Detecting Anomalous Human Interactions using Laser Range-finders

Anand Panangadan Computer Science Department University of Southern California Los Angeles, CA 90089 anandvp@robotics.usc.edu Maja Matarić
Computer Science Department
University of Southern California
Los Angeles, CA 90089
maja@robotics.usc.edu

Gaurav Sukhatme
Computer Science Department
University of Southern California
Los Angeles, CA 90089
gaurav@robotics.usc.edu

Abstract—We present a laser range-finder-based system for tracking people in an outdoor environment and detecting interactions between them. The system does not use identities of people for tracking. Observed tracks are automatically segmented into individual activities using an entropy-based measure (Jensen-Shannon divergence [12]). Two people situated close to each other throughout the duration of an activity represents an interaction. The observed activities are combined using a hierarchical clustering algorithm to generate a representative set. The frequency of occurrence of these activities is modeled by a Poisson distribution. During the monitoring phase, this model is used to compute the probability of observing the detected activities and interactions; an anomaly is flagged if this probability falls below a threshold. Experimental results from an outdoor courtyard environment are described where the system indicates anomalies when there is a sudden increase in the number of people in the environment or in the number of interactions. This detection occurs without giving the system any a priori concepts of space occupancy.

I. Introduction

The ability to detect and model interactions between people has many applications, especially in surveillance where the goal is to detect anomalous behavior. Such behavior may be defined as activity that occurs more or less frequently than is the norm for a given environment. Consequently, a system capable of detecting anomalous behavior must first identify the different types of activities for the given environment, including interactions between people. Next, it must build a model of the normal patterns of those activities. Anomalous behavior is then identified as activity not fitting the learned model.

We use laser range-finders for tracking people in an outdoor environment [7]. Unlike vision systems, laser range-finders cannot be readily used to uniquely identify people. The sequences of positions obtained from the laser-based tracker represent a series of different activities performed by a single person. We segment the track into distinct subsequences using a recursive algorithm that maximizes the Jensen-Shannon divergence [12] between the subsequences. By comparing the concurrent positions of different tracks, we can determine if the corresponding persons are interacting. In this work, we focus on modeling the frequency of occurrence of the observed activities and interactions. We assume that activities are generated by a Poisson process and model the number of

occurrences of each activity type as a Poisson distribution. This enables us to compute the probability of detecting a given number of activities per unit time and flagging anomalous observations based on low probability. We tested this system in a courtyard outside a lecture hall which features a large increase in people exiting the lecture hall after each class. Our system is able to detect this infrequent occurence without any pre-defined concept of the occupancy of an environment. Conversations between people standing in this courtyard occur rarely and thus when such conversations do occur, they are also flagged as anomalous. This is because we do not provide any high level notion of a "conversation". Instead, we let the system classify all activities based solely on location in an unsupervised manner.

Figure 1 shows the main components comprising our system. During the learning phase, the system segments tracks into distinct activities and clusters these into a set of representative activities seen in that environment. The rate of occurrence of those activities is also computed in this stage. During the monitoring phase, observed activities are classified into one of the representative activities. The probability of observing the activities is computed from the expected rate that was calculated during the learning phase. An activity is then marked as anomalous if this probability is very low. These steps are described in detail in the following sections. The Results section describes our experiments in the outdoor courtyard environment.

II. RELATED WORK

The problem of recognizing actions and activities of humans has been studied using vision [4, 9, 8, 5]. The tracks of foreground objects in an image are often classified into activity primitives such as hand gestures [4]. The recognition of these low level actions is usually performed by defining a hidden Markov model (HMM) for each action. More complex activities are defined as sequences of lower level ones, which may be considered as symbols serving as input into a higher level representation. Such higher level representations may be a stochastic grammar [4], Bayesian network [8], or another higher level HMM [9].

Detecting real-world interactions has been studied using mainly vision [13, 11]. The type of interactions that are

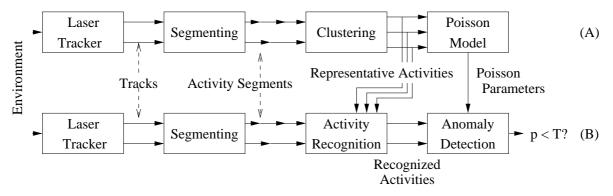


Fig. 1. Block diagram: (A) Learning Poisson model phase. (B) Anomaly detection phase.

detected include meetings between people in an open outdoor area [13] and pick-ups and drop-offs at a parking lot [11]. In these works, the tracks of individuals are segmented into predefined low level actions such as walking, stopping, and entering the parking lot. Thus, all the interactions that can be detected by the system are predefined. Moreover, no model of the observed frequency of activities or interactions is defined. Anomaly detection using these methods reduces to detecting if the observed behavior corresponds to one of the predefined models. Consequently, a behavior which happens much more often than expected will not be seen as unusual.

