Multi-camera localisation: a review

Rahul Raman*, Sambit Bakshi and Pankaj K. Sa

Department of Computer Science and Engineering,
National Institute of Technology Rourkela,
Rourkela – 769008, Odisha, India
E-mail: rahulraman2@gmail.com
E-mail: sambitbaksi@gmail.com
E-mail: pankajksa@gmail.com
*Corresponding author

Abstract: Past few years have seen exhaustive research in the field of camera localisation. As the era changed from single to multiple cameras, so is the paradigm shift from centralised to distributed algorithms. Euclidean geometry has been explored and the concept of Lie algebra has been touched for more subtle and non-Euclidean details. View overlaps in vision-based algorithms have been optimised and several depth measurement techniques have been implemented to extend the localisation from 2D to 3D space. LED-based techniques like triangulation and LED triangle have given depth measurement alternatives for 3D localisation whereas epipolar geometry has localised cameras with only image information. Multilateration-based approach has used anchor nodes for camera localisation whereas a few distributed algorithms (viz. DALT, DILOC) have used iterations for refinement of estimated locations. As the area under cover increased, wireless network has taken over and many algorithms have been developed for wireless networked cameras. Simultaneous existences of diverse algorithms belonging to different paradigms are needed to meet the requirement of deployment in diverse scenarios. This paper discusses the evolution from the localisation of non-camera equipped sensor network to the smart camera localisation in 3D environment that spans more than a decade.

Keywords: multi-camera; localisation; camera sensor networks; surveillance; vision graph; belief propagation.

Reference to this paper should be made as follows: Raman, R., Bakshi, S. and Sa, P.K. (2013) 'Multi-camera localisation: a review', *Int. J. Machine Intelligence and Sensory Signal Processing*, Vol. 1, No. 1, pp.91–109.

Biographical notes: Rahul Raman is currently pursuing his Masters research in Department of Computer Science and Engineering at National Institute of Technology, Rourkela, India. His research interests include algorithms for multi-camera network, video surveillance, and biometric security.

Sambit Bakshi received his Masters from National Institute of Technology, Rourkela, India in 2011. Currently, he is pursuing his Doctoral research at National Institute of Technology, Rourkela. His research interests include biometric authentication through non-cooperative images and videos.

Pankaj K. Sa earned his Masters and Doctoral from the National Institute of Technology, Rourkela, in the area of image processing. His research interests also include computer vision and computer graphics.

1 Introduction

The mode of vision-based security in the past decade has been centred with operation through single camera in small indoor environments. However, single camera surveillance gradually evolved to multi-camera sensor network due to following factors:

1 growing importance of video surveillance

2 environment of coverage areas becoming larger and more complex

3 decreased hardware cost of sensors.

For the operation of multi-camera sensor network, knowledge of location of other cameras is the pre-requisite for every camera. This process of establishing a relation among the camera coordinates is termed as camera localisation. Manual localisation methods of multi-camera network failed to handle large number of cameras in network. Automation of the localisation process started gaining importance to ascertain accuracy and real-time localisation. One of the primitive automated solutions to localisation has been through GPS (Hofmann-Wellenhof et al., 1997), but it has failed mostly due to the need of open environment having line-of-sight. Efforts have also been made towards developing localisation algorithms on single processor after collecting images from all the networked cameras in a single room (Davis et al., 1999; Kanade et al., 1997). But in practical scenario, large number of cameras producing high volume of images and video data makes the analysis time-consuming on single processor. The subsequent attempts of developing localisation algorithms deploy more than one processor concurrently to achieve real-time localisation. These approaches differ in variety of coverage areas, assumptions made on deployment of the nodes, and the way sensors work (Piovan et al., 2008).

The paper is organised as follows: Section 2 reveals the need and evolution of multi-camera network as an independent field of research followed by Section 3 describing main techniques that localised cameras based on vision captured from a camera. This section also describes formation of epipolar geometry. GPS- and LED-based techniques are discussed in this section and their bottlenecks are also discussed. Hence, evolved another genre of localisation techniques as illustrated in Section 4. Section 5 emphasises on recent need of localising wireless cameras and to solve the localisation in 3D plane. 3D localisation is more complex as it contains more number of unknown parameters. Latest techniques applied to partially solve 3D localisation are discussed concluding with a comparative analysis of existing works.

2 Pioneer works

Early automated localisation techniques for static sensors, viz. non-camera equipped networks have used ultra-sound, radio, or acoustic signals (Taylor et al., 2006). Likewise, moving sensors like robots have exploited LED-based techniques for their localisation. However all the methods proposed have been based on heuristic approaches and lagged theoretical foundation of network localisation until Aspnes et al. (2006) have identified specific problems and solved them theoretically. This work, motivated by previous work in Eren et al. (2004), has attempted to give systematic answer to the following questions:

1 conditions for unique network localisability

2 computational complexity of network localisation

3 complexity of localisation in typical network deployment scenario.

The authors have established the localisation problem in sparse graphs to be *NP*-hard unless $P = NP$. For dense graphs, localisation has been shown to be possible as explained by Biswas and Ye (2004).

The notion of centralised processing has been predominant in early camera sensor localisation techniques. Davis et al. (1999) have analysed human action in a closed environment. Stereoscopic reconstruction of virtual world based on depth calculation from multiple real scenes captured through multiple cameras has been attempted in Kanade et al. (1997). Aforementioned experiments revealed the importance of proper positioning and orientation of cameras for best coverage of view area. Various researches have attempted to solve the pose (location and orientation) (Funiak et al., 2006) of all cameras in the network. Funiak et al. (2006) have proposed a novel approach of relative over parameterisation (ROP) of the camera pose. However, some approaches have been successful to calculate relative locations only, but failed to estimate orientation of each camera. GPS-based approach (Hofmann-Wellenhof et al., 1997) have been successful in finding approximate relative location of cameras but the reasons of its failure were:

