2008 IEEE Intelligent Vehicles Symposium
Eindhoven University of Technology
Eindhoven, The Netherlands, June 4-6, 2008

Image Based Estimation of Pedestrian Orientation for Improving Path Prediction

Tarak Gandhi and Mohan Manubhai Trivedi

Abstract— **Pedestrian protection is an essential component of driver assistance systems. A pedestrian protection system should be able to predict the possibility of collision after detecting the pedestrian, and it is important to consider all the cues available in order to make that prediction. The direction in which the pedestrian is facing is one such cue that could be used in predicting where the pedestrian may move in future. This paper describes a novel approach to determine the pedestrian's orientation using Support Vector Machine (SVM) based scheme. Instead of providing a hard decision, this scheme estimates the discrete probability distribution of the orientation. A Hidden Markov Model (HMM) is used to model the transitions between orientations over time and the orientation probabilities are integrated over time to get a more reliable estimate of orientation. Experiments showing the performance of estimating orientations are described to show the promise of the approach.**

I. INTRODUCTION

Accidents involving pedestrians and other vulnerable road users such as bicyclists are one of the leading causes of death and injury around the world. In order to reduce these accidents, pedestrian protection systems needs not only to detect pedestrians, but also predict the possibility of collision based on the paths that the pedestrian and the vehicle are likely to take. The system should relay the information to the driver in efficient and non-distracting manner or to the control system of the vehicle in order to take preventive actions. For this purpose, it is necessary to consider of all the available data from the environment, vehicle dynamics, and driver that can be obtained using various sensors incorporated in vehicle and infrastructure [1].

In order to estimate the possibility of collision between a pedestrian and a vehicle, the trajectories that they are likely to take need to be predicted. At scenarios such as intersections and pedestrian crossings, pedestrians as well as vehicles change their speed and direction quite frequently. Also, unlike vehicles, pedestrians are capable of making sudden maneuvers in terms of the speed and direction of motion. Hence, it is appropriate to use a stochastic model that predicts the probability distribution of the pedestrian trajectories and estimate the likelihood that the pedestrian and vehicle may collide. In the case when the pedestrian is moving, one can infer the direction of motion as well as the orientation

This work was supported in part by Volkswagen Research Laboratory, UC Discovery Grant, and the U.S. Department of Defence.

T. Gandhi was with the Laboratory for Intelligent and Safe Vehicles, University of California San Diego, La Jolla, CA 92093 USA (e-mail: tgandhi@ucsd.edu).

M. M. Trivedi is with the Laboratory for Intelligent and Safe Vehicles, University of California San Diego, La Jolla, CA 92093 USA (phone: 858-822-0075; e-mail: mtrivedi@ucsd.edu)

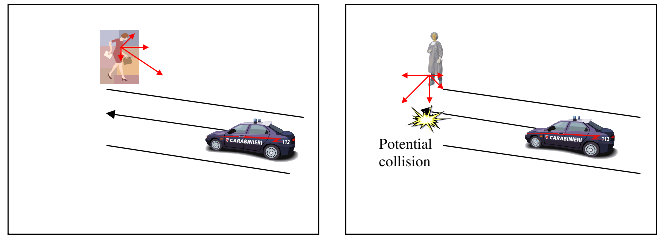


Fig. 1. Dependence of future pedestrian trajectories on orientation of the body. The red arrows denote the probabilities of anticipated movement in the respective directions. The pedestrian is more likely to move in the direction in which she is oriented.

of the pedestrian from the past trajectory. However, when the pedestrian is static, the direction of the motion is not defined, but the person is more likely in the future to move in the direction he/she is facing. Therefore, the orientation information can potentially improve the prediction of future trajectories that the pedestrian may take, which can result in improved prediction of the possibility of collision as seen in Figure 1.

In the subsequent sections, we address this problem of estimating the pedestrian's orientation and using it to predict future trajectories that the pedestrian is likely to take. Section II describes the related research in pedestrian detection, path prediction, and orientation estimation. Section III describes the proposed approach for estimation and prediction of pedestrian orientation based on feature extraction, classification, and temporal filtering with prediction. Section IV describes the experimental results based on video data collected at an intersection. Section V summarizes the article and explores avenues for future work.