Laser range-finders have been used to track people for activity modeling [2, 16, 1]. Bennewitz et al. [2] use the Expectation-Maximization algorithm to build models of tracks in an indoor environment. A model is a sequence of positions with an associated Gaussian probability distribution for each position. The positions that comprise a learned motion track are then used as states in a HMM which can be used to estimate the positions of people in that environment [6]. Patterson et al. [15] use this approach to model activity patterns at a city level. Yan and Matarić [16] use a laser range-finder tracker to study which parts of a laboratory environment are occupied the most. Arbuckle et al. [1] extend occupancy grids to take into account the differences in occupancy of a space over different time-scales. However, these systems neither consider interactions between people nor do they model time.

III. LASER TRACKING

Our tracking system has been described in previous work [14] and is summarized here. Two laser range-finders placed along the edges of a courtyard are used to track movements of people. The readings from different range-finders are transformed into a common coordinate system using *Mesh Relaxation* [10]. The range scans are used to maintain a model of the objects in the environment. The measurements are divided into background and foreground readings. Background readings arise from static objects such as walls; they are used to update a background model [7]. Readings that are not explained by the background model are assumed to come from objects that are

to be tracked. The foreground model consists of a set of *particle filters*, one for each object being tracked.

Each particle filter tracks an object using a probability distribution represented as a set of discrete samples or particles. Each particle is a 4-tuple $\langle x,y,\theta,v\rangle$ representing an estimate of the object's position (x,y), orientation θ , and velocity v. The position of the tracked object is given by the mean of the positions of all the particles comprising its probability distribution.

IV. SEGMENTING TRACKS

A person's track may span more than one activity. For instance, two people could stop and talk to each other in the courtyard before exiting the courtyard again. In this case, their tracks consist of three activities: entering the courtyard, having the conversation, and exiting. We have developed a method for automatic segmentation that considers every activity as a distinct probability distribution [14]. Activities within a track can be discovered by splitting the track in such a way as to maximize the difference (in a probabilistic sense) of the individual segments.

The output of the laser tracker is a sequence of x,y positions (in global coordinates) for each tracked object. Every sequence of positions is then converted into a sequence of displacements. Let $(x_i,y_i), i=0,1,\ldots,n$ be a sequence of positions obtained from the tracker. The corresponding displacements are $(r_i,\theta_i), i=1,2,\ldots,n$ where

$$r_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}, i > 0$$

$$\theta_i = \tan^{-1} \frac{y_i - y_{i-1}}{x_i - x_{i-1}}, i > 0$$

The displacements are then discretized into one of nine canonical values, as shown in Figure 2. If $r_i < 0.2m$, then it is discretized as displacement "0". If $r_i \geq 0.2m$ (minimum distance moved in one time-step), then it is discretized as one of displacements "1-8", depending on the sector in which θ_i lies. Thus, the continuous 2-dimensional space of displacements is reduced into a discrete space with 9 values.

Note that this scheme of discretizing displacements ignores the magnitude of the velocity of a person (provided

the displacement is above the threshold used to distinguish stationary people). We found that the velocity of the people in our environment did not vary much while walking. This is consistent with the observation made in [6].

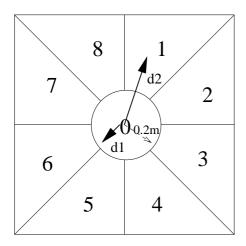


Fig. 2. Discretizing displacements into one of nine canonical displacements. Displacement d_1 is discretized into bin '0", while displacement d_2 is discretized into bin "1".

We denote the set of canonical displacements by \mathcal{X} . Given a sequence $D = \{d_1, d_2, \ldots, d_N\}, d_i \in \mathcal{X}$ of discrete displacements, the Maximum Likelihood (ML) probability distribution p is obtained by counting each type of displacement and dividing by the total number of displacements in the sequence. Let $C(i), i \in \mathcal{X}$ represent the number of occurrences of displacement i in the sequence D. Then

$$p(i) = C(i)/N$$

We define an *activity* to be a set of canonical displacements drawn from a fixed distribution. We assume that different activities give rise to different probability distributions of displacements. The task of segmentation is to then divide a track into is split into a number of consecutive subsequences such that these subsequences are distinct from each other in a probabilistic sense.

