Occlusion prediction algorithms for multi-camera network

Rahul Raman, Pankaj K. Sa, and Banshidhar Majhi
Department of Computer Science and Engineering
National Institute of Technology Rourkela
Odisha, India
rahulraman2@gmail.com,{pankajksa,bmajhi}@nitrkl.ac.in

***Abstract*—The mode of object tracking has evolved from single camera tracking to multi-camera tracking over the last few years. Even though multi-camera model overcomes limitations present in single camera system, it introduces complexities of handling network of multiple cameras. In this paper, we propose a novel real time occlusion prediction method for multi-camera network thus reducing complexities and cost of tracking in multi-camera model without losing track of the subject. The proposed method constitutes sequential phases as (a) estimation of direction of relative motion in real plane based on change in bounding pixel positions of tracked subject, (b) constructing algorithms for occlusion prediction, and (c) mitigation of occlusion by awaking a minimal subset of cameras in network. The method uses the pattern of change in the dimensions of bounding box with respect to frame number, obtained by applying background subtraction to the video of mutual motion of the subjects. On the basis of estimated direction of motion it uses proposed algorithms to decide the possibility and proximity of occlusion and thus awaking the minimal set of camera in the network that does not encounter occlusion. The proposed approach has been verified using various video samples and is observed that the proposed method successfully predicts the occurrence of occlusion.**

***Index Terms*—Occlusion prediction algorithm, Multi-camera network, Perspective view analysis**

I. Introduction

The efforts in technological growth have made way for the emergence of variety of methodologies for tracking objects in diverse contexts. Different algorithms have been designed for different requirements depending upon the mode of tracking, location, significance and specific needs. The earlier tracking approaches have implemented several image processing algorithms on the video output from a single camera. Contour based tracking, background subtraction based tracking, Gaussian based tracking, median filter based tracking, are some of the most studied and refined technologies among them [1]. These algorithms are simple in implementation, fast in processing and analysis. However, they are limited with constant field of view and suffer from occlusion of the tracked subject. As the demand for foolproof tracking algorithm prevailed so is the paradigm shifted from single to multi-camera model. These systems are more useful for tracking in crowded places and highly protected areas. This can be equipped with a variety of cameras and distributed processors to even amend the functionality of tracking. But multi-camera systems have their complexities and trade-offs. As compared with single camera tracking, multi-camera tracking needs additional processing, extra memory requirement, superfluous energy consumption, higher installation cost, and complex handling and implementation. Our novel approach attempts to bridge the gap between single and multi-camera tracking. The prime focus of this paper is to exploit multiple-camera network in such a way that it utilizes the resources efficiently as in a single camera network yet remain capable in providing strong tracking features as in multi-camera network. It uses a few of the cameras of the pre-calibrated multi-camera network to work always in active mode. This camera uses the results of background subtraction tracking and does further analysis on it for handling occlusion. Occlusion handling is one of the major problems in single camera based tracking. In the model proposed by Sinior et al. [2] background subtraction is used for object tracking and occlusion detection. It uses appearance based model to estimate the centroid of the moving object more accurately. This technique is although reliable but works with fixed background. Authors in[3], [4] have handled occlusion based on measurement error for each pixel. Authors in [5] have devised a motion based tracking algorithm that is adaptive with natural changes in appearance or variation in 3D pose and hence remain robust with occlusion but does not resolve or predict occlusion. In [6] two different approaches to cope occlusion are proposed; one using evaluation of correlation error in templates and other using infrared images to detect occluded region by human hand. Authors in [7] have exploited contextual information; it does better occlusion analysis but has tracking errors. It uses block motion vectors for calculating object boundary to predict occlusion. Amizquita et al. [8] have proposed an algorithm for auto detection of occlusion using motion based prediction of objects movement during the stages of entering occlusion, full occlusion, and exit occlusion.

On the other hand a multi-camera system can avoid occlusion and can provide robust tracking but are not as simple and energy-efficient as single camera systems. Although a camera system installed in master-slave mode [9], has the energy efficiency but the entire region under coverage should come under master cameras view. Towards making the multi-camera model as an efficient approach a few other works have also been proposed. Kulkarni et al. [10] have proposed an approach for efficient use of multiple cameras by devising multi-tier camera network called SensEye [11]. This approach is energy efficient although it has a complex hardware architecture and diverse software requirement.

