Person Reidentification and Recognition in Video

Rangachar Kasturi and Rajmadhan Ekambaram

Computer Science and Engineering,
University of South Florida, Tampa, Florida, USA
R1K@cse.usf.edu, rajmadhan@mail.usf.edu
http://figment.csee.usf.edu/

Abstract. Person recognition has been a challenging research problem for computer vision researchers for many years. A variation of this generic problem is that of identifying the reappearance of the same person in different segments to tag people in a family video. Often we are asked to answer seemingly simple queries such as 'how many different people are in this video? or 'find all instances of this person in these videos'. The complexity of the task grows quickly if the video in question includes segments taken at different times, places, lighting conditions, camera settings and distances since these could include substantial variations in resolution, pose, appearance, illumination, background, occlusions, etc. In some scenarios (airports, shopping centers, and city streets) we may have video feeds from multiple cameras with partially overlapping views operating under widely varying lighting and visibility conditions. Yet computer vision systems are challenged to find and track a person of interest as data from such systems have become ubiquitous and concern for security in public spaces has become a growing concern. While this is yet an unsolved challenge, much progress has been made in recent years in developing computer vision algorithms which are the building blocks for person detection, tracking and recognition. We consider several video capture scenarios, discuss the challenges they present for person re-identification and recognition as the complexity of the scene changes, and present pointers to recent research work in relevant computer vision areas in this paper.

Keywords: Video Surveillance, Multi-camera Tracking, Person Re-Identification.

1 Introduction

There has been much interest in developing computer vision algorithms for person detection and tracking in video since it has many interesting applications such as autonomous navigation [1], activity recognition [2], human-computer interactions [3], monitoring activities of elderly people [4], sports video analysis [5], surveillance [6], etc. In this paper we focus our attention on the problem of counting the number of people in a video clip or a collection of videos with

E. Bayro-Corrochano and E. Hancock (Eds.): CIARP 2014, LNCS 8827, pp. 280–293, 2014.

the requirement that multiple occurrences of the same person should only be counted once; i.e., we address the problem of answering the question "how many different people are in this video?" This makes it a much more challenging problem, since we now have to compare different instances of detected people to determine if they belong to the same person. Such a system is needed for example to tag people in family videos [7], track people across multiple camera views as they pass through different parts of a building or an outdoor complex [8], or for determining the activities of a suspect over an extended period of time in a surveillance application. While there are many solutions for person counting, counting number of different people in a video in an unconstrained environment is an unsolved problem. For example, we can count the number of people in a short video of a single event by simply detecting and tracking the people in the video. On the other hand, accurately counting the number of different people in a video compilation of events spanning many years would require biometric recognition to account for potential variations in appearance as well as aging.

A class of methods for surveillance and forensics applications have been developed in the recent past in which a persons identity between two tracks is reestablished through appearance continuity. Such methods are known as Re-identification and operate on low resolution data in the form of bounding boxes of people extracted from videos which are not adequate for biometric recognition. Vezzani, Baltieri, and Cucchiara et al. [9] have provided an excellent comprehensive survey of research advances in Re-identification over the past 15 years. Appearance variation of a person is induced by changes in any or a combination of the attributes such as clothing, hairstyle, eyeglass, hat, mask, etc. Within a short period of time (for example, a single day) the appearance of the person can be assumed to remain constant [10]. In [11] the current appearance based re-identification algorithms are classified as short-period re-identification methods. Obviously if a person changes any of the above mentioned attributes within the period of observation, the system is likely to assign her/him a different identity. We will need more robust methodologies based on biometrics to handle such situations.

In this paper we consider re-identification as well as other approaches for counting the number of different people. We briefly describe a number of scenarios in the next section to better appreciate the challenges presented by this problem. We then present a short overview of recent progress in relevant areas of computer vision that are the building blocks of such a system.

2 Scene Descriptions

In this section the characteristics of several typical video imaging scenes primarily from the surveillance domain are described. The entry-exit points in the scene are the locations through which people can enter or exit the camera view. Our goal is to count the number of different people in such a video feed within a specified time period.