1 inability to resolve camera orientation

2 need of direct line-of-sight to satellites

3 costly hardware requirement

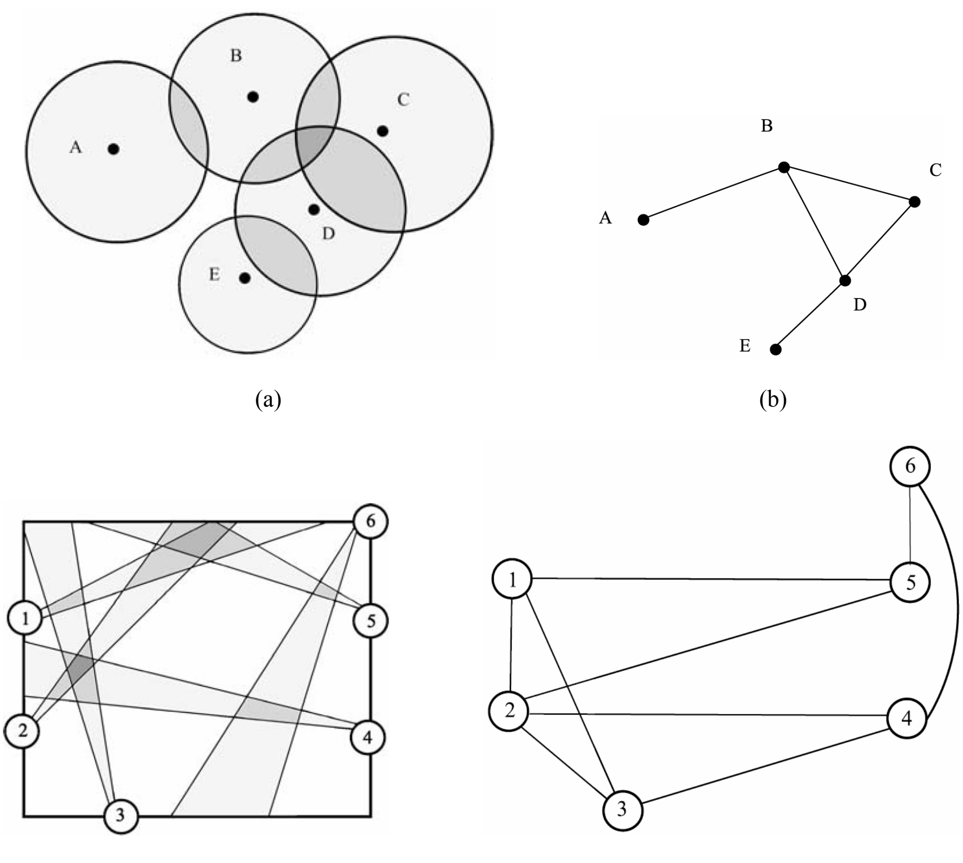
4 high power consumption.

Work in Liu et al. (2006) have presented a protocol that utilises GPS- and LED-based localisation. But this protocol needed human-assistance, which failed for large number of cameras deployed in a wide coverage area. Hence, several distributed computing algorithms (Mantzel et al., 2004; Funiak et al., 2006; Devarajan and Radke, 2007; Tron et al., 2008) have come into play to produce accurate and real-time localisation solution to large number of networked cameras.

3 Vision-based localisation

A stringent requirement of vision-based approach has been foreseen by the researchers as localisation through GPS was neither accurate nor able to provide orientation. The appeal of vision-based localisation is that it requires image data only. However, vision-based localisation algorithms impose a deployment constraint that there must be an overlap between view of cameras in the network. This constraint is analogous to the constraint in general transreceiver sensor network. Inspired by the graph theoretic representation (Bondy and Murty, 1982) of connectivity among sensors (as depicted in Figure 1), vision graph (Mantzel et al., 2004) with M networked cameras was introduced to be $G(V, E)$ defined on $V = \{V_i \mid i = 1,\dots,M\}$, and $E = \{E_{ij} \mid E_{ij} \in \{0, 1\}; i, j = 1,\dots,M\}$ representing cameras as vertices and vision overlap as edges respectively. Kurillo et al. (2008) introduced the concept of weighted vision graph, where each e_{ij} has been assigned a weight w_{ij} corresponding to the number of common points between i^{th} and j^{th} cameras. To serve the purpose of realigning all camera pose to a single network-wide coordinate frame, some researchers have come up with solutions that require triple-wise camera overlaps (Lowe, 2004; Mantzel et al., 2004), implying the need of densely deployed network, where as some researchers have proposed to position a camera in the network such that it is in view-overlap with all other cameras in the network (Lymberopoulos et al., 2005). Some researchers have used an LED-lit rod of known length to be placed at a position visible from all cameras to establish consistent scale (Medeiros et al., 2008; Kurillo et al., 2008). As the densely deployed network is not cost-optimised, researchers have come up with localisation solution for relatively sparsely deployed network (Kurillo et al., 2008; Ellis et al., 2003), and subsequently also for networks with non-overlap (Marinakis et al., 2005; Rahimi et al., 2004). The following Sections 3.1 and 3.2 explain visible and invisible LED-based techniques, and the formation of epipolar geometry behind resolving view-overlap respectively.

Figure 1 Analogy between formation of sensor connectivity graph and vision graph, (a) transreceiver range overlap of sensors (b) sensor connectivity graph (c) view overlap of networked cameras (d) vision graph



(a) (b)

(c) (d)

3.1 LED-based approaches to minimise view-overlap

Techniques based on LED (emitting visible or infrared spectrum) have reduced the view overlap leading to relatively sparsely deployed network. Use of LED reduces the view-overlap to be pair wise. A few recent works based on epipolar geometry have been done to reduce the density of overlap while maintaining the localisability of each camera.

In Medeiros et al. (2008) and Kurillo et al. (2008), two LED markers are placed on both ends of a fixed metal rod of known length. The time synchronised detection of LED provides correlated feature points. From the known length of the rod, unknown scale factor is resolved to consistent scale. Barton-Sweeney et al. (2006) and Farrell et al. (2007) have also exploited LEDs for modulated emission.

Depth measurement is required for 3D localisation. Since a camera cannot fetch depth information from a perspective view, hence an explicit distance measurement technique is essential. Lymberopoulos et al. (2005) have used three LED markers to form a triangle pattern to estimate distance measurement needed for 3D localisation. Barton-Sweeney et al. (2006) have experimentally verified that three LEDs in a triangle pattern with known dimensions can avoid explicit distance measurement, which had been in common practice before devising the said approach (Goldenberg et al., 2005; Sturm and Triggs, 1996). In Liu et al. (2006), global coordinates are taken from GPS-based calibration device for computing pose of camera, while image coordinates are calculated from LED of the camera.

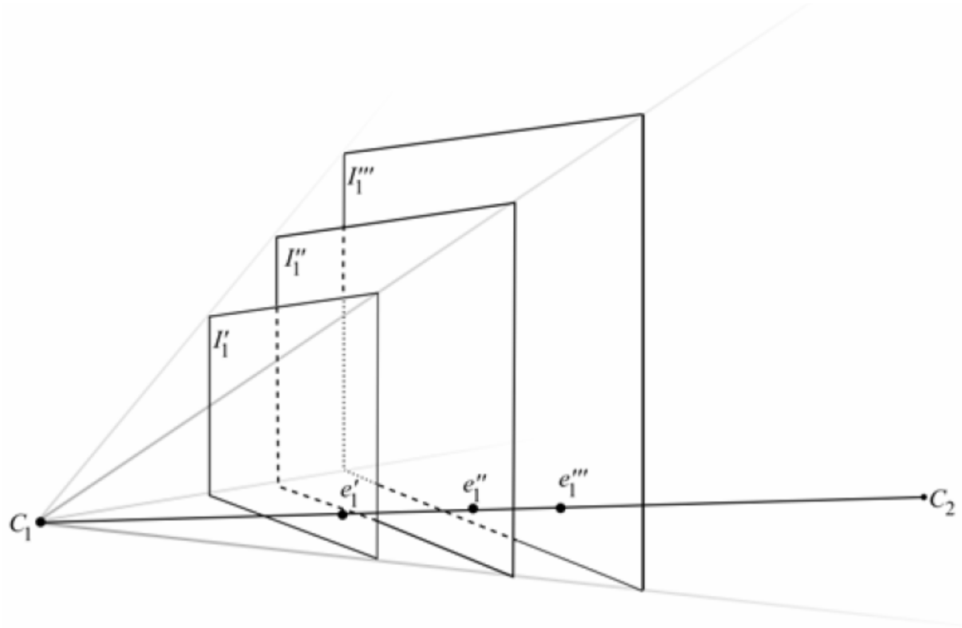
While most of the researches towards this direction employ visible LEDs to mark location and general cameras to sense the LEDs, techniques for localisation through invisible markers (sensed with IR sensors) also gained its importance as invisibility of markers do not impair the scenery. The invisible markers are made of translucent retro-reflectors which are visible only in IR illumination (Nakazato et al., 2005a). Localisation techniques through invisible markers are costlier than localisation through visible markers as they employ extra IR sensor along with general cameras that are intended to be localised (Nakazato et al., 2005b). Early invisible marker techniques have used infrared markers for estimating positions while orientations have been estimated through gyro meter only (Tenmoku et al., 2003; Maeda et al., 2004). However, later the known geometry of the invisible markers has been exploited to estimate both the position and orientation of the markers from its view projection (Kato and Billinghurst, 1999).