II. RELATED RESEARCH

Considerable research has been performed on detecting pedestrians from static as well as moving platforms as described in the survey papers [2], [3]. A few researchers have also addressed the problem of predicting collisions. In [4], [5], Antonini et al. use a "Discrete Choice Model" in which a pedestrian makes a choice at every step about the speed and direction of the next step. This model is integrated with person detection and tracking from static cameras in order to improve performance. Instead of making hard decisions about target presence on every frame, it integrates evidence from a number of frames before making a decision. In [6], the spatio-temporal relationships between people and vehicles are analyzed in semantic framework and categorized as safe or unsafe depending on the site context

978-1-4244-2569-3/08/$20.00 ©2008 IEEE.

such as walkway or driveway, as well as motion context in terms of trajectories. Abramson and Steux [7], apply particle filtering for simultaneously tracking the object and predicting the collision probability.

In [8], Wakim et al. model the pedestrian dynamics using Hidden Markov Model with four states corresponding to standing still, walking, jogging, and running. For each state, the probability distributions of absolute speed as well as the change of direction modeled as truncated Gaussians. Monte Carlo simulations are then used to generate a number of feasible trajectories and the ratio of the trajectories on collision course to total number of trajectories give the collision probability. The European project CAMELLIA [9] has conducted research in pedestrian detection and impact prediction partly based on [7], [8].

In this research, we note that the orientation of the pedestrian body can give useful information about the future direction of motion. Hence, estimating the pedestrian orientation can potentially improve the motion prediction and give better estimates of collision probability.

Human body pose estimation has recently been of great interest to the computer vision community. Agarwal and Triggs [10], develop a learning-based method for estimating the 3D body pose of persons from monocular images as well as video sequences. The approach uses histogram-of-shape-context descriptors as features and various regression methods for estimating the 3D pose. Cucchiara et al. [11] distinguish between various human postures such as standing, crouching, sitting, and laying using Probabilistic Projection Maps on 2-D silhouettes of the person. However, most of the systems use an accurate silhouette of the person, which may not always be available especially from moving platforms. Here, we perform classification based on bounding box enclosing the person without using the silhouettes.

Support Vector Machines (SVM) have been used effectively for classifying between two or more classes, or estimating a continuous variable using regression [12]. In [13], SVM regression is used with Local Gradient Orientation (LGO) Histogram to estimate the orientation of driver's face. Shimizu and Poggio [14] estimate the pedestrian orientation using multiple SVM classifiers on Haar wavelet coefficients, in order to distinguish between different orientations. In this research, we classify the orientation of pedestrian using SVM classification on Histogram of Oriented Gradient based features. However, instead of determining a single direction in which the pedestrian is oriented, we obtain the output in terms of probabilities of several discrete orientations. This approach enables integration and prediction over multiple frames in a probabilistic framework.

III. PROPOSED APPROACH

Figure 2 shows the block diagram of the proposed approach for pedestrian orientation estimation and prediction. The orientation of the pedestrian is discretized into $C = 8$ directions as shown in Figure 3. It is assumed that the pedestrian detection has already been addressed [15], hence it is assumed that the bounding boxes of the pedestrians are available. Characteristic features are extracted from bounding box containing the pedestrian and the feature vector is fed to a classifier to estimate the orientation of the pedestrian. Histograms of Oriented Gradients (HOG) proposed by Dalal and Triggs [13] are used as feature descriptors and the classifier is based on Support Vector Machine (SVM). Instead of performing only hard classification in terms of a single output corresponding to estimated orientation, probabilities of each discrete orientation are estimated by the classifier. Such soft classification is useful especially when the actual orientation direction does not have the best match, but still has significant probability associated with it. Furthermore, it enables integration from multiple image frames by using a Hidden Markov Model with the dynamic model of pedestrian. Prediction of future orientation probabilities is also performed by applying the dynamic model to the current orientation probabilities. The predicted orientations would be useful in predicting where the pedestrian may move in future and thus reflect the possibility of collision.

