IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 12, 2016, accepted September 3, 2016, date of publication September 13, 2016, date of current version October 6, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2608844

Direction Estimation for Pedestrian Monitoring System in Smart Cities: An HMM Based Approach

RAHUL RAMAN, PANKAJ KUMAR SA, (Member, IEEE), BANSHIDHAR MAJHI, (Member, IEEE), AND SAMBIT BAKSHI, (Member, IEEE)

Centre for Computer Vision and Pattern Recognition, Department of Computer Science & Engineering, National Institute of Technology Rourkela, Rourkela 769008, India

Corresponding author: S. Bakshi (sambitbaksi@gmail.com)

This work was supported by the Science and Engineering Research Board, Department of Science and Technology, Government of India, under Grant SB/FTP/ETA-0059/2014.

ABSTRACT The paper proposes a novel approach for direction estimation of a moving pedestrian as perceived in a 2-D coordinate of field camera. The proposed direction estimation method is intended for pedestrian monitoring in traffic control systems. Apart from traffic control, direction of motion estimation is also very important in accident avoidance system for smart cars, assisted living systems, in occlusion prediction for seamless tracking in visual surveillance, and so on. The proposed video-based direction estimation method exploits the notion of perspective distortion as perceived in monocular vision of 2-D camera co-ordinate. The temporal pattern of change in dimension of pedestrian in a frame sequence is unique for each direction; hence, the dimensional change-based feature is used to estimate the direction of motion; eight discrete directions of motion are considered and the hidden Markov model is used for classification. The experiments are conducted over *CASIA Dataset A*, *CASIA Dataset B*, and over a self-acquired dataset: *NITR Conscious Walk Dataset*. The balanced accuracy of direction estimation for these experiments yields satisfactory results with accuracy indices as 94.58%, 90.87%, and 95.83%, respectively. The experiment also justifies with suitable test conditions about the characteristic features, such as robustness toward improper segmentation, partial occlusion, and changing orientation of head or body during walk of a pedestrian. The proposed method can be used as a standalone system or can be integrated with existing frame-based direction estimation methods for implementing a pedestrian monitoring system.

INDEX TERMS Visual surveillance, occlusion handling, pedestrian direction estimation, perspective distortion, hidden Markov model.

I. INTRODUCTION

Direction estimation of a moving subject is an important task during many video processing and computer vision oriented applications such as behaviour analysis, motion analysis, traffic control systems, smart cars, gait based pedestrian identification and visual surveillance at secure public places. Motion of a pedestrian in a 3D global plane can be completely analysed in 2D camera plane by three factors, i.e. direction of motion, velocity of motion, and depth information. Therefore, information about direction of motion of pedestrian is very significant in motion analysis.

In many fields like accident avoidance mechanism in cars, traffic control system, and visual surveillance where dynamic decisions are needed to be taken, advance knowledge of direction of motion is very handy. This establishes the task of pedestrian direction estimation as an important domain of research. Specifically, a prior knowledge of most probable direction of pedestrian is crucial for accident avoidance and for different surveillance tasks as optimal camera placement [1] and for seamless object tracking [2].

Occlusion is a severe issue in pedestrian monitoring. Different authors have shown awareness towards problems arising due to occlusion and presented their views on treatment of occlusion. The survey by Yilmaz *et al.* [3] presents a section with some of the earlier research on treatment of occlusion. A recent survey on handling occlusion while object tracking is shown in [4].

Although a lot of research on occlusion treatment targets estimation of velocity, estimation of current location and its probable location of reappearance after occlusion, the other dimension of research towards occlusion has also been explored where the prediction of occlusion is targeted [2].

This prediction of occlusion will assist the existing methods of occlusion handling and occlusion avoidance with a prior knowledge of occlusion. Information about pedestrian's direction of motion thus becomes vital in such situations.

Apart from surveillance, direction of motion estimation has been used in many areas such as collision prediction mechanism in smart cars, traffic control system, assisted living, behaviour analysis and long term motion estimation. Markis and Ellis [5], Wakim *et al.* [6] and Antonini *et al.* [7], have used direction estimation for pedestrian behaviour model. Abramson and Steux [8] and Large *et al.* [9] have performed long term motion analysis and path prediction. In a similar work to occlusion prediction, Tsuji *et al.* [10] have used relative motions and used the knowledge of direction estimation for collision prediction.

Many works for direction estimation of a moving pedestrian have extracted the orientation information of the pedestrian in each frame and observed the orientation in subsequent frames to estimate the direction of motion. To cope with the challenges of low resolution images, complex and dynamic background, and illumination variation, many researches have used low-level features like Histogram of Gradient (HoG), Scale Invariant Feature Transform (SIFT) or Haar like features with Support Vector Machine (SVM), Neural Network (NN), Regressions, and Adaboost classifiers.

Shimizu and Poggio [11] have used SVM on Haar wavelet co-efficient to distinguish different orientation and estimate direction of motion with study of orientation over subsequent frames. Gandhi and Trivedi [12] have used multi-class SVM on HoG based features to distinguish different orientations and further used Hidden Markov Model (HMM) for integration from multiple frames to estimate direction of motion. Enzweiler *et al.* [13] have presented an integrated approach for single frame pedestrian classification and orientation estimation to predict the direction of motion. Zhao *et al.* [14] have used a Haar like feature vector subjected to Adaboost classifier for orientation. Cascaded orientation estimation is applied for body and head orientation estimations. Further, most frequent orientation estimate and rounded average of estimated sum are used for direction estimation.

Many other researchers have also attempted to segment different body parts and study their orientations individually. Recently, Flohr *et al.* [15] have proposed probabilistic pedestrian orientation system where head and body orientations are studied separately. The proposed research is intended to overcome faulty detection and provide robust orientation and direction estimation. Bensebaa *et al.* in their research [16], [17] have segmented different body parts i.e. head and shoulders, knees, feet, and body. The authors have attempted to study their orientation separately over their silhouettes to estimate the heading direction of pedestrian.

The articles so far have utilised the static cues and generalised it over multi frames. Liu *et al.* [18] have utilised RGB-D sensor where RGB sensing helps in illumination change and D sensing for depth of the subject. The article provides insight about how only static cues (intra frame) are not sufficient and needs to be complemented with dynamic cues (inter frame) for orientation and direction of motion estimation. The authors have proposed Dynamic Bayesian Network System (DBNS) to effectively employ the complementary nature of both static and motion cues, which motivates to incorporate dynamic cues for direction estimation. A lot of research for direction and orientation estimation of pedestrian has been performed over videos exploiting dynamic cues along with appearance patterns.

Goel and Chen [19] have called their classification method as Global Locale Motion Pattern Classification (GLMPC) where they have attempted to detect pedestrian in a video. In order to classify a pedestrian from a non-pedestrian, they have proposed 3 subclasses of pedestrian depending on 3 different walk directions. The algorithm classifies motion of pedestrian into 3 discrete directions. Andriluka *et al.* [20] have exploited body pose but not the motion cues in the temporal pattern and hence can only estimate the orientation and not the direction of motion. Chen *et al.* [21] have proposed head and body orientation based calculation of direction estimation in low resolution videos. Authors in this article have exploited intra-frame features for body cues and then the position exploiting temporal patterns in particle filtering framework. Both the researches have discretized the direction to 8 levels. Baltieri *et al.* [22] have proposed an orientation classifier approach exploiting only appearance model. Liu and Ma [23] have proposed an on-line orientation classifier approach on field camera. The article attempts to suppress the effect of perspective distortion and proposes a Reliable Motion Direction (RMD) determination method that assumes a constant apparent velocity of pedestrian walk. These methods are very limited in cases when pedestrians are moving slow, are stationary or suffering occlusion, however they work fine during constant motion of pedestrian.

Table 1 summarises a few landmark researches for direction estimation of moving subject. The study of existing survey for direction of motion of a subject reveals following facts:

- Researchers have used features like Haar, HoG or exploited silhouette over frames to estimate their orientation.
- This process is extended over multiple frames and then the trend over multiple frames are either modelled [12] or statistically concluded for direction estimation like selecting most frequent orientation or applying rounded average [14].
- The opposite pair of directions like approaching and departing from camera is always confusing if calculated over individual frame. A few articles [14], [24], [25] have reported them.
- The low resolution videos fail to capture many features, and factors like noisy environment, imprecise acquisition, or improper segmentation may further make the cause difficult.
- Most of the the researchers have considered 8 equiangular discrete directions as optimal

TABLE 1. A few landmark research on direction of motion estimation.

Sl no.	Article Details	Proposed Method	Experimental Details
1	**Article(Year)** Direction estimation of pedestrian from images (2003) [11] **Authors** Shimizu and Poggio	**Objective** Walking direction of pedestrians from field surveillance camera. Expected to be useful for autonomous cars and robots. **Approach** Orientation of subject is estimated individually in frames of a video, and voting based results used to estimate the direction of motion **Discrete walk directions** 16 equiangular discrete direction angles ranging between [0°, 360°] from view axis discrete directions are further divided into 2 classes with alternate angles in different class	**Modelling** Haar wavelets to generate feature vectors of the input frame and train 16 individual classifiers, SVM with linear kernel function used for classification over the extracted feature **Database** All directions are trained with 1k positive and 7k negative frame samples. All directions are tested with 150 frame samples, with no overlap between training and testing database **Constraints, Limitations and Assumptions** People are assumed to be detected. Field cameras are static. Directions that do not fall to any of the class is expected to fall in a class of closest proximity **Results** Recognition rates with (target and neighbouring directions) are claimed to be more than 90% for all the directions however the result of classification with only target direction is between 50 to 90% With more frames involved in direction estimation for each class, estimation results are found better. Average recognition rate with 10 frames is nearly 84%
2	**Article(Year)** A fast motion estimation algorithm based on direction of motion vectors (2007) [30] **Authors** Nisar and Choi	**Objective** Prediction of motion vector using spatio-temporal neighbourhood information to estimate the walk direction **Approach** Predicted Motion Vector (PMV over frames of a video to predict the direction of motion) **Discrete walk directions** 4 and 8 discrete walk directions presented as walk-sectors	**Modelling** Spatial and temporal neighbouring pixel blocks are used to calculate PMV over the same block of the frames to predict the direction of motion **Database** First 100 frame of Miss America and Coast Guard CIF video sequence, and Football SIF video sequence **Constraints, Limitations and Assumptions** PMV is estimated over 5 discrete numbers of spatial and temporal neighbouring motion vectors **Results** The correctness of direction estimation depends on the right search of the moving block in the neighbourhood. The search results are presented in terms of PSNR which is in the range of 37.36 to 25.24
3	**Article(Year)** Image based estimation of pedestrian orientation for improving path prediction (2008) [12] **Authors** Gandhi and Trivedi	**Objective** Path prediction of pedestrian based on orientation and direction estimation for avoiding collision on roads **Approach** Orientation of subject is estimated individually in frames of a video, and transition between orientations are modelled to estimate the direction of motion **Discrete walk directions** 8 equiangular discrete direction angles ranging [0°, 360°] from view axis	**Modelling** Image gradient orientations are used to form HOG features to train 8 separate classifiers using SVM, inter frame orientation transitions over time are modelled using HMM **Database** Training was performed using INRIA pedestrian dataset [31], from where 664 snapshots were selected for training. For testing the classifier, images acquired from self-acquired video footage taken at a signalized intersection are used, 427 frames are used for testing **Constraints, Limitations and Assumptions** Pedestrian detection is assumed as already addressed and subjects with their bounding box are available. The aspect ratio of bounding box of test data taken manually is resized to the size of training data to avoid scaling error. During direction estimation using HMM, it is assumed that transition probabilities depend only on the difference in orientation. **Results** Recognition rates with target orientation in single frame is claimed to be 41.7%, with target and neighbouring orientation in single frame are 75.4%, whereas direction estimation results with single and along with neighbouring frames are claimed to be 49.7% and 81.3%

Continued on next page...