3.2 Epipolar geometry to resolve view-overlap

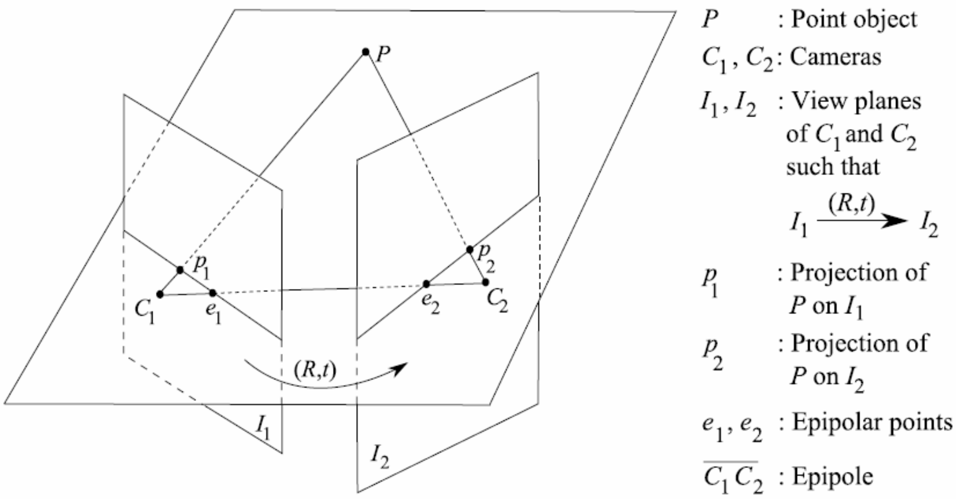
Epipolar geometry (Chum et al., 2003; Zhang, 1998) provides a 3×3 singular matrix describing the relation between two perspective images of the same rigid object from two cameras. Epipole is the line connecting any two cameras seeing the same object (depicted in Figure 2). The point where epipole meets the camera frame is epipolar point and hence epipole can also be realised as a collection of epipolar points between corresponding frames of two cameras [shown in Figure 2(a)].

Epipolar geometry has the basis that any object (in 3D coordinate) observed by two cameras and their projections are co-planar (Hartley and Zisserman, 2004) [shown in Figure 2(b)]. The essential matrix formulated from epipolar geometry is further used for localisation and camera calibration (Kurillo et al., 2008; Ma et al., 2004). Medeiros et al. (2008) and Kurillo et al. (2008) have employed epipolar geometry to resolve point correspondence problem (Liu et al., 1990) and unknown scale factor (Xu and Zhang, 1996).

Figure 2 Formation of epipolar geometry, (a) epipole as a collection of epipolar points (b) epipole and epipolar plane



(a)



(b)

In decentralised and distributed communication paradigm of multi-camera network, point correspondence problem can be solved through:

1 measurement correspondence (where features of an object seen from different cameras are wrapped into a common view prior to state estimation)

2 trajectory correspondence (where state estimates are computed independently in each view) (Spurlock and Souvenir, 2012; Taj and Cavallaro, 2011).

A recent work of Anjum (2011) exploits correspondence among trajectories estimated by Kalman filter to recover poses of non-overlapping cameras. Table 1 summarises few landmark researches towards solving point correspondence problem.

Table 1 Different approaches to solve point correspondence problem

Year	*Author*	*Approaches*
2004	Mantzel et al.	Time-synchronisation correlation of feature points (extracted by tracked motion)
2005	Lymberopoulos et al.	Deploying nodes with self-identifying lights (fails in bright or specular-filled environment)
2006	Devarajan et al.	Scale invariant feature transform (SIFT)-based feature point correlation
2008	Medeiros et al.	Time-synchronisation correlation of feature points (using LED rod) + recursion on fundamental matrix to refine camera positions
2008	Kurillo et al.	Time-synchronisation correlation of feature points (using LED rod) + bundle adjustment (Triggs et al., 1999) to refine camera positions
2010	Kassebaum et al.	3D target of known geometry and pairwise projection matrix estimation for point correspondence

Kurillo et al. (2008) have used it for camera position and orientation. Researchers in Mantzel et al. (2004) and Bulusu et al. (2000) have also used epipolar geometry for camera localisation. Lymberopoulos et al. (2005) have proposed sensor assisted camera localisation and have examined measured epipoles (ME) (Taylor, 2004) and estimated epipoles (EE) (Hartley and Zisserman, 2004). They have also formulated a more constrained optimisation problem [optimised estimated epipole (OEE)] to reduce the error in noisy EE.

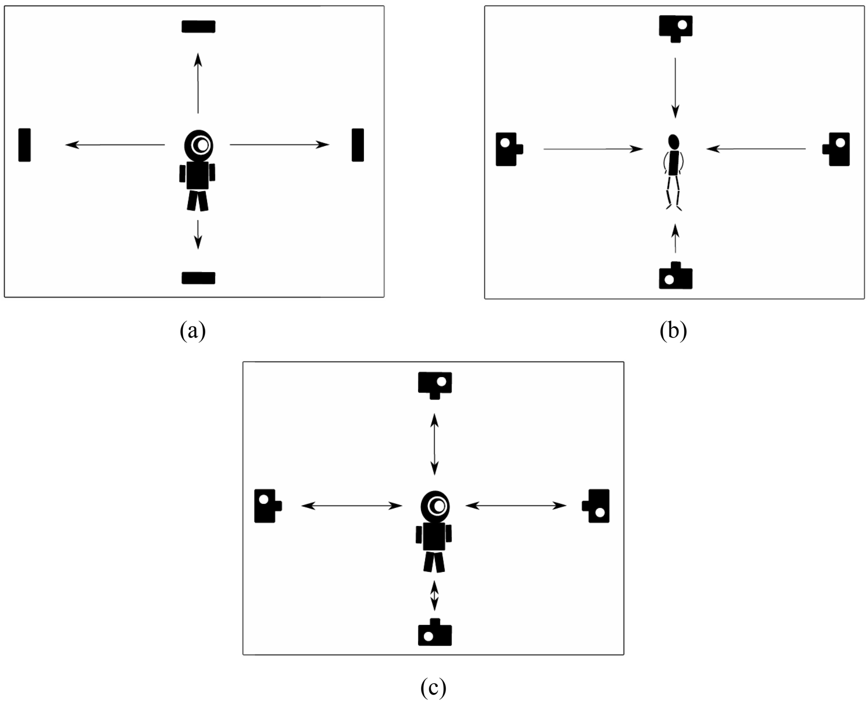
4 Consensus and belief propagation-based localisation

A consensus algorithm is an interaction rule that specifies the information exchange between an agent and all of its neighbours on the network. Consensus algorithms are used in many situations, viz. distributed formation control, synchronisation, rendezvous in space, distributed fusion in sensor, flocking theory (Olfati-Saber et al., 2007).

