Scientia Iranica A (2016) 23(4), 1641–1647

Sharif University of Technology
Scientia Iranica
Transactions A: Civil Engineering
www.scientiairanica.com

Improved multi-camera smart traffic surveillance system for resilient cities

Z. Kavehvash*

Department of Electrical Engineering, Sharif University of Technology, Tehran, P.O. Box 11155-4363, Iran.

Received 10 February 2015; received in revised form 11 May 2015; accepted 2 June 2015

KEYWORDS
Smart traffic surveillance system;
Multi-camera imaging;
Three-dimensional imaging;
Image processing;
Geometrical optics.

Abstract. In this paper, intelligent traffic surveillance system as an important part of a smart resilient city is reviewed. The smart traffic control system is very important in improving the lifestyle by decreasing the traffic saturation and air pollution. Still, multi-camera vision is more helpful in implementing different automatic traffic surveillance systems. Apart from their superior features, existing multi-camera surveillance systems suffer from computational complexity and degraded accuracy. The main reason for these shortcomings arises from image processing errors. These errors depend on the image content and thus are not predictable. To overcome this shortage, three-dimensional (3D) optical techniques for improved fusing of the multi-camera images and thus extracting 3D vehicle locations is proposed in this paper. In fact, the proposed multi-camera visionary system is a combination of image processing based and geometrical optics based methods. The result of 3D image reconstruction through the proposed technique shows its dominance in providing the 3D image information.

1. Introduction

The role of Information and Communication Technology (ICT) in creating a resilient city is to make an infrastructure system for the smart city which is cost- and resource-efficient. Monitoring and management of water cycle, energy usage, transportation and traffic, and air pollution are among the main issues to be considered in a smart city infrastructure. Among these issues, perhaps intelligent transportation system plays an important role.

An intelligent transportation system could be established in the city by integrating wireless sensors into its road infrastructure. The sensors might monitor traffic and adaptively control traffic signals in real time. The system can lead to reduction in traffic congestion and carbon emissions, thus enhancing public health. Furthermore, the data collected from the sensors could be precious for future transportation planning projects. This system might be designed to facilitate priority routes for emergency vehicles, to be detected in real-time. Vehicle traffic can also be monitored in order to modify the city traffic lights in a dynamic and adaptive way. Processing the images captured with the cameras installed at the crossroads helps in automatic estimation of the traffic congestion on each side. In this way, the length of green and red signals could be set accordingly. Understanding the flow and congestion of vehicular traffic is essential for efficient road systems in cities. Smooth vehicle flows reduce journey times, reduce emissions, and save energy. Monitoring traffic – whether road vehicles or people – is useful for operators of roads and transport hubs. The monitoring system can also be used to calculate the average speed of the vehicles which transit over a roadway by taking the time mark at two different points. This platform can help drivers to avoid congested roads through provision of real time warnings on electronic displays or via smart-phone applications. Such data can even be used

*. *Tel.: +98 21 66165927; Fax : +98 21 66023261*
E-mail address: kavehvash@sharif.edu

to assess the suitability of emergency evacuation plans. A smart city infrastructure can also help in solving atmospheric pollution. The quality of city life across the world is negatively impacted by atmospheric pollution and congested roads. Road congestion results in lost time for motorists and wasted fuel and is also a major cause of air pollution. A significant contribution to congestion arises from motorists searching for available parking spaces – often requiring a considerable time before they are successful – that is a major source of driver frustration. Providing accurate information to drivers on where to find available parking spaces helps traffic flow to be bettered. Motorists get timely information so that they can locate a free parking slot quickly, saving time and fuel. This information can reduce traffic jams and atmospheric pollution, improving the quality of life.

Traffic control is of great importance in Iran, especially in Tehran, considering the problems of air pollution and traffic congestion. In recent years, there has been a growing use of camera vision in different traffic monitoring systems implemented in Iran. The automatic speed control system, red light enforcement system, emergency green road, and traffic restricted zone management system are examples of using smart traffic surveillance system in Iran. An example of an image processed for red light enforcement at a cross-section in Tehran is shown in Figure 1. Still, broad aspects of an intelligent traffic control system could be implemented through camera vision and image processing, including traffic congestion estimation and traffic signal timer control. These aspects of camera vision are presently utilized in some smart cities such as Seoul, Singapore, and New York.

