.g. animals, cars, humans. Thus a multi-scale image representation is essential to the detection of multi-scale objects in VHR images. However, how to consider the intrinsic properties and different features of local objects are still key problems in VHR image interpretation.

--The semantic information mining in VHR images. Based on the hierarchical image representation of VHR image, several images of the same scene at different resolutions (such as low, medium, high and very high spatial resolution) are available. However, the extraction of semantic regions or urban patterns presents some challenges related to the image size, extraction accuracy and the computational complexity. In addition, how to describe these semantic regions in the hierarchical structure is also a difficulty for the information mining task.

1.2 Related Works

Regarding the solutions to the above three important tasks for UAV image processing, researchers have done many

* Corresponding author: E-mail: yangwen@whu.edu.cn

outstanding works. For the image stitching task, one of the most important work is the automatic panoramic image stitching algorithm using invariant local features (Brown M, 2007). The invariant feature based approach enables reliable matching of image sequences despite rotation, scale and illumination change in the input images, which results in high-quality image panorama without any manual correction and registration work. This fully automatic computer vision based panoramic image stitching method is applicable to large scale mapping of UAV image (Yuhua X, 2016). However, the SIFT based feature extraction and feature matching are time-consuming for large data volume of image pieces. Later many interesting work on 3D reconstruction are presented, with the foundation point cloud construction still applies the aforementioned stable but time-consuming features and complex matching algorithm based on kd-tree search. Meanwhile, some fast feature extraction and matching algorithm (Xin Y, 2014) are introduced in computer vision field but not used in the image stitching or 3D construction.

For the panoramic VHR image representation, it is very important to consider object differences at different scales, thus hierarchical structure representation is a practice of general. To represent images on multiple scales, three main methods are utilized, i.e., image pyramid (Binaghi E, 2003), wavelet transform (Baraldi A, 2004) and hierarchical image partitions (Burnett C, 2003). The pyramid structures capture both the absolute and relative spatial arrangements of objects, but the fixed regular shape and choice of analysis window size lack semantic difference. Wavelet decomposition is a low-pass filter convolution of VHR images that represents the multi-scale property by the coefficients in different bands. However, wavelet decomposition is a decimation of the original image which lacks consideration of the relationship between objects. By fully considering the semantic gap of different objects, some studies use object-based analysis methods and mathematical morphology to produce hierarchical image partitions (Gui-Song X, 2010; Salembier, 2000). Among these methods Binary Partition Tree (BPT) is a good structure with both local intrinsic properties and topological relationship can be well considered. For the semantic object mining from the hierarchical structure of VHR image, several approaches have been proposed. A representative method is the dynamic programming (DP) for a set of energy function (Salembier, 2015). As a greedy algorithm, DP traverses the tree structure in a bottom-up way to find the global optimal segmentation. However, the direct minimizing method concerns little about the balances of the overall information of the selected nodes. To overcome the limitation of under-segmentation at coarse levels and over-segmentation at fine levels, the uniform entropy slice (Chenliang X, 2013) is proposed to flatten the hierarchy into a single segmentation and seek a selection of objects that balances the objects' energy function and the relative level.

1.3 Our Contributions

Inspired by the excellent work in the aforementioned three aspects, an integrative framework fully addressing the image stitching, VHR image representation and optimal image segmentation is studied in this paper. For fast mapping, an efficient image stitching algorithm after the geometric and radiometric corrections is utilized based on our earlier work of UAV image features extraction and matching, which combines the local difference binary (LDB) descriptor and the local sensitive hashing (LSH) matching (Huai Y, 2016). To process the mosaicked large scale VHR image, we use a Binary Partition Tree (BPT) to represent the image by an object-based

hierarchical structure. The structure derives from the definition of an initial partition obtained by an over-segmentation (i.e. a super-pixel partition) by means of the simple linear iterative clustering (SLIC) algorithm. During the BPT construction, we fully consider the spectral, spatial information of the superpixels and their topological relationships, i.e. selecting a subset of the most interesting segmentations from the segmentation hierarchy, which can be obtained by monitoring the value of the merging threshold. Moreover, an optimal segmentation is achieved by dynamic programming and the uniform homogeneity slice. The remainder of this paper is organized as follows. Section 2 presents the image stitching algorithm based on LDB and LSH. In section 3, the BPT construction and the optimal segmentation algorithm are introduced in detail. Some representative experimental results are exhibited in section 4. Finally, we draw the conclusion of this work in section 5.

