An Efficient Data Driven Algorithm for Multi-Sensor Alignment

Feng Guo, Gaurav Aggarwal, Khurram Shafique, Xiaochun Cao, Zeeshan Rasheed, and Niels Haering

ObjectVideo Inc., Reston, VA 20191

Abstract. This paper describes how model-specific constraints and domain specific knowledge can be utilized to develop efficient sampling based algorithms for robust model estimation in the presence of outliers. As a special case, a robust algorithm for homography estimation is proposed that exploits the invariance of collinearity under homography to improve efficiency in noisy scenarios. Unlike most existing approaches, the proposed algorithm does not make any assumption regarding the data distribution, data specific properties or availability of large amount of data. The proposed estimation algorithm is applied for multiple applications involving large sensor networks. These include estimation and maintenance of geo-registration by fusing observations from different modalities, such as RADAR and Automatic Identification System (AIS), and data-driven estimation (using target observations) of the relative topology of cameras with overlapping fields of view. Qualitative and quantitative results are presented that show the ability of the proposed algorithm to handle large fraction of outliers in the data, spatial noise, and high traffic densities, which are defining characteristics of these applications.

1 Introduction

Modern automated video analysis systems consist of large networks of heterogeneous sensors with different output characteristics, for example, static surveillance cameras, Pan-Tilt-Zoom (PTZ) cameras, radars,infrared cameras, hyperspectral sensors [1–6]. These systems must not only perform content extraction on individual sensors but must also integrate and fuse the information from different sensors to effectively provide site-wide situational awareness. A critical step to analyze and fuse data for site-wide scene understanding is to map observations from multiple sensors to a common coordinate system. For example, target observations from multiple sensors may be mapped to the geodetic coordinate system and displayed on a map-based interface. Availability of such a mapping also enables critical operational tasks, such as fusion of multiple target measurements over the network, reasoning about target's relative or metric sizes and speeds, target hand-off across the sensors, tasking of PTZ and mobile sensors, and site-wide inference. Due to these benefits, many multi-sensor video analysis systems require geo-registration and/or inter-sensor calibration to be

performed at the time of installation [1, 3]. However maintaining registration among a large network of sensors is a daunting task as minor drifts in a sensor's geometry may introduce large localization errors. In this paper, we present a data driven approach for the estimation of homography in planar scenes for both semi-supervised and unsupervised multi-sensor (geo)registration in large sensor networks.

There is a large body of research that uses image features for supervised and unsupervised sensor registration. In its simplest form, for planar scenes and perspective cameras, registration can be performed by manually providing four or more corresponding points in the image and the reference image (or map) [7]. However feature based techniques are not feasible in many scenarios (for instance, cameras overlooking water regions or other terrains that do not contain many discriminative or well localizable features), and even manual registration require long setup time, specialized tools, and tedious labor. The shortcomings of feature-based methods and the need for automated methods to estimate and maintain sensor geometry has motivated data-driven approaches that use sensor observations (for example, pairs of simultaneous target detections) over time to infer sensor geometry and network topology [8–11] (The differences between the proposed approach with these existing methods are highlighted in Section 2). Since these methods use target observations as their primary feature, not only do they enable sensor registration in feature-less terrains, but are also agnostic to the sensor's output characteristics. In addition, they may also automatically adapt to the changes in sensor geometry [11]. The main challenge (that these methods face) is to automatically and robustly extract inliers from noisy observations. How they tackle this issue differentiates them from each other. The proposed approach exploits geometric invariance and domain specific knowledge in a RANSAC [12] based sampling framework to efficiently identify the inliers from noisy data. In particular, the method uses the invariance of collinearity among points under homographic transformation to obtain a reduced search space where the percentage of outliers (in the data) is significantly less than the original search space. We provide both analytical and quantitative results showing that the proposed approach outperforms the existing methods by efficiently and robustly estimating planar homography in very noisy scenarios (with more than 85% outliers) in the presence of lens distortion. We apply the proposed algorithm for multiple applications involving large sensor networks. These include, i) fusing data from targets that broadcast their geo-positioning information, such as dedicated mobile units (boats, people, vehicles) or targets that broadcast automatic idenficiation system (AIS) information and vehicles and people with RFID tags, for periodic geo-registration of the system, ii) fusing information from sensors that report geo-position of targets in their field of view (e.g., RADAR) with the visual sensors for automatic geo-registration and maintenance, and iii) alignment of multiple visual sensors with overlapping fields of view by fusing the target observations from each sensor. All of these applications can be viewed as special cases of data driven sensor alignment problem

formally defined in Section 4. In the rest of this paper, we collectively refer to these problems as data driven multi-sensor alignment.