It is observed from the literature that object tracking using a single camera is simple and energy efficient. Further, it has the ability to predict occlusion, but there is no scope for occlusion avoidance. To alleviate the occlusion occurrence, a multi-camera model is necessary where the field of view is tracked by multiple cameras. Generally, a multi-camera based approach utilizes the cameras always in the active mode. But this leads to energy inefficiency and more processing requirement. In our proposed work, we have assumed a multi-camera model for object tracking. Our algorithm is made to work on a single camera which is a part of multi-camera system and instead of handling occlusion; it predicts occlusion and avoids its occurrence. The multi-camera network may work in peer to peer model using simple cameras or in master-slave model using a wide view camera as master and other cameras as slaves. Our approach analyses the change in the dimension of the subject in camera coordinates as it moves in 3D world coordinate. It further analyses the data to decide the motion and direction of the subjects and their probability and proximity of occlusion. Based on this prediction further decision towards avoiding occlusion can be made.

Section II of this paper discusses our proposed work and algorithm for occlusion prediction and avoidance in detail. Section III shows experimental justification towards the proposed approach. Section IV draws the conclusion and articulates the future scope of work towards the complete solution of the aforementioned problems.

II. Proposed work and algorithm

Background subtraction is a reliable method for tracking motion of a subject with respect to a fixed background. The results of background subtraction are used in our approach for analysis. A few reasonable assumptions are made in our approach while considering motion of two human as different subjects, such as eight possible directions of motion and three levels of speed for any subject on move are considered. This assumption discretizes the approach at both the levels of direction as well as speed.

In the proposed work, cameras in the set can be of two types, one which always remain in active mode and other that remain in sleep mode until it is not required for tracking. Cameras with wide angle lenses can be used for monitoring, since they have extensive angular view. They always remain in active mode while high zoom cameras or general cameras can be used to work at specific time. Wide angle cameras can preferably be placed at locations having large view range. For implementation in general scenario, similar setup can be done all along with same camera. In the experiments conducted here all cameras of same type are used for the sake of simplicity. The proposed work can be well described by subdividing into following three steps:

(A) Direction of motion estimation

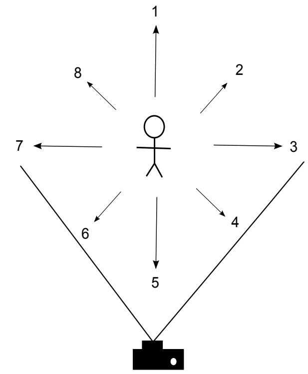

Fig. 1. Discrete Direction Vectors

(B) Algorithms for occlusion prediction

(C) Mitigation of occlusion

A. Direction of motion estimation

Motion of a subject can be in any direction but are discretized to 8 directions as shown in Fig. 1.

Since the direction of motion has to be utilized for approximating the chances of occlusion, discrete direction can be applied. The discrete direction will also provide faster computation which is needed for real time processing. To realize the direction of motion of a subject, change in the width, height, and location co-ordinates of the bounding box of the subject is studied. The pattern change in the subsequent frames of sample video during the motion in the perspective view of camera is shown in Fig. 2. Change in the width of bounding box versus change in frame, change in the height of bounding box versus change in frame and ratio of change in width to height versus frame change are plotted. This gives the pattern based on which the direction of motion can be explained .

In case of larger view area, high resolution wide angle camera is ideal as they provide wider field of view. However, such cameras have a wider angular width. Due to this, subjects at different angle from camera will make different view axis. Direction vectors generated with respect to this view axis is called local direction vectors. These local direction vectors vary from one location to another and have to be converted to global direction vectors before analyzing them through algorithms. Fig. 3 shows the case diagrammatically. However in the experiments conducted here, general cameras having common views are taken while all direction vectors obtained are global direction vectors. Therefore, no local to global direction vector updating is required. Further, existing curves for different directions are compared with real time graphs of each subject. In this way, based on least Euclidian distance of real time curve from the existing curves, their direction of motion is estimated.