TABLE 1. ***(Continued.)*** **A few landmark research on direction of motion estimation.**

Sl no.	Article Details	Proposed Method	Experimental Details
4	**Article(Year)** Pedestrian detection and direction estimation by cascade detector with multi-classifiers utilizing feature interaction descriptor (2011) [26] **Authors** Goto *et al.*	**Objective** Detection and direction of motion estimation of a pedestrian. The proposed method also attempts to estimate the distance of the moving subject from camera **Approach** Cascade approach with multi-classifiers using FIND. For improved efficiency and accuracy cascade approach with multi-classifiers are used that are trained in estimating direction of motion and distance from camera. This framework gives both the information simultaneously **Discrete walk directions** 4 equiangular discrete direction angles ranging [0°, 360°] from view axis	**Modelling** Feature Interaction Detector (FIND) are used as features. For classification, cascade approach with multi-classifiers is used to overcome computational disadvantage and to improve detection performance **Database** Self-acquired database with diverse conditions consists of 8166 pedestrians **Constraints, Limitations and Assumptions** Working on FIND features has high calculation cost **Results** The effectiveness of classifiers for direction estimation is presented in terms of ROC curves, whose analysis shows that the best case detection accuracies for back, front, left and right directions are upto 90%, 87%, 93%, and 87% respectively
5	**Article(Year)** Integrated pedestrian and direction classification using a random decision forest (2013) [27] **Authors** Tao and Klette	**Objective** Analysis of behaviour of pedestrian in a scene by simultaneous classification of pedestrians and their walking directions in mono-ocular and multi-view scenario **Approach** Random Decision Forest based approach is followed where presence of pedestrian and their direction of motion is attempted to be classified in a single module **Discrete walk directions** 4 equiangular discrete direction angles ranging [0°, 360°] from view axis	**Modelling** Image gradient orientations with three levels of cell size are adopted to form HOG like feature vector which are subjected to weighted sum functions with pedestrian classification term and direction classification term. Nine different modelling experiments with eight Random Decision Forest based modelling and one SVM based modelling are performed for simultaneous classification of pedestrian and walk direction **Database** Training: 4935 frames from TUD multi-view pedestrian dataset [32], 15000 non-pedestrian and 6000 pedestrian frames from Daimler Mono Pedestrian Detection Benchmark datasets [33]. Testing: 248 direction labelled frames from TUD multi-view pedestrian dataset, 10000 non pedestrian frames and 3000 direction labelled bounding boxes from Daimler Mono Pedestrian Detection Benchmark datasets **Constraints, Limitations and Assumptions** Aspect ratio of bounding box of test data is taken same as that of the size of training data to avoid scaling error. Directions are classified with 4 and tested with 8 discrete directions. Classification result of in-between directions is taken correct, if classified adjacent. Features could not handle similarity in head pose resulting a higher misclassification rate in opposite pair of directions **Results** The classification result claims that a direction estimation in the range of 26% to 95% is achieved with 4 direction classifiers
6	**Article(Year)** Joint probabilistic pedestrian head and body orientation estimation (2014) [15] **Authors** Flohr *et al.*	**Objective** Head and body orientation for estimating the direction of motion in the context of intelligent vehicle **Approach** Joint probabilistic estimation approach for head and body orientation **Discrete walk directions** 8 equiangular discrete direction angles ranging [0°, 360°] from view axis	**Modelling** Decoupled pedestrian tracker estimates the position of body. This track information used by orientation trackers of body and head jointly. Finally coupled single frame orientation estimates are integrated over time by particle filter **Database** 9300 manually contour labelled pedestrian samples from 6389 images with separate training and testing sets **Constraints, Limitations and Assumptions** Authors have constraint the body orientation to previous head orientation when body orientation is changing. Localization needs stereo vision **Results** Results are presented in terms of plots as error in angular estimation which is around 20° to 40° for head orientation and around 10° to 25° for body orientation at moderate distance of pedestrian from camera taken over 37 valid estimated walks

Continued on next page...

TABLE 1. ***(Continued.)*** **A few landmark research on direction of motion estimation.**

Sl no.	Article Details	Proposed Method	Experimental Details
7	**Article(Year)** Inferring heading direction from silhouettes (2015) [17] **Authors** Bensebaa *et al.*	**Objective** Inferring the heading direction of pedestrian from its silhouette's geometric features. Useful for low resolution image sequences **Approach** Geometric features of different body shapes like shoulder, head, feet, and knee's shapes are jointly inferred to predict the heading direction **Discrete walk directions** 24 equiangular discrete direction angles ranging [0°, 360°] from view axis	**Modelling** Geometric features of silhouettes of moving subject used, i.e. inflections of knees, direction of foot shape and variation of silhouette's width along the shape of head-shoulders and the length of each shoulder. Base line defined for feet's feature and Chetverikov's algorithms used to extract knee and head-shoulder features to estimate the inferring direction **Database** PETS dataset [34] is used **Constraints, Limitations and Assumptions** Using only R_w as feature creates confusion for the algorithm in front and backward movement, hence all features are needed to be used in combination, which makes the algorithm computationally expensive. Proper segmentation is required for proper silhouette extraction **Results** Some results in the form of video frames are presented however, no quantitative results are available
8	**Article(Year)** Online person orientation estimation based on classifier update (2015) [23] **Authors** Liu and Ma	**Objective** Online appearance based determination of pedestrian walk direction **Approach** LaRank (a low cost linear multi-class SVM) is applied over spatio temporal appearance feature for online direction classification. **Discrete walk directions** 8 equiangular discrete direction angles ranging [0°, 360°] from view axis	**Modelling** Spatio-temporal appearance features are used along with reliable motion direction determination to update orientation estimation thus producing online results of updated direction estimation **Database** Training: TUD multi-view pedestrian dataset Testing: PKU person orientation dataset [35] and 3DPeS dataset [36] **Constraints, Limitations and Assumptions** During online classifier update the effect of perspective distortion in monocular vision of camera is not included in estimation result and the effect is suppressed by linear transformation. During RMD determination, apparent velocity of pedestrian is constraint to be uniform (i.e. 1 *px*) **Results** The classification result claims the average accuracy of direction estimation using various proposed methods to be in the range of 69% to 91%

choice [2], [12], [14], [15], [18]. However, few articles have considered 4 discrete directions [26], [27], 16 discrete directions [11], and 24 discrete directions [17] as well.

We propose an HMM based method that operates over a video for pedestrian direction of motion and exploits the temporal terrains i.e. the pattern of change of width and pattern of change of height of bounding box of the identified pedestrian blob in the frame. It uses these features to train different HMMs to classify the movements among different direction classes. The robustness of dynamic cues selection is already presented in earlier work [28]. In this article, we have used 8 different classes for 8 equiangular discrete directions ranging [0°, 360°] from view axis (refer to Fig. 1).

The sinusoidal wave pattern with perspective affected distortion uniquely identifies a subject as pedestrian from any other moving object like vehicle or animal in the scene due to unique human gait patterns. Moreover, the method also

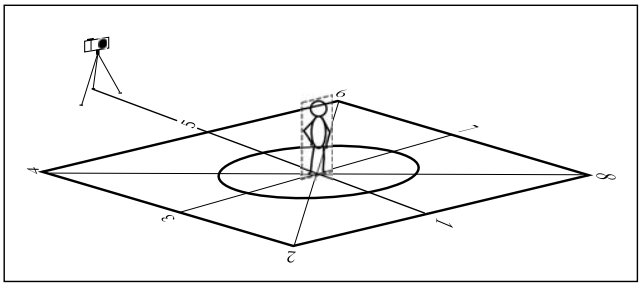


FIGURE 1. **8 equiangular discrete directions with respect to field camera.**

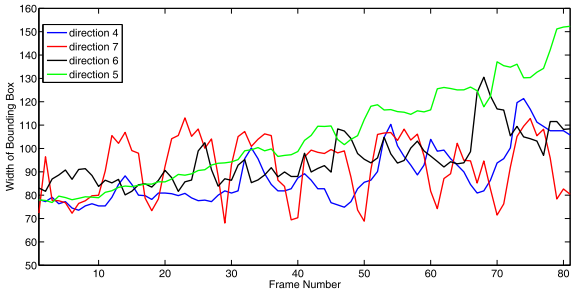
overcomes confusion between cases where subjects moving towards and away from camera. The proposed method is robust to various other issues like illumination changes, environmental factors, partial occlusion for few frames and low resolution of surveillance videos. The proposed method can either be used alone or with existing methods of orientation

estimation over consecutive frames to enhance the direction estimation results.

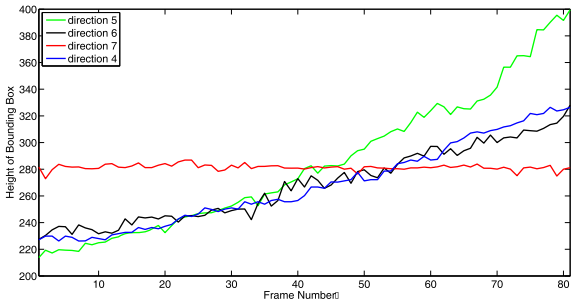
Problem formulation, theoretical proposal, feature selection, and HMM based training, testing and inferences are presented in the Section II. About the experimental environment and analysis which include assumptions, constraints, database details, evaluation parameters and steps involved in the experiment are discussed in the Section III. In Section III-E, results related to classification over different databases using various performance metrics and comparison of proposed method with some landmark research are presented. Section IV includes conclusion and scope for future work.

II. PROPOSED METHOD

We propose to exploit the dynamic cues of a surveillance video to estimate the direction of motion of a pedestrian. The temporal changes in the dimensions of bounding box are found to be unique as subject moves in different directions. We have chosen 8 equiangular discrete directions in the context of our proposed method as presented in Fig. 1. Changes in width and height with respect to a few selected directions are plotted in Figs. 2(a) and 2(b) that demonstrates unique patterns with respect to different directions of motion. These patterns can be exploited to estimate the same. Much elaboration on these graphs and inferences are presented in detail in Section III-D.



(a)



(b)

FIGURE 2. Plots representing unique pattern of change in dimensions as the pedestrian moves in different directions. (a) Pattern of change of width. (b) Pattern of change of height.

As the pattern of change of height and width along with the displacement of centroid appears to be enough to estimate the direction of motion, yet considering such features alone can handle a limited type of cases. A few cases of direction estimation that cannot be addressed with such features are:

- Partial static occlusion, where the pattern of change of width or pattern of change of height are not visible for few frames
- Segmentation errors, that can add noise and can also deform the shape of the pedestrian's blob
- Pedestrian's carrying conditions, clothing conditions, and unusual postures like putting hands in pocket etc can deform one of the feature

Such situation demands to exploit both width and height feature simultaneously. Moreover, the modelling technique used for the human motion patterns as captured in field camera has to be properly justified and robust enough to handle the cases of partial occlusion.