The organization of the paper is as follows. In the next section, we present a survey of the related work. In Section 3, we analyze how model-specific constraints can be utilized to speed-up model estimation in the presence of outliers. In Section 4, we present the technical details and analysis of the proposed inlier identification method for planar homography estimation. In Section 5, we demonstrate the performance of the proposed approach for different applications involving semi-supervised and unsupervised estimation of sensor alignment. The paper is concluded in Section 6.

2 Related Work

Data driven approaches for multi-sensor registration can be classified as either supervised [13–16] or unsupervised [9, 10, 17–19, 11] based on whether or not they require training data or user-supervision. They can be further classified based on the type of model they learn, for example, geometric models [14, 9–11] and correspondence models [13, 15, 17, 19, 16]. The geometric models explicitly describe the geometric relationship between sensors (or a sensor and a reference) and provide a common coordinate frame for sensor fusion and inter-sensor tracking. Geometric models can be used to fuse sensors with both overlapping and non-overlapping fields of view (the later providing that the geo-registration is available). The correspondence models however represent the likelihood of observations across sensors to be corresponding and are advantageous in inter-sensor tracking when the fields of view of sensors do not overlap and geo-registration is not available.

The method presented in this paper estimates geometric model, i.e., planar homography in an unsupervised fashion and hence is more closely related to [9], [10], and [11]. In [9], an unsupervised data driven method is presented that uses RANSAC on simultaneously occurring target observation pairs (detections) to obtain planar homography in sensors with overlapping fields of view. Although RANSAC can handle more than 50% of outliers in the data, it is known to become prohibitively expensive as this fraction increases [20] and at times not converging to a solution even after hundreds and thousands of iterations [10]. Recently, there have been many attempts to improve the efficiency of standard RANSAC for noisy data [21–25, 20]. Though efficient, most of these methods are limited in their handling of large number of outliers. For example, [20] describes an algorithm that can handle 70% of outliers for fundamental matrix estimation whereas [25] reports handling of $\approx 85\%$ outliers on simple line fitting examples. To account for large fraction of outliers in the data that is typical in multi-sensor alignment scenarios, the use of tracking information (instead of detections) was proposed in [10]. The method involved assigning a heuristic likelihood to each pair of co-occurring trajectories in two cameras based on the number of objects in the scene, number of possible pairs at the time, and the matching probability and non-uniformly sampling pairs of trajectories in a RANSAC based frame-

work. While this method improves over [9], its over-reliance on tracking is a disadvantage in noisy data and scenes with high traffic densities. In [11], a density based approach is presented that does not rely on tracking data and is able to handle large fraction of outliers in the data. A joint probability density of co-occurring target locations is estimated by kernel density estimation and the modes are obtained using mean-shift. The modes are weighted using Pearson coefficient and are non-uniformly sampled using RANSAC. The use of modes as opposed to original data allows the algorithm to reject most of the outliers before the application of RANSAC and thus the algorithm can work on noisy scenarios. A method to detect and adapt to the changes in topology based on the conformity of modes to the estimated homography was also presented. The density estimation step in [11] assumes availability of long period of target observations as well as redundancy in the observations. However, there are many practical scenarios where one or both of these assumptions do not hold, for example, obtaining redundant and consistent data is difficult in scenarios where the fields of view span wide areas or drift rather frequently.

The algorithm proposed here does not make these assumptions and can handle large fraction of outliers in small data sets by exploiting the geometric invariant properties to constrain the RANSAC sampling. As opposed to [21–25, 20] that attempt to improve the RANSAC algorithm for generic model estimation problems, the proposed algorithm utilizes model-specific knowledge that allows it to handle data with more than 95% of outliers. The efficiency of the proposed algorithm as well as minimal data requirement also enables adaptivity to the changes in network topology.

3 Using Model Specific Properties for Efficient Sampling

In this section, we present details of how model-specific properties can be chosen to find efficient sampling in problems involving model estimation in the presence of outliers.