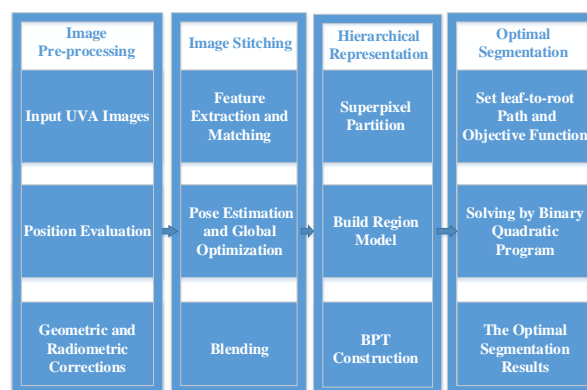


Figure 1. The integrative object-based image analysis framework for UAV images

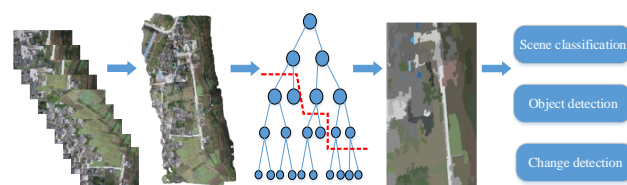


Figure 2. A brief schematic diagram for UAV image processing

2. UAV IMAGE STITCHING

2.1 UAV Image Pre-processing

With the rapid development of technique, UAV systems are equipped with high accurate inertial navigation system, which can be used to speed up the image registration. It means that we can restrict an image only matching the spatial neighbours. The position evaluation based on GPS and aerial information is used to determine the set of images adjacent to each image. By means of this method, the number of unnecessary matches is greatly decreased. The complexity of brute-force pairwise-match is $O(n^2)$, which can be reduced to $O(n)$ under the restriction of position evaluation.

In the low altitude UAV-based imaging system, image distortion correction is essential for promoting the quality of the final panoramic image. The imagery system of UAV is generally a central projection camera. To simplify the model, we just consider the radial distortion and tangential distortion, which can be described by Brown's distortion model (Brown D.C, 1971).

2.2 Feature Extraction and Matching

After the distortion correction, we are faced with the problem of image registration. Owing to the big volume and considerable overlap of UAV images, the conventional stitching methods, which use SIFT and kd-tree based nearest neighbour matching, are time-consuming and inapplicable for UAV image registration. Considering SURF (Bay H, 2006) is faster than SIFT while comparable in detection accuracy, we decide to utilize SURF to detect and locate feature points. However, float-type and high-dimension descriptors, e.g., SIFT and SURF descriptor, make the matching a time-consuming task. Thus, the local difference binary (LDB) algorithm (Xin Y, 2014) is used to describe the features for its computational efficiency. Relying on integral images, LDB computes a binary string using simple intensity and gradient difference tests on pairwise grid cells within the patch of detected point, which yield an ultrafast runtime. With LDB being a binary descriptor, local sensitive hashing (LSH) is chosen to replace the kd-tree for ANN search in the feature matching procedure (Andoni A, 2004). Because the similarity of binary descriptors can be easily measured by logical AND and OR, the hash function of LSH firstly maps similar descriptors into different buckets, then the NN search of a query descriptor can be easily obtained in the bucket it belongs to. After the NN search, the nearest neighbour (denote by fea_{n_1}) and second-nearest neighbour (denote by fea_{n_2}) of a query descriptor (denote by fea_o) can be obtained. A robust match can be implemented according to the formula (1):

$$\frac{\|fea_{n_1} - fea_o\|}{\|fea_{n_2} - fea_o\|} \leq \alpha, \quad (1)$$

where $\|\bullet\|$ is the mahalanobis distance, and α is the threshold value, which varies 0.5~0.9 in experiment. When the neighbours of a query descriptor satisfy eq. (1), the query descriptor and its nearest neighbour is accepted as a robust matching pair. For two overlapped image, we can get a set of pairwise matches. Subsequently, according to the affine transformation principle of 8 parameters, the RANdom SAmple Consensus (RANSAC) method (Fischler M, 1981) is utilized to get more robust matches.

