Learning and Predicting On-road Pedestrian Behavior Around Vehicles

Nachiket Deo and Mohan M. Trivedi
Laboratory for Intelligent and Safe Automobiles
University of California San Diego
La Jolla, USA
{ndeo, mtrivedi}@ucsd.edu

***Abstract*—Pedestrian safety is critical from the standpoint of autonomous vehicles. We address the problem of reliable long term ($\leq$ 5 sec) prediction of on-road pedestrian trajectories using track histories generated with two vehicle mounted cameras. Due to the relatively unconstrained nature of pedestrian motion, model based trajectory prediction approaches become unreliable for long term prediction. Data-driven approaches offer a viable alternative, but are susceptible to bias toward dominant motion patterns in the training set. This leads to poor prediction for under-represented motion patterns, which can be disastrous from a safety perspective.**

In this work, we extend the Variational Gaussian Mixture model (VGMM) based probabilistic trajectory prediction framework in [1]. We sub-categorize pedestrian trajectories in an unsupervised manner based on their estimated sources and destinations, and train a separate VGMM for each sub-category. We show that the sub-category VGMMs outperform a monolithic VGMM of equivalent complexity, especially for longer prediction intervals. We further analyze the errors made by the two models and the distributions learnt by them, to demonstrate that the sub-category VGMMs better model under-represented motion patterns.

***Index Terms*—Multi-perspective vision, pedestrian trajectory prediction, surround behavior analysis, Variational Gaussian Mixture Models (VGMM)**

I. Introduction

There are two key factors involved in deploying an autonomous vehicle in challenging settings such as city streets. It needs to ensure the safety of its passengers and other occupants of the road, while at the same time navigating smoothly without hindering traffic or causing discomfort to its passengers. Smooth navigation would require a system that is not purely reactive, but capable of planning an optimal and safe path over longer time intervals. The vehicle thus needs to have the capability of predicting the locations of moving entities on the road over long time intervals and assigning a degree of confidence to the predicted locations. Of particular importance is prediction of pedestrian trajectories, since pedestrians are far less constrained in their motion as compared to vehicles, as well as more vulnerable, wherein even slow speed accidents can prove deadly. In this work, we address long-term prediction of pedestrian motion using data captured with vehicle mounted cameras, for a challenging on-road setting.

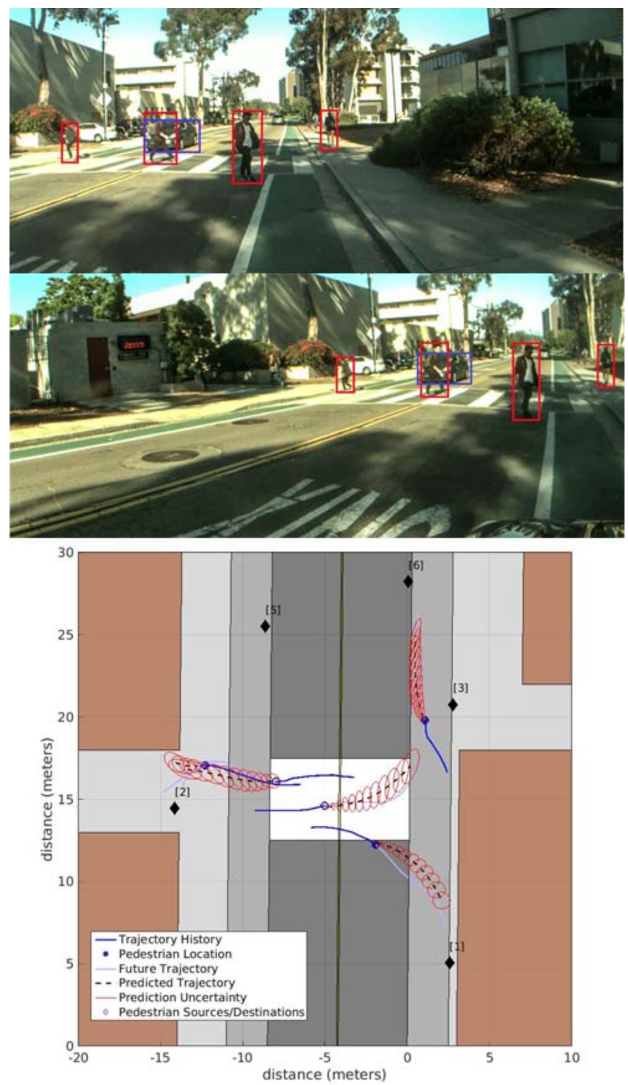


Fig. 1: The goal of this paper is to predict pedestrian trajectories based on their recent track histories and estimated destinations. The images show pedestrian locations in the two camera views at a particular instant. The plot shows locations, clustered destinations, track histories and predicted trajectories in the ground plane