Fig. 1. A typical indoor walkway [12]

Stationary Camera, Indoor Walkway: Consider a simple scenario in which a single, stationary, fixed focus camera is installed at one end of a walkway to capture the flow of people as shown in Figure 1 [12]. The primary entry-exit points are at the two ends of the walkway. In such a scenario people typically walk towards or away from the camera resulting in a steady flow of traffic. Typical processing steps would include person detection, person tracking, face detection as people walk towards the camera, and extraction of sufficient discriminating features from each tracked person, and continuously comparing the features to make sure a unique label is assigned to each person even if (s)he appears multiple times within the specified time period. Note that there are a number of challenges. For example, there are potential occlusions due to people crossing one another or stopping to have a conversation; we need a tracker that can maintain proper track identities through such events. Facial biometrics are only available for people walking towards the camera; for those walking away from the camera we need to depend upon other appearance-based features such as clothing, objects being carried such as backpack etc. Other biometric modalities such as gait and ear recognition have limited use in this configuration. While we can assume that a persons external appearance (clothing, backpacks, headwear etc.) are not likely to change if the interval of observation is short, we cannot depend on these characteristics if the observation period is longer; we then must depend on more reliable features such as those that can be extracted from a close-up facial image. Obviously this would require two cameras with one at each end of the walkway. Position and orientation of cameras can be controlled to capture full front view images for using re-identification based approaches. Similarly, high quality images can be captured at least when the person is closer to camera to use biometric recognition. In the example of Figure 1 there is some ambient light on the walkway from the outdoors on the right side of the image; it is possible to overcome the effects such illumination by installing additional lights as needed.

Stationary Multiple Cameras, Indoor Lobby: In this scenario multiple cameras are overlooking an indoor lobby. A lobby may have multiple entries and exits as well as space for people to pursue various activities resulting in uneven flow of traffic creating challenges for detection and tracking. For example, the observed 2D shape in the image may deviate significantly (e.g., when a person sits) from typical models based on a walking person used by many person detection and tracking algorithms. People may temporarily leave an object (backpack, rolling cart) creating multiple object instances; appearance based models would have difficulty in maintaining the same identity labels for people through such transitions. Often the cameras are mounted near the ceiling resulting in oblique view of people on the floor unlike cameras looking down the hallway. An example of a large indoor lobby is shown in Figures 2. Clearly the size of the image region covering each person is small because of the large area of the camera view and does not have sufficient resolution to recognize human faces. One possibility in such scenarios is to capture higher resolution images at the entry/exit points and use tracking to associate the identity as the person transitions through the lobby. Such an arrangement may not be feasible for a variety of reasons; for example, due to concerns for privacy. However this is an option in high security installations such as airports.

Fig. 2. Image of a large indoor lobby [13]

In such large indoor spaces it is possible to place cameras with partially overlapping fields of view with the potential for calculating the 3D position of an observed person at any given instance through appropriate calibration. By properly installing a number of cameras at appropriate locations it may be possible to have each point on the floor being observed by two or more cameras at all times; in such a case it is possible to create a single 3D track of each person over time even if the person moves from the field of view of one camera to another. This would altogether eliminate the need for any re-identification. Figure 3 shows views from four different cameras of an indoor lobby. A large part of the lobby is visible in all four views.

Fig. 3. Multiple views of an indoor lobby [14]

Stationary Multiple Cameras, Outdoor Scene: Unlike the indoor scene described in the previous section where we have much control over the placement of cameras, illumination, and the nature of people movement and their activities, an outdoor scene such as a market, town square, or busy street intersection represent an unconstrained environment. We can expect wide variations in illumination (day, night, overcast, rain/snow, and shadows), numerous obstacles (buildings, trees, and vehicles), large number of people engaged in a variety of activities, and many types of vehicular traffic. We also have limited access to structures (such as traffic light poles, street lights) for placing the cameras although occasionally we may have access to video feeds from surveillance cameras from neighboring buildings (especially during law enforcement and investigation situations). Large crowds make it difficult to track individuals. It is practically impossible to require all cameras to be synchronized and calibrated to create a 4D space-time model. Clearly such an environment imposes severe challenges for