Let $\mathbf{Z}$ be a set of observations and M be the model that we want to estimate from $\mathbf{Z}$. For each $z \in \mathbf{Z}$, let $\mathbf{I}_z$ be a binary indicator that is 1 if and only if z is an inlier and let $\mathbf{O}_z = 1 - \mathbf{I}_z$. The RANSAC based approaches (as in [9]) iteratively generate random samples from the set $\mathbf{Z}$, estimate a model M (such as homography) from the sample, and evaluate the model over the entire data set until a model is found that fits most of the elements of $\mathbf{Z}$ (defined by the probability of noise) or if a predefined maximum number of iterations is reached.

Using Bayes Theorem, the probability of a randomly sampling an inlier, s, from the set $\mathbf{Z}$ can be written as:

$$P(\mathbf{I}_s = 1) \propto P(\mathbf{I}_s = 1|s)\, P(s) \tag{1}$$

For example, for a sample drawn from uniform distribution (as in standard RANSAC), $P(\mathbf{I}_s = 1) = \sum_{z \in \mathbf{Z}} \mathbf{I}_z / |\mathbf{Z}|$. For simplicity of notation, we will omit the value of binary random variables in the rest of the paper and it will be assume to be 1 unless specified otherwise. Given the minimal number of

data points (samples), m, needed to estimate the model M and the probability of randomly picking an inlier, $P(\mathbf{I}_s)$, the required number of iterations, J, to obtain an outlier-free set of m samples with a probability ρ is given as $J = \lceil \ln(1-\rho)/\ln(1-P(\mathbf{I}_s)^m) \rceil$ [20]. Note that J serves as a lower bound on the number of iterations and has been found to be quite loose in practice where the required number of iterations to estimate a good model is generally much larger than J [21–23].

It is easy to see from the above analysis that improving efficiency of a RANSAC based algorithm in noisy scenarios requires finding a sampling strategy for which $P(\mathbf{I}_s)$ is better than the uniform sampling. This approach is adapted by [22], [23], and [11]. Wherein [22] and [23], the prior term in Equation (1) is modified to improve the probability with former assuming that the inliers are spatially closer to one another than the outliers and the later using likelihood of matches to define the sampling strategy. On the other hand, [11] assumes redundancy in data to find a smaller subset with improved likelihood term, i.e., $P(\mathbf{I}_s|s)$. In this paper, we define a sampling strategy that exploits the model-specific properties to improve the likelihood term.

Let $\mathbf{Z}^l$ be the set of all order l subsets of the elements of $\mathbf{Z}$. For each $z^l \in \mathbf{Z}^l$, let $\mathbf{I}_{z^l} = \prod_{z \in z^l} \mathbf{I}_z$ and $\mathbf{O}_{z^l} = 1 - \mathbf{I}_{z^l}$. Let Q be a property defined on $\mathbf{Z}^l$ and $\mathbf{Q}(z^l)$ be a binary variable that is true if and only if z^l satisfies the property Q. Further let $\mathbf{Z}^l(Q) \subseteq \mathbf{Z}^l$ be the set of all elements z^l of $\mathbf{Z}^l$ that satisfy Q, i.e., $\mathbf{Z}^l(Q) = \left\{ z^l \in \mathbf{Z}^l | \mathbf{Q}(z^l) = 1 \right\}$. From Equation (1) and by the definition of the set $\mathbf{Z}^l(Q)$, the probability of randomly picking an outlier-free set, s^l, from the set $\mathbf{Z}^l(Q)$ can be written as:

$$P(\mathbf{I}_{s^l}) \propto P\left(\mathbf{I}_{s^l} | \mathbf{Q}(s^l)\right) P\left(s^l\right)$$

where

$$\begin{aligned} P\left(\mathbf{I}_{s^l} | \mathbf{Q}(s^l)\right) &= \frac{P(\mathbf{I}_{s^l}) P\left(\mathbf{Q}(s^l) | \mathbf{I}_{s^l}\right)}{P(\mathbf{I}_{s^l}) P(\mathbf{Q}(s^l) | \mathbf{I}_{s^l}) + P(\mathbf{O}_{s^l}) P(\mathbf{Q}(s^l) | \mathbf{O}_{s^l})} \\ &= \frac{P(\mathbf{I}_s)^l}{P(\mathbf{I}_s)^l + \left(1 - P(\mathbf{I}_s)^l\right) \frac{P\left(\mathbf{Q}(s^l) | \mathbf{O}_{s^l}\right)}{P\left(\mathbf{Q}(s^l) | \mathbf{I}_{s^l}\right)}} \end{aligned} \tag{2}$$

