Supplementary Material for "Automatic Registration of Vector Road Maps with Wide Area Motion Imagery by Exploiting Vehicle Detections"

Ahmed Elliethy, *Student Member, IEEE* and Gaurav Sharma, *Fellow, IEEE*

This document provides supplementary material, for the paper [2]. In Section S.I, we provide geographical coordinates for the WAMI datasets used in our evaluation. In Section S.II, we summarize file format and other information for the datasets. In Section S.III, we describe how we align successive WAMI frames efficiently and detect tentative vehicle locations in a WAMI frame. In Section S.IV, we include images that visually illustrate and compare additional results for the proposed algorithm and alternatives. Section S.V presents results characterizing the dependence of the accuracy of the registration on the number of vehicle detections in the input. Section S.VI provides a description for the animated GIF included (separately) in the supplementary material for illustrating the LM iteration process. Finally, in Section S.VII, we present results for the proposed algorithm obtained with an alternative vehicle detection approach and also compare the performance against the preliminary version of our work presented in [1].

S.I. Test areas

Table S.I shows the approximate latitude and longitude bounds for each of the test areas for the datasets used in the evaluation in [2].

Data set	Test area	Latitude		Longitude		Area in Km^2
		from	to	from	to	
CORVUS (V)	Area 1	43.154	43.228	-077.697	-077.652	19.2
	Area 2	43.138	43.159	-077.697	-077.675	2.9
	Area 3	43.147	43.204	-077.734	-077.680	14.84
	Area 4	43.102	43.167	-077.682	-077.612	21.71
CORVUS (IR)	Area 5	27.977	28.002	-081.929	-081.901	3.91
	Area 6	27.977	28.000	-081.928	-081.904	4.14
	Area 7	27.978	28.000	-081.927	-081.906	3.96
	Area 8	27.980	28.003	-081.927	-081.906	3.98
WPAFB	Area 9	39.753	39.807	-084.146	-084.077	20.43
	Area 10	39.756	39.805	-084.149	-084.091	22.48
	Area 11	39.752	39.811	-084.158	-084.085	20.91
	Area 12	39.7528	39.809	-084.157	-084.085	21.23

TABLE S.I: Geographic locations of the test areas corresponding to the datasets used for evaluation in [2]. For the WPAFB dataset, the areas from 9 to 12 are corresponding to frames numbers 20091021202517-01000100-VIS, 20091021202530-01000117-VIS, 20091021202545-01000136-VIS, and 20091021202557-01000150-VIS respectively.

S.II. Data sets overview

Frames of all data sets (CORVUS(V), CORVUS(IR), and WPAFB) are stored using the NITF 2.1 format [3], which stores an encoded image and the associated meta-data within a single file. Table S.II gives more detail about each data set.

A. Elliethy is with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627, USA (e-mail: ahmed.s.elliethy@rochester.edu).

G. Sharma is with the Department of Electrical and Computer Engineering, Department of Computer Science, and Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14627, USA (e-mail: gaurav.sharma@rochester.edu).

Data set	Frame size	Number of bands	Encoding type	Single tiles availability (each camera)	Tile size
CORVUS (V)	4400×6600	3	JPEG 2000	No	-
CORVUS (IR)	7204×7330	1	JPEG 2000	Yes	2048×2048
WPAFB	$\approx 20000 \times 20000$	1	JPEG	Yes	4872×3248

TABLE S.II: Information about each data set used for our evaluation in [2].

The WAMI frames in the CORVUS (IR) dataset were already geo-registered before these were provided to us. In order to use this dataset in our evaluation, we synthetically introduce a positional error in the coordinates of the four corners of each WAMI frame in this dataset. We assume that this positional error is uniformly distributed over a range between a minimum and a maximum value. We estimate the positional error range from the CORVUS (V) dataset using a subset containing 50 frames, where in each frame, we measure the distance between the geo-coordinates of each corner provided by the meta-data and manually determined "ground truth" geo-coordinates and use the maximum and minimum to set the range of our positional errors for evaluating the registration for the IR dataset.

