DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

Vector Road Map Registration to Oblique Wide Area Motion Imagery by Exploiting Vehicle Movements

Ahmed Elliethy, Gaurav Sharma
University of Rochester, Rochester, NY 14627, USA

Abstract

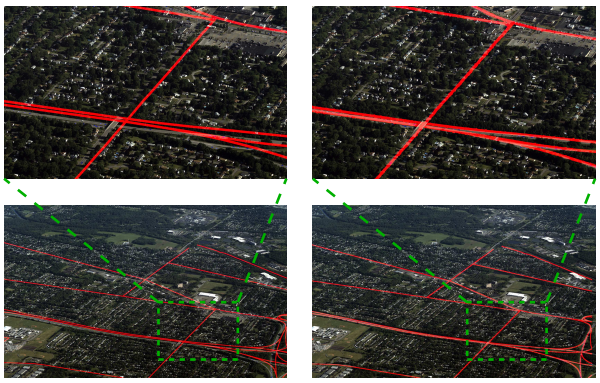
We present a novel methodology for accurately registering a vector road map to wide area motion imagery (WAMI) gathered from an oblique perspective by exploiting the local motion associated with vehicular movements. Specifically, we identify and compensate for global motion from frame-to-frame in the WAMI which then allows ready detection of local motion that corresponds strongly with the locations of moving vehicles along the roads. Minimization of the chamfer distance between these identified locations and the network of road lines identified in the vector road map provides an accurate alignment between the vector road map and the WAMI image frame under consideration. The methodology provides a significant improvement over the approximate geo-tagging provided by on-board sensors and effectively side-steps the challenge of matching features between the completely different data modalities and viewpoints for the vector road map and the captured WAMI frames. Results over a test WAMI dataset indicate the effectiveness of the proposed methodology: both visual comparison and numerical metrics for the alignment accuracy are significantly better for the proposed method as compared with existing alternatives.

Introduction

Recent technological advances have made available number of airborne platforms for capturing imagery [1,2]. One of the specific areas of emerging interest for applications is Wide Area Motion Imagery (WAMI) where images at temporal rates of 1–2 frames per-second can be captured for relatively large areas that span substantial parts of a city while maintaining adequate spatial detail to resolve individual vehicles [3]. WAMI platforms are becoming increasingly prevalent and the imagery they generate are also feeding a corresponding thrust in large scale visual data analytics. The effectiveness of such analytics can be enhanced by combining the WAMI with alternative sources of rich geo-spatial information such as road maps. In this paper we focus on near real-time registration of vector road-map data to WAMI and propose a novel methodology that exploits vehicular motion for accurate and computationally efficient alignment.

Registering road map vector data with aerial imagery leads to rich source of geo-spatial information, which can be used for many applications. One application of interest is moving vehicle detection and tracking in wide area motion imagery (WAMI). By registering the road network to aerial imagery, we can easily filter out the false detections that occurred off roads. Another interesting application is to detect and track a suspicious vehicle that goes off road. These applications depend on accurate road network alignment with the aerial imagery, which is the focus of this paper.

In general, successive WAMI video frames are related by both global and local motions. The global motion arises from the camera movement due to the aerial platform movement, and it can be parameterized as a homography between the spatial coordinates for successive frames under the assumption that the captured scene is planar. The local motion arises due to the local movement of objects in the scene. Local motion in WAMI for urban scenes is dominated by vehicle movements on the network of roads within the captured area. We exploit these vehicular movements to develop an effective registration scheme for aligning vector road maps data with the captured WAMI frame.



(a) (b)

Figure 1: Road network alignment. (a) using only aerial frame meta-data, (b) using our proposed algorithm.

WAMI frames are usually captured from platform equipped with Global Positioning System (GPS) and Inertial Navigation System (INS) which provide location and orientation information that are usually stored with the aerial image as meta-data. This meta-data can be used to align a road network extracted from external Geographic Information System (GIS) source. However, as illustrated by the example in Fig. 1(a), the accuracy of the meta-data is limited and only provides an approximate alignment.

Registering an aerial image directly with a geo-referenced vector road map is a challenging task because of the differences in the nature of the data in the two formats: in one case the data consists of image pixel values whereas in the other it is described as lines/curves connecting a series of points. Because of the inherent differences in the data formats, one cannot readily define low/mid-level features that are invariant to the representations and can be used for registration as conventional feature detectors, such as SIFT (Scale-Invariant Feature Transform) [4], are used for finding corresponding points in images. For static imagery, a lot of research has been done for aligning vector road maps to aerial imagery, normally referred to as the process of conflating. In general, conflation refers to a process that fuses spatial representation from multiple data sources to obtain a new superior representation. In [5–7], road vector data are aligned with an aerial image by matching the road intersection points in both representations. The crucial element in these prior works is the detection of road intersections from the aerial image. With the availability of hyper-spectral aerial imagery, spectral properties and contextual analysis are used in [5] to detect these road intersections in the aerial scene. However, road segmentation is not robust for different natural scenes specially when roads are obscured by shadows from trees and nearby buildings. In [6], a Bayes classifier used to classify pixels as on-road or off-road, then a localized template matching used to detect the road intersections. However, to get a reasonable accuracy with the Bayes classifier, a large number of manually labeled training pixels is required for each data set. In [7], corner detection is used to detect the road intersections, which is not reliable specially in high resolution aerial images, that contain enough wide roads where the simple corner detection fails.