We measure the difference between subsequences using an entropic measure known as the Jensen-Shannon divergence [12]. Let p_1, p_2 be two probability distributions defined over the discrete space of canonical displacements, \mathcal{X} . Define the weighted average distribution p_{12} as follows

$$p_{12}(x) = w_1 p_1(x) + w_2 p_2(x) \forall x \in \mathcal{X}$$

where the weights $w_1, w_2 \ge 0$ and $w_1 + w_2 = 1$. Let H(p) be the entropy for a distribution p, defined as

$$H(p) = -\sum_{x \in \mathcal{X}} p(x)log(p(x))$$

The Jensen-Shannon divergence between the two probability distributions p_1, p_2 is defined as

$$JS(p_1, p_2) = H(p_{12}) - (w_1H(p_1) + w_2H(p_2))$$

The Jensen-Shannon divergence has the property that it is always positive, symmetric, and zero only when the two distributions are equal.

We obtain all activity subsequences from a track of displacements by recursively splitting the track at the point that gives rise to maximum Jensen-Shannon divergence of the two component tracks. Let $D = \{d_1, d_2, \ldots, d_N\}, d_i \in \mathcal{X}$ be a sequence of displacements comprising a track. D is split into two subsequences D_l, D_r

$$D_l = \{d_1, d_2, \dots, d_k\}, D_r = \{d_{k+1}, d_{k+2}, \dots, d_N\}$$

such that $k = argmax_i(JS(p_{1,i}, p_{i+1,N}))$, where $p_{i,j}$ is the probability distribution over \mathcal{X} obtained from the displacements $d_i, d_{i+1}, \ldots, d_j$.

The Jensen-Shannon divergence can weight the two probability distributions differently. We utilize the lengths of the sequences giving rise to the probability distributions p_l, p_r to weight the divergence. Thus, $w_1 = k/N$ and $w_2 = (N-k)/N$. This choice of weighting the left and right subsequences in the calculation of the Jensen-Shannon divergence is also used for segmenting DNA sequences [3].

The recursive segmenting procedure is terminated when the confidence of seeing the Jensen-Shannon divergence between the left and right subsequences due to a real difference in their underlying probability distributions falls below a certain threshold. Since we assume that each displacement is an independent identically distributed random variable, we can use the approximation given in [3] to estimate the minimum value of the Jensen-Shannon divergence for a given confidence value.

A. Detecting interactions

In earlier work, we used an entropy-based method for detecting interactions in an indoor environment [14]. In our current work on outdoor environments, we use distance between two people to determine if they are interacting. Proximity between people is sufficient to detect interactions in the outdoor environment that we considered as interacting people were always close to each other, unlike indoors, where they could be close (for example, due to cluttered seating), but not interacting. In our current system, two people are said to be *interacting* if the distance between them always falls below a predefined threshold (2m). Examples of such interactions are people following one another, and holding a conversation while standing close to each other.

V. DETECTING ANOMALOUS ACTIVITIES

To identify anomalous behavior in the environment, we build a model of the frequency at which similar activities and interactions are observed. During the monitoring phase, we count the number of times different activities are observed in a small interval. If the probability of observing the detected number of activities in that interval falls below a certain threshold, an anomalous behavior is flagged.

To construct such a model, we first need to recognize similar activities. In our setup, two activities are said to be *similar* if their corresponding tracks are close. The number of points in any two tracks is generally different. Moreover, even two tracks arising from the same activity

do not exactly match, due to the probabilistic nature of the tracker. Hence, tracks are normalized before they are compared to each other: a cubic spline is fitted through all the points in a track. The fitted spline is then sampled at 20 equidistant points to generate the normalized track. Two normalized tracks are compared by measuring the sum of the distances between corresponding points on the two tracks. This divergence measure is then used to cluster similar tracks together.

We use hierarchical clustering to generate the activity clusters. At every step in the clustering algorithm, the two (normalized) tracks with the smallest divergence measure between them are replaced by a new track that is obtained by taking the mean of the corresponding points in the two tracks. This step is repeated until the minimum divergence among all pairs of tracks exceeds a pre-fixed threshold, ϵ . A track is said to belong to a cluster if the divergence measure between the track and that cluster is minimal relative to all other clusters. The final set of track clusters thus map all observed activities seen in the environment into a smaller number of representative activity types.

