IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 6, 2020, accepted September 13, 2020, date of publication September 18, 2020,
date of current version October 1, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3024568*

# Literature Survey on Multi-Camera System and Its Application

**ADESHINA SIRAJDIN OLAGOKE[1,2], (Student Member, IEEE),**
**HAIDI IBRAHIM[1], (Senior Member, IEEE),**
**AND SOO SIANG TEOH[1], (Senior Member, IEEE)**
[1]School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia
[2]Department of Computer Engineering Technology, Federal Polytechnic Mubi, Mubi 650101, Nigeria

Corresponding author: Haidi Ibrahim (haidi_ibrahim@ieee.org)

This work was supported by the Universiti Sains Malaysia, Research University, under Grant 1001/PELECT/8014052.

**ABSTRACT** A multi-camera system combines features from different cameras to exploit a scene of an event to increase the output image quality. The combination of two or more cameras requires prior settings in terms of calibration and architecture. Therefore, this paper surveys the available literature in terms of multi-camera systems' physical arrangements, calibrations, algorithms, and their advantages and disadvantages. We also survey the recent developments and advancements in four areas of multi-camera system applications, which are surveillance, sports, education, and mobile phones. In the surveillance system, the combination of multiple heterogeneous cameras and the discovery of Pan-Tilt-Zoom (PTZ) and smart cameras have brought tremendous achievements in the area of multi-camera control and coordination. Different approaches have been proposed to facilitate effective collaboration and monitoring among the camera network. Furthermore, the application of multi-cameras in sports has made the games more interesting in the aspect of analyses and transparency. The application of the multi-camera system in education has taken education beyond the four walls of the class. The method of teaching, student attendance enrollment, determination of students' attention, teacher and student assessment can now be determined with ease, and all forms of proxy and manipulation in education can be reduced by using a multi-camera system. Besides, the number of cameras featuring on smartphones is gaining noticeable recognition. However, most of these cameras serve different purposes, from zooming, telephoto, and wider Field of View (FOV). Therefore, future smartphones should be expecting more cameras or the development would be in a different direction.

**INDEX TERMS** Calibration, computerized monitoring, digital camera, educational technology, smart cameras, surveillance.

## I. INTRODUCTION

The capturing of still or moving images is an important step required in object recognition, object behavior analyses, and object monitoring processes. One of the possible methods of capturing an image is with the aid of a camera that could create a single image of an object (i.e., still image) or a sequence of images in rapid succession (i.e., video image). Currently, there are many image acquisition technologies (e.g., cameras). Most of the technologies are built on catadioptric or fisheye cameras, as well as image acquisition systems with static or moving parts. Camera classification or type can be based on one characteristic or another, from the FOV, image sensor, image quality, focusing properties, or power of projection. In this survey, the discussion will be based on unidirectional and omnidirectional cameras. There are six types of cameras considered in this survey; Pan-Tilt-Zoom (PTZ) camera [1], smart camera [2], orthographic camera [3], perspective camera (pinhole) [3], omnidirectional (panoramic) camera [4] and thermal camera [5] as shown in Figure 1.

The Pan-Tilt-Zoom (PTZ) cameras are usually applied in surveillance-based applications because of their high-resolution image output. They have a dynamic field of view and can be configured to monitor a specified area of coverage. PTZ cameras have the capability of zooming far distant object

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

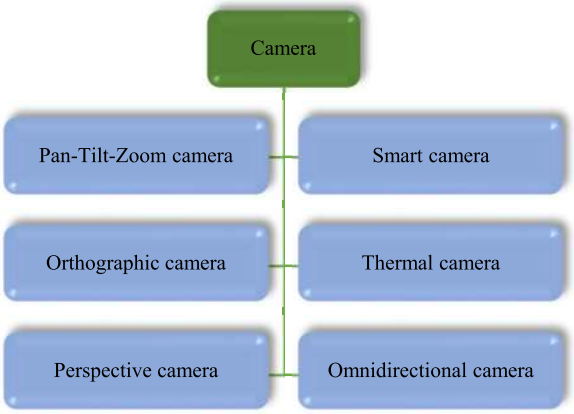


**FIGURE 1.** Types of cameras used in a multi-camera setup.

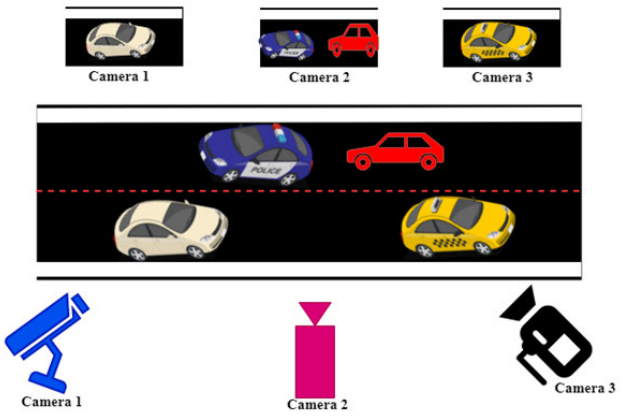


**FIGURE 2.** Heterogenous camera setup used to obtain a full-coverage view of a scene.

to display the detailed information required to analyze a captured image. However, they can only cover a limited area of the scene at one time. Smart cameras are intelligent cameras that came into the limelight around the mid-80s. They have the power of extracting application-based information from a captured image together with creating event information of an image or making an intelligent decision that will be applied in an automated process. At the earliest time of its invention, there was a limitation in its capabilities in terms of sensitivity and processing power, but later there are great improvements in its capabilities [6]. The orthographic camera captures an image without any perspective distortion. They produce a two-dimensional (2D) image output without any image depth. Perspective cameras are cameras that display an image in a real-world view. They produce a three-dimensional (3D) image with depth. All pinhole cameras are also referred to as perspective cameras. Omni-directional cameras can cover 360 degrees FOV with a high-resolution image of about $1600 \times 1200$ pixels [4]. They have the capability of covering images over a wide area FOV. A thermal camera (infrared camera or thermal imager) uses infrared radiation to create an image. It senses infrared light with a wavelength from about $1\mu$m to 14 $\mu$m.

A multi-camera system is an arrangement of sets of cameras used in capturing images or sequences of images of a scene. A multi-camera setup can be homogeneous (consist of the same types of camera setup) or heterogeneous (consist of different types of camera setup) that form a multi-camera system. The combination of two or more cameras can be employed to expand the span of the measuring area and when performing a high-precision measurement. The FOV of a multi-camera setup is more than that of the single camera. The advantages of high accuracy, and low cost of visual measurement techniques made the multi-camera system to be widely used in different areas of application. Figure 2 shows the coverage area of three heterogeneous cameras, which can never be obtained through a single camera.

According to Yilmaz *et al.* [7] and Mehmood [8], the omnipresence of high precision and low-cost cameras, with high computational resources, and the quest for automation in video analyses have generated great interest in multi-camera system developments. The awareness of multi-camera has started around 1884, where Triboulet [9] used multi-camera consisting of seven cameras tightened to a balloon (one camera attached to the mouth of the balloon and six others attached to the circumference of the balloon) mainly to perform aerial imaging. Similarly, in the mid-nineteen century, multi-camera systems were employed in industrial machine automation, where the application of robots taken charge of humans in the industry [10]. Since then, the research on multi-camera developments has been expanding. This can be shown by the number of papers on multi-camera applications. To proof this, a search on IEEExplore was done using the keyword ''multi-camera systems''. The result of this search is shown in Figure 3. The figure shows that the number of publications related to the multi-camera system is in the increasing trend, which indicates that research on the multi-camera system is still very popular in the present day.

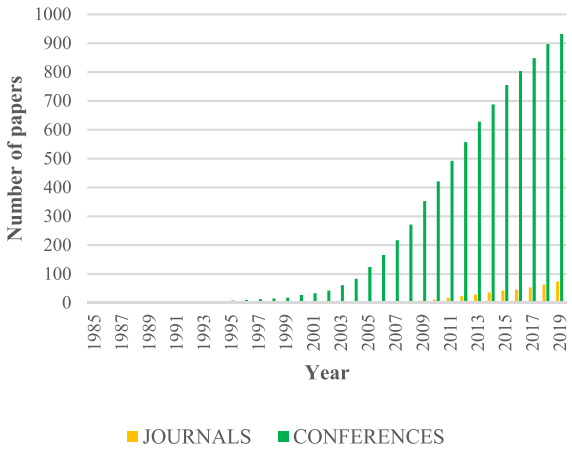


**FIGURE 3.** The number of publications related to the multi-camera system, per year (Data source: IEEExplore, search in October 2019).

Even though several surveys concerning multi-camera systems have been published, virtually all of them focused on a selected aspect of the multi-camera system and particular areas of its applications (e.g., tracking or surveillance) [11]–[17]. These surveys, of course, do not cover the wide spectrum of multi-camera system studies. Because of the literature shortfall, this survey paper comprehensively covers the multi-camera system in four different areas of application, with emphasis given on the previous works, identifying the problems with the existing strategies, the recent advancements and highlight the future direction.

This literature survey is divided into eight sections. Discussion on single and multi-camera systems and the description of their formations and calibration are presented in Section II. Section III surveys the basic architectural formation employed in literature for multi-camera setup with their advantages and shortcomings. Algorithms used in multi-camera systems and their classifications with their various approaches under the concept of person re-identification and tracking fusion are discussed in Section IV. Then, an elaborate discussion on the multi-camera systems in the surveillance applications is discussed in Section V. Description of the multi-camera application, its evolution and image overlay in sports analysis are presented in Section VI. Overview of the multi-camera application and its evolution in the educational system are given in Section VII. An overview of the progress made in the integration of multi-camera systems in mobile phones in the past decade and an attempt on the challenges and future outlooks in this ubiquitous field unraveled in Section VIII. Section IX unveils the various multi-camera application algorithms with their features and limitations. Our conclusion is presented in Section X.

## II. CALIBRATION IN A MULTI-CAMERA SYSTEM

The field-of-view (FOV) of the cameras in a multi-camera setup determines the calibration method to be used and the arrangement of the cameras. There are cameras arrangement with overlapping FOV and non-overlapping FOV. Figure 4 demonstrates the overlapping and non-overlapping FOV. In multi-camera systems with overlapping FOV, it is difficult for all cameras to simultaneously view the calibration target at the same time especially at the overlapped areas. In this type of scenario, the cameras in a multi-camera system are distributed across a wide area coverage. However, calibrating the cameras through the calibration target cannot be easily realizable. For this type of situation, Devarajan *et al.* [18] proposed a distributed algorithm for calibrating distributed cameras on a network. The algorithm determines the position of the camera, its orientation, and the focal length. This approach proffers solutions to complications, memory limitation, and networking constraint, which are commonly found problems with the centralized calibration, by introducing a scalable and parallel algorithm that design a complete framework that senses the visual overlap between cameras and finishes it with an accurate parameter estimate of all cameras on the network [18].

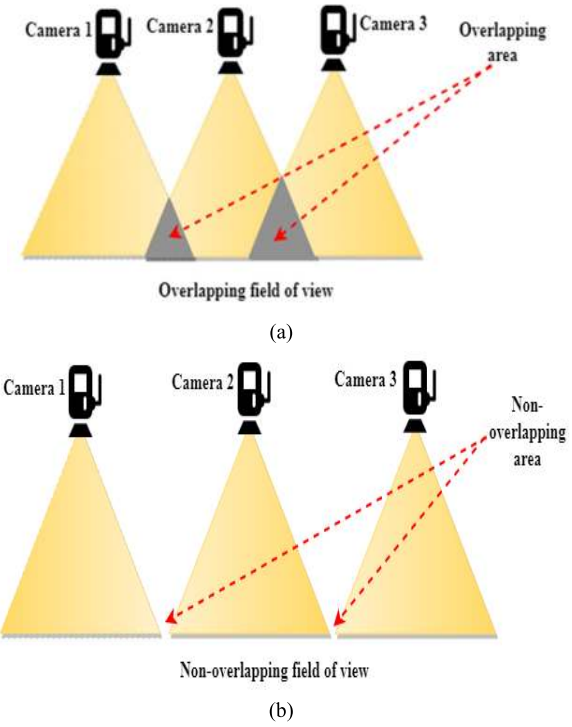


**FIGURE 4.** Multi-camera set up (a) overlapping FOV and (b) non-overlapping FOV.

A practical approach to video network surveillance was presented by Gemeiner *et al.* [19] where cameras are calibrated in a multi-camera system network. They perceived the problem as a localization problem, and then the image was treated as a 3D model assembled with a priori for a moving camera. The cameras are well distanced apart with non-overlapping FOV. Feng *et al.* [20] proposed a novel global planar calibration method whereby a box of translucent glass with one side covered with a pattern was made as a target. The cameras to be calibrated are arranged on both sides of the target glass box. Feng *et al.* [20] adopts Zheng's traditional method [21] to estimate the intrinsic and extrinsic parameters of the camera but addressed the calibration error (glass refraction error) in line with Zheng's method. The glass error was removed using the refractive projection model and ray tracing.

In a multi-camera system, it is critical to find the accurate position for cameras. This will serve as the key to accurate vision measurement for all the cameras [20], [22]. It will also enhance good performance at a collaborative level between cameras on a network and computer vision communication such as tracking of multiple objects together, 3D reconstruction of objects in a scene, and a combination of novel views [18]. The process of finding an appropriate position for a camera to achieve an accurate vision is referred to as camera calibration. Calibration of a camera can be done by determining the parameters of the camera using images obtained from a special calibrated pattern. The parameters are intrinsic (fixed to a camera), extrinsic (may change with

respect to the world frame), and distortion coefficients of the camera [23]. There are many different approaches to calibrate multi-cameras for a specific camera setup. Multi-camera calibration can be classified into two: (i) calibration method for overlapping FOV cameras [22] and (ii) calibration method for non- overlapping FOV cameras [24] as shown in Figure 5.

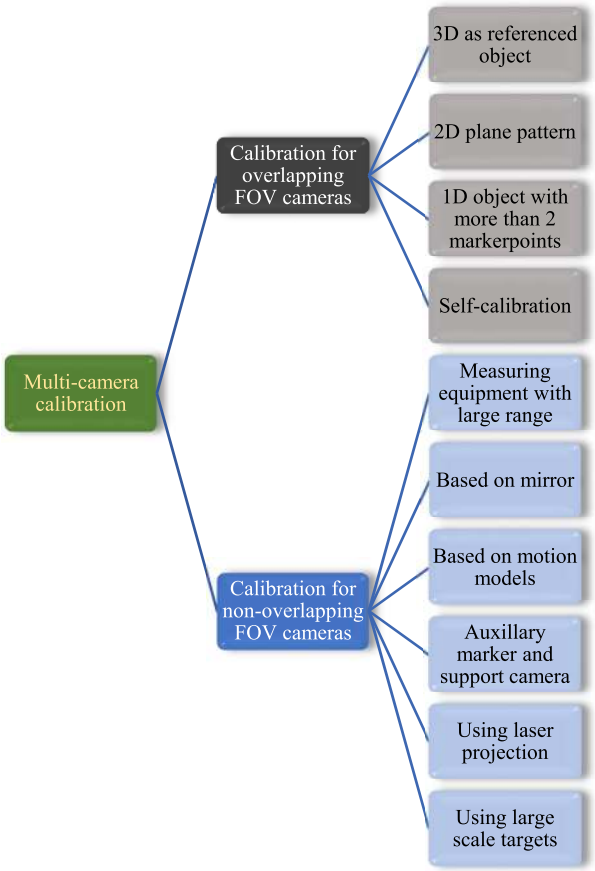


**FIGURE 5.** **Multi-camera calibration methods.**

Multi-camera calibration for overlapping FOV according to Shen *et al.* [22] are categorized into four:

(i) Calibration that is based on the object referenced to 3D [25], they used the calibration of a precise 3D geometry object as their referenced point.
(ii) Calibration based on a 2D plane object [26], it employed a plane pattern specifically designed for that purpose.
(iii) Calibration based on 1D object [27], it was built on a 1D stick with 3 and above points.
(iv) Self-calibration [28], this method was not built on any dimension of an object but uses the static scene to discover the intrinsic parameters of the camera.

Note, this categorization was based on cameras that are connected to a central node apart from self-calibration that is been used in a distributed network. The multi-camera calibration for non-overlapping FOV are classified into six according to Xia *et al.* [24]:

(i) Calibration using large range measuring equipment [29], [30].
(ii) Calibration using a mirror [31], [32].
(iii) Calibration using motion models [33], [34].
(iv) Calibration using auxiliary markers with supporting cameras [35].
(v) Calibration using laser projection [36], [37].
(vi) Calibration using a wide-range of targets [38].

Another calibration system was based on classical techniques used by Haraud *et al.* [39], it employed default patterns with fixed cameras. Smart camera calibrations by Basu and Ravi [40], and Borghese *et al.* [41] have received significant attention in research and development. This is because of their applications in a multi-camera system setup such as in surveillance, tracking, monitoring, and video conference applications. In multi-camera system setup, camera calibration can be performed either in a group (i.e., centralized architecture) or on individual single-camera (i.e., distributed architecture). In a centralized architecture, all the cameras in the multi-camera setup are connected through a central node that runs the calibration process to determine the parameters of all the cameras on that network. For a multi-camera system setup with centralized architecture, both extrinsic (i.e., relative position and direction) and the intrinsic parameters will be determined. In the case of the distributed architecture, the individual camera performs self-calibration which can only estimate the intrinsic parameters of the camera.

A one-dimensional calibration approach has been widely applied in many multi-camera systems [42]–[46]. The 1D calibration approach was first introduced by Zhang in 2002 [47]. In his approach, he used a one-dimensional calibration object that has three and above collinear points with a specified relative position. He demonstrated that to calibrate a camera, a moving one-dimensional calibration object cannot be used except it is stationary. To get an accurate result, he used the maximum likelihood (ML) approach to refine the estimate and performed computer simulation with real data obtained to test the algorithm which produced an encouraging result. One-dimensional camera calibration does not need 2D and 3D directions of marker to calibrate. It represents the most simplified method of calibration. One of the advantages of one-dimensional calibration in a multi-camera system is that they can be observed by all cameras and all the cameras in the system can be calibrated together simultaneously at the same time. Simultaneous calibration of cameras prevents error accumulation. However, 1D calibration has its peculiar disadvantages [48]:

(i) Due to assumptions taken in the model applied, the exact linearity of points cannot be ascertained.
(ii) The level of accuracy obtained in the extraction of the corner is below that of the 2D calibration method due to the tools employed in extracting the points.

A novel nonplanar target for fast calibration of a networked visual sensor was proposed by Shen and Hornsey [43]. They used two spheres built on a supporting rod as a calibration

target. The nonplanar target was applied to each camera separately to save time instead of applying to all the cameras simultaneously at once. A three-dimensional calibration method was proposed by Shin and Mun [49]. He used a direct linear transform (DLT) to determine the calibration parameters of cameras using a 3-axis frame. The main problem with his approach was the error generated due to ellipse fittings which were caused by lighting conditions and noise. Another shortcoming of the approach was that the precision achieved at the center extraction which could not be gotten at the edges [48]. Multiple planar patterns of three-dimensional calibration targets were proposed by Quan and Lan [50] and Xu *et al.* [51] because of its limitations in the way the multi-camera would be distributed, thereby limits its application.

For non-overlapping FOV camera calibration, Lu and Li [29] proposed a global calibration method that employed theodolite coordinate measurement system (TCMS). The system determines the 3D coordinates of the points on the calibrating target. Then, the global calibration of the cameras was performed in relation to the calibrating target position relative to that of the cameras. It employs the transformation matrix between TCMS and the cameras. Liu *et al.* [38] developed a global calibration method that uses multiple targets (MT). One sensor from the multiple vision sensor was used to obtain a global coordinate frame (GCF). The MT was placed in front of the selected sensor to capture the images of the corresponding sub-target for some time (at least four times). The coordinate frame for each sensor to GCF was used to obtain the transformation matrix.

In another approach, Lébraly *et al.* [31] used a planar mirror to calibrate two non-overlapping cameras placed on a vehicle for visual steersman ship. In this approach, the geometry of the scene was not determined. Xu *et al.* [32] proposed a multi-camera global calibration that unified the coordinates of two binocular vision pairs with non-overlapping views using a planar mirror. The method was then applied to obtain the global unification of the multi-camera system. Huang *et al.* [33] used a moving robot carrying a planar target to calibrate two fixed non-overlapping cameras. The relative position of the moving robot, the marker placed on it, and the image captured by the multi-cameras were used to calibrate the cameras. This type of approach can be applied in a large camera network because of its simplicity, low cost, and ease of implementation but the accuracy of the calibration will be very low. Wang and Liu [34] presented a tractable mechanism for choosing the best path a calibration target can take to enhanced the result in a camera network calibration.

Zhao *et al.* [35] used the chessboard augmented reality (AR) marker and supporting camera to calibrate non-overlapping cameras. The transformation was performed using the AR marker and the supporting camera to estimate between the marker and the cameras to be calibrated. The effectiveness of this method depends on the resolution of the supporting camera which will be seriously impaired when calibrating a well-spaced camera network.

Zou and Li [36] used a laser projection method to calibrate inward and outward-facing cameras in a vehicle. A laser pointer was mounted on the calibrating target. The two cameras were connected using the laser rays from the pointer and the pose of the ray was determined using the calibration target boards' coordinates. The problem with this method is setting the cameras in-line with the laser ray on the calibration target to avoid linearity error and this will affect the accuracy of the method. Xia et al. described a global method of calibrating multi-camera without overlapping FOV. Two planar targets are fixed together by a bar length equal to the distance between the positions of the two cameras that are to be calibrated. A photogrammetry method was used to determine the relative position of the planar targets. The initial transformation matrix for the two cameras coordinates were determined using a linear method which requires a single captured image by the two cameras. A global calibration for multiple captured images was obtained using the Levenberg-Marquardt nonlinear (LM) optimization method where the linear method's output served as the input.