Consensus algorithms are used for getting global pose of a camera in a network, and have been used for localisation with range measurements (Gotsman and Koren, 2004; Khan et al., 2009). Tron and Vidal (2009) have generalised the consensus algorithm for estimating pose of each node from noisy and inconsistent measurements.

On contrary to this, notion of belief propagation have also been proposed for establishing localisation (Devarajan and Radke, 2007). Belief propagation is a message passing technique for graphical network model which have been applied for scene estimation, shape finding, image segmentation, restoration, and tracking (Isard and Blake, 1998; Freeman and Pasztor, 1999; Coughlan and Ferreira 2002; Felzenszwalb and Huttenlocher, 2004; Sudderth et al., 2005). Belief propagation has originally been developed for trees. When applied for graphs with cycles, inferences (belief) might not converge, and even if convergence occurs, density is not guaranteed (Murphy et al., 1999; Pearl, 1988). The non-convergent form of belief propagation [loopy belief propagation (LBP)] (Murphy et al., 1999) is used in sharing localisation parameters in multi-camera localisation.

Figure 3 Simultaneous localisation techniques, (a) SLAM (b) SLAT (c) SPLAM



Paskin and Guestrin (2004) have presented a more robust algorithm than BP in several aspects. This approach has been extended by researchers in Dellaert et al. (2005) for localisation of robot in multi-camera scenario (SLAM: simultaneous localisation and mapping) (Paskin, 2003) where a robot observes all the landmarks and estimates its location and position of the landmarks. A similar concept has been proposed by Funiak et al. (2006) for camera localisation (SLAT: simultaneous localisation and tracking), where the camera replaces the landmarks and robot is replaced by a moving object. Robot observes the landmarks in SLAM [shown in Figure 3(a)], whereas cameras observe the object in SLAT [shown in Figure 3(b)]. Funiak et al. (2006) has also proposed ROP to represent the distribution in SLAT problem using single Gaussian.

There had been efforts to find the trajectory of object and pose of camera simultaneously (Funiak et al., 2006; Lee and Aghajan, 2005). In particular, Rekleitis et al. (2006) have addressed the issue of localisation in hybrid context of robot-camera network system, where object localisation takes place along with camera localisation (SPLAM: simultaneous planning localisation and mapping) [shown in Figure 3(c)]. Here, robot can localise itself treating cameras as its landmarks (similar to SLAM). Likewise, cameras can localise themselves treating the robot as moving object (similar to SLAT). Estimation, local planned behaviour, and data fusion are done for effective collaboration of camera network and robot in SPLAM.

5 Wireless and 3D localisation

With increasing coverage area and no. of cameras in a network, wireless mode of communication has grown its significance. Even though much work has been done over wireless sensor network, their localisation algorithms (Savvides et al., 2001; Savarese et al., 2001; Capkun et al., 2002; Galstyan et al., 2004; Moses et al., 2003; Patwari et al., 2003; Shang et al., 2003; Patwari and Hero, 2004) do not hold good for camera network due to two main reasons:

1 they do not achieve required accuracy for computer vision tasks

2 they do not provide orientation of a sensor.

Lee and Aghajan (2005) have proposed a wireless camera (connected by IEEE 208.11b protocol) localisation algorithm capable of estimating both camera pose and trajectory of the object. This work has been experimented in 2D plane with only five cameras, while Medeiros et al. (2008) have proposed four different localisation approaches simulated in a $20 \times 20 \times 20$ m^3 3D region with 50 randomly placed cameras. The system developed in Medeiros et al. (2008) can perform in fully-distributed scenario, and does not require anchor-nodes. This approach employs feature-based object trajectory estimation, and hence performs depending on robustness of the used feature-extraction algorithm.

3D image reconstruction has remained an active research area in computer vision for many years. Tomassi and Kanade (1992) have used matrix factorisation as a way for reconstructing a scene, as well as to estimate camera parameters and frame point localisation. This work has employed orthographic projection whereas Poelman et al. (1997) have used perspective projection to serve the same. Sturm and Triggs (1996) has also proposed more complete solution for measuring camera depth. Rahimi et al. (2004) have pre-computed the homographies between image plane of each camera, and a common ground plane leading to 3D localisation of cameras.

Lymberopoulos et al. (2005) have proposed an algorithm that combines a sparse set of distance measurements with image information from image planes of each camera. It uses three LED triangle of known geometry for depth measurement. Tron and Vidal (2009) have taken the work to distributed level, they have applied distributed consensus algorithm that enhances the work of Piovan et al. (2008) and have generalised it from 2D to 3D.

Latest works on 3D camera localisation include the work of Kassebaum et al. (2010). Kassebaum et al. (2010) have used 3D target. This is similar to the 2D targets like checker boards used earlier in Zhang (2000) and Heikkila and Silven (1997). The advantage of 3D target is that in one frame it provides all the feature points needed by a camera to determine its position and orientation relative to the target. On detected feature points, DLT (Hartley and Zisserman, 2000) is used to estimate projection matrix. The algorithm reduces the cost of feature point detection, number of overlaps and eliminates the unknown scale factor problem. Kassebaum et al. (2010) have experimented with error less than 1in when 3D target feature point fills only 2.9% of the frame.

Table 2 Review of related researches on multi-camera localisation

Approaches	*Algorithm*	*Assumption*	*Experimental setup*	*Constraints*	*Results*
Mantzel et al. (2004)	DALT (localisation through triangulation, refinement through retriangulation of 3D points through iterations	Assumes at least two or more cameras to be prelocalised	Not experimented practically; simulated using 20 actual views of checker board pattern with 156 corners (as feature points)	Each camera linked to eight to 16 other cameras; cameras were prelocalised	0.25% of planarity error; 14 mm error in 3 m scale
Lymberopoulos et al. (2005)	Pairwise view overlap and epipolar geometry based estimation; ME and EE are evaluated to propose OEE; refinement through iteration	Coordinate transformations to distribute rotation and transformation between camera pairs	Indoor setup: two camera, 16 non-camera nodes; outdoor setup: 80 nodes. Each camera node consists of COTS OV7649 camera module having motion detection and LED identification; all nodes carry Lumex CCI-CRS10SR omnidirectional LED	Resolutions used: 640 × 480 (VGA), 352 × 288 (CIF), 240 × 180, and 128 × 96 (SQCIF); cameras can observe LEDs up to 4 m.	Indoor experiment: error of 2–7 cm in a 6 × 6 m^2 room; outdoor experiment: error of 20–80 cm in an area of 30 × 30 m^2; maximum error at lowest resolution is 3.32 cm
Funiak et al. (2006)	Complex distribution of SLAT is represented using novel approach of single Gaussian model ROP; quality of the solution is represented explicitly by uncertainty in estimate of camera poses	Out of three position parameters and three angles, paper focuses on three parameters (x, y, θ) assuming rest to be known	Simulated in square area with 44 side-facing cameras tilted down about 35° and 50 downward-facing cameras with pose estimation within 95% confidence intervals; Experimented practically in real network of 25 overhead cameras, and a remote controlled toy-car carrying a colour marker moving around.	The subject is made to move in a circular path within the square area	Results of camera placements are shown in diagram for simulation as well as experiment in the article.