2.3 Pose Estimation and Global Optimization

After the feature matching, all the overlapped images can be connected together according to the feature point matches. In practice, we select the image located at the centre of the stitching area as the reference plane of the mosaicking result based on GPS information. To initialize the pose of every image, we use the 8-DOF affine transformation model to initialize the camera parameters (focal, principal point, rotation and translation). Then the global objective function for pose estimation is

$$E_{all-pairs} = \sum_i \sum_{jk} \|\tilde{x}_i(\hat{x}_{ij}; R_j, f_j) - \tilde{x}_i(\hat{x}_{ik}; R_k, f_k)\|^2, \quad (2)$$

where $\hat{x}_{ij}, \hat{x}_{ik}$ are the image point of the projection of a 3D point \tilde{x}_i , R and f are the rotation matrix and focal length respectively. It can be observed that the global optimization is a non-linear least squares problem which can be solved by using the Levenberg-Marquardt algorithm (Madsen K, 2004)

2.4 Blending

The last step of UAV image stitching is blending all image pieces to a panorama according to the estimated geometrical

transformation parameters. For two overlapped images, we first find the optimal seam line with the minimum cost. The cost consists of colour difference and geometrical difference, which are defined as:

$$E(x, y) = E_c(x, y)^2 + E_g(x, y), \quad (3)$$

where $E_c(x, y)$ is the colour difference of three channels, $E_g(x, y)$ is the geometrical difference which is computed by the convolution of sobel template. Subsequently, multi-band blending method (Brown M, 2007) is used to process the mosaicking artifacts generated by alignment errors in geometry and intensity differences.

3. HIERARCHICAL IMAGE REPRESENTATION

The big data volume and very high-resolution properties make the interpretation of UAV panorama a great challenge. The object-based image analysis (OBIA) methods not only preserve the useful information (including colour, contour and structure) in UAV images, but also decrease the data volume to be analysed. In addition, a hierarchical representation based on OBIA is an essential step for VHR image interpretation. This section mainly introduces the hierarchical representation of UAV panorama image.

3.1 Superpixels Partition

There are many superpixel partition algorithms in the literature, while each of them has its own strengths and weaknesses. Taking the computation speed and partition performance into consideration, the Simple Linear Iteration Clustering (SLIC) segmentation algorithm (Achanta R, 2012) is utilized to obtain the initial superpixels. This method can produce consistent superpixels with similar size and shape, as well as preserve objects' boundaries.

In this step, the VHR panorama image with complex boundaries is segmented into many superpixels. Each of superpixels is relatively homogeneous and the boundaries can be well preserved. Thus it is unnecessary to consider the information details of the internal superpixels. Furthermore, the superpixel-based description can speed up the later processing procedure and preserve the useful information. According to the principle of SLIC, the region size and regularity of superpixels can be set experientially.

3.2 Region Model and Similarity Criterion

After image partition, the description of superpixels is an important task, which directly relates to the measurement of similarity between superpixels. Because the major difference lies in the colour, we leave out size and shape information. The region model is characterised by colour names (Van De Weijer, 2009), which are linguistic colour labels based on the assignment of colours in the real world. The colour labels including 11 basic terms: black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow, are learned from Google images. The learning result is a partition of the colour space into eleven regions. To use this colour feature, the RGB area of superpixel is mapped to the colour attribute space. The colour names of superpixel are defined as follows:

$$CN_R = \{p_R(cn_1), p_R(cn_2), \dots, p_R(cn_{11})\}, \quad (4)$$

where

$$p_R(cn_i) = \frac{1}{N} \sum_{x \in R} p(cn_i | f(x)), \quad (5)$$

the variable $cn_i (i=1, \dots, 11)$ is the i -th colour name, N denotes the number of pixels in region R , $p(cn_i | f(x))$ denotes the probability of a colour name of pixel x . Colour names are more photometric invariant than other colour features because different shades of colour are mapped to the same colour names. The two important concepts in OBIA are region model and similarity measurement. The region model is characterised by colour names. For simplification, the union of some superpixels is modelled by the average of colour names. While the similarity of region R_1 and R_2 is measured by the weighted Euclidean distance:

$$d(R_1, R_2) = N_{R_1} \|M_{R_1} - M_{R_1 \cup R_2}\| + N_{R_2} \|M_{R_2} - M_{R_1 \cup R_2}\|, \quad (6)$$

where M_{R_1}, M_{R_2} denote the region models for R_1 and R_2 , respectively. N_{R_1}, N_{R_2} denote the data volumes of R_1 and R_2 , respectively.