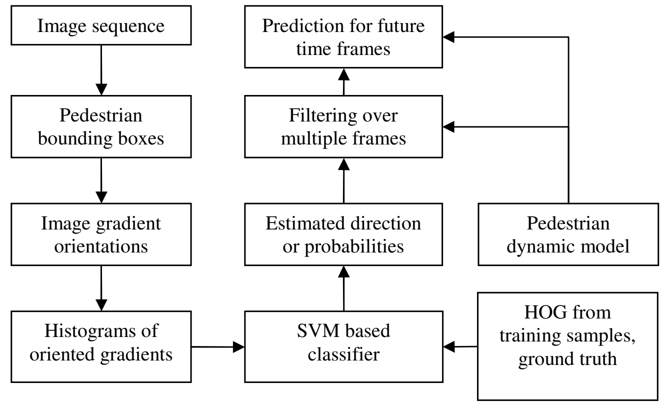


Fig. 2. Block diagram of the proposed orientation estimation and prediction system.

A. Feature Extraction

In order to distinguish between different orientations, characteristic features are extracted from the pedestrian images. Histograms of oriented gradients (HOG) have been proposed by Dalal and Triggs [13] to classify objects. This descriptor resembles the Scale-Invariant Feature Transform (SIFT) descriptor proposed by Lowe [16]. However, SIFT finds orientation histograms around characteristic feature points, whereas in HOG, the region of interest is subdivided into rectangular blocks and histogram of gradient orientations is computed in each block. This dense representation of orientation histograms obviates the need of having keypoint detection or creation of accurate silhouette of the person for finding SIFT features. In fact [13] note that most keypoint detection algorithms have not been able to extract human structures reliably. The steps for computing HOG are as follows:

1) Identify pedestrians in the image. Make a bounding box around the pedestrian with fixed aspect ratio and resize the bounding box to a fixed size but keeping

same aspect ratio to compensate for scaling.

2) Compute the image gradient direction at each pixel of the resized bounding box $F(x,y)$ containing the pedestrian in horizontal and vertical directions.
3) Subdivide the bounding box into $M \times N$ blocks and quantize the gradient directions ϕ into K bins each spanning $360°/K$ radians. For each bin (m,n,k), count the number of pixels in the block (m,n) having the gradient direction in bin k. This way, an $M \times N \times K$ array consisting of $M \times N$ local histograms is formed.
4) Apply smoothing to the histogram array by convolving with averaging kernels in position and orientation directions to reduce sensitivity to discretization.
5) Normalize the histogram array, such that the bin values of each of the MN histograms add to 1.
6) To reduce the effect of spurious image discontinuities, clip the bin values to a maximum threshold $H_{clip} = 0.2$ and renormalize as suggested by Lowe [16].
7) Stack the resulting array into a $B = M \times N \times K$ dimensional feature vector $\mathbf{x}$.

B. Single Frame Based Estimation

Support Vector Machine (SVM) is used as classifier in order to estimate the orientation of the pedestrian. The typical form of SVM is a two class classifier that forms a decision boundary between the classes by maximizing the 'margin' i.e. the separation between nearest samples on two sides of the boundary [12]. An SVM is trained using a large number of examples from both classes to obtain the decision function. After training, the classifier processes unknown samples and classifies them based on which side of the decision boundary the feature vector lies.

Let $\mathbf{x}_i$ denote the feature vector and y_i denote one of the two class labels in $\{0,\ 1\}$. Feature vector $\mathbf{x}_i$ is projected into a higher dimensional space using a mapping function Φ which allows non-linear decision boundary given by:

$$\mathbf{w}^T\Phi(\mathbf{x}) + b = 0 \tag{1}$$

Classification is formulated as ν-SVM problem where ν is the parameter to accommodate training errors [12]:

$$\min_{\mathbf{w},b,\xi,\rho} \frac{1}{2}\mathbf{w}^T\mathbf{w} - \nu\rho + \frac{1}{L}\sum_{i=1}^{L}\xi_i \tag{2}$$