Table 2 Review of related researches on multi-camera localisation (continued)

Approaches	*Algorithm*	*Assumption*	*Experimental setup*	*Constraints*	*Results*
Rekleitis et al. (2006)	SPLAM for both target and camera localisation; uses 3D markers over moving robots as feature points; information propagation among cameras using extended Kalman filter	The moving object is a robot	Seven camera nodes in a closed area consisted with rectangular loop triangular loop and a hall way of around 50 m length; robot traversed three times covering more than 360 m with five different movement patterns to perform ten trials each	Automated detection and calibration system allows 50 trials and 1,500 pattern detections; occurred in 3 hours using 3.2 GHz processor and Linux	Four different paths: stationary, two panel translation, rotation, and square are compared; standard deviation of MSE in square pattern is maximum as u_x and u_y are 2.4 and 13.9 while in two panel translation it is minimum as 3.6 and 5.0 respectively
Barton-Sweeney et al. (2006)	Based on OEE as an enhanced version of direct epipole observation (ME) and extracting epipole from fundamental matrix (EE); LED triangle of known geometry for depth measurement	Pair wise view overlap; modulated LED emission for unique identification	Camera used: imot2 nodes with COTS camera; two camera nodes and 16 non-camera nodes with blinking LEDs; indoor experiment in 6×6 m^2 area and outdoor experiment in 30×30 m^2 area	Cameras can see LEDs up to 4 m in test condition; node to node distance is taken as 85 cm (in indoor condition) and 297 cm (in outdoor condition)	Indoor experiment: OEE 7 cm and ME 2 cm with probability 90%; outdoor experiment: OEE 60 cm and ME 20 cm with probability 90%
Taylor et al. (2006)	Camera with controllable light source for signalling its position to other cameras for determining epipolar geometry; triangulation to determine the pose of non-camera nodes; refinement of pose values through bundle adjustment.	At least two camera nodes with light sources are required; rest of the node poses can be estimated using triangulation	Algorithm is only proposed; no simulations and practical experiment	Not simulated or experimented; hence no experimental setup	Only algorithm is proposed; hence no experimental results

Table 2 Review of related researches on multi-camera localisation (continued)

Approaches	*Algorithm*	*Assumption*	*Experimental setup*	*Constraints*	*Results*
Farrell et al. (2007)	Localises both camera and target; initially PTZ cameras are used for localisation, then motes are localised using magnetometers (a non-imaging sensor); The algorithm can perform in centralised as well as distributed scenario	PTZ cameras are used initially for localisation of nodes, once localised, non-imaging sensors are used further.	Simulated with 100 nodes distributed randomly in 100×100 m^2 area; a subset of 5, 10, 20 and 50 nodes are taken for simulation; experimented with 12 MicaZ motes with omnidirectional LEDs and two PTZ cameras (each with three position and three DOF rotation parameters); a subset of six motes is considered	For each node many PTZ parameters are obtained, their average is used for final location; noise is modelled synthetically to match observed noise	Simulation with different subsets of 100 nodes are taken, that shows the MSE is minimum of 11.73 cm with a subset of 50 nodes and maximum (96.25 m) with a subset of five nodes
Kurillo et al. (2008)	Pairwise view overlap is considered; epipolar geometry employed to calculate essential matrix for pose estimation; scale factor determined by markers on calibration bar; bundle adjustment for refinement	All cameras are pre-calibrated and synchronised	Simulated with five cameras. While experimenting practically cameras are internally calibrated using 10×15 checker board; 12 dragonfly firewire cameras with resolution 640×480 pixels are used in 4.0 m × 4.0 m × 2.5 m area	Two of the cameras (7th and 11th) are installed with 4 mm lens and rest with 6 mm lenses. In vision graph, camera # 3 is chosen as reference camera	Simulation errors are below 0.2% for noise levels of 0.6 pixels and less; in practical experiment image reprojection error varies from 0.0417 to 0.6750 as noise level changes from 0.0 to 0.7
Medeiros et al. (2008)	Pairwise view-overlap and epipolar geometry based estimation used; LED bars used for feature point detection and iterative refinement; four different centralised and distributed approaches are introduced	Cameras are pre-calibrated	Not experimented practically, simulated in an environment with the dimension of the area is $20 \times 20 \times 20$ m^3; 50 cameras on side planes and top plane are randomly placed; single target moves randomly in the area to calibrate the cameras	Bundle adjustment or any such refinement process is not applied to keep it portable to wireless setup; $8 \times \log^2 k$ bits are required for estimation of each parameter, where k is the number of objects used for calibration	Translation error < 60 mm and converges to around 30 mm when simulated for longer time; rotation error < 1.20 and converges to around 0.50 when simulated for longer time.

Table 2 Review of related researches on multi-camera localisation (continued)

Approaches	*Algorithm*	*Assumption*	*Experimental setup*	*Constraints*	*Results*
Piovan et al. (2008)	Node orientation calculated using least square estimate in a ring topology based on angle of arrival sensing; iterative estimation algorithm to reduce the effect of noise	A reference frame is assumed to be attached with each of the node, the first node is labelled as reference node	Simulated using complete graph with ten points (as ten different nodes) making 36 independent cycles; not experimented practically.	The graph representation of camera-nodes is considered to be planner; noise between a pair of nodes in both the directions is assumed to be different	Orientation localisability error (shown as mean square error) reduces with more iterations. As the number of independent cycle increases from 10 to 21 to 36, MSE reduces from 0.08 to 0.03 to approximately 0.025 respectively
Tron and Vidal (2009)	Consensus algorithm is generalised for estimating pose of camera nodes; optimisation of translation and rotation through iterations	Each camera extracts a set of 2D points from each image; neighbouring cameras can have point correspondence between them; all cameras are synchronised; communication among cameras is lossless	Seven cameras each of focal length 1 are distributed roughly in a circle of radius 8 f; cameras connected as four regular graph; 30 randomly distributed feature points in a cubic area of 4.5 f are taken; eight point algorithm used for point correspondence problem; optimisation of rotation with 600 iteration, optimisation of translation with 3,000 iteration and optimisation of overall variables with 100 iterations; experiment repeated for 100 times for each level of noise	Error in rotation and translation with zero-mean Gaussian noise and standard deviation of 0, 1, 2, and 3 pixels in 1,000 × 1,000 pixels	Error in rotation reduces from 4.809% (initial) to 0.393% (after iterations) when the image is corrupted with zero-mean Gaussian and 3 pixel standard deviation; error in translation reduces from 0.291% (initial) to 0.331% (after iterations) when the image is corrupted with zero-mean Gaussian and 3 pixel standard deviation; scale error remained between 1.000% to 1.005% as the deviation ranges from 0 to 3 pixels