If m_l is the minimal number of elements of $\mathbf{Z}^l$, needed to estimate the model M, the lower bound on the number of iterations to obtain an outlier-free set with the probability ρ is given as $J(l,Q) = \lceil \ln(1-\rho)/\ln(1-P(\mathbf{I}_{s^l})^{m_l}) \rceil$. For a given model, an optimal sampling strategy is the one that minimizes $J(l,Q)$ over all l and Q. In practice, any l and Q can be chosen for which $J >> J(l,Q)$, i.e., $m_l \log P(\mathbf{I}_{s^l}) >> m \log P(\mathbf{I}_s)$. This observation along with Equation (2) imply that the property Q should be chosen such that $P\left(\mathbf{Q}(s^l) | \mathbf{I}_{s^l}\right) >> P\left(\mathbf{Q}(s^l) | \mathbf{O}_{s^l}\right)$. In the next section, we provide the data-driven homography estimation prob-

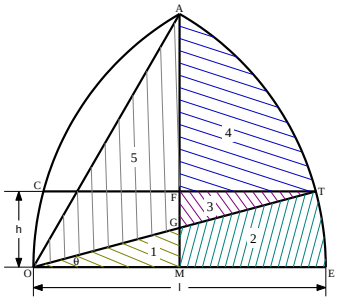

Fig. 1. The diagram illustrates estimation of the probability of a set of uniformly sampled random three points to be collinear. A is the point of intersection of two circles with radius $|OE|$ and centered at O and E respectively

lem as an example of finding efficient sampling strategy using model-specific properties.

4 Efficient Homography Estimation using Collinearity Constraint

The data driven homography estimation problem mentioned in the Section 1 can be formally defined as follows: Let $\mathbf{p}_i$ and $\mathbf{p}_j$ be the corresponding observations in two sensors C_i and C_j in homogeneous coordinate system. Assuming planarity of activity area in the scene (a typical assumption in surveillance scenarios), the geometric relationship between the observations is defined by a 3x3 linear transformation $\mathbf{H}_{ij}$ (homography), i.e., $\mathbf{p}_i = \mathbf{H}_{ij}\mathbf{p}_j$. Let $\mathbf{Z}_i$ be the set of observations of the sensor C_i, where $\mathbf{Z}_i = \bigcup_t \mathbf{Z}_i(t)$, $\mathbf{Z}_i(t) = \left\{ Z_i^1(t), Z_i^2(t), \ldots, Z_i^{k(t)}(t) \right\}$, and $\mathbf{Z}_i(t)$ is the set of $k(t)$ observations in the sensor C_i at time t. The set of all pairs of co-occurring observations between a given sensor pair (C_i, C_j) is defined as $\mathbf{Z}_{ij} = \bigcup_t \mathbf{Z}_i(t) \times \mathbf{Z}_j(t)$. A pair of observations $z_{ij} = \left(Z_i^a(t_k), Z_j^b(t_k) \right) \in \mathbf{Z}_{ij}$ is called an inlier if both observations $Z_i^a(t_k)$ and $Z_i^b(t_k)$ belong to the same target in the world, z_{ij} is called an outlier otherwise. For a given pair of sensors (C_i, C_j), the homography $\mathbf{H}_{ij}$ can be estimated using standard RANSAC by uniformly sampling four pairs at a time from $\mathbf{Z}_{ij}$. We will refer to this algorithm as RANSAC4 in the rest of this paper. If ν denotes the fraction of inliers in $\mathbf{Z}_{ij}$, i.e., $P(\mathbf{I}_s) = \nu$, the lower bound on the number of iterations needed to find an outlier-free set of samples with probability ρ is given as: $J_{RANSAC4} = \left\lceil ln\left(1-\rho\right) / ln\left(1-\nu^4\right) \right\rceil$

Consider a set $\mathbf{Z}_{ij}^3$ of all subsets of order 3 (triplets) of $\mathbf{Z}_{ij}$. We note that collinearity is invariant under homography transformation, i.e., the corresponding points of collinear points in one sensor are also collinear. Lets call this property (invariance of collinearity), Q_H, and define $\mathbf{Z}_{ij}^3(Q_H)$ to be the set of all triplets for which the points in both the sensors are collinear. Note that for each

Table 1. Lower Bounds on the Number of RANSAC Iterations.

