

Person Detection from Overhead View: A Survey

Misbah Ahmad¹, Imran Ahmed², Kaleem Ullah³, Iqbal Khan⁴, Ayesha Khattak⁵, Awais Adnan⁶

Center of Excellence in Information Technology

Institute of Management Sciences (IMSciences) Hayatabad, Peshawar (Pakistan)

Abstract—In recent years, overhead view based person detection gained importance, due to handling occlusion problem and providing better coverage in scene, as compared to frontal view. In computer vision, overhead based person detection holds significant importance in many applications including person detection, person counting, person tracking, behavior analysis and occlusion free surveillance system, etc. This paper aims to provide a comprehensive survey on recent development and challenges related to person detection from top view. To the best of our knowledge, it is the first attempt which provides the survey of different overhead person detection techniques. This paper provides an overview of state of the art overhead based person detection methods and guidelines to choose the appropriate method in real life applications. The techniques are divided into two main categories: the blob-based techniques and the feature-based techniques. Various detection factors such as field of view, region of interest, color space, image resolution are also examined along with a variety of top view datasets.

Keywords—Person detection; background subtraction; segmentation; blob based techniques; feature based techniques

I. INTRODUCTION

In video surveillance, one of the key tasks is to detect, identify, and monitor person in crowded and public scenes such as airports, train stations, and supermarkets. The problem of locating a person in the surveillance images and videos sequences from overhead view has been actively researched since last decades. Top view based person detection got importance, due to its better handling occlusion and providing wide coverage of the scene as compared to a frontal view. Overhead based person detection has many applications in various fields, however, the most significant is surveillance systems. Other applications include person detection in indoor and outdoor surveillance systems, person counting [1] (including pedestrian [2], [3], [4], & passenger counting in railway stations [5], shopping malls, airports, buses [6], person tracking [7], [8], [9], [6] [10], [11], behavioural understanding [12], action recognition, person posture characterization [13], crowd analysis [14], industrial work flow [15], [10], provision of large or more coverage area. Furthermore it is also helpful in search and rescue situations. Privacy issues can also be reduced by using an overhead camera because instead of face images overhead view of the person body is captured [16].

Detecting person in overhead videos and images is a challenging job because of the following factors: person body appearance, the wide range of poses, complex backgrounds, unconstrained illumination conditions and self-occlusion. For detecting the person in top view videos and images the understanding of the shape, structure and features of overhead viewed person body is mandatory. Once a person is detected the application system can be further improved.

A variety of top view person detection algorithms are available however these algorithms are still far from human ability to detect the person in images and videos just using a single clue.

One of the major hurdles in the person detection task is the flexible nature of person body, such as variation in poses, size, orientation and direction of person body. The variation in hair colour and texture, clothes colour and texture also add to the complexity. Similarly, the complex environment i.e. cluttered background, crowd, and lighting conditions, also create hurdles for researchers.

Various overhead person detection techniques have been developed in the last few years. In this paper, the significance of overhead view person detection techniques have been studied. To the best of our knowledge, this paper is the first review and survey that objects to cover the most significant advances reported in the literature until now.

From the broad literature of overhead person detection, a representative sample of papers have been selected. This paper classify developed techniques into two groups; blob based and features based.

As discussed earlier that task of person detection can be done using two different perspectives; frontal and overhead. The content of the paper and general frame work of overhead person detection has been summarized in Fig. 1. It can be seen that categorically the developed techniques are divided into two groups: Blob based techniques and feature based techniques.

The different detection factors including camera field of view (narrow or wide), the region of interest (person head, head-shoulder, whole overhead body), colour space (RGB, Depth, HSV & YCbCr), the device used for video and image recordings, recording environments (indoor and outdoor) are also shown in Fig. 1. Furthermore the challenges and applications of overhead based person detection are also depicted in Fig. 1.

The remainder of this paper is structured as follows: Section II begins by introducing the overhead based person detection techniques in general. It is divided into two subsections; the first reflects the different blob-based techniques used by most of the researchers. While the second discusses the features-based techniques. Different datasets used in the literature are also examined in Section II. The discussion on the existing techniques, dataset, issues related to the existing overhead based person detection are covered in Section III. Section IV concludes and summarize the paper and provide some future directions.