Work on registration of (non-static) WAMI frames to geo-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

referenced vector road maps has received comparatively less attention, even though the capability for performing such registration in a computationally efficient manner is crucial for a number real/near real-time analysis applications for WAMI, as already mentioned. Some of the prior work on this problem overcomes the problem posed by fundamentally different modalities of the WAMI and vector datasets by using an auxiliary geo-referenced image that is already aligned with the vector road map. The aerial image frames are then aligned to the auxiliary geo-referenced image by using conventional image feature matching methods. For example, in [8], for the purpose of vehicular tracking, the aerial frame is geo-registered with a geo-reference image and then a GIS database is used for road network extraction. This road network is used to regularize the matching of the current vehicle detections to the previous existing vehicular tracks. In an alternative approach that relies on 3D geometry, in [9], SIFT is used to detect correspondences between the ground features from a small footprint aerial video frame and geo-referenced image. This geo-registration helps to estimate the camera pose and depth map for each frame, and this depth map is used to segment the scene into building, foliage, and roads using a multi-cue segmentation framework. The process is computationally intensive and the use of the auxiliary geo-referenced image is still plagued by problems with identification of corresponding feature points because of the illumination changes, different capturing times, severe view point change in aerial imagery, and occlusion. State of the art feature point detectors and descriptors such as SIFT (Scale-Invariant Feature Transform) [4], and SURF (Speeded Up Robust Features) [10], often find many spurious matches that cause robust estimators such as RANSAC [11] to fail when estimating a homography. Also, these methods cannot work directly if the aerial video frames have a different modality (infra-red for example) than the geo-referenced image. Last, but not least, a single homography represents the relation between two images when the scene is close to planar [12]. In WAMI, aerial video frames usually taken from oblique camera array to cover large ground area from moderate height and the scene usually contains non ground objects such as building, trees, and foliage. Thus the planar assumption does not necessarily hold across the entire imagery, although it is not unreasonable for the road network.

In this paper, we propose an algorithm that accurately aligns a vector road network to WAMI aerial video frames by detecting the locations of moving vehicles and aligning the detected vehicle locations with the network of roads in the vector road map. The vehicle locations are readily detected by performing frame-to-frame registration using conventional image feature matching methods and computing compensated frame differences to identify local motion that differs significantly from the overall global motion resulting from the camera movement. Such local motion is predominantly due to moving vehicles and the regions where the compensated frame differences are large correspond (predominantly) to vehicle locations. We align the WAMI frames to the vector road map by estimating the projective transformation parameters that, after appropriate application of the transformation, minimize a metric defined as the sum of minimum squared distances from the detected vehicle locations to the corresponding nearest points on the network of roads. This metric is the well known chamfer distance, which can be efficiently computed via the distance transform [13]. The chamfer distance serves as an ideal quantitative metric for the degree of misalignment because it does not require any feature correspondences or computation of displaced frame differences, both of which are inappropriate for our problem setting because of the different modalities of the data. By exploiting vehicle detections and using the vector road network, we implicitly transfer both the aerial image and the geo-referenced one to a representation that can be easily matched. In other words, unlike traditional methods, our algorithm does not directly estimate any feature correspondence between the WAMI image frames and the vector road maps. Instead, it aligns two binary images representing the vehicle detections and the network of road lines identified in the vector map, thereby providing a more accurate and robust alignment. A sample result from our algorithm is shown in Fig. 1(b), where it can be appreciated that the method provides an accurate alignment to the road network. Our main assumption here is that the investigated scene should contain a forked road network which is reasonable assumption for WAMI, which covers a city scale ground area within each frame. Our algorithm does not depend on the aerial camera sensor type; for example, it can be used directly with infra-red aerial camera.

This paper is organized as follows. The next section explains our proposed algorithm. Results and a comparison against alternative methods are presented in the following section. The final section summarizes concluding remarks.

Proposed algorithm for vehicle motion based WAMI alignment

A high level overview of the proposed algorithm is shown in block-diagram format in Fig. 2 using illustrative example images. Our algorithm consists of three major parts. First, we do frame to frame registration to align temporally adjacent WAMI frame, denoted by I_t, and I_{t+1}, into the common reference frame for I_t and compute the displaced frame-to-frame difference [14] between them. The regions of significant magnitude in these frame-to-frame differences correspond predominantly to the locations of moving vehicles. Then we use the meta-data associated with I_t along with the vector road network to generate a road network coarsely aligned with I_t. Finally, we estimate the final alignment between the aligned road network and the vehicle detections by minimizing the chamfer distance [13] between them which corresponds to minimizing the sum of the squared distances between each vehicle detection and corresponding nearest point on the road network. The chamfer distance measures how close the vehicle detections are to the road network and therefore applies nicely to our problem.