$$\text{subject to } y_i\left[\mathbf{w}^T\Phi(\mathbf{x}_i) + b\right] \geq \rho - \xi_i$$
$$\xi_i \geq 0,\ i = 1\ldots L,\ \rho \geq 0 \tag{3}$$

where ν is the parameter to accommodate training errors and ξ is used to account for some samples that are not separated by the boundary. The problem is converted into the dual form which is solved using quadratic programming [12]:

$$\min_{\alpha_i} \sum_{i=1}^{L}\sum_{j=1}^{L}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x}_j) y_j \alpha_j \tag{4}$$

$$\text{subject to } 0 \leq \alpha_i \leq \frac{1}{L},\ \sum_{i=1}^{L}\alpha_i \geq \nu,\ \sum_{i=1}^{L}\alpha_i y_i = 0 \tag{5}$$

Here, $K(\mathbf{x}_i,\mathbf{x}_j) = \Phi(\mathbf{x}_i)^T\Phi(\mathbf{x}_j)$ is the kernel function derived from the mapping function Φ, and represents the distance in the high-dimensional space. It should be noted that the kernel function is usually much easier to compute than the mapping function Φ. Here, the radial basis function is used as kernel:

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp\left(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \tag{6}$$

The decision for new sample with feature vector $\mathbf{x}$ is based on the value f given by:

$$f = \sum_{i=1}^{L}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x}) + b \tag{7}$$

In this system, the HOG feature vector for the pedestrian is fed to the classifier [12] with $C = 8$ orientation classes. Instead of providing a single class as an output, of probabilities of all orientation classes are estimated using the approach proposed in [17]. This scheme uses $C(C-1)$ individual binary SVMs that output the likelihood ratios distinguishing between each pair of classes p and q as:

$$L_{pq} = \frac{P(c = p|c \in \{p,q\},\mathbf{x})}{P(c = q|c \in \{p,q\},\mathbf{x})} \approx \exp\left(-Af_{pq} - B\right) \tag{8}$$

where f_{pq} is the decision value outputted by the particular SVM. The parameters A and B are obtained using cross-validation from training data. The pairwise class probabilities are then given by:

$$r_{pq} = P(c = p|c \in \{p,q\},\mathbf{x}) = \frac{L_{pq}}{1 + L_{pq}} \tag{9}$$

These pairwise probabilities satisfy the following equation:

$$r_{qp}P_p = r_{pq}P_q \tag{10}$$

Hence, to generate the posterior probabilities of the classes, the following constrained minimization problem is solved using an iterative method described in [12].

$$\min_{\mathbf{P}} \frac{1}{2}\sum_{p=1}^{C}\sum_{q\neq p}\left(r_{qp}P_p - r_{pq}P_q\right)^2 \tag{11}$$

$$\text{subject to } \sum_{p=1}^{C}P_p = 1,\ P_p \geq 0 \tag{12}$$

C. Multi-frame based integration and prediction

In order to improve the reliability of classification, the single frame classification probabilities computed above are integrated in a Bayesian framework of a Hidden Markov Model (HMM). The evolution of pedestrian orientation over time is modeled as a Markov chain governed by transition probabilities between different orientation states. The transition probabilities between orientation classes over a single time step are given by the matrix of values in $P(c_t|c_{t-1})$. The class conditional probabilities $P(o_t|c_t)$ are proportional to the posterior probabilities $P(c_t|o_t)$ of the classes given by the SVM classifier. At each time step t, the class probabilities estimated by integrating t image frames are computed recursively from those from $t-1$ frames using the equations:

Prediction Step:

$$P(c_t|o_{1:t-1}) = \sum c_{t-1} P(c_t|c_{t-1}) P(c_{t-1}|o_{1:t-1}) \quad (13)$$

Correction Step:

$$P(c_t|o_{1:t}) = \frac{P(o_t|c_t)P(c_t|o_{1:t-1})}{\sum_{c_t} P(o_t|c_t)P(c_t|o_{1:t-1})} \quad (14)$$