We assume that the number of similar activities performed in a given time period follow a Poisson distribution. The probability p(n) of observing n activities in a given time interval is then given by

$$p(n) = \frac{\mu^n e^{-\mu}}{n!}$$

where μ is the mean number of activities expected to occur in that time interval. The Poisson distribution is completely specified by the parameter μ . This parameter is computed for each activity cluster by counting the total occurrences of activities from that cluster and dividing by the duration of the observation period.

Interactions are also assumed to follow a Poisson distribution. Since interactions are defined as two people maintaining a close distance, two interacting tracks are also mapped into the same cluster. Thus, the number of distinct types of interactions detected by our system is equal to the number of activity clusters. The μ parameter (mean number of interactions per unit time) is computed for all distinct interaction types by counting the total number of interactions mapped into that cluster divided by the duration of the observation period.

VI. RESULTS

We monitored people in an outdoor courtyard measuring approximately $10m \times 10m$ (Figure 3). The courtyard is enclosed on all sides. People can enter/exit the courtyard through from the doors of the surrounding building or through a walkway passing through the courtyard. Most of the activities in this environment consisted of people crossing the courtyard from one of the entrances to one of the exits. Such activities often happened in small groups. Occasionally, a small group of people stopped in the courtyard for a conversation before exiting the courtyard.

Two SICK laser range-finders were placed at the corners of the courtyard, at waist level, to capture position data. The



Fig. 3. People crossing the courtyard where we monitored activity.

motion capture session lasted 3.5 hours. Normal activity proceeded throughout the tracking session. The number of occupants of the environment during the experiment varied widely from none to tens of people crossing the courtyard at the end of a class.

The resulting tracks were segmented (recursive segmentation was continued until the confidence in the Jensen-Shannon divergence fell below 90%) and clustered into 20 representative activity segments using the hierarchical clustering algorithm (Figure 4). The laser tracker does not track perfectly, especially when there is a large number of people in the environment. Moreover, objects such as large bicycles are not always distinguishable from people. Hence, not all the clustered segments correspond to complete tracks. A detailed analysis of the quality of segmentation is given in our previous work [14].



Fig. 4. The thick lines indicate the 20 clusters of activity in the environment. The shaded areas indicate the building and bicycle rack enclosing the courtyard. Activity 11 is depicited in Figures 5 and 6. The interactions depicted in Figures 7 and 8 also occur along this activity.

To test our system, we computed the mean number of activities performed per unit time for each activity cluster by dividing the total number of such activities by the total duration of the experiment. The Poisson distribution then gave us the probability of seeing a particular number of activities performed in any 5-minute interval. Figure 5 shows the number of activities performed in 5-minute intervals for one of the 20 activity clusters shown in Figure 4. Figure 6 shows the corresponding probability of seeing that number of activities. The probability drops sharply at t=4850s, where the number of activities in the 5-minute interval goes up to 30, well over the number of activities seen at other times. This time corresponds to the end of a class and the number of students exiting the building into the courtyard suddenly increases.

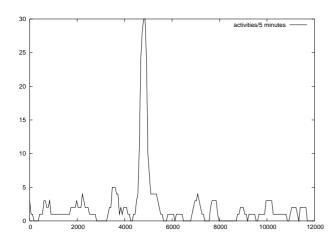


Fig. 5. Number of activities observed in 5-minute intervals (x-axis shows time in seconds).

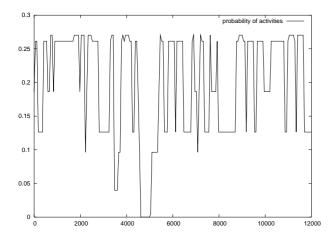


Fig. 6. Expected probability of seeing the number of activities in 5-minute intervals according to the learned Poisson distribution (x-axis shows time in seconds).

A similar analysis can be performed for the observed interactions. The mean number of interactions observed per unit time for each cluster type is calculated from the total number of interactions observed of that type divided by the total duration. Figure 7 shows the number of interactions observed in 5-minute intervals along the track shown in

Figure 4 (corresponds to two people exiting the building together). Figure 8 shows the corresponding probability of seeing that number of interactions. The probability drops sharply at t=4800s when the number of activities in the 5-minute interval goes up to 6, a higher number of interactions than seen at other times. Note that the sudden increase in interactions at t=4800s corresponds to the end of the class as in the previous experiment. This is expected since, in our system, interactions are defined by proximity. Thus, a large number of people in a confined area is likely to lead to an increase in detected interactions.