S.III. Successive frame alignment & vehicle detection

For the visual datasets (CORVUS(V) and WPAFB), we align the WAMI frame I with the immediately temporally preceding frame I_0 with an efficient alignment strategy. First, we use the enhanced version of FAST (Features from Accelerated Segment Test) [4] algorithm proposed in [5] to detect key-points in both images. The enhancement proposed in [5], allows FAST to have a good measure of cornerness and overcomes limitations for multi-scale features, while maintaining low computational complexity. Then, we extract the descriptors associated with the detected key-points using the FREAK (Fast Retina Keypoint) descriptor [6]. 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, using RANSAC, we filter out the false matches and estimate a planar homography that aligns the two frames.

Because the infra-red images tend to be lower resolution, using the above procedure for aligning successive frames can introduce relatively large misalignment errors, which introduce many spurious vehicle detections. To overcome this problem, we use the SURF features [7] instead for the IR dataset, which perform better under the lower resolution.

For the data sets that provide individual tiles that captured by each camera in the WAMI system (CORVUS(IR) and WPAFB), one could create a full mosaiced frame first, then try to detect vehicles in the mosaiced frame. However, generating a high quality large scale mosaiced frame from tiles that are captured by different cameras is a challenging task. Any mis-registration or visible seam between the tiles, introduces many spurious detections. Instead, we detect vehicles in each tile separately by aligning successive tiles (*from the same camera*) using the alignment method described above, then estimate vehicle detections using the compensated frame difference. Finally, we transform these vehicle detections from all tiles into a common mosaiced frame. We estimate the geometric transformation that aligns each tile to the common mosaiced frame using the method in [8]. In other words, we create a mosaiced frame from only the estimated vehicle detections in each tile, instead of creating the mosaiced frame using the tiles directly and then detecting vehicles inside that mosaiced frame. Thus, we avoid the challenges associated with mosaic generation and detect vehicles with good accuracy.

S.IV. Additional visual results

Figures S.1, S.2, and S.3 show additional visual results on the data sets used in our evaluation. Captions for these figures are self-explanatory.

S.V. Dependence on Number of Vehicle Detections

A semi-synthetic experiment was performed to assess the impact of the number of vehicles on the accuracy of registration. For this purpose, the number of vehicle detections was reduced by random subsampling of our initial set of detections used for the results presented in Fig. 7 (a). Subsampled detections corresponding to $95\%, 90\%, \ldots 15\%, 10\%$ of the complete (100%) set of original detections were used in our algorithm for estimating

the registration parameters and the accuracy of the registration obtained with the corresponding input was quantified in terms of the chamfer distance metric already described in Section IV of the main manuscript [2]. Figure S.4 summarizes the results of this experiment. The results in Fig. S.4 show that the chamfer distance metric remains close to the small value obtained for the full data set, even as the number of detections is subsampled to include as few as 20% of the original detections, although when the number of detections is further reduced to 15% the registration error increases rapidly. Thus the accuracy of the registration is maintained even when the number of detections is a relatively small fraction of the number in our original experiments.

S.VI. Animated GIF showing LM iterations

The GIF file "SampleLMIterationsElliethyTIP2016.gif" included (separately) in the supplementary materials provides a visual demonstration of the LM iteration process for minimization of (10) via equations (11)–(14). This sample result corresponds to Test Area 4. The sequence of frames in the animated GIF show the roadmap and two versions of the mapped versions of the vehicle detections. The first in green corresponds to the initial alignment transform estimated from the meta-data (MBA) and the second in red corresponding to the current iteration estimate of the alignment transform for the LM iterations. The iterative process demonstrates how the LM iterations converge to a value of the registration transform parameters that aligns the detected locations with the road network.

S.VII. Results obtained using vehicle detection by background modeling

In this section, we report alignment results obtained for the registration of the same frames that were used in Section IV of the parent manuscript([2]), where instead of the compensated frame difference we make use of a background modeling approach for detecting vehicles. Specifically, we first estimate the background using the median filter, then we detect vehicles by subtracting the WAMI frame form the estimated background. This method is used in [9], [10] to detect vehicles in WAMI frames for the purpose of tracking.