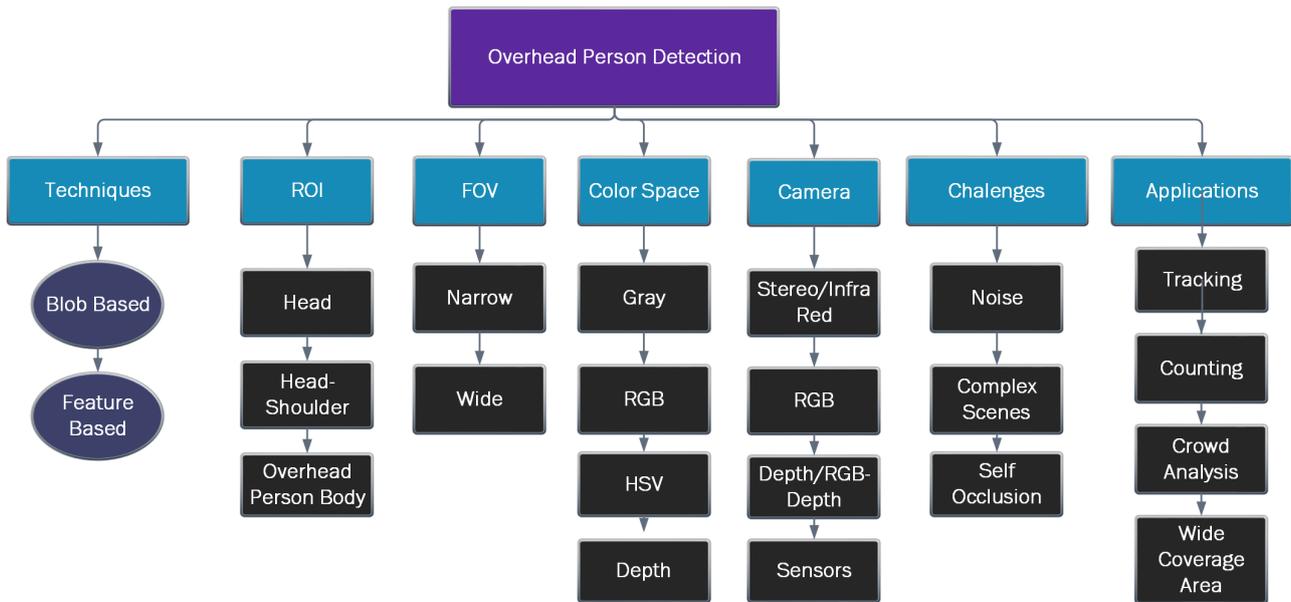


Fig. 1. Overview of the paper and general frame work of Person detection.

II. OVERHEAD BASED PERSON DETECTION TECHNIQUES

Locating a person in top view images and videos is merely a two step process; defining the ROI and localizing the person in the input image and video sequence. In overhead, the person in the image can be localized by its head, head-shoulder or sometimes, the whole overhead body is considered. In below Fig. 2 the overhead view of the person body can be clearly seen. As in the Fig. 2 it can be clearly perceived how the shape, size and body orientation of the person changes from an overhead view. In computer vision, it is important to understand how the person is detected in overhead images and videos. This chapter mainly focuses on algorithms which involved people detection in given overhead videos or images. This chapter broadly divides the relevant literature into two practices. In section, two different practices have been discussed, the first is based on simple blob based algorithms, while the second subsection discusses the feature based techniques used for person detection.



Fig. 2. Some of the images captured from Overhead surveillance cameras [3], [17], [18], [2].

A. Blob based Techniques

In this section, blob based techniques for overhead view person detection is debated. The general framework of blob-based techniques is shown in Fig. 3. These involve background subtraction, foreground extraction, segmentation and pre-processing techniques. Usually, in blob based techniques a foreground image is obtained from background subtraction. Different pre-processing techniques are also used to remove noise, shadow and illumination. A threshold is set to get the desired foreground image.

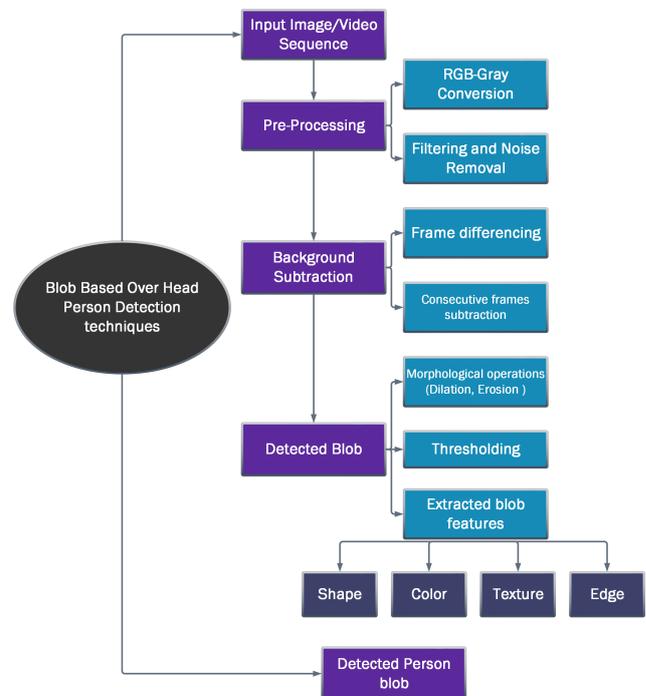


Fig. 3. Blob based Person Detection.

From the foreground image the blob is extracted which is further classified into categories like, a single person, number of the person or another object, etc. The classification is based on the shape, color, motion or other feature of the blob. The review of the techniques which detect a person from top view using blob and background subtraction based techniques is as follows:

Cohan et al. [19] and other studied out in this paper adopted a basic background subtraction and segmentation method to detect the person in overhead images. Table I illustrates that majority of the researcher adopted background subtraction and segmentation based methods to detect the individual in the overhead image and video sequences. In pre-processing, morphological operators including erosion and dilation are typically used by mainstream researchers to detect the required blob. For noise removing Gaussian and median filter methods have been used as revealed in Table I Connected component labeling method for blob extraction is used by [2], [20], [18] and [13] shown in Fig. 4. By examining Table I it can be clearly stated that after background subtraction maximum of the researchers used the extracted blob information to detect the person. The basic features of blob including shape, color, edges and size are considered.



Fig. 4. Connected Component Labeling Method used by [2], [18] and [13].

The blob shape feature is used by majority of the researchers as shown in Table I. Zhang et al. [7] considered the cylindrical shape blob to detect individual body in images. The author in [8] used the blob shape information to reconstruct the hemi ellipsoid head model of the person with the help of image stitching seen in Fig. 5(a). Ozturk et al. [21] used the elliptical shape blob to detect the person in input images (Fig. 5(b)).

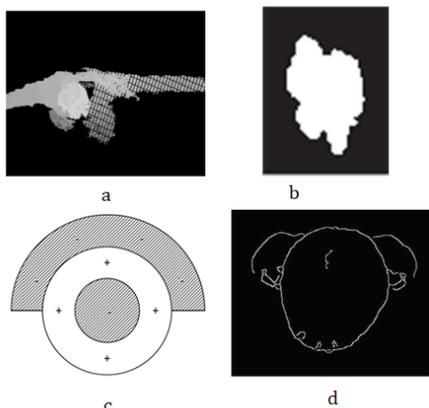


Fig. 5. Images of some shape based features [8], [21] used elliptical based shape, [16] two spherical bounds to detect the person.

