2014 IEEE 17th International Conference on
Intelligent Transportation Systems (ITSC)
October 8-11, 2014. Qingdao, China

Pedestrian Path Prediction based on Body Language and Action Classification

R. Quintero, I. Parra, D. F. Llorca and M. A. Sotelo

***Abstract*— Safety-related driver assistance systems are becoming mainstream and nowadays many automobile manufacturers include them as standard equipment. For example, pedestrian protection systems are already available in a number of commercial vehicles. However, there is still work to do in the improvement of the accuracy of these systems since the difference between an effective and a non-effective intervention can depend on a few centimeters or on a fraction of a second. In this paper, we use the 3D pedestrian body language in order to perform accurate pedestrian path prediction by means of action classification. To carry out the prediction, we propose the use of GPDM (Gaussian Process Dynamical Models) that reduces the high dimensionality of the input vector in the 3D pose space and learns the pedestrian dynamics in a latent space. Instead of combining a reduced number of subjects in a single model that will have to deal with the stylistic variations, we propose a much more scalable approach where all the subjects are separately trained in individual models. These models will be then hierarchically separated according to their action (walking, starting, standing, stopping) and direction of the motion. Finally, for a test sequence, the appropiate model will be selected by means of an action classification system based on the similarity of the 3D poses transitions and the joints velocities. The estimated action will constrain the models to use for the prediction, taking into account only the ones trained for that action. Experimental results show that the system has the potential to provide accurate path predictions with mean errors of 7 cm, for walking trajectories, 20 cm, for stopping trajectories and 14 cm for starting trajectories, at a time horizon of 1 s.**

I. Introduction and Related Work

Pedestrian path prediction is a hot research topic in different application contexts such as robotics, surveillance or human-machine interaction, but it is in the Advanced Driver Assistance Systems (ADAS) context where it is a matter of the utmost importance. Pedestrian detection, collision avoidance or near collision warning systems require accurate information about the current and future positions of the pedestrians. A difference of 30 cm in the estimated lateral position of a pedestrian can make the difference for a successful collision avoidance maneuver [1]. Moreover, accident analysis in [2] showed that initiating an emergency braking 0.16 s in advance could reduce the severity of accident injuries up to 50%. Early recognition of pedestrian intent can lead to much more accurate active interventions in last second automatic maneuvers. As a consequence, over the last few years a lot of effort has been put into understanding the pedestrian intentions.

R. Quintero, D. F. Llorca and M.A Sotelo are with the Computer Engineering Department, University of Alcalá, Alcalá de Henares, Spain {raul.quintero,llorca,sotelo}@aut.uah.es

I. Parra is with the Signals, Systems and Radiocommunication Department, Polytechnic School of Madrid, Madrid, Spain i.parra@upm.es

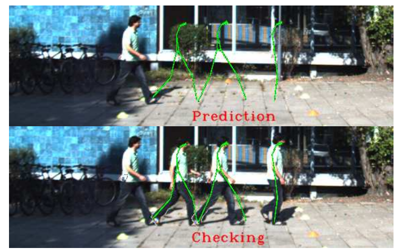


Fig. 1. Pedestrian path and pose prediction.

Early approaches for pedestrian detection and tracking used Kalman Filters in a trajectory-based framework [3], including interacting multiple model filters [4] [5], in order to account for different motion dynamics. Nonetheless, the sole consideration of trajectory is clearly insufficient for predicting the pedestrian path in an accurate manner in situations with changing motion dynamics. Empirical studies [6] have demonstrated that when only the trajectory of the pedestrian is available, a higher error rate is produced in drivers judgment regarding the pedestrian intentions. Other systems use the whole pedestrian body language to provide an early indicator of the pedestrian intentions [7] [8]. A common approach is to learn the dynamics for different actions (walking, running, stopping, starting) using probabilistic frameworks that reduce the dimensionality of the input data in the so-called latent space [9] [10] [11]. In [12] a non-linear model with stylistic variation (multiple people walking) is learned using Local Linear Embedding [13] by first building individual models and then using non-linear regression to align the manifold and build a unified model. However, it is not clear that this combined models approach can deal with complex motions or with many subjects. In [14], models for different activities are learned within a shared latent space, along with transitions between activities. They proposed a constrained combined model that learns smooth transitions between models without the need of including these transitions in the training data. However these constrains made the training process very complex, especially with noisy data.

