Towards Group-Level Social Activity Recognition for Mobile Robots

Billy Okal Kai O. Arras

***Abstract*— Robots in human populated environments need to perceive and understand the social context they are in for a variety of tasks. One key element to this understanding are group-level activities of people in the vicinity of the robot. In this paper, we employ supervised learning to recognize such activities from a robot-centric first-person perspective for the task of navigation in human crowds. We develop and compare several feature descriptors that encode spatiotemporal motion information of surrounding people using histograms and use Random forests for classification. Extensive comparative experiments in simulation reveal that adding additional information such as velocity and speed to the histograms gives best performance given that some activities are indistinguishable by mere density counts. We also observe that directional information in velocity dominates speed. We obtain a 77% classification accuracy for five activity classes.**

I. Introduction

Group-level activity recognition and analysis is a key skill for assistive and service robots in human environments as it provides the capacity to reason about human behavior and derive suitable robot behaviors. Robots with this level of understanding can, for example, generate socially compliant behavior by mirroring activities or plan actions that are particularly task-efficient for the robot while being socially normative at the same time.

Previous work in this area is typically carried out in the computer vision community motivated by applications such as monitoring or surveillance. Such work usually make use of overhead cameras that overlook the entire scene [1], [2], [3], [4], [5]. Such conditions are not met with a mobile robot operating in-scene and perceiving the surrounding people from a first-person perspective as shown in [6] although no robots are used. Thus, in this paper, we develop methods and models that encode the relevant information on the surrounding people in a robot-centric way and perform extensive experiments across several features and spatial descriptors. We focus on activity classes that are common in public spaces such as queueing, standing in a group, shopping, walking, walking in groups, walking in flows, etc. as shown in Fig. 2. The ability to understand social context is also important in the application scenario that also motivates this work. In this scenario a mobile service robot is deployed to guide groups of delayed transfer passengers through the crowds of a busy airport within EU FP7-project SPENCER[1].

In related work, typical approaches compute large numbers of low-level features on visual data over multiple frames and employ different coding schemes like bag-of-words to derive a reduced feature representations which is then used for classification [7], [8], [2]. The importance of social role in such activity discovery was studied in [9] albeit only considering pairwise interactions. Choi *et al.* [10] propose a novel feature descriptor to encode spatial relations between people for the task of group activity recognition. The descriptor studied here is based on this idea. When visual information is not available as in our case, features need to be extracted from motion state primitives such as poses and orientations along trajectories. Non-visual approaches to human activity recognition typically uses wearable sensors such as accelerometers as in [11]. However, this group of works aims at recognizing activities of individuals only.

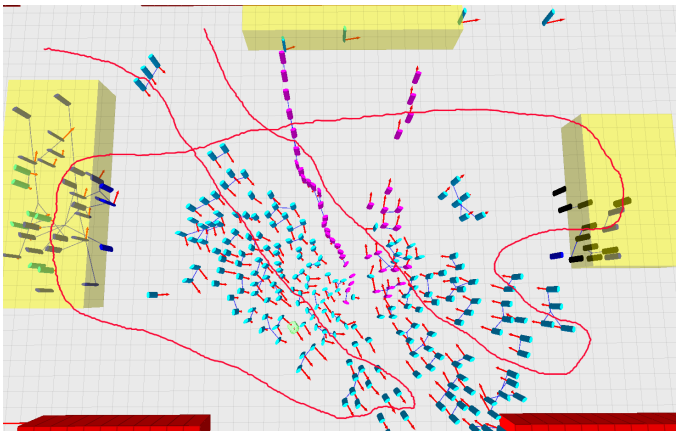

Fig. 1: Airport-like scenario with 175 simulated agents engaged into the activities standing (black), walking (cyan), walking in groups (linked with blue lines), shopping (blue within yellow blocks) or queuing (pink). Yellow blocks indicate attractions (shopping, info-desks etc) while walls are red blocks. Robot is shown in green while its sample path is shown in red.

The paper is structured as follows: section II introduces our approach including the formulation we use, section III desribes our simulator, section IV discusses the experiments we undertook in this work, section IV-B shows our results and finally in section V we give some conclusions and future outlook.