The prediction step smoothes the probability estimates obtained from previous observations whereas the observation step weights the probabilities according to the current observation. For predicting the future orientation probabilities, the prediction equation is applied repeatedly over a number of time steps without performing the observation step. The transition probabilities can be learned based on observing state transitions in a given sequence of pedestrians.

$$P(c_t|c_{t-1}) = \frac{\#\text{Transitions from state } c_{t-1} \text{ to } c_t}{\#\text{Total time steps}} \quad (15)$$

However, if it is assumed that the transition probabilities depend only on the difference in orientation Δc and are independent of the actual initial orientation, the probabilities can be estimated as:

$$P(\Delta c) = \frac{\#\text{Transitions from with orientation change of } \Delta c}{\#\text{Total time steps}} \quad (16)$$

In this case, the transition probabilities can be expressed as a vector and the prediction step becomes a circular convolution.

IV. EXPERIMENTAL ANALYSIS AND VALIDATION

The orientation estimation approach described above was tested in realistic conditions using images obtained from a signalized intersection with people standing and crossing at the crosswalks. Training was performed using the INRIA pedestrian dataset [18]. This dataset consists of 96×160 snapshots of a number of pedestrians standing or walking in different directions and background as shown in Figure 3. From this dataset, 664 snapshots were selected and the ground truth orientation in 8 directions was manually determined. In order to provide invariance to small translation and rotation, 6 samples were generated from each snapshot by subjecting it to an affine transformation. The central 48×108 pixels which contained the pedestrian were used for the computation of HOG features. The HOG contained $6 \times 12 \times 8$ bins in horizontal, vertical, and orientation directions respectively, forming a feature vector of length $B = 576$. The classifier was trained using the feature vector from these samples. In order to optimize the classification accuracy, cross-validation was performed by splitting the training set into partitions, training on one and testing on another. This was repeated multiple times in order to obtain optimal selection of parameters.

For testing the classifier, images were acquired from a video camera placed at a signalized intersection. The pedestrian bounding boxes were extracted manually and their orientation was noted as ground truth. The bounding boxes

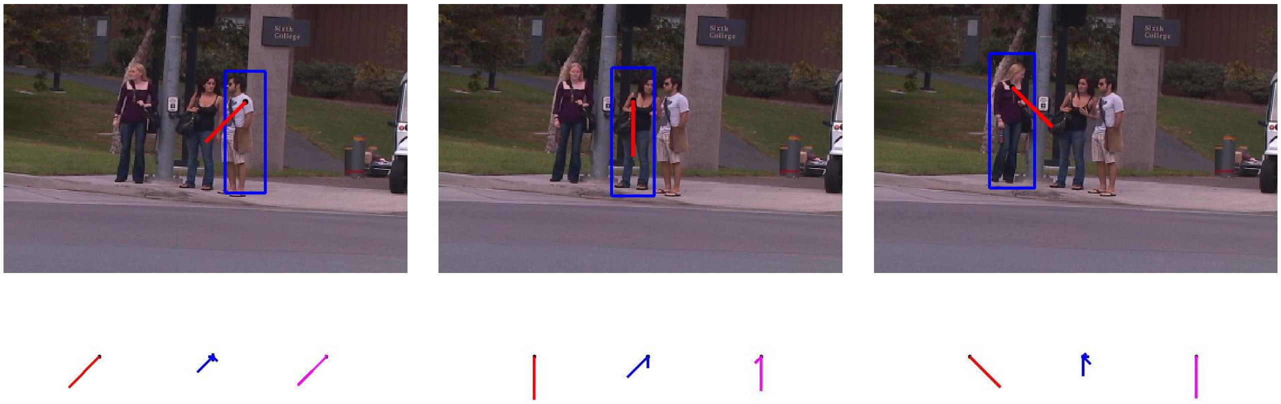
Fig. 3. Images of people from INRIA person database [18] in various orientations. The orientations are marked as yellow arrows and the orientation code between 0 and 7 is specified.

were resized to 48×108 pixels. Note that the bounding boxes were selected in such a way that they also would have the same aspect ratio $(48 : 108)$. This was done in order to preserve the information given by the aspect ratio about the orientation of the person, e.g. the person is wider in front than in side view. The bounding boxes were fed to the classifier that gives the probability values for each of the 8 discrete orientations using single frame as well as multiple frames as shown in Figure 4 (a). Similar experiment was conducted with video taken from a moving vehicle. Figure 4 (b) show sample images from the moving camera video.