## III. MULTI-CAMERA SYSTEM ARCHITECTURE

This is a carefully designed structure of cameras in multi-camera system formation. A multi-camera arrangement could be of different forms, as obtained in different works of literature. They can be classified as (i) Centralized (ii) Distributed (iii) Hybrid, and (iv) Multi-tier [52]–[54]. In a centralized multi-camera architecture, analysis and accumulation of data are performed at the central unit whereby all the detailed information gathered by the individual cameras will be sent to a central system, which is usually a work station or a camera node. In this type of arrangement, no autonomous decision or processing will be done by any of the distributed cameras. The function of the camera could either be any or all of the followings: (i) integrating the image or data collected from the distributed cameras on the network, (ii) taking in or dissecting the information gathered from the subordinate cameras and (iii) controlling of other cameras as a means of a remote access point to them [11]. Many approaches have used this method, such as Kang *et al.* [54], Lim *et al.* [55], Kattnaker and Zabih [56], Lu and Payande [57], Evert *et al.* [58], Sommerlade and Reid [59], and Piciarelli *et al.* [60]. All the cameras are connected to the central work station. Therefore, when two or more cameras are exchanging data, it must go through the central work station that connected them. Figure 6 (a) shows the connection setup of centralized architecture.

In a distributed architecture, each camera does the processes locally. Smart cameras that have the capability of sensing, processing digital signal, and communication components are usually employed in a distributed architecture. This approach was used by the following researchers; Kim *et al.* [61], Quaritsch*et al.* [62], Rinner *et al.* [63], Fleck and Strasser [64], and Fleck *et al.* [65]. A distributed network could also be a PC based method where the distributed cameras are seen as an autonomous body on

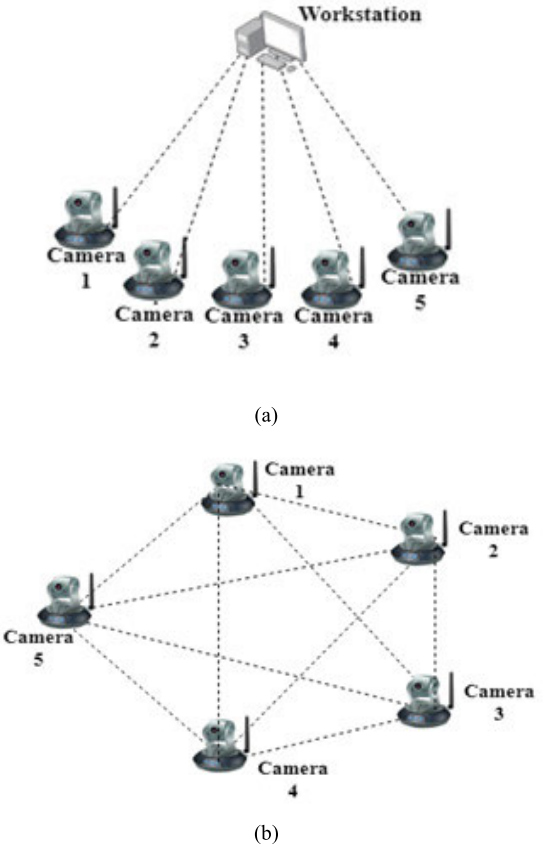


**FIGURE 6.** Multicamera system architecture (a) Centralized [11], and (b) Distributed cameras.

the network. Examples of this type of arrangement can be found in Qureshi and Terzopoulos [66], Micheloni *et al.* [67], Morioka *et al.* [68], Park *et al.* [69], Hodge and Kamel [70], Li and Bhanu [71], and Song *et al.* [72]. Figure 6 (b) shows the diagrammatical illustration of distributed architecture.

The integration of the features of centralized architecture and distribution architecture form the hybrid architecture. In this case, the subordinate cameras perform monitoring and capturing of the target image and at the same time processing the image before sending it to the central system or workstation. The workstation articulates all the raw data or images gotten from the subordinate cameras to form the information. Subordinate cameras make certain decisions at their level while the higher-level decision would be left for the central controlling system or workstation. Prati *et al.* [73] used hybrid architecture in an integrated multi-sensor setup. Mult-tier architecture can also be called hierarchical architecture in the sense that the level of the decision making depended on the level of the hierarchy of the subordinate cameras. Multi-tier architecture was applied by Matsuyama and Ukita [74] and Bamberger *et al.* [75] as an alternative to hybrid, centralized and distribution architectures. Figure 7 shows a diagrammatic illustration of hybrid and multi-tier architectures.

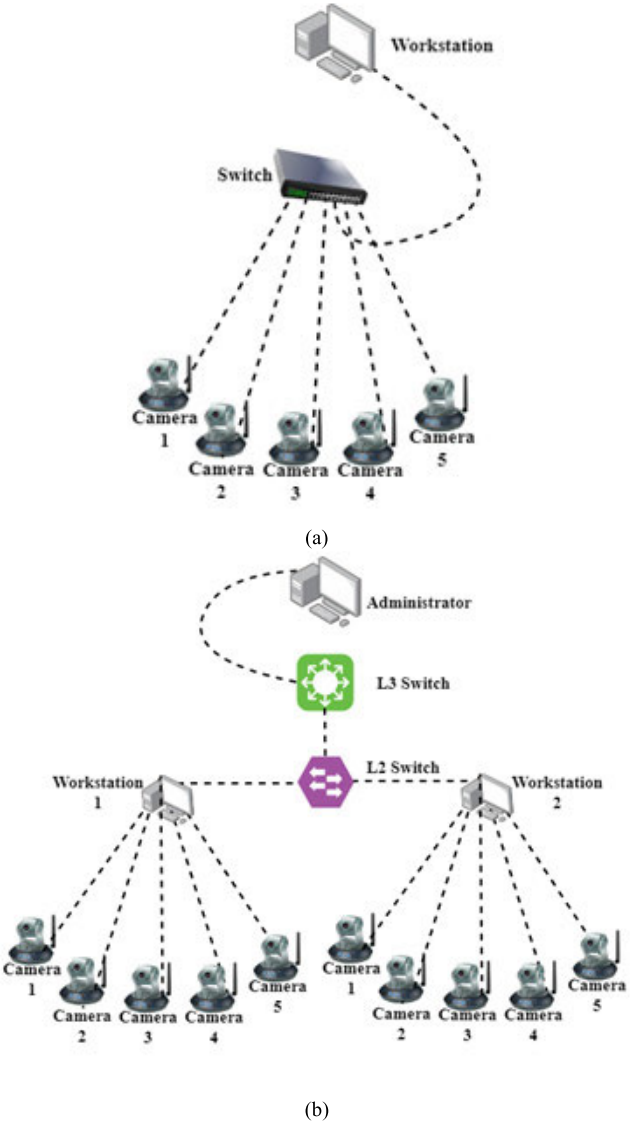


**FIGURE 7.** Multicamera system set up in (a) Hybrid architecture, and (b) Multi-tier camera architecture.

The advantages and disadvantages of these architectures are stated in Table 1.

The ubiquitous nature of cameras made it applicable in every aspect of human endeavor except in some areas where it has not been exploited. In this paper, the application of the multi-camera system is centered on the following areas: surveillance and tracking system, sports and entertainment, education, and mobile phones.

## IV. MULTI-CAMERA SYSTEM ALGORITHMS

Human tracking is seen as automatic monitoring of the trajectories of an individual and the record of the timely ordered sequence of location data of the tracked person. Tracking using a multi-camera system provides complete coverage of human movement from many angles of view. Therefore, it produces comprehensive information on the scene which

**TABLE 1. Advantages and disadvantages of camera architectures.**

| Camera Architecture | Advantages | Disadvantages |
|---|---|---|
| Centralized | • Easy to control because all the subordinate cameras take instruction from a single source.<br>• There is a collaboration among cameras.<br>• It has a simple hardware arrangement. | • It has a bottleneck communication path. Therefore, it needed a bandwidth increment.<br>• The central system (i.e. workstation) that connects all the cameras should have higher computational capability.<br>• No redundant path. |
| Distributed | • No bottle-neck communication path. Therefore, requires less bandwidth.<br>• It has many redundant paths.<br>• It requires less computational time. | • Less collaboration among the cameras. |
| Hybrid | • It combined the advantages of both centralized and distributed architecture.<br>• Faster in terms of processing and communication. | • It forms a huge task when it involves multiple cameras. |
| Multi-tier | • It has a scalable network.<br>• It has a redundancy communication path.<br>• It reduced the number of cameras require and has better object monitoring and control. | • A calibration of cameras may be necessary.<br>• It consumes more time than necessary. |

solves some of the single-camera problems such as occlusion, FOV limitation, and temporary disappearance [76]. A multi-camera tracking can be online (in real-time) or offline (in a stationary image(s) and motion pictures) [77]. In a multi-camera system tracking, the correspondences of the tracked individuals across the multiple cameras are very important unlike that of a single camera that is ignored. Most of the algorithms used in multi-camera tracking are adopted from single-camera tracking with a modification that caters to multiple cameras. There are many approaches adopted from literature towards classifying a multi-camera tracking system, some are based on cameras' FOV [17], motion detection, and camera architecture [16]. Figure 8 depict the multi-camera tracking classification.

Multi-camera tracking using overlapping FOV was employed by many researchers. For example, Cai and Aggarwal [78] used the intensity and geometry features to track people in an overlapping multi-camera setup. Multivariate normal distribution was used to determine the

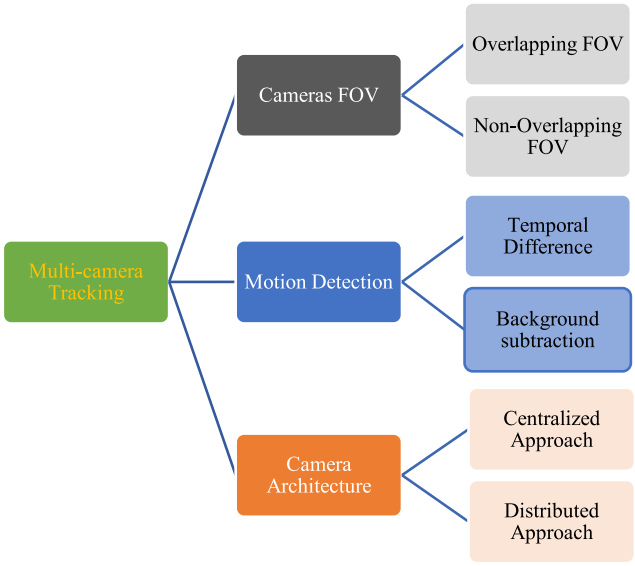


**FIGURE 8. Multi-camera tracking classification.**

densities of the features and Bayesian classification schemes for matching points. Liem and Gavrila [79] presented multiple persons tracking systems. The three-body region (head-shoulder, torso, and legs) is projected using overlapping camera views. A color histogram algorithm was used to describe the appearance and the tracking measurement was modeled using the Kalman filter. Calderara *et al.* [80] used overlapping multi-cameras to track objects using consistent labeling methods. The position of the object in the multiple views of the cameras is determined using the homograph of the first detection of the camera view. Generally, most of the above-mentioned tracking methods, appearance, and motion modeling was carried out using a Kalman filter and Bayesian network. Kalman filter allows more precise motion estimation especially for objects that are viewed by different cameras at the same time. The main setback of the Kalman filtering based approach is the arbitrary assumption of a linear dynamic system and posteriors as Gaussian problems in nature. The pattern of motion demonstrated by people are generally non-Gaussian in nature. This can be addressed by the Extended Kalman filter (EKF) which presumes linearization of a model and approximates the posterior estimate to be Gaussian [81]–[83]. Another setback for the Kalman filter is how to define the strategy for identifying sinking nodes for a distributed multi-camera network.

Tracking objects within overlapping FOV cameras is only possible for objects moving within the vicinity of the cameras' view. Non-overlapping camera tracking is required in blind areas that are out of the cameras' FOV coverage. Makris *et al.* [84] proposed a topographical model of a multi-camera network that performs unsupervised learning of activity around the camera view and then creates data from the large set of observations. The learning experience from data created is used for tracking targets across the whole area of the camera network. The method has no reference

for correspondence, entry, and exit of the target from a camera's view. They are learned using the expectation-maximization algorithm and the zones are formed by a Gaussian Mixture Model (GMM). Stauffer [85] proposed a linking structure for non-overlapping cameras. This method established a technique of modeling correlations based on inter-arrival times instead of the constant rate assumption taken by Makris *et al.* [84]. It also creates a means of estimating the validity of a link based on the total distribution instead of using the statistical value. In addition to these approaches, Rahimi *et al.* [86] presented a method of recalculating the path taken by a target within the view of non-intersecting cameras and then reconstruct the calibration of the cameras. The location of the moving target was determined using the ground plane coordinates of the series of cameras field of view on the network. The performance of the technique was not tested on the online application. Zhang *et al.* [87] presented a framework for tracking multiple targets in a multi-camera setup. They applied the Structural Support Vector Machine (SSVM) to obtain status information of the targets as a group or individual. The tracking of multiple targets involved was converted to a network flow problem which was resolved using the K-shortest paths algorithm. Tesfaye *et al.* [88] tried to solve the problem of tracking an object in multiple non-overlapping cameras with a single three-layer approach. A version of standard quadratic optimization (constrained dominant sets clustering) was used to solve the tracking problem. An algorithm based on dynamics in evolutionary game-theory was used to solve the inter-camera tracking which is performed by merging the tracks of the target in all cameras.

The motion detection approach of multi-camera tracking is classified into temporal difference [89] and background subtraction [90]–[92]. The temporal difference technique is the subtraction between two consecutive frames and then set the output on thresholding value. The resulting pixels with the difference value higher than the threshold are referred to as the foreground pixels. This approach has a setback on changing background with time, therefore, it does not accommodate overlapping parts of the camera while detecting moving persons. The background subtraction approach is a technique of removing a background model from the frames of a video scene to determine the foreground pixels. This technique requires an update of the model for every change in the video scene. Also, it is generally employed in a fixed multi-camera tracking setup [93]. One of the researchers that employed background subtraction is Senior *et al.* [94], where they used 2D and 3D ground plane to determine the positional information of a tracked person and other three different algorithms. They compared the tracking performance of four different algorithms that is background subtraction, face detection-based tracker, feature matching particle filter, and edge alignment of a cylindrical model. The background subtraction method uses multiple Gaussian color tracking methods. The particle filtering tracker used frame differencing which is susceptible to noise and distraction. The face detection approach depends on the faces detected to track and the accuracy of the approach can be seriously affected in the case of occlusion, distance from the camera, and light intensity. Lastly, the edge-alignment-based tracker uses a 3D graphical human model designated with cylinders connected using kinematic chains. The method experiences failure after initialization which requires a re-initialization strategy. Liang *et al.* [95] presented multi-camera collaboration using head detection and the trifocal tensor pointer transfer method. They used Kalman and PDA algorithms for tracking people and then applied background subtraction to detect the head position.

Classification based on camera architecture is based on how data acquired by the multi-camera sensors are been processed. There are two categories; in the first category data detection and tracking is carried out after fusing the information acquired from different sensors. This is referred to as a centralized approach. In the second category, called a distributed approach, each camera on the network performs their detection and tracking separately and the output of the cameras is combined to obtain the final trajectories of the tracking. Figure 9 depicts the representation of the classification.

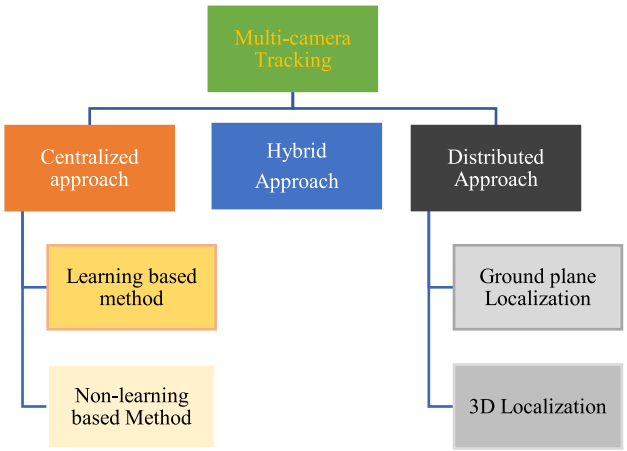


**FIGURE 9.** Classification of multi-tracking base camera architecture.

### A. CENTRALIZED APPROACH

In this approach, the data obtained from the different cameras on the network are first combined before performing the tracking of the people. The data obtained from each camera are usually preprocessed before been fused. Data fusion is used to obtain scene information from different camera views. Some used data fusion to establish ground plane occupancy map, for examples, the works by Khan and shah [96], Figueira *et al.* [97], Baltieri *et al.* [98], Minh *et al.* [99], and Chen *et al.* [100]. Others used it to establish 3D geometry of the tracked person on the scene such as the works by Li *et al.* [101], Hirzer *et al.* [102], Bouma *et al.* [103], Wang *et al.* [104], Wen *et al.* [105] and Brendel *et al.* [106].

A centralized multi-tracking approach has a well-coordinated and control center which makes the system to be more effective. Some of its limitations include cameras synchronization, redundancy problem and it is not suitable for an online application. Figure 10 shows a diagrammatic representation of the distributed approach. The centralized approach is usually applied in overlapping multi-camera setup to reduce noise and occlusion problems. Table 2 gives a summary of the centralized approaches used by different researchers.

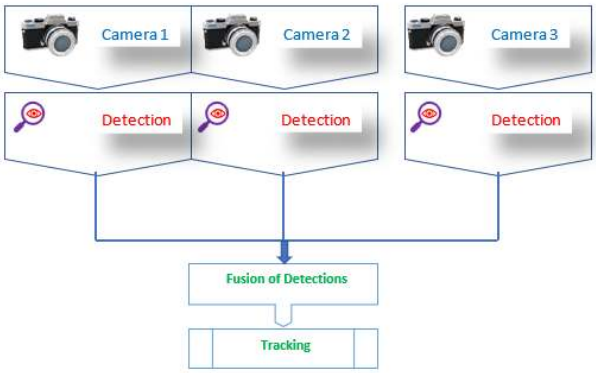


**FIGURE 10.** Centralized multi-camera tracking.

**TABLE 2.** some of the centralized approach in literature and the methods.

| Authors | Feature | Data fusion method |
|---|---|---|
| Huang and Wang, 2012 [107] | • Binary map | • Plane occupancy map |
| Khan and Shah, 2006 [96] | • Binary map | • Homography constraints |
| Liem and Gavrila, 2014 [108], Guan et al., 2010 [109] | • Color appearance. | • 3D reconstruction |
| Kim and Davis, 2006 [110], Du and Piater, 2006 [111]. | • Binary map | • The intersection of the central vertical axis |
| Yao et al., 2008 [112] | • Color appearance. | • 3D model |
| Hofmann et al., 2013 [113] | • Color Appearance | • 3D model projection |
| Jiang et al., 2014 [114] Balteiri et al., 2011 [98] | • Color appearance | • Deformable part model and 3D project |
| Wen et al., 2017 [105] | • Appearance and motion | • 3D position |
| Du and Piater, 2006 [115] | • Color appearance | • Probabilistic method |

### B. HYBRID APPROACH

This approach balanced the limitations and advantages of centralized and distributed approaches. It allows the creation of multi-cameras in groups to form clusters where detection, tracking, and data fusion are done within the clusters of cameras using a centralized approach. Person re-identification is carried out by combining the feature signatures obtained from different clusters on the camera network to create the trajectories of the tracked persons. Among the studies that embraced this approach is Kim *et al.* [61], where they built a clustered wireless network of cameras that employed a Kalman filter to track people. Information communication is tracked based on the cluster location of the camera which determines the position of the tracked person. The cluster head of each cluster in the network then transmits the gathered information to the central station where the trajectories of the tracked individuals will be built. Besides, Agrawal and Davis [25] have proposed a hybrid formation when extracting features from images. They used an occupancy map (ground plane) at the fusion node point using the centralized approach and then performed the ground plane color mapping for tracked persons using a color likelihood for each camera in a distributed approach. The color functions are then combined at the fusion node to produce a color multiview function. The trajectories of the tracked people are obtained by combining the result of the occupancy map and color map of each individual.

### C. DISTRIBUTED APPROACH

The distributed approach consists of autonomous camera nodes that perform detection and tracking of people independently and then collate the correspondences of the tracked individuals from the different cameras' views. The collation of the correspondences is done in such a way that the trajectories of the tracked persons obtained in the first camera are compared with the correspondences in other cameras before fusing them to obtain the final trajectory. The image obtained from the first camera used to compensate for images obtained from other cameras is called the probe image and the process is referred to as person re-identification (Re-ID). The distributed approach is generally applied in a non-overlapping multi-camera setup. Figure 11 shows a diagrammatic representation of the distributed approach.

The major advantage of the distributed multi-tracking approach is its scalability which makes it to be easily deployable in a disjointed multi-camera set up (Non-overlapping setup) without reference to the geometric relationship between the cameras. It is suitable for an online application. The major setback of this approach is that the failure of one camera renders the section of network non-trackable and also occlusion is difficult to manage using this approach. Table 3 shows a list of studies that applied the distributed approach.

Recognizing and tracking people across spatially disconnected cameras requires the knowledge of the transition from one camera to another. Person Re-ID provides a means of identifying people across disjointed cameras. Person Re-ID can be classified into two; (i) learning-based and (ii) non-learning based method.