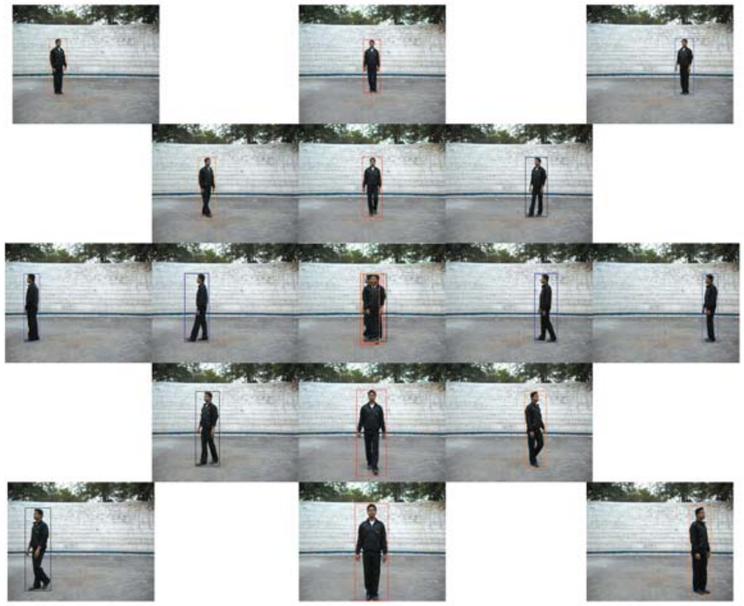

Fig. 2. Discrete direction estimation from dimension change in bounding box

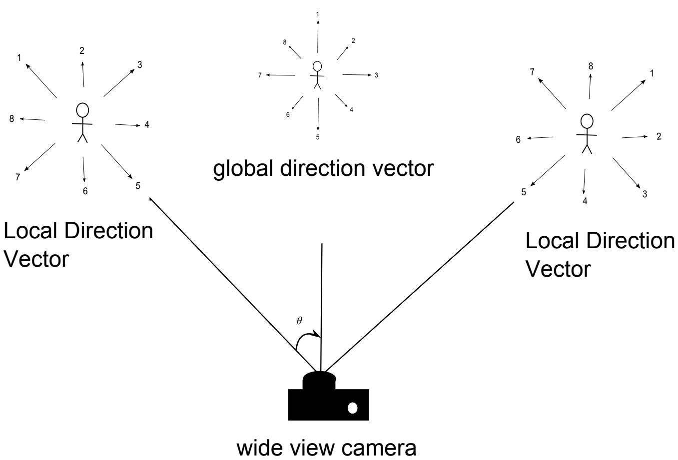

Fig. 3. Wide angle lenses with different line of sights

B. *Algorithms for occlusion prediction*

Successful estimation of direction of subject's motion takes the work to the next level where the direction of motion estimates is used in the proposed algorithm to predict the chances of occlusion in near proximity. This prediction is a two-step process and separate algorithms are proposed for each of them. Algorithm: 1 predicts the possibility and proximity of occlusion based on the direction of motion.

Algorithm: 1 gives probability regarding the occlusion due to the subject's direction. Subjects are very likely to occlude when they are approaching each other while if they are moving parallel to the camera's view axis or moving away from each other occlusion is not possible. In other scenario, chances of occlusion can be predicted based on the speed of subjects.

Algorithm 2 deals with situations where decision is taken based on the speed of the subject. When the subjects are moving in the same direction, their speed is crucial to decide the chances of occlusion. If subjects are moving in the same direction and the chasing subject has speed higher than that of the other then occlusion may occur. Algorithm 2 also decides the proximity of occlusion when subjects are approaching each other. These two algorithms give the probability of

---

Algorithm 1 Occlusion prediction based on subjects direction of motion

---

Input: Direction vectors of subject 1 and subject 2
Output: Probability of occlusion due to direction of subjects $\Pr\{Occ_{direction}\}$

Step 1: Get the direction vectors of sub1 and sub2
Step 2: If subjects are approaching each other (i.e. direction$(sub1) = (2, 3, 4)$ and direction$(sub2) = (6, 7, 8)$)
return $\Pr(Occ_{direction})$
goto step 6
Step 3: If sub1 and sub2 are moving parallel to the view of camera
i.e. direction$(sub1) = (1, 5)$ and direction$(sub2) = (1, 5)$
$\Pr(Occ) = 0$
goto step 6
Step 4: If sub1 and Sub2 are in same direction
(i.e. direction for both subjects is (2,3,4) or (6,7,8))
return $\Pr(Occ_{direction})$
goto step 6
Step 5: If sub1 and sub2 are in opposite direction
(i.e. directions are either ((2,3,4) and (6,7,8)) or ((6,7,8) and (2,3,4)))
Step 6: Exit