Fraction of Inliers ν	2.5%	5%	10%	20%	25%
$J_{RANSAC4}$	17683850	1105237	69074	4313	1764
J_{CONSAC}	13200922	206261	3219	46	9

triplet $T \in \mathbf{Z}_{ij}^3$, $P\left(\mathbf{Q}_H(T)|\mathbf{I}_T\right) = 1$. Hence, from Equation (2), the probability of uniformly sampling an outlier-free triplet, T, from the set $\mathbf{Z}_{ij}^3(Q_H)$ is given by: $P\left(\mathbf{I}_T\right) = \nu^3 / \left[\nu^3 + \left(1-\nu^3\right) P\left(\mathbf{Q}_H(T)|\mathbf{O}_T\right)\right]$.

The term $P\left(\mathbf{Q}_H(T)|\mathbf{O}_T\right)$ in the equation represents the probability of the triplets composed of one or more outliers to be collinear. This can be estimated by finding the probability of three random points being collinear. Note that in practice even three inlier points may not be exactly collinear due to noise or sensor's limited resolution. Thus, we define three points to be collinear if the ratio between the shortest height and the longest edge of the triangle formed by the points is less than a suitably defined threshold γ. Assume without loss of any generality that OE represents the largest side of the triangle (Figure 1). Thus, the third point must lie within the region OAE (otherwise OE will not be the largest side of the resulting triangle). Given γ, the shortest height allowed by the model is defined by $h = \gamma|OE|$. Let C be a point on the arc OA such that the distance $d(OE, C) = h$, then the third point must lie within the region $OCTE$. Thus from Figure 1, the probability of three random points being collinear can be estimated as $\left(\Delta OTE - \Delta OGM + \Delta GFT\right) / \left(\Delta OAE - \Delta OAM\right) \approx \gamma / \left(\pi/3 - \sqrt{3}/4\right)$, where ΔX represents the area of the region X. Also note that the order of a triplet should remain the same under homography, i.e., the point in the middle must remain in the middle after transformation. Hence, the probability of collinearity given a random triplet is given by $P\left(\mathbf{Q}_H(T)|\mathbf{O}_T\right) \approx \gamma / \left[3\left(\pi/3 - \sqrt{3}/4\right)\right] = 0.54\gamma$. Thus, $P\left(\mathbf{I}_T\right) = \nu^3 / \left[\nu^3 + 0.54\gamma\left(1-\nu^3\right)\right]$.

Note that a homography can be estimated by using only two triplets (6 pairs of points). Thus the modified RANSAC, called Constrained Random Sample Consensus (CONSAC) in the rest of the paper, uniformly samples 2 triplets in each iteration from the set $\mathbf{Z}_{ij}(3, Q_H)$ and the lower bound J_{CONSAC} is given as: $J_{CONSAC} = \left\lceil ln\left(1-\rho\right) / ln\left(1 - P\left(\mathbf{I}_T\right)^2\right)\right\rceil$. It is easy to show that $J_{CONSAC} < J_{RANSAC4}$ for all $P((I)_s) = \nu > 0.54\gamma$. In our implementation, γ is chosen to be 0.04. For this value of γ, the proposed approach will perform better as long as $\nu > 0.02$, i.e., there are at least 2% of inliers in the data. The values of J_{CONSAC} and $J_{RANSAC4}$ for different values of ν (with $\rho = 0.999$ are shown in Table 1. It can be seen that the theoretical bounds for the triplet based algorithm are order of magnitude better than the standard algorithm. The quantitative results in Section 5 show that the practical impact of the proposed method is even more significant.

5 Results and Applications

In this section, we present quantitative and qualitative results of the proposed algorithm on synthetic data as well as real data from multi-sensor alignment applications. In addition to planar homography, we also model and estimate lens distortion for different applications. The details of lens distortion estimation are omitted due to space considerations.