II. Problem Statement and Approach

Our goal is to recognize different social activities of people in the vicinity of the robot, and we assume to have only very limited information, namely their relative position, orientation and motion state. This assumption is realistic

Billy Okal and Kai. O. Arras are with the Social Robotics Lab, University of Freiburg {okal, arras}@cs.uni-freiburg.de

This work has been partly supported by the European Commission under contract number FP7-ICT-600877 (SPENCER).

[1]http://www.spencer.eu

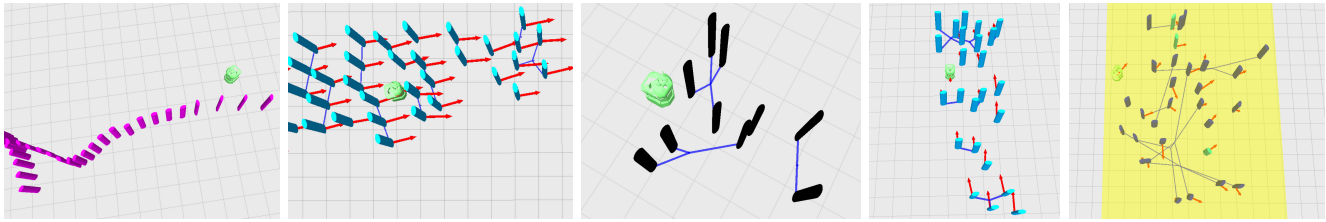

Fig. 2: The different group activities considered here (from left to right): queueing, moving with flow, standing, moving against flow and shopping. Connection lines between agents (in blue) show pedestrians that belong to groups while red arrows show velocity vectors.

given the capacities of today people tracking systems from on-board sensors and it allows us to study the problem under particularly difficult conditions.

A. Problem Formulation

Consider a robot traveling between two points in a populated environment. The environment is modeled as a bi-dimensional Euclidean space $\mathcal{C}$ which is the union of three sets, static obstacles $\mathcal{C}^{\mathsf{s}}_{\mathsf{obs}}$, dynamic obstacles (including people) $\mathcal{C}^{\mathsf{d}}_{\mathsf{obs}}$ and free space $\mathcal{C}_{\mathsf{free}}$. Without loss of generality we can assume all dynamic obstacles to be human pedestrians and write $\mathcal{C}^{\mathsf{d}}_{\mathsf{obs}}$ as a set of N pedestrians $\mathcal{C}^{\mathsf{d}}_{\mathsf{obs}} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N\}$. In our case, each pedestrian is represented by a position and velocity state vector $\boldsymbol{p} = [x_p, y_p, \dot{x}_p, \dot{y}_p]^T$ which is also the representation of the robot $\boldsymbol{r} = [x_r, y_r, \dot{x}_r, \dot{y}_r]^T$. Groups of pedestrians are given by sets $\mathcal{G}_j = \{\boldsymbol{p}_i \mid \psi(j,i) = 1\}$ with $\psi(\cdot,\cdot)$ being an arbitrary group membership function.

We define a neighborhood $\mathcal{N}$ to be a locality centered at the robot or any other agent. We will consider different methods to define this neighborhood in the next subsection. Our group activity recognition task can then be framed as extracting relevant features from sequences of relative pose vectors $\xi_{1..T}$ of length T

$$\xi_{1..T} \quad = \quad \begin{bmatrix} \boldsymbol{p}_1^1, \ldots, \boldsymbol{p}_{N_1}^1 \\ \boldsymbol{p}_1^2, \ldots, \boldsymbol{p}_{N_2}^2 \\ \vdots \\ \boldsymbol{p}_1^T, \ldots, \boldsymbol{p}_{N_T}^T \end{bmatrix}. \qquad (1)$$

Note that the number of pedestrians in the neighborhood, N_t, may change at any time t when people enter or leave $\mathcal{N}$.

A single frame at time t consists of the positions and motion states of the robot and the pedestrians in the robot's vicinity as shown in Fig. 3. We will consider neighborhoods centered at both the robot and at other agents and refer to the combined state information of all agents in $\mathcal{N}$ as *context* or *social context* of the focal agent.