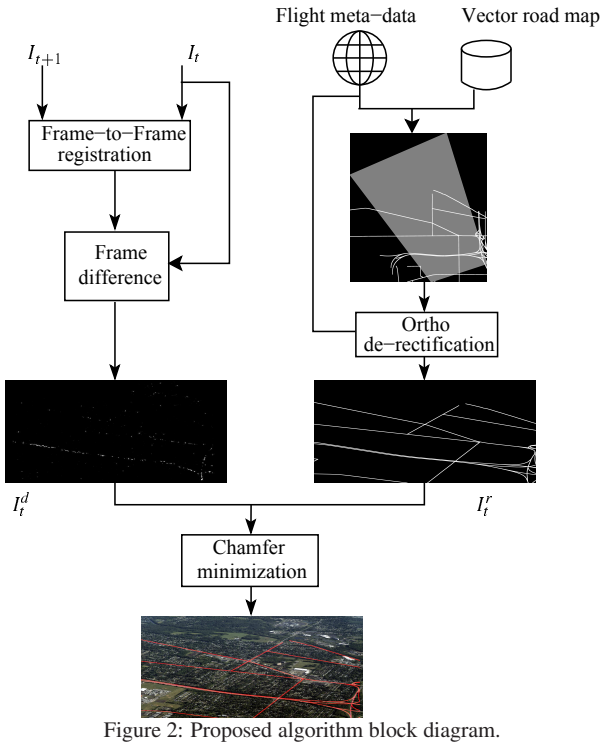


Figure 2: Proposed algorithm block diagram.

We use a projective transformation [12] for our alignment, where the 2D point $\mathbf{p}_1 = (x, y)$ in the input image is mapped to the 2D point $\mathbf{p}_2 = (u, v)$ in the target image, by the transformations

$$u = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + 1}, \quad v = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + 1},$$

where the transformation is specified by the parameters $\beta = [h_1, \ldots, h_8]^T$

DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

and can be equivalent represented as the matrix multiplication

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{H}_\beta \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad (1)$$

where $[x,y,1]^T$ and $[u,v,1]^T$ are the homogeneous coordinate representation of $\mathbf{p}_1$ and $\mathbf{p}_2$ respectively. The projective transformation has 8 degree of freedoms and the only invariant property of this transform is the cross ratio of any four collinear points [12].

Frame to frame alignment

Frame to frame alignment is essential step before obtaining the moving car detections. By estimating the projective transformation that align successive frames, I_{t+1} and I_t, we can use it to compute the aligned image $\tilde{I}_{t+1}$ which is aligned with I_t. Then we compute the local differences between $\tilde{I}_{t+1}$ and I_t using frame difference. Specifically, we compute the binary image

$$I_t^d(\mathbf{x}) = \begin{cases} 1, & \text{if } |I_t(\mathbf{x}) - \tilde{I}_{t+1}(\mathbf{x})| \geq \tau \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where τ is a suitably determined threshold that trades-off the detection of true regions of local motion versus inevitable noise and other sources of variations in the images. These detection points are presumed to correspond to the locations of moving vehicles in our algorithm.

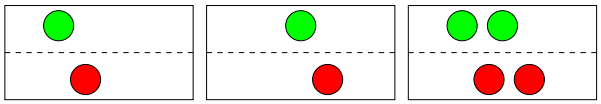


Figure 3: Frame difference result in a favorable two blobs for each moving vehicle due to low frame rate in WAMI.

As illustrated in Fig. 3, each moving vehicle results in two blobs in I_t^d due to the low frame rate in WAMI. One of these blobs can be eliminated using three frame difference [8]. However, in our case, because the two blobs still reside on the road network, we use both to our advantage. In other words, I_t^d contains blobs at locations of the vehicles' in the current frame and in the (compensated) past frame. The total number of such blobs approximates two times the number of vehicles in the scene and using both locations helps improve the accuracy of the subsequent chamfer based alignment in our algorithm.

To align a frame I_{t+1} with the immediately temporally preceding frame I_t, we use efficient alignment strategy. First, we use the enhanced version of FAST (Features from Accelerated Segment Test) [15] algorithm proposed in [16] to detect key-points in both images. The enhancement proposed in [16], allows FAST to have a good measure of cornerness and overcome its limitations for multi-scale features, while keeping its low computational complexity. Then, we extract the descriptors associated with the detected key-points using FREAK (Fast Retina Keypoint) descriptor [17]. Unlike, SIFT or SURF, FREAK yields an efficiently computed binary descriptor which can be matched with much lower computational complexity using a simple Hamming distance measure. Finally we filter out the false matches and estimate the projective transformation that align the two frames using RANSAC.

Road network extraction

The vector map provides the locations of the road segments in geo-referenced coordinates and, for use in our registration process, we are interested in extracting the approximate segments of the road network that lie within the field of view for our WAMI frame.

Since we already have the approximate geographic positions of the four corners of I_t from meta-data, we can compute the projective transformation matrix $\mathbf{H}_g$ that transforms I_t from its coordinate system to the geo-referenced image's system as shown in Fig. 4. We estimate the parameters of $\mathbf{H}_g$ by solving the system of linear equations

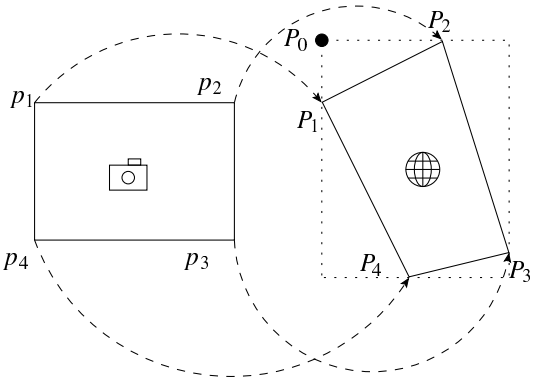


Figure 4: Mapping I_t to a geo-referenced coordinate system.

$$\tilde{\mathbf{P}}_i = s(\mathbf{P}_i - \mathbf{P}_0) = \mathbf{H}_g \mathbf{p}_i, \quad (3)$$

where $i \in \{1,\ldots,4\}$, $\mathbf{p}_i$ are the coordinates of the i^{th} corner point in I_t coordinate system, $\mathbf{P}_i$ are the coordinates of the i^{th} corner point in geo-referenced coordinate system, $\mathbf{P}_0$ is a common reference point, s is a reasonable scaling factor to relate the resolutions of the two coordinate systems. Both $\mathbf{p}_i$ and $\mathbf{P}_i$ are represented in homogeneous coordinate system, and we use the direct linear transformation algorithm (DLT) [12] to compute $\mathbf{H}_g$.