Table 2 Review of related researches on multi-camera localisation (continued)

Approaches	*Algorithm*	*Assumption*	*Experimental setup*	*Constraints*	*Results*
Kassebaum et al. (2010)	Localisation through feature point detection of a 3D target moved through the network; DLT method used for estimating projection matrix, further decomposed to get position and orientation parameters	Connected vision graph for pairwise view overlap	A 3D target moved for feature point collection; five smart cameras, other nodes are COTS webcams of 640 × 480 pixel resolution; simulated with five intrinsic parameters and 14 lens distortion parameters (estimated using Zhang's algorithm)	Experimented three times with feature points occupying less than 3% of frame area; 16, 24, or 32 out of 48 available feature points per grid are considered	Position error < 1 inch when the 3D target feature points fill only 2.9% of the frame
Anjum (2011)	Camera localisation using trajectory estimation (CLUTE) is proposed; works on distributed network of non-overlapping cameras; uses Kalman filter to recover pose of camera	Known intrinsic parameters of cameras; cameras aligned with respect to presumed reference camera during registration	Simulated with four and eight camera network, experimented with four camera networks; to analyse in noisy environment, 5% Gaussian noise is introduced in the field of view of cameras	Four cameras used in real time experiment with cameras placed 3–4 m apart; field of view of cameras are limited to square region of 1.5 m^2 coverage area	Through simulation, minimum translation error: 0.13 unit and rotation error: 1.29°; through experiment with real data: minimum translation error: 0.7 unit and rotation error: 10.33°

6 Conclusions

Networked communication in early days used to exploit sound, radio and other acoustic signals for localisation of static sensors. However, with the development of multi-camera network, it gradually became stringent to localise the nodes for initialisation of a camera-network. There are several method devised depending on different types of coverage area, different strength (number) of cameras in network, different types of camera used, and different purpose of the camera-network. The variation has been as wide as ranging from the work of Mantzel et al. (2004) using 2D object (checkerboard) to be feature for localisation till latest work of Kassebaum et al. (2010) employing 3D target with error less than 2.9% and with decreased cost of feature point detection. Table 2 illustrates and compares few landmark researches to portray the variety of algorithms used, assumptions, experimental setups and results thus obtained. There has also been change in application domain of camera-localisation and hence the need of precise localisation. 3D localisation addresses the issue of localising more number of unknown parameters, whereas previous 2D localisation dealt with less number of unknown parameters considering few parameters to be known. Sensing the availability of low-cost cameras, parallel research is going to make the localisation algorithms distributed rather than centralised. Researches have also been perceived in the direction of accurate localisation in presence of noisy environments, e.g., less number of available feature points, feature points on the visual boundaries of the cameras, etc. These kind of algorithms are useful when number of cameras in a network is very high. There is still research going on whether all the unknown parameters (including intrinsic and extrinsic) to determine 3D pose of a camera can be localised.

References

Anjum, N. (2011) 'Camera localization in distributed networks using trajectory estimation', *Journal of Electrical and Computer Engineering*, Vol. 2011, No. 2011, pp.1–13, doi: 10.1155/2011/604647.

Aspnes, J., Eren, T., Goldenberg, D.K., Morse, A.S., Whiteley, W., Yang, Y.R., Anderson, B.D.O. and Balhumeur, P. (2006) 'A theory of network localization', *IEEE Transactions on Mobile Computing*, Vol. 5, No. 12, pp.1663–1678, doi: 10.1109/TMC.2006.174.

Barton-Sweeney, A., Lymberopoulos, D. and Sawides, A. (2006) 'Sensor localization and camera calibration in distributed camera sensor networks', *3rd International Conference on Broadband Communications, Networks and Systems (BROADNETS)*, doi: 10.1109/BROADNETS.2006.4374301.

Biswas, P. and Ye, Y. (2004) 'Semidefinite programming for ad hoc wireless sensor network localization', *Proceedings of 3rd International Symposium on Information Processing in Sensor Networks, IPSN'04*, ACM, pp.46–54, doi: 10.1145/984622.984630.

Bondy, J.A. and Murty, U.S.R. (1982) *Graph Theory with Applications*, 5th ed., Elsevier Science Publishing Co., New York.

Bulusu, N., Heidemann, J. and Estrin, D. (2000) 'GPS-less low-cost outdoor localization for very small devices', *Personal Communications*, Vol. 7, No. 5, pp.28–34, IEEE, doi: 10.1109/98.878533.

Capkun, S., Hamdi, M. and Hubaux, J.P. (2002) 'GPS-free positioning in mobile ad-hoc networks', *Cluster Computing*, Vol. 5, No. 2, pp.157–167.

Chum, O., Matas, J. and Obdrzalek, S. (2003) 'Epipolar geometry from three correspondences', *Computer Vision Winter Workshop (CVWW)*.

Coughlan, J.M. and Ferreira, S.J. (2002) 'Finding deformable shapes using loopy belief propagation', in Heyden, A., Sparr, G., Nielsen, M. and Johansen, P. (Eds.): *Proceedings of 7th European Conference on Computer Vision – Part III (ECCV'02)*, Springer-Verlag, London, UK, pp.453–468.

Davis, L., Borovikov, E., Cutler, R., Harwood, D. and Horprasert, T. (1999) 'Multi-perspective analysis of human action', *Proceedings of 3rd International Workshop on Cooperative Distributed Vision*, Kyoto, Japan.

Dellaert, F., Kipp, A. and Krauthausen, P. (2005) 'A multifrontal QR factorization approach to distributed inference applied to multirobot localization and mapping', in Cohn, A. (Ed.): *Proceedings of 20th National Conference on Artificial Intelligence (AAAI'05)*, AAAI Press, Vol. 3, pp.1261–1266.

Devarajan, D. and Radke, R. (2007) 'Calibrating distributed camera networks using belief propagation', *EURASIP Journal of Applied Signal Processing*, Vol. 2007, pp.1–10, doi: 10.1155/2007/60696.

Devarajan, D., Radke, R.J. and Chung, H. (2006) 'Distributed metric calibration of ad hoc camera networks', *ACM Transactions on Sensor Networks*, Vol. 2, No. 3, pp.380–403, doi: 10.1145/1167935.1167939.

Ellis, T.J., Makris, D. and Black, J. (2003) 'Learning a multicamera topology', *IEEE International Workshop on Visual Surveillance & Performance Evaluation of Tracking & Surveillance*, pp.165–171.

Eren, T., Goldenberg, D.K., Whiteley, W., Yang, Y.R., Morse, A.S., Anderson, B.D.O. and Belhumeur, P.N. (2004) 'Rigidity, computation, and randomization in network localization', *23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004)*, Vol. 4, pp.2673–2684, doi: 10.1109/INFCOM.2004.1354686.