Some of the previous works have focused on learning activity specific models and try to combine them in a latent space with natural transitions from one activity to another. In general, this activity specific models fail to generalize when there are large stylistic variations and are hard to adjust for the different activities transitions. To overcome the problem of combining different activities in the same model [1] proposed to use separate models (one for walking, one for stopping) and keeping a continuous estimation of

978-1-4799-6077-4/14/$31.00 ©2014 IEEE

their probabilities in an Interacting Multiple Model Filter framework. This solves the problem of combining different activities, but it is computationally expensive when the number of possible activities increases. This approach used augmented motion features derived from dense optical flow instead of the pedestrian body language. That can make their models less sensitive to stylistic variations but also less accurate in the prediction as there is less information available in the image optical flow than in the pedestrian 3D pose time sequence.

In this paper, we propose a novel approach to perform pedestrian path prediction for walking, starting, stopping and standing actions based on the pedestrian body language. This new approach is based on our previous system described in [15]. In our system, instead of combining a reduced number of subjects in a single model that will have to deal with the stylistic variations, we propose a much more scalable approach where all the subjects are separately trained in individual models. These models are then hierarchically divided according to their motion and action (left/right and walking, starting, stopping, standing). For the selection of the different actions, a continuous estimation is maintained based on the similarity of the 3D poses transitions and the joints velocities. Finally, the appropriate model will be selected from the detected action sub-set, using a pose-based hierarchical search in the 3D space that allows us to easily introduce new subjects in the database. The feasibility of this new approach has been tested using the publicly available CMU data-set [16].

The remaining of the paper is organized as follows: Section II provides a description of the system. The sequences of 3D pedestrian poses are used to create individual low-dimensional embeddings as illustrated in section II-A. The data-sets used to create the models are presented in section II-B. Then section II-C describes the naive-Bayes classifier used to perform the action classification. Finally, the appropriate model will be selected, among those corresponding to the detected action, by using a pose-based search in the 3D space as explained in section II-D. Experimental results are presented in section III. We discuss our conclusions and future work in section IV.

II. System Description

Our final goal is to develop a pedestrian path prediction system set up in a moving vehicle equipped with stereo cameras and LIDAR. In this paper, we will first test the feasibility and limits of our approach by using the high frequency and low noise data-sets from CMU [16]. In the future, we will apply the learnt models to a stereovision-based pedestrian pose extraction system, as explained in [15], in order to obtain massive quantitative conclusions for vision-based systems. So far, in this paper we concentrate on studying the performance limits of the proposed GPDM-based pedestrian pose prediction system using quasi noise-free measurements.

A. GPDM

One of the most important ways of modeling in statistic and machine learning is to achieve a dimensionality reduction of a high-dimensional data. Several approaches have been followed in the technical literature for this purpose, such as PCA, GPLVM or GPDM. The latter has been applied in pedestrian path prediction sucessfully in the latest years using different types of features, i.e. 3D joints and velocity [9] [12] [15] or dense optical flow [1]. GPDM provides a framework for transforming a sequence of feature vectors, which are related in time, into a low dimensional latent space. In order to apply this transformation, the observation and the dynamics mapping are computed separately in a non-linear form, marginalizing out both mappings and optimizing the latent variables and the hyper-parameters of the kernels. The conditional probability of Y given X, θ and W for the observation mapping is defined in (1)

$$p(Y|X,\theta,W) = \frac{|W|^N}{\sqrt{(2\pi)^{ND}|K_Y|^D}} exp\,(-\frac{1}{2}tr\,(K_Y^{-1}YW^2Y^T)) \tag{1}$$

where Y is the centred observed data-set, X represents the latent positions on the model, K_Y is the kernel matrix, $\theta = [\theta_1, \theta_2, ..., \theta_N]$ contains the kernel hyper-parameters, N is the number of samples, D is the dimension of the data-set, and W is the scaling matrix (to account for different variances in the different data dimension). The elements of kernel matrix for the observation mapping are computed using (2).