3.3 BPT Construction

Based on superpixels segmentation, the bottom level of hierarchical structure is composed by the original superpixels. To build the hierarchical representation, binary partition tree structure based on superpixels is constructed. Every node and every level of the hierarchical structure BPT contain semantic information. The leaves represent the original superpixels and the root represents the entire image. we can reconstruct the tree on the condition of the parent, siblings and sons of every node are available. The building process is as follows:

- (1) build a priority queue to store all the superpixel-pairs (superpixel and its neighbour) in an ascending sequence according to similarity criterion (Eq. 6);
- (2) pop out the top pairs of the queue to merge into a new region until the queue is empty;
- (3) update the adjacencies of the merged region then enter the queue.
- (4) output the tree structure, as figure 2.

An example of BPT hierarchical representation of UAV image is shown figure 3

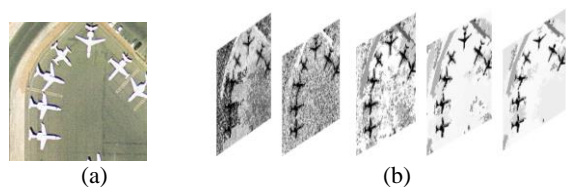


Figure 3 representation of several levels on BPT.
(a) An UAV image (b) several levels representation of the image.

4. OPTIMAL SEGMENTATION

The hierarchical structure, i.e., binary partition tree, represents the UAV panorama in multiple spatial scales. Based on this structure, hierarchical segmentation algorithms can analyse images at different scales simultaneously, whose output is a set of regions that captures different partitions of different scales. The optimal segmentation based on hierarchical structure can overcome the limitation of over-segmentation at fine levels and under-segmentation at coarser levels, which means the partition of multiple scale meaningful regions can be exactly achieved. Under this conception, we design an optimal segmentation methods based on the hierarchical structure.

4.1 Setting leaf-to-root Path and Objective Function

Considering a given UAV panorama I . Based on BPT construction, we denote the max hierarchical level by m , the node set of each level T^i is denoted by V^i , the entire tree is denoted by $T = \{T^1, T^2, \dots, T^m\}$, and the individual node s at level i is denoted by V_s^i . The only node V_0^1 at T^1 is the root of T . Because not all original superpixels are at the bottom level, we copy the nodes at upper level to bottom level. Thus each level of BPT corresponds to a partition of the image. During the process of BPT construction, once two nodes are merged, their parent node is at a new level.

A segmentation is a non-overlapped division of an image with the union restore the image in its entirety. Thus a partition in the hierarchy is a set of nodes satisfying the principle that there is one and only one node selected at each leaf-to-root path in the hierarchy. For example, Fig.4 shows valid tree slices of a particular BPT.

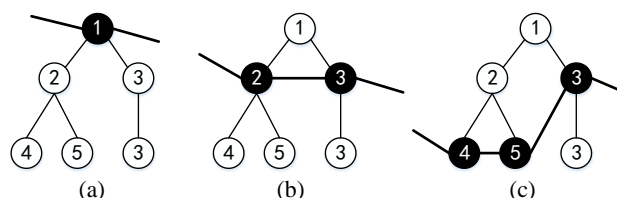


Figure 4. All valid tree slices of a particular BPT. Each slice is highlighted a black curve, the nodes on the slice are darkened.

In the following, we formulate the above constraint. Let P denotes a $p \times n$ binary matrix, where p is the number of leaf nodes in T , n is the total number of nodes in T . Each row of P denotes a leaf-to-root path. If a node is in the path, the value of corresponding location at P takes 1, otherwise takes 0. The corresponding path matrix of Fig.4 is show in Fig.5 (b). There are three rows in P , which represents the three leaf-to-root paths in the BPT. For instance, node sequence (V_1, V_2, V_5) is the path P_2 . Therefore the value of the second row of P is $[1, 1, 0, 0, 1]$.

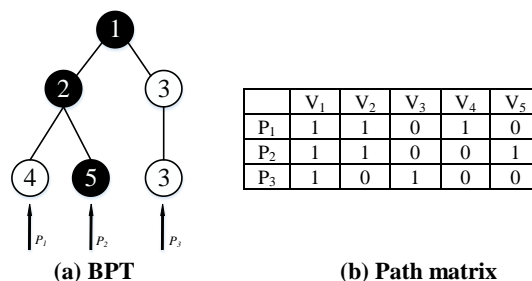


Figure 5. BPT and the corresponding path matrix

Because a valid tree slice x consists of one and only one node in each path, the valid tree slice satisfies the following formula:

$$Px = l_p, \quad (7)$$

where l_p is a $p \times 1$ column vector. x is a $1 \times n$ vector, if a node is selected as a partition, the value of corresponding location is set as 1, otherwise set 0. Thus any x satisfies the Eq.7 provides a possible partition of BPT, which corresponds to a plausible segmentation of the UAV panorama.