Farrell, R., Garcia, R., Lucarelli, D., Terzis, A. and Wang, I.J. (2007) 'Localization in multi-modal sensor networks', *3rd International Conference on Intelligent Sensors, Sensor Networks and Information (ISSNIP)*, pp.37–42, doi: 10.1109/ISSNIP.2007.4496816.

Felzenszwalb, P.F. and Huttenlocher, D.R. (2004) 'Efficient belief propagation for early vision', *Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, Vol. 1, pp.I-261–I-268, doi: 10.1109/CVPR.2004.1315041.

Freeman, W.T. and Pasztor, E.C. (1999) 'Learning low-level vision', *Proceedings of 7th IEEE International Conference on Computer Vision*, Vol. 2, pp.1182–1189, doi: 10.1109/ICCV.1999.790414.

Funiak, S., Guestrin, C., Paskin, M. and Sukthankar, R. (2006) 'Distributed localization of networked cameras', *Proceedings of 5th International Conference on Information Processing in Sensor Networks (IPSN 2006)*, ACM, pp.34–42, doi:10.1145/1127777.1127786.

Galstyan, A., Krishnamachari, B., Lerman, K. and Pattem, S. (2004) 'Distributed online localization in sensor networks using a moving target', *3rd International Symposium on Information Processing in Sensor Networks (IPSN)*, IEEE, pp.61–70, doi: 10.1109/IPSN.2004.1307324.

Goldenberg, D.K., Krishnamurthy, A., Maness, W.C., Yang, Y.R., Young, A., Morse, A.S. and Savvides, A. (2005) 'Network localization in partially localizable networks', *Proceedings of 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005)*, IEEE, pp.313–326, doi: 10.1109/INFCOM.2005.1497902.

Gotsman, C. and Koren, Y. (2004) 'Distributed graph layout for sensor networks', *International Symposium on Graph Drawing*, pp.273–284.

Hartley, R. and Zisserman, A. (2000) *Multiple View Geometry in Computer Vision*, 1st ed., Cambridge Univ. Press.

Hartley, R. and Zisserman, A. (2004) *Multiple View Geometry in Computer Vision*, 2nd ed., ISBN: 978-0521540513, Cambridge University Press, Cambridge.

Heikkila, J. and Silven, O. (1997) 'A four-step camera calibration procedure with implicit image correction', *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pp.1106–1112, doi: 10.1109/CVPR.1997.609468.

Hofmann-Wellenhof, B., Lichtenegger, H. and Collins, J. (1997) *Global Positioning System: Theory and Practice*, 4th revised ed., ISBN: 3211828397, Springer-Verlag.

Isard, M. and Blake, A. (1998) 'CONDENSATION – conditional density propagation for visual tracking', *Computer Vision*, Vol. 29, No. 1, pp.5–28.

Kanade, T., Rander, P. and Narayanan, P.J. (1997) 'Virtualized reality: constructing virtual worlds from real scenes', *Immersive Telepresence, IEEE Multimedia*, Vol. 4, No. 1, pp.34–47, doi: 10.1109/93.580394, IEEE.

Kassebaum, J., Bulusu, N. and Feng, W. (2010) '3-D target-based distributed smart camera network localization', *IEEE Transactions on Image Processing*, Vol. 19, No. 10, pp.2530–2539, doi: 10.1109/TIP.2010.2062032.

Kato, H. and Billinghurst, M. (1999) 'Marker tracking and HMD calibration for a video-based augmented reality conferencing system', *2nd IEEE and ACM International Workshop on Augmented Reality (IWAR)*, pp.85–94, doi: 10.1109/IWAR.1999.803809.

Khan, U.A., Kar, S. and Moura, J.M.F. (2009) 'Distributed sensor localization in random environments using minimal number of anchor nodes', *IEEE Transactions on Signal Processing*, Vol. 57, No. 5, pp.2000–2016, doi: 10.1109/TSP.2009.2014812.

Kurillo, G., Li, Z. and Bajcsy, R. (2008) 'Wide-area external multi-camera calibration using vision graphs and virtual calibration object', *2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2008)*, ACM, doi: 10.1109/ICDSC.2008.4635695.

Lee, H. and Aghajan, H. (2005) 'Collaborative self-localization techniques for wireless image sensor networks', *39th Asilomar Conference on Signals, Systems and Computers*, pp.1566–1570, doi: 10.1109/ACSSC.2005.1600030.

Liu, X., Kulkarni, P. and Shenoy, P. (2006) 'Snapshot: a self-calibration protocol for camera sensor networks', *3rd International Conference on Broadband Communications, Networks, and Systems (BROADNETS)*, doi: 10.1109/BROADNETS.2006.4374302.

Liu, Y., Huang, T.S. and Faugeras, O.D. (1990) 'Determination of camera location from 2-D to 3-D line and point correspondences', *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 12, No. 1, pp.28–37, doi: 10.1109/34.41381.

Lowe, D.G. (2004) 'Distinctive image features from scale-invariant keypoints', *International Journal of Computer Vision*, Vol. 60, No. 2, ACM, doi:10.1023/B:VISI.0000029664.99615.94.

Lymberopoulos, D., Sweeny, A.B. and Savvides, A. (2005) 'Sensor localization and camera calibration using low power cameras', ENALAB Technical Report 090105.

Ma, Y., Soatto, S., Kosecka, J. and Sastry, S.S. (2004) 'An invitation to 3D vision: from images to geometric models', *Interdisciplinary Applied Mathematics*, Vol. 26, ISBN: 978-0-387-00893-6, Springer.

Maeda, M., Ogawa, T., Kiyokawa, K. and Takemura, H. (2004) 'Tracking of user position and orientation by stereo measurement of infrared markers and orientation sensing', *8th IEEE International Symposium on Wearable Computers (ISWC)*, IEEE computer Society, pp.77–84, doi: 10.1109/ISWC.2004.46.

Mantzel, W.E., Choi, H. and Baraniuk, R.G. (2004) 'Distributed camera network localization', *38th Asilomar Conference on Signals, Systems and Computers*, Vol. 2, pp.1381–1386, doi: 10.1109/ACSSC.2004.1399380.

Marinakis, D., Dudek, G. and Fleet, D. (2005) 'Learning sensor network topology through Monte Carlo expectation maximisation', *Proceedings of IEEE International Conference on Robotics & Automation*, Spain.

Medeiros, H., Iwaki, H. and Park, J. (2008) 'Online distributed calibration of a large network of wireless cameras using dynamic clustering', *2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2008)*, IEEE, doi: 10.1109/ICDSC.2008.4635698.

Moses, R.L., Krishnamurthy, D. and Patterson, R. (2003) 'A self-localization method for wireless sensor networks', *EURASIP Journal on Applied Signal Processing*, Vol. 2203, No. 4, pp.348–358.

Murphy, K.P., Weiss, Y. and Jordan, M.I. (1999) 'Loopy belief propagation for approximate inference: an imperical study', *Proceedings of Uncertainty in Artificial Intelligence UAI'99*, pp.467–475.