$$k(x_i,x_j) = \theta_1 exp\,(\frac{-\theta_2}{2}(x_i-x_j)^T(x_i-x_j)) + \theta_3\delta_{i,j} \tag{2}$$

where $\delta_{i,j}$ is the Kronecher delta function.

The dynamic mapping from the latent coordinates is defined in (3),

$$p(X|\beta) = \frac{p(x_1)}{\sqrt{(2\pi)^{(N-1)d}|K_X|^d}} exp\,(-\frac{1}{2}tr\,(K_X^{-1}X_{out}X_{out}^T)) \tag{3}$$

where $X_{out} = [x_2,...,x_N]^T$, d is the model dimension, and K_X is the kernel matrix constructed from $\{x_1,...,x_{N-1}\}$ using the kernel function provided in (4)

$$k(x_i,x_j) = \beta_1 exp\,(\frac{-\beta_2}{2}(x_i-x_j)^T(x_i-x_j)) + \beta_3 x_i^T x_j + \beta_4\delta_{i,j} \tag{4}$$

where β_1 to β_4 are the kernel hyper-parameters.

The goal is to minimize the negative log-likelihood function $-ln\,p(X,\theta,\beta,W|Y)$ that is given in (5)

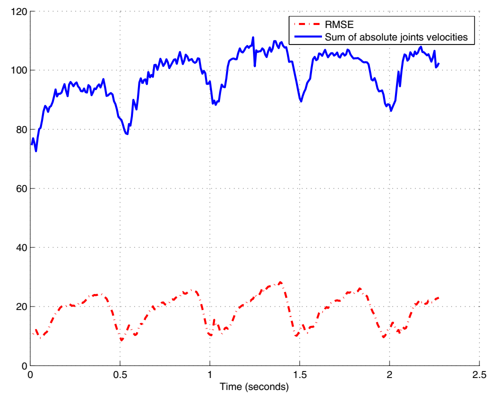
$$\mathscr{L} = \mathscr{L}_Y + \mathscr{L}_X + \sum_j ln\,\theta_j + \frac{1}{2\kappa^2}tr\,(W^2) + \sum_j ln\,\beta_j \tag{5}$$

where

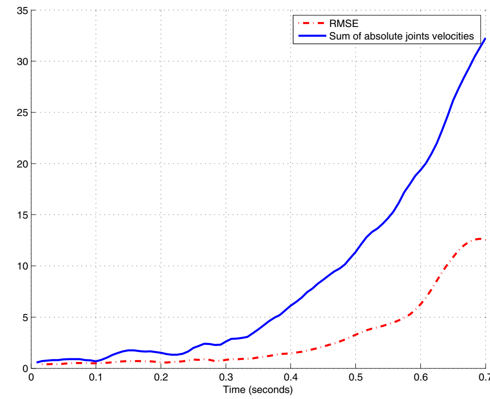
$$\mathscr{L}_Y = \frac{D}{2}ln\,|K_Y| + \frac{1}{2}tr\,(K_Y^{-1}YW^2Y^T) - N\,ln\,|W| \tag{6}$$

$$\mathscr{L}_X = \frac{d}{2}ln\,|K_X| + \frac{1}{2}tr\,(K_X^{-1}X_{out}X_{out}^T) + \frac{1}{2}x_1^T x_1 \tag{7}$$