According to the constraint above, there are still many feasible tree slices that are proper segmentation of the image. However, our purpose is to find the optimal partition that is most meaningful. In this paper we propose a meaningful criterion named minimal heterogeneity, which defined as follows,

$$\phi_R = \frac{\sum_{i \in R} \|CN_i - CN_R\|}{\|CN_R\|}, \quad (8)$$

where R is a node in the BPT which consists of several adjacent superpixels. CN_R is the region model, CN_i is the model of i -th superpixel. By using this criterion, we can obtain the entire heterogeneity of region R , which measures the meaning of this region.

Thus the segmentation objective is to seek a slice that balances the overall heterogeneity of selected nodes,

$$\mathbf{x}^* = \arg \min \sum_{V_s \in T} \phi(V_s), \quad (9)$$

which subjects to Eq.7

4.2 Solving by Binary Quadratic Program

However, directly solving Eq.9 is difficult, which requires to enumerate all tree slices and the answer is a degenerate minimum which selects all leaf nodes because their heterogeneity are all zero. We add a penalty term which tends to select nodes at coarse level. Although nodes in the coarser levels have relatively higher heterogeneity than nodes in the finer level, the number of coarser level nodes is much less than those at the finer levels. Thus the objective function is formulated as follows:

$$\begin{aligned} & \text{minimize } \sum_s \alpha_s x_s + \sigma \sum_{s,t} \beta_{s,t} x_s x_t \\ & \text{subject to } \mathbf{P}\mathbf{x} = \mathbf{I}_p \end{aligned} \quad (10)$$

$$\mathbf{x} = \{0,1\}^n$$

where α_s denotes to a n -length vector, which makes the slice prefer coarser levels in BPT. α_s is defined as follows:

$$\alpha_s = |V^i|, \text{ if } V_s \in V^i, \quad (11)$$

where $|V^i|$ is the total number of nodes in i -th level of the tree.

$\beta_{s,t}$ is an entry in an $n \times n$ matrix, which is specified based on neighbourhood structure, and defined as follows:

$$\beta_{s,t} = |\phi(V_s) - \phi(V_t)| |V_s| |V_t|, \quad (12)$$

where $|V_s|$ and $|V_t|$ denote the number of superpixels in V_s and V_t . By adding the volume factors, we can evaluate the total difference of two nodes.

The objective is a typical binary quadratic program problem. We solve the QP by using an open source solver (OPTI toolbox).

5. EXPERIMENTAL RESULTS

In this section, we provide experimental results on processing the post-seismic UAV images of the 2013 Ya'an earthquake to demonstrate the effectiveness of our algorithm. The particular location is Yuxi village, Baosheng town, Lushan county, and the geographic coordinate is around 30.29 N, 103.04 E. The sensor onboard is Canon 5D Mark II and 93 image pieces were acquired in total. Fig.6 shows 4 typical example images.



Figure 6. Typical experiment datasets

The experimental platform consists of Intel Core i7-4790 CPU and 32G RAM. We use Microsoft Visual Studio 2010, Opencv 2.4.10 and MATLAB software.

5.1 Experiment settings

For the stitching algorithm, the proposed method is compared with the conventional stitching method using SURF descriptor and kd-tree based ANN matching. The performance metrics include the stitching time and root mean square error (RMSE) of all matching points.

$$\text{RMSE} = \sqrt{\frac{1}{2N} \sum_{i=1}^N (\|\tilde{H}_i \tilde{x}_i - \hat{x}\| + \|\hat{H}_i \hat{x}_i - \tilde{x}_i\|)}, \quad (13)$$

Where N is the total number of matching points, \tilde{H}_i, \hat{H}_i is the projection matrix of \tilde{x}_i, \hat{x}_i , respectively.