To model this unique temporal pattern of the dimensions of moving subject in field camera, authors are motivated to use HMM for direction classification. The justification to use HMM, formulation of feature and preprocessing are discussed next. The involved steps are:

- Preprocessing
- Feature formulation for HMM
- Problem modelling through HMM
- Machine learning and classification

A. PREPROCESSING

Field camera records video footage that is needed to be pre-processed before feature extraction. The step of pre-processing can be further divided into following sub-steps:

- Background subtraction for pedestrian selection
- Connected component generation through morphological operations
- Frame rectification and unwanted blob removal
- Bounding box fitting over pedestrian blob

1) BACKGROUND SUBTRACTION FOR PEDESTRIAN SELECTION

Through the method of background subtraction, blobs of non-stationary pixels are separated from stationary pixels. Moving blobs are categorised as foreground while rest as background. The existing databases (as presented in Table 2), which are already background separated do not require this step. However, we have used Visual Background Extractor (ViBe) [29] for background subtraction for self-acquired dataset to ensure fast and accurate computation, citing the in-time processing being the prime requirement of the proposed method. ViBe is a robust method of background subtraction. This is also an adaptive method as it works over different environmental constraints. Hence, this method has been adopted for the proposed method. The set of background pixels also contains undesired blobs other than that of pedestrian and frames need morphological operation followed by a certain rectifications as discussed next.

TABLE 2. Different databases used, their parametric properties, and special cases present.

	CASIA Dataset A	CASIA Dataset B	NITR conscious walk database
Parametric Properties			
Environment	Outdoor	Indoor	Outdoor
Number of Subjects	20	124	21
Number of Sequences	240	13640	524
Walk Directions	6	11	8
Resolution	$352 \times 240 \times 8$	$352 \times 240 \times 8$	$1280 \times 720 \times 24$
Color Model	Binary	Binary	RGB
Time of Release	Dec. 2001	Jan. 2005	Jan. 2015
Special Cases Present			
Partial Static Occlusion	√	√	√
Different Head and Body Orientation			√
Carrying and Clothing Conditions		√	
Varying Walk Velocities	√	√	√

2) CONNECTED COMPONENT GENERATION THROUGH MORPHOLOGICAL OPERATION

After background subtraction, frame may contain undesired blobs identified as foreground. This may happen due to improper segmentation, partial occlusion or due to presence of noise. Dilation is performed to connect the nearby blobs to overcome the unwanted separation of connected foreground as different blobs. Different body parts of a pedestrian might be identified as different blobs in the binary frames. Disk dilation has been performed iteratively over such frames on different blobs until the nearby blobs form a single connected component. In the article, we have used dilation operation with disk as structuring element with radius 7 running for 5 iterations. The objective of this morphological operation is to fill unwanted holes in the foreground and to strengthen the foreground pixel near the body joint regions. However, for severely cluttered images the authors have adopted the background subtraction methods by Yao and Odobez [37] and Reddy *et al.* [38]. The resulting connected foreground helps in selection of single largest blob.

3) FRAME RECTIFICATION AND UNWANTED BLOB REMOVAL

After morphological operation largest connected component has been chosen as desired foreground while deleting rest of the foreground blobs. With fixed background and moving foreground, optical flow [39], [40] based methods may also be applied to overcome improper segmentation.

Figs. 3(a) and 3(b) show two frame sequences where the pedestrians are moving in discrete direction 6 and direction 3 respectively. Figs. 3(c) and 3(d) show corresponding frame sequences after noise removal.

4) BOUNDING BOX FITTING OVER PEDESTRIAN BLOB

The temporal change in the dimensions of a moving blob can be defined efficiently by the change in the dimension of bounding box fitted over the blob. With this motive a rectangular bounding box has been put over the blob and their temporal change in subsequent frames are recorded. Figs. 3(e) and 3(f) present same set of frame sequences after fitting rectangular bounding box over the identified pedestrian blob.

B. FEATURE FORMULATION FOR HMM

The temporal change in consecutive frames is the output after pre-processing. 8 equiangular discrete directions of motion, D_1 through D_8 that a moving subject may achieve with respect to field camera is considered. The change in the dimension of a moving object as perceived in a camera view is perspective in nature and unveils distinct patterns. This fact has been exploited to model the patterns for each of the 8 distinct directions of motion using HMM. The temporal pattern of change of width and height of a moving subject is unique for each of the discrete direction. Figs. 2(a) and 2(b) show the pattern of change of width and height over the frames in different directions respectively. These unique patterns are now needed to be trained using a machine learning algorithm. The temporal changes in the pattern are stateless and hence follow Markovian property; this gives a good reason to use HMM for training.

The features identified to be unique for each direction are the temporal change pattern in height and width of the bounding box along with the displacement direction of its centroid. Thus, the aggregated feature is formed as a 1D array

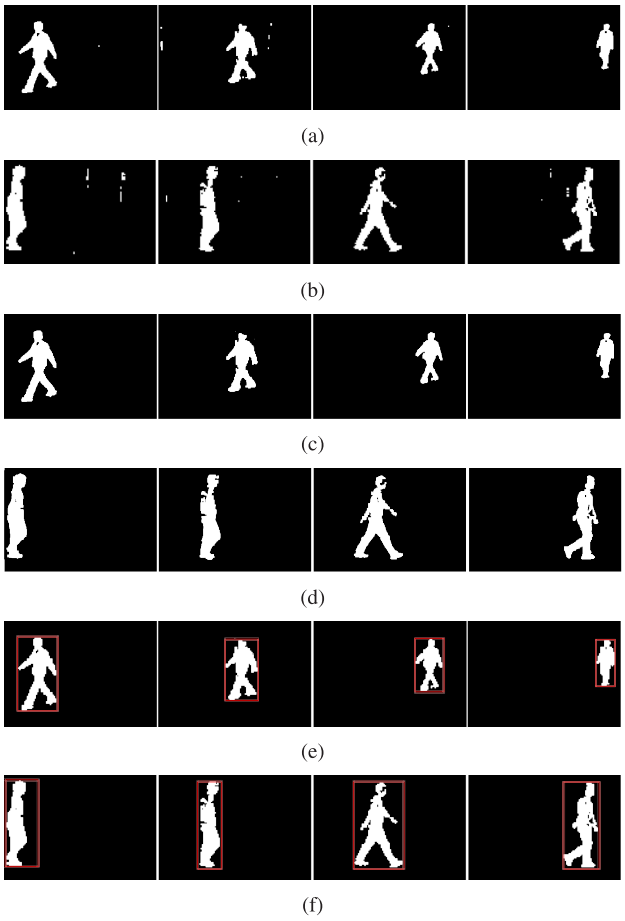


FIGURE 3. Steps of preprocessing. (a) Sample frame sequence # 1 with noise. (b) Sample frame sequence # 2 with noise. (c) Sample frame sequence # 1 after noise removal. (d) Sample frame sequence # 2 after noise removal. (e) Sample frame sequence # 1 after boundary fitting. (f) Sample frame sequence # 2 after boundary fitting.

which is linear concatenation of temporal pattern of change in height, padding (for displacement direction of centroid) and temporal pattern of change in width.

Temporal change patterns of the dimensions of the bounding box exploits perspective distortion and can classify directions among (direction 1), (directions 2 *and* 8), (directions 3 *and* 7), (directions 4 *and* 6), and (direction 5). The displacement direction of centroid of bounding box can differentiate between pair of directions with similar perspective distortions but opposite displacement direction of centroid i.e. (directions 2 *and* 8), (directions 3 *and* 7), and (directions 4 *and* 6) (refer to Fig. 4). Thus the aggregated feature can classify among 8 direction classes as proposed in this article (cases available in CASIA Dataset A and NIT Conscious Walk Dataset, discussed in Section III-B).

In cases where all the available discrete directions of pedestrian have same displacement direction of centroid (as in cases of CASIA Dataset B, discussed in Section III-B), the displacement direction of centroid is not required, as among the available direction classes in this database, pair of directions with similar perspective

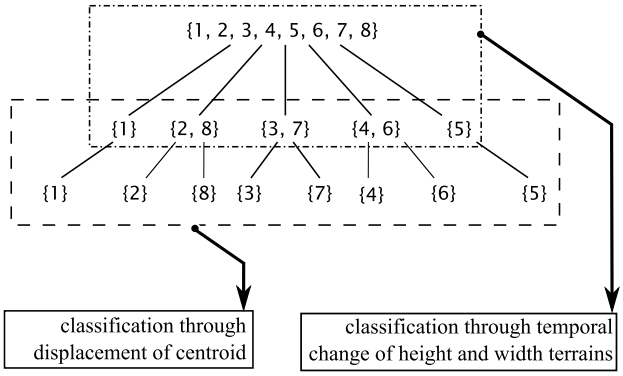


FIGURE 4. Genesis of the aggregated feature vector.

distortion but opposite displacement direction of centroid are not present and padding may be removed from the aggregated feature.

Corresponding feature vector can be formulated in two possible ways as follows:

$$\textit{Case I:}\ \ \Delta h|_t + pd + \Delta w|_t$$
$$\textit{Case II:}\ \ \Delta h|_t + \Delta w|_t$$

where,

$\Delta h|_t$: temporal change of height (h) of the bounding box fitted on the silhouette of the moving pedestrian over time t extracted from consecutive frames in video

$\Delta w|_t$: temporal change of width (w) of the bounding box fitted on the silhouette of the moving pedestrian over time t extracted from consecutive frames in video

pd : sequence denoting padding to discriminate $\Delta h|_t$ and $\Delta w|_t$

The padding (pd) can be represented as:

$$pd = \begin{cases} 0^n & \text{if } x_{f_0} > x_{f_c} \\ 1^n & \text{if } x_{f_0} \leq x_{f_c} \end{cases}$$

where, in the 2D Cartesian coordinate system,

x_{f_0} : abscissa of the centroid of the foreground in first frame

x_{f_c} : abscissa of the centroid of the foreground in current frame

Hence pd should be an array filled with n zeros when the pedestrian moves in directions 6, 7, and 8. In other cases (for movements along directions 1, 2, 3, 4, and 5), pd should be filled with ones (refer to Fig. 6). Thus pd will not only act as a unique separator between height terrain and width terrain, but will also bear partial discriminating power with respect to movement direction. Fig. 5 illustrates two typical sample cases (for directions 3, and 7) graphically depicting the change in centroid and selection of pd accordingly.

However it is not sufficient only to choose whether pd will be filled with 0 or 1. The number of zeros or ones (n) has

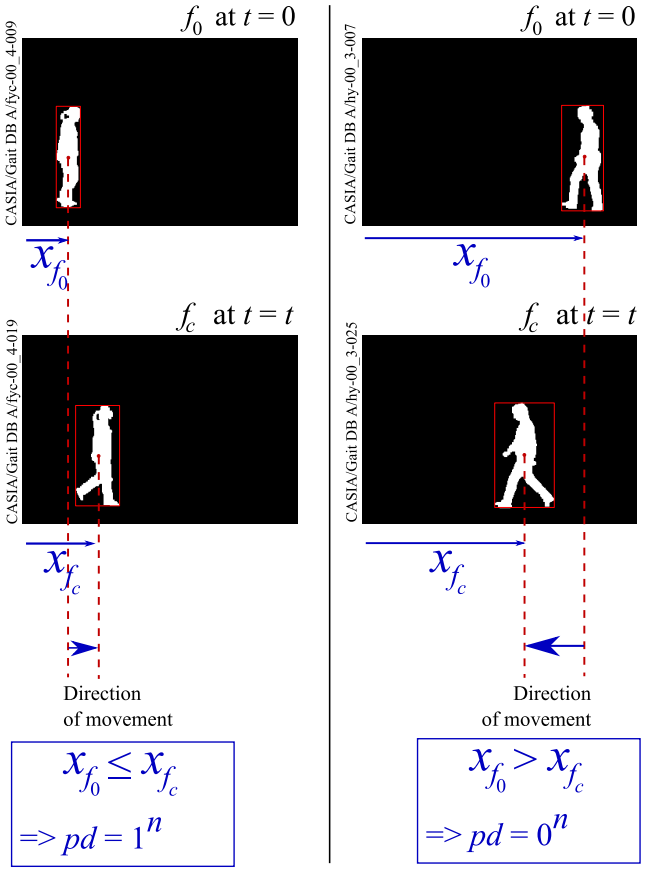


FIGURE 5. Illustration for finding *pd* in different cases.