B. Features

We now seek features that capture the relevant attributes and relationships of and between pedestrians, as well as the relationships between pedestrians and robot. Our approach utilizes feature descriptors that involve computing histograms in the neighborhood of the focal agent (robot or any other pedestrian). This method is inspired by the shape context descriptor [12], popular in computer vision, as well as the spatio-temporal volume descriptor proposed in [10]. These descriptors are chosen ad-hoc and do not incorporate domain knowledge from the cognitive or social sciences. It appears, however, necessary to account for knowledge from these fields which is why, in this work, we extend those features to exploit empirical findings in the way the histograms are computed and the information that they represent.

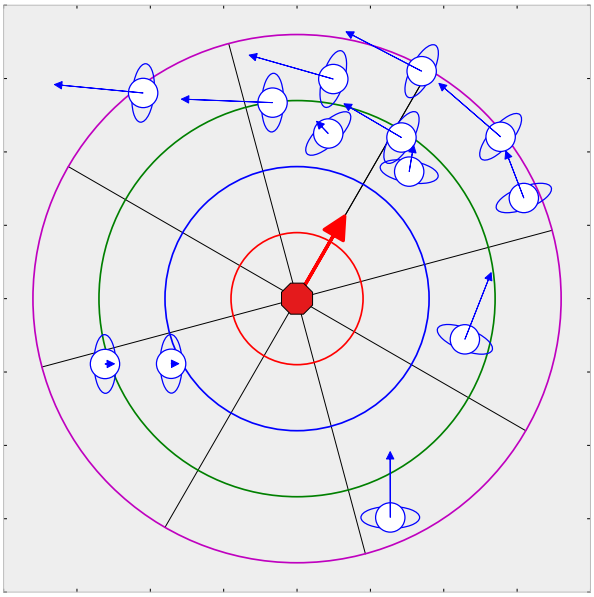

Fig. 3: Neighborhood $\mathcal{N}$ and example context with the robot being the focal agent. Uniform binning, arrows show velocity vectors.

The basic form of our feature is computed as follows: given the set of state vectors of surrounding agents $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_t}\}$ and the focal agent (the robot state $\boldsymbol{r}$ for example), we compute the relative distances and orientations of surrounding pedestrians with respect to the focal agent as $\boldsymbol{\rho} = \{\rho_1, \ldots, \rho_{N_t}\}$ and $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_{N_t}\}$. We then histogram the $\boldsymbol{\rho}$'s and $\boldsymbol{\theta}$'s into different bins. Like the shape context descriptor, we use uniformly sized bins along $\boldsymbol{\theta}$ but have three types of bins along $\boldsymbol{\rho}$, namely uniform, Proxemics, and anisotropic bins described next:

1) Uniform binning: This descriptor divides the radius vector into K equally sized intervals, inducing K concentric

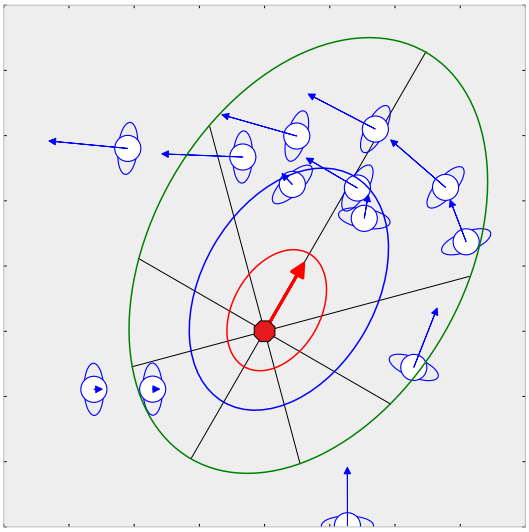

Fig. 4: Proxemics binning descriptor. The elliptical regions correspond to the intimate (red), personal (blue), social (green), and public (beyond green) spaces.