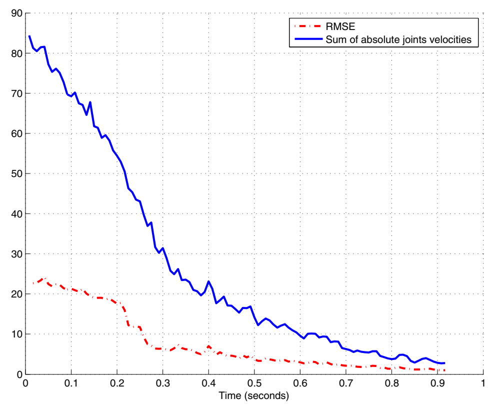
In order to increase the smoothness of the learned trajectories in the latent space, a modified version of GPDM can



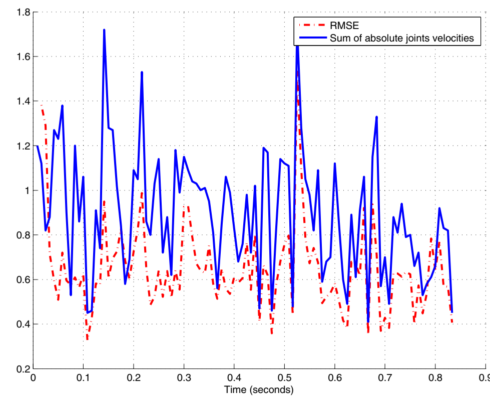
(a) Walking sequence



(b) Starting sequence



(c) Stopping sequence



(d) Standing sequence

Fig. 2. Difference between consecutive poses (red dash-dot) and sum of absolute joints velocity (blue solid) for walking, starting, stopping and standing actions.

be used by changing the weight of $\mathscr{L}_X$ by means of a λ element. A value for λ of $\frac{D}{d}$ is recommended in [9]. This modification is known as Balanced GPDM.

Given a latent position the original feature vector can be recovered as described in (8).

$$\mu = Y^T K_Y^{-1} k_Y(x) \tag{8}$$

where Y is the centred data-set, K_Y^{-1} the inverse matrix of the kernel for the observation mapping (see 2) and $k_Y(x)$ is a column vector with elements $k_Y(x, x_j)$ for all other latent position x_j in the model.

GPDM also provides the grounds for predicting the next position in the latent space based on the current latent position. Thus, the next latent position can be obtained as described in (9)

$$\mu_X(x) = X_{out}^T K_X^{-1} k_X(x) \tag{9}$$

where $X_{out} = [x_2, ..., x_N]^T$, K_X is the kernel matrix constructed from $\{x_1, ..., x_{N-1}\}$ using the kernel function provided in (4) and $k_X(x)$ is a column vector with elements $k_X(x, x_j)$ for all other latent position x_j in the model. A prediction at a time horizon of N latent positions ahead can be obtained computing (9) iteratively.

B. Data-set description

In our experiments we use the publicly available data-set from Carnegie Mellon University (CMU) [16]. It is composed of different pedestrian sequences captured using a high accuracy and high frequency (120 Hz) motion capture system (CMU mocap). We learn our individual models using these high frequency and low noise sequences to get high quality models. The high frequency will help the GPDM to properly learn the dynamics of the different actions and will increase the probability of finding a similar test pose in the trained data without missing intermediate poses. In addition, these low noise models will improve the prediction when working with noisy test samples.

The CMU data-set contains the 3D coordinates of 41 joints along the body. In our experiments we use a sub-set of the most relevant joints (shoulders, clavicle, sternum, hips, knees and ankles). Our feature vector is composed of the 3D pose and the joints velocities, removing the 3D body translation parameters. The joints velocities are included in the model because it was observed to increase the accuracy in the estimation of the reconstructed displacement. We reduce the dimensionality of the feature vector using the described GPDM to construct a latent space.

The pedestrian motions from the CMU data-set are hierarchicaly divided into eight sub-sets. The first division is based on the direction, left-to-right and right-to-left. The second one is based on the action (standing, starting, stopping and walking). To capture the dynamics of the different actions, the beginning and end of the sequences were cropped manually trying that all the poses in a sequence were representative of their action (see Fig.2). This is a key aspect for the early detection of the pedestrian intentions as the prediction will be based on the similarity of the pedestrian action with these training sequences. As shown in Table I our data-set is composed of 195 sequences from 27 different subjects divided according to the action and direction.