Still, almost all of the existing smart traffic surveillance systems are based on processing of single camera images. Unfortunately, the view of a single camera is finite and limited by scene structures. In order to monitor a wide area, such as tracking a vehicle traveling through the road network of a city or analyzing the global filled and empty locations in a parking lot, video streams from multiple cameras have to be used. To this end, many intelligent multi-camera video surveillance systems have been developed [1-14]. Multi-camera vision is a multidisciplinary field related to computer vision, pattern recognition, signal processing, communication, embedded computing, and imaging sensors.

Multi-camera tracking requires matching tracks obtained from different camera views according to their visual and spatio-temporal similarities. Matching the appearances of image regions is studied in object re-identification. The spatio-temporal reasoning requires camera calibration and the knowledge of topology. In commonly in use smart traffic surveillance systems via multi-camera vision, image fusion is completely performed through image processing techniques. The calibration and registration of different images, taken from variant viewpoints of a scene, are performed through extracting different image features and mathematically matching them. The most commonly method, used for merging multi-camera images and extracting the depth information is line-path triangulation. However, inferring the visible surface depth from two or more images using light-path triangulation is an ill-posed problem due to the multiplicity of ways for establishing point correspondences in the input images. To alleviate this problem of ambiguity, different rectification algorithms [5,6] have been proposed to rearrange image pixels so that the corresponding points (that result from the projection of the same 3D point) will lie on the same image scan line. Even with this powerful constraint in hand, identifying stereo correspondences is still a very challenging problem. A great number of stereo algorithms have been proposed in the past few decades and many of them are surveyed in [7,8]. Among all these algorithms, Dynamic Programming (DP)-based

Figure 1. Red light enforcement through image processing at a cross-road in Tehran [https://ir.linkedin.com/pub/farshid-babakhani/42/7a7/b59].

optimization techniques are often used due to their simplicity and efficiency. DP is an efficient algorithm that constructs globally optimal solutions by reuse and pruning. One common way to achieve global optimality and stabilize the DP-based stereo matching results is to impose the continuity (or smoothing) constraint. This constraint indicates that the neighboring points should have similar stereo disparity values, as they view 3D points that lie close to each other. However, this constraint is only applicable to a single or a few neighboring scan lines, and applying this constraint often results in undesired striking effects. Furthermore, without using any Ground Control Points (GCPs) or reliable anchor points as guidance, DP is very sensitive to the parameters chosen for the continuity constraint [9]. In other multi-view fusion methods used in video surveillance, different view images are registered and fused through geometric constraints of the roads [10]. Among these methods, single Plane probability Fusion Map (PFM) is perhaps the most successful [11-13]. However, since the coplanarity of image points is not strictly true, the single plane PFM is subject to distortions, which lead to less accurate measurement of target positions and dimensions. More work has been done in terms of 3D model matching for tracking and vehicle classification [14,15]. However, in these approaches, there should be no occlusion in the first few frames for correct matches to be established. Otherwise, any subsequent tracking is suspect. Therefore, all image processing based approaches for multi-camera image fusion are error-prone to a high extent because of different reasons such as occlusion, anchor point misidentification, and non-coplanarity of corresponding image points, to name a few. In other words, for a growing number of cameras, the existing methods become infeasible for two reasons: (a) The number of required camera-to-camera homographies increases dramatically, which gives rise to the computational complexity; and (b) The information fusion becomes more and more complicated requiring more processing utilities and taking more time.

Therefore, in this manuscript, a new multi-camera image fusion technique is proposed for traffic surveillance systems. In the proposed approach, the geometrical optics concepts used in three-dimensional integral imaging systems are utilized in order to improve the final reconstructed 3D image. To the best of our knowledge, this is the first study in utilizing geometrical optics concepts together with image processing image fusion techniques in order to improve the resultant 3D reconstructed image. The proposed improved multi-camera image fusion technique for traffic surveillance is presented and described in Section 2. The performance of the reconstructed 3D images is compared with that of image processing based multi-camera fusion techniques in Section 3. Finally, the conclusions are made in Section 4.

2. The proposed improved multi-camera visionary system for smart traffic control system

As was mentioned in Section 1, in the existing multi-camera image fusion techniques, calibration and registration of different images, taken from variant viewpoints of a scene, are performed through extracting different image features, such as corners, and mathematically matching them. Thus, the image fusion procedure is completely based on image processing and it is influenced by intrinsic image processing errors. These errors arise from different corner and line detection algorithms and multiplicity of ways in establishing point and line correspondences.