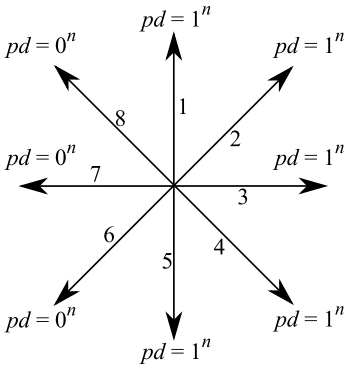


FIGURE 6. Selection of *pd* reflecting classification through displacement of centroid.

to be chosen wisely. The value of n is chosen satisfying two contradictory objectives:

i. The n-unit padding should be long enough so that it can act as a unique separator between height and width terrain
ii. The n-unit padding should be minimal in length to reduce the total feature vector length (as a long feature vector will cost more computation time)

With a trade-off between the two factors outlined above, the value of n has empirically been found to be 5. This possibly happens due to the presence of partial occlusion or segmentation error (where height or width is not visible) persisting through consecutive 4 frames. Through this understanding, the size of padding can be modelled based on the nature of the database (and the efficiency of background subtraction applied on it) on which the method is supposed to work.

Thus it can be well justified why an aggregate feature vector with padding (*Case I* shown above) will bear better directional discrimination ability than a feature vector constructed without padding (*Case II* shown above). Hence we have chosen a padded aggregate feature for proposed modelling.

1) PROBLEM MODELLING THROUGH HIDDEN MARKOV MODEL

The time varying changes in the dimension of bounding box follows dependencies from the previous frame. i.e., $P(q_{t+1}|q_t, y_t) = P(q_{t+1}|q_t)$ where, q_{t+1} and q_t are the dimensions at time t and $(t+1)$ respectively and y_t being direction of motion to reach current dimensions.

The problem thus satisfies the Markovian property and this is the reason Hidden Markov Model is used for training the model for estimation of direction of motion. To model a problem to HMM, their observation states, hidden states, transition and emission probabilities are needed to be defined. Following are the definitions with respect to the proposed method.

Observed State or visible state (denoted by $V(t)$): are the visible features that is accessible from an event sequence to be modelled. A sequence of observed state forms **observed state sequence** denoted as

$$V_1^t = \{V_1, \cdots\cdots, V_t\}$$

In the proposed method, observable feature sequence is defined as change in dimensions of the bounding box surrounding the foreground blob. As stated before, the observed state sequence is formed as $\Delta h|_t + pd + \Delta w|_t$.

While the size of pd is already discussed to be 5 units, $\Delta h|_t$ and $\Delta w|_t$ are constructed with $m\ (= f \times t)$ units (where, f is the number of frames elapsed per second in the video). In our implementation, we consider change of terrains within a time gap of $t = 1s$ and the video is considered to move with 30 FPS. Hence, $\Delta h|_t$ and $\Delta w|_t$ are both $1 \times 30 = 30$ units long in size. Thus, the aggregated feature size turns out to be $(30 + 5 + 30) = 65$ units.

Since the discrete directions of motion are mirror image of each other along the view axis, the pattern of change of dimensions for pair of directions 2 and 8, 3 and 7, and 4 and 6 are identical in camera projection, however their direction of motion is exactly opposite. To differentiate this a padding of 0 or 1 denoting rightwards or leftwards movement is introduced. Many of the existing researches are producing erroneous results while handling opposite pair of directions reported in the survey of Section I.

Hidden State (denoted by $\omega(t)$): are the states not observed directly rather needs interpretation from observed sequence. The perceiver does not have access to the hidden state, instead algorithm measures some property of the observed state to infer hidden state. In the proposed method 4 hidden states are taken due to sinusoidal nature observed at the selected feature vector. As the articles suggested [41], [42], brute force selection of number of hidden states near the estimated proximal value of number of hidden states, the simulation of the HMM are also performed with 3 and 5 hidden states however the training lasted for 100-120 iterations as compared to 50-60 iterations in case of 4 hidden states. This empirically justifies the selection of number of hidden states.

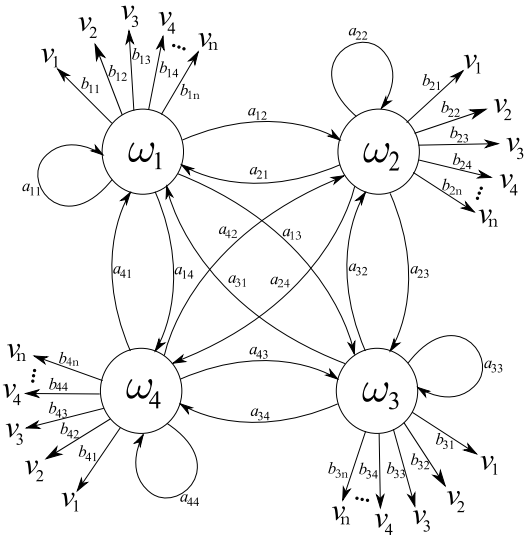


FIGURE 7. State transition diagram of the HMM of proposed method.

2) TRANSITION AND EMISSION PROBABILITIES

As depicted in Fig. 7, the transition and emission probabilities are defined as: Transition among hidden states, i.e., $a_{ij} : P(\omega_j(t+1)|\omega_i(t))$ Emission of a visible state, i.e., $b_{jk} : P(v_k(t)|\omega_j(t))$ with limiting conditions as, $\sum_j a_{ij} = 1 \ \forall i$ and $\sum_k b_{jk} = 1 \ \forall j$. Projected problem is a learning problem of HMM. The problem states that given a set of observed state sequence V^T and any hidden state as given by $\omega(t)$, the task is to determine the probabilities a_{ij} and b_{jk} using forward backward algorithm. We start with the above defined initial arbitrary values of a_{ij} and b_{jk} and find more accurate values of a_{ij} and b_{jk} at the end of Baum-Welch or forward-backward algorithm as illustrated below.

The probability that the model produces a sequence V^T of visible states is

$$P\left(V^T\right) = \sum_{r=1}^{r_{max}} P\left(V^T|\omega_r^T\right) P(\omega_r^T) \tag{1}$$

where each r indexes a particular sequence $\omega_r^T = \{\omega(1), \omega(2), \ldots, \omega(t)\}$ of T hidden states. In the general case of c hidden states, there will be $r_{max} = c^T$ possible terms in the sum of Eq. (1), with respect to all possible sequence of length T.

As we are dealing with first-order Markov process, the factors in Eq. (1) can be written as Eq. (2) and Eq. (3).

$$P(\omega_r^T) = \prod_{t=1}^{T} P(\omega(t)|\omega(t-1)) \tag{2}$$

$$P(V^T|\omega_r^T) = \prod_{t=1}^{T} P(v(t)|\omega(t)) \tag{3}$$

Combining the results of Eq. (2) and Eq. (3), previously described Eq. (1) can be rewritten as Eq. (4).

$$P(V^T) = \sum_{r=1}^{r_{max}} \prod_{t=1}^{T} P(v(t)|\omega(t)) P(\omega(t)|\omega(t-1)) \tag{4}$$

We denote our model - the a's and b's - by θ and using Bayes formula, probability of the model given observed sequence is given by Eq. (5).

$$P(\theta \mid V^T) = \frac{P(V^T \mid \theta)\, P(\theta)}{P(V^T)} \tag{5}$$

Now, $\alpha_j(t)$ and $\beta_i(t)$ can be defined as shown in Eq. (6) and Eq. (7).

$$\alpha_j(t) = \begin{cases} 0 & t = 0 \text{ and } j \neq \text{ initial state} \\ 1 & t = 0 \text{ and } j = \text{ initial state} \\ \sum_i \alpha_i(t-1) a_{ij}] b_{jk} v(t) & \text{otherwise} \end{cases} \tag{6}$$

$$\beta_i(t) = \begin{cases} 0 & \omega_i(t) \neq \omega_0 \text{ and } t = T \\ 1 & \omega_i(t) = \omega_0 \text{ and } t = T \\ \sum_j \beta_j(t+1) a_{ij} b_{jk} v(t+1) & \text{otherwise} \end{cases} \tag{7}$$

Where, $\alpha_j(t)$ represents the probability that the model is in hidden state $\omega_j(t)$ having generated first t elements of V^T and $\beta_i(t)$ represents the probability that the model is in hidden state $\omega_i(t)$ and will generate rest of the target sequence from $(t+1)$ to T. The notation $b_{jk}v(t)$ (as mentioned in Eq. (6)), and $b_{jk}v(t+1)$ (as mentioned in Eq. (7)) denotes the transition probabilities b_{jk} selected by visible state emitted at time t and $(t+1)$ respectively. However, this way of determining $\alpha_j(t)$ and $\beta_i(t)$ are mere estimates of their true values, as we do not know the actual values of a_{ij} and b_{jk} in Eq. (6) and Eq. (7). We can calculate improved values of $\alpha_j(t)$ and $\beta_i(t)$ by defining $\gamma_{ij}(t)$ (shown in Eq. (8)) which is the probability of transition between $\omega_i(t-1)$ and $\omega_j(t)$, given the model generated the entire training sequence V^T by any path.

$$\gamma_{ij}(t) = \frac{\alpha_i(t-1) a_{ij} b_{jk} \beta_j(t)}{P(V^T \mid \theta)} \tag{8}$$

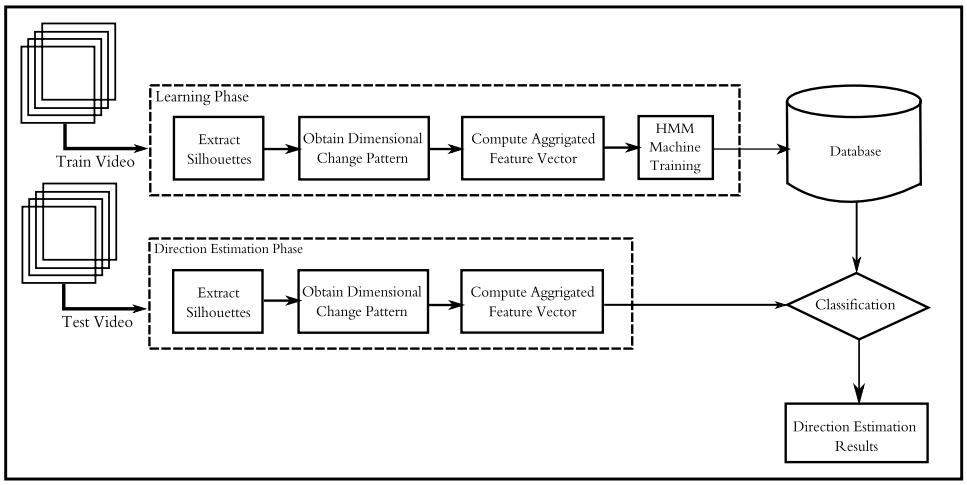


FIGURE 8. Block diagram of the proposed method.