TABLE I
NUMBER OF SEQUENCES FOR EACH TYPE OF ACTION

	Standing	**Starting**	**Stopping**	**Walking**
Left-to-Right	21	26	9	70
Right-to-Left	12	21	9	27
Total	33	47	18	97

C. Action classification

As explained before, the early detection of the transitions between the different actions is a key point for an ADAS because it is in the transitions where it is critical getting an accurate path prediction.

In this paper we propose a naive-Bayes classifier based on the similarity between consecutive poses and the joints velocities to classify the pedestrian action into walking, stopping, standing or starting. Fig. 2 plots an example of the difference between consecutive poses (red dash-dot) and the sum of absolute joints velocities (blue solid) for the four different actions. As can be seen in the Figure, this information is distinctive, and can be used to estimate the action. The velocities will help us to distinguish between actions that at some points are similar in pose but show different velocity trends such as starting and stopping.

Let p_i be the pedestrian 3D pose at time i and $\dot{p}_i$ their velocity. We define a new feature vector $C = \{|p_i - p_{i-1}|, |\dot{p}_i - \dot{p}_{i-1}|\}$ where the first term captures the changes in the pose and the second term describes how fast these changes happened.

Then, the mean and variance are computed for all the actions in the training data-set and a naive-Bayes classifier is constructed. To estimate the action we use the *maximum a posteriori* (MAP) decision rule. Finally, the models of the action with the highest probability are used to predict the pedestrian path. Fig. 3 shows the different actions probabilities for a starting sequence. At first, the system is clearly classifying the action as standing, but as the pedestrian increases the speed, the action starts to be more similar to stopping and starting. At this point the velocity local variation becomes important to decide between starting and stopping.

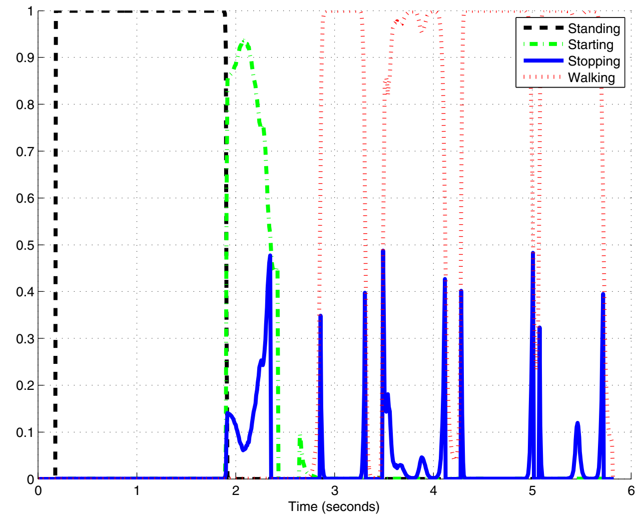


Fig. 3. Probabilities for a starting sequence (standing black dashed, starting green dot-dash, stopping blue solid and walking red dotted).

D. Model selection

Once we have estimated the pedestrian action we focus on selecting the appropiate model to deal with the stylistic variation and the different speeds of the tracked pedestrians. In our system, all the subjects are separately trained in individual models. Then, the appropriate model will be selected using a pose-based search in the 3D space that allows us to easily introduce new subjects in the database. With this approach we avoid the problems of learning a unified model or learning the transitions between models. However, we face the problem of selecting the most appropiate model to perform the path prediction.

To select the model we search for similar 3D poses in the corresponding action training sub-set, and use the most similar pose to start the search in the latent space. Once the latent position has been estimated, a prediction at a time horizon of N poses ahead can be done using (9) iteratively.