On the other side, optical three-dimensional (3D) imaging techniques are able to extract the 3D image information and thus 3D objects location through the principles of geometrical and diffractive optics from the same set of multi-camera images [16-18]. The main advantage of using optical 3D imaging concepts in addition to image processing techniques is to make use of optical concepts when image processing fails to give accurate response. This improved accuracy is due to non-dependent nature of these processes on the image contents. Furthermore, optical multi-camera image integration techniques do not depend on the image contents, making the process much faster. Integral imaging is perhaps the most powerful optical 3D imaging system based on multi-camera vision and geometrical optics. In integral imaging, the 3D image information could be extracted from the array of 2D images computationally captured with a set of cameras from different viewpoints [16]. In this structure, the 3D information is extracted from the recorded intensities in different directions through the principles of geometrical optics. Given that the captured image of each camera gives the information of a different viewing angle of the 3D object, the information of the third dimension – depth information – would be available in the captured images. Therefore, it is possible to reconstruct the 3D image in each depth distance and in each arbitrary view-angle from the set of multi-camera captured images. These depth slice images are utilized here in order to improve the quality of 3D image fusion technique. This is done through extracting the approximate depth information of each object in the scene from the optically reconstructed depth and view-point slices. From the approximated depth information, the approximate amount of disparity between each pair of cameras could be estimated. These approximate amount of disparity, in turn, helps in better estimation of the camera view angles and thus the correspondences

of image features. More accurate image correspondence means improved 3D image reconstruction. The main steps of the proposed algorithm are explained in the following.

Consider a multi-camera imaging system capturing N images in N different view angles, $(\theta_1, \theta_2, ..., \theta_n)$, named I_1 to I_N, respectively, in the x direction. The camera image is assumed to be in the $x-y$ plane. The general steps in all existing image processing based multi-camera image fusion techniques are as follows: We should first find some seed points such as corners in each images named C_1^i to C_P^i for the ith image, for example through Harris corner detection algorithm [19]. Then, these seed points are used to extract the relative geometrical transformation between each pair of images through extracting the correspondence between them. This is performed through some well-known algorithms such as RANSAC [20]. This correspondence is again derived based on image processing techniques and thus may contain some wrongly detected matched points depending on the image content. Due to the aforementioned fusion deficiencies, the resultant 3D image suffers from some drawbacks. These drawbacks are mainly results of incorrect extracted correspondence between point or line pairs. Thus, in the resultant 3D image, some incorrect parts may be seen between two merged consequent angle images.

To solve this problem, in the next step, the same set of images is used to reconstruct depth slice images in a set of depth values, Z_1 to Z_n, named I_{z1} to I_{zn}, based on geometrical optics relations [16]:

$$I_{z1} = \sum_{q=1}^{N} \frac{I_p\left(-\frac{x}{M} + \left(1+\frac{1}{M}\right)s_x q, \frac{y}{M}\right)}{(Z_1+g)^2 + [(x-s_x q)^2 + (y)^2]\left(1+\frac{1}{M}\right)^2}, \quad (1)$$

where M is the magnification factor equal to z/g, g is the distance between the camera lens and sensor, and s_x is the size of the camera sensor in the x direction. The resultant images are processed based on object detection algorithms to extract the intended objects in each depth image. Object extraction is performed through the well-known pattern recognition algorithm, Support-Vector-Machine (SVM). In this algorithm, a linear classifier, $f(x)$, where:

$$f(x) = W^T X + b, \quad (2)$$

is learned through solving an optimization problem over W. Therefore, each extracted object will have a specific depth value z_i where its image has been reconstructed. In other words, for each object, the approximate depth location will be specified. The corresponding depth values help in specifying the relative camera view angle with regard to each object. For instance, if the extracted depth location of the first object O_1 is z_1, then for the ith camera in location x_i and y_i, the corresponding view angles would be:

$$\theta_{1i,x} = \tan^{-1}(x_i/z_1),$$
$$\theta_{1i,y} = \tan^{-1}(y_i/z_1). \quad (3)$$