Similarly [22] used Hough circle to detect the spherical blob in the image Fig. 5(c). The blobs are further divided into two sub-spherical bounds with a same central point. Inner spherical was used to detect a person head while the outer one was used to detect the head shoulder of the person as shown in Fig. 5(c). Nakatani et al. [16] used the shape information of hair whorl shape to detect the person in overhead images shown in Fig. 5(d).

Moreover blob color information is also considered by the researchers e.g. Cohan et al. [19] used color information to detect the human in topview images. Similarly, [17], [23], [16], [21], [24] also used color information to detect the person in overhead images.

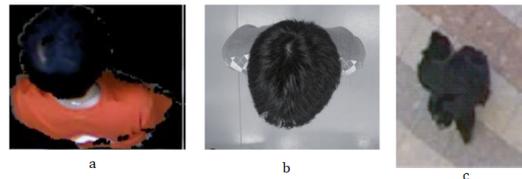


Fig. 6. Color information used by some of the discussed used by some of the authors a [23], b[16], c[21].

As seen in Fig. 7 some of the researchers [25], [7], [16], [21], [9] & [26] took advantage of the edge information of the blob to detect the person. Ozturk et al. [21] and Garcia et al. [9] used the Sobel edge method while Mukherjee et al. [26] used canny edge methods for detecting the person in overhead images as shown in Fig. 7(a), Fig. 7(b) and Fig. 7(c), respectively.

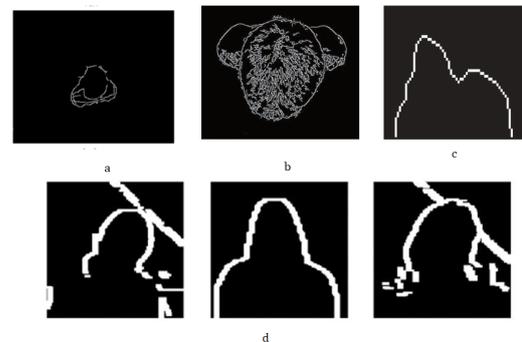


Fig. 7. Sobel edge detector used by [21] and [9], Canny Edge detector [26].

The texture information is also a significant factor, some of the work discussed in this paper also used texture feature to detect the person for example Cohan et al. [19], Chia et al. and Nakatani [16] used hair texture information to detect the person head in overhead images as shown in Fig. 8 below.



Fig. 8. Hair texture information used by [16] [24].

It can be seen from Table I that to detect the person in overhead images there exists a variation in region of interest ROI. Some of the researchers assumed the person head while others considered head-shoulder as ROI to detect the person. Few of them also considered whole overhead person body as ROI for detection. Some examples of considered ROI used by some of the researchers discussed in this paper has been depicted in Fig. 9. In first Row researcher [16], [9] considered Head as ROI. In the second row some [27], [25], [23] & [18] considered Head-shoulders as ROI. Some of them are discussed in this paper [1], [27], [21]& [28] also considered whole overhead body as ROI.

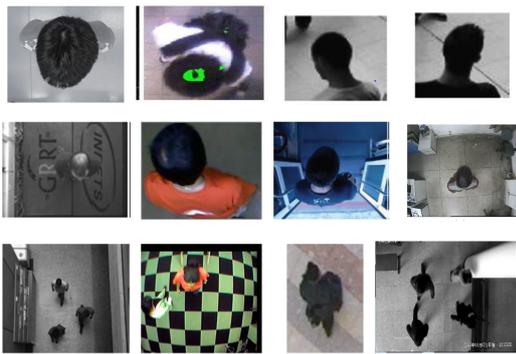


Fig. 9. The First Row shows Head as ROI [16], [9] the second row whose Head-shoulder [27],[25],[23] & [18] as ROI while the third in third row [1],[27],[21]& [28] whole overhead body has been considered as ROI, respectively.

Table I demonstrate that most of the researchers except [12], [16], [22], [22], [8] & [19] used overhead background subtraction methods for people counting. Authors in [25], [29], [28] & [18] have taken account of virtual lines for counting people in top-view images as shown in Fig. 10 below.

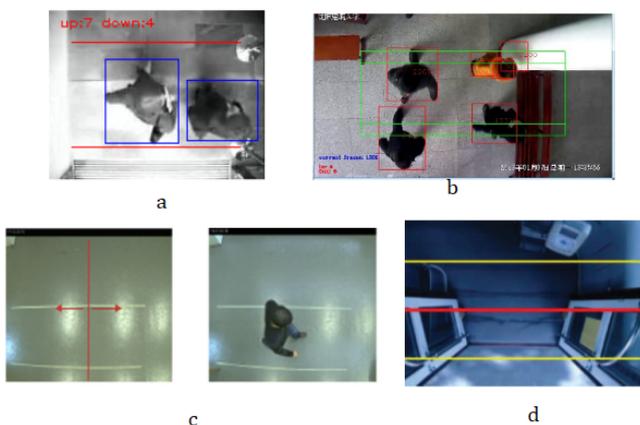


Fig. 10. Virtual lines used for person counting e.g (a) [25] (b) [28] (c) [29] & [18].

In the surveillance system, overhead person detection methods along with tracking has also been developed. For tracking the person, overhead images and video sequences has been considered. For tracking purposes different tracking algorithms are adopted as reflected in Table I.

Authors in [17], [1], [9], [4], & [23] used the Kalman filter whereas [8],[17], [24], [21], & [20] used a particle filter to track the person. Vera et al. [2] used Hungarian algorithm for person tracking. Moreover [24] used median filter and Velipasalar [27] considered mean shift algorithm for tracking. Burbano et al. [6] uses graph structuring based method for tracking people. Some of the examples of different tracking algorithm has been shown in Fig. 11.