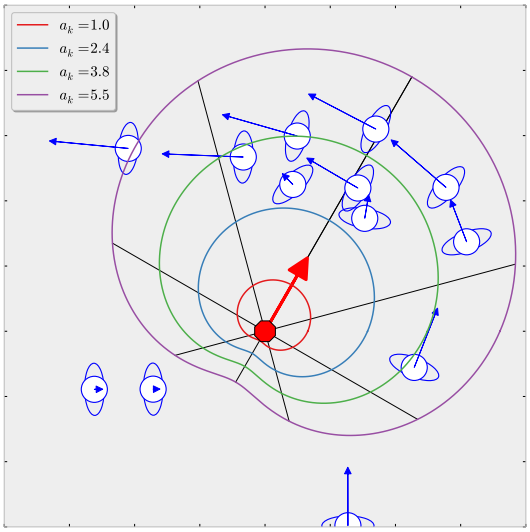

Fig. 5: Feature descriptor with anisotropic binning. The regions are obtained by varying size parameter a

circles. We found $K = 5$ to be best in our experiments.

2) Proxemics binning: Uses the elliptic Proxemics regions as illustrated in Fig 4 as proposed by Hall [13]. The Proxemics theory is an empirical model that relates interpersonal body distance during interaction. The four different spaces are called *intimate* (up to $0.45m$ from the focal agent), *personal* (between $0.45m$ to $1.2m$), *social* (between $1.2m$ to $3.6m$) and *public* ($3.6m$ or more).

3) Anisotropic binning: This model is proposed by Helbing *et al.* [14] and describes the observation that influences are not isotropic given the limited field of view of humans. Instead of elliptic regions as in Hall's Proxemics model, this model scales influences from other agents by an anisotropic factor defined as;

$$a \cdot \exp\left(\frac{r_{ij} - d_{ij}}{b}\right)\boldsymbol{n}_{ij}\left(\lambda + (1-\lambda)\frac{1+\cos(\varphi_{ij})}{2}\right).$$

where index i denotes the focal agent and index j the interacting pedestrian. a, b are parameters that control the size of the regions and λ defines the strength of the anisotropic factor that controls the region's circularity. $\boldsymbol{n}_{ij}$ is the normalized vector pointing from pedestrian j to the focal agent i, φ_{ij} is the relative orientation of pedestrian j with respect to the line through the centers of the focal agent i and pedestrian j, r_{ij} is the sum of radii of the focal and interacting agents (assuming circular and similar body shape), and d_{ij} is the Euclidian distance between the two agents. The regions are illustrated in Fig 5.

The shape context descriptor and the descriptor proposed in [10] is obtained by counting the occurrences of events (pixels along the shape or pedestrians) that falls into each respective bin. In this way, each bin holds a density estimate of those events. Here, we exploit more information on the surrounding pedestrians than relative position and incorporate also information on their relative velocity direction and magnitude. Similar features have been to encode coherent motion indicators as found in large-scale empirial studies by [15].

Concretely, for each bin, we compute the average relative direction (called *direction*) and the average relative velocity magnitude (called *speed*) of all pedestrians that fall into that bin. This is done for all three descriptor shapes described above. The combination of density, direction and speed distributions over the bins form the basis feature representation in this work.

One goal of this paper is finding the best coding, modeling and learning schemes for the task of group-level activity recognition for mobile robots. Thus, we systematically compare the seven feature combinations *density*, *direction*, *speed*, *density–direction*, *density–speed*, *speed–direction* and *density–direction–speed*.

Many group-level activities unfold over time and can only be distinguished over sequences of observations. Thus, instead of considering only the context of a focal agent at a single frame, we integrate information over time by tracking the motion state of the surrounding pedestrians and aggregating feature values over the last T frames. Short state sequences of this kind are typically called *tracklets*. In the experiments, we will evaluate the contribution of this form of temporal smoothing against the limit case $T = 1$ (no smoothing).

Finally, we also perform smoothing over descriptor space to diminish quantization artefacts from the histogram binning. Without smoothing, small changes in the position of a pedestrian at the edge of two bins may have large and unwanted effects onto the feature encoding. We smooth the raw feature values by convolution with a Gaussian kernel. Fig 6 illustrates an example feature histogram with uniform binning.