Hence we find an improved estimation of a_{ij} and b_{jk} as $\hat{a}_{ij}$ and $\hat{b}_{jk}$ through Eq. (9) and Eq. (10) respectively.

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} \gamma_{ij}(t)}{\sum_{t=1}^{T} \sum_{k} \gamma_{ik}(t)} \tag{9}$$

$$\hat{b}_{jk} = \frac{\sum_{t=1, v(t)=v_k}^{T} \sum_{l} \gamma_{jl}(t)}{\sum_{t=1}^{T} \sum -l\gamma_{jl}(t)} \tag{10}$$

3) PROBABILITY OF THE MODEL

If we denote the *a's* and *b's* of our model by θ and use Bayes formula, **probability of the model given observed sequence** is

$$P(\theta \mid O_1^t) = \frac{P(O_1^t \mid \theta)\, P(\theta)}{P(O_1^t)} \tag{11}$$

where, *a's* and *b's* of HMM of the proposed model is denoted by θ.

In this way 8 HMMs for each discrete direction are modeled. The successful supervised training of models are depicted by log likelihood graphs as shown further in Figs. 10, 11 and 12. Further, test sequences are classified with highest probability and above a minimal threshold. The highest probability classifies the direction while minimal threshold segregates human from non-human based on motion patterns. Formal definition and explanation of HMM in the context of proposed model is discussed among the evaluation parameters in Section III.[1]

[1]For general understanding of HMM, and its implementation in time sequential image data, readers may refer to Rabiner and Juang [43] and Yamato *et al.* [44].

C. MACHINE LEARNING AND CLASSIFICATION

The proposed method of classification through HMM comprises two phase viz. learning phase and direction estimation phase. In the learning phase training video samples undergo various steps to obtain feature vectors for each video that uniquely defines respective classes. Hence different HMMs are formed each representing a unique direction. Log likelihood graphs are used to represent their proper learning. Section III-C defines log likelihood graph as an evaluation parameter and log likelihood graphs related to performed experiments are presented in Section III-E. Further, in the direction estimation phase, test videos are taken, their feature vectors are extracted and are classified among different classes as defined in the learning phase, thus resulting into an estimated direction for each subjects testing samples. From both the subsets (training as well as testing video) of database; steps of extracting silhouette of pedestrian, temporal change in dimension of their bounding box and their feature extract are performed. The set of training feature vectors for each of the direction classes are kept in the database. Newly arriving test videos undergoes similar procedure for its feature extraction and are subjected to be classified among any of the direction classes. Fig. 8 presents a block diagram depicting overall description of the proposed method.

III. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method for direction estimation is experimentally justified. This section elaborates the experiment and related experimental environment prepared for conducting those experiments. This includes:

- Constraints and assumptions in the experiment
- Details about various databases used
- Evaluation parameters
- Experiments conducted for direction estimation
- Results
- Analysis

A. CONSTRAINTS AND ASSUMPTIONS IN THE EXPERIMENT

The experiments are conducted in a constrained environment with certain assumptions. These constraints depicts the robustness as well as environmental limitations of the proposed method. They are presented as follows.

- The proposed method is constraint to work over binary image sequences with human subject as foreground silhouettes
- Preprocessing to achieve the same is not elaborated and assumed to be minimal
- Sudden and frequent changes in the direction of walk are not assumed in the experiment
- The proposed method is constraint to work with static field camera
- The method is constraint to work in visible spectrum
- The proposed method is not robust to full and longer exposure to occlusion, such cases are not considered in the experiment
- The method is proposed to handle direction estimation of multiple human subjects present in a scene assuming no mutual occlusion however experiments demonstrated here have cases with only single pedestrian in the scene

B. VARIOUS DATABASES USED

With the limitations of existing databases with respect to our proposed experimental requirements, three different databases are used. They are CASIA dataset A and dataset B [45] and NITR conscious walk dataset [46].

These datasets are used with an intention to bring many scenarios in the purview of our proposed method. Subjects involved as pedestrian in the experiments have diversified cases as:

- Participating pedestrians are both female and male, and have diverse physical build, age group and ethnicity
- The velocity of walk within a video footage is not necessarily constant
- The velocity of walk in different video footage may be different
- Capturing environment and walking surfaces are different
- Pedestrians are not strictly following a direction during their walk
- Pedestrians have diverse carrying and clothing conditions that affects the pattern of temporal changes in the dimensions of silhouette
- Depth of the pedestrian with respect to field camera is varying from video to video
- In few cases pedestrians are under partial static occlusion
- There are some cases with different orientation of head and body of a pedestrian to that of actual direction of their motion

To further clarify the need of different databases used in the experiment, their parametric properties and special cases available in the database are presented in Table 2. Selected frames from different databases to support the diversity of special cases are presented in Fig. 9.

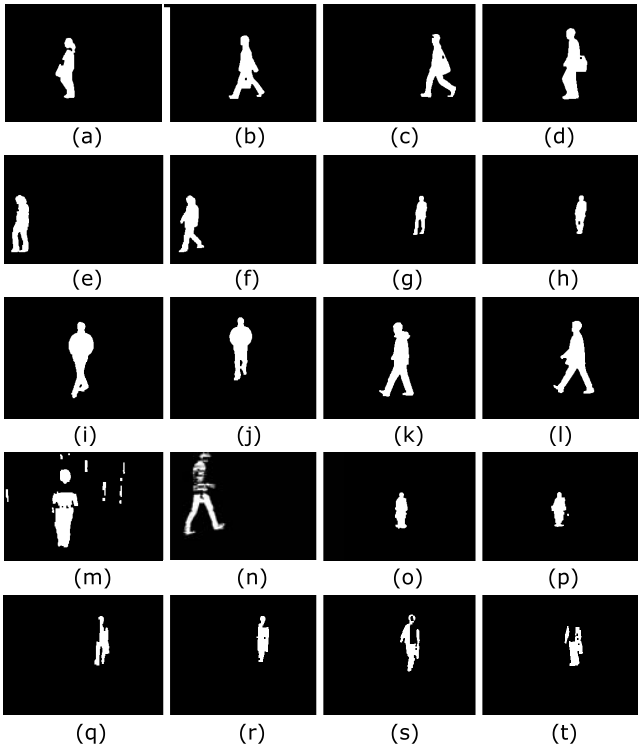


FIGURE 9. Frame sequences depicting the diverse cases with pedestrian carrying bag (a, b, c, d), having different head orientation during a walk (e, f, g, h), walk with hands in the pockets (i, j), wearing coat (k, l), improper segmentations (m, n, o, p), and partial static occlusion with incomplete width or height information (q, r, s, t). Pedestrian frames are from CASIA dataset A, CASIA dataset B [45] and NITR conscious walk dataset [46] and have diverse physical build, walking velocity, age group, ethnicity and gender. Respective frame locations in the database available in .png format are: (a) CASIA/Gait Dataset B/012-bg-02-090-073 (b) CASIA/Gait Dataset B/007-bg-02-090-070 (c) CASIA/Gait Dataset B/124-bg-01-090-043 (d) CASIA/Gait Dataset B/001-bg-01-090-057 (e) CASIA/Gait Dataset A/fyc-45_1-092 (f) CASIA/Gait Dataset A/fyc-45_1-085 (g) CASIA/Gait Dataset A/fyc-45_1-014 (h) CASIA/Gait Dataset A/fyc-45_1-011 (i) CASIA/Gait Dataset B/032-cl-02-000-069 (j) CASIA/Gait Dataset B/032-cl-02-000-059 (k) CASIA/Gait Dataset B/001-cl-02-090-059 (l) CASIA/Gait Dataset B/020-cl-02-090-060 (m) CASIA/Gait Dataset A/syj-90_3-089 (n) NITR conscious walk db/1001D2S1F065 (o) CASIA/Gait Dataset A/fyc-90_1-002 (p) CASIA/Gait Dataset A/fyc-90_1-006 (q) CASIA/Gait Dataset B/002-bg-01-018-042 (r) CASIA/Gait Dataset B/002-bg-01-018-034 (s) CASIA/Gait Dataset B/002-bg-01-018-059 (t) CASIA/Gait Dataset B/002-bg-01-018-050.

C. EVALUATION PARAMETERS

- **Log Likelihood Graph**: Log Likelihood Graph in a supervised learning depicts the Log likelihood (on Y axis) of a learning model with the increase of expectation maximization iterations (on X axis). In the context of the proposed 4-state Hidden Markov Model based training, corresponding plots in different log likelihood graph shows monotonic increase, followed by convergence. This depicts the successful training of all the HMMs ending up with maximum likelihood within

limited iterations and hence the suitability of the parameter set to form aggregated features for training.

- **Balanced Accuracy**: Accuracy can be defined as proportion of total number of correct predictions. The formula for accuracy is given by

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

Where,

TP: True Positive
FP: False Positive
TN: True Negative
FN: False Negative

However, in multi-class classification, balanced accuracy is used and is given by the arithmetic mean of class specific accuracies.

$$\text{Balanced Accuracy} = \frac{\sum_{all\ classes} \text{Accuracy}}{\text{Number of classes}} \tag{13}$$

- **Recall**: Recall (also known as TP Rate, True Positive Rate, or Sensitivity) can be defined as the proportion of positive cases that were correctly identified and is given by

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

In multi-class classification, as in the proposed case the corresponding information is given by average recall gives, which is arithmetic mean of all positive cases that were correctly identified.

$$\text{Average Recall} = \frac{\sum_{all\ classes} \text{Recall}}{\text{Number of classes}} \tag{15}$$

- **Precision**: Precision can be defined as the predicted positive cases that were correct and is given by

$$\text{Precision} = \frac{TP}{TP + FP} \tag{16}$$

In the proposed multi-class classification, precision of classification can be defined with a more suitable parameter called average precision which is given by:

$$\text{Average Precision} = \frac{\sum_{all\ classes} \text{Precision}}{\text{Number of classes}} \tag{17}$$

- ***F*-Measure**: *F*-Measure is the harmonic mean of precision and recall, and it is a measure to judge the accuracy of a classifier. In the harmonic mean, when equal weight is given to recall and precision, it is more precisely called as F_1-Measure.

$$F_1 - \text{Measure} = \frac{2 \times \text{Precision} \times \text{recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

In the proposed multi-class classification, F_1-Measure of classification can be defined with a more suitable parameter called average F_1-Measure which is given by:

$$\text{Average } F_1 - \text{Measure} = \frac{\sum_{all\ classes} F - \text{Measure}}{\text{Number of classes}} \tag{19}$$

- **False Positive Rate**: FP Rate can be defined as the proportion of positive cases that were incorrectly identified and is given by

$$\text{FP Rate} = \frac{FP}{FP + TN} \tag{20}$$

In multi-class classification, as in the proposed scenario, a more suitable parameter called average FP Rate yields the corresponding information.which represents the average of all false rejections and is given by:

$$\text{Average FP Rate} = \frac{\sum_{all\ classes} \text{FP Rate}}{\text{Number of classes}} \tag{21}$$

- **Error Rate**: Error rate is given by (1 − Accuracy). In the proposed multi-class classification, error rate is presented as percentage of incorrectly classified instances and is given by the average of misclassification for each individual classes.

$$\text{Average Error Rate} = \frac{\sum_{all\ classes} \text{Error Rate}}{\text{Number of classes}} \tag{22}$$

- **Confusion Matrix**: In supervised learning, a confusion matrix or an error matrix is a tool of statistical classification that lets the visualization of mislabelling of classification data in the form of *false positive* and *false negative*, while the correct labelling are present in the forms of *true positive* and *true negative*.
- ***V*-Fold Cross Validation**: In such validation method, database is randomly divided into *V* equal sized samples and each time any one of the sample is utilised as testing sample while all other samples are utilised as training sample. This technique of model verification also proves a dataset to be unbiased.