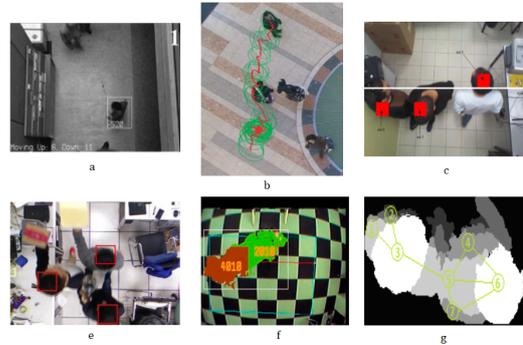


Fig. 11. Results of Different tracking algorithms used by authors discussed in this paper (a) Kalman filter [17], (b,c,d) particle filter [21], [17] [20] and (e) [27] uses mean shift tracking algorithm, (f) [6].

Majority of the work considered for this paper, used depth information to detect the person in images as depicted by Table I and Fig. 12. Few of them also considered RGB images. Authors in [17] and [21] have taken account of HSV model while only [12] used YCbCr model to detect the person in overhead images as shown in Fig. 12. For recording purposes, multiple researchers used variety of recording devices and sensors.

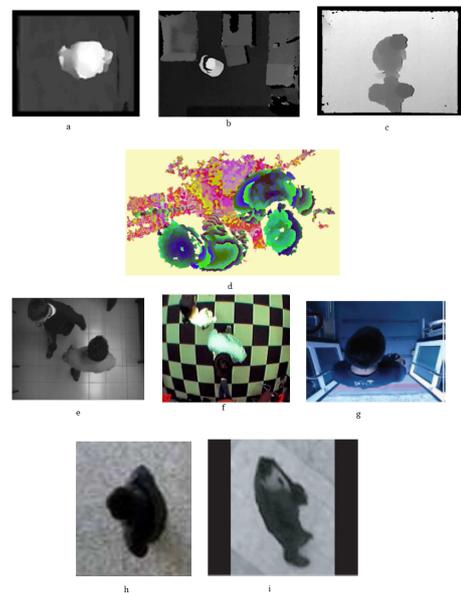


Fig. 12. First Row shows the example of depth images, (d)[30], third row contains images of RGB color space while last row include YCbCr [12]and HSV [21] color model, respectively.

TABLE I. BLOB BASED OVERHEAD PERSON DETECTION TECHNIQUES

S.No	Ref	Year	Purpose	Methodology	Field of View	Region of Interest	Color Space	Data Set	Number of Images/Videos
1	[19]	2000	Detection, Identification	BS, Segmentation, Gaussian Probability, Blob Color/Texture Features, Hair(Color, Texture), DTC, Bayesian Network	Narrow	Head	Gray	Indoor images, Low Resolution,	720
2	[17]	2005	Detection, Counting, Tracking	BS, Segmentation, Blob shape and color features, RGB-Gray conversion, Blob statistical features, Kalman Filter	Narrow	Overhead Body	RGB and HSV	Indoor images, 384 × 288	13138
3	[1]	2008	Detection, Counting, Tracking	BS, Segmentation, Binary Image Conversion, Morphological (Opening), Kalman Filter	Narrow	Head-Shoulder	Depth	Indoor	150 Sequences
4	[27]	2006	Detection, Counting	BS, Blob Statistical Features (Mean & Standard Deviation), Mean Shift Tracking Algorithm	Narrow	Overhead Body	RGB	Indoor, 320 × 240	6 Video Sequences (3-5 min)
5	[25]	2008	Detection, Counting	BS, Foreground Background Edge Model, RGB-Gray conversion, Canny edge detector, Bounding Box, Virtual lines for counting, Tracking (Mass Center)	Narrow	Overhead Body	RGB	Indoor, 320 × 240, 640 × 480	Multiple Video clips
6	[7]	2008	Detection, Counting	BS, Frame Differencing, blob based 3D human cylindrical model(Head, torso, leg) Camera Calibration, Matching Score(size, shape, visibility)	Narrow	Overhead Body	RGB	Indoor, Multiple Sensor Based Cameras , 320 × 240	38 hrs video clip 50fps
7	[12]	2012	Detection, Behaviour Understanding	BS, GMM, Geometric transformation, Hu moments, Behaviour feature (squat, stand, standing)	Narrow	Overhead body	YCbCr	Indoor images	300(100 sample of each posture)
8	[23]	2013	Detection, Counting	BS, segmentation, GMM, Blob color features, U-Disparity, Kalman & particle filter	Narrow	Head	RGB, Depth	Indoor RGB and Depth images	Video sequences
9	[8]	2014	Detection, Tracking	BS, Graph based segmentation, Image stitching, Head hemiellipsoid model, geodesic model distance map, particle tracking filter	Narrow	Head-shoulder	RGB, Depth	Indoor images	888
10	[16]	2012	Detection, Identification	BS, Shadow & Noise removal, Blob features (size, color), Hair (color, whorl)	Narrow	Head	Gray	Indoor images, 400 × 500	240