Using the computed $\mathbf{H}_g$ from (3), we project the vector road network into the coordinate system of I_t. Consider the j^{th} road segment characterized by the geographical coordinates of both start and end points. Then, we compute the corresponding pixel locations by the relation

$$\mathbf{p}_j = s\mathbf{H}_g^{-1}(\mathbf{P}_j - \mathbf{P}_0), \quad (4)$$

for the start and end points of the j^{th} road segment in the WAMI image. In other words, we map the geographical coordinate of the start and end points for each road segment into the corresponding pixel locations and then draw a single pixel width line between these points. We use the standard line clipping algorithm [18] to clip the line outside the image region. Thus we obtain a binary image I_t^r which contain the road network represented as series of line segments that are coarsely aligned with I_t using the WAMI meta-data.

Aligning vehicle detections to the road network

To align the binary images I_t^d and I_t^r (obtained as described in the previous sections), we define a distance $f(\beta)$ between them for the alignment specified by the projective transformation with parameters β. We motivate and develop this distance next, where we drop the subscript t to simplify notation. Specifically, let $\mathbf{p}_i^d$ denote the coordinates of the non-zero pixels in I^d, i.e. $\mathbf{p}_i^d = \{\mathbf{x} : I^d(\mathbf{x}) \neq 0\}$, where $i \in \{1,\ldots,N_d\}$, and N_d is the total number of non-zero pixels in I^d and, similarly, let $\mathbf{p}_k^r = \{\mathbf{x} : I^r(\mathbf{x}) \neq 0\}$ be the set of N_r coordinates for which I^r is nonzero, where both $\mathbf{p}_i^d$ and $\mathbf{p}_k^r$ are represented in homogeneous coordinates. We then define the distance $f(\beta)$ as

$$f(\beta) = \frac{1}{N_d} \sum_{i=1}^{N_d} \min_k d(\mathbf{p}_k^r, \mathbf{H}_\beta \mathbf{p}_i^d), \quad (5)$$

where the transformation $\mathbf{H}_\beta$ is as defined in (1) and $d(\mathbf{a},\mathbf{b}) \equiv \|\mathbf{a}-\mathbf{b}\|_2^2$. The nonzero locations in I^r correspond to positions located on the road. Under the (reasonable) assumption that most of the nonzero locations in I^d correspond to vehicle detection locations, this metric can be clearly seen to be intuitively meaningful as the sum of the minimum squared-distances between the vehicle detection locations and the corresponding

DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

nearest points in the road network. Computationally, $f(\beta)$ represents the chamfer distance between I^d and I^r under the projective alignment specified by the parameters β, which can be computed efficiently using distance transform [19]. To align the vehicle detection locations I^d with the road network I^r_t, we therefore seek the optimal projective transformation parameters β^* that minimize the chamfer distance $f(\beta)$.

To compute the optimal parameters, we adopt the Levenberg-Marquardt (LM) [20] non-linear least squares optimization algorithm which minimizes (5) in iterative fashion. In each iteration, the LM algorithm estimate the parameter update vector $\delta \in \mathbb{R}^{8\times 1}$ such that the value of the objective function is reduced when moving from β to $\beta+\delta$ with the parameters converging to a minimum of the objective function with the progression of iterations. The parameters update vector δ is obtained by solving the following system of equations:

$$(\mathbf{A}+\lambda\mathbf{I})\delta = -\mathbf{b}(\beta), \tag{6}$$

where $\mathbf{b} \in \mathbb{R}^{8\times 1}$ is the residual vector which computed as

$$\mathbf{b} = \frac{\partial f}{\partial \beta} = -\frac{2}{N_d}\sum_{i=1}^{N_d} \mathbf{J}_i^T [\min_k (\mathbf{p}_k^r - \mathbf{H}_\beta \mathbf{p}_i^d)], \tag{7}$$

and $\mathbf{J}_i \in \mathbb{R}^{2\times 8}$ is the Jacobian matrix computed at each transformed point $\mathbf{H}_\beta \mathbf{p}_i^d$, which computed as

$$\mathbf{J}_i = \frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial \beta} = \left[\frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_1}, \ldots, \frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_8}\right], \tag{8}$$

and $\mathbf{A} \in \mathbb{R}^{8\times 8}$ is the approximation to the Hessian matrix, obtained as