D. EXPERIMENTS CONDUCTED FOR DIRECTION ESTIMATION

The proposed method is validated with conducted experiments. As already shown in Fig 2, the temporal pattern of width and height of the bounding box of randomly selected video having directions of walk as 4, 5, 6 and 7 (refer to Fig. 1) are plotted. Study of the graph reveals following observations:

- The temporal changes in the dimensions of bounding box are unique with respect to each direction
- Field camera follow perspective geometry, hence the temporal patterns shows perspective affected scaling distortion
- Height of the pedestrian remains constant as it moves orthogonal to the view axis (i.e. in direction 3 or 7).

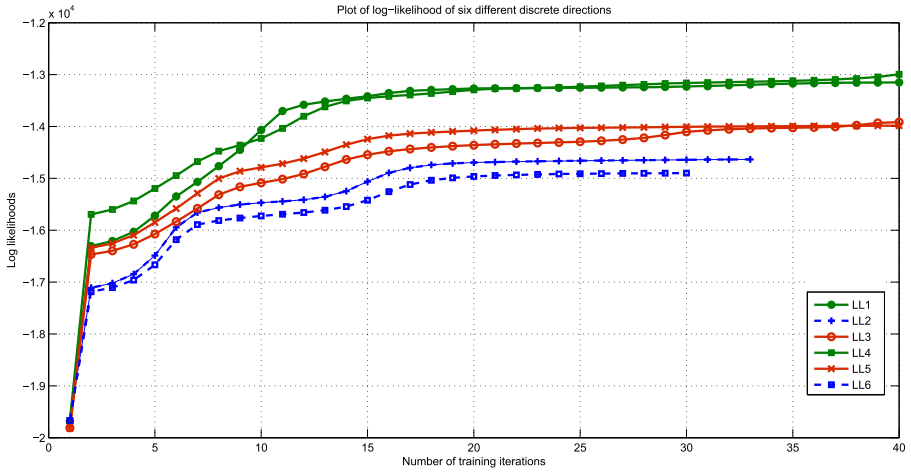


FIGURE 10. Log Likelihood graph presenting machine learning of 6 discrete directions using HMM over CASIA Dataset A.

This is due to the fact that depth of pedestrian remains constant with respect to field camera

- Temporal change in width of the pedestrian in the same pair of directions shows sinusoidal nature; this is due to cyclic motion of limbs during a gait cycle
- The temporal increment in the dimension of pedestrian's bounding box is maximal when it directly approaches towards the camera (i.e. direction 5) and minimal in the opposite case (i.e. direction 1)
- The temporal changes along the direction 4 and 6 were expected to show moderate increment in dimensions

The temporal pattern of change in the dimensions of a pedestrian moving along the direction 8, 1, 2, and 3 are mirror images of directions 5, 4, 6, and 7 respectively along x axis. They have not been included in order to avoid clumsiness in the graph.

All the results discussed so far are consistent with the anticipated pattern. With this motivation, further experiments are conducted over different databases (refer to Table 2) that are discussed in subsequent paragraphs.

Experiment # 1 is conducted over CASIA Dataset A. This database has overall 240 binary frame sequences where pedestrians are moving along 0°, 45°, and 90° with respect to view axis. This covers 6 of the 8 discrete directions that are discussed in the proposed method. The missing two directions (i.e. direction 4 and 8), are mirror images of directions 6 and 2 respectively and shows similar temporal pattern due to similar perspective distortion. The direction of pedestrian motion in this experiment are classified among 6 classes. Each of the directions have 40 frame sequences in the database. Further, 10 fold division of the database is created in such a way that each fold contains equal number of randomly chosen samples from all the walk directions. All the experiments are conducted with 9 folds for training and 1 fold for testing the result. These experiments are conducted 10 times with each fold tested exactly once. The random selection of samples uses entire database for training as well as testing, supports the unbiased nature of database.

Fig. 10 presents a sample case where different convergent plots of log likelihood graph presents HMM based training of different classes over CASIA Dataset A. Related experimental environment is summarised in Table 4. Results related to this experiment are discussed in the Section III-E. CASIA Dataset A however does not contain enough discrete directions.

TABLE 3. Merging 11 different walk directions of CASIA Dataset B into five discrete directions.

Direction	Angle from view axis
Direction 1	{0°, 18°}
Direction 2	{36°, 54°}
Direction 3	{72°, 90°, 108°}
Direction 4	{126°, 144°}
Direction 5	{162°, 180°}

Experiment # 2 is conducted over CASIA Dataset B. This dataset has overall 13640 binary frame sequences with 11 different walk directions. The dataset contains various carrying and clothing variations since the pedestrians are carrying bag, wearing coat or having a regular outfit without carrying anything.Walk directions are ranging 0° to 180° from view axis of camera instead of varying 0° to 360° hence these directions are needed to be classified among 5 of the 8 discrete direction classes. Table 3 shows merging of the directions. Direction 3 which clubs 3 different walking cases has 3720 walk samples while rest of the directions have 2420 walk samples. However, for uniform training and testing, 2420 randomly selected samples are considered for direction 3 making an overall 12100 walk samples available for experiment. These samples further undergoes 10 fold cross validation where each fold are tested while other 9 folds are used for training the model. Fig. 11 presents a sample case where different convergent plots represents HMM based training over CASIA Dataset B. Related experimental setup

TABLE 4. Experimental environment for Experiment# 1.

Experiment # 1	
Database Used	CASIA Dataset A
Discrete Directions Covered	6 i.e. directions 1,2,3,5,6, and 7 (Refer to Fig. 1)
Applied Training Mechanism	HMM with 6 classes
Database Set-up	10 fold division. Equal distribution of each direction in each fold. Random selection for each folds
Experiment Iterations	10 times, all results presented are averaged over 10 iterations
Balanced accuracy achieved	94.58%

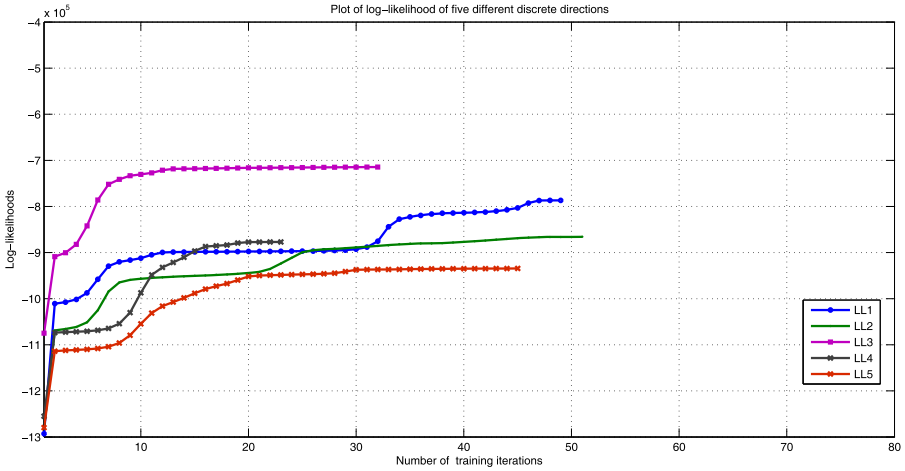


FIGURE 11. Log Likelihood graph presenting machine learning of 5 discrete directions using HMM over CASIA Dataset B.

TABLE 5. Experimental environment for Experiment# 2.

Experiment # 2	
Database Used	CASIA Dataset B
Discrete Directions Covered	5 i.e. directions 1,2,3,4, and 5 (Refer to Fig. 1)
Applied Training Mechanism	HMM with 5 classes
Database Set-up	10 fold division. Equal distribution of each direction in each fold. Random selection for each folds
Experiment Iterations	10 times, all results presented are averaged over 10 iterations
Balanced accuracy achieved	90.87%

TABLE 6. Experimental environment for Experiment# 3.

Experiment # 3	
Database Used	NITRKL Conscious Walk Database
Discrete Directions Covered	8 i.e. directions 1,2,3,4,5,6,7, and 8 (Refer to Fig. 1)
Applied Training Mechanism	HMM with 8 classes
Database Set-up	10 fold division. Equal distribution of each direction in each fold. Random selection for each folds
Experiment Iterations	10 times, all results presented are averaged over 10 iterations
Balanced accuracy achieved	95.83%

is summarised in Table 5 and related results are discussed in Section III-E.

So far, the classification over existing database samples are limited to 5 and 6 different directions. To classify among all the proposed 8 discrete directions, a new outdoor database is acquired.

Experiment # 3 is conducted with NITR Conscious Walk Database. This database consists of 21 subjects contributing 3 sample walks in each of the 8 directions spanning from [0°-360°]. The pedestrians are consciously made to walk in the presumed discrete directions with a few little deviations, so as to have better machine training. 10 fold cross validation is again performed over this dataset. Fig. 12 presents a sample case where different convergent plots represents HMM based training over NITR Conscious Walk Dataset. Experimental environment of the same is presented in Table 6. Related results are discussed in the Section III-E.

All the experiments are performed using MATLAB Simulink software over workstation having Intel Xeon processor with duel processing core where each core has a clock speed of 2.4 *GHz*. The system works on a 64-bit operating system and it has 8 *GB* of volatile memory. The detailed classification results over different databases are presented in Section III-E.

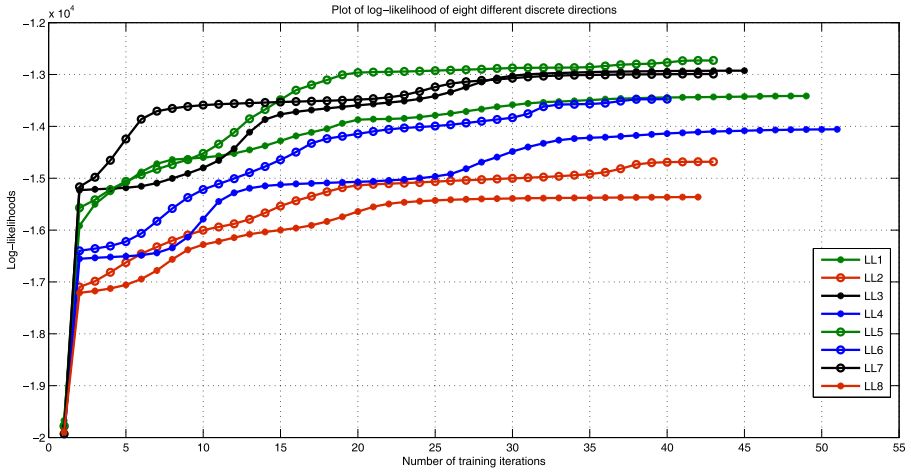


FIGURE 12. Log Likelihood graph presenting machine learning of 8 discrete directions using HMM over NITR Conscious Walk Dataset.

TABLE 7. Confusion matrix for casia dataset A.