S.No	Ref	Year	Purpose	Methodology	Field of View	Region of Interest	Color Space	Data Set	Number of Images/Videos
11	[21]	2009	Detection, Tracking	BS, Blob shape features (Elliptical), Blob(color, orientation), Sobel edge detection, Morphological(opening), Median & particle filter	Narrow	Overhead	HSV	Indoor, 1440 × 1080 HD, Human Model(70 × 90)	Video clips
12	[9]	2011	Detection, Counting, Tracking	BS, Segmentation, Blob Shape, Gaussian Distribution, Morphological, Blob separation, Connected component labeling	Narrow	Head-Shoulder	RGB, Depth	Indoor RGB, Depth, Color images	Video clip (292 frames)
13	[2]	2016	Detection, Counting	BS, Segmentation, Blob Shape, Morphological(Filling), Camera Calibration, hungarian Tracking Algorithm	Narrow	Head	Depth, RGB	Indoor RGB & Depth images, 240 × 320	2384
14	[31]	2016	Detection, Counting	BS, Gaussian Smoothing, Virtual line	Narrow	Head	RGB, Depth	Indoor and Outdoor, 640 × 480, 320 × 240, 160 × 120	Video Sequences
15	[32]	2017	Detection	BS, Connected component morphological(opening), Blob(split merge)	Narrow	Overhead body	RGB, Depth	Indoor and outdoor images 640 × 480	Indoor = 33599 , Outdoor = 37799
16	[33]	2018	Detection, Counting	BS, Gaussian filter, MOG, ROC	Narrow	Overhead	RGB, Depth	Indoor images	4500 Sequences
17	[20]	2018	Detection, Counting , Tracking	BS, Depth images subtraction, Gaussian Filter, particle Filter	Narrow	Head-shoulder	Depth	Indoor images	
18	[26]	2011	Detection, Counting	Segmentation, Consecutive frame differencing, Aspatio-temporal background subtraction AMBS, Blob (shape,edge) features, Hough Circle, Canny edge detector, Optical flow(tracking)	Narrow	Head	RGB	Indoor, 480 × 640	Video Sequences (2000 & 5000) samples
19	[34]	2010	Detection, Counting	BS, Segmentation, Blob features(shape, size, velocity, orientation) extraction,spatial temporal map, linear regression model,optical flow	Narrow	Overhead body	RGB	Indoor&Outdoor 640 × 240&240 × 320	Video Sequences (30 min 25fps & 2hr 30fps)
20	[9]	2013	Detection, Counting, Tracking	BS, Segmentation, Equalized difference image, Dilation, Erosion, Circular patterns, Sobel Edge Detector, Kalman Filter	Narrow	Head	Gray	Indoor, 320 × 240	Video Sequence
21	[9]	2013	Detection, Counting	BS, For ground Extraction, Segmentation, Kalman Filter	Narrow	Head	RGB	Indoor	-

S.NO	Ref	Year	Purpose	Methodology	Field of View	Region of Interest	Color Space	Data Set	Number of Images/Videos
22	[35]	2014	Detection, Tracking	BS, MOG, Forgound Mask, Erode, Dilate, contour threshold	Narrow	Overhead body	RGB, YUV	Indoor 320 × 240	Video sequences
23	[5]	2014	Detection, Counting	BS, Segmentation, Head candidate detection, Local maxima detection	Narrow	Head	Depth	Indoor 640 × 480	500 images
24	[6]	2015	Detection, Counting, Tracking	BS, Light weight segmentation, Estimation of relative position & orientation, Blob size features, graph structure representation, camera calibration	Narrow	Head-shoulder	Depth	Indoor 640 × 480	
25	[14]	2015	Detection, Counting	Color Based BS, Segmentation, Morphological Operations (Dilation, Erosion), Local Minima estimation, Blob grouping	Narrow	Head-shoulder, torso, leg	Color Depth	Indoor	268 images
26	[28]	2016	Detection, Counting	BS, Forgound extraction, Blob merge-split, Gaussian model for noise removal, counting zone	Narrow	Overhead Body	RGB	Indoor 1080 × 720	HD Video Sequence
27	[4]	2016	Detection, Counting	Adaptive background model, Segmentation, Reconstruction of 3D points, Kalman Filter	Narrow	Body	Depth	Outdoor Thermal Images 640 × 480	Two Five min videos
28	[24]	2016	Detection, Counting	Color based BS, Water Filling, GMM, Detected Regions Shape, Color gradient model, Median filter	Narrow	Overhead body	Depth	Indoor	11890 images
29	[18]	2016	Detection, Counting	BS, Frame differencing, Morphological operations(Dilation, Erosion), Connected component labeling, Counting zone, Virtual line, Euclidean Distance	Narrow		Overhead Body RGB	Indoor 320 × 240	Video 30fps
30	[13]	2016	Detection, Action recognition	BS, Segmentation, Forgound extraction, Connected component labeling, Action recognition (sitting down, standing up, lying down, waking up, bending and falling on)	Narrow	Overhead Body	Depth	Indoor	Video sequence

B. Feature based Techniques

In recent decades, with the advancement of computer vision and machine learning, various feature based techniques gained importance due to their robustness and efficient detection accuracy. Some of the feature based techniques considered for overhead view person detection has been highlighted in Table II. The general frame work of feature based person detection is shown in Fig. 13 below.

These techniques operate on features extracted from overhead view videos and images. The extracted features are further used to classify person and non-person images. The extracted features contain shape (in the form of contours, edges, corners or other descriptors), color and texture (of hair or clothes), direction, orientation or motion information and sometimes combinations of these. Some used features based algorithms including SIFT and HOG algorithms to detect person in overhead view images as reflected in Table II. The images are often divided into samples for training and testing. These samples are further fed into machine learning classifiers i.e. SVM, adaboost, KNN.

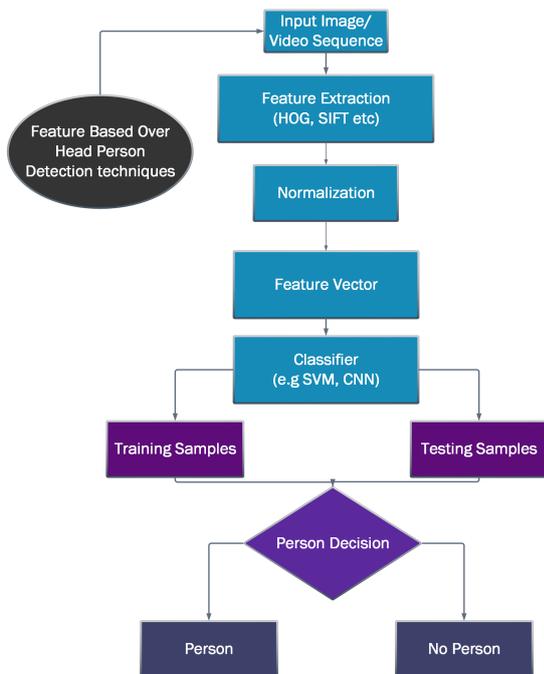


Fig. 13. General Frame Work of Feature based Person Detection.