Table S.III shows the chamfer distance between the ground truth road network and the aligned road network generated from our proposed method and the method in [1], using vehicle detections from the above described background modeling method. By comparing the results shown in Table I in the parent of manuscript ([2]) and in Table S.III, we can draw two important conclusions. First, our proposed method is general and can be used with different vehicle detection techniques, i.e. it is not dependent on the compensated frame difference used in [2]. Second, the results in Table S.III reinforce the conclusion drawn from Fig. 6 in [2], that our proposed method is more robust against vehicle detection errors compared to the method in [1], which shows a large performance change in Table S.III compared with Table I (in [2]).

Data set	Test area	method in [1]	Proposed method
CORVUS (V)	Area 1	16.15	**6.02**
	Area 2	14.43	**3.16**
	Area 3	13.91	**2.96**
	Area 4	8.95	**5.137**

TABLE S.III: Chamfer distance between the ground truth road network and the aligned road network obtained using the preliminary version of our work presented in [1] and using the proposed method (detailed in [2]). The test areas are as defined in Table S.I.

References

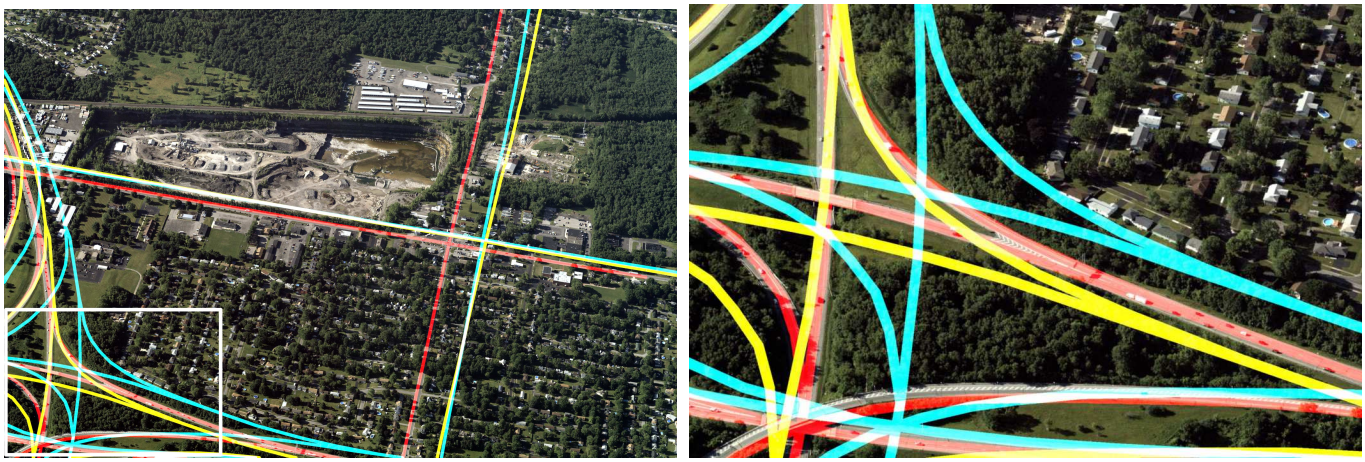
[1] A. Elliethy and G. Sharma, "Vector road map registration to oblique wide area motion imagery by exploiting vehicles movements," in *IS&T Electronic Imaging: Video Surveillance and Transportation Imaging Applications*, San Francisco, California, 2016, pp. VSTIA–520.1–8. [Online]. Available: http://ist.publisher.ingentaconnect.com/contentone/ist/ei/2016/00002016/00000003/art00008

[2] ——, "Automatic registration of vector road maps with wide area motion imagery by exploiting vehicle detections," *IEEE Trans. Image Proc.*, accepted August 2016.