Prediction frameworks based on motion models such as Kalman Filters or Linear Regressors lead to poor results as

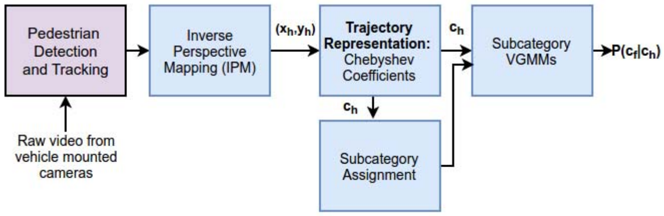


Fig. 2: Overview of the proposed system

the prediction horizon increases, as is also shown in our results section III-B. This is because they fail to identify pedestrians as decision making entities. Some approaches [9], [10], [12] termed 'social forces' explicitly model *human-human* interactions such as collision avoidance, intra-group attraction, as well as attraction to goals. However, these models fail to capture certain 'rules of the road' that affect on-road pedestrian trajectories. For example, pedestrians tend to walk on sidewalks and demarcated crossings, tend not to loiter around on crossings as opposed to other locations and know that vehicles typically yield to pedestrians at uncontrolled crossings. All of these factors can be implicitly modeled through data-driven approaches for trajectory learning [1]–[5]. In particular, as in [1], [2] we formulate trajectory prediction as estimating the posterior distribution over the future trajectory of pedestrians given a snippet of their most recent trajectory history. It must be noted that approaches such as [7], [8] which use scene information to learn a prior distribution on future pedestrian trajectories irrespective of their past trajectory can be considered complementary to our approach.

One of the major drawbacks of data-driven approaches to trajectory prediction is that they are susceptible to bias toward the dominant motion patterns in data, and tend to poorly model some of the rarer motion patterns. This can be particularly disastrous for the on-road setting. While every possible anomalous trajectory cannot be captured by a training dataset, it is desirable for a model to best utilize the rare motion patterns that are present in the data. While one work-around is to increase the model complexity, this often leads to overfitting. A better approach is to leverage sub-category learning [14] , where the data is clustered into disjoint subsets based on a simple theme, and separate probabilistic models are learnt for each subset as opposed to a single monolithic model. Sub-category learning has been extensively used in visual detection, particularly in the context of vehicle detection [15]–[17]. Here, we extend the Variational Gaussian Mixture Model (VGMM) based trajectory prediction framework described in [1] by incorporating subcategories based on estimated sources and destinations of pedestrians in the scene. We work with the VGMM based approach since it was shown to effectively model the highly non-linear trajectories of vehicles at intersections, while being less susceptible to overfitting due to Bayesian estimation.

It is important to evaluate the method using a dataset that is faithful to the setting being considered. Most commonly used datasets [12], [13] involve pedestrian data captured using surveillance cameras in a scene not cohabited by vehicles. We evaluate our method using a new dataset collected using vehicle mounted cameras at an uncontrolled crowded intersection on a university campus. This introduces a number of additional challenges. For example, the view angle of vehicle mounted cameras make the IPM less reliable as compared to surveillance cameras and leads to more pedestrian occlusions in crowded scenes. This makes it harder to capture subtle adjustments to pedestrian paths due to social interactions. An uncontrolled intersection is also perhaps the most challenging setting for an autonomous vehicle [11], since there are no traffic lights, and effective trajectory prediction is the only way to avoid halting traffic or injuring crossing pedestrians.

Thus, the two key contributions of this work could be summarized as:

1) An extension to the VGMM based probabilistic trajectory prediction framework by incorporating subcategories based on pedestrian sources and destinations
2) Evaluation using a challenging on-road dataset collected using vehicle mounted cameras

II. Model

Figure 2 shows the complete system to be used for predicting on-road pedestrian trajectories. The raw video feed from two vehicle mounted cameras is used for pedestrian detection and tracking. This block is outside the scope of this work. We work with the ground truth tracks provided in the dataset described in section III-A. These tracks can in practice be generated using state of the art pedestrian detectors [20], [21] and multi-view trackers [18]. The tracks are projected to the ground plane using an Inverse Perspective Mapping (IPM) to generate ground plane trajectories. At each instant, the most recent 3 second snippet of the ground plane trajectory is used for predicting the future trajectory up to a prediction horizon of 5 seconds. The trajectory snippets are represented in terms of Chebyshev coefficients of their x and y ground plane co-ordinates. The trajectory history is used for assigning the trajectory to a subcategory by predicting the pedestrian's intended destination. A trained VGMM corresponding to the assigned subcategory is then used for estimating the conditional distribution on the future trajectory snippet, given the past trajectory snippet.