One of the attractive, preferred and widely used feature based algorithm for person detection is Histogram of Oriented Gradient [36]. This feature based method counts occurrences of gradient orientation localized in image portion. For overhead person detection, Pang et al. in [37] proposed efficient HoG based features detection method for human detection as shown in Fig. 14. Pang et al. [38] proposed another human feature extraction technique using both top view images. The proposed method mainly focused on co-occurrence histograms of oriented gradients of HOG-based human detection method. Rauter et al. [11] proposed a method based on local ternary patterns features for human detection and tracking. The main features used for training SVM was human head and shoulder.

Ahmed et al. [15] used Rotated-HoG algorithm in

overhead images in industrial environment as shown in Fig. 14. Another robust algorithm is proposed by Ahmed et al. [39] in which wide angle lens camera is used for capturing overhead view images of person. The proposed algorithm used bounding boxes with variable size having different orientations. Features from detected windows or bounding box images were extracted using RHOG for building a learning or training model with the help of a linear Support Vector machine. [24] also used HOG features to detect after background subtraction to detect and count the people in complex environments.

Another popular feature based algorithm is SIFT [40]. Ozturk et al. [21] used optical flow of SIFT features to observe the orientation change of person body and head. From Table II it can be clearly comprehended that SVM is the most popular classifier. To detect a person in image Table II shows that scholars used a kernel or sample called detection window (as shown in Fig. 14). The size of the detection window varies with height and width of the person. Likewise [37] used 80×120 size detection window to detect person body in overhead images. The author in [15] used variable size detection window due to the variations in the size of the person as the distance of the person form the camera increases. The author in [11] used 64×128 size detection window along with CoHOG to detect person in overhead images. The authors in [41] and [3] used various size detection windows including 4×4 , 8×8 & 12×12 and 64×96 & 96×96 to detect head-shoulder and person body in overhead images, respectively.

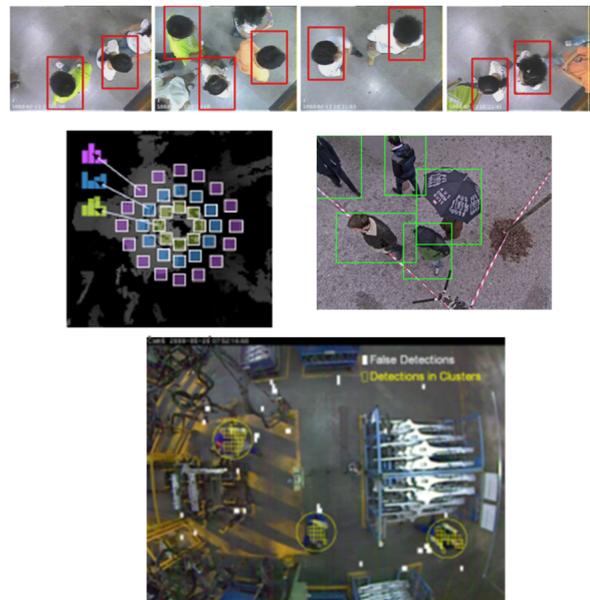


Fig. 14. Example of feature based person detection images taken from [37][41] [3], [15], respectively.

Table II shows that to reduce the feature vector after feature extraction [11] PCA has been used. To reduce feature vector size, [41] used statistical measurement, like mean and standard deviation. For normalization, non maxima suppression NMS was used which improved the performance of the proposed models. In feature based person detection techniques majority of the work consider in this chapter used Non Maxima Suppression in multiple ways (as shown in Table II).

TABLE II. FEATURE BASED OVERHEAD PERSON DETECTION TECHNIQUES

S.no	Ref	Year	Purpose	Methodology	Classifier	Field of View	Region of Interest	Color Space	Data Set	Number of Images/Videos
1	[21]	2009	Detection, Tracking	Motion Vector Detection, SIFT Optical Flow(Around head region) SOFV vector, particle filter.		Narrow	Overhead	HSV	Indoor, 1440 × 1080 HD, Human Mode(70 × 90)	Video clips
2	[37]	2011	Detection	HoG, Detection Window Size 80 × 120, Sub-cell based interpolation algorithm, Non-maximum suppression	SVM	Narrow	Overhead Body	RGB	Indoor and Semi outdoor images, 320 × 240	Almost 1000 images
3	[15]	2012	Detection	Rotated HoG, Variable Size Bounding Box, Non-maximum suppression, Disparity Check between actual and detected positions	SVM	Wide	Overhead body	RGB	Indoor, Industrial SCOVIS data set, 640 × 480	13000 images
4	[38]	2012	Detection	Co-HoG, Sliding Window, CoHOG feature vector is very large (200 000), Incremental principal component analysis (IPCA), non maximum-suppression, Detection Window Size 64 × 128)	SVM	Narrow	Overhead body	RGB	Indoor images 320 × 240, INRIA + & Overhead	6680 images
5	[11]	2013	Detection, Tracking	Depth map for local maxima Local Ternary patterns, Non-Maxima Suppression, Head trajectories	SVM	Narrow	Head-shoulder	Depth	Indoor images	30780 images
6	[24]	2016	Detection, Counting	HoG, Color gradient Model, Median filter	SVM & KNN	Narrow	Head	Depth	Indoor	15188 images & 11890 images
7	[41]	2016	Detection	SLTP and Rauter. Sliding detection window, Statistical Features 4 × 4, 8 × 8, $amd12 \times 12$	SVM	Narrow	Head-shoulder	Depth	Indoor	2382 images
8	[3]	2017	Detection	bounding box, Original Greedy Non-maxima suppression	CNN, R-CNN	Narrow	Overhead body	RGB-Depth	Indoor 800 × 600	850 images
9	[39]	2018	Detection	HoG, RHoG, Bounding box, 64 × 96, 96 × 96, Non-maxima suppression	SVM	Wide	Overhead body	RGB	Indoor SCOVIS industrial, 640 × 480	26032

To increase the accuracy and efficiency of the feature based person detection techniques. Christian Ertler et al. [3] suggested two extensions to the state of the art Faster R-CNN detection model. The RGB and depth representations were fused at different layers of the proposed model. Their experiment shows that the most eligible layer for the fusion was the mid layer.