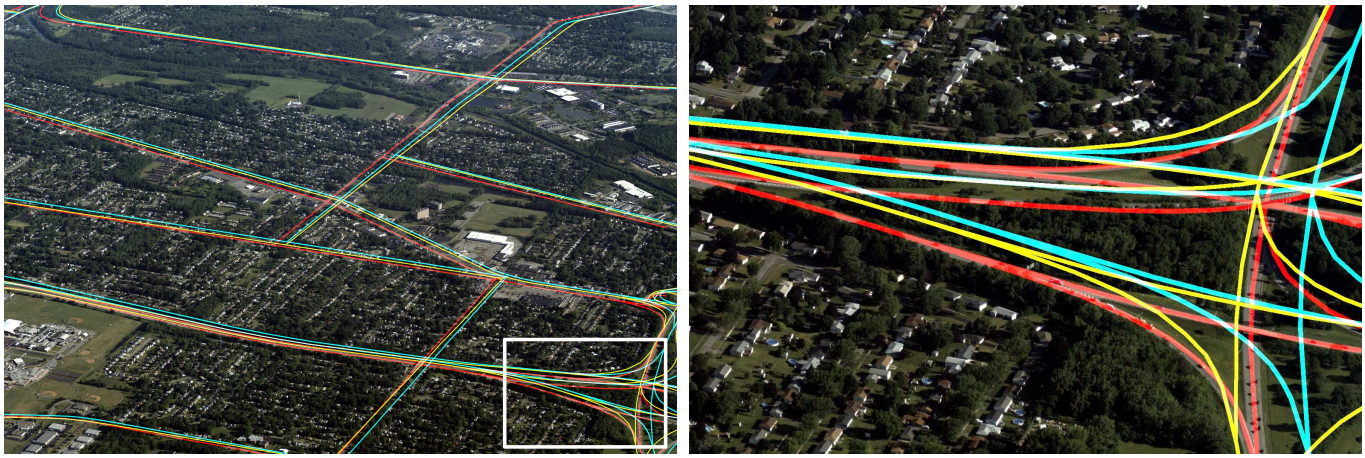
[3] National imagery transmission format (NITF) standard, version 2.1. [Online]. Available: http://www.gwg.nga.mil/ntb/baseline/index.html

[4] 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.

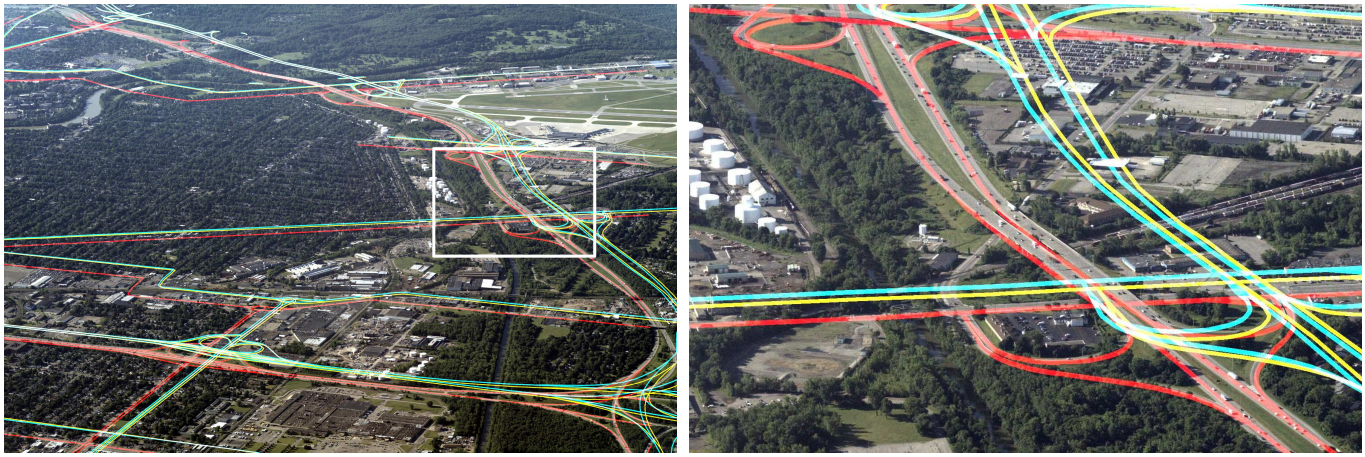
[5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *IEEE Intl. Conf. Comp. Vision.*, Nov 2011, pp. 2564–2571.
[6] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2012, pp. 510–517.
[7] 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.
[8] J. Prokaj and G. Medioni, "Accurate efficient mosaicking for wide area aerial surveillance," in *IEEE Workshop on Appl. of Comp. Vision.*, Jan 2012, pp. 273–280.
[9] V. Reilly, H. Idrees, and M. Shah, "Detection and tracking of large number of targets in wide area surveillance," in *Proc. European Conf. Computer Vision*, 2010, vol. 6313, pp. 186–199.
[10] X. Shi, P. Li, H. Ling, W. Hu, and E. Blasch, "Using maximum consistency context for multiple target association in wide area traffic scenes," in *IEEE Intl. Conf. Acoust., Speech, and Signal Proc.*, May 2013, pp. 2188–2192.