A. Problem formulation

The IPM gives ground plane co-ordinates of the pedestrian tracks. For any instant t, the probabilistic trajectory prediction framework may be formulated as estimating the conditional distribution $P(\mathbf{x_f}, \mathbf{y_f}|\mathbf{x_h}, \mathbf{y_h})$ where $\mathbf{x_h}$, $\mathbf{y_h}$ represent the recent track history over the last N_h frames given by:

$$\mathbf{x_h} = (x_{t-N_h+1}, x_{t-N_h+2}, ..., x_t)$$
$$\mathbf{y_h} = (y_{t-N_h+1}, y_{t-N_h+2}, ..., y_t)$$

and $\mathbf{x_f}$, $\mathbf{y_f}$ represent the future trajectory for a prediction horizon of N_f frames

$$\mathbf{x_f} = (x_{t+1}, x_{t+2}, ..., x_{t+N_f})$$
$$\mathbf{y_f} = (y_{t+1}, y_{t+2}, ..., y_{t+N_f})$$

In particular we are interested in $\mathbb{E}[\mathbf{x_f}, \mathbf{y_f}|\mathbf{x_h}, \mathbf{y_h}]$ which gives the predicted trajectory and $\mathbf{cov}(\mathbf{x_f}, \mathbf{y_f}|\mathbf{x_h}, \mathbf{y_h})$ which gives the prediction uncertainty

B. Trajectory Representation

$\mathbf{x_f}, \mathbf{y_f}, \mathbf{x_h}$ and $\mathbf{y_h}$ are represented in terms of their Chebyshev coefficients since they provide a succinct representation of the trajectory whilst also smoothing out noise due to the detector, tracker or IPM. The Chebyshev coefficients also allow for data captured at different frame rates to be represented using a fixed length feature vector. The Chebyshev coefficients $\mathbf{c_{h,x}}$ for $\mathbf{x_h}$ are obtained using the expression:

$$c_{h,x}(n) = \frac{2}{N_h} \sum_{k=0}^{N_h-1} x_h(k) T_n(k)$$

and similar expressions for $\mathbf{c_{h,y}}$, $\mathbf{c_{f,x}}$ and $\mathbf{c_{f,y}}$. Here, T_n is the n^{th} Chebyshev polynomial of the first kind. The first two Chebyshev polynomials are given by

$$T_0(k) = 1$$
$$T_1(k) = k/N_h$$

with the higher order polynomials given by the recursive expression:

$$T_{n+1}(k) = 2kT_n(k) - T_{n-1}(k)$$

The trajectory co-ordinates can be reconstructed using the expression:

$$x_h(k) = \sum_{n=0}^{M} c_{h,x}(n) T_n(k) - \frac{1}{2} c_{h,x}(0)$$

which can be expressed in matrix form as

$$\mathbf{x_h} = \mathbf{T}\mathbf{c_{h,x}}$$

This gives us

$$\mathbb{E}[\mathbf{x_f}, \mathbf{y_f}|\mathbf{x_h}, \mathbf{y_h}] = \mathbf{T}\mathbb{E}[\mathbf{c_{f,x}}, \mathbf{c_{f,y}}|\mathbf{c_{h,x}}, \mathbf{c_{h,y}}]$$
$$\mathbf{cov}[\mathbf{x_f}, \mathbf{y_f}|\mathbf{x_h}, \mathbf{y_h}] = \mathbf{T}\mathbf{cov}[\mathbf{c_{f,x}}, \mathbf{c_{f,y}}|\mathbf{c_{h,x}}, \mathbf{c_{h,y}}]\mathbf{T^T}$$