Having the camera view angles for each object in hand, the seed point matching algorithm could be implemented for each object with the relative view angle of its specific cameras. Unlike common multi-camera vision approaches where object occlusion may cause error in image fusion process, here, this phenomenon does not have such an effect. In this approach, if any object occludes other objects in the images of some cameras, it will appear in the other view point images and thus will be reconstructed in its corresponding depth location z_i. This could be concluded from geometrical optics relation of the output 3D image and the set of input elemental images shown in Eq. (1). Here, we see that each camera image, I_p, has an equal impact on the reconstructed image in a specific depth distance, z_i. Thus, although the corresponding object is missed in some images of the cameras due to occlusion, it will appear in other view point images and thus I_{zi} will contain the reconstructed image of this object. This reconstructed object will be used in consequent 3D image fusion based on Eq. (3), which prevents the image fusion process to be affected by this obstruction. Put differently, having the relative view angle of each camera with respect to a specific extracted object in depth z_i, the approximate location of each seed point in each camera image could be predicted. Thus, the lost seed points due to occlusion could be easily detected and excluded from the calibration and matching process. Therefore, for each seed point, there is always a set of camera images containing that seed point. Thus, it can take part in the process of image matching and fusion. On the other hand, in ordinary multi-camera visionary systems, if a seed point is missed in a camera image due to occlusion, it should be eliminated from the set of matched seed points. This diminished set of seed points obviously degrades the quality of final image fusion. Thus, the proposed multi-camera image fusion technique will lead to more accurate image calibration and thus more precise seed point correspondence. This more accurate correspondence leads to more accurate 3D image reconstruction. This theory will be verified through simulations in the next section.

3. Experimental simulations

In order to demonstrate feasibility of the proposed method in improving the performance of multi-camera surveillance systems, a simulation scenario has been

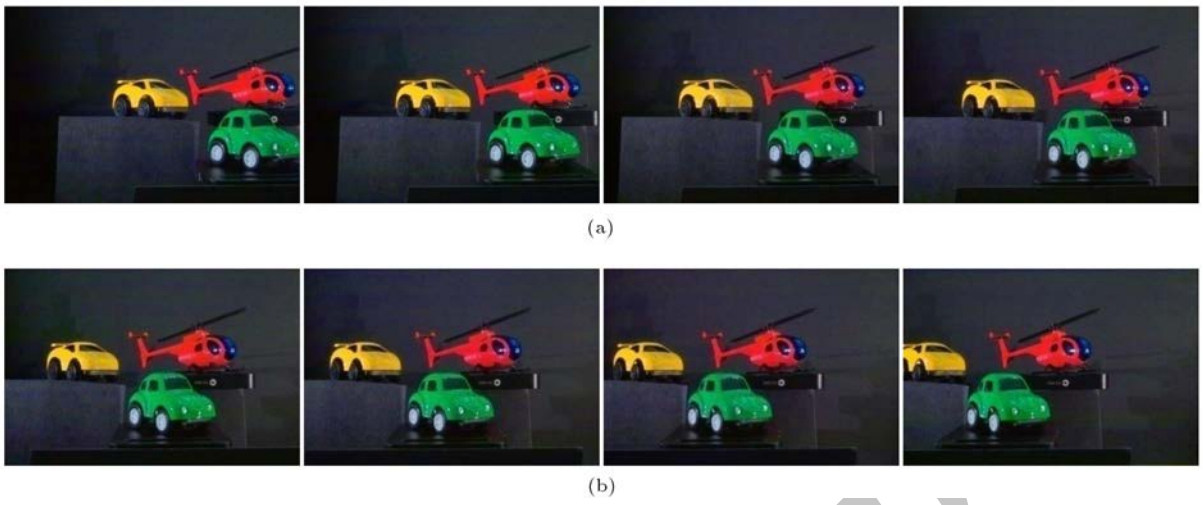

Figure 2. An asset of viewpoint images captured by 8 cameras: (a) The first to fourth images; and (b) the fifth to eighth images.