(a)

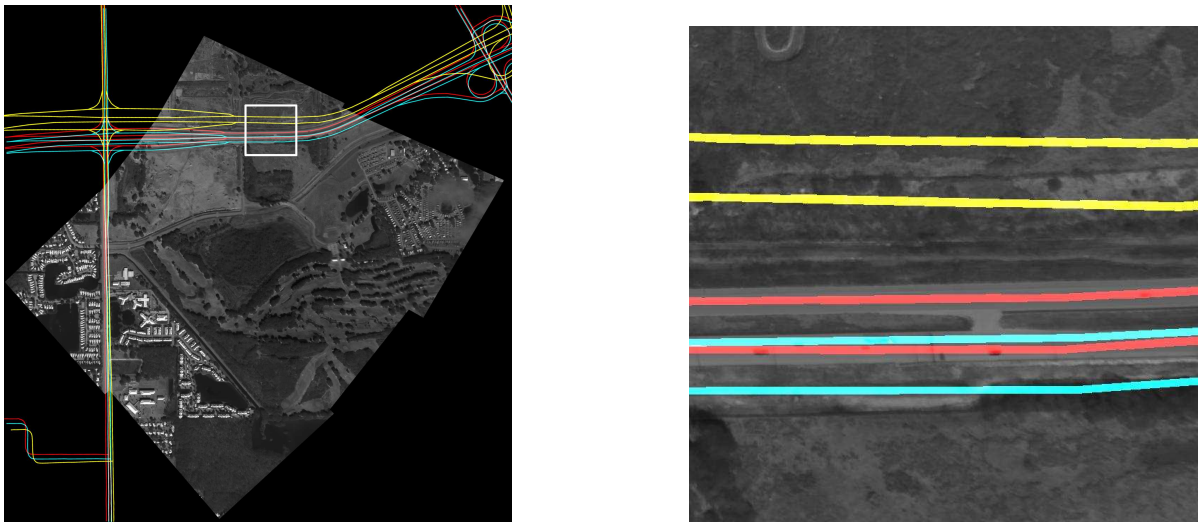


(b)

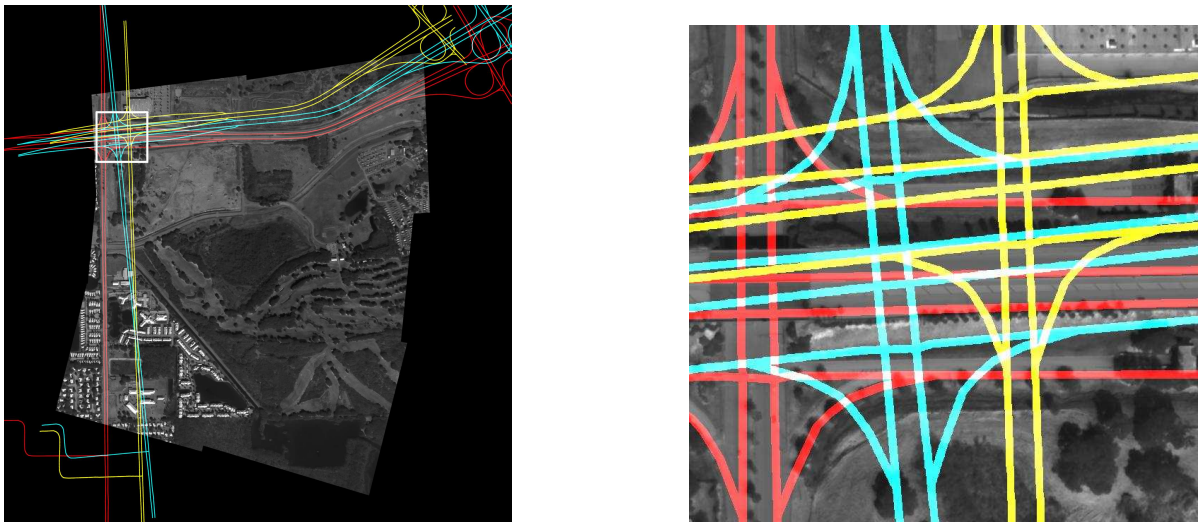


(c)