C. Variational Gaussian Mixture Models (VGMMs)

Variational Gaussian Mixture Models (VGMMs) are used for learning the joint distribution $P(\mathbf{c_h}, \mathbf{c_f})$ over the training data, which can then be used to infer the conditional distribution $P(\mathbf{c_f}|\mathbf{c_h})$ for a new trajectory history $\mathbf{c_h}$. VGMMs are the Bayesian analogue to standard GMMs, where the model parameters, $\{\pi, \mu_1, \mu_2, ..., \mu_K, \mathbf{\Lambda_1}, \mathbf{\Lambda_2}, ..., \mathbf{\Lambda_K}\}$ are given conjugate prior distributions. The prior over mixture weights π is a Dirichlet distribution

$$P(\pi) = Dir(\pi|\alpha_0)$$

The prior over each component mean $\mu_\mathbf{k}$ and component precision $\mathbf{\Lambda_k}$ is an independent Gauss-Wishart distribution

$$P(\mu_\mathbf{k}, \mathbf{\Lambda_k}) = \mathcal{N}(\mu_\mathbf{k}|\mathbf{m_{0_k}}, (\beta_{0_k}\mathbf{\Lambda_k})^{-1})\mathcal{W}(\mathbf{\Lambda_k}|\mathbf{W_{0_k}}, \nu_{0_k})$$

Parameters, $\{\alpha, \mathbf{m_1}, .., \mathbf{m_K}, \beta_1, .., \beta_K, \mathbf{W_1}, .., \mathbf{W_K}, \nu_1, .., \nu_K\}$ of the posterior distributions are estimated using the Variational Bayesian Expectation Maximization algorithm [19]. The predictive distribution for a VGMM is given by a mixture of Student's t-distributions

$$P(\mathbf{c_h}, \mathbf{c_f}) = \frac{1}{sum(\alpha)} \sum_{k=1}^{K} \alpha_k St(\mathbf{c_h}, \mathbf{c_f}|\mathbf{m_k}, \mathbf{L_k}, \nu_k + 1 - d)$$

where d is the number of degrees of freedom of the Wishart distribution and

$$\mathbf{L_k} = \frac{(\nu_k + 1 - d)\beta_k}{1 + \beta_k}\mathbf{W_k}$$

For a new trajectory history $\mathbf{c_h}$, the conditional predictive distribution $P(\mathbf{c_f}|\mathbf{c_h})$ is given by:

$$P(\mathbf{c_f}|\mathbf{c_h}) = \frac{1}{sum(\hat{\alpha})} \sum_{k=1}^{K} \hat{\alpha_k} St(\mathbf{c_f}|\mathbf{c_h}, \hat{\mathbf{m_k}}, \mathbf{L_k}, \nu_k + 1 - d)$$

where

$$\begin{aligned}
\hat{\nu_k} &= \nu_k + 1 - d \\
\hat{\alpha_k} &= \frac{\alpha_k St(\mathbf{c_h}|\mathbf{m_{k,c_h}}, \mathbf{L_{k,c_h}}, \hat{\nu_k})}{\sum_{j=1}^{K} \alpha_j St(\mathbf{c_h}|\mathbf{m_{j,c_h}}, \mathbf{L_{j,c_h}}, \hat{\nu_j})} \\
\hat{\mathbf{m_k}} &= \mathbf{m_{k,c_f}} + \mathbf{\Sigma_{k,c_f c_h}} \mathbf{\Sigma^{-1}_{k,c_h c_h}} (\mathbf{c_h} - \mathbf{m_{k,c_h}}) \\
\hat{\mathbf{L_k}}^{-1} &= \frac{\hat{\nu_k}}{\hat{\nu_k} + d - 2} \left(1 + \Delta_k^T \frac{\mathbf{\Sigma_{k,c_h c_h}}}{\hat{\nu_k}} \Delta_k\right) \mathbf{\Sigma_k}^* \\
\Delta_k &= (\mathbf{c_h} - \mathbf{m_{k,c_h}}) \\
\mathbf{\Sigma_k}^* &= \mathbf{\Sigma_{k,c_f c_f}} - \mathbf{\Sigma_{k,c_f c_h}} \mathbf{\Sigma^{-1}_{k,c_h c_h}} \mathbf{\Sigma_{k,c_h c_f}} \\
\mathbf{\Sigma_k} &= \frac{\hat{\nu_k} + d - 2}{\hat{\nu_k} + d} \mathbf{L_k^{-1}}
\end{aligned}$$

D. Sub-Category VGMMs

Every scene tends to have a handful of locations from which pedestrians enter and exit it. Typically, every pedestrian source also tends to be a viable pedestrian destination. These could be entry points for lanes running across the scene, the entrance to buildings or enclosures or even a point on the sidewalk at the scene boundaries. A pedestrian's source and intended destination seems to have a much greater influence on their trajectories as compared to factors such as social interactions. A pedestrian's destination in particular can be greatly predictive of the trajectory that they follow. We determine the sources and destinations in the scene by clustering the start and endpoints of all trajectories in the training data. This is done by learning a Gaussian Mixture Model over them. Figure 3 shows the start and endpoints of all the trajectories in the training data and the means and variances of the components of the GMM learned over them. The trajectories in the training data can now be sub-categorized by assigning them to a source-destination pair, where the source and destination are determined as the cluster