III. Experimental Results

The described system was tested using the CMU data-set with 27 subjects and 195 sequences. To test the generalization ability of the system a one vs. all strategy was adopted. This means that all the models generated by one test subject were removed from the training data while performing tests on this subject. All the tests were performed using the high frequency and low noise sequences from the CMU data-set to prove the feasibility of this approach. Therefore we consider this results as the best case scenario and further experimentation is required to test how this approach performs with noisy test sequences. A hint of how low frequency and noisy test samples can affect the prediction can be seen in [15].

A. Results on action classification

For pedestrian path estimation systems based on action classification, the sucessful detection of the action is essential, as the prediction will use only the models of the detected

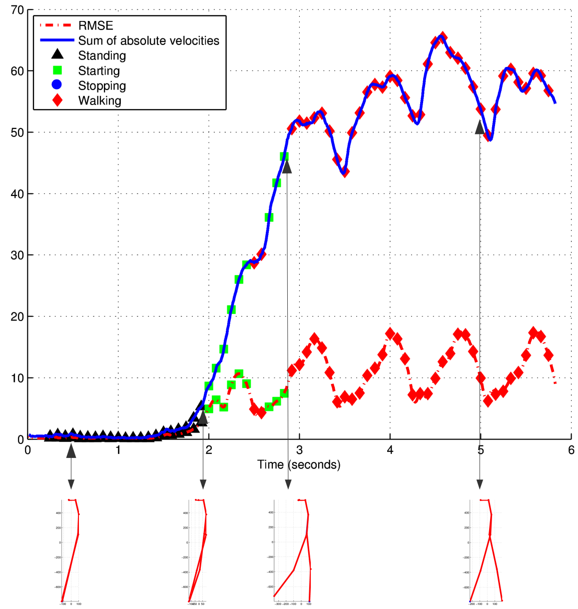


Fig. 4. Starting sequence, action classification results

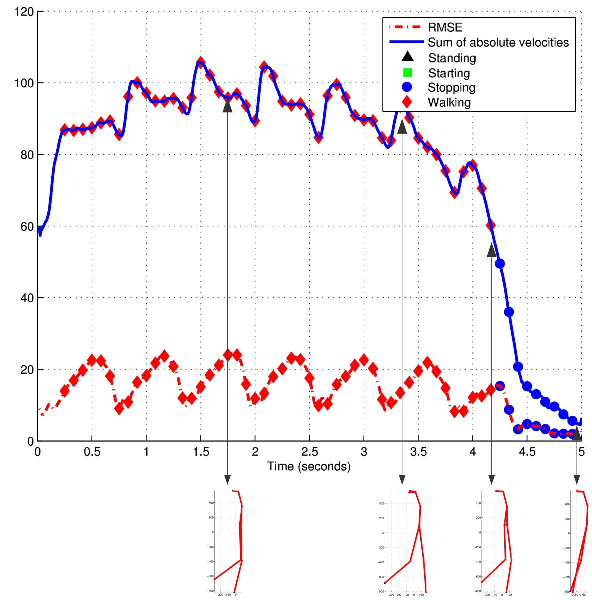


Fig. 5. Stopping sequence, action classification results

action. Also, an early detection of the transitions between actions will allow these systems to deliver more accurate pedestrian path estimations.

To test the performance of the proposed action classification algorithm the actions were manually labelled on 6 sequences (3 starting to walk and 3 stopping) by a human expert. In all, the sequences added up to 3373 pedestrian poses that were used as input for the action classification algorithm. Table II summarizes the classification results on a confusion matrix. The overall detection rate is a 95.20% for the 4 different actions.

TABLE II
CONFUSION MATRIX FOR ACTION CLASSIFICATION ALGORITHM

		Classification			
		Standing	**Starting**	**Stopping**	**Walking**
Actual	**Standing**	950	22	4	10
	Starting	0	157	0	60
	Stopping	12	0	98	26
	Walking	0	27	1	2006

The missclassifications are mainly starting actions classified as walking actions and the other way around. These errors are produced when the speed variation in a starting action decreases or when during a walking action the pedestrian increases the speed. However these missclassificatios are not critical from the point of view of the path estimation as both actions have similar dynamics and the path predictions will be also very similar. On the other hand, standing actions classified as starting (2.23%) and walking actions classified as stopping (0.05%) are due respectively to delays and early detections between the manually labelled actions and the action classification system.