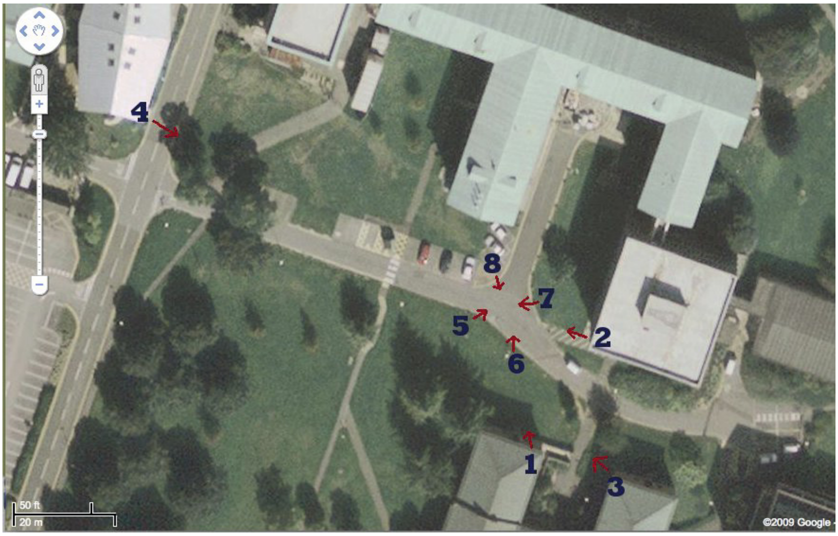

Fig. 4. PETS 2010 Dataset [15]. Layout of the site and the placement of cameras (marked 1 to 8) and their direction of observation

Fig. 5. PETS 2010 Dataset [15] Images captured from the cameras shown in figure 4

vision algorithms designed for detecting, tracking and recognizing people. Thus, to keep the complexity of the outdoor scene to practical limits, PETS (Performance Evaluation of Tracking and Surveillance) workshops [15], have been using an outdoor space shown in Figure 4 for creating the challenge datasets.

PETS dataset [15] consists of videos taken from 8 cameras with overlapping views. The dataset also contains calibration information for all the cameras. The location and the tasks were chosen to simulate surveillance in a real world environment. The dataset contains sequences of increasing complexity for several scenarios such as person counting, density estimation, people tracking, flow analysis and event recognition.

Moving Cameras: In the three scenarios we described so far, we assumed that the cameras are stationary. While this is typical in surveillance applications, large fraction of video produced today are captured using moving cameras. Much of family videos, news stories, TV shows and movies are all from non-stationary cameras. Consider creating an app for tagging people in your family video collection. This requires the app to detect, track, and recognize people; after an initial period of interaction with you to acquire names for detected people, the app should automatically tag all reoccurrences of these people. Re-identification methods can be deployed to track and tag all individuals at a single event based on their appearance. However, if your collection includes multiple events with people in different clothes and make up, some indoors and some outdoors, taken with cameras of varying quality, capturing a variety of expressions, and spanning extended periods of time potentially going back to your childhood, we need a highly accurate biometric recognition system that works across all these dimensions. It is not surprising that there has not been a video tagging app yet that meets all these expectations!