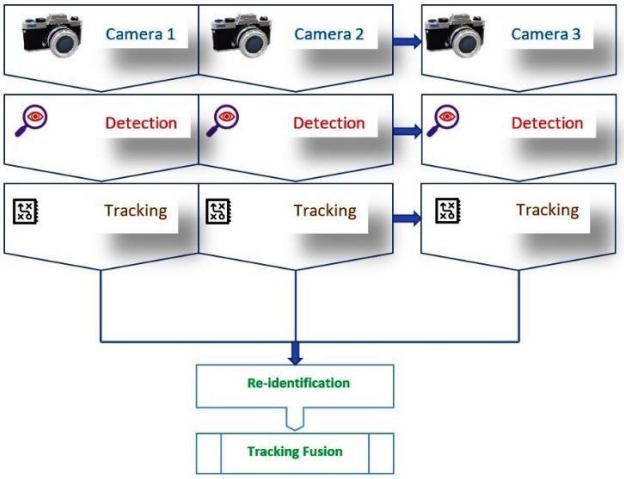


**FIGURE 11.** Distributed multi-camera tracking.

**TABLE 3.** Some of the studies that applied distributed.

| Authors | Feature | Data fusion method |
|---|---|---|
| Brendel et al., 2011 [106] | • Position, size, color, histograms, intensity, gradient, motion | • MWIS algorithm |
| Bouma et al., 2012 [103] | • Color and spatial information | • Histogram intersection |
| Chen et al., 2014 [116] | • Color spectrum. | Tracklet association |
| Li et al., 2014 [117] | • HSV, HOG, spatial information. | • Graph matching |
| Mazzon and Cavallaro, 2013 [118] | • Position and velocity | • Social force model |
| Li et al., 2012 [101] | • Color Appearance | • 3D model projection |
| Cong et al., 2010 [119] | • Color appearance | • 2D model and k-NN |
| Javed et al., 2008 [120] | • Color appearance | • Epipolar geometry |
| Bak et al., 2010 [121] | • Covariance descriptor | • SVM |
| Farenzena et al., 2010 [122] | • Appearance features | • The asymmetric and symmetric axis of the human body model |
| Prosser et al., 2010 [123] | • Color, texture | • Rank SVM |
| Bazzani et al., 2012 [124] | • histogram and epitome | • K-NN |
| Avraham et al., 2012 [125] Kenk et al., 2015 [126] | • HOG and color histogram | • SVM |

### 1) LEARNING-BASED RE-ID

Person Re-ID in this approach is based on learning the distinct features of an individual from a dataset, extract the features which are then used to re-identify the person from the image of the tracked people. What differentiates the learning-based approaches is the learning metrics (classifier) that are used to identify or re-identify a person. Some of the metrics or classifiers that are normally been utilized are the shape model [127], boosting method [128], deep learning [123], [129], support vector machine (SVM) [130], [131], descriptors learning methods [132]–[135], distance learning methods [136], [137], nearest neighbor approach [108] and least square reduction [138].

A learning algorithm was used by Bak *et al.* [128] for person Re-ID based on the Adaboost and it uses Haar-like features. They employed histogram of oriented gradients (HOG) for tracking and detection. After detections and tracking of people, the Haar-like features of the tracked people are used to obtain the visual signature using the Adaboost scheme. The extracted visual signature was used to re-identify people in other camera detections. Su *et al.* [139] presented 3 stages of deep CNN learning metrics for person Re-ID, while Chen *et al.* [116] used a deep learning approach that employed the fusion pyramid multi-scale metric for Re-ID.

Another learning approach is the descriptors-based type. The approaches under this method combine the appearance and shape features to track people. The learning metric is used to identify the features to be extracted from the tracked individual. Huang *et al.* [33] presented a discriminative approach that used an algorithm to learn and extract appearance-based descriptors from the body of each tracked person. The extracted features are used to create a model using a multi-scale feature learning framework obtained from the work of Qian *et al.* [140]. Also, Kuo *et al.* [141] have presented an approach that used a discriminative appearance learning feature together with the Adaboost algorithm for a group of appearance features extracted from tracked individuals in a different location. The obtained discriminative feature is used to identify who has a close affinity to a tracklet.

Avraham *et al.* [125] introduced an algorithm that used correspondences of two people obtained from two overlapping camera views to train. The algorithm can differentiate or identify two tracked individuals using color histograms which shows positive pair if the detected people in the camera views are the same and a negative pair for two different individuals. The binary SVM classifier was used to differentiate between the two detected pairs. Similarly, Nakajima *et al.* [142] used a multi-class SVM classifier to detect the full-body of a person. The SVM algorithm classifier was trained using colored and shape-based features obtained from the tracked individual. Another study was done by Martinel *et al.* [143] that presented a method for training multiple re-identifications using features obtained from pairs of images for the same or different individuals. The model is used to determine from the new pairs of images obtained depending on the result of the models.

A learning-based distance approach used similarity measures in terms of the distance between pairs of images or features of an individual in different camera views. The distances (differences) between images that represent the same

person (positive pair) and that of different persons (negative pair) are used for Re-ID instead of learning the individuals' visual features. One of the studies under this method was proposed by Hirzer *et al.* [136] where the concept of person Re-ID is classified into three stages; (1) feature extraction (2) metric learning and (3) classification.

Features like color, Haar-like, shape-based, body parts and texture are distinct characteristics for human identification. HSV, Lab color channels, and Local Binary Patterns can be used to extract the local features to create a global image representation. At the learning stage, an algorithm (for example PCA) is used to reduce dimensionality and noise. However, a small dataset or low dimensional representation is enough. During the evaluation, the distance between two samples for example $X_m$ *and* $X_n$ are calculated, where $X_m$ *and* $X_n$ describe the camera views of the same person. At the classification stage, the work is to find the image (probe image) that was first obtained from one of the cameras' views in all the images (gallery images) obtained from all other cameras. Similarly, Makris *et al.* [84] used relative distance comparison learning between a pair of true images and a pair of false images match to perform re-identification. The comparison model was used on the texture and color histograms extracted features of the tracked people.

Gou *et al.* [144] presented a comprehensive performance evaluation of person Re-ID (feature extraction and metric learning) algorithms. Table 4a and Table 4b summarize the algorithms. It was demonstrated that kLFDA and kMFA were the best metric learning methods because they resolve the problems of eigenvalue in a scattered matrices data and the best algorithms for feature extraction to be GOG, LDFV and LOMO as shown in Table 4b.

Finally, Shen and Hornsey [43] presented an approach that is based on space-time and the appearance between every two cameras view. It is used to determine whether the resulting output of two camera views will be the same or different. The appearance relationship is made up of the BTF between the two cameras' view while the information about the entrance point, exit point, direction of movement, and the required time for the tracked person to move from one camera view to the other is in the space-time correspondence.

### 2) NON-LEARNING-BASED RE-ID

These are person Re-ID methods that directly extract discriminative features from a tracked person without any training data. Non-learning-based Re-ID relied on the signature used in representing the tracked person, which can be extracted by local or global features. Appearance features are the common features employed in non-learning person Re-ID which include shape, texture, and color. One of the relevant studies that embraced this method is the work of Aziz *et al.* [153] that used the front or back appearance of the tracked individuals to perform person re-identification. They used SIFT and SURF algorithms to extract the discriminative signature of the segmented three-body level (i.e head, legs, and torso). For peoples' re-identification, the extracted signatures are matched with the body parts of the detected individuals. The effectiveness of the approach can be improved by employing a pose estimation approach, for example, the one presented by

**TABLE 4. (a) Re-Id metric learning algorithms. (b) Re-Id feature extraction algorithms.**

(a)

| Metric Algorithms | Type | Source |
|---|---|---|
| Discriminant analysis | • Fisher Discriminant Analysis (FDA) [133].<br>• Kernel Local Fisher Discriminant Analysis (kLFDA) [134] | • AE 1936<br>• ECCV 14 |
| Discriminant analysis | • Local Fisher Discriminant Analysis (LFDA) [135] | • CVPR 13 |
| Discriminant analysis | • Manifold Discriminant Analysis (MFA) [145].<br>• Kernel Marginal Fisher Analysis (kMFA) [145]. | • PAMI 07<br>• ECCV 14 |
| SVM | • Support Vector Machine Metric Learning (SVMML) [131].<br>• Ranking Support Vector Machine (RankSVM) [123]<br>• Regularized Pairwise Constrained Component Analysis (rPCCA) [134].<br>• Pairwise Constrained Component Analysis (PCCA) [146].<br>• Kernel Pairwise Constrained Component Analysis (kPCCA) [146]. | • CVPR 13.<br>• BMVC 10<br>• ECCV 14<br>• CVPR 12<br>• CVPR 12 |
| Statistical inference perspective | • Keep it simple and straightforward metric (KISSME) [137]. | • CVPR 12. |
| Discriminant analysis | • Cross-view Quadratic Discriminant Analysis (XQDA) [147]. | • CVPR 15. |

(b)

| Feature Extraction Algorithms | Features | Source |
|---|---|---|
| An ensemble of Localized Features (ELF) [148] | • Extract color features from the color channels (RGB, YCbCr, and HSV).<br>• Extract texture features Kernel from Schmid and Gabor filters outputs. | • ECCV 08 |
| Local Descriptors encoded by Fisher Vector (LDFV) [149] | • It uses local pixel-level information encoded by the Fisher vector representation. | • ECCVW 12 |
| Gabor filter Biologically inspired Covariance (gBiCov) [150] | • It uses multi-scale biological-inspired features encoded covariance descriptor | • BMVC 12 |
| Dense Color SIFT [151] | • It employs dense SIFT features. | • CVPR 13. |
| HistLBP [134] | • It replaces the outputs of Schmid and Gabor filter with LBP features | • ECCV 14 |
| Local Maximal Occurrence (LOMO) [147] | • It extracts color feature from HSV histogram and uses scale-invariant LBP features | • CVPR 15. |
| Gaussian Of Gaussian (GOG) [152] | • It uses pixel-level descriptors encoded by hierarchical Gaussian modeling. | • CVPR 16 |

Hong *et al.* [154], [155] before performing person re-identification. Truong Cong *et al.* [119] used localized features of the body parts and extract their descriptors for person re-identification; where Alahi *et al.* [156] presented an approach for creating object descriptor built using cascaded grids of descriptors.

Wang *et al.* [157] presented an approach that combines shape and appearance features to prepare a matrix that indicates the descriptor of the tracked people. The descriptor is built based on the captured closed relation between the appearance labels. The same approach was used to build a framework for real-time computation of the occurrence matrix. Besides, Truong Cong *et al.* [119] presented a color-position histogram obtained from the silhouette of the tracked persons. The presented color-position signature was classified into regions where RGB values were extracted from each to create a discriminative signature employed for matching tracked individuals from different camera views.

Jüngling *et al.* [127] presented the person Re-ID approach in a multi-camera setup that used the Implicit Shape Model (ISM) and SIFT features. The approach performs human detection and tracking, where SIFT feature models are built during tracking. ISM is used for human detection and tracking, while SIFT serves the purpose of feature extraction. The approach is not sensor independent (homogenous or heterogeneous) because it does not use color or other sensor specifics features for re-identification. Also, Hamdoun *et al.* [158] presented a method of identification using interest points collated from different motion images. SURF algorithm was used to obtain the feature points with descriptors to compute a signature for the tracked persons.

Approaches based on local features, Wong *et al.* [17] presented a method that used local features like the color histogram, size, the intensity gradient, and position of the tracked individuals to perform tracking using data association algorithms. Another approach under this is the work of Piciarelli *et al.* [60] used appearance features obtained from the upper body level of the detected person and the spatial position for re-identification in other camera views.

### D. INFORMATION FUSION IN A MULTI-CAMERA SYSTEM

The fusion of information or data aims to produce something better than what can be obtained separately. However, information fusion in a multi-camera system is the combination of different features extracted from the same object, different instances of the same object, or the same scene of an object from different views of cameras. Views of cameras in a multi-camera system are best explained in the overlapping and non-overlapping form, also information fusion as a multi-source data fusion process from different views of cameras. Therefore, information fusion can be well seen from the perspective of the overlapping and non-overlapping camera.

#### 1) INFORMATION FUSION IN OVERLAPPING CAMERAS

Many information fusions approach or methods have been applied by researchers for overlapping cameras. Some of the approaches are particle filtering (Sequential Monte Carlo Methods), Bayesian estimation, support vector machine (SVM), and Kalman filtering. For example, Lu and Li [29] presented a novel image fusion approach that integrated particle filters and belief propagation as a unified framework. A dedicated particle-filter tracker was applied in each camera's view. Each local tracker in different views works together with other views through belief propagation as a means of passing genuine and important information across the view. Also, Zhao *et al.* [30] presented a distributed Bayesian formulation using multiple interactive trackers. The approach avoids joint state-space rendition to prevent tedious and complex joint data union. The method has a better approach for an online real-time tracking application. However, a conventional Bayesian tracking framework was modeled for problems like proximity, occlusion, or interactions between the objects' observation. Lébraly *et al.* [31] integrated Bayesian particle filtering with Dempster–Shafer theory to fuse the pieces of evidence obtained from multiple heterogeneous and unreliable sensors. The approach was used to solve the problem of tracking people in a multi-camera indoor environment. Xu *et al.* [32] also used the Bayesian framework to track a variable number of 3D persons in an overlapping multi-camera setup. In their approach, they employed joint multi-object state-space formulation, each object states are defined in 3D. This approach is not applicable in real-time applications because of the computational complexity and large joint data required. Huang *et al.* [33] presented a system that fuses tracking information in an overlapping multi-camera set up using an approach called map-view mapping. A particle filtering algorithm was adopted for target tracking and information fusion was performed according to reliability and weighting of each source of information.

#### 2) INFORMATION FUSION IN NON-OVERLAPPING CAMERAS

Other researchers applied information fusion in a non-overlapping camera setup. For example, D'Orazio *et al.* [159] proposed tracking of people in a multi-camera set up using appearance similarity. The color histogram mapping of the same object from different views was performed. The mean brightness transfer function (MBTF) and cumulative brightness transfer function (CBTF) were used to determine the appearance similarity but the performance of the two is quite similar in the phase of simple association problem. The two methods experienced a setback when it comes to the detection of a new entry scenario. In this method, further research can be centered on the major differences between two similar bodies so that a technique can be applied based on the differences. Chilgunde *et al.* [160] presented a real-time tracking system for targets in a non-overlapping multi-camera setup. Kalman filter was used to select the targets based on shape and motion from each cameras' view. For areas that are not covered by the camera, a prediction of each cameras' view is made using the Kalman filter prediction. Gaussian distributions of the tracking parameters were computed for each camera to determine the target position and motion.

Lin and Huang [161] also presented a framework that applied the client-server system for tracking targets in a multicamera setup. The client aspect manages single cameras' object tracking and detection and the server is responsible for collaboration between the multiple cameras. Kalman filter was applied for object tracking and detection for a single camera view and homograph was implemented to determine the FOV lines for each camera's view. The features identified from the single-camera object tracking are fused for object matching and the FOV lines are used for switching between the cameras. The framework was designed to unify object tracking for overlapping and non-overlapping multi-camera systems. Leoputra *et al.* [162] proposed a unified framework that used a particle filter to track target objects in a non-overlapping multicamera system. The blind section between the camera views is predicted by switching between the tracking predictions and visual tracking. The trajectory information of the target object as it moves through the blind areas is mapped by the Particle Filter algorithm.

Bauml *et al.* [163] presented a system for online capture and recognition that uses facial features for person re-identification in a non-overlapping multi-camera system. The system combines the support vector machine (SVM) and Discrete Cosine Transform (DCT) to determine the facial features and extraction. The challenges of face appearance feature for person re-identification like low resolutions facial image, lighting and pose are addressed. Since people's faces are sometimes very similar, future techniques should integrate face tracker with body tracker or associate face tracking with other close term features such as clothing for better identification. Avraham *et al.* [125] proposed an approach that modeled a camera with a transfer function for a multi-valued transformation for pedestrian re-identification. The system is metric independent and did not depend on learning object appearance from one domain to another. The approach used the radial basis function kernel (RBF) binary SVM classifier to reveal the unapparent differences between the camera's view. Prosser *et al.* [123] described a person re-identification as a sequencing problem rather than a distance measuring problem. They introduced a novel ensemble Rank SVM algorithm to perform the actual sequencing match and minimize the computational time suffered by the original SVM approach. The approach is more scalable, and it requires less memory.

## V. APPLICATION OF MULTI-CAMERA IN SURVEILLANCE SYSTEMS

In literature, many papers have been written on the application of multi-camera, especially in surveillance systems. Table 5 shows the list of studies that have been conducted on the application of multi-cameras in the surveillance system and their description.

The earliest research carried out on the multi-camera system was performed with fixed view cameras. The resolution of the cameras employed was a low type because of their

**TABLE 5. Summary of recent studies on the application of multi-camera in surveillance systems.**

| Authors | Topic/Area | Description |
|---|---|---|
| Valera and Velastin, 2005 [164] | Computer vision and tracking technique | Evaluated the generations of intelligent surveillance systems and reported previous works on motion detection, image enhancement and interpretation, recognition, tracking, human behavioral investigation, and method storing. |
| Morris and Trivedi, 2008 [165] | Activity analysis | Study objects path-way activities over time to predict a behavioral probabilistic model. The data generated is used for surveillance analysis. |
| Abidi et al., 2008 [166] | Multi-sensor integration | Examined the available technologies under sensor and surveillance systems for a wide area network. Also, it described the modalities required for selecting, planning, and fusing data obtained from a sensor in a multiple sensor environment. |
| Sheikh et al., 2009 [167] | Computer vision and tracking technique | Described the characteristics for tracking objects in the multi-camera system based on changes in shape, motion, and size. Since the object's shape, motion, and size are not constantly fixed, object tracking was described as a region tracking problem. The region is seen as a 2D image plane display of an object. |
| Aghajan and Cavallaro, 2009 [168] | Computer vision, multi-view geometry, and multiple camera networks | Described the basic multi-camera configurations from image development and camera model, the geometry of stereo vision and camera matrix, the projective form of transformation and n-camera construction. Other features discussed are detection and matching, basic assessment algorithms in multi-camera systems. |
| Sommerlade and Reid, 2009 [59] | Computer vision techniques. | This work created a framework for cameras of different make with overlapping views of a scene of the same location. The output of the cameras was homogenized such that there was no need for supervising camera or creating a constraint for the parameters of the camera. |
| Kim et al., 2010 [169] | Computer vision techniques. | This work centered on analysis and elucidation of object responses, detection and tracking to perceive the look of the scene. It also speculates about wide-area control of cameras used in tracking an object that is moving in a multi-camera system. |
| Seema and Reisslein, 2011 [170] | Wireless video sensor platforms. | The paper conducted a hardware and software review on the platforms of wireless video sensor nodes for surveillance. Based survey outcome, it produced a novel wireless video sensor network at low-cost framework design. |
| Castaneda et al., 2011 [171] | Computer vision and vehicle tracking. | It described a real-time multi-camera-based vehicle tracking system for tunnel surveillance. An algorithm (i.e inter-camera matching) was used to measure the correlation of projections between the images of the vehicles. |
| Tavli et al., 2012 [172] | Visual sensor network platforms | Highlighted the various visual sensor networks platform that was available and the methods of compressing an image, coding of video, and computer vision methods applicable on those platforms. |

**TABLE 5.** ***(Continued.)*** **Summary of recent studies on the application of multi-camera in surveillance systems.**

| Authors | Topic/Area | Description |
|---|---|---|
| Roy-Chowdhury and Song, 2012 [173] | Wide-area camera network | It highlighted the challenges faced in comprehending images created in a multi-camera system network. Then, it presented how cameras handle overall decision making at the network layer in a multi-camera system network. It also discussed in detail the most recent methodologies in the multi-camera system network. |
| Winkler and Rinner, 2012 [174] | Security and privacy protection. | Discussed the various methods of privacy protection, major challenges, warning, and attack that can be faced in visual sensor networks where the visual sensors have inbuilt modules for processing and communication. |
| Vezzani et al., 2013 [175] | People re-identification. | Identified the pragmatic issues involved concerning people's recognition process and the several methods of implementation. |
| Song et al., 2013 [289] | Computer vision methods for cameras in a sparse network | Reviewed the various techniques used by sparse camera network surveillance applications. It also discussed the various activities involved in terms of Intra-camera and Inter-camera ranging from space modeling, motion partitioning, and target tracking. It discussed further cameras without overlapping FOVs, how they relate and recover when speculating a target. |
| Wang, 2013 [176] | Computer vision, pattern recognition, image sensors, and signal processing | This paper reviewed multi-camera systems in terms of camera calibration, topology, tracking, object recognition, activity analysis and collaborative video surveillance for linear and smart cameras. It discussed the various challenges faced and compared the available solutions in all the areas of discussion. |
| Sanmiguel et al., 2014 [177] | Smart camera network | This paper examined the challenges faced by autonomous smart cameras with self-reconfiguration. It identified five pillars of self-reconfiguration which are camera, network, environment, tasks, and performance. |
| Jiang et al., 2014 [114] | Computer vision, tracking, and optimization | A two-stage graph-based tracking system that used multi-camera was proposed. The data obtained was converted to an optimization problem using the MAP algorithm to determine the mini-cost path. |
| Natarajan et al., 2015 [11] | Multi-camera coordination and control strategies | They provide a comprehensive survey on state-of-the-art multi-camera coordination and control ($MC^3$) strategies or approaches used in the surveillance. |
| Jin, An, and Bhanu, 2017 [178] | Computer and pedestrian tracking | This paper proposed a framework used in tracking pedestrian including the group structural information. It updated the cross-camera model using a structured vector machine. |

wider view angle to cover wide areas. Eventually, with the emergence of Pan-Tilt-Zoom (PTZ) cameras, a clear image of the surveillance object can now be obtained. Initially, the problem of non-overlapping FOV of the static cameras has prevented the solution to occlusion and localization of images in the real world. With the invention of PTZ cameras, a particular area of interest (AOI) can now be focused. PTZ cameras are mostly configured as a master-slave system to quickly sense and capture a human face and objects' shape in any direction.

There are series of development in the area of multi-camera system control and coordination, especially for the heterogeneous camera system. PTZ cameras are configured as a master-slave setup to collaborate, sense and monitor targets in a multi-camera arrangement. Several approaches like the game theory, decision theory, and control theory were used to create collaborative monitoring among cameras, while others applied optimization structures [11]. The discovery of smart cameras contributed a lot in the area of object tracking, on-board processing and getting detailed information of the captured image from the scene. In distributed smart cameras (DSCs) information is shared between the individual cameras with distributed sensing and processing capability in a smart camera network. Pervasive smart cameras (PSCs) create autonomous and adaptation functioning of smart cameras which provides easy usage and operation in the various areas of application. With this development in smart cameras, the attention of the researchers was directed towards self-control and collaboration among distributed cameras [52].