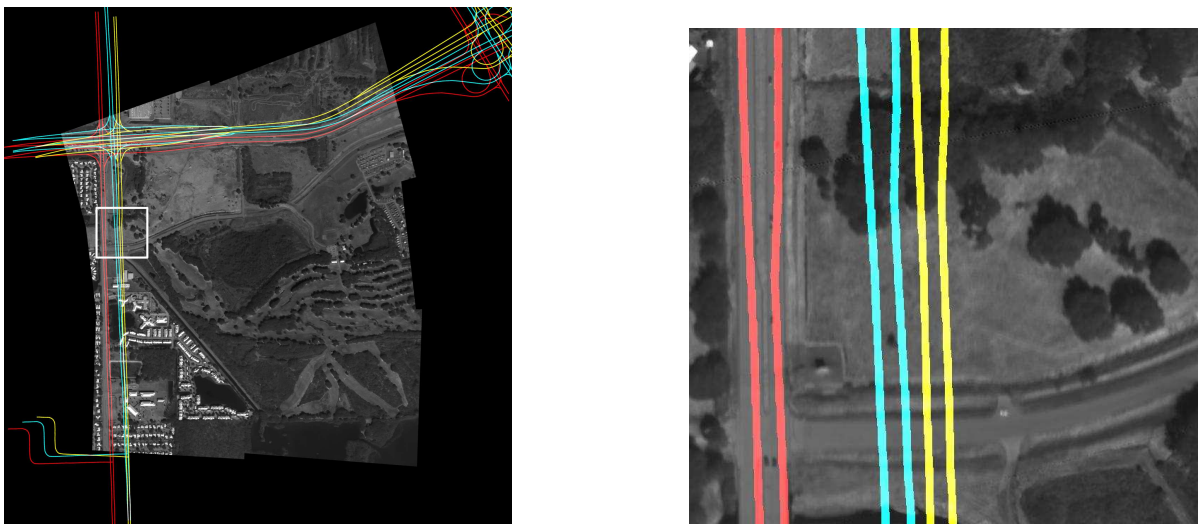
Fig. S.1: Road network alignment results on three CORVUS(V) frames using the different methods for: (a) Area 2, (b) Area 3, and (c) Area 4. The initial road network obtained from the WAMI frame meta-data is shown in cyan, while the result of the SBA method and our proposed method appear in yellow and red colors, respectively. Left column is the full WAMI frame, while the right column shows a smaller cropped region that is marked on the corresponding full frame by a white rectangle.



(a)