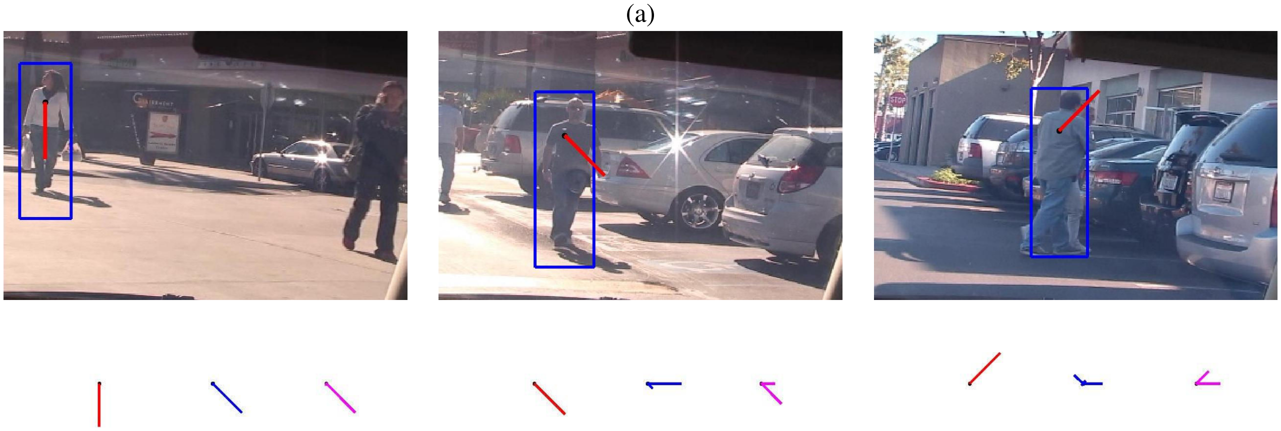
In order to measure the performance of the classifier, the orientation with maximum probability was compared with the ground truth. For each of the 427 examples, the difference between the ground truth orientation bin c_0 and the estimated orientation bin c with highest probability was noted. Histograms of these errors were plotted for single-frame estimation and multi-frame integration as shown in Figure 5 and those for multi-frame prediction are shown in Figure 6. For single frame estimation with static camera video and 8 discrete orientation bins, 41.7% of the orientations fall in the same bin as the ground truth, whereas 75.4% of orientations fall in the same or adjacent bins as shown in Table I. For multi-frame integration using HMM, 49.7% fall in same bin as ground truth and 81.3% fall in same or adjacent bins. The results deteriorate as expected as one performs prediction ahead in time. It is also seen that in the case of single frame analysis, the errors with orientation difference of 4 bins or $180°$ is more frequent than other errors. The reason for this could be that the classifier may be finding it difficult to discriminate between poses of person such as front and back. In the case of video from moving camera, the accuracies are somewhat lower. However, the error distribution trend is quite similar, having larger number of errors with orientation bin difference of 4 bins for single frame analysis. The accuracy results are also given in Table I.

V. CONCLUSION AND FUTURE WORK

Pedestrian protection is an important component of intelligent driver support system to enhance safety and convenience. A robust system that detects pedestrians and predicts the possible paths of the vehicle and pedestrians is essential for predicting the collision probability and take necessary steps to warn the driver. This paper addressed an important problem in pedestrian path prediction: To determine the orientation of the pedestrian in order to get a head-start on the knowledge of direction that the person may move in. Support



(a)



(b)

Fig. 4. Examples of orientation estimation using single-frame and multi-frame classification for video taken from (a) stationary camera (b) vehicle-mounted camera. Red line shows ground truth orientation, blue lines show orientation estimated from single frame, and magenta lines show the orientation probabilities estimated from current and previous frames of the person. Blue and magenta lines in all orientation directions have lengths proportional to probabilities of respective orientations.

