HIGH PRECISION MAP CONFLATION OF FLEET SOURCED TRAFFIC SIGNS

Vaibhav Kango, Hesham M. Eraqi, and Mohamed Moustafa

{kangov, heraqi, mmoustm}@amazon.com

ABSTRACT

The ever-increasing demand for digital maps in various domains amplifies the importance of having accurate and up-to-date maps. To address this, the proposed system pervasively conflates large volume of sign detections recorded by a transportation fleet of vehicles into map database. Detected and geo-localized sign objects collected from the fleet over a time period are passed through a context-aware clustering for aggregation and followed by map matching with a new Hidden Markov Model (HMM) that utilizes vehicle GPS and compass sensors. Eventually, a resolution model utilizes features from detection, traversal, and surrounding map context to assign a confidence score for identified new signs for ingestion via a rapid resolution route. On a test data across the USA geography with large volume of detections, the proposed system could add 51% stop signs and 21% of traffic lights to the map at average precision of 99.55% via rapid resolution path.

Index Terms— Conflation, Fusion, Mapping, Traffic Sign, Map Matching, Last Mile, Pervasive Systems, Rapid Resolution

1. INTRODUCTION

In recent years, the demand for digital maps has been ever-increasing in various domains such as logistics, delivery, fleet management, autonomous driving, and ride-sharing, ensuring that these maps are accurate and up-to-date is important [1]. Accurate and feature-rich maps are key to enable safe and efficient last mile delivery planning and navigation.

While there has been extensive amount of literature on detection [2][3][4] and geo-localization [5] of traffic signs, there is very limited work [6][7] on fusion of the detected signs by real-time systems into map on a large scale. Integration of the identified map entity that includes geo-referencing, identifying underlying context, and aligning to appropriate layer with respect to the defined guidelines is referred to as conflation [8][9]. Conflation of traffic signs plays a crucial role in maintaining high quality and consistency to digital map. Even when detection of traffic signs is accurate, automated fusion of detected signs pose major challenges. At large scale, vehicles driving on similar trajectories detect same sign several times with different levels of accuracy in their estimated geo-location on the ground. Aggregation and deduplication of these detections becomes complex and requires understanding of the underlying map context and large scale human effort. Some of these challenges are illustrated in figure 1.

To overcome these challenges, our proposed system extracts and utilizes contextual features to automatically conflate traffic signs detections with or without manual intervention. The system, first, aggregates multiple instances of sign detections through context-aware clustering and then associates signs to the underlying road network through a novel Hidden Markov Model (HMM)-based map matching model that utilizes compass sensors in addition to vehicle GPS traces. At large scale operations, human effort to determine and review which signs should be added or updated to the map wastes huge cost and time. The proposed system carries out spatial and attributes-based comparison with existing signs present in the map and an auto-reviewer model utilises detection, vehicle and contextual features to flag high confidence signs for immediate resolution.

(a) (b) (c) (d)

Fig. 1: Key challenges. (a) Geo-locations variability of same sign detected by multiple vehicles; (b) Localization error due to sun glare (here); (c) Importance of road segment context from which speed limit is detected; (d) Modeling traffic light on a complex intersection. Yellow dots are detection locations, brown arrow are vehicle location and heading.

2. PROPOSED SYSTEM

Over a time period, traffic sign objects from fleet camera images are detected and geo-localized using deep learning based detection algorithm and Structure-from-Motion (SfM) 3D reconstruction [10], respectively. As shown in figure 2, the proposed conflation system is composed of four components: 1) context-aware clustering that aggregates locations of detections and assigns a representative location for the real sign posts, 2) map matching traffic signs to the underlying road network through a novel HMM model, 3) conflict detection to determine new signs based on spatial and attribute-based comparison with existing map, and finally 4) the auto-reviewer model classifies signs for prompt resolution.