3 Recent Progress in Relevant Areas of Computer Vision

The scenarios described in the previous section presents many challenges to computer vision algorithms due to the large number of parameters such as the viewpoint, illumination, scale, object deformation, occlusion, background clutter, motion, camera characteristics etc. In this section we briefly describe some of the key operations that form the building blocks for *counting the number of people, excluding reappearances, in video.*

Person Detection and Tracking: Full body detection and tracking methods are proposed in [16] and [17] respectively. Availability of full body images depend on camera view and crowd density. The method in [18] addresses person detection in crowded scenes under occlusion conditions. A part based detector method is proposed in [19] to address tracking under occlusion scenarios. The method proposed in [20] addresses tracking in complex situations such as people moving in small groups, have inter-occlusions, cast shadows on the ground, and reflections. When multiple calibrated cameras are available people detection and tracking can be directly done in 3D co-ordinate space or using 3D models for tracking using methods proposed in [21] and [22]. Tracking in uncalibrated multiple cameras with overlapping views is addressed in [8]. Tracking and recognition in multiple camera scenarios is addressed in [23] by exploiting the real world constraints such as the presence of a world plane, the presence of a three-dimensional scene model, consistency of motion across cameras, and the color and texture properties.

Biometric Recognition: Biometric Recognition is the science of uniquely identifying a person by analyzing various biometric traits such as face, ear, voice, gait, etc. Introduction to Biometrics [24] is a good introductory text book in this field. Biometric modalities that are available in a video are face, ear, gait and voice. Face recognition requires good quality images captured when the subjects are close to the camera. Several potential methods are proposed in [25–27] for face recognition in videos. A method described in [28] addresses face recognition in low resolution images. Face recognition in unconstrained scenarios are more complex than the constrained scenarios that can be found in popular databases such as FERET [29] and Extended Yale B [30]. Some promising results for face recognition under unconstrained scenarios have been recently described in [31–33] using Labelled Faces in the Wild (LFW) dataset [34] and YouTube face dataset [35]. Gait recognition is a potential method for recognition at a distance. Simple gait recognition method based on averaged silhouette is proposed in [36]. Frontal view gait recognition method for surveillance video scenarios is proposed in [37]. Gait recognition method for occlusion scenarios is proposed and tested using synthesized occlusions in [38]. A review of recent gait recognition approaches can be found in [39]. Ear Recognition has been shown to be a useful biometric modality in certain applications [40]. A few methods [41, 42] were recently proposed for 3D ear recognition. A recent survey on ear recognition algorithms can be found in [43]. Voice [44], facial marks [45], tattoos [46], etc. have also been shown to be useful for person identification.

Camera Calibration and 3D Scene Reconstruction: In multi-camera systems with overlapping views, person detection and tracking can be performed directly in 3D co-ordinate space after calibration and 3D reconstruction as suggested in [21] and [22]. An excellent review of the existing methods in camera calibration and topology estimation is provided in the recent survey paper [47] on intelligent multi-camera video surveillance. Also the recent book [48] on multi-camera networks has a separate section on multi-camera calibration and topology estimation. The problem of 3D reconstruction from large collection of internet images is explored in [49]. Approaches for urban 3D reconstruction from video together with GPS and inertial measurements are proposed in [50, 51]. An interesting comparison of depth estimation between LIDAR measurements and from applying 3D modelling techniques on high resolution images is done in [52].

Re-identification: There has been much work on re-identification in the past few years. Researchers have addressed the problems using different approaches such as brightness transfer functions [53–56], attribute based matching [57–59], saliency matching [60, 61], discriminative classification [62–64], and transfer learning [65–67]. The popular datasets [68–70] for this research are a collection of image regions of people cropped from the video data. Some example are shown in Figure 6. As noted earlier a comprehensive survey has been recently published [9].

Fig. 6. Sample images from popular datasets used for re-identification problem. A pair of images for each person is shown. Images in the first two columns are from VIPeR dataset [68], images in the middle two columns are from iLIDS dataset [69] and the images in the last two columns are from CAVIAR4REID dataset [70].

Commercial People Counters: There are many commercial products for counting people. However, these products do not discriminate multiple appearances of the same person; i.e., a person is counted again if (s)he reappears during the counting period. With no re-identification or recognition, the technical requirements are simpler. These products use many imaging technologies such as thermal, laser, infrared and stereoscopic in addition to standard cameras. A brief description of the technologies and applications of people counters is at [71].