Nakazato, Y., Kanbara, M. and Yokoya, N. (2005a) 'A localization system using invisible retro-reflective markers', *IAPR Conference on Machine Vision Applications*, pp.140–143.

Nakazato, Y., Kanbara, M. and Yokoya, N. (2005b) 'Localization of wearable users using invisible retro-reflective markers and an IR camera', *SPIE Electronic Imaging*, pp.1234–1242.

Olfati-Saber, R., Fax, J.A. and Murray, R.M. (2007) 'Consensus and cooperation in networked multi-agent systems', *Proceedings of the IEEE*, pp.215–233, doi: 10.1109/JPROC. 2006.887293.

Paskin, M.A. (2003) 'Thin junction tree filters for simultaneous localization and mapping', *Proceedings of 18th international joint conference on Artificial intelligence IJCAI'03*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp.1157–1164.

Paskin, M.A. and Guestrin, C.E. (2004) 'Robust probabilistic inference in distributed systems', *Proceedings of 20th conference on Uncertainty in artificial intelligence UAI'04*, AUAI Press, USA, pp.436–445.

Patwari, N. and Hero, A.O., III (2004) 'Manifold learning algorithms for localization in wireless sensor networks', *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, Vol. 3, pp.iii-857–iii-860, doi: 10.1109/ICASSP.2004.1326680.

Patwari, N., Hero, A.O., III, Perkins, M., Correal, N.S. and O'Dea, R.J. (2003) 'Relative location estimation in wireless sensor networks', *IEEE Transactions on Signal Processing*, Vol. 51, No. 8, pp.2137–2148, doi: 10.1109/TSP.2003.814469.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers Inc., USA, ISBN: 1558604790.

Piovan, G., Shames, I., Fidan, B., Bullo, F. and Anderson, B.D.O. (2008) 'On frame and orientation localization for relative sensing networks', *47th IEEE Conference on Decision and Control (CDC 2008)*, IEEE, pp.2326–2331, doi: 10.1109/CDC.2008.4738809.

Poelman, C.J., Albuquerque, N.M. and Kanade, T. (1997) 'A paraperspective factorization method for shape and motion recovery', *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 19, No. 3, pp.206–218, doi: 10.1109/34.584098.

Rahimi, A., Dunagan, B. and Darrell, T. (2004) 'Simultaneous calibration and tracking with a network of non-overlapping sensors', *Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, IEEE, Vol. 1, pp.I-187–I-194, doi: 10.1109/CVPR.2004.1315031.

Rekleitis, I., Meger, D. and Dudek, G. (2006) 'Simultaneous planning, localization, and mapping in a camera sensor network', *Journal of Robotics and Autonomous Systems*, Vol. 54, No. 11, pp.921–932, ISSN: 0921-8890, doi: 10.1016/j.robot.2006.05.009, Elsevier.

Savarese, C., Rabaey, J. M. and Beutel, J. (2001) 'Location in distributed ad-hoc wireless sensor networks', *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2001)*, Vol. 4, pp.2037–2040, doi: 10.1109/ICASSP.2001.940391.

Savvides, A., Han, C-C. and Strivastava, M.B. (2001) 'Dynamic fine-grained localization in ad-hoc networks of sensors', *Proceedings of 7th Annual International Conference on Mobile Computing and Networking (MobiCom'01)*, ACM, pp.166–179, doi: 10.1145/381677.381693.

Shang, Y., Ruml, W., Zhang, Y. and Fromherz, M.P.J. (2003) 'Localization from mere connectivity', *4th ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc'03)*, ACM, pp. 201–212, doi: 10.1145/778415.778439.

Spurlock, S. and Souvenir, R. (2012) 'Dynamic subset selection for multi-camera tracking', *50th Annual Southeast Regional Conference*, ACM, pp.244–249, doi: 10.1145/2184512.2184570.

Sturm, P.F. and Triggs, B. (1996) 'A factorization based algorithm for multi-image projective structure and motion', in Buxton, B.F. and Cipolla, R. (Eds.): *Proceedings of 4th European Conference on Computer Vision (ECCV'96)*, Vol. 2, Springer-Verlag, London, UK, pp.709–720.

Sudderth, E.B., Mandel, M.I., Freeman, W.T. and Willsky, A.S. (2005) 'Distributed occlusion reasoning for tracking with nonparametric belief propagation', in Saul, L.K., Weiss, Y. and Bottou, L. (Eds.): *Advances in Neural Information Processing Systems*, Vol. 17, pp.1369–1376, MIT Press, Cambridge.

Taj, M. and Cavallaro, A. (2011) 'Distributed and decentralized multicamera tracking', *Signal Processing Magazine*, Vol. 28, No. 3, pp.46–58, doi: 10.1109/MSP.2011.940281, IEEE.

Taylor, C., Rahimi, A., Bachrach, J., Shrobe, H. and Grue, A. (2006) 'Simultaneous localization, calibration, and tracking in an ad hoc sensor network', *5th International Conference on Information Processing in Sensor Networks (IPSN 2006)*, ACM, pp.27–33, doi: 10.1109/IPSN.2006.244053.

Taylor, C.J. (2004) 'A scheme for calibrating smart camera networks using active lights', *Proceedings of 2nd international conference on Embedded networked sensor systems (SenSys'04)*, ACM, pp.322–322, doi: 10.1145/1031495.1031565.

Tenmoku, R., Kanbara, M. and Nakazato, Y. (2003) 'Augmented reality system using positioning infrastructures and a pedometer', *7th IEEE International Symposium on Wearable Computers (ISWC)*, IEEE Computer Society, pp.110–117.

Tomassi, C. and Kanade, T. (1992) 'Shape and motion from images streams: a factorization method', Carnegie Mellon Technical Report CMU-CS-92-104.

Triggs, B., McLauchlan, P.F., Hartley, R.I. and Fitzgibbon, A.W. (1999) 'Bundle adjustment – a modern synthesis', in Triggs, B., Zisserman, A. and Szeliski, R. (Eds.): *Proceedings of International Workshop on Vision Algorithms: Theory and Practice (ICCV'99)*, Springer-Verlag, pp.298–372, ISBN: 3-540-67973-1.

Tron, R. and Vidal, R. (2009) 'Distributed image-based 3D localization of camera sensor networks', *Proceedings of 48th IEEE Conference on Decision and Control*, IEEE, pp.901–908, doi: 10.1109/CDC.2009.5400405.

Tron, R., Vidal, R. and Terzis, A. (2008) 'Distributed pose averaging in camera networks via consensus on SE(3)', *2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2008)*, IEEE, doi: 10.1109/ICDSC.2008.4635701.

Xu, G. and Zhang, Z. (1996) *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*, ISBN: 978-0792341994, Springer.

Zhang, Z. (1998) 'Determining the epipolar geometry and its uncertainty: a review', *International Journal of Computer Vision*, Vol. 27, No. 2, pp.161–195, doi: 10.1023/A:1007941100561.

Zhang, Z. (2000) 'A flexible new technique for camera calibration', *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 22, No. 11, pp.1330–1334, doi: 10.1109/34.888718.