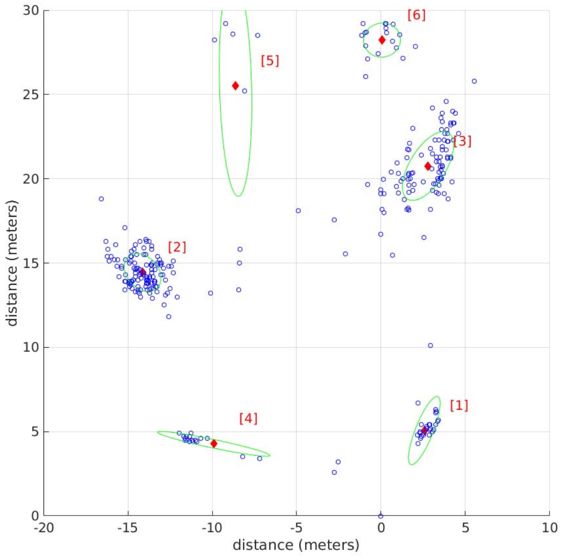


Fig. 3: Clustered sources and destinations of pedestrians

centers closest to the start and endpoints of the trajectory respectively. This assignment can be made deterministically for all the training trajectories since their start and endpoints are known beforehand. We train a separate VGMM for each sub-category of trajectories.

During test time the source and destination need to be estimated based on the available track history for an active track. The source is first assigned as the cluster center closest to the start of the trajectory. However, the destination needs to be predicted, since the end of the track is not known. It is reasonable to expect the correct subcategory VGMM to better fit snippets $\mathbf{c_h}$ of the active trajectory than the other subcategory VGMMs. Also, the most recent snippet can be considered most useful for determining the destination. We thus calculate the marginal probability $P(\mathbf{c_h}|i)$ for all subcategory VGMMs corresponding to the determined source, using the most recent trajectory snippet. The trajectory is assigned to the subcategory i with the maximum value of $P(\mathbf{c_h}|i)$. The marginal probability is given by

$$P(\mathbf{c_h}|i) = \frac{1}{sum(\alpha_\mathbf{i})} \sum_{k=1}^{K} \alpha_{k_i} St(\mathbf{c_h}|\mathbf{m_{i_k,c_h}}, \mathbf{L_{i_k,c_h}}, \nu_{i_k} + 1 - d)$$
$$i^* = \text{argmax}_i P(\mathbf{c_h}|i)$$

III. Evaluation

A. Dataset

We evaluate our system using a challenging dataset captured at a crowded uncontrolled intersection on campus with two vehicle mounted cameras. The data is annotated with pedestrian as well as vehicle bounding boxes in both views as seen in Figure 1. Additionally, track Ids are assigned to pedestrians and cars in both the views. While the raw video is captured at a frame rate of 30 fps, the annotations are available for every 10th frame, effectively making the frame rate of the annotated dataset 3 fps. The homography matrices corresponding to the IPM for the two views is made available for the data. We can see from Figure 1 that using both the camera views is essential to capture the entire scene and completely reconstruct the ground plane tracks in front of the vehicle. For example, the point where a horizontal lane intersects the main street can only be seen in the left view. This point is a major source of pedestrians (source 2 in Figure 3). Similarly, a major portion of the right curb can only be seen in the right view. . The entire dataset consists of 5 minutes of data with 152 pedestrian tracks and 15 vehicle tracks. While outside the scope of this work, the dataset can also be used for analyzing vehicle-pedestrian interactions at an uncontrolled intersection. We evaluate our model on the pedestrian trajectories by using 3-fold cross validation.