Fig. 4 shows an example of the results of the action classification algorithm for a starting sequence. The output of the action classification algorithm (Standing black triangles, starting green squares, stopping blue circles and walking red diamonds) was overlaid on the difference between consecutive poses (red dashed) and the velocity (blue solid) plots for the starting sequence. A few 3D poses at significant points in the sequence were also introduced to get and idea of when the transitions are detected. For the sake of clarity, only one out of ten action classification markers have been plotted.

As shown in Fig. 4 the system detects the transition from standing to starting very early, when the pedestrian starts leaning and opening their legs. Around second 2.5 there is an example of missclassification where a reduction in the speed of the pedestrian action creates a walking classification for some frames. This is a missclassification that could lead to a slight underestimation of the velocities in the prediction but it is not a serious problem, as path predictions for walking and starting models are very similar. The system is adjusted to sharply detect standing to starting and walking to stopping transitions, where missclassifications would lead to serious errors in the path prediction.

Fig. 5 shows the results of the action classification algorithm for a stopping sequence. As can be seen, the transition from walking to stopping is detected as the stride gets reduced and the velocity starts to drop. The algorithm detects the stopping action approximately 0.83 s prior to the actual full stop.

B. Results on model selection and path prediction

Table III shows the mean combined longitudinal and lateral path prediction error and standard deviation (cm) for different prediction horizons obtained for the CMU data-sets described in Table I. As explained before, a one vs. all strategy was adopted to test the path prediction performance. The model was first selected from each one of the action data-sets as explained in section II-D and then a path prediction estimation is performed using the selected model as explained in section II-A. Path predictions are not obtained for standing actions as we assume no motion in this state.

TABLE III
MEAN COMBINED LONGITUDINAL AND LATERAL PREDICTION ERROR±STD (CM) FOR DIFFERENT PREDICTION HORIZONS (SECONDS)

	0 sec.	0.23 sec.	0.5 sec.	0.78 sec.	1 sec.
Walking	0.56 ±0.15	3.68 ±0.94	4.55 ±1.24	6.15 ±1.67	6.8 ±2.01
Stopping	1.55 ±0.81	9.33 ±5.3	14.69 ±7.6	16.92 ±6.74	19.38 ±7.94
Starting	0.82 ±0.36	5.77 ±3.48	10.31 ±6.29	12.35 ±6.51	14.04 ±6.57

As can be observed, prediction accuracy at 1 s is higher for walking sequences (6.8 cm) than for stopping (19.38 cm) or starting (14.04 cm). This could be due to the fact that walking sequences have a much more regular dynamics, making it easier for the model to predict the future positions. Also, the number of sequences for stopping and starting actions is much lower (see Table I), being more difficult for the system to cope with the stylistic variations.

IV. Conclusions and Future Work

We have developed a system for accurate pedestrian path prediction by means of action classification in a limited time horizon up to 1 s. For such purpose, we propose a naive-Bayes classifier based on the similarity of transitions in the 3D space and their velocity. Once the action has been classified the most similar pose is found on the 3D space and a prediction at a time horizon of 1 second ahead can be done using a GPDM model that reduces the high dimensionality of the input feature vector and learns the dynamics in a latent space. The system has the potential to provide accurate path predictions with mean errors of 7 cm, for walking trajectories, 20 cm, for stopping trajectories and 14 cm for starting trajectories, at a time horizon of 1 s. These results were obtained using the models created with the high accuracy and high frequency (120 Hz) CMU data-sets [16]. Our final goal is to develop a pedestrian path prediction system set up in a moving vehicle equipped with stereo cameras and LIDAR. The work presented in this paper can be considered as the best case scenario and further experimentation will be carried out to test how this approach performs with noisy test sequences.