$$\mathbf{A} = \sum_{i=1}^{N_d} \mathbf{J}_i^T \mathbf{J}_i, \tag{9}$$

where

$$\begin{aligned}
&\frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_1} = \left[\frac{x_i^d}{w}, 0\right]^T, \quad \frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_2} = \left[\frac{y_i^d}{w}, 0\right]^T, \\
&\frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_3} = \left[\frac{1}{w}, 0\right]^T, \quad \frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_4} = \left[0, \frac{x_i^d}{w}\right]^T, \\
&\frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_5} = \left[0, \frac{y_i^d}{w}\right]^T, \quad \frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_6} = \left[0, \frac{1}{w}\right]^T, \\
&\frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_7} = \left[-\frac{x_i^d z}{w^2}, -\frac{x_i^d z}{w^2}\right]^T, \quad \frac{\partial \mathbf{H}_\beta \mathbf{p}_i^d}{\partial h_8} = \left[-\frac{y_i^d z}{w^2}, -\frac{y_i^d z}{w^2}\right]^T, \\
&w = x_i^d h_7 + y_i^d h_8 + 1, \quad z = x_i^d h_1 + y_i^d h_2 + h_3.
\end{aligned} \tag{10}$$

At each iteration, the parameters β is updated to the value $\beta+\delta$, and the process is continued until convergence.

Results

We evaluated our algorithm on a WAMI data set recorded using CorvusEye 1500 Wide-Area Airborne System [3] for the Rochester, NY region. For the vector road map, we use OpenStreetMap (OSM) [21]. OpenStreetMap, is a collaborative project, which uses free data sources such as Volunteered Geographic Information (VGI) [22] to create a free editable map of the world. The map data from OSM is available in a vector format. For example, each road in a road network for a given area is represented by multiple road segments connecting start and end points specified in the map data by their latitude and longitude coordinates. Additionally, many other properties of each road such as its type (highway, residential, etc) and its number of lanes, etc are included in the data.

Our WAMI frames are each 4400×6600 pixels, and stored using NITF 2.1 format [23], which stores a JPEG 2000 encoded image and meta-data within a single file. We parse these files to extract the four approximate geographical coordinates for the corners associated with each aerial frame.

We compare our proposed method with two alternative methods which we will refer to as "Meta-data Based Alignment (MBA)" and "SIFT matching with auxiliary geo-referenced image (SBA)". The MBA method simply uses the aerial frame meta-data to get the aligned road network. The SBA method tries to match SIFT features between the aerial image and an auxiliary geo-referenced image taken from Google Maps, where aerial image meta-data is used to first orhto-rectify the aerial image, and correspondences between this ortho-rectified image and the geo-referenced image are obtained by SIFT feature matching. Specifically, we extract SIFT features from the ortho-rectified aerial image and the geo-referenced image, then for each feature point in one image, we search for the corresponding point in the other image within a circle with radius r, where center of the circle is determined by the approximate alignment parameters from the metadata and the radius of the search is set by determining the maximum spatial error for the approximate alignment provided by the meta-data. After obtaining these putative correspondences, we use RANSAC to filter out the incorrect matches and to estimate the final transformation between the geo-referenced image and the aerial image. We apply this transformation to the vector road network, then undo the ortho-rectification to get the final result. Visual comparison for the road network in some frames aligned with proposed method, and both the SBA method[1], and the meta-data method, are shown in Fig. 8 and Fig. 9. From these images, we can see that the proposed method offers a significant enhancement over MBA which depends only on the meta-data to get an aligned road network and over SBA which uses SIFT and auxiliary geo-referenced Google map image. The MBA method has significant errors because of the inaccuracy of the meta-data parameters due to the limited accuracy of on-board navigation devices. The SBA method does not improve significantly because of spurious correspondences found by the SIFT matching between the aerial image and the Google map image which have significant differences due to severe view point change, different illumination, and different capturing times. Our proposed method does not face the challenges associated with aligning images captured under these different conditions because it aligns vehicle detections to the road network by transforming both into a binary representation that then allows efficient computation of the chamfer distance as a meaningful metric.

To provide quantitative comparison between the methods, we manually generate ground truth road network for few frames[2] and calculate three measures to quantify the accuracy of alignment. First, the chamfer distance between the ground truth road network and the road network generated from our method, the SBA method, and the MBA method are shown in Table 1. The results in the table reinforce the conclusions seen from the visual images. The proposed method has a much lower value for the chamfer distance highlighting the fact that the proposed method offers a significant improvement over both the MBA and SBA methods.

The second quantitative measure that we use is the precision-recall performance. Because the lines in our road network have single pixel width, to obtain a more meaningful metric, we dilate the roads in the road network image by progressively increasing amounts and compute the precision and recall for each dilation amount. For each dilation width, we estimate the true positives (TP), the false positives (FP), and the false negatives (FN), which are shown in Fig. 5 as a function of the dilation width. After specifying TP,FP, and FN, calculating precision and recall is straightforward. Precision-recall plot for frame no. 820 is shown in Fig. 6. Once again, the significant improvement offered by the proposed method over both the MBA and SBA methods is apparent from the plots.

[1] All results of SBA method, are reported using the radius r that gives the best result.

[2] We generate the ground truth for only 4 frames because it is very tedious to manually draw roads in WAMI image.

DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

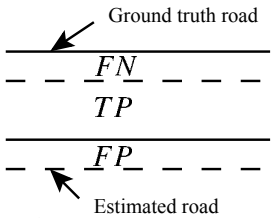


Figure 5: Calculating TP,FP, and FN.

Finally, Fig. 7 shows the precision plot that shows the percentage of accurately estimated road pixels for which the estimated road pixel location is within some threshold distance of the ground truth. These percentages are averaged over the same frames used to report the result in this paper.