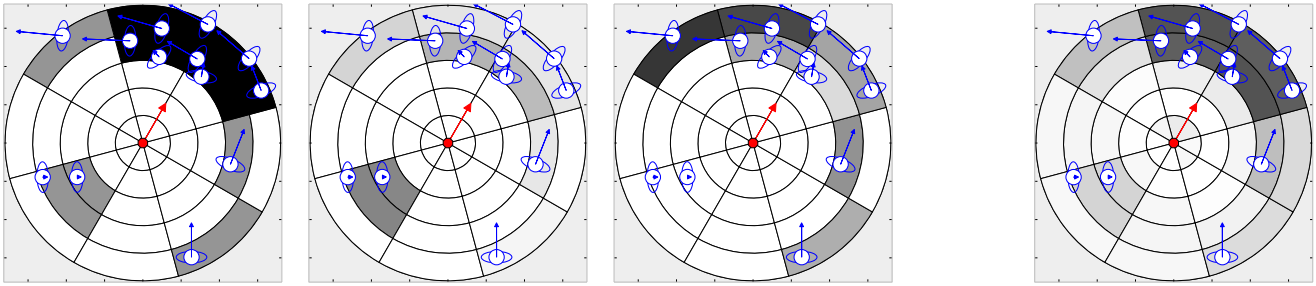

Fig. 6: Example feature histograms using uniform binning. **Left** (from left to right): density, direction and speed respectively. **Right:** Gaussian smoothing example shown with the left-most density histogram. Robot is shown in red while arrows indicate direction and speed.

C. Classification model

At every time index t, the descriptors produce feature vectors $\boldsymbol{\phi}_t$ that contain the stacked values of all flattened histogram bins. As we take a supervised learning approach, we also have discrete activity labels y_t from the label set $\{1, \dots k\}$ for k different activities. We seek to learn a model $\mathcal{M}$ so as to be able to predict activities from unseen feature vectors. Once the model is learned, we estimate the probability distribution over activity labels, $P(A_t \mid \mathcal{M}, \boldsymbol{\phi}_t)$, for example over observation sequences from a test trajectory of the robot through the environment. The predicted activity at frame t is then found in a maximum a posteriori fashion to be the one with the highest probability.

We use Random forests [16] for recognizing the activities shown in Fig. 2. Random forests is an ensemble classification method that combines multiple tree estimators trained using different subsets of the training data in a 'bagging' fashion. Unlike other bagging approaches, random forests strive to decorrelate the base estimators and hence reduce variance based on randomly chosen subsets of input variables and training samples which usually gives good predictive accuracy as demonstrated in many applications. For a rigorous treatment of random forests theory, we refer the reader to [17]. We make use of the standard implementation of random forest algorithm available from the scikit-learn library [18] and determine the classifier parameters (number of estimators and split criteria) via k-fold cross validation.

III. Simulator

One goal of this paper is to systematically study and compare different feature coding and modeling schemes for which we require repeatability of experimental conditions. Because this is hard to find in real-world experiments, we make use of simulation to generate data for our experiments. The integration with a person tracker based on real sensory data is planned as future work. For now, the simulator allows us to analyze the problem in controlled conditions and receive upper performance bounds.

Our pedestrian simulator is an extension of the PedSim simulator [19] to which we have added several functionalities such as the group-level activities shown in Fig. 2. Pedestrian motion is generated via a series of waypoints, guided by the social force model [14] which posits that motion is governed by a combination of three forces: a social force from other agents, a repulsive force from obstacles and a desired force in goal direction. We have extended the system to also generate group behavior using an extension of the social force model by Moussaid *et al.* [20] which adds three group forces: a group coherence force, a group repulsion force to keep distances between group members and a gaze force for orienting heading towards the center of the group. Alltogether the pedestrians are driven by the equations of motion specified in Eq. 2, the reader is referred to [14], [20] for details of the derivations. The individual forces are weighted to tune their relative influences.

$$\begin{aligned} \frac{d\boldsymbol{v}_i}{dt} &= \boldsymbol{f}^i_{\text{desired}} + \boldsymbol{f}^i_{\text{obstacle}} + \sum_j \boldsymbol{f}^{ij}_{\text{social}} \\ &\quad + \boldsymbol{f}^i_{\text{gaze}} + \boldsymbol{f}^i_{\text{repulsion}} + \boldsymbol{f}^i_{\text{coherence}} \end{aligned} \tag{2}$$