TABLE I: Trajectory prediction results: L_2 error in meters vs. prediction horizon for the 3 methods being compared

Prediction Horizon (s)	**Kalman Filter**	**Monolithic VGMM**	**Sub-cat. VGMMs**	**Sub-cat. VGMMs (ground truth)**
0.33	0.32	**0.14**	0.19	0.19
0.67	0.53	**0.28**	0.35	0.35
1	0.75	**0.43**	0.51	0.50
1.33	0.95	**0.58**	0.67	0.65
1.67	1.15	**0.73**	0.81	0.78
2	1.34	**0.86**	0.93	0.89
2.33	1.53	1.03	**1.03**	0.98
2.67	1.72	1.13	**1.12**	1.04
3	1.91	1.25	**1.20**	1.10
3.33	2.10	1.37	**1.28**	1.16
3.67	2.29	1.50	**1.36**	1.23
4	2.49	1.63	**1.46**	1.30
4.33	2.69	1.75	**1.55**	1.38
4.67	2.89	1.88	**1.64**	1.45
5	3.08	2.01	**1.72**	1.50
Average	1.72	1.10	**1.06**	0.97

B. Trajectory Prediction Results

Trajectories are predicted over a prediction horizon of 5 seconds based on 3 seconds of the most recent track history. Trajectory prediction results are evaluated in terms of the L_2 norm of the error between the predicted trajectory and the ground truth vs. the prediction horizon. We compare three different systems for pedestrian trajectory prediction: (1) A constant velocity Kalman Filter as the baseline. (2) A monolithic VGMM consisting of 110 mixture components. (3) Sub-category VGMMs.

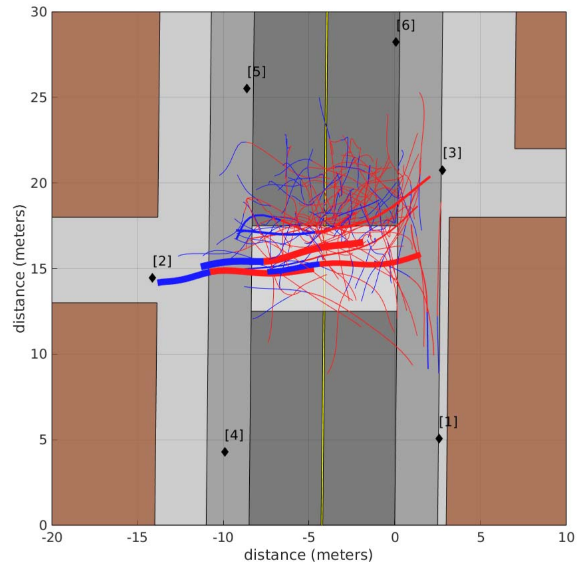
Since clustering yields 11 sub-categories for each of the three folds of the training set, we train each sub-category VGMM with 10 mixture components in order to keep the model complexity comparable to that of the monolithic VGMM. Table I shows the prediction errors for the 3 frameworks considered for different lengths of the prediction horizon, as well as the average L_2 error of the predictions over the complete dataset. To further highlight the effect of sub-categorizing the trajectories, we also report the prediction error using the ground-truth sub-category assignments. These results can be interpreted as the lower bound on prediction error with the sub-category VGMMs.

We observe that VGMM based trajectory prediction yields significant improvements over the Kalman filter based prediction, almost halving the average L_2 error. We also observe that the sub-category VGMMs further reduce the average L_2 error. In particular, we note that the sub-category VGMMs significantly outperform the monolithic VGMM for longer prediction horizons greater than 3 seconds. Finally we note from the last column of the table that if the sub-categories were perfectly assigned for all the test trajectory snippets, we would get a further reduction of about 0.1 m in average L_2 error, with a more pronounced improvement of about 0.2 m for long term predictions, 5 s into the future. This calls for a more sophisticated approach to assigning sub-categories based on observed trajectory histories. With our current approach, we get a subcategory assignment accuracy of about 91.31% on the entire training data

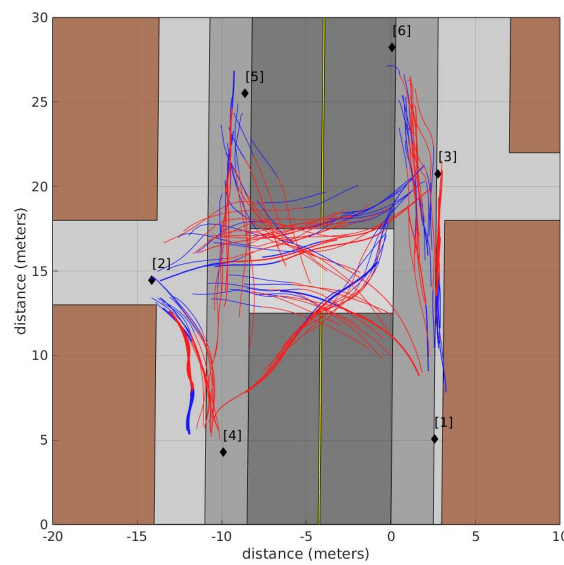
C. Analysis of errors and learnt distributions