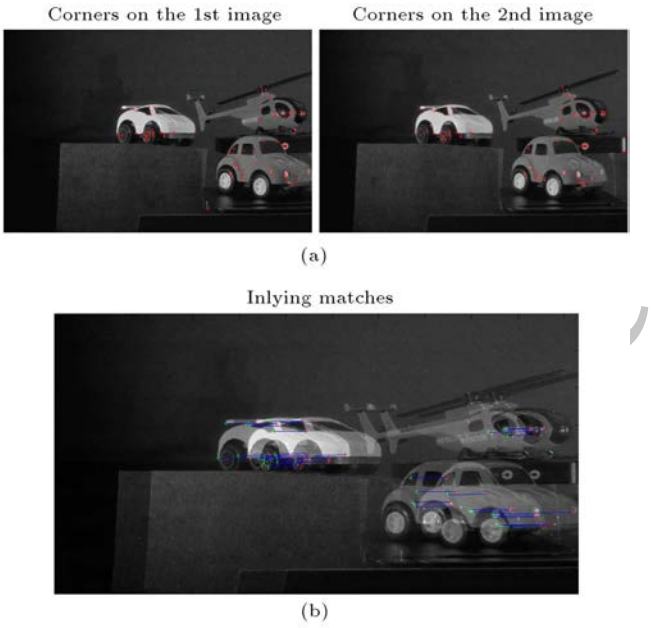

Figure 3. (a) Detected corners in the first image pair, and (b) the correspondence between them.

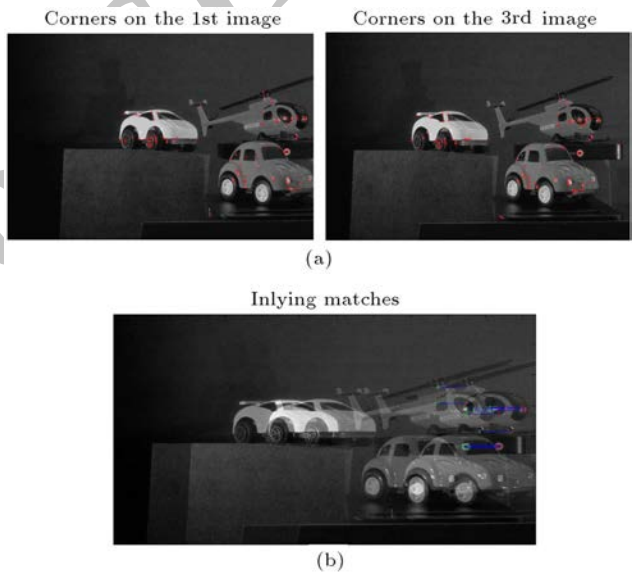

Figure 4. (a) Detected corners in the second image pair, and (b) the correspondence between them.

run on a set of experimentally captured multi-camera images. The set of captured images from 8 horizontal view angles uniformly spaced at $\theta_1 = -30°$ to $\theta_8 = 40°$ is shown in Figure 2.

Let us consider the first pair. In this example of camera image pair, we first extract the corner of objects using Harris corner detection technique. In the next step, the corresponding corner pairs are computed using the well-known RANSAC algorithm [20]. The extracted corners and the corresponding corner points are shown in Figure 3 (a) and (b), respectively.

In the second example, the first and third images in the 8 image set are considered as the camera image pairs. Again, the corresponding corners are extracted based on RANSAC algorithm. The extracted corners and corresponding points are shown in Figure 4(a) and (b), respectively. As could be seen in the figure, in this case, just a few matched corner pairs are detected unfortunately while there are many missed corresponding seed points because of image processing errors.

Therefore, in merging all images to reconstruct the final 3D image, these insufficient correspondences will degrade the quality of the final 3D image. To solve this problem, the 3D image is reconstructed in 8 different depth slices as shown in Figure 5. The corresponding objects in each depth slice have been extracted through object detection algorithms.

Now, with regard to the approximate depth value of each object in the scene and also the lateral location