TABLE I

PERFORMANCE OF ORIENTATION ESTIMATION AND PREDICTION FOR STATIC AND MOVING CAMERA VIDEO

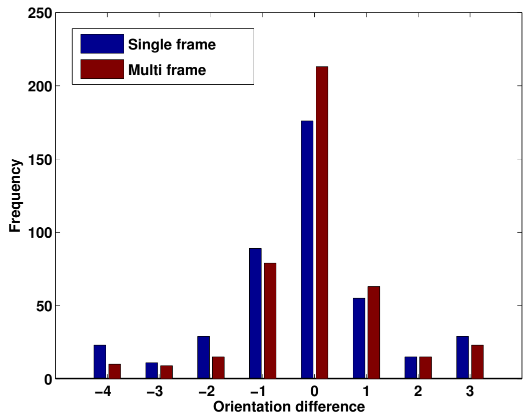
Estimation or Prediction scheme	Samples with estimated and actual orientation in same bin	Samples with error within 1 bin
Static camera video		
Single frame est.	41.7 %	75.4 %
Multi frame est.	49.9 %	83.1 %
Pred. (1 sec.)	43.3 %	74.7 %
Pred. (2 sec.)	37.0 %	73.3 %
Pred. (3 sec.)	38.4 %	72.8 %
Moving camera video		
Single frame est.	25.2 %	61.2 %
Multi frame est.	29.2 %	66.0 %

Vector Machines were used for classifying the orientation probabilities in eight directions and a Hidden Markov Model was used for integrating the estimates over time. Based on this knowledge, the prediction of pedestrian direction was made for a number of look-ahead times.

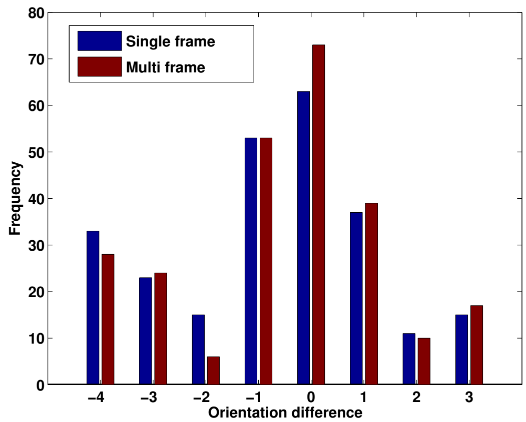
Future work consists of improving the classification results in order to get more reliable prediction. For predicting trajectories, scene context, such as whether pedestrian is on sidewalk, crosswalk or middle of road, traffic signal state, etc. can be incorporated. The transition probabilities between gait and orientation of pedestrians were learned from the training data. However, in real world, pedestrians may behave differently and it would be helpful to adaptively adjust the parameters in order to reflect the tendencies of individual pedestrians. Finally, the system would need to be integrated with driver monitoring systems so that appropriate warnings are generated according to the state of the driver.

VI. ACKNOWLEDGMENTS

The authors thank the UC Discovery Grant with several automobile companies, Department of Defense Technical Support Working Group, National Science Foundation for the sponsorship of the research. The authors also thank the members in the Computer Vision and Robotics Research Laboratory including Dr. Shinko Cheng, Dr. Stephen Krotosky, Brendan Morris, and Erik Murphy-Chutorian for their contributions.



(a)



(b)

Fig. 5. Histogram of difference between ground truth and estimated orientation with static camera video for single frame and multi frame estimation for (a) static camera (b) vehicle-mounted camera. It is seen that the error in orientation estimation is reduced in the multi-frame estimation.