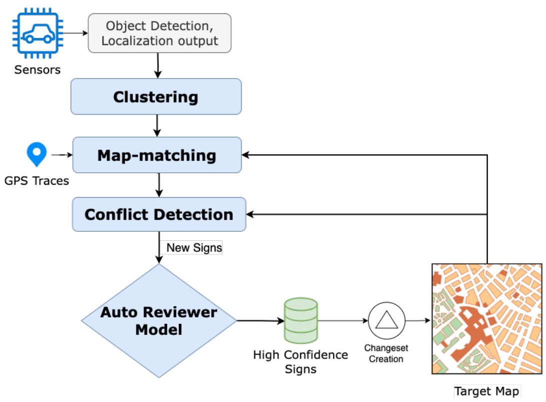

Fig. 2: System Overview

2.1. Clustering

Introducing contextual features into clustering enables 1) error-tolerance to geo-localization inaccuracies, 2) having additional cue about input detections accuracy, and 3) differentiating detections from different sign posts in a close proximity. The latter case is common around intersections where sign posts of the same type, such as traffic light heads or stop signs signaling to different incoming ways of traffic, can be in a close distance in presence of localization approximations. Without contextual features, under-clustering (grouping of detections from different sign posts as in Figure 3 a1 and a2) is inevitable which may lead to inaccurate estimation of signs actual locations on the ground. On the contrary, allowing over-clustering leads to significant amount of duplicate signs wasting considerable review effort and cost. Any sub-optimal cluster may cause an inaccurate sign location to be positioned and pose infiltration risk of bad data during map ingestion. Our intuition is that detections originating from same sign post should have similar orientation; i.e.; comparable poses in the bird's-eye-view perspective and comparable relative position with respect to the location and heading of the vehicle at the time of the detection event. First, we build an aggregated confidence score for each detection calculated as the weighted average of confidences derived from: (a) angle of detection, (b) distance of detection from vehicle, (c) object detection, and (d) localization. The detection angle confidence is measured linearly scaled from 0 to 1 corresponding to the angle between the line of sight (vehicle to detection vector) and vehicle heading vector ranging between 90° (detection orthogonal to vehicle direction) and 0° (in front of vehicle), respectively. For the detection distance feature, beyond a certain distance farther a detection is localized from vehicle position the less confidence it gets assigned. Next, we adopt a conditional and variable distance-based clustering. The clustering distance threshold is inversely proportional to detections average confidence score. This intuition characterizes outlier detections indicated by their lower confidence score and they 1) tend less to form separate clusters and 2) contribute least to estimating signs representative locations on the ground. In addition, to distinguish detections from different sign posts localized within close distance, the feature vector having 1) sign pose, 2) detection angle, and 3) vehicle heading angle is required to be similar when compared with cosine similarity. Sign pose feature is excluded for traffic lights in absence of their planar shape. Finally, the weighted median of the confidence scores of all detections in a cluster is used to elect a representative location that corresponds to the "most likely" coordinates of the physical sign post. Detected signs are localized by 3D-reconstructing its surrounding scene. 2D feature correspondences are computed through multiple images and build a 3D point cloud model using SfM [10]. In that process, a sign produces multiple 3D points that should form a plane for planar sign boards. We fit such 3D points of a sign to a line in bird's-eye view; the sign pose vector is orthogonal to it, heading towards the vehicle. RANSAC is used to due to robustness against outliers. Figure 4 illustrates the estimated poses of signs.

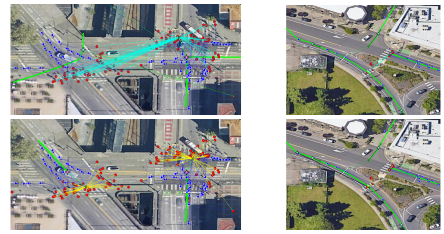

Fig. 3: Results of clustering without (upper) and with contextual features (lower). Blue arrows show vehicle positions and headings and red dots connected to them are detections. Lines colored as light blue and yellow show individual clusters.