Many works of literature have identified different problems associated with surveillance using multi-camera systems while some have proffer solutions to the problems identified by others. Presently, surveillance problems using multi-camera systems rest on multi-camera coordination when it involves multiple targets to be observed. The number of targets that a particular camera can observe without compromising the quality of the captured image has not been stipulated. Especially when the camera to target ratio is less. Also, the problem of occlusion in a multi-camera setup is a persistent issue for areas where there are physical obstructions and privacy issues. Furthermore, there is a problem with computation or processing, especially for real-time applications. Surveillance is a real-time activity that demands a lot of processing power from cameras on a surveillance network. Presently, the processing capabilities of smart cameras are still fell short when it involved many cameras. Even if the smart cameras can process faster, the communication medium still limits the rate of data transfer from one camera to another because of the traffic load.

## VI. APPLICATION OF MULTI-CAMERA AND IMAGE OVERLAY IN SPORT ANALYSIS

A multi-camera system has a wider area of applications in the human endeavor or factually in every aspect of human activities [179] and entertainment is not an exception. This area ranging from film making, sports, TV shows and other performances or activities that keep people's attention. This study will concentrate on the applications of the multi-camera

system in sports where football and other games will be the center of discussion. For example, the automatic display of views that draws special attention in sports programs, video-on-demand display [180], and other aspects like tracking of football as a way of making comments by pundits in sports videos [181] will be the areas considered for review. Needham and Boyle [182] applied a monocular non-moving camera to monitor players in an indoor 5-aside soccer fiesta. The tracking of the players was done with an algorithm called ''CONDENSATION''. This was carried out with a single camera that has low-resolution.

The single-camera view system has some shortcomings ranging from low-resolution image output, lack of total coverage of the pitch due to limited FOV of the camera, and players occlusion problem. Especially, when the players queue up in–line in the direction of the camera. Also, the position of the ball on the pitch or when the ball is rolling on the ground creates constraints on the 3D view of the ball while using a single camera. This will further demean the application of a single-camera view because the direction of the movement of the ball draws the special attention of the viewer [183]. Table 6 shows a summary of the applications of multi-camera systems in sports.

The earliest research according to Table 6 on the application of camera(s) in entertainment or sporting activities was a graphical overlay on a calibrated camera image. This is done by placing a sensor on the camera stand and lens which allows the recording of the cameras' panning, tilting, and zooming. This sensor allows images to be overlaid on the moving video and placed as a background for the video recorded by the camera. The method can also be used to create graphics that can measure the distance of the player to the goal post and mark the off-side line by calibrating the position of the camera with regards to the scene manually. This was one of the reasons why video image and single-camera were the most common applicable sensor as the first type of graphic technology in sport [191], [195]–[197]. Furthermore, a typical application of a mechanical sensor on-camera stand together with other groups of cameras was applied [185]. The 'FoxTrax' was a puck-tracking system that used graphics overlay to display the movement of an ice hockey puck. A 20 IR LED was implanted into the puck, the signals of the LEDs are picked by the set of IR cameras which were hanged up at the roof. However, the position of the puck was made to be located using the images of the IR cameras. For the easy location of the puck during motion, the system added a 'comet tail' and a blue trailing light [198].

The introduction of image processing into sporting display brought about a color segmentation algorithm (Chromakeyer) which allows the segmentation of the foreground color from the background. This technology was used to make the background color of a sporting field uniform, for example, the green background color in a soccer pitch. A typical application of this can be found in '1st and Ten$^{TM}$' [199] American football and ski race is shown as if they were physically present.

**TABLE 6. Summaries of the application of multi-camera systems in sport.**

| Authors | Sport | Sensor | Description |
|---|---|---|---|
| Cai and Aggarwal, 1996 [184] | Indoor sports | Multiple fixed perspective cameras. | It presented a framework for monitoring human movement in an indoor environment using images captured from multiple fixed cameras. |
| Cavallaro, 1997 [185] | Ice hockey puck tracking system to improve visibility | Thermal camera, IR repeater, radar, and broadcast camera. | This design has a set of eight infra-red cameras to track the movement of a hockey puck that has a set of infra-red embedded in it. The triangulation method was used to compute the position of the puck from the cameras. |
| Bebie and Bieri, 1998 [186] | Soccer game. | Video sequences acquired from television. | It generated 3D scenes from a soccer game video sequence. The 3D scene has a virtual viewpoint viewer. |
| Ohno et al., 2000 [187] | Tracking of players and the ball in a soccer sport. | Fixed Multi-cameras. | It automatically tracks players and ball using fixed cameras. It can determine the position of the players and ball in the 3D view. |
| Prandoni et al. 2004 [188] | Tracking the trajectory of the ball, Ice skating, diving or athlete gesture. | Video Multi-cameras. | It described stroboscope techniques for analyzing the development of an athlete trajectories over time or space. Multi-camera can also be used to cover a wider FOV or for many sequences comparison. |
| Grau et al., 2002 [189] | Analysis of sport scene. | Two or more PTZ cameras. | This approach presented a virtual 3D-shape reconstruction in the studio. It was an expansion of the chroma-key technology. |
| Owens et al., 2003 [190] | Tennis ball tracking. | Broadcast multiple cameras. | It described a passive system for tracking a tennis ball using 3D reconstitution techniques from 2D images captured by different cameras. |
| McIlroy, 2008 [191] | Tennis ball tracking. | Multi-camera. | It described a camera-based ball-tracking system that monitors the touchlines of a tennis playing court. |
| Reusens et al., 2009 [192] | Downhill ski race. | Video multi-cameras. | It generated a composite video sequence from two or more video sequences where the composite video contained the elements of each video sequence. It also allowed the combination of a video sequence with that of audio. |

The application of the multi-camera view system to soccer game coverage or sports analysis in general tackles the single-

camera shortcomings. Multi-camera system application in sports coverage and analysis provides a wider FOV that will cover the whole pitch of play, reduces the dynamic occlusion, and allows camera output integration or collaboration. Though a multi-camera system application in sports coverage requires camera calibration and the camera position is also paramount [200]. Generally, the use of the multi-camera system has found its purpose in various sports applications such as football, ice hockey, snooker, diving and ice skating, down-hill sky race, tennis ball, and tracking of an athlete.

In the event of viewing sports activities, it became difficult to show an athlete's motion or objects trajectories such as ice hockey puck, football, tennis, and others' evolution over time and space. This is because the FOV of a fixed photographic camera cannot capture the entire spatial and time of an athlete's motion. A new idea was employed in which a set of cameras are arranged along the path of the athlete or object to be monitored to snapshot the athletes or objects as they pass. The resulting images of the athletes or objects' motion are joined together to form the total view of the event. A typical example of this was applied [188] to show athletes' motion in sports like ice skating and swimming. This method compared to image overlay improves player resolution and provides full coverage of the total sporting event with the application of multiple fixed cameras [183].

Steins [201] and Del Bimbo *et al.* [202] reportedly computed homography transformation for images obtained from the field of view of two overlapping uncalibrated cameras. All the targets on the images are considered to be on the same plane. Therefore, the homography computation result was meant for all the targets between the FOV of the two cameras. Khan *et al.* [203] and a similar work presented by Cai and Aggarwal [184] monitor a target using uncalibrated multi-cameras. It was highlighted in their approach that for cameras with overlapping FOV, the current camera should hand the field of view to the neighbor camera once the target leaves its field of view.

A different approach to the above method was the use of calibrated cameras to create a 3D position of an athlete or ball in a sporting event. Several approaches have been discussed in the literature. First, using the foreground segmented image of the athlete together with calibrated cameras. The lowest point of the segmented image of the athlete from a calibrated camera should be assumed to be in direct contact with the ground. Another approach was to place the foreground segmented image of the player on a 3D model of a stadium. This allowed a seamless creation of a virtual view of the game other than the actual view captured by the physical camera which can be tilted to a different direction of view other than that of the physical camera. This approach was employed by the Red Bee Media (RBM) [204] in segmenting player's images from the pitch.

The problem with the player or stadium modeling approach was that there are limitations in its applicability when using a single camera. The degree of tilting the virtual camera view relative to that of the physical camera is too small and the players' occlusion problem is also difficult to resolve. An alternative approach to the above method was the use of the multi-camera method. This approach made use of high frame rate calibrated broadcast cameras to capture the pitch or field of the sport. The system was first used for tracking tennis game in a 3D view [190].

One more peculiar problem in sport was the determination of the location and tracking of the individual players and the ball. There are a lot of approaches to this problem in the literature. Though it was a general assumption that tracking players are more difficult than the ball because of some obvious reasons. The ball rolls on a pitch alone and it has definite shape while players are many, running after the ball [193]. Therefore, there will be a problem of occlusion between players. The ball has a particular pattern of movement which can easily be modeled but players move erratically following the direction of the ball. One more thing was that players need to be identified whether by number or jersey color before been tracked.

Distributed multiple fixed cameras at different locations on the pitch of the game was one of the methods [193] applied in tracking players and the ball, though it required more cameras and somehow costly to implement. For commercial purpose, most multi-camera-based player tracking system employs the use of automated cameras together with manual camera tracking approach. Areas of spectacular tackling and accurate passes are been identified and manually logged into live and latter highlighted those areas for analysis at the end of the match.

Another method is the use of two cameras positioned to work together, then applying the triangulation method to determine the points in 3D space. Here, the matrices of the cameras (camera projection function from 3D to 2D) involved must be known.

A lot has been done in the application of multi-camera in the area of sports and entertainment from image overlay, single-camera with a mechanical sensor, single-camera with an image overlay and multiple cameras. Monitoring the position of the limbs of an athlete during sporting activities remains a challenge in this field of study. Marker-based motion approach [200] was sometimes proposed but this only applicable during training and exercise, in live matches analysis, factually not possible. Future studies should focus on automated systems for tracking athletes' limbs during live matches or on recorded videos.

## VII. APPLICATION OF MULTI-CAMERA IN EDUCATION

The wind of the digital revolution blowing across all the facets of our life has brought tremendous growth and ease into the way we do things traditionally. Environmental digitalization was at the current leading front of creating new ways of human relations. The effort of carrying out unobtrusive monitoring, distinct and trackable actions with large numbers of the audience gave us the chance of looking at the diverse facet of human undertakings. The ubiquitous presence of

cameras in most automation applications especially those that deal with light sensing, photo capturing, monitoring and tracking system has made cameras relevant in every facet of our life. For example, in the area of education, technological inventions gave us the idea of considering a scenario where teaching can be carried out without a teacher (intelligent teaching systems) [205], or in another perspective; teaching unreal students (distance learning) [206].

The latest technological innovation was focusing on the level under which lectures can hold, and the possibility of transmitting it with quality to large numbers of students (Massive Open Online Courses (MOOCs) [207], [208]. All these are made possible due to advancements in the area of video capturing, processing, compression and delivery techniques. In the late '60s, it all started outside the class which can now be referred to as distance learning. Distance education was based on noncontiguous communication between a school (represented by the teacher) and its students [209]. In other words, this was two-way communication. The first way was the communication from the teacher or supporting organization to the student through the sending of instructional materials and other learning materials. The other way was represented as feedback from the students back to the teacher. Other areas where cameras are applicable in education are in the area of estimation of students' attention and smart attendance monitoring system, teacher's evaluation/appraisal exercise and feedback on student performance, automatic students face recognition/detection and teachers and student security and protection. All these can be achieved through technological innovations in education and they are majorly carried out by the application of camera or multi-camera systems. There are several papers related to cameras or multi-camera applications in educational settings found in the literature. Table 7 summarizes the related survey papers on camera or multi-camera applications in education. As described by most papers, the majority of the work focus on lecture capturing and broadcasting, students' attention estimation, teacher's evaluation, protection and security of students and teachers, attendance monitoring, face recognition and detection, motion detection, and behavior analysis during lectures where cameras or multi-camera systems are applied.

**TABLE 7. Summary of related survey papers on multi-camera applications in education.**

| Authors | Sensor | Description |
|---|---|---|
| Rui et al., 2001 [210] | Multi-camera | The paper presented an automatic audio-visual tracking of lectures, lecturer and audience systems. It automatically performed video editing and produced an output of the same quality as that of humans. |
| Bianchi and Madison, 2004 [211] | Multi-camera | It presented an automatic auditorium system that can be used in businesses and academic environments. It created a real-time video of lectures captured using multi-cameras without any external control apart from turning it on and off. |
| Rui et al., 2004 [212] | Multi-camera | It summarized a system where capturing and broadcasting of lectures are automated and then sent to an online audience. |
| Zhang et al., 2007 [213] | Multi-camera, gyroscopes, accelerometer | It presented a system that used wearable devices like head-motion, pen-motion, and visual-focus modules to determine the level of student attention in class. |
| Zhang et al., 2008 [214] | Multi-camera | It presented an automated lecture capturing system that synchronizes captured lectures together with the audio, video and slide presentations. It automatically broadcast lectures which are accessible by the remote audience. |
| Li et al., 2009 [215] | Multi-camera | The paper presented an algorithm with a multi-camera multi-touch technique designed for the educational system. It has a high processing speed with low CPU consumption. |
| Xu et al., 2010 [216] | Multi-camera | The paper proposed an eye-tracking system to detect objects. It used a multi-focal camera system to mimic visual function. Several objects were detected by introducing incremental top-down information of the target object into the bottom-up model. |
| Napoletano and Tisato, 2014 [217] | Multi-camera | It dynamically displayed the content of a camera from all available cameras in a multi-camera setup. It mimics the human visual system that employed bottom-up and top-down clues |
| Kalaivani et al., 2017 [218] | Multi-camera | It used a machine-learning algorithm to determine the behavior of students on distance learning. The algorithm used the histogram of optical flow orientation, magnitude, and entropy (HOFOME) |
| Mothwa et al., 2018 [219] | Multi-camera | The paper designed a model for face recognition AI-based student attendance monitoring system. |

### A. EVOLUTION OF CAMERA AND MULTI-CAMERA IN EDUCATIONAL SYSTEM

This section briefly discusses the evolution of multi-camera in education. The application of multi-camera started around 1985 where the majority of its usage was in the area of object detection and recognition. The evolution of multi-camera was due to the limitation in the application of the single-camera in terms of area of coverage, resolution of the image and accuracy of the image output. However, this led to the introduction of multi-camera where the accuracy of the camera can be improved by fusing the image obtained from multiple cameras. Thereby increasing the resolution of the output image and allow a wider area of coverage. Earlier work of multi-camera focuses on object detection and image classification [220]. Up till early 2000, the majority of the cameras used in surveillance and tracking of an object are fixed cameras which are of low resolution. Though, they have a wider view but cannot capture objects that are of far distance. Due to these shortcomings, Pan Tilt Zoom (PTZ) was invented to produce a high-resolution image with the power of capturing far distance target objects. The first set

of PTZ cameras are manually configured to capture specific target areas. For example, focusing on the presentation slide board, surveillance of the entry and exit point and tracking of the individual students' faces for recognition. In a multi-camera setup, the PTZ cameras are later configured automatically to focus on specific areas based on defined features or configured as master-slave for automatic control and coordination [11].

In early 2010, smart cameras and wearable devices are introduced to carry out intelligent monitoring of events and tracking of human limbs. For example, Bianchi and Way [221] described the concept of the automatic auditorium (AutoAuditorium system), where audiovisual lectures produced in a lecture auditorium can be automatically captured in real-time without the support of any human control apart from turning it on and off. Erdmann and Gabriel [222] applied an automated smart camera system to perform audio-visual tracking of the lecturer and audience in a lecture hall. The system also performs automatic video editing with quality near to the one performed by a human coordinated system. This setup completely automates events or lecture capturing for distribution and easy access by students. In the aspect of human action recognition and position monitoring, generally, they can be observed in two ways. There are wearable sensor-based devices and vision monitoring devices. For wearable sensor-based devices, the devices are to be worn by the target to determine or monitor his / her activities [223]. This approach used action models to infer the behaviors and actions of the target. For example, Zhang *et al.* [213] used wearable devices like head, pen, and eyes-focus modules to analyze students' attention. These modules gathered information through smart cameras, gyroscope, and accelerometer embedded in those wearable devices. Another approach [224] was the use of an eye-tracker to determine the attention level of the student. However, this approach will have a serious negative impact on the eyes of the target [225] and the gadgets are quite expensive. The use of wearable devices has tendencies of higher accuracy but there are some major challenges like:

- The number of human features to be used for inference, each feature will require a sensor and each sensor will generate data. So, there will be a diverse set of data to handle together.
- After gathering the data from different features, selecting the feature to use to determine the best result will be difficult because it involves physical and emotional feedback.
- Another challenge is individual differences. The reaction or behavior of an individual might be different towards different things at the same time. An individual might pretend to be focused but his / her attention is somewhere else.

The vision monitoring devices employ a camera or multi-camera-based system to detect the actions of the targets [226]. Steriadis et al used video cameras [227] to monitor students' behavior and their facial expressions to determine their level of attention [216]. A lot of factors will affect the correctness of this approach like lighting intensity and image background obstruction. It also requires high processing computer capability, especially in real-time applications.

Mothwa et al presented a conceptual model of a face recognition student attendance monitoring system. The authors make use of a full multi-camera view to capture and detect the faces of the students. The system is designed to perform periodic real-time recognition of students during lectures and performs update recognition after a specific interval of time to ensure that some of the students did not leave the lecture hall after the first capture. The captured images of the students are compared against the database image of the student already stored. This is then used to determine the presence or absence of a student in the lecture. The approach used centralized architecture where all the cameras are connected to a single facial recognition. In this type of architecture, there is no redundancy in the setup when the central interface that connected the multi-camera is down. The authors employ histogram equalization, bilateral filtration, and elliptical cropping to perform preprocessing tasks on the captured images.

### B. OBJECT TRACKING ALGORITHM AND IMAGE RECOGNITION

This subsection highlights the various tools and algorithms used in object detection or recognition from images or video frames. Object recognition in image or video of multiple frames is performed using object tracking techniques. Most of the object tracking algorithms composed of three features; object representation, dynamic model and search procedure [228]. There are two ways of representing objects; holistic and local description object representation. Examples of the holistic descriptor are values of the raw pixels and color histogram while that of the local descriptor is local histogram and color information. The dynamic model narrates the motion between two successive frames thereby reducing the computational problem in object recognition. Also, the searching procedure in the object recognition algorithm is seen as a problem of optimization. A deterministic and stochastic approach is always employed to solve it. The deterministic approach is only applicable when there are no local minima involved otherwise Stochastic or sampling methods are employed [228], [229]. However, there are a lot of object recognition algorithms proposed by researchers, some of them are good for real-time (online method) object detection which is mostly used in educational applications while others are used for object recognition in still-image. Examples of the algorithms are Tracking-Learning-Detection (TLD) [230], resonant tunneling device (RTD) [231], Multiple Instance Learning (MIL) [232], incremental visual tracker (IVT) [233], Beyond semi-supervised (BeSemiT) [234], L1 tracker (L1T) [229], visual tracking decomposition (VTD), semi-supervised tracker (SemiT) [235], variance ratio tracker (VRT) [236], fragment-based tracker (FragT) [237], and online boosting tracker (BoostT) [238]. It was shown that TLD was the most reliable and robust online object

tracking algorithm when tested on different video frames and images [228].

Wang *et al.* [228] in Table 8 tested the performance of the algorithms based on the modeling of their motion, the state vector movement between two successive frames (dynamic model), and their searching methods. Boris et al. presented multiple instant learning (MIL) to separate an object from its background. The researchers employed the appearance model which composed of the discriminative classifier, the classifier load itself and then extract the examples (positive and negative) from the most recent frame. A little change in the tracker can cause incorrect labeling of the trained examples and this debases the classifier. However, the motion model employed in the approach was too simple. It could be replaced with a more robust and sophisticated one like a particle filter. Kainz *et al.* [239] used LTD to determine the number of the student attending a lecture and later on applied face recognition algorithm to identify the particular student in the detected faces.

Feature extraction is an essential part of face recognition. It deducts salient features subsets from the main data following some rules. The advantage of feature extraction is that it increases the speed of machine training and reduces space complications. The different categories of feature extraction methods are the holistic method, feature-based and hybrid matching methods. The holistic methods are the most widely used because they employed the whole face as input data. Examples of the holistic methods are Linear Discriminant Analysis (LDA) [240], Principle Component Analysis (PCA) [241], Independent Component Analysis (ICA) [242] and Local Binary Patterns (LBP). Mothwa *et al.* [219] used the Viola and Jones face detection algorithm [243], this algorithm employed Haar cascade to identify faces and it can work in real-time.

Mothwa *et al.* [219] employed three different cameras of the same quality to capture student's faces and used PCA, LDA, and LBP to extract the features of the face images. The Euclidean distance was calculated to determine the accuracy between the image tested and the train data. The accuracy of the feature extraction algorithms was tested and compared. High recognition accuracy was obtained when the PCA and LDA were combined and used on the first camera.

All these algorithms mentioned above can be found in computer vision tools. Computer vision tools are software for implementing image and video frame processing, examples of these tools are OpenCV, VXL, LTI, OpenTLD, MatLab and fast CV (produced by Qualcomm). The fast CV tool can be operated on mobile devices. A comparison of these vision libraries shows that open CV is faster when used on computers of the same specifications [244].

Fuzail *et al.* [245] implemented real-time detection algorithms for managing student attendance. They used the Haar classifier for face detection and it was implemented in the OpenCV computer vision tool. The recognition aspect was done using an algorithm in python named pyfaces. Apart from its fast speed of recognition, this algorithm relies on the pose of the image, scale, and color to differentiate between the compared image and the one stored in the database. However, the system cannot identify the individual student available in class. Tamimi *et al.* [246] presented a real-time group face detection system. This approach is similar to Fuzail et al.'s approach. The face detection algorithm was implemented in MatLab. 2012, which is another example of computer vision software. This system cannot identify the individual student in the captured image.