III. DISCUSSION

The above discussion shows that most of the work done in literature is based on background subtraction techniques. Most of the researchers used simple background subtraction and segmentation methods to detect the ROI in overhead images including (person head, head shoulder and body). These techniques are effective but only on their own developed dataset. As seen in Table I that most of their researcher developed their own indoor based data set for overhead based person detection. Some of the data sets have been recorded in a constrained environment against simple backgrounds as seen in Fig. 10, Fig. 11 and Fig. 12, etc. Some detect the person in overhead images when the person lies exactly below the camera as shown in Fig. 7, Fig. 8 and Fig. 9. Mostly considered the basic blob features i.e. shape, texture, color and size information to detect the person in overhead images. The comparison of the accuracy of developed blob based techniques cannot be made as every method has been tested on their own developed data set according to environments requirement, illumination conditions and application needs.

Background subtraction based methods performed accurately but it mostly suffers from issues including gradual(day-night) and sudden illumination changes, background geometry (entrance of other objects in scene), camera related issues including (camera oscillation, noise, motion blur), shadow factors, same cloths colors as background and some time self occlusion of the own person body. Some feature based overhead person detection methods has also been discussed in this paper. Feature based methods, increases the accuracy and robustness of overhead based person detection methods. But because of the self generated data sets the accuracy of feature based method also cannot be more meaningful as they have also been recorded against simple backgrounds, low illumination condition. Some of the researcher used very complex industrial environment as shown in Fig. 14. With the recent trend of using a feature-based method to increase the efficiency of the algorithm and with the advancement of machine learning algorithms, this work can be extended and make more robust using deep learning. Feature based methods are robust but sometimes it increases the computation cost for researchers. Also, sometimes the size of feature vector is too large which creates complexity for the researchers.

In this paper, we consider methods used different recording devices for capturing topview images, some used single sensor based device e.g. Kinect, while others used multiple sensors. Majority of the researchers consider in this paper used narrow field of view as shown in Fig. 10, Fig. 11 and Fig. 12.

IV. CONCLUSION

In this paper, different overhead person detection methods found in the literature has been reviewed. Differ-

ent issues faced during topview person detection methods have also been discussed. The literature considered for this paper has been divided in to two categories based on blob and features. We have analyzed that most of the work considered for this paper is based on background subtraction and very few discussed feature based methods for overhead person detection. Different challenges and issues related to the data set are also pointed out including noise, complex and cluttered background and self occlusion. From above discussion, we concludes that overhead person detection is widely open to new research and development. The accuracy of discussed methods can not be evaluated because of self recorded datasets. The paper, also suggests new directions for future research. Some of the possible suggestion would be: The developed techniques should be implemented on same benchmark data set, most of the data set has been recorded in constrained environments and illumination conditions. Different feature based methods might be considered for future work. For feature classification advanced machine learning algorithms might be used which makes person detection more efficient and robust.

ACKNOWLEDGMENT

We are thankful to the Institute of Management Sciences and Higher Education Commission (HEC) Pakistan, funding for this research work via NRPU project under project number 5840/KPK/NRPU/RND/HEC/2016.

REFERENCES

- [1] T. Yahiaoui, C. Meurie, L. Khoudour, and F. Cabestaing, "A people counting system based on dense and close stereovision," in *International Conference on Image and Signal Processing*. Springer, 2008, pp. 59–66.
- [2] P. Vera, S. Monjaraz, and J. Salas, "Counting pedestrians with a zenithal arrangement of depth cameras," *Machine Vision and Applications*, vol. 27, no. 2, pp. 303–315, 2016.
- [3] C. Ertler, H. Posseger, M. Optiz, and H. Bischof, "Pedestrian detection in rgb-d images from an elevated viewpoint," in *22nd Computer Vision Winter Workshop*, 2017.
- [4] M. Kristoffersen, J. Dueholm, R. Gade, and T. Moeslund, "Pedestrian counting with occlusion handling using stereo thermal cameras," *Sensors*, vol. 16, no. 1, p. 62, 2016.
- [5] F. Malawski, "Top-view people counting in public transportation using kinect," *Challenges of Modern Technology*, vol. 5, 2014.
- [6] A. Burbano, S. Bouaziz, and M. Vasiliu, "3d-sensing distributed embedded system for people tracking and counting," in *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2015, pp. 470–475.
- [7] Z. Zhang, P. L. Venetianer, and A. J. Lipton, "A robust human detection and tracking system using a human-model-based camera calibration," in *The Eighth International Workshop on Visual Surveillance-VS2008*, 2008.
- [8] T.-E. Tseng, A.-S. Liu, P.-H. Hsiao, C.-M. Huang, and L.-C. Fu, "Real-time people detection and tracking for indoor surveillance using multiple top-view depth cameras," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 4077–4082.
- [9] J. García, A. Gardel, I. Bravo, J. L. Lázaro, M. Martínez, and D. Rodríguez, "Directional people counter based on head tracking," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 9, pp. 3991–4000, 2013.
- [10] I. Ahmed, A. Ahmad, F. Piccialli, A. K. Sangaiah, and G. Jeon, "A robust features-based person tracker for overhead views in industrial environment," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1598–1605, 2018.
- [11] M. Rauter, "Reliable human detection and tracking in top-view depth images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 529–534.