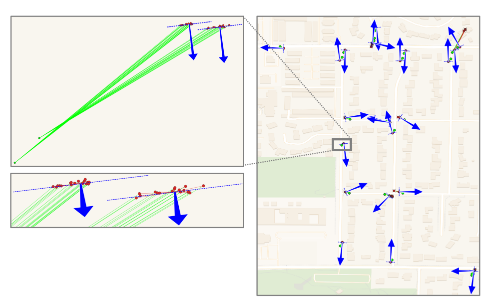

Fig. 4: Sign pose estimation. Each blue arrow shows the estimated pose of a sign detection. Green lines connect 3D points (red dots) to vehicle location (green dot).

2.2. Map Matching

Inherently, a traffic sign affects the flow of traffic on the road segment it can be directly "seen" from. It is important to identify the road segment, a map context, for which the traffic is affected by the detected sign. Our approach to infer the affected road segment is by map matching vehicle GPS trajectory around detection event timestamp using an upgraded version of the model in [11]. We upgrade the model by incorporating the vehicle heading information from the compass sensor which can be a strong cue to improve map matching accuracy. Without heading information, the original model causes road direction ambiguity such as in U-turns. The measurement distribution models vehicle heading and GPS points as zero-mean multivariate Gaussian as in equation 1. Our intuition is that for a moving vehicle, its heading is expected to be aligned with the road segment driving direction following zero-mean Gaussian noise [12].

$$p(m_t|r_i) = \frac{1}{2\pi\sqrt{|\Sigma_M|}} e^{-0.5 M_t^T \Sigma_M^{-1} M_t},$$
$$\Sigma_M = \begin{bmatrix} \sigma_m^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix}, M_t = \begin{bmatrix} ||m_t - x_{t,i}||_{great\_circle} \\ |sin(\frac{\delta_{\theta_t, r_i}}{2})| \end{bmatrix}. \quad (1)$$

m_t is the GPS measurement at time t, r_i is road i, $x_{t,i}$ is m_t projection on road i, δ_{θ_t, r_i} is the angular difference between vehicle orientation θ_t at time t and the directed orientation of the road r_i around x_i, the covariance matrix Σ_M having σ_m and σ_θ are estimated using Median Absolute Deviation (MAD) as in [11]. Figure 5 shows an example of the notation and figure 6 illustrates map matching.

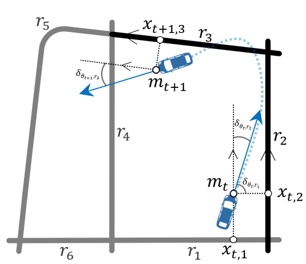

Fig. 5: HMM measurement distribution notation

In addition to road segment, for a category of intersection-based signs such as stop or traffic lights, we determine intersection as another map context. This is evaluated using spatial search of intersections in small vicinity and angle of view relative to vehicle position and direction. As a result, each clustered sign is available with map context of road segment and/or intersection, if applicable.

2.3. Conflict Detection

A considerable proportion of detections may already be present in the existing version of map. To expand sign coverage, the goal of conflict detection module is to classify clustered and map-matched signs as "new" or "existing" based on their location and attributes. Intersection-based signs such as stop and traffic lights are added to map on incoming road segments of the intersection junction. Therefore, we interpolate their locations at a fixed distance on incoming road segments to the intersection node. This provides a final derived location of the sign along with map context previously evaluated. Way-based signs such as speed limits and height or weight restrictions are projected to the nearest point to their map-matched way geometry. The derived location of sign, their map context (road segment or intersection) and other special attributes (label value such as speed limit value) are matched spatially with the geometry of the existing map attributes.