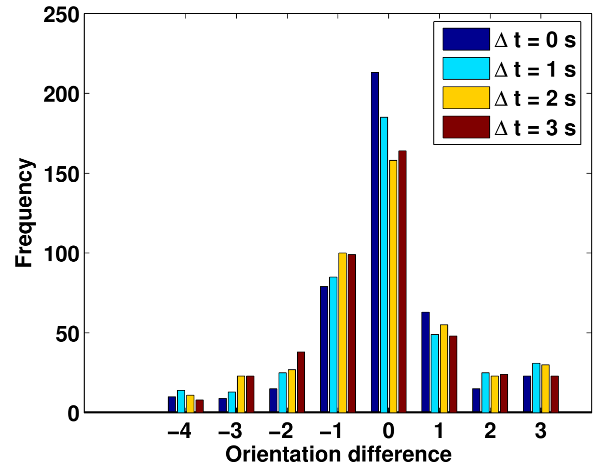


Fig. 6. The orientation error histograms for HMM based prediction with static camera video. The first set of bars represents the estimation at the same time without prediction. The other sets of bars represent the prediction 1, 2, and 3 seconds ahead. The errors generally increase with the prediction look-ahead time.

REFERENCES

[1] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer vision based enhanced vehicle safety," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 108–120, March 2007.

[2] T. Gandhi and M. M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 413–430, September 2007.

[3] D. M. Gavrila, "Sensor-based pedestrian protection," *IEEE Intelligent Systems*, vol. 6, no. 5, pp. 77–81, 2001.

[4] G. Antonini, S. Venegas, J. P. Thiran, and M. Bierlaire, "A discrete choice pedestrian behavior model for pedestrian detection in visual tracking systems," in *Proc. Advanced Concepts for Intelligent Vision Systems*, September 2004.

[5] G. Antonini, S. V. Martinez, M. Bierlaire, and J. P. Thiran, "Behavioral priors for detection and tracking of pedestrians in video sequences," *International Journal of Computer Vision*, vol. 69, no. 2, pp. 159–180, 2006.

[6] S. Park and M. M. Trivedi, "Analysis and query of person vehicle interactions in homography domain," in *Proc. ACM Workshop on Video Surveillance and Sensor Networks*, October 2006.

[7] Y. Abramson and B. Steux, "Hardware-friendly pedestrian detection and impact prediction," in *IEEE Intelligent Vehicle Symposium*, June 2004, pp. 590–595.

[8] C. Wakim, S. Capperon, and J. Oksman, "A markovian model of pedestrian behavior," in *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics*, October 2004, pp. 4028–4033.

[9] "Deliverable 3.3b report on initial algorithms 2," CAMELLIA: Core for Ambient and Mobile intELLigent Imaging Applications, Tech. Rep. IST-2001-34410, December 2003.

[10] A. Agarwal and B. Triggs, "Recovering 3d human pose from monocular images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 44–58, January 2006.

[11] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic posture classification for human-behavior analysis," *IEEE Transactions on Systems, Man, And Cybernetics, Part A: Systems and Humans*, vol. 35, no. 1, pp. 42–54, 2005.

[12] C.-C. Chang and C.-J. Lin, *LIBSVM : A Library for Support Vector Machines*, Last updated June 2007, http://www.csie.ntu.edu.tw/ cjlin/papers/libsvm.pdf.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2005.

[14] H. Shimizu and T. Poggio, "Direction estimation of pedestrian from images," Massachusetts Institute of Technology, Cambridge, MA, AI Memo 2003-020, August 2003, http://cbcl.mit.edu/cbcl/publications/ai-publications/2003/AIM-2003-020.pdf.

[15] S. J. Krotosky and M. M. Trivedi, "On color-, infrared-, and multimodal-stereo approaches to pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 4, pp. 619–629, December 2007.

[16] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,*, vol. 60, no. 2, pp. 91–110, 2004.

[17] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004, http://www.csie.ntu.edu.tw/˜cjlin/papers/svmprob/svmprob.pdf.

[18] "INRIA person dataset," a large set of marked up images of standing or walking people, used to train Navneet Dalal's CVPR 2005 human detector http://pascal.inrialpes.fr/data/human/.