4 List of Datasets, Special Issues, and Focused Workshops

A comprehensive list of datasets for person re-identification and multi-camera tracking is included in [9]. Most popular datasets for person re-identification research are image based such as those shown in Figure 6 effectively reducing the re-identification problem to comparison between images. Recently, a multi-camera object tracking challenge has been announced in conjunction with an upcoming workshop [72] and a dataset of non-overlapping videos has been released. Another video dataset with non-overlapping views for people detection, tracking, action analysis, and trajectory analysis is the 3DPeS dataset [73]. Camera calibration parameters and identity of detected people are also included in the 3DPeS dataset. A book [10] inspired by the First International Workshop on Reidentification [74] has been published.

5 Conclusions

Person re-identification and recognition have been active computer vision research areas. The popular datasets [68–70] for re-identification research use a small number of cameras with similar sensor characteristics. While there has been much progress in the development of computer vision algorithms in the past ten years as demonstrated on such datasets, in real-world person re-identification challenges such as those faced by Boston bombing investigators [75], the data comes from disparate sources with widely varying sensor characteristics; automatically analyzing and finding the persons of interest in such a large collection is way beyond the capabilities of current computer vision systems. We anticipate an increased interest in this area of research and we hope that more realistic datasets become available to researchers to solve this important problem.

References

1. Massimo, B., Broggi, A., Cellario, M., Fascioli, A., Lombardi, P., Porta, M.: Artificial vision in road vehicles. Proceedings of the IEEE 90(7), 1258–1271 (2002)
2. Pavan, T., Chellappa, R., Subrahmanian, V.S., Udrea, O.: Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video Technology 18(11), 1473–1488 (2008)
3. Fakhreddine, K., Alemzadeh, M., Saleh, J.A., Arab, M.N.: Human-computer interaction: Overview on state of the art (2008)
4. Hammadi, N.-C., McKenna, S.J.: Activity summarisation and fall detection in a supportive home environment. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 4, pp. 323–326. IEEE (2004)
5. Atrey, P.K., Hossain, M.A., El Saddik, A., Kankanhalli, M.S.: Multimodal fusion for multimedia analysis: a survey. Multimedia Systems 16(6), 345–379 (2010)
6. Weiming, H., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 34(3), 334–352 (2004)
7. Abowd, G.D., Gauger, M., Lachenmann, A.: The Family Video Archive: an annotation and browsing environment for home movies. In: Abowd, G.D. (ed.) Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval, pp. 1–8. ACM (2003)
8. Sohaib, K., Shah, M.: Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(10), 1355–1360 (2003)