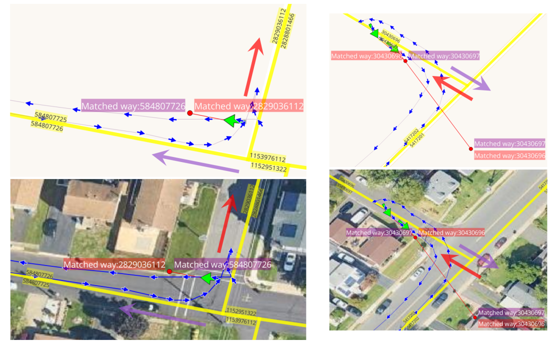

Fig. 6: Map matching illustration. Detection locations in red dots, road segment as yellow, vehicle traces as blue arrows and vehicle position/headings as green arrows at detection time. The red and magenta matched segments are those matched using [11] and our models, respectively. Our upgraded model provides accurate matching particularly in complex cases.

2.4. Auto-Reviewer Model

While sign conflation aggregates and removes existing signs on the map, we take a step further to identify high quality signs that can be auto-resolved into map. Automated reviewer system utilizes newly proposed sign details and predicts if they can be added directly into map.

Our model uses Gradient Boosted decision tree trained with contextual features to predict a confidence score for every sign that reflects its quality in terms of ingestion into map without requiring any edit. These features are derived from following attributes: (i) detected image attributes: detection confidence, bounding box size and location of the detection in image; (ii) sensor attributes: heading, speed, GPS accuracy; (iii) location attributes: number of raw detections grouped in cluster, sparseness of detection locations in cluster, distance between detection and vehicle position; (iv) contextual attributes: intersection complexity, road class, occlusion flag if detection sight is obstructed by any road geometry and angle of road geometry to line of detection sight. To avoid potential risk of ingesting bad edits into map at scale, only the signs

Table 1: Comparing clustering precision (%) with and without contextual features

Region	Traffic Light		Stop Sign	
	Context-aware	Fixed dist.	Context-aware	Fixed dist.
Region A	91.97	79.21	99.50	96.96
Region B	88.89	71.57	97.45	96.01
Average	**90.43**	75.39	**98.48**	96.49

Table 2: Auto-Reviewer audit precision (%) evaluation for stop sign and traffic light

	Stop Sign						Traffic Light
	Region A	Region B	Region C	Region D	Region E	Total	Mixed regions USA
Sign Type (Detection)	99.91	99.91	100	100	99.94	99.95	99.68
Derived direction attribute	98.53	99.63	98.97	99.64	99.55	99.30	100
Derived sub-type (all versus minor)	98.81	99.63	98.97	99.64	99.48	99.33	NA
Derived Sign Location	99.82	99.91	99.77	99.64	99.55	99.72	99.55
Precision (all attributes)	**98.53**	**99.63**	**98.97**	**99.64**	**99.48**	**99.28**	**99.55**

with confidence score higher than a configured threshold, are routed to a rapid resolution path. The threshold to route signs for rapid resolution is carefully selected through model validation, and iterations of audits to meet high precision requirements. The model operates for stop sign and traffic-lights at 99%+ precision to keep high quality and consistency of the map.

Additional attributes for each sign type such as stop type (All or minor), affected traffic direction (forward or backward) are derived using the direction of vehicle and map context. These details are finally sent for changeset generation before ingestion to map database.

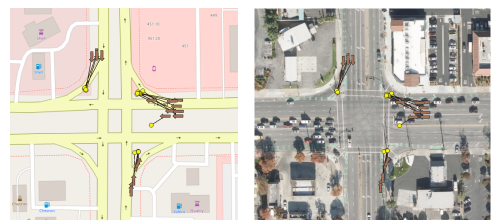

Fig. 7: Detections at complex intersections including slip/via lane involves decision to first identify and next target multiple intersection nodes are rejected by auto-reviewer model. Legend: yellow circles as detection locations, brown arrow as vehicle location and heading, respectively.

3. EVALUATION

3.1. Clustering and Map Matching

We evaluate the context aware clustering method by comparing it against spatial clustering based on a fixed distance threshold. Table 1 summarizes results comparing both methods on test data from 2 regions. The results demonstrate the effectiveness of utilizing contextual features. Figure 3, in previous section, compares the clustering method output with respect to baseline method with fixed distance. Pose estimation achieves a Mean Absolute Error (MAE) of 18.78° on annotated test set of 200 signs. For map matching, on manually collected test dataset of 886 challenging detection cases from region A, the upgraded model improves accuracy from 26.41% to 84.20% compared to baseline method [11]. Figure 6, in previous section, illustrates the map matching output from the two models.