Further insight about the results can be gained by analyzing the distributions over the motion patterns learnt by the monolithic and Sub-category VGMMs and by looking at the cases where each method fails. Both the monolithic and Sub-category VGMMs have 110 mixture components in total. The means $\mathbf{m_k}$ of the posterior distributions of the mixture components can be considered representative of the types of motion patterns that the model has learnt. The mixture weights α_k determine how much importance the model assigns to these mixture components. The Chebyshev reconstruction formula from Section II-B can be used to convert these mean values to x and y co-ordinates in the ground plane. Figures 4a and 4b show the mixture component means for the monolithic and sub-category VGMMs plotted in the ground plane. The blue segments represent the history snippets while the red segments represent the prediction snippets. The thickness of the segments is proportional to the weights of the mixture components. We note that the monolithic VGMM has a high concentration of mixture components at the center of the scene, corresponding to the most dominant motion pattern in the dataset: that of pedestrians walking across the demarcated crossing. We also note the large mixture weights assigned to the three components along the middle of the scene. We also note that the monolithic VGMM completely fails to represent certain motion patterns such as movements along the left curb between points 2 and 4, or those between points 3 and 4, across the street. As opposed to this, the subcategory VGMMs learn a more diverse set of motion patterns, given the same model complexity, with relatively equal weight assigned to all motion patterns.

Analysis of the errors made by each model re-affirms the insights gained from analyzing the mixture components. Figure 5 shows average L_2 errors in meters for each trajectory in the test set for one fold of the cross validation for the monolithic and sub-category VGMMs. We note that the monolithic VGMM makes many more drastic errors (red lines) than the sub-category VGMM on the test set. In particular, these errors correspond to the rarer motion patterns discussed in the previous paragraph, that the model fails to capture. For example, we see that the monolithic VGMM makes lots of errors on test trajectories between points 2 and 4. This clearly shows that learning sub-categories leads to better utilization of the training data and better generalizability of the model.



(a) Mixture component means for monolithic VGMM

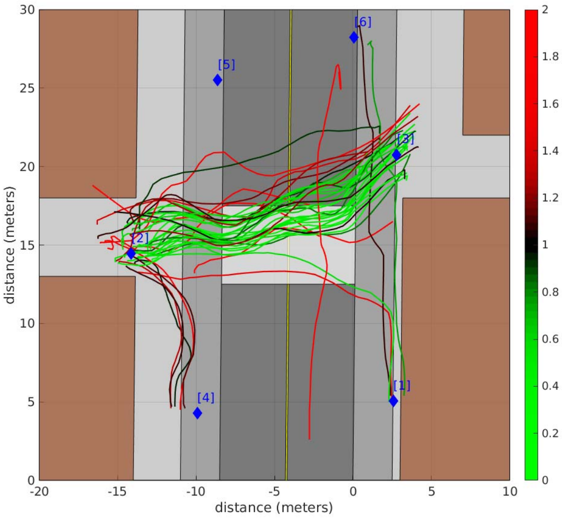


(b) Mixture component means for Sub-Category VGMM

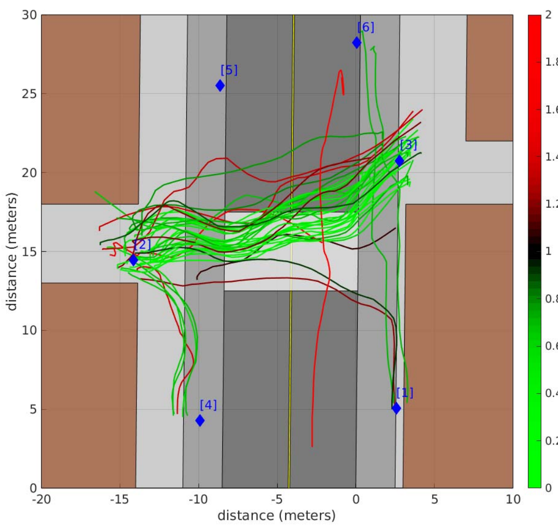
Fig. 4: Analysis of learnt distributions: The sub-category VGMM distribution captures a more diverse set of motion patterns as compared to the monolithic VGMM

IV. Conclusion and Future Work

In this work we showed the viability of the VGMM based probabilistic trajectory prediction framework for the challenging task of long term prediction of pedestrian trajectories at an uncontrolled intersection on a city street. We showed that the probabilistic framework significantly outperforms the Kalman Filter baseline. We also showed that clustering the



(a) Errors for monolithic VGMM



(b) Errors for Sub-Category VGMM

Fig. 5: Analysis of errors made by the models: Test set errors for each trajectory show that the monolithic VGMMs make more errors for motion patterns that are away from the dominant path, along the crossing. The subcategory VGMMs make fewer such errors

training data into subcategories based on pedestrian sources and destinations, and learning a separate model for each sub-category leads to better prediction results over longer prediction horizons as compared to a monolithic model. An analysis of learnt distributions and errors made by the two models shows that the sub-category VGMMs better represent the motion patterns in the training data, especially the rarer motion patterns, and leads to better generalization on the test set. Future work would focus on incorporating interactions between pedestrians and vehicles into the framework.