5.1 Quantitative Evaluation

We use synthetic data with varying amount of fraction of outliers, variance of spatial noise, and traffic density to evaluate the proposed approach and compare it with standard RANSAC. The plots in Figure 2 show the comparisons of the execution speed and the estimation error of both algorithms with respect to these parameters. It can be seen from the figure that the proposed algorithm is: i) significantly efficient than the standard RANSAC, ii) able to handle large fraction of outliers in the data, and iii) robust to traffic densities and spatial noise.

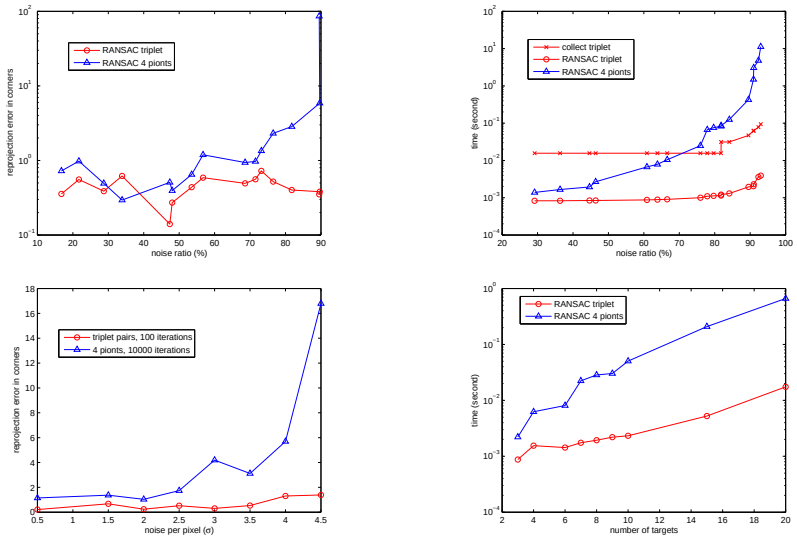

Fig. 2. Comparison of speed and estimation error of the proposed algorithm with the standard RANSAC based approach with respect to the fraction of outliers (top row), variance of spatial noise (bottom left), and traffic density (bottom right)

5.2 Semi-supervised Sensor Alignment

In scenarios where the terrain does not have many well localized and/or discriminative features, even manual sensor alignment or geo-registration is a challenging task. One of the solutions to this problem is use of dedicated mobile units

(people, vehicles, boats, etc.) that report their geo-location while moving in the sensor's field of view. A real-world system utilizing this approach for establishing and maintaining geo-registration in a large visual sensor network is described in [6]. Note that in an uncontrolled environment, neither all targets in the scene transmit geo-positional data nor do all targets transmitting this data have to be visible in one or more sensors. Therefore one-to-one correspondences between target metadata and geo-positional data is not available. As a matter of fact, in our experience, accidental correspondences (noise) significantly out-number the true correspondences (typically noisy data is more than 85% of the total data).

In our experiments, we use a GPS carrier that follows a pre-specified pattern. This provides additional domain-specific constraints (priors) that can be utilize to improve the efficiency of the proposed sampling scheme for matching view detections with the geo-locations of the controlled carrier and estimating geo-registration. Moving direction is one such constraint. When a target moves clockwise around the camera, its corresponding image moves from left to the right in the view and vice-versa. To best utilize the moving direction constraint, motion toward or away from the camera should be eliminated. Thus, the GPS carrier should move horizontally across the FOV of the camera. To form a continuous route, it should follow a zigzag curve (Fig. 3). Under this condition, half of the outlier view points can be eliminated using the direction of motion constraint. Note that the moving direction constraint is useful even for the traditional four-point RANSAC algorithm. Horizontal motion across the FOV of the camera provides additional benefits for the triplet-based algorithm. As direction of motion constraint is most effective when the GPS carrier moves horizontally across the FOV of the camera, the traces of GPS carrier are designed to be horizontal. Therefore, vertical straight lines are drawn to increase the chance to intersect them to obtain possible collinear inlier triplets (Fig 4). A vertical line is expected to intersect horizontal GPS traces with higher probability than non-horizontal target trace of any outlier target. Assuming the direction of outlier target traces to be uniformly distributed, the effective mean horizontal projection of the outlier traces is $\left(\int_0^{\pi/2} \cos\theta d\theta\right) / \left(\int_0^{\pi/2} d\theta\right) = 2/\pi$. Therefore, the proposed approach has $\pi/2$ higher chance of obtaining an inlier triplet point than an outlier triplet point due to the horizontal motion of the GPS carrier. Fig. 3 shows the result of the proposed approach to geo-register four camera views at a port. The estimated field-of-views are overlaid on a satellite snapshot of the port. The maximum error is around 15 meters at a distance of 1500 meters from the cameras. We observe that the geo-registration errors can mostly be attributed to the inaccurate localization of the targets by the tracking algorithm.