For the purpose of optimal segmentation, the uniform homogeneity slice is compared with dynamic programming (Salembier, 2016), which is a greedy algorithm starting from the initial superpixels to extract the optimal partition by minimizing the criterion $C = \sum_R \phi_R$.

$$\phi_R = \frac{\sum_{i \in R} \|CN_i - CN_R\|}{\|CN_R\|} + \lambda, \quad (14)$$

where λ is a constant value as data regularization term which encourages the optimization to find partitions with a reduced number of regions. In the experiment we set $\lambda = 2$ according to the author's recommendation. Considering that there is no ground truth of these datasets, the performance comparison is in the form of qualitative analysis.

5.2 Experiment comparisons

5.2.1 Stitching methods analysis

The details about stitching methods comparisons are shown in table 1. Where t_F denotes the feature extraction time, t_M denotes the feature matching time, t_T denotes the total time of stitching, N denotes the number of matching points and RMSE is the root mean square error.

Metric	Our method	Conventional method
t_F (s)	14.69	16.97
t_M (s)	28.34	153.96
t_T (s)	148.53	282.36
N	73163	27531
RMSE	1.8	3.01

It can be observed that our method works slightly faster in feature extraction procedure, while extremely faster in the matching step. Under the same condition, there are much more correct matching points in our method, resulting in a smaller image stitching error. These observations demonstrate that the LDB descriptor is more efficient than the conventional descriptors in UAV image stitching. Furthermore, the LSH search based matching algorithm can effectively improve the efficiency of UAV image stitching by using LDB descriptor. The final stitching panorama is shown in figure 7.

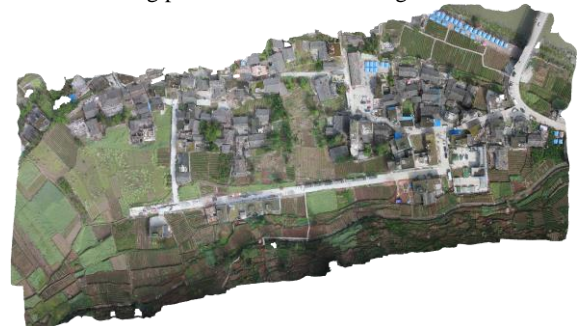


Figure 7. The stitching panorama of the post-seismic UAV images of the 2013 Ya'an earthquake

5.2.2 Optimal segmentation results

As a stitching of 93 image pieces, the data volume of the panorama is very large. For convenience, we choose the inscribed rectangle to conduct the segmentation experiment. Figure 8 presents the experimental dataset.



Figure 8. The experimental dataset (3845x1038)

It should be noticed that the scene is very complicated and consists of several different semantic areas. For example, buildings around roads show very similar spectral information, shadows of building have strong impacts on the detection of the obscured regions. Because there is no ground truth about this area, here we only give a qualitative comparison as follows.



Figure 9. The optimal segmentation using dynamic programming



Figure 10. The corresponding partition of Fig.9



Figure 11. The optimal segmentation using uniform homogeneity slice



Figure 12. The corresponding partition of Fig.11

In general, the two methods reach almost the same optimal segmentation, which demonstrate that these two methods are all applicable in the optimal segmentation on hierarchical image representation. However, dynamic programming uses a constant value λ as regularization, which encourages to find partitions with a reduced number of regions. This value actually set a hard threshold for the semantic gap of a node between its two descendants. However, the semantic gap is rising from fine level to coarse level, which means the value actually selects levels that satisfy the constant value. Being different with this criterion, the uniform homogeneity slice uses a penalty term which tends to select nodes at coarse level which are relevant to the total number of nodes in i -th level. What we need to do is to change the weight of penalty term. In this perspective, the uniform homogeneity slice is more generalizable.

For UAV image interpretation, this optimal segmentation scheme of UAV panorama image can be used in subsequent tasks, such as image classification, object detection and recognition, change detection. Figure 13 shows the segmented area of grassy land and buildings. The building area in Fig.13 is finely distinguished, which can be used for for extracting residential areas.