ACKNOWLEDGMENT

The authors would like to thank their colleagues at the Lab for Intelligent and Safe Automobiles (LISA), UC San Diego for their support and useful discussions and feedback.

REFERENCES

[1] J.Wiest, M. Hffken, U. Kreel, and K. Dietmayer. Probabilistic trajectory prediction with gaussian mixture models. In *Intelligent Vehicles Symposium (IV)*, pp. 141-146. IEEE, 2012.

[2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 961-971. 2016.

[3] B. T. Morris, and M. M. Trivedi. Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach. *IEEE transactions on pattern analysis and machine intelligence 33*, no. 11 (2011): 2287-2301.

[4] B. Zhou, X. Wang, and X. Tang. Random field topic model for semantic region analysis in crowded scenes from tracklets. *In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3441-3448. IEEE, 2011.

[5] K. Kim, D. Lee, and I. Essa. Gaussian process regression flow for analysis of motion trajectories. *In Computer vision (ICCV), 2011 IEEE international conference on,* pp. 1164-1171. IEEE, 2011.

[6] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund. Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations. *In Intelligent Vehicles Symposium (IV), 2015 IEEE,* pp. 330-335. IEEE, 2015.

[7] K. Kitani, B. Ziebart, J. Bagnell, and M. Hebert. Activity forecasting. *Computer Vision-ECCV 2012* (2012): 201-214.

[8] T. Gandhi, and M. M. Trivedi. Image based estimation of pedestrian orientation for improving path prediction. *In Intelligent Vehicles Symposium, 2008 IEEE,* pp. 506-511. IEEE, 2008.

[9] D. Helbing, and P. Molnar. Social force model for pedestrian dynamics. *Physical review E 51*, no. 5 (1995): 4282.

[10] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going?. *In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 1345-1352. IEEE, 2011.

[11] M. S. Shirazi, and B. T. Morris. Looking at Intersections: A Survey of Intersection Monitoring, Behavior and Safety Analysis of Recent Studies. *IEEE Transactions on Intelligent Transportation Systems* 18, no. 1 (2017): 4-24.

[12] S. Pellegrini, A. Ess, K. Schindler and L. Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 261-268). IEEE.

[13] A.Lerner, Y. Chrysanthou and D.Lischinski. Crowds by example. *In Computer Graphics Forum* (Vol. 26, No. 3, pp. 655-664). Blackwell Publishing Ltd.

[14] S. Divvala, A. Efros, and M. Hebert. How important are "deformable parts in the deformable parts model?. *In Computer VisionECCV 2012. Workshops and Demonstrations*, pp. 31-40. Springer Berlin/Heidelberg, 2012.

[15] E. Ohn-Bar, and M. M. Trivedi. Learning to detect vehicles by clustering appearance patterns. *IEEE Transactions on Intelligent Transportation Systems* 16, no. 5 (2015): 2511-2521.

[16] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. *In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1-8. IEEE, 2008.

[17] Y.Xiang, W. Choi, Y. Lin, and S. Savarese. Subcategory-aware convolutional neural networks for object proposals and detection. *arXiv preprint arXiv:1604.04693* (2016).

[18] J.V.Dueholm, M. S. Kristoffersen, R. K. Satzoda, T. B. Moeslund, and M. M. Trivedi. Trajectories and Maneuvers of Surrounding Vehicles with Panoramic Camera Arrays. *IEEE Transactions on Intelligent Vehicles* 1, no. 2 (2016): 203-214.

[19] C. M. Bishop. Pattern recognition. *Machine Learning* 128 (2006): 1-58.

[20] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. *In European Conference on Computer Vision*, pp. 354-370. Springer International Publishing, 2016.

[21] R. Rajaram, E. Ohn-Bar, M. Trivedi. Looking at pedestrians at different scales: A multiresolution approach and evaluations. *IEEE Transactions on Intelligent Transportation Systems*, 2016.