9. Roberto, V., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics: A survey. ACM Computing Surveys (CSUR) 46(2), 29 (2013)
10. Gong, S., Cristani, M., Yan, S., Loy, C.C.: Person Re-Identification. Springer (2014)
11. Apurva, B.-G., Shah, S.K.: A survey of approaches and trends in person re-identification. Image and Vision Computing 32(4), 270–286 (2014)
12. EC Funded CAVIAR project/IST 2001 37540, `http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`
13. `https://www.flickr.com/photos/marshallstudentcenter/3092267497/`
14. Liem, M.C., Gavrila, D.M.: A comparative study on multi-person tracking using overlapping cameras. In: Chen, M., Leibe, B., Neumann, B. (eds.) ICVS 2013. LNCS, vol. 7963, pp. 203–212. Springer, Heidelberg (2013)
15. Ferryman, J., Ellis, A.: PETS2010: Dataset and challenge. In: Ferryman, J. (ed.) 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 143–150. IEEE (2010)
16. Navneet, D., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
17. Mykhaylo, A., Roth, S., Schiele, B.: People-tracking-by-detection and people-detection-by-tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
18. Bastian, L., Seemann, E., Schiele, B.: Pedestrian detection in crowded scenes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 878–885. IEEE (2005)
19. Bo, W., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. International Journal of Computer Vision 75(2), 247–266 (2007)
20. Tao, Z., Nevatia, R.: Tracking multiple humans in complex situations. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9), 1208–1221 (2004)
21. Ivana, M., Santini, S., Jain, R.: Video processing and integration from multiple cameras. In: Proceedings of the 1998 Image Understanding Workshop, vol. 6. Morgan-Kaufman, San Francisco (1998)
22. Jian, Y., Odobez, J.-M.: Multi-Person Bayesian Tracking with Multiple Cameras. Multi-Camera Networks: Principles and Applications 363 (2009)
23. Sankaranarayanan, A.C., Veeraraghavan, A., Chellappa, R.: Object detection, tracking and recognition for multiple smart cameras. Proceedings of the IEEE 96(10), 1606–1624 (2008)
24. Jain, A.K., Ross, A.A., Nandakumar, K.: Introduction to biometrics. Springer (2011)
25. Kuang-Chih, L., Ho, J., Yang, M.-H., Kriegman, D.: Video-based face recognition using probabilistic appearance manifolds. In: Proceedings 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I–313. IEEE (2003)
26. Unsang, P., Jain, A.K., Ross, A.: Face recognition in video: Adaptive fusion of multiple matchers. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8. IEEE (2007)
27. Xiaoming, L., Chen, T.: Video-based face recognition using adaptive hidden markov models. In: Proceedings. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. I–313. IEEE (2003)
28. Gunturk, B.K., Batur, A.U., Altunbasak, Y., Hayes, M.H., Mersereau, R.M.: Eigenface-domain super-resolution for face recognition. IEEE Transactions on Image Processing 12(5), 597–606 (2003)

29. Jonathon, P.P., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
30. Lee, K.-C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(5), 684–698 (2005)
31. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: Computer Vision and Pattern Recognition, CVPR (2014)
32. Hu, J., Lu, J., Tan, Y.-P.: Discriminative Deep Metric Learning for Face Verification in the Wild. In: Computer Vision and Pattern Recognition, CVPR (2014)
33. Sun, Y., Wang, X., Tang, X.: Deep Learning Face Representation from Predicting 10,000 Classes. Computer Vision and Pattern Recognition, CVPR (2014)
34. Huang, G.B., Ramesh, M., Berg, T., Learned-miller, E.: Labeled faces in the wild: A database for studying face recognition in un-constrained environments. In: ECCV Workshop on Faces in Real-life Images (2008)
35. Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matched background similarity. In: CVPR, pp. 529–534 (2011)
36. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: Averaged silhouette. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 4, pp. 211–214. IEEE (2004)
37. Olivier, B., Van Droogenbroeck, M.: Frontal-view gait recognition by intra-and inter-frame rectangle size distribution. Pattern Recognition Letters 30(10), 893–901 (2009)
38. Chen, C., Liang, J., Zhao, H., Hu, H., Tian, J.: Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters 30(11), 977–984 (2009)
39. Wang, J., She, M., Nahavandi, S., Kouzani, A.: A review of vision-based gait recognition methods for human identification. In: 2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 320–327. IEEE (2010)
40. Chang, K., Bowyer, K.W., Sarkar, S., Victor, B.: Comparison and combination of ear and face images in appearance-based biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1160–1165 (2003)
41. Chen, H., Bhanu, B.: Human ear recognition in 3D. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4), 718–737 (2007)
42. Yan, P., Bowyer, K.W.: Biometric recognition using 3D ear shape. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(8), 1297–1308 (2007)
43. Abaza, A., Ross, A., Hebert, C., Harrison, M.A.F., Nixon, M.S.: A survey on ear biometrics. ACM Computing Surveys (CSUR) 45(2), 22 (2013)
44. Kinnunen, T., Li, H.: An overview of text-independent speaker recognition: from features to supervectors. Speech Communication 52(1), 12–40 (2010)
45. Jain, A.K., Park, U.: Facial marks: Soft biometric for face recognition. In: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 37–40. IEEE (2009)
46. Lee, J.-E., Jain, A.K., Jin, R.: Scars, marks and tattoos (SMT): Soft biometric for suspect and victim identification. In: Biometrics Symposium, BSYM 2008, pp. 1–8. IEEE (2008)
47. Wang, X.: Intelligent multi-camera video surveillance: A review. Pattern Recognition Letters 34(1), 3–19 (2013)