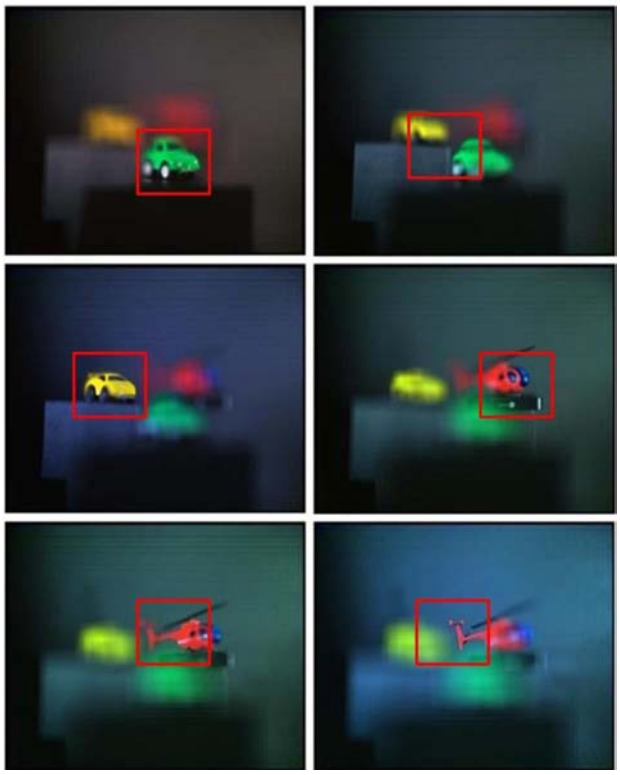

Figure 5. The fused images in different depths based on geometrical optics (integral imaging).

of each camera in the $x-y$ plane, the corresponding view angle of each camera with regard to each object could be extracted through Eq. (2). Therefore, before running Harris corner detection, the objects, and thus the corners, are extracted for each object separately. Then, the RANSAC algorithm for extracting the homogeneity matrix is run for each different object separately with regard to the camera angles extracted for that object through optical depth extraction. These camera view angles in fact determine the size of matching window in the RANSAC algorithm. The resultant matched corner pairs for the same image pairs are shown in Figure 6. As could be seen in this figure, the correspondences of corner points are now determined more completely. For example, the left white car has some corresponding points in two images which are missed in the previous extracted matching pattern.

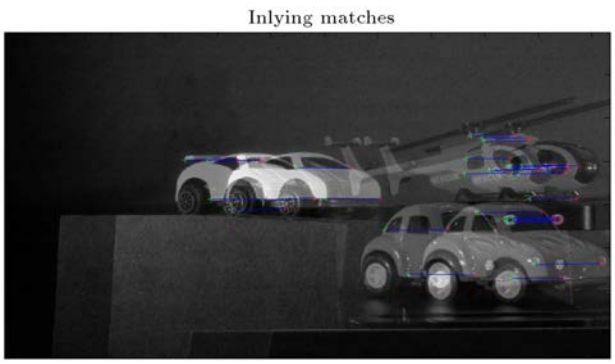

Figure 6. The corresponding corner points between the second pair of camera images extracted through the proposed method.

From these results, it is obvious that the proposed method is more successful than the existing fully image processing based methods in extracting the seed point correspondence between camera image pairs. Consequently, multi-camera image fusion could be performed more efficiently, which in turn helps in more precise extraction of 3D object locations in the scene. Extracting the 3D location of different vehicles is an essential part of smart traffic monitoring and control systems.

4. Conclusions

By employing distributed camera networks, traffic video surveillance systems substantially extend their capabilities and improve their robustness through data fusion and cooperative sensing. With multi-camera traffic surveillance systems, activities in wide areas are analyzed and the accuracy and robustness of vehicle tracking are improved by fusing data from multiple camera views. Multi-camera vision is also advantageous in preparing for and rescuing from the emergency conditions by providing the location of ruined regions and rescue centers.

As the sizes and complexities of camera networks increase, there are higher requirements for robustness, reliability, scalability, transferability, self-adaptability, and less human intervention in intelligent multi-camera traffic video surveillance systems. Recent studies show that different modules actually should support each other. For example, activity modeling can improve inter-camera tracking and multi-camera tracking, providing information for camera calibration and inference of the topology of camera views. Notwithstanding this fact, this interference in turn gives rise to error propagation resulted from different image processing and pattern recognition techniques. In order to lessen these errors, in this paper, a new multi-camera stereo image fusion technique has been proposed, which combines the concepts of geometrical optics and image processing for 3D image fusion. These fusion techniques do not completely depend on image processing and object recognition and tracking. This, in turn, lessens the amount of resulted error propagation and thus improves the accuracy of multi-camera traffic surveillance systems while lessening the computational complexity to a large extent. The feasibility of the proposed approach in improving the extraction of multi-camera image correspondences was shown through running simulations on experimental data.

Acknowledgements

The author gratefully acknowledges financial support from Iran National Science Foundation (INSF). She also acknowledges the research deputy of Sharif University of Technology for supporting this work.