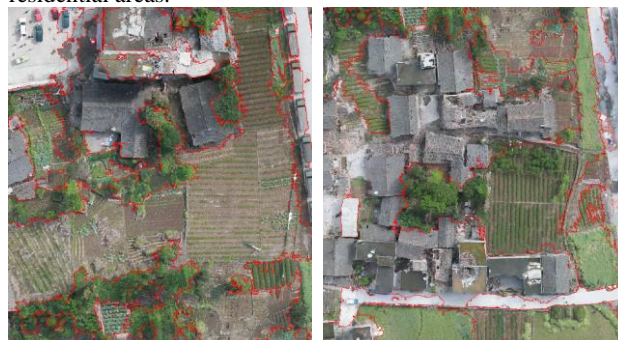


Figure 13. The distinguished area of grassy land and buildings in the optimal segmentation results

6. CONCLUSION

In this paper, an integrative object-based image processing workflow has been proposed to understanding UAV images. This framework fully addresses three important tasks of UAV image interpretation, i.e., UAV image stitching, large scale image hierarchical representation and optimal segmentation, and provides substantial possibility for large scale VHR image classification, object detection, change detection and etc. The experimental results demonstrate that BPT representation based on superpixels is a very effective hierarchical structure for processing and analysing large scale UAV images. Furthermore, the uniform homogeneity slice and dynamic programming are all applicable for the semantic object segmentation of large scale UAV images. The future work will focus on target detection and scene classification.

ACKNOWLEDGEMENTS

This work was supported in part by the National Key Basic Research and Development Program of China under Contract 2013CB733404 and the grant from Shenzhen basic scientific research program of China (No. JCYJ20150422150029095)

REFERENCES

- Brown M, Lowe D G, 2007. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1), pp. 59-73.
- Xu Y, Ou J, He H, et al, 2016. Mosaicking of Unmanned Aerial Vehicle Imagery in the Absence of Camera Poses. *Remote Sensing*, 8(3), pp. 204.
- Turner, Darren, Lucieer, et al, 2012. An Automated Technique for Generating Georectified Mosaics from Ultra-High Resolution Unmanned Aerial Vehicle (UAV) Imagery, Based on Structure from Motion (SfM) Point Clouds. *Remote Sensing*, 4(5), pp. 1392-1410.
- Binaghi E, Gallo I, Pepe M, 2004. Cognitive pyramid for contextual classification of remote sensing images. *IEEE Transactions on Geoscience & Remote Sensing*, 41(12), pp. 2906-2922.

- Baraldi A, Bruzzone L, 2004. Classification of high spatial resolution images by means of a Gabor wavelet decomposition and a support vector machine. In Proceedings of the International Society for Optics and Photonics, Remote Sensing, Maspalomas, Canary Islands, Spain, pp. 19–29.
- Burnett C, Blaschke T, 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling*, 168(3), pp. 233-249.
- Xu C, Whitt S, Corso J, 2013. Flattening supervoxel hierarchies by the uniform entropy slice. *Proceedings of the IEEE International Conference on Computer Vision*, Sydney, Australia, pp. 2240-2247.
- Yu Huai, Yang Wen, 2016. A Fast Feature Extraction and Matching Algorithm for Unmanned Aerial Vehicle Images. *Journal of Electronics and Information Technology*, 38(3), pp. 509-516.
- Xia G. S., Delon J., & Gousseau Y, 2010. Shape-based invariant texture indexing. *International Journal of Computer Vision*, 88(3), 382-403.
- Yu H., Yang W., Xia G. S., & Liu G, 2016. A Color-Texture-Structure Descriptor for High-Resolution Satellite Image Classification. *Remote Sensing*, 8(3), 259.
- Salembier P, Garrido L, 2000. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing*, 9(4): 561-576.
- Salembier P, 2015. Study of binary partition tree pruning techniques for polarimetric SAR images. *Mathematical Morphology and Its Applications to Signal and Image Processing*, pp. 51-62.
- D.C. Brown, 1971. Close-Range Camera Calibration. *Photogrammetric Engineering*, 37(8), pp. 855-866.
- Yang X, Cheng K T, 2014. Local difference binary for ultrafast and distinctive feature description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1), pp. 188-194.
- Bay H, Tuytelaars T, Van Gool L, 2006. Surf: Speeded up robust features. *Proceedings of European Conference on Computer Vision*, Graz, Austria, pp. 404-417.
- Andoni A, Indyk P, 2008. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of the ACM*, 51(1), pp. 117-122
- Fischler M A, Bolles R C, 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp. 381-395.
- Madsen K, Nielsen H B, Tingleff O, 2004. Methods for non-linear least squares problems, 2nd edition, *informatics and mathematical modelling*, Technical University of Denmark, pp. 24-29
- Achanta R, Shaji A, Smith K, et al, 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), pp. 2274-2282.
- Van De Weijer J, Schmid C, et al, 2009. Learning colour names for real-world applications. *IEEE Transactions on Image Processing*, 18(7): 1512-1523.