3.2. Auto-Reviewer

To maintain high quality and consistency of map, signs are required to meet high precision for rapid addition at a large scale. To measure accuracy, signs are marked as true positives only when all attributes including sign class type (from upstream detection) and conflation derived features including derived location, specific derived attributes (all vs minor or forward vs backward) are validated. We audit statistically significant sample over a set of mixed regions to evaluate precision and recall. Table 2 show results for sign types stop and traffic lights from recent evaluations. Recall for stop sign and traffic light, at precision above 99%, are 51.2% and 20.1%, respectively. Modeling traffic lights in complicated situations such as complex intersections, slip lanes requires nuanced human judgement (Figure 7).

4. CONCLUSION

We propose a cost-effective and scalable system for automated fusion of traffic sign attributes for map enrichment. Our system addresses key challenges of maintaining quality of signs detected from sensor equipped vehicles. Across the US, we evaluate our precision as 99+% for intersection based signs such as stop sign and traffic light for map ingestion without human involvement. This allows the end-to-end process of fusion in a rapid manner thereby minimizing the time needed for manual review and corrections.

5. REFERENCES

[1] Gorisha Agarwal, *A principled approach to automated road network conflation*, Ph.D. thesis, University of British Columbia, 2021.

[2] Yanmei Jin, Yusheng Fu, Wenqin Wang, Jinhong Guo, Chunhui Ren, and Xin Xiang, "Multi-feature fusion and enhancement single shot detector for traffic sign recognition," *IEEE Access*, vol. 8, pp. 38931–38940, 2020.

[3] Christine Dewi, Rung-Ching Chen, Xiaoyi Jiang, and Hui Yu, "Deep convolutional neural network for enhancing traffic sign recognition developed on yolo v4," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 37821–37845, 2022.

[4] Weidong Min, Ruikang Liu, Daojing He, Qing Han, Qingting Wei, and Qi Wang, "Traffic sign recognition based on semantic scene understanding and structural traffic sign location," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15794–15807, 2022.

[5] A. Chigorin and A. Konushin, "A system for large-scale automatic traffic sign recognition and mapping," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3/W3, pp. 13–17, 2013.

[6] Yifang Yin, Jagannadan Varadarajan, Guanfeng Wang, Xueou Wang, Dhruva Sahrawat, Roger Zimmermann, and See-Kiong Ng, "A multi-task learning framework for road attribute updating via joint analysis of map data and gps traces," in *Proceedings of The Web Conference 2020*, New York, NY, USA, 2020, WWW '20, p. 2662–2668, Association for Computing Machinery.

[7] Homayoun Jamshidi, Thomas Lukaszewicz, Amin Kashi, Ansel Berghuvud, Hans-Jürgen Zepernick, and Siamak Khatibi, "Fusion of digital map traffic signs and camera-detected signs," in *2011 5th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2011, pp. 1–7.

[8] Ting L. Lei, "Large scale geospatial data conflation: A feature matching framework based on optimization and divide-and-conquer," *Computers, Environment and Urban Systems*, vol. 87, pp. 101618, 2021.

[9] Juan J Ruiz, F Javier Ariza, Manuel A Urena, and Elidia B Blázquez, "Digital map conflation: a review of the process and a proposal for classification," *International Journal of Geographical Information Science*, vol. 25, no. 9, pp. 1439–1466, 2011.

[10] Shimon Ullman, "The interpretation of structure from motion," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 203, no. 1153, pp. 405–426, 1979.

[11] Paul Newson and John Krumm, "Hidden markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2009, pp. 336–343.

[12] Britta Hummel, "Map matching for vehicle guidance," in *Dynamic and Mobile GIS*, pp. 211–222. CRC Press, 2006.