References

1. Tessens, L., Morbee, M., Aghajan, H. and Philips, W. "Camera selection for tracking in distributed smart camera networks", *ACM Transactions on Sensor Networks (TOSN)*, **10**(2), p. 23 (2014).
2. Aghajan, H., Cavallaro, A. and Queen M. "Multi-camera Networks", *Principles and Applications*, 1st Ed., Academic Press (2009).
3. Aghajan, H. and Cavallaro, A. (Eds.), *Multi-Camera Networks: Concepts and Applications*, Academic Press (2009).
4. Valera, M. and Velastin, S.A. "Intelligent distributed surveillance systems: A review", *IEE Proceedings*, **152**, pp. 193-204 (2004).
5. Pollefeys, M., Koch, R. and Gool, L.V. "A simple and efficient rectification method for general motion", In: *Proceedings of International Conference on Computer Vision*, pp. 496-501 (1999).
6. Hartley, R.I. "Theory and practice of projective rectification", *International Journal of Computer Vision*, **35**, pp. 115-127 (1999).
7. Brown, M.Z., Burschka, D. and Hager, G.D. "Advances in computational stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**, pp. 993-1008 (2003).
8. Scharstein, D. and Szeliski, R. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision*, **47**, pp. 7-42 (2002).
9. Bobick, A.F. and Intille, S.S. "Large occlusion stereo", *International Journal of Computer Vision*, **33**, pp. 181-200 (1999).
10. Douret, J. and Benosman, R. "A volumetric multi-cameras method dedicated to road traffic monitoring", *IEEE IV, Parma*, Italy, pp. 14-17 (2004).
11. Lemosa, F., Uchimura, K. and Hu, Z. "Vehicle detection using probability fusion maps generated by a multi-camera system", *Proc. IWAIT* (2007).
12. Hu, Z., Wang, C. and Uchimura, K. "3D vehicle extraction and tracking from multiple viewpoints for traffic monitoring by using probability map", *IEEE Intelligent Transportation Systems Conference (ITSC 2007)*, pp. 30-35 (2007).
13. Wang, J. and Yang, D. "A traffic parameters extraction method using time-spatial image based on multicameras", *International Journal of Distributed Sensor Networks* (2013).
14. Calavia, L., Baladrón, C., Aguiar, J.M., Carro, B. and Sánchez-Esguevillas, A. "A semantic autonomous video surveillance system for Dense camera networks in smart cities", *Sensors*, **12**(8), pp. 10407-10429 (2012).
15. Ferryman, J.M., Maybank, S. and Worrall, A. "Visual surveillance for moving vehicles", *International Journal of Computer Vision*, **37**(2), pp. 187-197 (2000).
16. Hong, S.H., Jang, J.S. and Javidi, B. "Three-dimensional volumetric object reconstruction using computational integral imaging", *Optics Express*, **12**(3), pp. 483-491 (2004).
17. Kavehvash, Z., Corral, M.M., Mehrany, K., Bagheri, S., Saavedra, G. and Navarro, H. "Three-dimensional resolvability in an integral imaging system", *Journal of OSA A*, **29**(4), pp. 525-530 (2012)
18. EsnaAshari, Z.H., Kavehvash, Z. and Mehrany, K. "Diffraction influence on the field of view and resolution of three-dimensional integral imaging", *IEEE/OSA Journal of Display Technology*, **10**(7), pp. 553-559 (2014).
19. Harris, C. and Stephens, M. "A combined corner and edge detector", *Proceedings of the 4th Alvey Vision Conference*, University of Manchester, pp. 147-151 (1988).
20. Hast, A., Nysjö, J. and Marchetti, A. "Optimal RANSAC - towards a repeatable algorithm for finding the optimal set", *Journal of WSCG*, **21**(1), pp. 21-30 (2013).

Biography

Zahra Kavehvash was born in Kermanshah, Iran, in 1983. She received the BSc, MSc, and PhD degrees all in Electrical Engineering from Sharif University of Technology (SUT), Tehran, Iran, in 2005, 2007, and 2012, respectively. She joined Sharif University of Technology, EE Department, in 2013 as a faculty member. Her research interests include optical and millimeter wave imaging devices, three-dimensional imaging systems, biomedical imaging systems, and optical signal processing.