Frame no.	MBA	SBA	Proposed method
1	28.22	17.1	**6.36**
300	122.28	83.09	**9.30**
416	36.95	26.49	**8.69**
820	87.35	87.29	**6.68**

Table 1: Chamfer distance between the ground truth road network and the road network generated using the MBA method, the SBA method, and our proposed method.

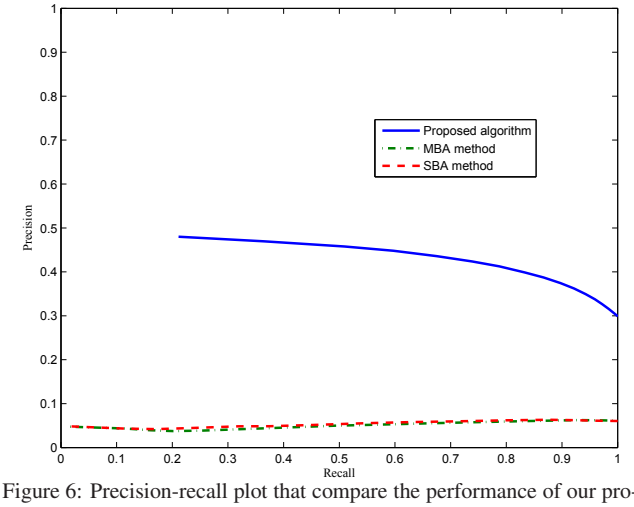


Figure 6: Precision-recall plot that compare the performance of our proposed method with other methods for frame no. 820.

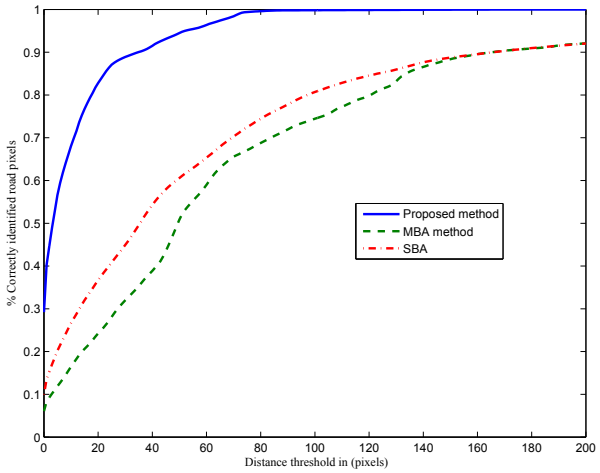


Figure 7: Precision plot that compare the performance of our proposed method with the SBA method and the MBA method.

Our method, implemented in C++ using OpenCV [24], takes ($5 \sim 10$) seconds to align the vector road network with a WAMI frame. The expensive part in our current implementation is the LM minimization, which can be further speeded-up with the support of GPU processing, particularly by parallelizing the Jacobian calculations as these are computed independently for each pixel. Such parallelization, while beyond the scope of the present work, has the potential for making the process real-time allowing deployment of the method on airborne WAMI platforms for real-time applications.

Conclusion

The framework proposed in this paper offers a methodology for accurately registering vector road maps to wide area motion imagery (WAMI) by exploiting vehicular motion. Specifically, local motion observed in the WAMI image frames, after compensation of global motion via standard techniques, corresponds strongly with moving vehicles and by minimizing the chamfer distance between the vehicle locations identified from the local motion and the lines corresponding to the network of roads in the vector map data, we provide an effective method for aligning the two that does not require direct feature matching between these very different data modalities and also eliminates the need for a geo-referenced image as an intermediary. Results obtained for our test datasets show the effectiveness of the proposed methodology. Both visually and in terms of numerical metrics for alignment accuracy, the proposed method offers a very significant improvement over available alternatives. Our future work focuses on exploiting the registered vector road network for vehicle tracking in WAMI, and enhancing both the vector road network registration and vehicle tracking in a joint framework. By leverage locations and directions of trajectories formed by vehicle tracking in road network registration, and by exploiting vector road network in vehicle tracking, both problems can benefit from each other, and result in more accurate and robust solution, an approach that we are pursuing in ongoing work [25].

Acknowledgment

We would like to thank Bernard Brower of Harris Corporation (previously Exelis Inc.) for making available the CorvusEye [3] WAMI datasets used in this research.

References

[1] K. Palaniappan, R. M. Rao, and G. Seetharaman, "Wide-area persistent airborne video: Architecture and challenges," in *Distributed Video Sensor Networks*. Springer, 2011, pp. 349–371.

[2] E. Blasch, G. Seetharaman, S. Suddarth, K. Palaniappan, G. Chen, H. Ling, and A. Basharat, "Summary of methods in wide-area motion imagery (WAMI)," in *Proc. SPIE*, vol. 9089, 2014, pp. 90 890C–90 890C–10.

[3] "CorvusEye™ 1500," http://www.exelisinc.com/solutions/corvuseye1500/Pages/default.aspx.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] W. Song, J. Keller, T. Haithcoat, and C. Davis, "Automated geospatial conflation of vector road maps to high resolution imagery," *IEEE Trans. Image Proc.*, vol. 18, no. 2, pp. 388–400, Feb 2009.