**TABLE 8. Summary of related survey papers on multi-camera applications in education.**

| Algorithm | Motion Model | Object representation | Dynamic model | Searching mechanism | Characteristics |
|---|---|---|---|---|---|
| IVT | Affine transform | Holistic gray-scale image vector | Gau | PF | Gen |
| FragT | Similarity transform | Local gray-scale histograms | - | Sam | Gen |
| VRT | Translational motion | Holistic color histograms | - | MS | Dis |
| BoostT | Translational motion | Holistic representation based on Haar-like HOG and LBP descriptors | - | Sam | Dis |
| SemiT | Translational motion | Holistic representation based on Haar-like descriptor | - | Sam | Dis |
| BeSemiT | Translational motion | Holistic representation based on Haar-like, HOG and color histograms | - | Sam | Dis |
| L1T | Affine transform | Holistic gray-level image vector | Gau | PF | Gen |
| MILT | Translational motion | Holistic representation based on the Haar-like descriptor | - | Sam | Dis |
| VTD | Similarity transform | Holistic representation based on hue, saturation, intensity, and edge template | Gau | PF | Gen |
| TLD | Similarity transform | Holistic representation based on Haar-like descriptor | - | Sam | Dis |

**Abbreviations:**
Gau = Gaussian; PF = Particle filter; Sam = Sampling; MS = Mean-shift; Gen = Generative; Dis = Discriminative

Zhang *et al.* [213] demonstrated student attention level determinant using wearable devices built on four modules; they are head movement, pen movement, visual focus, and Apps modules. In the head movement module, Speed-Up Robust Feature (SURF) [247] algorithm was used to determine the correlation between the position of the students' head (whether up, down, or center) and the information on the white marker board. SURF was implemented in an open CV library.

## VIII. APPLICATION OF MULTI-CAMERA IN MOBILE PHONES

The growing competition in the mobile communication world has prompted many manufacturers of mobile phones to introduce more features in their mobile products like imaging-related functions such as video recording, video calls, and video conferencing [248], [249], [222]. The importance attached to the camera on smartphones by manufacturers and users has driven most mobile phone manufacturers to work hard on how to improve the features and image qualities of their products. The initial concern of most mobile phone manufacturers was the resolution of the camera on the smartphone which lead to the production of lots of megapixels cameras. Another possible feature was the introduction of object digital zooming in on mobile phones which was entirely based on software interpolation when it came out initially due to the limitation in the camera focal lens and the size of the mobile phone. The thin shape of the smartphone body [250] and the design of the lens makes it complicated to fit a zoom lens. The implementation of multi-cameras on mobile phones has made several features possible and much easier such as object zooming (through optical), portrait mode, 3D, better high dynamic range (HDR) and low-light photography. The most popular smartphone manufacturers have moved to stereo camera design. However, not all cameras on multi-camera smartphones are serving the same function. Some smartphones have a primary camera that does the actual capturing of images while the secondary camera is either a telephoto lens (thick and wider) or monochrome with a wider field of view (FOV). A telephoto lens has a longer focal length and it can bring a distant object or scene closer as shown in Figure 12. The advantage of using a dedicated telephoto camera module on a mobile phone is that it solves the problem of zoom-in and produces better images at the long focal length. The result will be better than using cropping and scaling the image output of the main camera. The output of a two-camera module can be integrated to produce an improved image. Though, this poses an image processing challenges such as white balance and focus distance problem because the two images will slightly offset each other.

According to the literature, this area has not been well studied. There will be a need for alignment in terms of the geometry and photometric properties of the images produced by the two asymmetric lenses. Anirudh *et al.* [250] proposed a computational algorithm that can solve the problem of

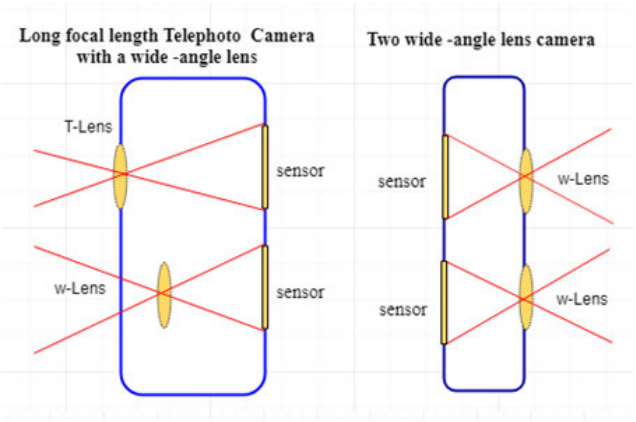


**FIGURE 12.** Mobile phones with a telephoto lens (usually wide) and a wide-angle lens.

fusion of images produced by multiple asymmetrical cameras (tele-wide) in terms of color and brightness integration. The images obtained from the telephoto and wider view lenses were first set to the same scale and FOV (refer to as global image registration) using different algorithms (Oriented Fast and Rotated (ORB), Library for Approximate Nearest Neighbour (FLANN), Random Sample Consensus (RANSAC), and affine transform). The brightness and color of the two images are then matched since they are both in the same FOVs. The brightness and color of the resulted image were then corrected at Y-channel and U and V-channels respectively. However, the algorithm was tested on static image and scene with occlusion but not on motion pictures. The approach of Liu and Zhang [251] was performed on six symmetric wide-view cameras and the cameras are all calibrated. A global correction gain was calculated for each camera view using the appropriate color index that matches the global brightness and color. The differences between the corresponding overlapping sample of the views were minimized by the joint optimization of the optimal gain. Then, the tone marking curve was applied to remove the photometric misalignment. In this approach, a lot of assumptions were made and the method is not adjustable to different lighting conditions.

The focal length limitation imposed on the mobile phone by the thin size flat body shape can be resolved using folded optics. This is a method whereby a mirror or prism is used to direct the light reflection from the scene to the lenses and the sensor. Tremblay *et al.* [252] examine the performance of conventional miniature refractive lenses used by mobile cameras and multiple fold reflective optics of less thickness, huge light collection, and high resolution. In miniature conventional optics, the focal length of the lens is smaller compared to the size of the pixel and its array (pixel pitch). A smaller pixel pitch means higher pixel density and higher resolution. Pixel pitch is important because it influences viewing distance. The smaller the pixel pitch, the closer the viewing distance. Folded optic creates a longer focal length

without increasing the physical distance between the lens and the image sensor. It also increases the diameter of the image sensor thereby increases the light collection of the aperture area. The effective aperture diameter of a circular folded optic lens compared to that of the miniature conventional is given as:

$$d_{eff} = d_{outer}\sqrt{1 - o^2} \quad (1)$$

where $o$ is the inner aperture diameter divided by the outer aperture diameter (obscuration ratio), $d_{outer}$ is the outer diameter (OD) of the folded optic and $d_{eff}$ is the diameter of an unobscured circular aperture of the same aperture as the folded optic. Multi-camera arrangement at the front side and the rear back of the smartphones are usually fixed at the top. Each camera lens is fixed on the same module with the sensor. The rear cameras are usually placed in a cluster whether at the top right of the phone or top center [253].

### A. MULTI-CAMERA PHONE DEPTH ESTIMATION

The presence of the homogenous or heterogeneous cameras on a phone made it possible for the camera to be able to perform depth estimation of the objects in the scene. Depth estimation is a method of using images obtained from two or more cameras to carry out survey triangulation and estimate distance [254]. This process is used to determine the distance of the object in the image from the two cameras (refers to as parallax). The objects closer to the cameras will be quite far apart in the image and those far away from the cameras will look closer in the image. One of the advantages of the depth of the object in a scene is that it allowed the introduction of special portrait modes in multi-camera phones which makes the image sharp and displays a nice blurring background. Another method of blurring the background of an image is referred to as ''Bokeh'', it is the natural method of blurring the background by using wide aperture optics. This is usually done through the hardware but hardly difficult to replicate computationally. Depth estimation is determined using the different methods; one is carried out with two or more input images (stereo vision [255] and shape from a motion [256]) and the second one is from a single monocular image which has been recently proposed [257]–[259]. The first methods produce the most accurate depth information. Guo *et al.* [260] used the Markov Random Field (MRF) model to estimate the depth map and determine the relationships between the different parts of a single image. The model of the MRF was trained using supervised learning. Then, the estimated depth details and the geometric information were used to generate a pedestrian candidate. Saxena *et al.* [254] combine monocular and stereo methods (triangulation) to estimate depth information. Markov Random Field was used to obtain the monocular cues and the result was incorporated into the stereo formation. The advantage of this approach is that the result can be applied in both areas where any of the systems (stereo and multi-camera) performs poorly. Other depth estimation approaches are summarized in Table 9.

**TABLE 9. Summary of some past works of literature on depth information estimation.**

| Authors | Sensor | Description |
|---|---|---|
| Woetzel and Koch, 2004 [261] | Stereo camera | In this research, a generalized sweep approach was employed to determine the depth map for each referenced camera. The implementation was run entirely on GPU. |
| Feldmann et al., 2010 [262] | 4-HD cameras | The approach used a depth map to estimates the related disparity between a 2D projected image of the 3D model of local conferees and a deterministic background in a real-time multi-party, multi-user terminal conference system. |
| Lee and Ho, 2011 [263] | 1-TOF camera and 5-HD video cameras | The approach obtained multi-view video with depth data using hybrid cameras (1-time of flight plus 5-video cameras). A 3D warping was performed to obtain the initial depth map at each viewpoint. The final depth map was recomputed at each segment using pairwise stereo matching using a cost function. |
| Kovacs and Zilly, 2012 [264] | 4-HD cameras | In this approach, a real-time 3D video rendering for multi-view, stereoscopic and light-field was demonstrated. A Hybrid Recursive Matcher (HRM) was used as a depth estimator. Due to its recursive structure, it can only generate a depth map for a smaller resolution. The recursive structure of HRM was broken to a line-wise level which leads to a reasonable increase in the speed of execution. |
| Stefanowski et al., 2013 [265] | Feature diffusion | The feature diffusion algorithm that was presented for fast and accurate processing of high-resolution data was applied to image-based depth estimation. Extended to cover the depth of two views. |
| G. Marin et al., 2019 [266] | Multi-camera. | The approach employs time-of-flight (ToF) sensors and stereo vision cameras as acquisition devices. The acquired dataset was carried out with a multi-camera consist of a Microsoft Kinect v2 ToF sensor, an Intel RealSense R200 active stereo sensor and a Stereolabs ZED passive stereo camera. The depth data was acquired for each scene using a line laser. |

### B. IMAGE DETAILS ENHANCEMENT THROUGH MORE CAMERAS

The image details can be enhanced when multi-camera is employed in mobile phones. For example, in image demosaicing, a typical camera sensor cannot record color on its own except an array of the color filter is laid over it. Each photosite displays in one color (Red, Green, or Blue). From this, an RGB image can be produced through a process called demosaicing. Though, this method has disadvantages such as a reduction in resolution, and less sensitive image. However, most smartphones use a color camera together with a monochrome sensor that can capture the available light and in full resolution. The image output of the two (color camera

and monochrome sensor) is then combined which produced a better-detailed image. Also, some manufacturer uses two high-resolution color cameras and then combined their output image but the process of combining and aligning is complex, the result of the combination is not as detailed as that of the monochrome sensor. For low light and high contrast condition, the combination of the image from a more light-sensitive monochrome camera with image from high color camera produce better image output. In this combination, there is a result of artifacts due to the alignment difference in the output of the two cameras.

Multi-camera phones have the advantage of using the differences in images produced by their cameras to create a depth map. This map can be used to enhance various augmented reality (AR) applications. Although it is not only through a multi-camera module that depth information can be measured, some manufacturers used a dedicated depth sensor that makes use of the time of flight (TOF) or other technologies to generate the depth map required for AR improvement.

However, the addition of more cameras to mobile phones or smartphones is in no doubt makes smartphones to be more robust but some challenges accompany it. The problem of cost and space are not the only limitations, the processing power of smartphones is also a thing of concern. The processing of multiple flows of images is significantly complex than working with images obtained from a single camera. Additional work is required to align the images obtained from multiple cameras properly to reduce the ghosting artifacts and other actions required to create a quality and detailed image output from the cameras.

## IX. MULTI-CAMERA APPLICATION ALGORITHMS

The ubiquitous nature of multicamera has made it to be widely applied in different areas of human life ranging from surveillance, sport, mobile phone, and other areas. Most multi-camera applications are based on tracking, matching, and surveillance. Human and Objects can be correctly observed, tracked, or identified using scale, appearance, and shape changes, provided they reveal enough texture. Local features are the most generally used features because they are good invariants. They are categorized into two: (i) local features based on absolute value and (ii) the one based on relative value or discriminative descriptor.

One of the most common local features techniques built on absolute value is the scale-invariant feature transform (SIFT). SIFT is invariant to image scale, rotation angle, and brightness, it can build a histogram with gray and gradient quantization. Some of the advantages of SIFT are:

- **Locality:** It recognizes all features as local, and therefore, it is a good option for occlusion and clutter (no early subdivision).
- **Distinctiveness:** Each feature can be integrated into a large database of objects.
- **Quantity:** It can work with many features.
- **Efficiency:** Its performance is comparable to that of the real-time.
- **Extensibility:** It has a wide range of support for many feature types, with each adding robustness.

However, SIFT is computationally intensive, and it is practically not implementable for real-time applications like visual odometry and low-power devices, for example, mobile phones because of its computational demand. Ke and Sukthankar [267] described a principal component analysis (PCA-SIFT) technique that substituted the histogram in SIFT which improves the computational speed. Speeded up robust features (SURF) [268] is another algorithm that has better performance than SIFT, especially in terms of computational speed. Oriented FAST and Rotated BRIEF (ORB) [269] is binary descriptors based on BRIEF, it is a robust algorithm that is illumination and rotation invariant. It is highly resistant to noise and can compute 10 times faster than SURF, but it has a scale variation problem. In general, local feature algorithms based on relative value have challenges in local feature points identification and description, local features description capability, and computational intensity. Examples of the relative value-based methods include Binary Robust Independent Elementary Features [270], modified feature point descriptor based on Binary Robust Independent Elementary Features (MBRIEF) [271], Ordinal Spatial Intensity Distribution [272], and Binary Robust invariant scalable keypoints [273].

Some researchers used the texture of an object in tracking non-rigid objects. Nummiaro *et al.* [274] implemented color distribution in particle filtering together with edge-based image features to perform real-time tracking of non-rigid objects. The method is not susceptible to partial occlusion, rotation, and scale variation and it is computationally balanced. Mathes and Piater [275] combined low-level features to form a model of shape and appearance of an object. The resulting model performs very well in serious partial occlusions images. The approach is built to detect and track texture object in a clumsy scene for non-static cameras. The algorithm was tested on 160 frame soccer sequence tracking 6 players as targets.

Others perform tracking using 2D or 3D geometry of the objects' shape. These approaches are employed in surveillance, human-computer interface (HCI), and communication support services. Senior *et al.* [276] applied tracking technology in 2D and 3D ground plane to determine the positional information of a tracking object. In their approach, four different algorithms were used in tracking a person in an indoor environment. One of the approaches is particle filtering, this method used particle trackers adapted from the work of Nickel *et al.* [277]. This method does not use a background model subtraction because of its limitation in handling moving targets. It employed frame differencing which is also susceptible to noise and distraction. The algorithms employed are background subtraction, par-

**TABLE 10. Summary of some past works of literature on depth information estimation.**

| Authors | Tracking Algorithm | Tracking features |
|---|---|---|
| Liem and Gavrila, 2014 [108] | Kalman filter | 3D position detection and appearance |
| Possegger et al., 2013 [281] | Particle filter | 3D position detection and proximity appearance. |
| Berclaz et al., 2011 [282] | Batch mode flow optimization | Probabilistic Occupancy Map (POM) [283] detections |
| Du and Piater, 2007 [115] | Multiple particle filters | Position. |
| Arsic´ et al., 2008 [284] | Kalman filter | Position and appearance |
| Calderara et al., 2008 [285] | Tracking by detection | Position and appearance |
| Hu et al., 2006 [286] | Kalman filter | Position |
| Kim and Davis, 2014 [110] | Particle filter | Position and appearance |
| Otsuka and Mukawa, 2004 [287] | particle filters | 2D detections based on visual angles. |
| Kang et al., 2004 [288] | Kalman filter | 2D and 3D positions and appearance |
| Mittal and Davis, 2003 [280] | Kalman filter | Position and velocity |

ticle filter, face detection and edge alignment of a cylindrical model. The second approach is face detection which depends on the faces detected to perform tracking. This method can be seriously affected in the case of occlusion, distance from the camera and light intensity. Next is the background subtraction approach which works on keeping a good background model that cannot be certainly guaranteed. Lastly, the edge-alignment method, it used a 3D graphical model of a human represented by cylinders coupled in kinematic chains aligned in multi-camera views. This method requires a re-initialization strategy whenever it fails. Therefore, it is not suitable for an online application. Liang *et al.* [95] presented object matching with multi-camera collaboration using head detection and trifocal tensor pointer transfer method. They used Kalman and PDA algorithms for tracking people and then applied background subtraction to detect the head position. Using the corresponding head points, trifocal tensor transfer was used to detect objects in the upper view of the two cameras. Straw *et al.* [278] body inclination of animals by tracking their 3D motion and location using the Kalman filter and the nearest neighbor filter algorithm. The system was designed to study the neurobiological behaviors of freely flying animals. The system used 11 cameras to track three flies simultaneously at 60 fps in real-time.

Subsequently, Dockstader and Tekalp [279] proposed a method of tracking persons in motion using multi-view implementation of the Bayesian belief network which integrates the 2D features of each camera view. Sparse motion image estimate and Kalman-like state propagation were used for observation and filtering respectively. Mittal and Davis [280] presented a suitable method for tracking people in a cluttered area using synchronized cameras. The system employed a region-based stereo algorithm in detecting the 3D points and Bayesian classification for segmentation. Calderara *et al.* [80] used overlapping multi-cameras to track people using consistent labeling methods. The position of the people in the multiple views of the cameras is determined using the homograph of the first detection of the camera view. Table 10 shows a summary of multi-camera application tracking algorithms.

## X. CONCLUSION

Multi-camera systems have gained significant attention during the past few years especially in the area of surveillance, tracking, image recognition, image sensor, and computer vision. This becomes an excellent opportunity where we combine the techniques (multi-camera system algorithms) and advancement in these fields with that of multi-camera systems to proffer sustainable solutions to the problems of humanity. In this survey paper, we discussed the aspect of camera calibration and architecture in a multi-camera formation because of their importance. Also, there has been a focus on the application of multi-camera systems in the area of surveillance, sports, education, and mobile phones. Multi-camera systems application algorithms are also discussed with references to the area's application. More importantly, we have discussed the current challenges faced, progresses made, and potential directions for the future to guide the researchers and scientists who are in need to understand how this area of research is evolving.

## REFERENCES

[1] VideoSurveillance.com. (2020). *What is a PTZ Camera?*. Accessed: Aug. 28, 2020. [Online]. Available: https://www.videosurveillance.com/tech/ptz-technology.asp
[2] B. Rinner, T. Winkler, W. Schriebl, M. Quaritsch, and W. Wolf, "The evolution from single to pervasive smart cameras," in *Proc. 2nd ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC)*, Stanford, CA, USA, Sep. 2008, vol. 96, no. 10, pp. 1–10, doi: 10.1109/ICDSC.2008.4635674.
[3] P. Sturm, "Camera models and fundamental concepts used in geometric computer vision," *Found. Trends Comput. Graph. Vis.*, vol. 6, nos. 1–2, pp. 1–183, 2010, doi: 10.1561/0600000023.
[4] R. Khoshabeh, T. Gandhi, and M. M. Trivedi, "Multi-camera based traffic flow characterization & classification," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Seattle, WA, USA, Sep./Oct. 2007, pp. 259–264, doi: 10.1109/ITSC.2007.4357750.
[5] O. Ozdil, B. Demirel, Y. E. Esin, and S. Ozturk, "SPARK detection with thermal camera," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, Izmir, Turkey, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404631.
[6] B. Rinner and W. Wolf, "An introduction to distributed smart cameras," *Proc. IEEE*, vol. 96, no. 10, pp. 1565–1575, Oct. 2008, doi: 10.1109/JPROC.2008.928742.
[7] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, Dec. 2006, doi: 10.1145/1177352.1177355.
[8] M. O. Mehmood, "People detection methods for intelligent multi-Camera surveillance systems," Ph.D. dissertation, Dept. Comput. Sci., Signal Image Process., Ecole Centrale de Lille, Villeneuve-d'Ascq, France, 2015.

[9] G. Tissandier, *La Photographie en Ballon*. Paris, France: Gauthier-Villars, 1884.

[10] *Robots and the Workplace of the Future*, IFR, Int. Fed. Robot., Frankfurt, Germany, Mar. 2018, pp. 1–35. [Online]. Available: https://ifr.org/downloads/papers/IFR_Robots_ and_the_Workplace_of_the_Future_Positioning_Paper.pdf

[11] P. Natarajan, P. K. Atrey, and M. Kankanhalli, "Multi-camera coordination and control in surveillance systems: A survey," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 11, no. 4, pp. 1–30, Jun. 2015, doi: 10.1145/2710128.

[12] E. Pons, "A comparison of a GPS device and a multi-camera video technology during official soccer matches: Agreement between systems," *PLoS ONE*, vol. 14, no. 8, 2019, Art. no. e0220729, doi: 10.1371/journal.pone.0220729.

[13] D. Wang, "A study on camera array and its applications," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 10323–10328, 2017, doi: 10.1016/j.ifacol.2017.08.1662.

[14] P. Kelly, C. O Conaire, C. Kim, and N. E. O'Connor, "Automatic camera selection for activity monitoring in a multi-camera system for tennis," in *Proc. 3rd ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC)*, Como, Italy, Aug. 2009, pp. 1–8, doi: 10.1109/ICDSC.2009.5289353.

[15] D. Wierzbicki, "Multi-camera imaging system for UAV photogrammetry," *Sensors*, vol. 18, no. 8, p. 2433, Jul. 2018, doi: 10.3390/s18082433.

[16] R. Iguernaissi, D. Merad, K. Aziz, and P. Drap, "People tracking in multi-camera systems: A review," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 10773–10793, Apr. 2019, doi: 10.1007/s11042-018-6638-5.