Algorithm 2 Occlusion prediction based on subjects speed

Input: Direction vectors and speeds of subject 1 and subject 2
Output:Probability of occlusion due to speed of subjects i.e. Pr(Occ_{speed})

step 1:Get direction vectors and speed of sub1 and sub2
Step 2: If subjects are approaching each other (i.e. direction$(sub1) = (2, 3, 4)$ and direction$(sub2) = (6, 7, 8)$)
return Pr(Occ_{speed})
(used for detecting proximity of occlusion)
goto step 5
step 3: If sub1 and sub2 are in same direction
(i.e. direction for both subjects is (2,3,4) or (6,7,8))
return Pr(Occ_{speed})
(Occlusion highly depends on relative speed of subjects)
goto step 5
step 4:In all other cases return Pr(Occ_{speed})=0
goto step 5
step 5: Exit

occlusion with respect to directions and speed which can be mathematically modeled as:

- Pr$\{Occ\}$: overall probability of occlusion
- Pr$\{Occ\}$=$W_1 \times Pr(Occ_{direction})$+$W_2 \times Pr(Occ_{speed})$
- Pr$\{Occ_{direction}\}$: Probability of occlusion due to the subjects direction given by algorithm 1
- Pr$\{Occ_{speed}\}$: Probability of occlusion due to subjects speed given by algorithm 2

Where W_1 and W_2 are the weights assigned to the probabilities and depend upon the factors such as distance of subjects from each other, distance of subject from camera etc. Fig. 6 to Fig. 9 shows different scenarios where occlusion is caused due to the direction and/or speed of the subjects.

C. Mitigation of occlusion

Based on the direction of motion and speed of the subjects, their chances of occlusion can be predicted. Up to this level, the system acts as a single camera system. Once a conclusion is drawn about the proximity of occlusion, then steps towards mitigation of occlusion takes place. Multiple cameras in the systems are precalibrated and localized, hence the poses of cameras are known to each other. In case of any possible occlusion, the active camera selects minimal number of cameras in the network that does not encounter occlusion. The camera will awake those set of cameras and they start further tracking of subject. This will not only let the continuous tracking of subjects possible but also let the cameras to work efficiently in terms of energy and processing complexity.

III. EXPERIMENTAL EVALUATION

To justify the proposed approach towards occlusion prediction, different test cases are experimented. Sample videos are captured with various possible motions in different directions depicting the scenario of occlusion and non-occlusion. During the motion, subject may be nearer to or far from the camera and hence may appear larger or smaller in size, but the characteristics of the graph will remain unchanged irrespective of its appearance. Many different sample cases are considered where subjects are either approaching or moving away from each other. In these cases the bounding boxes of each of the subjects are studied for its width and height change. As shown earlier, motion has given discrete direction from 1 to 8. Direction vectors 8, 1, 2 and 3 are exactly opposite in behavior with respect to direction vectors 4, 5, 6 and 7 respectively. Corresponding graphs are plotted in the Fig. 6 and Fig. 7 that depicts the change in width of bounding box and change in height of bounding box with the frame sequence. Since opposite behavior of direction vectors appears as mirror image of other graph, only four of the distinct graphs are compared.

Fig. 5 shows change in the width of bounding box with change in frame number. A constant increase in the width (as shown by green curve) shows the motion of the subject in direction 5. Red curve in the graph shows a pattern repeated with constant frequency depicting the motion towards direction 7. Similarly Directions 4 and 6 are shown by blue and black curves depicting their behavior.

Fig. 4 shows change in the height of bounding box with change in frame number. During motion towards direction vector 7, height is constant while during motion towards direction vector 5 heights is constantly increasing in the perspective view which is shown through red and green curve respectively. Motion along direction 4 and direction 6 has very similar increase in height and is visible in the blue and black direction curves respectively. Direction of motion of a subject is estimated by comparing the nature of the subjects motion curve with existing curves (for both width and height) and minimal Euclidian distance of the subjects curve from existing one is considered for assigning a direction to the subject. However another graph with change in the ratio of width to height versus frame number can also be plotted and compared for each subject for another level of comparison.