[6] C.-C. Chen, C. A. Knoblock, and C. Shahabi, "Automatically conflating road vector data with orthoimagery," *GeoInformatica*, vol. 10, no. 4, pp. 495–530, 2006.

[7] C.-C. Chen, C. A. Knoblock, C. Shahabi, Y.-Y. Chiang, and S. Thakkar, "Automatically and accurately conflating orthoimagery and street maps," in *Proc. ACM Int. Workshop on Geographic Information Systems*. ACM, 2004, pp. 47–56.

[8] J. Xiao, H. Cheng, H. Sawhney, and F. Han, "Vehicle detection and tracking in wide field-of-view aerial video," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2010, pp. 679–684.

[9] J. Xiao, H. Cheng, F. Han, and H. Sawhney, "Geo-spatial aerial video processing for scene understanding and object tracking," in *IEEE Intl. Conf.*

DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

Comp. Vision, and Pattern Recog., June 2008, pp. 1–8.

[10] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comp. Vis. and Image Understanding.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

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

[12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[13] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proc. Int. Joint Conf. Artificial Intell.*, 1977, pp. 659–663.

[14] A. M. Tekalp, *Digital Video Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[15] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. European Conf. Computer Vision*, ser. Lecture Notes in Computer Science, 2006, vol. 3951, pp. 430–443.

[16] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *IEEE Intl. Conf. Comp. Vision.*, Nov 2011, pp. 2564–2571.

[17] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2012, pp. 510–517.

[18] J. D. Foley, R. L. Phillips, J. F. Hughes, A. v. Dam, and S. K. Feiner, *Introduction to Computer Graphics*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1994.

[19] G. Borgefors, "Distance transformations in digital images," *Comp. Vis., Graphics and Image Proc.*, vol. 34, no. 3, pp. 344–371, Jun. 1986.

[20] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[21] "OpenStreetMap," http://www.openstreetmap.org.

[22] M. F. Goodchild, "Citizens as voluntary sensors: spatial data infrastructure in the world of web 2.0," *Intl. J. of Spatial Data Infrastructures Research*, vol. 2, pp. 24–32, 2007.

[23] NITFS baseline documents. [Online]. Available: http://www.gwg.nga.mil/ntb/baseline/index.html

[24] OpenCV library. [Online]. Available: http://opencv.org/

[25] A. Elliethy and G. Sharma, "A joint approach to vector road map registration and vehicle tracking for wide area motion imagery," submitted to IEEE ICASSP 2016.

DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

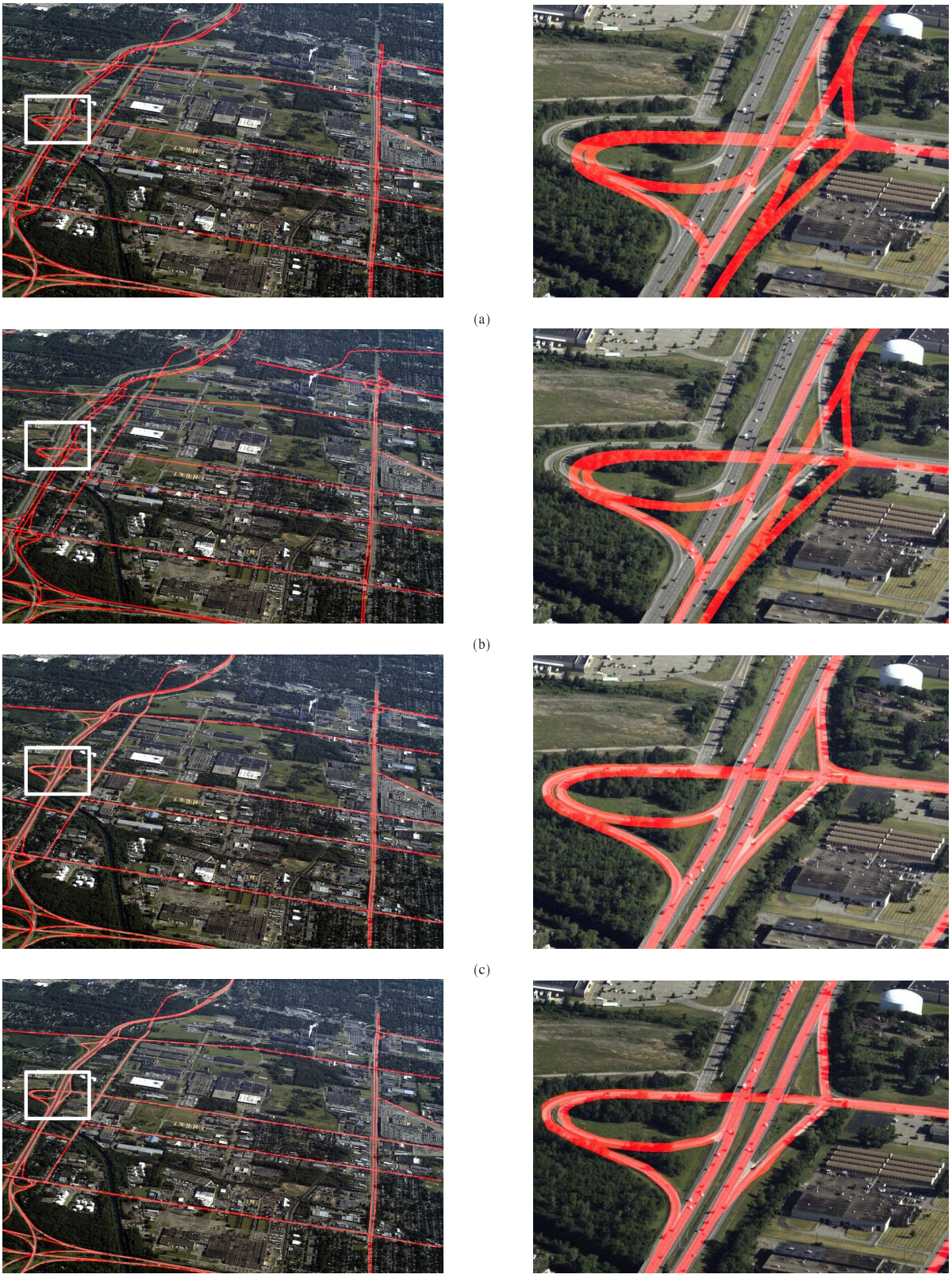


Figure 8: Road network alignment results for frame no. 1 using different methods: (a) MBA, (b) SBA, (c) proposed alignment algorithm, (d) ground truth. Left column is the full frame, while right one shows a smaller cropped region.

DOI: 10.2352/ISSN.2470-1173.2016.3.VSTIA-520

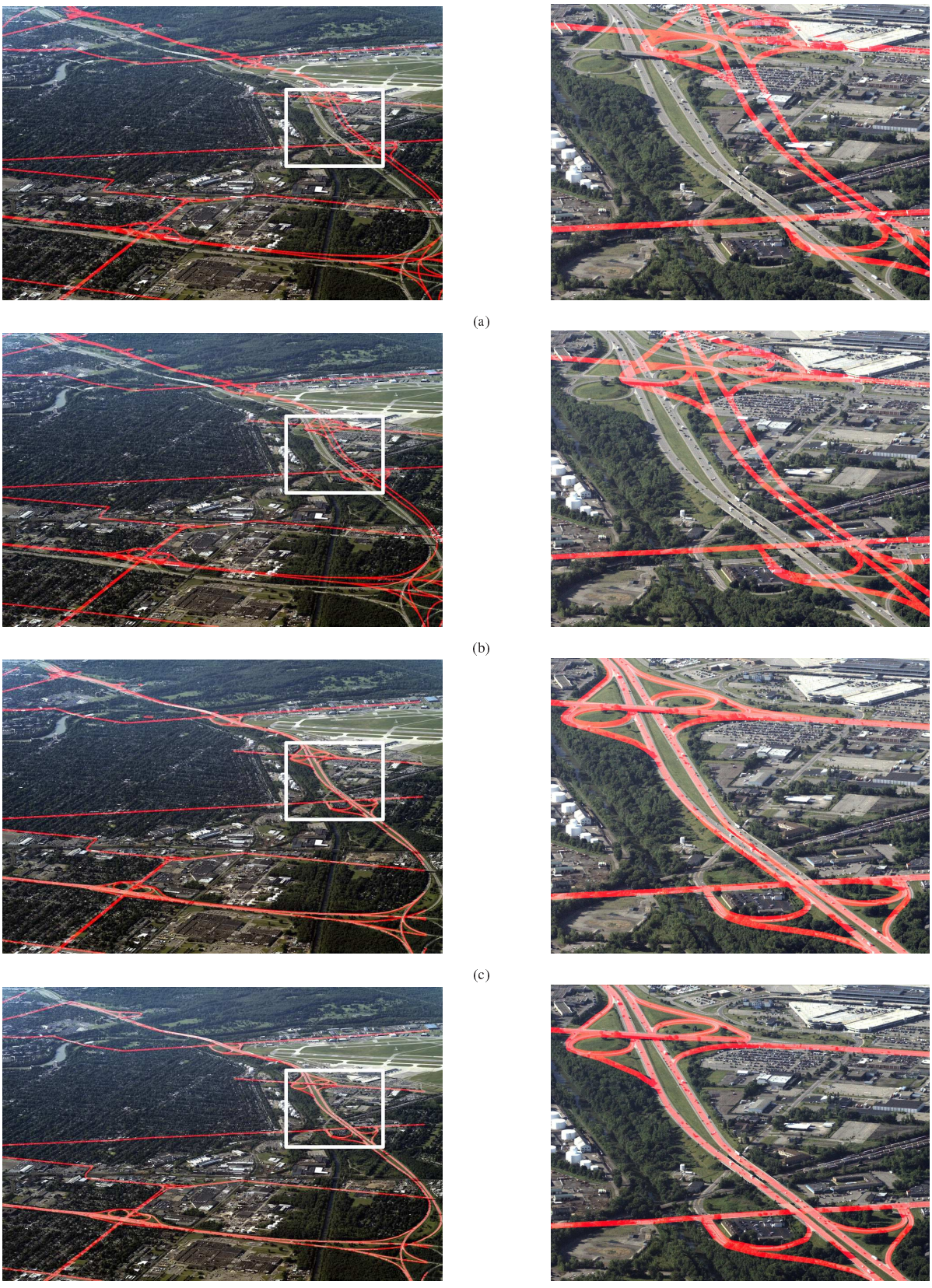


Figure 9: Road network alignment results for frame no. 820 using different methods: (a) MBA, (b) SBA, (c) proposed alignment algorithm, (d) ground truth. Left column is the full frame, while right one shows a smaller cropped region.