[17] Y. Wang, K. Lu, and R. Zhai, "Challenge of multi-camera tracking," in *Proc. 7th Int. Congr. Image Signal Process.*, Dalian, China, Oct. 2014, pp. 32–37, doi: 10.1109/CISP.2014.7003745.

[18] D. Devarajan, Z. Cheng, R. J. Radke, and B. D. Devarajan, "Calibrating distributed camera networks," *Proc. IEEE*, vol. 96, no. 10, pp. 1625–1639, Oct. 2008, doi: 10.1109/JPROC.2008.928759.

[19] P. Gemeiner, M. Branislav, and P. Roman, "Calibration methodology for distant surveillance cameras," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 2014, vol. 79, no. 8, pp. 162–173, doi: 10.1007/978-3-319-16199-0_12.

[20] M. Feng, X. Jia, J. Wang, S. Feng, and T. Zheng, "Global calibration of multi-cameras based on refractive projection and ray tracing," *Sensors*, vol. 17, no. 11, p. 2494, 2017, doi: 10.3390/s17112494.

[21] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000, doi: 10.1109/34.888718.

[22] R. Shen, I. Cheng, and A. Basu, "Multi-camera calibration using a globe," in *Proc. 8th Workshop Omnidirectional Vis., Camera Netw. Non-Classical Cameras (OMNIVIS)*, Marseille, France, Oct. 2008, pp. 1–11. [Online]. Available: http://hal.inria.fr/inria-00325386/

[23] MathWorks. (2020). *Camera Calibration and 3-Vision*. Accessed: Sep. 24, 2019. [Online]. Available: https://www.mathworks.com/help/vision/camera-calibration-and-3-d-vision.html

[24] R. Xia, M. Hu, J. Zhao, S. Chen, and Y. Chen, "Global calibration of multi-cameras with non-overlapping fields of view based on photogrammetry and reconfigurable target," *Meas. Sci. Technol.*, vol. 29, no. 6, Jun. 2018, Art. no. 065005, doi: 10.1088/1361-6501/aab028.

[25] Agrawal and Davis, "Camera calibration using spheres: A semi-definite programming approach," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 2. Nice, France, Oct. 2003, pp. 782–789, doi: 10.1109/ICCV.2003.1238428.

[26] P. F. Sturm and S. J. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Fort Collins, CO, United States, 1999, pp. 432–437, doi: 10.1109/CVPR.1999.786974.

[27] P. Hammarstedt, P. Sturm, A. Heyden, "Degenerate cases and closed-form solutions for camera calibration with one-dimensional objects," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, vol. 1. Beijing, China, Oct. 2005, pp. 317–324, doi: 10.1109/ICCV.2005.68.

[28] R. I. Hartley, "Self-calibration of stationary cameras," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 5–23, 1997, doi: 10.1023/A:1007957826135.

[29] R. S. Lu and Y. F. Li, "A global calibration method for large-scale multi-sensor visual measurement systems," *Sens. Actuators A, Phys.*, vol. 116, no. 3, pp. 384–393, Oct. 2004, doi: 10.1016/j.sna.2004.05.019.

[30] Y. Zhao, F. Yuan, Z. Ding, and J. Li, "Global calibration method for multi-vision measurement system under the conditions of large field of view," *Yingyong Jichu yu Gongcheng Kexue Xuebao/J. Basic Sci. Eng.*, vol. 19, no. 4, pp. 679–688, 2011, doi: 10.3969/j.issn.1005-0930.2011.04.018.

[31] P. Lébraly, C. Deymier, O. Ait-Aider, E. Royer, and M. Dhome, "Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: Application to vision-based robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Taipei, Taiwan, Oct. 2010, pp. 5640–5647, doi: 10.1109/IROS.2010.5651552.

[32] Z. Xu, Y. Wang, and C. Yang, "Multi-camera global calibration for large-scale measurement based on plane mirror," *Optik*, vol. 126, no. 23, pp. 4149–4154, Dec. 2015, doi: 10.1016/j.ijleo.2015.08.015.

[33] H. Huang, N. Li, H. Guo, Y.-L. Chen, and X. Wu, "Calibration of non-overlapping cameras based on a mobile robot," in *Proc. 5th Int. Conf. Inf. Sci. Technol. (ICIST)*, Changsa, China, Apr. 2015, pp. 328–333, doi: 10.1109/ICIST.2015.7288991.

[34] Q. Wang and Y. Liu, "A tractable mechanism for external calibration in non-overlapping camera network," in *Proc. 6th Int. ICST Conf. Commun. Netw. China (CHINACOM)*, Harbin, China, Aug. 2011, pp. 893–898, doi: 10.1109/ChinaCom.2011.6158281.

[35] F. Zhao, T. Tamaki, T. Kurita, B. Raytchev, and K. Kaneda, "Marker based simple non-overlapping camera calibration," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 1180–1184, doi: 10.1109/ICIP.2016.7532544.

[36] W. Zou and S. Li, "Calibration of nonoverlapping in-vehicle cameras with laser pointers," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1348–1359, Jun. 2015, doi: 10.1109/TITS.2014.2361666.

[37] W. Zou, "Calibrating non-overlapping cameras with a laser ray," Ph.D. dissertation, Dept. Elect. Electron. Eng., Tottori Univ., Tottori, Japan, 2015.

[38] Z. Liu, G. Zhang, Z. Wei, and J. Sun, "A global calibration method for multiple vision sensors based on multiple targets," *Meas. Sci. Technol.*, vol. 22, no. 12, Dec. 2011, Art. no. 125102, doi: 10.1088/0957-0233/22/12/125102.

[39] R. Horaud, R. Mohr, and B. Lorecki, "Linear camera calibration," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 2. Nice, France, Apr. 2004, pp. 1539–1544, doi: 10.1109/ROBOT.1992.220033.

[40] A. Basu and K. Ravi, "Active camera calibration using pan, tilt and roll," *IEEE Trans. Syst., Man Cybern., B (Cybern.)*, vol. 27, no. 3, pp. 559–566, Jun. 1997, doi: 10.1109/3477.584964.

[41] N. A. Borghese, F. M. Colombo, and A. Alzati, "Computing camera focal length by zooming a single point," *Pattern Recognit.*, vol. 39, no. 8, pp. 1522–1529, Aug. 2006, doi: 10.1016/j.patcog.2006.01.011.

[42] L. Wang, W. Wang, C. Shen, and F. Duan, "A convex relaxation optimization algorithm for multi-camera calibration with 1D objects," *Neurocomputing*, vol. 215, pp. 82–89, Nov. 2016, doi: 10.1016/j.neucom.2015.07.158.

[43] E. Shen and R. Hornsey, "Multi-camera network calibration with a non-planar target," *IEEE Sensors J.*, vol. 11, no. 10, pp. 2356–2364, Oct. 2011, doi: 10.1109/JSEN.2011.2123884.

[44] Z. Liu, F. Li, and G. Zhang, "An external parameter calibration method for multiple cameras based on laser rangefinder," *Measurement*, vol. 47, pp. 954–962, Jan. 2014, doi: 10.1016/j.measurement.2013.10.029.

[45] Q. Fu, K.-Y. Cai, and Q. Quan, "Calibration of multiple fish-eye cameras using a wand," *IET Comput. Vis.*, vol. 9, no. 3, pp. 378–389, Jun. 2015, doi: 10.1049/iet-cvi.2014.0181.

[46] M. E. Loaiza, A. B. Raposo, and M. Gattass, "Multi-camera calibration based on an invariant pattern," *Comput. Graph.*, vol. 35, no. 2, pp. 198–207, Apr. 2011, doi: 10.1016/j.cag.2010.12.007.

[47] Z. Zhang, "Camera calibration with one-dimensional objects," in *Lecture Notes Computer Science (Including Subser. Lect. Notes Artifical Intelligent Lecture Notes Bioinformatics)*, vol. 2353. Berlin, Germany: Springer, Dec. 2002, pp. 161–174.

[48] J. A. de França, M. R. Stemmer, M. B. D. M. França, and J. C. Piai, "A new robust algorithmic for multi-camera calibration with a 1D object under general motions without prior knowledge of any camera intrinsic parameter," *Pattern Recognit.*, vol. 45, no. 10, pp. 3636–3647, Oct. 2012, doi: 10.1016/j.patcog.2012.04.006.

[49] K.-Y. Shin and J. H. Mun, "A multi-camera calibration method using a 3-axis frame and wand," *Int. J. Precis. Eng. Manuf.*, vol. 13, no. 2, pp. 283–289, Feb. 2012, doi: 10.1007/s12541-012-0035-1.

[50] L. Quan and Z. Lan, "Linear N-point camera pose determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 774–780, Aug. 1999, doi: 10.1109/34.784291.

[51] G. Xu, X. Zhang, X. Li, J. Su, and Z. Hao, "Global calibration method of a camera using the constraint of line features and 3D world points," *Meas. Sci. Rev.*, vol. 16, no. 4, pp. 190–196, Aug. 2016, doi: 10.1515/msr-2016-0023.

[52] B. Rinner and W. Wolf, "Toward pervasive smart camera networks," in *Multi-Camera Networks*. New York, NY, USA: Academic, 2009, pp. 483–496, doi: 10.1016/B978-0-12-374633-7.00022-7.
[53] Y. Li and B. Bhanu, "A comparison of techniques for camera selection and handoff in a video network," in *Proc. 3rd ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC)*, Como, Italy, Aug. 2009, pp. 1–8, doi: 10.1109/ICDSC.2009.5289342.
[54] S. Kang, J.-K. Paik, A. Koschan, B. R. Abidi, and M. A. Abidi, "Real-time video tracking using PTZ cameras," *Proc. SPIE*, vol. 5132, May 2003, Art. no. 514945, doi: 10.1117/12.514945.
[55] S.-N. Lim, L. S. Davis, and A. Elgammal, "Scalable image-based multi-camera visual surveillance system," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Miami, FL, USA, Jul. 2003, pp. 205–212, doi: 10.1109/AVSS.2003.1217923.
[56] V. Kettnaker and R. Zabih, "Bayesian multi-camera surveillance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Fort Collins, CO, USA, Jun. 1999, pp. 117–123, doi: 10.1109/CVPR.1999.784638.
[57] Y. Lu and S. Payandeh, "Cooperative hybrid multi-camera tracking for people surveillance," *Can. J. Electr. Comput. Eng.*, vol. 33, nos. 3–4, pp. 145–152, 2008, doi: 10.1109/CJECE.2008.4721631.
[58] I. Everts, N. Sebe, and G. A. Jones, "Cooperative object tracking with multiple PTZ cameras," in *Proc. 14th Int. Conf. Image Anal. Process (ICIAP)*, Modena, Italy, 2007, pp. 323–330, doi: 10.1109/ICIAP.2007.46.
[59] E. Sommerlade and I. Reid, "PhD forum: Probabilistic surveillance with multiple active cameras," in *Proc. 3rd ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC)*, Como, Italy, Aug. 2009, pp. 1–2, doi: 10.1109/ICDSC.2009.5289403.
[60] C. Piciarelli, C. Micheloni, and G. L. Foresti, "PTZ camera network reconfiguration," in *Proc. 3rd ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC)*, Como, Italy, Sep. 2009, pp. 1–7, doi: 10.1109/ICDSC.2009.5289419.
[61] H.-S. Kim, C. Nam, K.-Y. Ha, O. Ayurzana, and J.-W. Kwon, "An algorithm of a real time image tracking system using a camera with pan/tilt motors on an embedded system," in *Proc. ICMIT: Inf. Syst. Signal Process.*, vol. 6041, Dec. 200, Art. no. 604112, doi: 10.1117/12.664317.
[62] M. Quaritsch, M. Kreuzthaler, B. Rinner, H. Bischof, and B. Strobl, "Autonomous multicamera tracking on embedded smart cameras," *EURASIP J. Embedded Syst.*, vol. 2007, Jan. 2007, Art. no. 092827, doi: 10.1155/2007/92827.
[63] B. Rinner, M. Jovanovic, and M. Quaritsch, "Embedded middleware on distributed smart cameras," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, p. 1381, doi: 10.1109/ICASSP.2007.367336.
[64] S. Fleck and W. Strasser, "Adaptive probabilistic tracking embedded in a smart camera," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Sep. 2005, p. 134, doi: 10.1109/cvpr.2005.404.
[65] S. Fleck, F. Busch, P. Biber, and W. Straßer, "3D surveillance a distributed network of smart cameras for real-time tracking and its visualization in 3D," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, New York, NY, USA, Jun. 2006, p. 118, doi: 10.1109/CVPRW.2006.6.
[66] F. Z. Qureshi and D. Terzopoulos, "Multi-camera control through constraint satisfaction for persistent surveillance," in *Proc. IEEE 5th Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2008, pp. 211–218, doi: 10.1109/AVSS.2008.37.
[67] C. Micheloni, G. L. Foresti, L. Snidaro, "A network of co-operative cameras for visual surveillance," *IEE Proc.-Vis., Image Signal Process.*, vol. 152, no. 2, pp. 205–212, 2005, doi: 10.1049/ip-vis:20041256.
[68] K. Morioka, S. Kovacs, J.-H. Lee, P. Korondi, and H. Hashimoto, "Fuzzy-based camera selection for object tracking in a multi-camera system," in *Proc. Conf. Hum. Syst. Interact.*, Krakow, Poland, May 2008, pp. 767–772, doi: 10.1109/HSI.2008.4581538.
[69] J. Park, P. C. Bhat, and A. C. Kak, "A look-up table based approach for solving the camera selection problem in large camera networks," in *Proc. Workshop Distrib. Smart Cameras*, 2006, pp. 1–5.
[70] L. Hodge and M. Kamel, "An agent-based approach to multisensor coordination," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 33, no. 5, pp. 648–662, Sep. 2003, doi: 10.1109/TSMCA.2003.817397.
[71] Y. Li and B. Bhanu, "Utility-based camera assignment in a video network: A game theoretic framework," *IEEE Sensors J.*, vol. 11, no. 3, pp. 676–687, Mar. 2011, doi: 10.1109/JSEN.2010.2051148.
[72] B. Song, C. Soto, A. K. Roy-Chowdhury, and J. A. Farrell, "Decentralized camera network control using game theory," in *Proc. 2nd ACM/IEEE Int. Conf. Distrib. Smart Cameras*, Stanford, CA, USA, Sep. 2008, pp. 1–8, doi: 10.1109/ICDSC.2008.4635735.
[73] A. Prati, R. Vezzani, L. Benini, E. Farella, and P. Zappi, "An integrated multi-modal sensor network for video surveillance," in *Proc. 3rd ACM Int. Workshop Video Surveill. Sensor Netw. (VSSN)*, 2005, pp. 95–102, doi: 10.1145/1099396.1099415.
[74] T. Matsuyama and N. Ukita, "Real-time multitarget tracking by a cooperative distributed vision system," *Proc. IEEE*, vol. 90, no. 7, pp. 1136–1150, Jul. 2002, doi: 10.1109/JPROC.2002.801442.
[75] M. Bramberger, M. Quaritsch, T. Winkler, B. Rinner, and H. Schwabach, "Integrating multi-camera tracking into a dynamic task allocation system for smart cameras," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Como, Italy, Sep. 2005, pp. 474–479, doi: 10.1109/AVSS.2005.1577315.
[76] N. Hideyuki, A. Hamid, and A. Juan Carlos, *Handbook of Ambient Intelligence and Smart Environments*. New York, NY, USA: Springer, 2010, doi: 10.1007/978-0-387-93808-0.
[77] M. Evans, C. J. Osborne, and J. Ferryman, "Multicamera object detection and tracking with object size estimation," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Krakow, Poland, Aug. 2013, pp. 177–182, doi: 10.1109/AVSS.2013.6636636.
[78] Q. Cai and J. K. Aggarwal, "Tracking human motion in structured environments using a distributed-camera system," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 11, pp. 1241–1247, Nov. 1999, doi: 10.1109/34.809119.
[79] M. Liem and D. M. Gavrila, "Multi-person tracking with overlapping cameras in complex, dynamic environments," in *Proc. Procedings Brit. Mach. Vis. Conf.*, 2009, pp. 1–10, doi: 10.5244/C.23.87.
[80] S. Calderara, P. Andrea, V. Roberto, and C. Rita, "Consistent labeling for multi-camera object tracking," in *Image Analysis and Processing (ICIAP)* (Lecture Notes in Computer Science), vol. 3617. Berlin, Germany: Springer, 2005, pp. 1133–1139, doi: 10.1007/11553595_139.
[81] M. Taj and A. Cavallaro, "Distributed and decentralized multicamera tracking," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 46–58, May 2011, doi: 10.1109/MSP.2011.940281.
[82] H. Medeiros, J. Park, and A. Kak, "Distributed object tracking using a cluster-based Kalman filter in wireless camera networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 4, pp. 448–463, Aug. 2008, doi: 10.1109/JSTSP.2008.2001310.
[83] J. Yoder, H. Medeiros, J. Park, and A. C. Kak, "Cluster-based distributed face tracking in camera networks," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2551–2563, Oct. 2010, doi: 10.1109/TIP.2010.2049179.
[84] D. Makris, T. Ellis, and J. Black, "Bridging the gaps between cameras," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Washington, DC, USA, Jun./Jul. 2004, p. 11, doi: 10.1109/cvpr.2004.1315165.
[85] C. Stauffer, "Learning to track objects through unobserved regions," in *Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION)*, vol. 1. Breckenridge, CO, USA, Jan. 2005, pp. 96–102, doi: 10.1109/ACVMOT.2005.69.
[86] A. Rahimi, B. Dunagan, and T. Darrell, "Simultaneous calibration and tracking with a network of non-overlapping sensors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Washington, DC, USA, Jun./Jul. 2004, p. 1, doi: 10.1109/cvpr.2004.1315031.
[87] S. Zhang, Y. Zhu, and A. Roy-Chowdhury, "An online learned elementary grouping model for multi-target tracking," *Comput. Vis. Image Understand.*, vol. 134, pp. 64–73, May 2015, doi: 10.1016/j.cviu.2015.01.002.
[88] Y. T. Tesfaye, E. Zemene, A. Prati, M. Pelillo, and M. Shah, "Multi-target tracking in multiple non-overlapping cameras using fast-constrained dominant sets," *Int. J. Comput. Vis.*, vol. 127, no. 9, pp. 1303–1320, Sep. 2019, doi: 10.1007/s11263-019-01180-6.
[89] C. H. Anderson, P. J. Burt, and G. S. van der Wal, "Change detection and tracking using pyramid transform techniques," in *Intelligent Robots and Computer Vision IV*, vol. 0579. Cambridge, MA, USA: SPIE, 1985, pp. 72–78, doi: 10.1117/12.950785.
[90] M. Casares, S. Velipasalar, and A. Pinto, "Light-weight salient foreground detection for embedded smart cameras," *Comput. Vis. Image Understand.*, vol. 114, no. 11, pp. 1223–1237, Nov. 2010, doi: 10.1016/j.cviu.2010.03.023.

[91] I. Haritaoglu, D. Harwood, and L. S. Davis, "W$^4$: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 809–830, Aug. 2000, doi: 10.1109/34.868683.

[92] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000, doi: 10.1109/34.868677.

[93] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, Dec. 2006, doi: 10.1145/1177352.1177355.

[94] A. W. Senior, G. Potamianos, S. Chu, Z. Zhang, A. Hampapur, and Y. Heights, "A comparison of multicamera person-tracking algorithms," in *Proc. IEEE Int. Workshop Vis. Surveill. (VS/ECCV)*, May 2006, pp. 1–7. [Online]. Available: http://andrewsenior.com/papers/SeniorVS06.pdf

[95] H. Liang, Y.-H. Liu, and X.-P. Cai, "Multi-camera collaboration based on trifocal tensor transfer," *J. Softw.*, vol. 20, no. 9, pp. 2597–2606, Nov. 2009, doi: 10.3724/SP.J.1001.2009.03571.

[96] S. M. Khan and M. Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," in *Lecture Notes Computure Science (Including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 3954. 2006, pp. 133–146, doi: 10.1007/11744085_11.

[97] D. Figueira, L. Bazzani, H. Q. Minh, M. Cristani, A. Bernardino, and V. Murino, "Semi-supervised multi-feature learning for person re-identification," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Krakow, Poland Aug. 2013, pp. 111–116, doi: 10.1109/AVSS.2013.6636625.

[98] D. Baltieri, R. Vezzani, R. Cucchiara, A. Utasi, C. Benedek, and T. Sziranyi, "Multi-view people surveillance using 3D information," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Barcelona, Spain. Piscataway, NJ, USA: IEEE Press, Nov. 2011, pp. 1817–1824, doi: 10.1109/ICCVW.2011.6130469.

[99] H. Q. Minh, L. Bazzani, and V. Murino, "A unifying framework for vector-valued manifold regularization and multi-view learning," in *Proc. 30th Int. Conf. Mach. Learn. (ICML)*, vol. 28, no. 2. Phoenix, AZ, USA: PMLR, 2013, pp. 100–108.

[100] W. Chen, L. Cao, X. Chen, and K. Huang, "A novel solution for multi-camera object tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 2329–2333, doi: 10.1109/ICIP.2014.7025472.

[101] W. Li, Y. Wu, M. Mukunoki, and M. Minoh, "Common-near-neighbor analysis for person re-identification," in *Proc. 19th IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 1621–1624, doi: 10.1109/ICIP.2012.6467186.

[102] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Lecture Notes Computer Science (Including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 6688. Berlin, Germany: Springer, Jul. 2014, pp. 91–102, 2011, doi: 10.1007/978-3-642-21227-7_9.

[103] H. Bouma, S. Borsboom, R. J. M. den Hollander, S. H. Landsmeer, and M. Worring, "Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination," *Proc. SPIE*, vol. 8359, Jun. 2012, Art. no. 83590Q, doi: 10.1117/12.918576.