The agents are spawned with different maximum speeds which are drawn from a Guassian distribution given by $\nu \sim \mathcal{N}(\mu = 1.34, \sigma = 0.26)$ whose parameters were taken from [21]. We control the sizes of groups by drawing samples from a Poisson distribution $|\mathcal{G}| \sim \text{Poi}(\lambda)$ subject to the number of available agents. Queues are modeled as special waypoints with waiting times t drawn from an Erlang distribution (implemented as Gamma distribution with shape parameter as an integer) $t \sim \Gamma(\alpha, \beta)$ while the distance between queueing pedestrians d is drawn from a uniform distribution $d \sim \text{Uni}(a, b)$.

Shops are modeled as rectangular attraction regions so that a pedestrian switches from walking to shopping by a switch variable s drawn from a Bernoulli distribution so that $s \sim \text{Bern}(p)$ whose parameter p encodes a shop's attraction strength. Once in shopping mode, a pedestrian wanders around the shop by moving to randomly selected poses in varying intervals of time as if the agent was browsing through articles. Poses are drawn from $\text{Uni}(a, b)$ parameterized by the shop dimensions.

We have determined the parameters of the distributions to reproduce realistic human behavior to the best of our knowledge. However, we can expect that the exact choices

of these values have little effect on the studied aspects in this paper. Our simulator runs on a simulated clock at 25Hz.

IV. Experiments

The objectives of the experiments are twofold: first, to investigate the impact of the three different descriptors *uniform*, *Proxemics* and *anisotropic* on the activity recognition task, and second, to compare the baseline descriptor that includes only density counts with our extended descriptor that also incorporates relative direction and speed features.

A. Setup

Using the simulator described in Sec. III, we have developed scenarios with the five activities *walking in a flow*, *walking against a flow*, *queueing*, *standing* and *shopping* as shown in Fig. 2. For learning, we collect training data in a set of scenes where each behavior is presented individually. The robot moves around the scene encountering agents with the different activities and perceives the scene from a robot-centric perspective using a limited sensor field of view defined to be a circle with radius R. Along those trajectories, it maintains sequences of relative pose vectors ξ_t from which the different features are computed. This phase produces the activity classification model $\mathcal{M}$.

In the test phase, the robot moves through a novel scene that simulates a complex airport-like environment with 175 agents in which all activities are present at the same time (see Fig. 1). In each step, the model $\mathcal{M}$ is used to classify the activity of every pedestrian within the robot's field of view by making each of them the focal agent and classifying its social context. We denote the set of visible pedestrians at every step $\mathcal{V} = \{\boldsymbol{p} \mid \text{dist}(\boldsymbol{p}, \boldsymbol{r}) \leq R\}$ where $\text{dist}(\boldsymbol{p}_i, \boldsymbol{p}_j)$ is the Euclidean distance of agents $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ based on their positions.

As performance metric, we compute the frame accuracy as the ratio of correct label matches to the number of classified pedestrians

$$acc = \frac{\sum_{\boldsymbol{p} \in \mathcal{V}} \mathbb{1}(\hat{a} = a)}{|\mathcal{V}|} \tag{3}$$

where $\hat{a}$ is the predicted activity and a the true activity of pedestrian $\boldsymbol{p}$, respectively, and $\mathbb{1}(\cdot)$ the indicator function. We do this over an entire test trajectory and compute the trajectory accuracy as the average over the frame accuracies. We compute averages over 5 exemplar trajectories and use 5-fold cross validation to find the optimal parameters of our classification model.

We also show confusion matrices illustrating how our model performs in distinguishing the five activities from each other. We show results obtained by considering single frames, $T = 1$, and by aggregating multiple frames together, $T = 10$, to see how the consideration of the temporal information which is inherent in the activities impacts the classification performance.

We use 45 base estimators (trees) for the Random forest classifier and entropy as the tree splitting criterion. In the uniform binning model, we use a radius of $R = 3.6m$ as descriptor size which is inspired by Proxemics theory as the radius of the social space. In the two other binning models, we scale relevant parameters so that the furthest pedestrians covered in the space is within the $3.6m$ radius. We then use 5 bins for $\boldsymbol{\rho}$ and 8 bins for $\boldsymbol{\theta}$ to give a 40-dimensional vector of either density, direction and speed features. In case of combinations, the feature vector is 80 or 120-dimensional.