As future work we propose to create a bigger data-set in order to include a signicative number of sequences for the different actions that will help the system to cope with the stylistic variations. In addition, experiments with pedestrian joints extraction systems in real conditions, will be performed to test the real predictive power of the system with noisy samples. We plan to continue the work on vision-based joints extraction using a 3D point cloud and fuse this information with a LIDAR to increase the accuracy. Finally, a CPU time profiling for the different algorithms of the system is needed to evaluate the implementation of some of them in GPUs or hardware.

V. ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Economy under Grant ONDA-FP TRA2011-27712-C02-02 and the "Juan de la Cierva" programme JCI-2012-14671.

References

[1] C. G. Keller and D. M. Gavrila, "Will the pedestrian cross? a study on pedestrian path prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. In Press, 2014.

[2] M. M. Meinecke, M. Obojski, D. M. Gavrila, E. Marc, R. Morris, M. Tons, and L. Lettelier, "Strategies in terms of vulnerable road user protection," EU Project. SAVE-U," Deliverable D6, 2003.

[3] D. F. Llorca, M. A. Sotelo, A. M. Hellín, A. Orellana, M. Gavilán, I. G. Daza, and A. G. Lorente, "Stereo regions-of-interest selection for pedestrian protection: A survey," *Transportation Research Part C*, vol. 25, pp. 226–237, 2012.

[4] M. Farmer, L. H. Rein, and A. Jain, "Interacting multiple model (imm) kalman filters for robust high speed human motion tracking," in *Int. Conf. on Pattern Recognition (ICPR)*, 2002, pp. 20–23.

[5] Y. Boers and J. Driessen, "Interacting multiple model particle filter," *IET Radar, Sonar Navigation*, vol. 150, no. 5, pp. 344–349, 2003.

[6] S. Schmidt and B. Farber, "Recognising the action intentions of humans," *Transportation Research Part F*, vol. 12, pp. 300–310, 2009.

[7] C. C. Lee, C. H. Chuang, J. W. Hsieh, M. X. Wu, and K. C. Fan, "Frame difference history image for gait recognition," in *International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 4, 2011, pp. 1785–1788.

[8] S. Kohler, M. Goldhammer, S. Bauer, S. Zecha, K. Doll, U. Brunsmann, and K. Dietmayer, "Stationary detection of the pedestrianś intention at intersections," *IEEE Intelligent Transportation Systems Magazine*, pp. 87–99, 2013.

[9] R. Urtasun, D. J. Fleet, , and P. Fua, "3d people tracking with gaussian process dynamical models," in *IEEE International Conference on Computer vision and pattern recognition (CVPR)*, 2006, pp. 238–245.

[10] A. Yao, J. Gall, L. Van Gool, and R. Urtasun, "Learning probabilistic non-linear latent variable models for tracking complex activities," in *NIPS*, 2011.

[11] L. Raskin, M. Rudzsky, and E. Rivlin, "Dimensionality reduction using a gaussian process annealed particle filter for tracking and classification of articulated body motions," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 503 – 519, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314210002651

[12] A. Elgammal and C.-S. Lee, "Separating style and content on a nonlinear manifold," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, June 2004, pp. I–478–I–485 Vol.1.

[13] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000. [Online]. Available: http://www.sciencemag.org/content/290/5500/2323.abstract

[14] R. Urtasun, D. J. Fleet, A. Hertzmann, and P. Fua, "Priors for people tracking from small training sets," *Computer Vision, IEEE International Conference on*, vol. 1, pp. 403–410, 2005.

[15] R. Quintero, J. Almeida, D. F. Llorca, and M. A. Sotelo, "Pedestrian path prediction using body language traits," in *IEEE Intelligent Vehicles Symposium (IV)*, 2014.

[16] CMU, "Cmu graphics lab motion capture database," http://mocap.cs.cmu.edu/subjects.php.