These values are further used for processing through described algorithms. Fig. 6 and Fig. 7 shows sequence of frames where the directions of two subject are estimated based on earlier mentioned method. On that basis their probability and proximity of occlusion is estimated. However occlusion also depends on subjects speed. Fig. 8 and Fig. 9 shows two frame sequences in which the direction of subjects are same but Fig. 8 resulting occlusion while Fig. 9 avoids occlusion. The direction in both the cases is found to be direction 3. Algorithm 1 cannot decide the chances of occlusion here hence their speeds are compared through Algorithm 2. This speed is the apparent speed, as nearer subjects motion appears much more

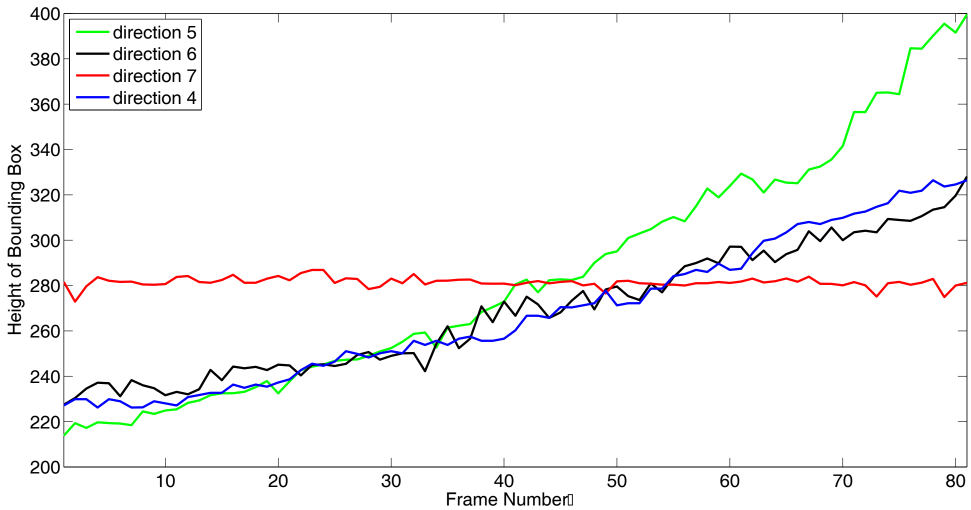

Fig. 4. Change in height of Bounding Box with frame sequence with respect to different directions

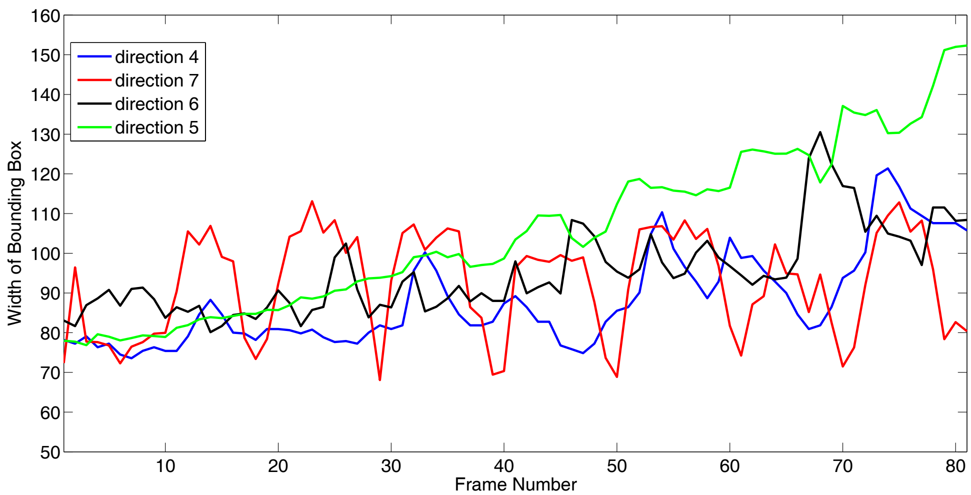

Fig. 5. Change in width of Bounding Box with frame sequence with respect to different directions

than farther subjects. The difference in the subjects' speed level is found zero in Fig. 8 resulting non-occlusion while the speed level difference is found to be two in Fig. 9 resulting occlusion.

IV. Conclusion and Future Work

The proposed work shows efficient and effective way for predicting occlusion and avoiding it in a multi-camera network. The approach lets the system to work in single camera mode avoiding complexity and higher operation cost as long as tracking is possible from one camera. It awakes one or more cameras in case of any occlusion and does not compromise in losing the track of the subject. It finds its implementation in places like dedicated roads, office corridor, etc where multiple cameras are installed. The concept can also be implemented in larger setup with little modification in the algorithm and addition of dedicated cameras. Even though, the proposed work is basically designed for predicting occlusion and awaking a minimal number of cameras to start tracking and thus avoiding occlusion, it can further be extended for best view synthesis while tracking a subject from more than one cameras. The minimal set of cameras that are rightly placed to avoid occlusion can thus compare their view and based on the consensus, the camera with best view can track hence forth. The curves obtained for different individual may vary from person to person and this even amend the chances of adding persons identification mechanism based on gait pattern analysis.