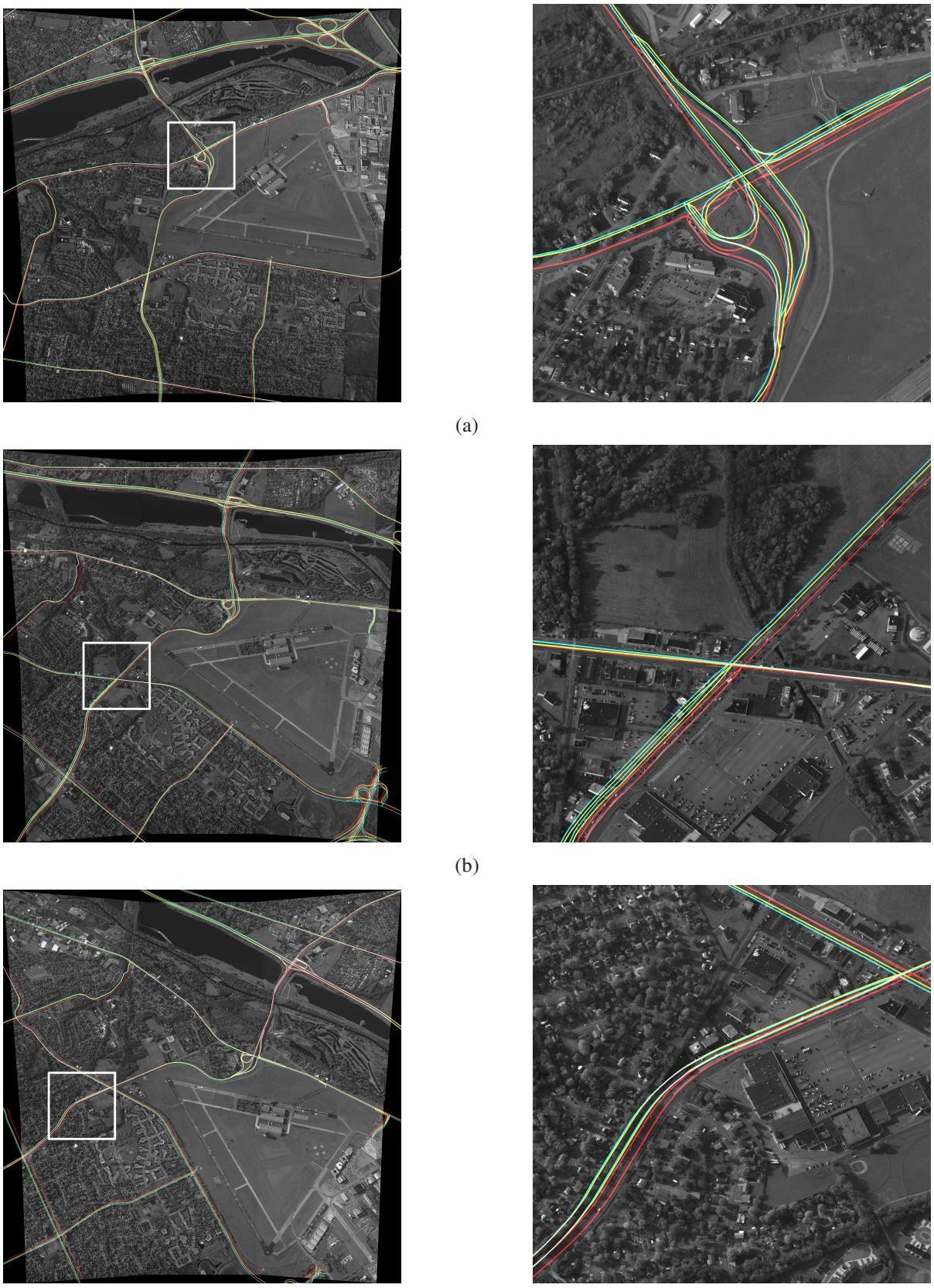


(b)



(c)

Fig. S.2: Road network alignment results on three CORVUS(IR) frames using the different methods for: (a) Area 5, (b) Area 7, and (c) Area 8. The initial road network obtained from the WAMI frame meta-data is shown in cyan, while the result of the SBA method and our proposed method appear in yellow and red colors, respectively. Left column is the full WAMI frame, while the right column shows a smaller cropped region that is marked on the corresponding full frame by a white rectangle.



(a)

(b)

(c)

Fig. S.3: Road network alignment results on three WPAFB frames using the different methods for: (a) Area 10, and (b) Area 11, and (c) Area 12. The initial road network obtained from the WAMI frame meta-data is shown in cyan, while the result of the SBA method and our proposed method appear in yellow and red colors, respectively. Left column is the full WAMI frame, while the right column shows a smaller cropped region that is marked on the corresponding full frame by a white rectangle.

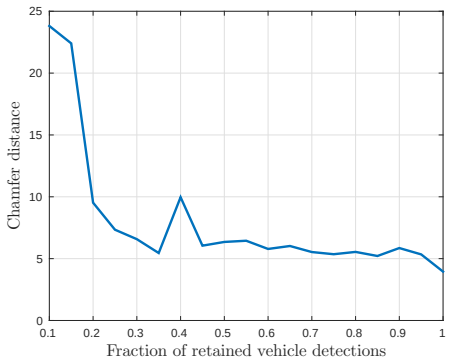


Fig. S.4: Chamfer distance between the ground truth road network and the road network for the proposed method as a function of the fraction of detections retained from the original detections. Results summarized here correspond to the data of Fig. 7 (a).