Directions	D_1	D_2	D_3	D_5	D_6	D_7	Accuracy(%)
D_1	36	4	0	0	0	0	90
D_2	1	37	2	0	0	0	92.5
D_3	0	0	40	0	0	0	100
D_5	0	0	0	38	2	0	95
D_6	0	0	0	1	37	2	92.5
D_7	0	0	0	0	1	39	97.5
					Balanced	**Accuracy**	**94.58**

TABLE 8. Confusion matrix for Casia dataset B

Directions	D_1	D_2	D_3	D_4	D_5	Accuracy(%)
D_1	2252	142	14	9	3	93.06
D_2	121	2202	69	21	7	90.99
D_3	15	140	2129	122	14	87.97
D_4	9	23	118	2172	98	89.75
D_5	4	7	21	142	2241	92.60
				Balanced	**Accuracy**	**90.87**

TABLE 9. Confusion matrix for NITR conscious walk dataset.

Directions	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	Accuracy(%)
D_1	61	2	0	0	0	0	0	0	96.83
D_2	2	59	2	0	0	0	0	0	93.65
D_3	0	0	62	1	0	0	0	0	98.41
D_4	0	0	1	60	2	0	0	0	95.24
D_5	0	0	0	2	61	0	0	0	96.83
D_6	0	0	0	0	1	59	3	0	93.65
D_7	0	0	0	0	0	0	63	0	100
D_8	1	0	0	0	0	0	4	58	92.06
							Balanced	**Accuracy**	**95.83**

E. RESULTS

This section presents detailed evaluation of the proposed method and its comparison with state-of-the-art and few recent research in the domain. This results are further elaborated in following sub-sections:

1) QUANTITATIVE EVALUATION AND COMPARISON

The quantitative experimental results of the proposed method over different databases and its quantitative comparison with some existing work are presented in this subsection. The frames are rectified for feature extraction in the pre-processing stage, as already presented earlier in Fig.3. Tables 4, 5, and 6 summarises experimental environment. Results related to experiment # 1, experiment #2 and enperiment #3 are presented in the form of confusion matrix in Tables 7, 8, and 9, respectively. The direction estimation accuracy for CASIA Dataset A using a 10 fold cross validation with 9:1 training to testing ratio for 6 directions are found in the range of 90% to 100%, with an average balanced accuracy of 94.58%.

The direction estimation accuracy for the experiment conducted over CASIA Dataset B with 10-fold cross validation

TABLE 10. Summary of evaluation results for direction estimation.

Database Name	Balanced Accuracy (%)	Precision (%)	Recall (%)	F_1 Measure (%)	Error Rate (%)	False Positive Rate (%)
CASIA dataset A	94.58	94.32	94.58	94.45	5.42	1.22
CASIA dataset B	90.87	94.84	90.87	92.81	9.13	1.91
NITRKL conscious walk dataset	95.83	95.99	95.83	95.91	4.17	0.63

TABLE 11. Quantitative comparison of direction estimation results over different database.

	CASIA Dataset A		CASIA Dataset B		NITR Conscious Walk Dataset	
	Balanced Accuracy (%)	False Positive (%)	Balanced Accuracy (%)	False Positive (%)	Balanced Accuracy (%)	False Positive (%)
Andriluka *et al.* (2010) [20]	61.91	2.41	59.73	3.91	64.72	1.84
Chen *et al.* (2011) [21]	66.42	2.13	64.67	3.30	72.45	1.65
Baltieri *et al.* (2012) [22]	77.67	1.86	72.50	2.65	79.18	1.20
Raman *et al.* (2012) [2]	82.28	1.94	80.46	3.04	84.17	1.41
Liu and Ma (2015) [23]	90.32	1.73	84.61	2.72	91.37	1.13
Proposed Method	94.58	1.22	90.87	1.91	95.83	0.63

with 5 discrete directions is found to be in the range of 89% to 94%, with an average balanced accuracy of 90.87%. This drop in the balanced accuracy is due to the merging of 11 intermediate directions of walk into 5 discrete directions. However, an overall balanced accuracy over such a large dataset with several walk directions justifies the robustness of the proposed method.

These dataset does not possess all the 8 discrete directions to be classified among. Hence experiment 3 is conducted covering all the 8 proposed discrete directions. The direction estimation accuracy for the experiment conducted over NITR Conscious Walk Dataset with 10 fold cross validation with 8 directions are found to be in the range of 92 to 100% with an average balanced accuracy of 95.83%. In multi-class classification over different directions, results with different parameters of evaluation i.e. *Balanced Accuracy, Precision, Recall,* F_1 *Measure, Error Rate* and *False Positive Rate* are presented in Table 10.

The proposed method is further compared with a few parallel researches with common experimental platform. The methods compared in this section, works on video dataset captured from field camera. All the methods are trained and tested over 3 different datasets (as discussed in Table 2) and uniformly underwent 10 fold cross validation for fair comparison. Their comparison results with two evaluation parameters: *Balanced Accuracy* and *False Positive Rate* is presented in Table 11. The quantitative comparison of the proposed method with some existing research in the domain shows that the proposed method out performs these existing methods with better *Balanced Accuracy* and *False Positive Rate*.

2) COMPARISON OF INTRINSIC PROPERTIES

This subsection compares intrinsic properties of the proposed method from some of the existing method in the domain that witnesses a vast diversity towards approaching estimation of pedestrian direction. These diversities are mainly due to different research requirements. Some of them are view angles of the recording camera (top camera, field camera), camera position (stationary, moving camera), extra sensor based requirement (infra-red sensor, depth sensor, monocular, binocular or multi-vision based camera), different data input requirement (frame based and video based) and different objectives (traffic, surveillance, home assistance). Due to this, a direct comparison from many other research is not feasible. However, we have attempted to compare their intrinsic properties in this subsection.

The proposed method can handle the situation of static partial occlusion since temporal patterns of both width and height contribute to the aggregate feature and at least one of the patterns remain available during partial occlusion. Fig. 9 shows a few example cases available in the dataset where only height or width information of the pedestrian blob is available in the frame and such cases are handled.

Proposed method uses motion feature over the bounding box of the pedestrian blob and hence even if the head or body is oriented in other direction as that of actual direction of motion, the overall direction estimation is unaffected. The head and body orientation based direction estimation methods [14], [15], [17], have reported to get affected in such scenario.

Motion feature can capture the human gait motion and are used for human detection [19], [47]–[49]. The set of feature selected in the proposed method for pedestrian direction estimation segregates human from a non-human under motion. This is due to the properties of the feature, like unique human gait pattern and height to width ratio of human, that it is segregated from vehicles, animals and other moving objects in the scene. Being less than a matching threshold, a non-human blob is not classified in any of the 8 direction classes;

TABLE 12. Comparison of intrnsic properties of proposed method with some existing research on direction estimation.

	Number of discrete Directions	Correct estimation of opposite pairs	Handling static partial occlusion	Handling different head orientation	Detection and direction estimation of pedestrian motion	Orientation update per frame	Camera position
Andriluka *et al.* (2010) [20]	8		√		√	√	Field Camera
Goto *et al.* (2011) [26]	4		√	√	√	√	Field Camera
Chen *et al.* (2011) [21]	8	√		√		√	Field Camera
Raman *et al.* (2012) [2]	8	√		√			Field Camera
Guangzhe *et al.* (2012) [14]	8					√	Car Camera
Baltieri *et al.* (2012) [22]	8		√			√	Field Camera
Tao and Klette (2013) [27]	4			√	√	√	Field Camera
Liu *et al.* (2013) [18]	8	√	√	√		√	RGB-D Field Camera
Flohr *et al.* (2014) [15]	8	√			√		Car Camera
Bensebaa *et al.* (2015) [17]	24	√	√			√	Field Camera
Liu and Ma (2015) [23]	8	√		√		√	Field Camera
Proposed Method	8	√	√	√	√		Field Camera

hence the direction estimation is performed only over those moving blobs which are identified as a human, based on their motion feature. Goel and Chen [19] have also attempted to perform human detection based on their motion feature using SVM and GLMPC over synthesized CASIA Dataset A claiming an accuracy of 97.4%, however, they have classified the directions only among 3 direction classes i.e. front, left and right (direction 3, 5, and 7). On the similar experimental conditions the proposed method is found to achieve estimation accuracy of 97.75% with all the 8 discrete direction classes.

In the proposed method, exploiting motion feature gives a robust direction estimation result yet a sudden change in the direction of motion takes few frames to update and the overall result updates in such conditions are slower. This slowness in the update is comparable to other existing methods [14], [15], [26], [27], which generates direction estimation results by statistical combination of different orientation results over a few frames. However, such methods gives an orientation estimates in each frame and regards it as a probable direction of motion, which our proposed method does not.

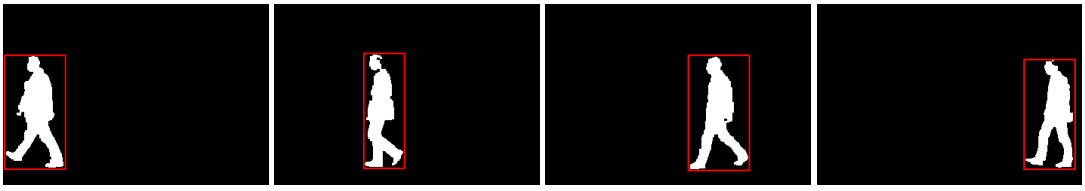
Due to large deviation in the environmental constraints, experimental set-up, and database types, a quantitative comparison of the proposed method with some landmark researches over a common platform is not feasible; however, an attempt has been made to compare the intrinsic features of the proposed method with few existing landmark researches across the platform. Table 12 presents the same. The table gives a fair idea about the robustness of the proposed model over existing researches and about its intrinsic features.

F. ANALYSIS

This subsection evaluates the claims of the proposed method with different results and comparisons. The quantitative results compared over common platform justifies the robustness of the proposed method over other researches. The result convincingly shows the proposed method to be working

#Frame	120	90	60	30
D_{gt}	D_6	D_6	D_6	D_6
D_{est}	D_6	D_6	D_6	D_6

(a)



#Frame	60	45	30	15
D_{gt}	D_7	D_7	D_7	D_7
D_{est}	D_7	D_7	D_7	D_7

(b)

FIGURE 13. Qualitative results with successful direction estimates. (a) Sample frame sequence with successful direction estimation by proposed method as direction 6. (b) Sample frame sequence with successful direction estimation by proposed method as direction 7.

well even when the cases of static partial occlusion, varying head orientation, diverse clothing and carrying condition, and improper foreground segmentation are present. The proposed method out performs some of the existing methods in the domain with better *Balanced Accuracy* and *False Positive Rate*.

Fig. 13 presents two frame sequences from CASIA dataset A for presentation of qualitative result of direction estimation by the proposed method. In Fig. 13, D_{gt} represents the ground truth direction and D_{est} represents estimated direction by the proposed method. The results on two different pedestrian videos (with frame gap of 30 and 15 frames) are shown respectively in Figs. 13(a) and 13(b), depicting the correct direction estimation at each frame in both the sequences. The result also shows the correct estimation of direction with different apparent velocity as both the frame sequences have different apparent velocity due to different depth from camera in two frame sequences. With different experimental results and comparisons, the accuracy of direction estimation and its intrinsic features are presented. The proposed method is robust since it uses motion features, on the other hand methods that performs direction estimation using orientation information gives an additional information about the orientation of pedestrian in each frame, that the proposed method does not. The proposed method can be used as a standalone system or can be integrated with orientation estimation based methods to produce faster yet accurate direction estimation results along with orientation information.