48. Aghajan, H., Cavallaro, A. (eds.): Multi-Camera Networks: Concepts and Applications. Elsevier (2009)
49. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from internet photo collections. International Journal of Computer Vision 80(2), 189–210 (2008)
50. Pollefeys, M., Nistér, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., et al.: Detailed real-time urban 3d reconstruction from video. International Journal of Computer Vision 78(2-3), 143–167 (2008)
51. Akbarzadeh, A., Frahm, J.-M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., et al.: Towards urban 3D reconstruction from video. In: Akbarzadeh, A.J.-M. (ed.) Third International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 1–8. IEEE (2006)
52. Strecha, C., von Hansen, H., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
53. Porikli, F.: Inter-camera color calibration using cross-correlation model function. In: ICIP (2003)
54. Javed, O., Shafique, K., Shah, M.: Appearance modeling for tracking in multiple non-overlapping cameras. In: CVPR (2005)
55. Prosser, B., Gong, S., Xiang, T.: Multi-camera Matching Using Bi-Directional Cumulative Brightness Transfer Functions. In: BMVC (2008)
56. Chen, K.-W., Lai, C.-C., Hung, Y.-P., Chen, C.-S.: An adaptive learning method for target tracking across multiple cameras. In: CVPR (2008)
57. Vaquero, D.A., Feris, R.S., Tran, D., Brown, L., Hampapur, A., Turk, M.: Attribute-based people search in surveillance environments. In: Vaquero, D.A. (ed.) 2009 Workshop on Applications of Computer Vision (WACV), pp. 1–8. IEEE (2009)
58. Layne, R., Hospedales, T.M., Gong, S.: Towards person identification and re-identification with attributes. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 402–412. Springer, Heidelberg (2012)
59. Liu, X., Song, M., Zhao, Q., Tao, D., Chen, C., Bu, J.: Attribute-restricted latent topic model for person re-identification. Pattern Recognition 45(12), 4204–4213 (2012)
60. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3586–3593. IEEE (2013)
61. Zhao, R., Ouyang, W., Wang, X.: Person re-identification by salience matching. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 2528–2535. IEEE (2013)
62. Douglas, G., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (2007)
63. Schwartz, W.R., Davis, L.S.: Learning discriminative appearance-based models using partial least squares. In: 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI), pp. 322–329. IEEE (2009)
64. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning implicit transfer for person re-identification. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 381–390. Springer, Heidelberg (2012)

65. Wu, Y., Li, W., Minoh, M., Mukunoki, M.: Can feature-based inductive transfer learning help person re-identification? In: ICIP, pp. 2812–2816 (2013)
66. Layne, R., Hospedales, T.M., Gong, S.: Domain transfer for person re-identification. In: Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream, pp. 25–32. ACM (2013)
67. Ma, A.J., Yuen, P.C., Li, J.: Domain transfer support vector ranking for person re-identification without target camera label information. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 3567–3574. IEEE (2013)
68. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (2007)
69. Zheng, W.-S., Gong, S., Xiang, T.: Associating Groups of People. In: Proc. British Machine Vision Conference (BMVC), London (September 2009)
70. Dong, S.C., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom Pictorial Structures for Re-identification. In: BMVC (2011)
71. http://en.wikipedia.org/wiki/People_counter
72. http://www.vs-re-id-2014.org/
73. Baltieri, D., Vezzani, R., Cucchiara, R.: 3dpes: 3d people dataset for surveillance and forensics. In: Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, pp. 59–64. ACM (2011)
74. http://profs.sci.univr.it/~cristanm/reid2012/
75. http://www.theverge.com/2013/4/16/4230820/in-boston-bombing-flood-of-digital-evidence-is-a-blessing-and-a-curse