[104] H. Wang, Y. Yan, J. Hua, Y. Yang, X. Wang, X. Li, J. R. Deller, G. Zhang, and H. Bao, "Pedestrian recognition in multi-camera networks using multilevel important salient feature and multicategory incremental learning," *Pattern Recognit.*, vol. 67, pp. 340–352, Jul. 2017, doi: 10.1016/j.patcog.2017.01.033.

[105] L. Wen, Z. Lei, M. C. Honggang, and Q. Siwei, "Multi-camera multi-target tracking with space-time-view hyper-graph," *Int. J. Comput. Vis.*, vol. 122, no. 2, pp. 313–333, 2016, doi: 10.1007/s11263-016-0943-0.

[106] W. Brendel, M. Amer, and S. Todorovic, "Multiobject tracking as maximum weight independent set," in *Proc. CVPR*, Providence, RI, USA, Jun. 2011, pp. 1273–1280, doi: 10.1109/CVPR.2011.5995395.

[107] C. C. Huang and S. J. Wang, "A Bayesian hierarchical framework for multitarget labeling and correspondence with ghost suppression over multicamera surveillance system," *IEEE Trans. Automat. Sci. Eng.*, vol. 9, no. 1, pp. 16–30, Jan. 2012, doi: 10.1109/TASE.2011.2163197.

[108] M. C. Liem and D. M. Gavrila, "Joint multi-person detection and tracking from overlapping cameras," *Comput. Vis. Image Understand.*, vol. 128, pp. 36–50, Nov. 2014, doi: 10.1016/j.cviu.2014.06.003.

[109] L. Guan, J.-S. Franco, and M. Pollefeys, "Multi-view occlusion reasoning for probabilistic silhouette-based dynamic scene reconstruction," *Int. J. Comput. Vis.*, vol. 90, no. 3, pp. 283–303, Dec. 2010, doi: 10.1007/s11263-010-0341-y.

[110] K. Kim and L. S. Davis, "Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering," *Lecture Notes Computer Science (including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 3953. Berlin, Germany: Springer, Apr. 2014, pp. 98–109, 2006, doi: 10.1007/11744078_8.

[111] W. Du and J. Piater, "Data fusion by belief propagation for multi-camera tracking," in *Proc. 9th Int. Conf. Inf. Fusion*, Florence, Italy, Jul. 2006, pp. 1–8, doi: 10.1109/ICIF.2006.301712.

[112] J. Yao and J.-M. Odobez, "Multi-camera multi-person 3D space tracking with MCMC in surveillance scenarios," in *Proc. Eur. Conf. Comput. Vis., Work. Multi Camera Multi-modal Sens. Fusion Algorithms Appl.*, Marseille, France, 2008, pp. 1–12.

[113] M. Hofmann, D. Wolf, and G. Rigoll, "Hypergraphs for joint multi-view reconstruction and multi-object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 3650–3657, doi: 10.1109/CVPR.2013.468.

[114] X. Jiang, M. Körner, D. Haase, and J. Denzler, "A graph-based map solution for multi-person tracking using multi-camera systems," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, Lisbon, Portugal, Jun. 2014, pp. 343–350.

[115] W. Du and J. Piater, "Multi-camera people tracking by collaborative particle filters and principal axis-based integration," *Lecture Notes Computer Science (including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 4843. Berlin, Germany: Springer, 2007, pp. 365–374, doi: 10.1007/978-3-540-76386-4_34.

[116] Y. Chen, X. Zhu, and S. Gong, "Person re-identification by deep learning multi-scale representations," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Venice, Italy, Oct. 2017, pp. 2590–2600, doi: 10.1109/ICCVW.2017.304.

[117] M. Li, L.-L. Zhang, and Z.-P. Wang, "Single/cross-camera multiple-person traffic tracking by graph matching," in *Proc. Int. Conf. Electron., Control, Autom. Mech. Eng.*, Sanya, China, vol. 2017, pp. 479–482.

[118] R. Mazzon and A. Cavallaro, "Multi-camera tracking using a multi-goal social force model," *Neurocomputing*, vol. 100, pp. 41–50, Jan. 2013, doi: 10.1016/j.neucom.2011.09.038.

[119] D.-N. Truong Cong, L. Khoudour, C. Achard, C. Meurie, and O. Lezoray, "People re-identification by spectral classification of silhouettes," *Signal Process.*, vol. 90, no. 8, pp. 2362–2374, Aug. 2010, doi: 10.1016/j.sigpro.2009.09.005.

[120] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views," *Comput. Vis. Image Understand.*, vol. 109, no. 2, pp. 146–162, 2008, doi: 10.1016/j.cviu.2007.01.003.

[121] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Boston, MA, USA, Aug. 2010, pp. 435–440, doi: 10.1109/AVSS.2010.34.

[122] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 2360–2367, doi: 10.1109/CVPR.2010.5539926.

[123] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *Proc. Brit. Mach. Vis. Conf.*, 2010, pp. 1–11, doi: 10.5244/C.24.21.

[124] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 898–903, May 2012, doi: 10.1016/j.patrec.2011.11.016.

[125] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 381–390.

[126] V. S. Kenk, R. Mandeljc, S. Kova i , M. Kristan, M. Hajdinjak, and J. Perš, "Visual re-identification across large, distributed camera networks," *Image Vis. Comput.*, vol. 34, pp. 11–26, Feb. 2015, doi: 10.1016/j.imavis.2014.11.002.

[127] K. Jungling, C. Bodensteiner, and M. Arens, "Person re-identification in multi-camera networks," in *Proc. CVPR WORKSHOPS*, Colorado Springs, CO, USA, Jun. 2011, pp. 55–61, doi: 10.1109/CVPRW.2011.5981771.

[128] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Person re-identification using Haar-based and DCD-based signature," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Boston, MA, USA, Aug. 2010, pp. 1–8, doi: 10.1109/AVSS.2010.68.

[129] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *Proc. XXII Brazilian Symp. Comput. Graph. Image Process.*, Oct. 2009, pp. 322–329, doi: 10.1109/SIBGRAPI.2009.42.

[130] L. Patino and J. Ferryman, "Multicamera trajectory analysis for semantic behaviour characterisation," in *Proc. 11th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Seoul, South Korea, Aug. 2014, pp. 369–374, doi: 10.1109/AVSS.2014.6918696.

[131] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portlland, OR, USA, Jun. 2013, pp. 3610–3617, doi: 10.1109/CVPR.2013.463.

[132] D. Forsyth, "Object detection with discriminatively trained part-based models," *Computer*, vol. 47, no. 2, pp. 6–7, Feb. 2014, doi: 10.1109/MC.2014.42.

[133] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 1, no. 1, pp. 1–8, 1954.

[134] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Using kernel-based metric learning methods," in *Proc. Eur. Conf. Com. Vis. (ECCV)*, 2014, pp. 1–16, doi: 10.1007/978-3-319-10584-0_1.

[135] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portlland, OR, USA, Jun. 2013, pp. 3318–3325, doi: 10.1109/CVPR.2013.426.

[136] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," *Lecture Notes Computer Science (including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 7577. Berlin, Germany: Springer, 2012, pp. 780–793, doi: 10.1007/978-3-642-33783-3_56.

[137] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 2288–2295, doi: 10.1109/CVPR.2012.6247939.

[138] A. Schumann and E. Monari, "A soft-biometrics dataset for person tracking and re-identification," in *Proc. 11th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Seoul, South Korea, Aug. 2014, pp. 193–198, doi: 10.1109/AVSS.2014.6918667.

[139] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Multi-type attributes driven multi-camera person re-identification," *Pattern Recognit.*, vol. 75, pp. 77–89, Mar. 2018, doi: 10.1016/j.patcog.2017.07.005.

[140] X. Qian, Y. Fu, Y.-G. Jiang, T. Xiang, and X. Xue, "Multi-scale deep learning architectures for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5409–5418, doi: 10.1109/ICCV.2017.577.

[141] C.-H. Kuo, C. Huang, and R. Nevatia, "Multi-target tracking by on-line learned discriminative appearance models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA,Jun. 2010, pp. 685–692, doi: 10.1109/CVPR.2010.5540148.

[142] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio, "Full-body person recognition system," *Pattern Recognit.*, vol. 36, no. 9, pp. 1997–2006, 2003, doi: 10.1016/S0031-3203(03)00061-X.

[143] N. Martinel, C. Micheloni, and G. L. Foresti, "Kernelized saliency-based person re-identification through multiple metric learning," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5645–5658, Dec. 2015, doi: 10.1109/TIP.2015.2487048.

[144] M. Gou, S. Karanam, W. Liu, O. Camps, and R. J. Radke, "DukeMTMC4ReID: A large-scale multi-camera person re-identification dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1425–1434, doi: 10.1109/CVPRW.2017.185.

[145] Y. Chang, "Graph embedding and extensions: A general framework for dimensionality reduction," Dept. ECE, Northeastern Univ., Boston, MA, USA, Mar. 2014. Accessed: Jul. 30, 2020. [Online]. Available: http://www1.ece.neu.edu/~ychang/notes/dim_reduction.pdf

[146] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 2666–2672, doi: 10.1109/CVPR.2012.6247987.

[147] H. Dong, P. Lu, S. Zhong, C. Liu, Y. Ji, and S. Gong, "Person re-identification by enhanced local maximal occurrence representation and generalized similarity metric learning," *Neurocomputing*, vol. 307, pp. 25–37, Sep. 2018, doi: 10.1016/j.neucom.2018.04.013.

[148] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 5302, 2008, pp. 262–275, doi: 10.1007/978-3-540-88682-2_21.

[149] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by Fisher Vectors for person re-identification," *Lecture Notes Computer Science (including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 7583. 2012, pp. 413–422, doi: 10.1007/978-3-642-33863-2_41.

[150] B. Ma, Y. Su, and F. Jurie, "BiCov: A novel image representation for person re-identification and face verification," in *Proc. Brit. Mach. Vis. Conf.* London, U.K.: BMVA Press, 2012, pp. 1–11, doi: 10.5244/C.26.57.

[151] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 3586–3593, doi: 10.1109/CVPR.2013.460.

[152] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1363–1372, doi: 10.1109/CVPR.2016.152.

[153] K.-E. Aziz, D. Merad, and B. Fertil, "People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation," in *Proc. 8th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Klagenfurt, Austria, Aug. 2011, pp. 303–308, doi: 10.1109/AVSS.2011.6027341.

[154] C. Hong, J. Yu, D. Tao, and M. Wang, "Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3742–3751, Jun. 2015, doi: 10.1109/TIE.2014.2378735.

[155] C. Hong, J. Yu, J. Wan, D. Tao, and M. Wang, "Multimodal deep autoencoder for human pose recovery," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5659–5670, Dec. 2015, doi: 10.1109/TIP.2015.2487860.

[156] A. Alahi, P. Vandergheynst, M. Bierlaire, and M. Kunt, "Cascade of descriptors to detect and track objects across any network of cameras," *Comput. Vis. Image Understand.*, vol. 114, no. 6, pp. 624–640, Jun. 2010, doi: 10.1016/j.cviu.2010.01.004.

[157] Z. Wang, R. Hu, C. Liang, Y. Yu, J. Jiang, M. Ye, J. Chen, and Q. Leng, "Zero-shot person re-identification via cross-view consistency," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 260–272, Feb. 2016, doi: 10.1109/TMM.2015.2505083.

[158] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *Proc. 2nd ACM/IEEE Int. Conf. Distrib. Smart Cameras*, Stanford, CA, USA, Sep. 2008, pp. 1–6, doi: 10.1109/ICDSC.2008.4635689.

[159] T. D'Orazio, P. L. Mazzeo, and P. Spagnolo, "Color brightness transfer function evaluation for non overlapping multi camera tracking," in *Proc. 3rd ACM/IEEE Int. Conf. Distrib. Smart Cameras (ICDSC)*, Como, Italy, Aug. 2009, pp. 1–9, doi: 10.1109/ICDSC.2009.5289365.

[160] A. Chilgunde, P. Kumar, S. Ranganath, and H. Weimin, "Multi-camera target tracking in blind regions of cameras with non-overlapping fields of view," in *Proc. Brit. Mach. Vis. Conf.* Fountain Valley, CA, USA: Kingston, 2004, p. 42, doi: 10.5244/c.18.42.

[161] D.-T. Lin and K.-Y. Huang, "Collaborative pedestrian tracking and data fusion with multiple cameras," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1432–1444, Dec. 2011, doi: 10.1109/TIFS.2011.2159972.

[162] W. Leoputra, T. Tan, and F. Lee Lim, "Non-overlapping distributed tracking using particle filter," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, 2006, pp. 181–185, doi: 10.1109/ICPR.2006.862.

[163] M. Bauml, K. Bernardin, M. Fischer, H. K. Ekenel, and R. Stiefelhagen, "Multi-pose face recognition for person retrieval in camera networks," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Boston, MA, USA, Aug. 2010, pp. 441–447, doi: 10.1109/AVSS.2010.42.

[164] M. Valera and S. A. Velastin, "Intelligent distributed surveillance systems: A review," *IEE Proc.-Vis., Image Signal Process.*, vol. 152, no. 2, pp. 192–204, 2005, doi: 10.1049/ip-vis:20041147.

[165] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1114–1127, Aug. 2008, doi: 10.1109/TCSVT.2008.927109.

[166] B. R. Abidi, N. R. Aragam, Y. Yao, and M. A. Abidi, "Survey and analysis of multimodal sensor planning and integration for wide area surveillance," *ACM Comput. Surv.*, vol. 41, no. 1, pp. 1–36, Jan. 2009, doi: 10.1145/1456650.1456657.

[167] Y. Sheikh, O. Javed, and M. Shah, *Object Association Across Multiple Cameras*. Amsterdam, The Netherlands: Elsevier, 2009.

[168] H. Aghajan and A. Cavallaro, *Multi-Camera Networks: Principles and Applications*. New York, NY, USA: Academic, 2009.

[169] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi, and S. G. Kong, "Intelligent visual surveillance–A survey," *Int. J. Control, Automat. Syst.*, vol. 8, no. 5, pp. 926–939, 2010, doi: 10.1007/s12555-010-0501-4.

[170] A. Seema and M. Reisslein, "Towards efficient wireless video sensor networks: A survey of existing node architectures and proposal for a flexi-WVSNP design," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 462–486, 3rd Quart., 2011, doi: 10.1109/SURV.2011.102910.00098.

[171] J. N. Castaneda, V. Jelaca, A. Frias, A. Pizurica, W. Philips, R. R. Cabrera, and T. Tuytelaars, "Non-overlapping multi-camera detection and tracking of vehicles in tunnel surveillance," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, Noosa, QLD, Australia, Dec. 2011, pp. 591–596, doi: 10.1109/DICTA.2011.105.

[172] B. Tavli, K. Bicakci, R. Zilan, and J. M. Barcelo-Ordinas, "A survey of visual sensor network platforms," *Multimedia Tools Appl.*, vol. 60, no. 3, pp. 689–726, Oct. 2012, doi: 10.1007/s11042-011-0840-z.

[173] A. K. Roy-Chowdhury and B. Song, *Camera Networks: The Acquisition and Analysis of Videos over Wide Areas*. California, CA, USA: Morgan & Claypool, 2012.

[174] T. Winkler and B. Rinner, "Security and privacy protection in visual sensor networks?: A survey," *ACM Trans. Sens. Netw.*, vol. 47, no. 1, pp. 1–42, 2012.

[175] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–37, Nov. 2013, doi: 10.1145/2543581.2543596.

[176] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, Jan. 2013, doi: 10.1016/j.patrec.2012.07.005.

[177] J. C. Sanmiguel, C. Micheloni, K. Shoop, G. L. Foresti, and A. Cavallaro, "Self-reconfigurable smart camera networks," *Computer*, vol. 47, no. 5, pp. 67–73, May 2014, doi: 10.1109/MC.2014.133.

[178] Z. Jin, L. An, and B. Bhanu, "Group structure preserving pedestrian tracking in a multicamera video network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2165–2176, Oct. 2017, doi: 10.1109/TCSVT.2016.2565998.

[179] M. Weiser, "The computer for the 21st century," *Sci. Amer.*, vol. 265, no. 3, pp. 94–103, 1991.

[180] R. Ren and J. Joe, "General highlight detection in sport videos," in *Proc. 15th Int. Conf. Multimedia Modeling*, 2009, vol. 9, no. 3, pp. 27–38.

[181] F. Yan, W. Christmas, and J. Kittler, "Layered data association using graph-theoretic formulation with application to tennis ball tracking in monocular sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1814–1830, Oct. 2008.

[182] C. J. Needham and R. D. Boyle, "Tracking multiple sports players through occlusion, congestion and scale," in *Proc. Brit. Mach. Vis. Conf.* Manchester, U.K.: Univ. Manchester, 2001, p. 11, doi: 10.5244/c.15.11.

[183] M. Xu, J. Orwell, L. Lowey, and D. Thirde, "Architecture and algorithms for tracking football players with multiple cameras," *IEE Proc.-Vis., Image Signal Process.*, vol. 152, no. 2, pp. 232–241, Apr. 2005, doi: 10.1049/ip-vis:20041257.

[184] Q. Cai and J. K. Aggarwal, "Tracking human motion using multiple cameras," in *Proc. 13th Int. Conf. Pattern Recognit.*, Vienna, Austria, 1996, pp. 68–72, doi: 10.1109/ICPR.1996.546796.

[185] R. Cavallaro, "The FoxTrax hockey puck tracking system," *IEEE Comput. Graph. Appl.*, vol. 17, no. 2, pp. 6–12, Mar./Apr. 1997, doi: 10.1109/38.574652.

[186] T. Bebie and H. Bieri, "SoccerMan-reconstructing soccer games from video sequences," in *Proc. IEEE Int. Conf. Image Process.*, Chicago, IL, USA, Oct. 1998, pp. 898–902, doi: 10.1109/ICIP.1998.723665.

[187] Y. Ohno, J. Miura, and Y. Shirai, "Tracking players and estimation of the 3D position of a ball in soccer games," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, vol. 1. Barcelona, Spain, Sep. 2000, pp. 145–148, doi: 10.1109/ICPR.2000.905293.

[188] P. Prandoni, E. Reusens, M. Vetterli, L. Sbaiz, and S. Ayer, "Automated stroboscoping of video sequence," U.S. Patent 2004 0 017 504 A1, Jan. 29, 2004.

[189] O. Grau, M. Price, and G. A. Thomas, "Use of 3-D techniques for virtual production," BBC R&D, Greater Manchester, U.K., White Paper 033, 2002. Accessed: May 27, 2020. [Online]. Available: http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP033.pdf

[190] N. Owens, C. Harris, and C. Stennett, "Hawk-eye tennis system," in *Proc. Int. Conf. Vis. Inf. Eng. (VIE)*, Guildford, U.K., Jul. 2003, pp. 182–185, doi: 10.1049/cp:20030517.

[191] P. McIlroy, "Hawk-eye: Augmented reality in sports broadcasting and officiating," in *Proc. 7th IEEE/ACM Int. Symp. Mixed Augmented Reality*, Cambridge, MA, USA, Sep. 2008, doi: 10.1109/ismar.2008.4637309.

[192] E. Reusens, V. Martin, A. Serge, and B. Victor, "Coordination and combination of video sequences with spatial and temporal normalization," U.S. Patent 8 675 021 B2, Dec. 31, 2009.

[193] J. Ren, M. Xu, J. Orwell, and G. A. Jones, "Multi-camera video surveillance for real-time analysis and reconstruction of soccer games," *Mach. Vis. Appl.*, vol. 21, no. 6, pp. 855–863, Oct. 2010, doi: 10.1007/s00138-009-0212-0.

[194] C. Hego. *Image Tracking System*. Accessed: Oct. 8, 2019. [Online]. Available: http://www.tracab.com/technology.asp

[195] X. Yu, C. Xu, H. W. Leong, Q. Tian, Q. Tang, and K. W. Wan, "Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video," in *Proc. 11th ACM Int. Conf. Multimedia (MULTIMEDIA)*, 2003, pp. 11–20.

[196] T. D'Orazio, C. Guaragnella, M. Leo, and A. Distante, "A new algorithm for ball recognition using circle Hough transform and neural classifier," *Pattern Recognit.*, vol. 37, no. 3, pp. 393–408, 2004, doi: 10.1016/S0031-3203(03)00228-0.

[197] J. Martinez, J.-C. Nebel, D. Makris, and C. Orrite, "Tracking human body parts using particle filters constrained by human biomechanics," in *Proc. Brit. Mach. Vis. Conf.*, 2008, p. 31, doi: 10.5244/C.22.31.

[198] R. J. Frayne, R. B. Dean, and T. R. Jenkyn, "Improving ice hockey slap shot analysis using three-dimensional optical motion capture: A pilot study determining the effects of a novel grip tape on slap shot performance," *Proc. Inst. Mech. Eng., P, J. Sports Eng. Technol.*, vol. 229, no. 2, pp. 136–144, Jun. 2015, doi: 10.1177/1754337114562096.

[199] J. Zedalis. 2013. *Hamilton Technology Company Changes the Way College and Professional Sports are Broadcast*. Times Trenton. Accessed: Oct. 11, 2019. [Online]. Available: http://www.nj.Degenerate cases and closed-form solutions for camera calibration with one-dimensional objectscom/mercer/index.ssf /2013/09/hamilton_technology_company_changes_the_way_sports_are_broadcast.html

[200] G. Thomas, *Visual Analysis of Humans*. London, U.K.: Springer, 2011, pp. 563–579, doi: 10.1007/978-0-85729-997-0.

[201] G. P. Stein, "Tracking from multiple view points: Self-calibration of space and time," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Fort Collins, CO, USA, Jun. 1999, pp. 521–527, doi: 10.1109/CVPR.1999.786987.

[202] A. Del Bimbo, F. Dini, F. Pernici, and A. Grifoni, *Pan-Tilt-Zoom Camera Networks*. Amsterdam, The Netherlands: Elsevier, 2009.

[203] S. Khan, O. Javed, Z. Rasheed, and M. Shah, "Human tracking in multiple cameras," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Vancouver, BC, Canada, Jul. 2001, pp. 331–336, doi: 10.1109/ICCV.2001.937537.