IV. CONCLUSION

Proposed method is a motion feature based direction estimation method. Due to motion feature, the proposed method is robust to partial occlusion, tolerable segmentation errors, and different head and body orientation. The method can estimate the direction of human motion through his gait patterns.

The proposed method finds its usage in the domain of traffic safety and management, visual surveillance in smart cities, assisted living in smart homes and human computer interaction. Its potential usage in different aspects towards development of smart cities motivates to take the research further for more complicated and challenging environments like shopping malls, subways, and railway station to explore emerging issues like crowd behaviour analysis for business intelligence, monitoring, and surveillance.

ABBREVIATIONS

3DPeS Dataset	3D People Surveillance Dataset
CASIA Dataset	Institute of Automation Chinese Academy of Sciences Dataset
DBNS	Dynamic Bayesian Network System
FIND	Feature Interaction Detector
GLMPC	Global Locale Motion Pattern Classification
HMM	Hidden Markov Model
HoG	Histogram of Gradient
INRIA Dataset	Institut National de Recherche en Informatique et en Automatique Dataset
NITR Dataset	National Institute of Technology Rourkela Dataset
NN	Neural Network
PETS Dataset	Performance Evaluation of Tracking and Surveillance Dataset
PKU Dataset	Peking University Dataset

RGB-D Sensors	Red Green Blue and Depth Sensor
RMD	Reliable Motion Direction
SIFT	Scale Invariant Feature Transform
SVM	Support Vector Machine
TUD Dataset	Technische Universitat Darmstadt Dataset
ViBe	Visual Background Extractor.

REFERENCES

[1] R. Raman, P. K. Sa, S. Bakshi, and B. Majhi, "Towards optimized placement of cameras for gait pattern recognition," *Procedia Technol.*, vol. 6, pp. 1019–1025, 2012.

[2] R. Raman, P. K. Sa, and B. Majhi, "Occlusion prediction algorithms for multi-camera network," in *Proc. IEEE/ACM Int. Conf. Distrib. Smart Cameras*, Oct./Nov. 2012, pp. 1–6.

[3] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, pp. 1–45, 2006.

[4] B. Y. Lee, L. H. Liew, W. S. Cheah, and Y. C. Wang, "Occlusion handling in videos object tracking: A survey," in *Proc. Int. Symp. Digit. Earth*, vol. 18. 2014, pp. 12–20.

[5] D. Makris and T. Ellis, "Spatial and probabilistic modelling of pedestrian behaviour," in *Proc. Brit. Mach. Vis. Conf.*, vol. 2. 2002, pp. 557–566.

[6] C. F. Wakim, S. Capperon, and J. Oksman, "A Markovian model of pedestrian behavior," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2004, pp. 4028–4033.

[7] G. Antonini, M. Bierlaire, and M. Weber, "Discrete choice models of pedestrian walking behavior," *Transp. Res. B, Methodol.*, vol. 40, no. 8, pp. 667–687, 2006.

[8] Y. Abramson and B. Steux, "Hardware-friendly pedestrian detection and impact prediction," in *Proc. IEEE Intell. Vehicle Symp.*, Jun. 2004, pp. 590–595.

[9] F. Large, D. Vasquez, T. Fraichard, and C. Laugier, "Avoiding cars and pedestrians using velocity obstacles and motion prediction," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 375–379.

[10] T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka, "Development of night-vision system," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 3, pp. 203–209, Sep. 2002.

[11] H. Shimizu and T. Poggio, *Direction Estimation of Pedestrian From Images*, document AI Memo 2003-020, Massachusetts Institute of Technology, 2003, pp. 1–11.

[12] T. Gandhi and M. M. Trivedi, "Image based estimation of pedestrian orientation for improving path prediction," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2008, pp. 506–511.

[13] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. M. Gavrila, "Multi-cue pedestrian classification with partial occlusion handling," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 990–997.

[14] G. Zhao, M. Takafumi, K. Shoji, and M. Kenji, "Video based estimation of pedestrian walking direction for pedestrian protection system," *J. Electron.*, vol. 29, nos. 1–2, pp. 72–81, 2012.

[15] F. Flohr, M. Dumitru-Guzu, J. F. P. Kooij, and D. M. Gavrila, "Joint probabilistic pedestrian head and body orientation estimation," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 617–622.

[16] A. Bensebaa, S. Larabi, and N. M. Robertson, "Head direction estimation from silhouette," in *Proc. 17th Int. Conf. Image Anal. Process. (ICIAP)*, 2013, pp. 340–350.

[17] A. Bensebaa, S. Larabi, and N. M. Robertson, "Inferring heading direction from silhouettes," in *Developments in Medical Image Processing and Computational Vision*, vol. 19. The Netherlands: Springer, 2015, pp. 319–334.

[18] W. Liu, Y. Zhang, S. Tang, J. Tang, R. Hong, and J. Li, "Accurate estimation of human body orientation from RGB-D sensors," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1442–1452, Oct. 2013.

[19] D. Goel and T. Chen, "Pedestrian detection using global-local motion patterns," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2007, pp. 220–229.

[20] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3D pose estimation and tracking by detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 623–630.

[21] C. Chen, A. Heili, and J. Odobez, "Combined estimation of location and body pose in surveillance video," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug./Sep. 2011, pp. 860–867.

[22] D. Baltieri, R. Vezzani, and R. Cucchiara, "People orientation recognition by mixtures of wrapped distributions on random trees," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 270–283.

[23] H. Liu and L. Ma, "Online person orientation estimation based on classifier update," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1568–1572.

[24] T. Gandhi and M. M. Trivedi, "Pedestrian collision avoidance systems: A survey of computer vision based recent studies," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2006, pp. 17–20.

[25] S. Pierard and M. Van Droogenbroeck, "Estimation of human orientation based on silhouettes and machine learning principles," in *Proc. Int. Conf. Pattern Recognit. Appl. Methods (ICPRAM)*, 2012, pp. 51–60.

[26] K. Goto, K. Kidono, Y. Kimura, and T. Naito, "Pedestrian detection and direction estimation by cascade detector with multi-classifiers utilizing feature interaction descriptor," in *Proc. IEEE Intell. Vehicle Symp. (IV)*, Jun. 2011, pp. 224–229.

[27] J. Tao and R. Klette, "Integrated pedestrian and direction classification using a random decision forest," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV) Workshops*, Dec. 2013, pp. 230–237.

[28] R. Raman, P. K. Sa, and B. Majhi, "Direction prediction for avoiding occlusion in visual surveillance," *Innov. Syst. Softw. Eng.*, vol. 12, no. 3, pp. 1–14, 2016.

[29] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[30] H. Nisar and T.-S. Choi, "Fast motion estimation algorithm based on spatio-temporal correlation and direction of motion vectors," *Electron. Lett.*, vol. 42, no. 24, pp. 1384–1385, Nov. 2007.

[31] *INRIA Person Dataset*. (Jan 05, 2016). [Online]. Available: http://pascal.inrialpes.fr/data/human/

[32] *TUD Multiview Pedestrian Dataset*. (Jan 15, 2016). [Online]. Available: https://www.d2.mpi-inf.mpg.de/node/428

[33] *Daimler Mono Pedestrain Detection Benchmark Dataset*. [Online]. (Feb 13, 2016). Available: http://www.gavrila.net/Datasets/Daimler_Pedestrian_Benchmark_D/Daimler_Mono_Ped__Detection_Be/daimler_mono_ped__detection_be.html

[34] *PETS 2009 Dataset*. [Online]. (Jan 19, 2016). Available: http://www.cvg.reading.ac.uk/PETS2009/a.html

[35] *PKU Person Orientation Dataset*. [Online]. (Feb 04, 2016). Available: https://github.com/mlq513773348/PKU-Person-Orientation.git

[36] *3DPeS Dataset*. [Online]. (Feb 19, 2016). Available: https://www.openvisor.org/3dpes.asp

[37] J. Yao and J.-M. Odobez, "Multi-layer background subtraction based on color and texture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[38] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved foreground detection via block-based classifier cascade with probabilistic decision integration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 83–93, Jan. 2013.

[39] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[40] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. DARPA Image Understand. Workshop*, 1981, pp. 121–130.

[41] S. M. Siddiqi, G. J. Gordon, and A. W. Moore, "Fast state discovery for HMM model selection and learning," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2007, pp. 492–499.

[42] G. Celeux and J.-B. Durand, "Selecting hidden Markov model state number with cross-validated likelihood," *Comput. Statist.*, vol. 23, no. 4, pp. 541–564, 2008.

[43] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

[44] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1992, pp. 379–385.

[45] *CASIA Dataset*. [Online]. (Dec 05, 2015). Available: http://www.csbr.ia.ac.cn/english/Gait%20databases.asp

[46] *NITR Conscious Walk Dataset*. [Online]. (Dec 22, 2015). Available: http://www.nitrkl.ac.in/Academic/Academic_Centers/Data_Computer_Vision.aspx

[47] A. Mahapatra, T. K. Mishra, P. K. Sa, and B. Majhi, "Human recognition system for outdoor videos using hidden Markov model," *AEU-Int. J. Electron. Commun.*, vol. 68, no. 3, pp. 227–236, 2014.

[48] S. Zhang, D. A. Klein, C. Bauckhage, and A. B. Cremers, "Fast moving pedestrian detection based on motion segmentation and new motion features," *Multimedia Tools Appl.*, vol. 75, no. 11, pp. 6263–6282, 2016.
[49] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2003, pp. 734–741.

RAHUL RAMAN received the Master in Technology degree from the National Institute of Technology Rourkela Rourkela, India in 2013. He is currently a Doctoral Research Scholar with the Department of Computer Science and Engineering, National Institute of Technology Rourkela, India. His area of interest includes image processing, computer vision, visual surveillance, and biometrics. His research publications include international journals and conferences, book chapters, and patents in the same domain. He has also served as the Guest Editor and Reviewer in many peer-reviewed international conferences and journals.

PANKAJ KUMAR SA (M'07) received the Ph.D. degree in Computer Science in 2010. He is currently serving as an assistant professor with the Department of Computer Science and Engineering, National Institute of Technology Rourkela, India. His research interests include computer vision, biometrics, visual surveillance, and robotic perception. He has coauthored a number of research articles in various journals, conferences, and book chapters. He has co-investigated some research and development projects that are funded by SERB, DRDO-PXE, DeitY, and ISRO. He was a recipient of prestigious awards and honors for his excellence in academics and research. Apart from research and teaching, he conceptualizes and engineers the process of institutional automation.

BANSHIDHAR MAJHI (M'07) is a Professor with the Department of Computer Science and Engineering, National Institute of Technology Rourkela, India. He has successfully executed various Research and Development projects being funded by agencies such as MHRD, ISRO, DRDO, and DeitY. He has authored hundreds of articles in reputed journals and conferences. His current research interests include image processing, computer vision, biometric security, and pattern recognition. He has been conferred with prestigious awards and honors for his contribution towards scientific research and academic excellence.

SAMBIT BAKSHI (M'13) is currently with the Centre for Computer Vision and Pattern Recognition, National Institute of Technology Rourkela, India. He also serves as an Assistant Professor with the Department of Computer Science and Engineering, National Institute of Technology Rourkela, India. He received the Ph.D. degree in computer science and engineering in 2015. He serves as an Associate Editor of the *International Journal of Biometrics* (2013). He is a member of the IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence. He received the prestigious Innovative Student Projects Award - 2011 from the Indian National Academy of Engineering for his master's thesis. He has authored or coauthored over 30 publications in journals, reports, and conferences.