5.3 Unsupervised Sensor Alignment

We now show the robustness of the proposed approach in unsupervised scenarios in which alignment is done using completely uncontrolled targets. The example of such a scenario includes utilizing ships and boats that broadcast automatic identification system (AIS) information (in some cases mandated by port au-

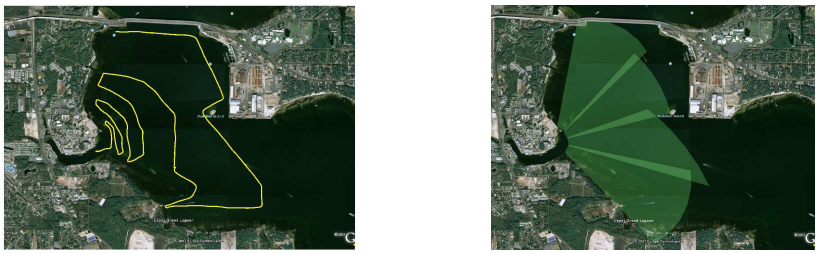

Fig. 3. Semi-supervised Sensor Alignment: (Left) A typical GPS track designed to be horizontal for better alignment. (Right) The field-of-views of four cameras obtained using the proposed approach at a port.

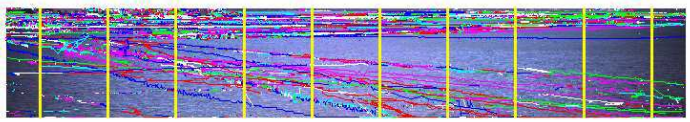

Fig. 4. A snapshot of the camera view with color-coded target tracks overlaid with a set of vertical line segments used during matching in the proposed approach.

thorities), RADAR bearing and distance observations of targets at a port, and vehicles and people with RFID tags [6]. Multi-sensor alignment with overlapping fields of view using target observations is another example of unsupervised sensor alignment [11]. For our experiments, we used RADAR information to calibrate hi-resolution cameras installed at the port. Fig. 5 (left) shows the color-coded RADAR tracks. The field-of-views of cameras estimated using the proposed alignment approach is shown on the right. As with semi-supervised alignment, the maximum error is around 15 meters at a distance of 1500 meters from the cameras. Figure 6 shows an alignment of two visual sensors obtained by the proposed approach based on their target observations during a time span of 5 minutes. The estimated homography is consistent with the one reported in [11] using target observations from hour-long clips of the same video sequences as well as the manually obtained results by visually matching feature points in both the images.

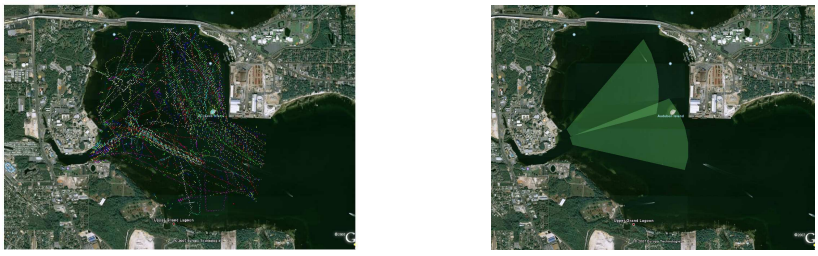

Fig. 5. Unsupervised Sensor Alignment: (Left) Color coded RADAR tracks at a port. (Right) The field-of-views of two high resolution cameras obtained using the proposed approach.

Fig. 6. Unsupervised alignment of visual sensors with overlapping fields of view: (Top Row) Camera views, (Bottom Left) Camera Setup (Bottom Right) Warped Image using the homography estimated from only 5 minutes of target observations.