- [12] Q. Lin, C. Zhou, S. Wang, and X. Xu, "Human behavior understanding via top-view vision," *AASRI Procedia*, vol. 3, pp. 184–190, 2012.
- [13] T.-W. Hsu, Y.-H. Yang, T.-H. Yeh, A.-S. Liu, L.-C. Fu, and Y.-C. Zeng, "Privacy free indoor action detection system using top-view depth camera based on key-poses," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, pp. 004 058–004 063.
- [14] D. Ryan, S. Denman, S. Sridharan, and C. Fookes, "An evaluation of crowd counting methods, features and regression models," *Computer Vision and Image Understanding*, vol. 130, pp. 1–17, 2015.
- [15] I. Ahmed and J. N. Carter, "A robust person detector for overhead views," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 2012, pp. 1483–1486.
- [16] R. Nakatani, D. Kouno, K. Shimada, and T. Endo, "A person identification method using a top-view head image from an overhead camera," *JACIII*, vol. 16, no. 6, pp. 696–703, 2012.
- [17] L. Snidaro, C. Micheloni, and C. Chiavedale, "Video security for ambient intelligence," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 35, no. 1, pp. 133–144, 2005.
- [18] J.-W. Perng, T.-Y. Wang, Y.-W. Hsu, and B.-F. Wu, "The design and implementation of a vision-based people counting system in buses," in *2016 International Conference on System Science and Engineering (ICSSE)*. IEEE, 2016, pp. 1–3.
- [19] I. Cohen, A. Garg, and T. S. Huang, "Vision-based overhead view person recognition," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 1. IEEE, 2000, pp. 1119–1124.
- [20] R. Iguernaissi, D. Merad, and P. Drap, "People counting based on kinect depth data," in *ICPRAM*, 2018, pp. 364–370.
- [21] O. Ozturk, T. Yamasaki, and K. Aizawa, "Tracking of humans and estimation of body/head orientation from top-view single camera for visual focus of attention analysis," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. IEEE, 2009, pp. 1020–1027.
- [22] T. Van Oosterhout, S. Bakkes, B. J. Kröse *et al.*, "Head detection in stereo data for people counting and segmentation," in *VISAPP*, 2011, pp. 620–625.
- [23] C. Wateosot, N. Suvonvorn *et al.*, "Top-view based people counting using mixture of depth and color information," in *The second asian conference on information systems, ACIS*. Citeseer, 2013.
- [24] C. Gao, J. Liu, Q. Feng, and J. Lv, "People-flow counting in complex environments by combining depth and color information," *Multimedia Tools and Applications*, vol. 75, no. 15, pp. 9315–9331, 2016.
- [25] S. Yu, X. Chen, W. Sun, and D. Xie, "A robust method for detecting and counting people," in *2008 International Conference on Audio, Language and Image Processing*. IEEE, 2008, pp. 1545–1549.
- [26] S. Mukherjee, B. Saha, I. Jamal, R. Leclerc, and N. Ray, "Anovel framework for automatic passenger counting," in *2011 18th IEEE International Conference on Image Processing*. IEEE, 2011, pp. 2969–2972.
- [27] S. Velipasalar, Y.-L. Tian, and A. Hampapur, "Automatic counting of interacting people by using a single uncalibrated camera," in *2006 IEEE International Conference on Multimedia and Expo*. IEEE, 2006, pp. 1265–1268.
- [28] J. Cao, L. Sun, M. G. Odoom, F. Luan, and X. Song, "Counting people by using a single camera without calibration," in *2016 Chinese Control and Decision Conference (CCDC)*. IEEE, 2016, pp. 2048–2051.
- [29] L. Del Pizzo, P. Foggia, A. Greco, G. Percannella, and M. Vento, "A versatile and effective method for counting people on either rgb or depth overhead cameras," in *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–6.
- [30] J. Nalepa, J. Szymanek, and M. Kawulok, "Real-time people counting from depth images," in *International Conference: Beyond Databases, Architectures and Structures*. Springer, 2015, pp. 387–397.
- [31] L. Del Pizzo, P. Foggia, A. Greco, G. Percannella, and M. Vento, "Counting people by rgb or depth overhead cameras," *Pattern Recognition Letters*, vol. 81, pp. 41–50, 2016.
- [32] V. Carletti, L. Del Pizzo, G. Percannella, and M. Vento, "An efficient and effective method for people detection from top-view depth cameras," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2017, pp. 1–6.
- [33] H. Song, S. Sun, N. Akhtar, C. Zhang, J. Li, and A. Mian, "Benchmark data and method for real-time people counting in cluttered scenes using depth sensors," *arXiv preprint arXiv:1804.04339*, 2018.
- [34] Y. Benabbas, N. Ihaddadene, T. Yahiaoui, T. Urruty, and C. Djeraba, "Spatio-temporal optical flow analysis for people counting," in *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2010, pp. 212–217.
- [35] L. Maddalena, A. Petrosino, and F. Russo, "People counting by learning their appearance in a multi-view camera environment," *Pattern Recognition Letters*, vol. 36, pp. 125–134, 2014.
- [36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *international Conference on computer vision & Pattern Recognition (CVPR'05)*, vol. 1. IEEE Computer Society, 2005, pp. 886–893.
- [37] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient hog human detection," *Signal Processing*, vol. 91, no. 4, pp. 773–781, 2011.
- [38] Y. Pang, H. Yan, Y. Yuan, and K. Wang, "Robust cohog feature extraction in human-centered image/video management system," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 458–468, 2012.
- [39] I. Ahmed and A. Adnan, "A robust algorithm for detecting people in overhead views," *Cluster Computing*, pp. 1–22, 2017.
- [40] D. G. Lowe, "Object recognition from local scale-invariant features," in *iccv*. Ieee, 1999, p. 1150.
- [41] T.-W. Choi, D.-H. Kim, and K.-H. Kim, "Human detection in top-view depth image," *Contemporary Engineering Sciences*, vol. 9, no. 11, pp. 547–552, 2016.