[204] Red Bee Media. (2004). *The PieroTM Sports Graphics System*. http://www/redbeemedia.com/piero/

[205] J. R. Anderson, C. F. Boyle, A. T. Corbett, and M. W. Lewis, "Cognitive modeling and intelligent tutoring," *Artif. Intell.*, vol. 42, no. 1, pp. 7–49, 1990, doi: 10.1016/0004-3702(90)90093-F.

[206] B. Holmberg, The Evolution, *Principles and Practices of of Distance Education*, vol. 11, no. 4. Oldenburg, Germany: BIS-Verlag der Carl von Ossietzky Universität Oldenburg Postfach Oldenburg, 2005.

[207] D. Clow, "MOOCs and the funnel of participation," in *Proc. 3rd Int. Conf. Learn. Anal. Knowl. (LAK)*, 2013, pp. 185–189, doi: 10.1145/2460296.2460332.

[208] K. K. Wollard, "Thinking, fast and slow," *Develop. Learn. Organizations: Int. J.*, vol. 26, no. 4, pp. 38–39, Jun. 2012, doi: 10.1108/14777281211249969.

[209] R. M. Delling, "Towards a theory of distance education," *ICDE Bull.*, vol. 13, pp. 21–25, Jan. 1987.

[210] Y. Rui, L. He, A. Gupta, and Q. Liu, "Building an intelligent camera management system," in *Proc. ACM Int. Multimedia Conf. Exhib.*, vol. 2001, pp. 2–11, doi: 10.1145/500144.500145.

[211] M. Bianchi, "Automatic video production of lectures using an intelligent and aware environment," in *Proc. 3rd Int. Conf. Mobile Ubiquitous Multimedia (MUM)*, vol. 83, 2004, pp. 117–123, doi: 10.1145/1052380.1052397.
[212] Y. Rui, A. Gupta, J. Grudin, and L. He, "Automating lecture capture and broadcast: Technology and videography," *Multimedia Syst.*, vol. 10, no. 1, pp. 3–15, Jun. 2004, doi: 10.1007/s00530-004-0132-9.
[213] X. Zhang, C.-W. Wu, P. Fournier-Viger, L.-D. Van, and Y.-C. Tseng, "Analyzing students' attention in class using wearable devices," in *Proc. IEEE 18th Int. Symp. A World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Macau, China, Jun. 2017, pp. 1–9, doi: 10.1109/WoWMoM.2017.7974306.
[214] C. Zhang, Y. Rui, J. Crawford, and L.-W. He, "An automated end-to-end lecture capture and broadcasting system," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 4, no. 1, pp. 1–23, Jan. 2008, doi: 10.1145/1324287.1324293.
[215] X. Li, Y. Wang, Y. Liu, and J. Xie, "Analyzing algorithm of multi-camera multi-touch system for educational application," in *Proc. 2nd Int. Conf. Edu. Technol. Training*, Sanya, China, Dec. 2009, pp. 90–94, doi: 10.1109/ETT.2009.64.
[216] T. Xu, T. Zhang, K. Kühnlenz, and M. Buss, "Attentional object detection with an active multi-focal vision system," *Int. J. Humanoid Robot.*, vol. 7, no. 2, pp. 223–243, Jun. 2010, doi: 10.1142/S0219843610002076.
[217] P. Napoletano and F. Tisato, "An attentive multi-camera system," *Proc. SPIE*, vol. 9024, May 2014, Art. no. 90240O, doi: 10.1117/12.2042652.
[218] P. Kalaivani and M. Annalakshmi, "A comprehensive framework for learning events," *Online J. Distance Educ. e-Learning*, vol. 5, no. 3, pp. 1–17, 2017.
[219] L. Mothwa, J.-R. Tapamo, and T. Mapati, "Conceptual model of the smart attendance monitoring system using computer vision," in *Proc. 14th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Las Palmas de Gran Canaria, Spain, Nov. 2018, pp. 229–234, doi: 10.1109/SITIS.2018.00042.
[220] N. Nandhakumar and J. K. Aggarwal, "Integrated analysis of thermal and visual images for scene interpretation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 469–481, Jul. 1988, doi: 10.1109/34.3911.
[221] M. H. Bianchi and L. Way. (2009). *10 Years of Television Presentations Without a Crew*. [Online]. Available: http://www.autoauditorium.com/PressRelease/AutoAuditorium_10years.pdf
[222] L. Erdmann and K. J. Gabriel, "High-resolution digital integral photography by use of a scanning microlens array," *Appl. Opt.*, vol. 40, no. 31, p. 5592, Nov. 2001, doi: 10.1364/ao.40.005592.
[223] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1166–1172, Sep. 2010, doi: 10.1109/TITB.2010.2051955.
[224] Z. Ye, Y. Li, A. Fathi, Y. Han, A. Rozga, G. D. Abowd, and J. M. Rehg, "Detecting eye contact using wearable eye-tracking glasses," in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, 2012, pp. 699–704, doi: 10.1145/2370216.2370368.
[225] W. R. Pruehsner and J. D. Enderle, "Infra-red radiant intensity exposure safety study for the eye tracker," Tech. Pap. ISA 455, May 2005, pp. 299–304,
[226] V. D. Nguyen, M. T. Le, A. D. Do, H. H. Duong, T. D. Thai, and D. H. Tran, "An efficient camera-based surveillance for fall detection of elderly people," in *Proc. 9th IEEE Conf. Ind. Electron. Appl.*, Hangzhou, China, Jun. 2014, pp. 994–997, doi: 10.1109/ICIEA.2014.6931308.
[227] S. Asteriadis, K. Karpouzis, and S. Kollias, "The importance of eye gaze and head pose to estimating levels of attention," in *Proc. 3rd Int. Conf. Games Virtual Worlds Serious Appl.*, May 2011, pp. 186–191, doi: 10.1109/VS-GAMES.2011.38.
[228] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, "An experimental comparison of online object-tracking algorithms," *Proc. SPIE*, vol. 8138, Sep. 2011, Art. no. 81381A, doi: 10.1117/12.895965.
[229] X. Mei and H. Ling, "Robust visual tracking using $\ell 1$ minimization," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep./Oct. 2009, pp. 1436–1443, doi: 10.1109/ICCV.2009.5459292.
[230] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 2756–2759, doi: 10.1109/ICPR.2010.675.
[231] J. Kwon and K. M. Lee, "Visual tracking decomposition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 1269–1276, doi: 10.1109/CVPR.2010.5539821.
[232] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami Beach, FL, USA, Jun. 2009, pp. 983–990, doi: 10.1109/CVPR.2009.5206737.
[233] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 125–141, May 2008, doi: 10.1007/s11263-007-0075-7.
[234] S. Stalder, H. Grabner, and L. V. Gool, "Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops, ICCV Workshops*, Kyoto, Japan, Sep. 2009, pp. 1409–1416, doi: 10.1109/ICCVW.2009.5457445.
[235] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," *Lecture Notes Computer Science (including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 5302. Berlin, Germany: Springer, 2008, pp. 234–247, doi: 10.1007/978-3-540-88682-2-19.
[236] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1631–1643, Oct. 2005, doi: 10.1109/TPAMI.2005.205.
[237] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. New York, NY, USA, Jun. 2006, pp. 798–805, doi: 10.1109/CVPR.2006.256.
[238] H. Grabner and H. Bischof, "On-line boosting and vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. New York, NY, USA, Jun. 2006, pp. 260–267, doi: 10.1109/CVPR.2006.215.
[239] O. Kainz, D. Cymbalak, J. Lamer, and F. Jakab, "Visual system for student attendance monitoring with non-standard situation detection," in *Proc. IEEE 12th IEEE Int. Conf. Emerg. eLearning Technol. Appl. (ICETA)*, Stary Smokovec, Slovakia, Dec. 2014, pp. 221–226, doi: 10.1109/ICETA.2014.7107589.
[240] G. G N, S. Naika C. L., and P. K. Das, "Face recognition using MB-LBP and PCA: A comparative study," in *Proc. Int. Conf. Comput. Commun. Informat.*, Jan. 2014, pp. 1–6, doi: 10.1109/ICCCI.2014.6921773.
[241] E. Hidayat, N. A. Fajrian, A. K. Muda, C. Y. Huoy, and S. Ahmad, "A comparative study of feature extraction using PCA and LDA for face recognition," in *Proc. 7th Int. Conf. Inf. Assurance Secur. (IAS)*, Dec. 2011, pp. 354–359, doi: 10.1109/ISIAS.2011.6122779.
[242] J. P. Papa, A. X. Falcao, A. L. M. Levada, D. C. Correa, D. H. P. Salvadeo, and N. D. A. Mascarenhas, "Fast and accurate holistic face recognition using Optimum-Path Forest," in *Proc. 16th Int. Conf. Digit. Signal Process.*, Santorini-Hellas, Greece, Jul. 2009, pp. 1–6, doi: 10.1109/ICDSP.2009.5201217.
[243] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Kauai, HI, USA, Dec. 2001, p. 1, doi: 10.1109/CVPR.2001.990517.
[244] K. Adrian and B. Gary, *Learning OpenCV 3 Computer Vision in C++ with the OpenCV Library*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2017.
[245] M. Fuzail, H. M. Fahad, O. M. Muhammad, R. Binish, T. Awais, and T. Muhammad Waqas, "Face detection system for attendance of class' students," *Int. J. Multidisciplinary Sci. Eng.*, vol. 5, no. 4, pp. 6–10, 2014.
[246] A. A. Tamimi, O. N. A. AL-Allaf, and M. A. Alia, "Real-time group face-detection for an intelligent class-attendance system," *Int. J. Inf. Technol. Comput. Sci.*, vol. 7, no. 6, pp. 66–73, May 2015, doi: 10.5815/ijitcs.2015.06.09.
[247] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *Lecture Notes Computure Science (including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)*, vol. 3001. Berlin, Germany: Springer, 2004, pp. 1–17, doi: 10.1007/978-3-540-24646-6_1.
[248] E. R. Fossum, "CMOS image sensors: Electronic camera-on-a-chip," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1689–1698, Oct. 1997, doi: 10.1109/16.628824.
[249] T. D. Binnie, "Fast imaging microlenses," *Appl. Opt.*, vol. 33, no. 7, p. 1170, Mar. 1994, doi: 10.1364/ao.33.001170.

[250] N. Anirudth, B. H. P. Prasad, A. Jain, and V. Peddigari, "Robust photometric alignment for asymmetric camera system," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2018, pp. 1–4, doi: 10.1109/ICCE.2018.8326314.

[251] Y. Liu and B. Zhang, "Photometric alignment for surround view camera system," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 1827–1831, doi: 10.1109/ICIP.2014.7025366.

[252] E. J. Tremblay, R. A. Stack, R. L. Morrison, and J. E. Ford, "Ultrathin cameras using annular folded optics," *Appl. Opt.*, vol. 46, no. 4, pp. 463–471, 2007, doi: 10.1364/AO.46.000463.

[253] O. Burggraaff, "Standardized spectral and radiometric calibration of consumer cameras," *Opt. Express*, vol. 27, no. 14, pp. 19075–19101, 2019, doi: 10.1364/oe.27.019075.

[254] A. Saxena, S. Jamie, and A. Y. Ng, "Depth estimation using monocular and stereo cues," in *Proc. 20th Int. Joint Conf. Artif. Intell.*, 2007, pp. 2197–2203.

[255] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proc. IEEE Workshop Stereo Multi-Baseline Vis. (SMBV )*, VOL. 1, Kauai, HI, USA, Dec. 2001, pp. 131–140, doi: 10.1109/SMBV.2001.988771.

[256] F. David A. and P. Jean, *Computer Vision: A Modern Approach*, 2nd ed. London, U.K.: Pearson, 2012.

[257] O. Vogel, L. Valgaerts, M. Breuß, and J. Weickert, "Making shape from shading work for real-world images," in *Proc. Joint Pattern Recognit. Symp.*, vol. 5748, 2009, pp. 191–200, doi: 10.1007/978-3-642-03798-6_20.

[258] J. Michels, A. Saxena, and A. Y. Ng, "High speed obstacle avoidance using monocular vision and reinforcement learning," in *Proc. 22nd Int. Conf. Mach. Learn. (ICML)*, 2005, pp. 593–600, doi: 10.1145/1102351.1102426.

[259] E. Delage, H. Lee, and A. Y. Ng, "A dynamic Bayesian network model for autonomous 3D reconstruction from a single indoor image," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. New York, NY, USA, Jun. 2006, pp. 2418–2428, doi: 10.1109/CVPR.2006.23.

[260] Y. Guo, S. Zou, and H. Li, "Depth estimation from a single image in pedestrian candidate generation," in *Proc. IEEE 11th Conf. Ind. Electron. Appl. (ICIEA)*, Hefei, China, Jun. 2016, pp. 1005–1008, doi: 10.1109/ICIEA.2016.7603729.

[261] J. Woetzel and R. Koch, "Multi-camera real-time depth estimation with discontinuity handling on PC graphics hardware," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, Cambridge, MA, USA, Aug. 2004, pp. 741–744, doi: 10.1109/ICPR.2004.1334296.

[262] I. Feldmann, W. Waizenegger, N. Atzpadin, and O. Schreer, "Real-time depth estimation for immersive 3D videoconferencing," in *Proc. 3DTV-Conf.: True Vis.-Capture, Transmiss. Display 3D Video*, Tampere, Finland, Jun. 2010, pp. 1–4, doi: 10.1109/3DTV.2010.5506312.

[263] E.-K. Lee and Y.-S. Ho, "Generation of high-quality depth maps using hybrid camera system for 3-D video," *J. Vis. Commun. Image Represent.*, vol. 22, no. 1, pp. 73–84, Jan. 2011, doi: 10.1016/j.jvcir.2010.10.006.

[264] P. T. Kovacs and F. Zilly, "3D capturing using multi-camera rigs, real-time depth estimation and depth-based content creation for multi-view and light-field auto-stereoscopic displays," in *Proc. ACM SIGGRAPH Emerg. Technol. (SIGGRAPH)*, 2012, doi: 10.1145/2343456.2343457.

[265] N. Stefanoski, C. Bal, M. Lang, O. Wang, and A. Smolic, "Depth estimation and depth enhancement by diffusion of depth features," in *Proc. IEEE Int. Conf. Image Process.*, Melbourne, VIC, Australia, Sep. 2013, pp. 1247–1251, doi: 10.1109/ICIP.2013.6738257.

[266] G. Marin, G. Agresti, L. Minto, and P. Zanuttigh, "A multi-camera dataset for depth estimation in an indoor scenario," *Data Brief*, vol. 27, Dec. 2019, Art. no. 104619, doi: 10.1016/j.dib.2019.104619.

[267] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Washington, DC, USA, Jun./Jul. 2004, pp. 2–9, doi: 10.1109/cvpr.2004.1315206.

[268] J. Luo and G. Oubong, "A comparison of SIFT, PCA-SIFT and SURF," *Int. J. Image Process.*, vol. 3, no. 4, pp. 143–152, 2009, doi: 10.1080/00420980020014884.

[269] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.

[270] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Binary robust independent elementary features," *Lecture Notes Compute Science (including Subseries Lecture Notes Artifical Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 6314. Berlin, Germany: Springer, 2010, pp. 778–792, doi: 10.1007/978-3-642-15561-1_56.

[271] F. Zhang, F. Ye, and Z. Su, "A modified feature point descriptor based on binary robust independent elementary features," in *Proc. 7th Int. Congr. Image Signal Process.*, Dalian, China, Oct. 2014, pp. 258–263, doi: 10.1109/CISP.2014.7003788.

[272] F. Tang, S. H. Lim, N. L. Chang, and P. Alto, "A novel feature descriptor invariant to complex brightness changes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 2631–2638, doi: 10.1109/CVPR.2009.5206550.

[273] S. Leutenegger, M. Chli, and Y. S. Roland, "Research collection," *Bin. Robust Invariant Scalable Keypoints*, vol. 15, no. 3, pp. 12–19, 2011, doi: 10.3929/ethz-a-010782581.

[274] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "Color features for tracking non-rigid objects," *Zidonghua Xuebao/Acta Automatica Sinica*, vol. 29, no. 3, pp. 345–355, 2003.

[275] T. Mathes and J. H. Piater, *Robust Non-rigid Object Tracking Using Point Distribution Manifolds* (Lecture Notes in Computer Science), vol. 3465, 28th ed. Berlin, Germany: Springer, 2006, pp. 533–790.

[276] A. W. Senior, G. Potamianos, S. Chu, Z. Zhang, A. Hampapur, and Y. Heights, "A comparison of multicamera person-tracking algorithms," in *Proc. Vis. Surveill.*, 2006, pp. 1–7. [Online]. Available: http://andrewsenior.com/papers/SeniorVS06.pdf

[277] K. Nickel, T. Gehrig, R. Stiefelhagen, and J. McDonough, "A joint particle filter for audio-visual speaker tracking," in *Proc. 7th Int. Conf. Multimodal Interfaces (ICMI)*, 2005, pp. 61–68, doi: 10.1145/1088463.1088477.

[278] A. D. Straw, K. Branson, T. R. Neumann, and M. H. Dickinson, "Multi-camera real-time three-dimensional tracking of multiple flying animals," *J. Roy. Soc. Interface*, vol. 8, no. 56, pp. 395–409, Mar. 2011, doi: 10.1098/rsif.2010.0230.

[279] S. L. Dockstader and A. M. Tekalp, "Multiple camera fusion for multi-object tracking," in *Proc. IEEE Workshop Multi-Object Tracking*, Jul. 2001, pp. 95–102, doi: 10.1109/MOT.2001.937987.

[280] A. Mittal and L. S. Davis, "$M_2$Tracker: A multi-view approach to segmenting and tracking people in a cluttered scene," *Int. J. Comput. Vis.*, vol. 51, no. 3, pp. 189–203, 2003, doi: 10.1023/A:1021849801764.

[281] H. Possegger, S. Sternig, T. Mauthner, P. M. Roth, and H. Bischof, "Robust real-time tracking of multiple objects by volumetric mass densities," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 2395–2402, doi: 10.1109/CVPR.2013.310.

[282] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using K-Shortest paths optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1806–1819, Sep. 2011, doi: 10.1109/TPAMI.2011.21.

[283] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 267–282, Feb. 2008, doi: 10.1109/TPAMI.2007.1174.

[284] D. Arsic, E. Hristov, N. Lehment, B. Hornler, B. Schuller, and G. Rigoll, "Applying multi layer homography for multi camera person tracking," in *Proc. 2nd ACM/IEEE Int. Conf. Distrib. Smart Cameras*, Stanford, CA, USA, Sep. 2008, pp. 1–9, doi: 10.1109/ICDSC.2008.4635731.

[285] S. Calderara, R. Cucchiara, and A. Prati, "Bayesian-competitive consistent labeling for people surveillance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 354–360, Feb. 2008, doi: 10.1109/TPAMI.2007.70814.

[286] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, and S. Maybank, "Principal axis-based correspondence between multiple cameras for people tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 663–671, Apr. 2006, doi: 10.1109/TPAMI.2006.80.

[287] K. Otsuka and N. Mukawa, "Multiview occlusion analysis for tracking densely populated objects based on 2-D visual angles," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Washington, DC, USA, Jun./Jul. 2004, p. 1, doi: 10.1109/cvpr.2004.1315018.

[288] J. Kang, I. Cohen, and G. Medioni, "Tracking people in crowded scenes across multiple cameras," in *Proc. Asian Conf. Comput. Vis.*, Jan. 2004, pp. 1–6.
[289] M. Song, D. Tao, and S. J. Maybank, "Sparse camera network for visual surveillance—A comprehensive survey," 2013, *arXiv:1302.0446*. [Online]. Available: https://arxiv.org/abs/1302.0446

**ADESHINA SIRAJDIN OLAGOKE** (Student Member, IEEE) was born in March 1981. He received the B.Eng. degree in electrical and electronic engineering from the Federal University of Technology Yola (FUTY), Adamawa, in 2005, and the master's degree in signal processing from the University of Maiduguri, in 2015. He is currently pursuing the Ph.D. degree in image processing and with the School of Electrical and Electronic Engineering, Universiti Sains Malaysia.

He is also a Graduate Research Assistant with the School of Electrical and Electronic Engineering, Universiti Sains Malaysia. He is also a native of Offa, in Kwara, Nigeria. His current research interest is in the field of face recognition and detection. He has experience in computer networks and system maintenance. From 2007 to 2010, he worked as a Network Engineer and a System Administrator in Yaysib Wireless Networks and Computers. He is also a certified Cisco Associate and CompTIA. He is a Lecturer with the Department of Computer Engineering, Federal Polytechnic Mubi, Adamawa, Nigeria. He already published some articles. He is a member of the Nigerian Society of Engineering, and a registered Engineer with the Council for Regulation of Engineering and Engineering Practice in Nigeria.

**HAIDI IBRAHIM** (Senior Member, IEEE) received the B.Eng. degree in electrical and electronic engineering from Universiti Sains Malaysia, Malaysia, and the Ph.D. degree in image processing from the Centre for Vision, Speech, and Signal Processing (CVSSP), University of Surrey, U.K., in 2005. His research interest includes digital image and signal processing and analysis.

**SOO SIANG TEOH** (Senior Member, IEEE) received the B.E. degree in electronic engineering from University Putra Malaysia, in 1993, the M.S. degree in digital electronics from the University of Manchester, U.K., in 1995, and the Ph.D. degree in computer engineering from the University of Western Australia, in 2012.

From 1995 to 2008, he has worked as an Electronic Engineer in several multinational companies in Malaysia and Finland. Since 2012, he has been a Senior Lecturer with the School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Penang. He has authored or coauthored more than 15 journal and conference articles. His research interests include image processing and machine learning for industrial and biomedical applications.

Dr. Teoh is a member of the Institution of Engineers Malaysia. He was a recipient of the Best Paper Award in the IEEE Conference on System, Process, and Control, in 2016, and the IEEE International Conference on Control System, Computing, and Engineering, in 2015.