6 Conclusion

An efficient and robust solution for data-driven multi-sensor alignment is proposed. The proposed solution exploits model-specific constraints, in particular, invariance of collinearity under homography transformation to speed-up the model estimation in the presence of outliers. The robustness and efficiency of the proposed solution in scenarios involving large fraction of outliers, spatial noise, and varying traffic densities is shown using both real and synthetic data.

References

1. Collins, R., Lipton, A., Fujiyoshi, H., Kanade, T.: Algorithms for cooperative multisensor surveillance. Proceedings of the IEEE **89** (2001)
2. Espina, M., Velastin, S.: Intelligent distributed surveillance systems: A review. IEE Proceedings - Vision, Image, and Signal Processing **152** (2005) 192–204
3. Shah, M., Javed, O., Shafique, K.: Automated visual surveillance in realistic scenarios. IEEE Multimedia **14** (2007)

4. Cucchiara, R., Prati, A., Vezzani, R., Benini, L., Farella, E., Zappi, P.: An integrated multi-modal sensor network for video surveillance. Journal of Ubiquitous Computing and Intelligence (2005)
5. Taj, M., Cavallaro, A.: Multi-camera scene analysis using an object-centric continuous distrubtion hidden markov model. In: ICIP. (2007)
6. Rasheed, Z., Cao, X., Shafique, K., Liu, H., Yu, L., Lee, M., Ramnath, K., Choe, T., Javed, O., Haering, N.: Automated visual analysis in large scale sensor networks. In: ACM/IEEE International Conference on Distributed Smart Cameras. (2008)
7. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. (Cambridge University Press, 2000)
8. F. Lv, T.Z., Nevatia, R.: Self-calibration of a camera from video of a walking human. In: International Conference on Pattern Recognition. (2002)
9. Lee, L., Romano, R., Stein, G.: Monitoring activities from multiple video streams: Establishing a common coordinate frame. IEEE Trans. on PAMI (2000)
10. Stauffer, C., Tieu, K.: Automated multi-camera planar tracking correspondence modeling. In: IEEE CVPR. (2003)
11. Shafique, K., Hakeem, A., Javed, O., Haering, N.: Self calibrating visual sensor networks. In: IEEE Workshop on Applications of Computer Vision. (2008)
12. Fischler, M.A., Bolles, R.C.: Ransac: A paradigm for model fitting with applications to image analysis and automated cartography. In: ECCV. (1996)
13. Kettnaker, V., Zabih, R.: Bayesian multi-camera surveillance. In: International Conference on Computer Vision and Pattern Recognition. (1999)
14. Khan, S., Javed, O., Rasheed, Z., Shah, M.: Human tracking in multiple cameras. In: International Conference on Computer Vision. (2001)
15. Rahimi, A., Dunagan, B., Darrell, T.: Simultaneous calibration and tracking with a network of non-overlapping sensors. In: IEEE CVPR. (2004)
16. Javed, O., Shafique, K., Rasheed, Z., Shah, M.: Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. Computer Vision and Image Understanding Journal **109** (2008)
17. Makris, D., Ellis, T., Black, J.: Bridging the gaps between cameras. In: International Conference on Computer Vision and Pattern Recognition. (2004)
18. Stauffer, C.: Learning to track objects through unobserved regions. In: IEEE Workshop on Motion. (2005)
19. Tieu, K., Dalley, G., Grimson, W.: Inference of non-overlapping camera network topology by measuring statistical dependence. In: IEEE ICCV. (2005)
20. Zhang, W., Kosecka, J.: A new inlier identification procedure for robust estimation problems. In: Robotics: Science and Systems Conference. (2006)
21. Matas, J., Chum, O.: Randomized ransac with t(d,d) test. In: BMVC. (2002)
22. Myatt, D., Torr, P., Nasuto, S., Craddock, R.: Napsac: High noise, high dimensional robust estimation - its in the bag. In: British Machine Vision Conference. (2002)
23. Tordoff, B., Murray, D.: Guided sampling and consensus for motion estimation. In: European Conference on Computer vision. (2002)
24. Nister, D.: Preemptive ransac for live structure and motion estimation. In: International Conference on Computer Vision. (2003)
25. Wang, H., Suter, D.: Robust adaptive-scale parametric model estimation for computer vision. IEEE Trans. on PAMI **26** (2004)