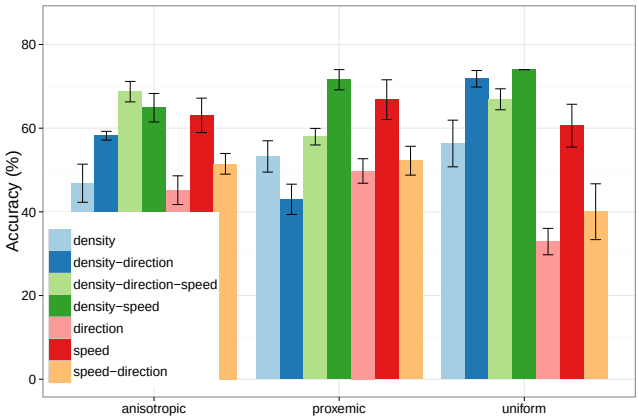

Fig. 7: Frame accuracy for every feature combination on the test airport scene **without** temporal smoothing.

B. Results

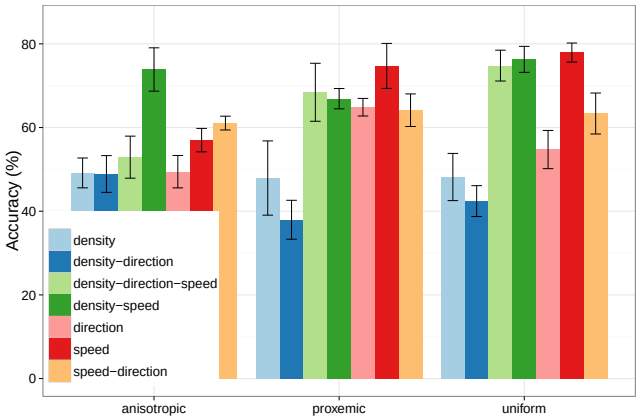

Fig. 8: Frame accuracy for every feature combination on the test airport scene **with** temporal smoothing of 10 frames.

The average frame accuracies per feature averaged over 5 test trajectories and standard deviations on the test airport scene are shown in Fig. 8 with temporal smoothing and Fig. 7 without temporal smoothing. We generally observe a good classification accuracy in particular when the *density–direction–speed* features are considered. We also observe increased accuracy for most combinations when integrating information over time which is a clear indication that the five activities possess inherent temporal information that needs to be accounted for. The corresponding confusion matrices are shown in Fig. 9 and similar effect of smoothing it observed.

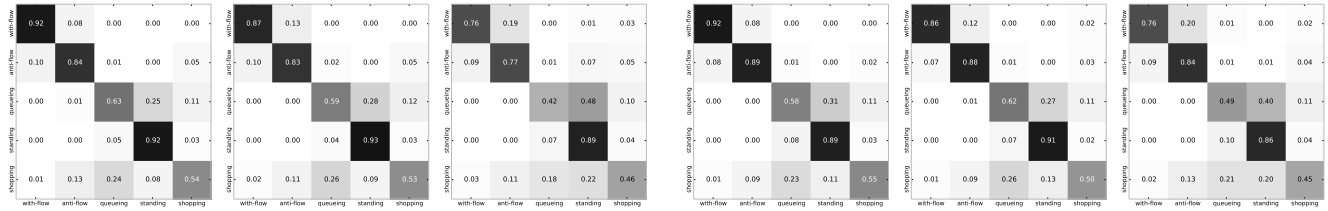

Fig. 9: Confusion matrices for different binning spaces and temporal smoothing settings. The **left** three figures show the confusion percentages for the uniform, proxemics, and anisotropic binning model for the best feature combination of density-velocity-speed without temporal smoothing ($T = 1$). The **right** three figures show the matrices for the same three binning models and the same feature combination with temporal smoothing ($T = 10$).

V. Conclusion

In this paper we have studied the problem of recognizing group-level social activities for mobile robots that perceive surrounding humans only through trajectory data of positions and velocities. The goal of this paper was to frame the problem and compare different feature coding, modeling and learning schemes in large-scale simulation in order to find the best-performing method in terms of classification accuracy.