Further scope towards implementation of the proposed work lies in developing algorithms that can decide the control flow from one camera to another. Mitigation of occlusion can be done when a robust algorithm can utilize the pose (position

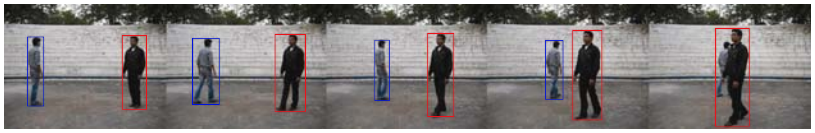

Fig. 6. Occlusion predicted with subjects directions as 2 and 6

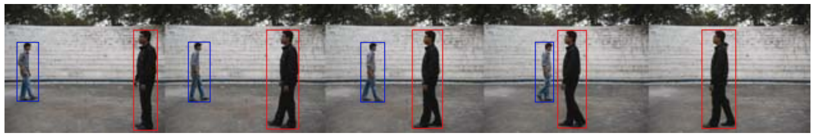

Fig. 7. Occlusion predicted with subjects directions as 3 and 7

Fig. 8. Occlusion predicted with subjects directions as 3 and 3 but velocity levels as 3 and 1

Fig. 9. No occlusion predicted with subjects directions as 3 and 3 but velocity levels as 1 and 1

with orientation) information of pre-localized and calibrated cameras, and based on these values, it can awake the camera which does not encounter occlusion at all. Since the algorithm works in real time environment, and finds its implementation in vital issues as surveillance, fast and optimized working along with high accuracy is a concern that can always be explored.

ACKNOWLEDGMENT

The authors would like to thank all the co-researchers at Department of Computer Science and Engineering, National Institute of Technology Rourkela for their active cooperation towards manifestation of this research.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Survey*, vol. 38, no. 4, Dec 2006, doi: 10.1145/1177352.1177355.

[2] A. Senior, A. Hampapur, Y. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance models for occlusion handling," *Image and Vision Computing*, vol. 24, pp. 1233–1243, 2006, doi: 10.1016/j.imavis.2005.06.007.

[3] H. Nguyen and A. Smeulders, "Fast occluded object tracking by a robust appearance filter," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1099 –1104, Aug. 2004, doi: 10.1109/TPAMI.2004.45.

[4] S. Zhou, S. Kevin, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," vol. 13, pp. 1491–1506, 2004, doi: 10.1109/TIP.2004.836152.

[5] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi, "Robust online appearance models for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1296–1311, 2003, doi: 10.1109/TPAMI.2003.1233903.

[6] K. Ito and S. Sakane, "Robust view-based visual tracking with detection of occlusions," in *IEEE International Conference on Robotics and Automation, 2001. Proceedings 2001 ICRA*, vol. 2, 2001, pp. 1207–1213, doi: 10.1109/ROBOT.2001.932775.

[7] K. Hariharakrishnan and D. Schonfeld, "Fast object tracking using adaptive block matching," *IEEE Transactions on Multimedia*, vol. 7, no. 5, pp. 853–859, Oct. 2005, doi: 10.1109/TMM.2005.854437.

[8] N. Amezquita, R. Alquezar, and F. Serratosa, "Dealing with occlusion in a probabilistic object tracking method," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference*, june 2008, pp. 1–8, doi: 10.1109/CVPRW.2008.4563060.

[9] H. Aghajan and A. Cavallaro, *Multi-Camera Networks Principles and Applications*. Academic Press, 2009, ISBN: 978-0123746337.

[10] P. Kulkarni, D. Ganesan, and P. Shenoy, "The case for multi–tier camera sensor networks," in *Proceedings of the international workshop on Network and operating systems support for digital audio and video*, ser. NOSSDAV, 2005, pp. 141–146, doi: 10.1145/1065983.1066016.

[11] P. Kulkarni, D. Ganesan, P. Shenoy, and Q. Lu, "Senseye: a multi-tier camera sensor network," in *Proceedings of the 13th annual ACM international conference on Multimedia*, 2005, pp. 229–238, doi: 10.1145/1101149.1101191.