We have extended related work by two novel histogram descriptors and velocity-related features and demonstrated their viability in our experiments. Temporal smoothing improves recognition performance which we attribute to the fact that most group-level activities unfold over time. We observed that density count features are limited in terms of accuracy as some activities like moving with or against a flow of pedestrians are only distinguishable through relative direction information. We also observed that using non-uniform binning regions improves recognition as in the case with Proxemics and anisotropic binning which focus on 'relevant' parts of the surrounding.

In future work, we plan to integrate the classification procedure with a people tracker from either 2D laser data or RGB-D data and study the robustness of the method with respect to errors in perception. We also intend to learn temporal graphical models that are able to represent more complex temporal group-level activities and perform spatial smoothing over neighboring pedestrians as well as developing feature descriptors to take advantage of other social cues.

References

[1] T. Lan, Y. Wang, W. Yang, and G. Mori, "Beyond actions: Discriminative models for contextual group activities." in *Proceedings of the Conf. on Neural Information Processing Systems (NIPS)*, vol. 4321, 2010, pp. 4322–4325.

[2] R. Li, P. Porfilio, and T. Zickler, "Finding group interactions in social clutter," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[3] B. Ni, S. Yan, and A. Kassim, "Recognizing human group activities with localized causalities," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1470–1477.

[4] K. H. Lee, M. G. Choi, Q. Hong, and J. Lee, "Group behavior from video: A data-driven approach to crowd simulation," in *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, 2007, pp. 109–118.

[5] W. Choi, K. Shahid, and S. Savarese, "Learning context for collective activity recognition," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2011, pp. 3273–3280.

[6] A. Fathi, J. K. Hodgins, and J. M. Rehg, "Social interactions: A first-person perspective," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1226–1233.

[7] W. Choi and S. Savarese, "A Unified Framework for Multi-target Tracking and Collective Activity Recognition," in *European Conference on Computer Vision*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 215–230.

[8] T. Lan, Y. Wang, W. Yang, S. N. Robinovitch, and G. Mori, "Discriminative Latent Models for Recognizing Contextual Group Activities," *IEEE Transactions on Pattern Analysis and Machine Intell. (PAMI)*, vol. 34, no. 8, pp. 1549–1562, 2012.

[9] V. Ramanathan, B. Yao, and L. Fei-Fei, "Social Role Discovery in Human Events," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[10] W. Choi, K. Shahid, and S. Savarese, "What are they doing?: Collective activity classification using spatio-temporal relationship among people," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 2009, pp. 1282–1289.

[11] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors," in *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, vol. 2, 2001, pp. 747–752 vol.2.

[12] S. Belongie, J. Malik, and J. Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in *Proceedings of the Conf. on Neural Information Processing Systems (NIPS)*, vol. 2, 2000, p. 3.

[13] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannan, A. R. Diebold Jr, M. Durbin, M. S. Edmonson, J. Fischer, D. Hymes, S. T. Kimball, *et al.*, "Proxemics," *Current anthropology*, pp. 83–108, 1968.

[14] D. Helbing and P. Molnar, "A social force model for pedestrian dynamics," *Physical Review E*, vol. 51, pp. 4284–4286, 1995.

[15] Z. Ycel, F. Zanlungo, T. Ikeda, T. Miyashita, and N. Hagita, "Modeling indicators of coherent motion," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 2134–2140.

[16] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT Press, 2012.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research (JLMR)*, vol. 12, pp. 2825–2830, 2011.

[19] C. Gloor, "PedSim: A Microscopic Pedestrian Crowd Simulation System," http://pedsim.silmaril.org, 2012, [Online; accessed Feb-2014].

[20] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The walking behaviour of pedestrian social groups and its impact on crowd dynamics," *PloS one*, vol. 5, no. 4, p. e10047, 2010.

[21] D. Helbing, P. Molnar, I. J. Farkas, and K. Bolay, "Self-organizing pedestrian movement," *Environment and planning B*, vol. 28, no. 3, pp. 361–384, 2001.