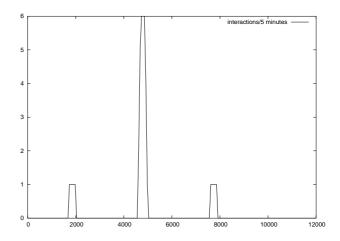


Fig. 7. Number of interactions in a 5-minute interval (x-axis shows time in seconds).

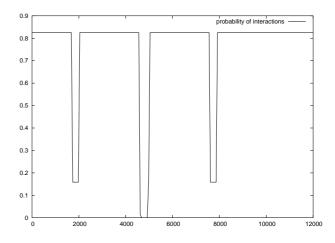


Fig. 8. Expected probability of seeing the number of interactions in a 5-minute interval according to the learned Poisson distribution (x-axis shows time in seconds).

Conversations between stationary people occurred only thrice during the length of the experiment (from t=4100s to t=4241s, from t=6232 to t=6557, and from t=10632 to t=10819). These occurred at different locations in the courtyard and were thus clustered as separate activities (with only one occurrence each). During the monitoring phase, when these conversations took place, they were all shown as low probability activities.

VII. CONCLUSION

We presented a system that tracks people in an outdoor environment and automatically segments tracks of individuals into sub-tracks representing distinct activities. The tracked movements were classified into a small number of representative activities and interactions. We constructed a model of the occurrences of these activities and interactions in an outdoor courtyard over a large time-scale (3.5 hours). The model was then used to identify periods of time when the observed frequency of activities or interactions exceeded what was usually observed, thus flagging an anomalous occurrence. The detected anomalies corresponded to sudden increases in the number of people in the courtyard and also a related increase in the number of interactions (that occurred due to a real-world event: students leaving a large lecture hall after a class). However, all interactions are clustered based on location, and therefore our system does not have the ability to recognize all conversations as belonging to one activity type. Thus, every conversation that took place was marked as anomalous.

ACKNOWLEDGEMENT

The authors thank Helen Yan for developing the laser tracking system. This work was supported in part by the DoE RIM grant DE-FG03-01ER45905, and in part by the ONR MURI grant SA3319.

REFERENCES

- [1] Daniel Arbuckle, Andrew Howard, and Maja J Matarić. Temporal Occupancy Grids: a method for classifying spatio-temporal properties of the environment. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 409–414, EPFL, Switzerland, September 30-October 4, 2002.
- [2] M. Bennewitz, W. Burgard, and S. Thrun. Learning motion patterns of persons for mobile service robots. In *Proceedings of the IEEE International Conference* on Robotics and Automation (ICRA), pages 3601– 3606, Washington, DC, USA, May 11-15, 2002.
- [3] P. Bernaola-Galvan, R. Roman-Roldan, and J. Oliver. Compositional segmentation and long-range fractal correlations in dna sequences. *Physical Review E*, 53(5):5181–5189, 1996.
- [4] A. F. Bobick and Y. A. Ivanov. Action recognition using probabilistic parsing. In *Conference on Com*puter Vision and Pattern Recognition (CVPR), pages 196–202, Santa Barbara, USA, June 23-25, 1998.
- [5] Matthew Brand, Nuria Oliver, and Alex Pentland. Coupled hidden markov models for complex action recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 994–999, San Juan, Puerto Rico, June 17-19, 1997.
- [6] G. Cielniak, M. Bennewitz, and W. Burgard. Where is ...? learning and utilizing motion patterns of persons with mobile robots. In *Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 909–914, Acapulco, Mexico, August 9-10, 2003.

- [7] Ajo Fod, Andrew Howard, and Maja J Matarić. A laser-based people tracker. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3024–3029, Washington, DC, USA, May 11-15, 2002.
- [8] Somboon Hongeng, Francois Bremond, and Ram Nevatia. Bayesian framework for video surveillance application. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 1, pages 1164–1170, Barcelona, Spain, September 3-8, 2000.
- [9] Somboon Hongeng, Francois Bremond, and Ram Nevatia. Representation and optimal recognition of human activities. In *Conference on Computer Vision* and Pattern Recognition (CVPR), pages 1818–1825, Hilton Head, SC, USA, June 13-15, 2000.
- [10] Andrew Howard, Maja J Matarić, and Gaurav S. Sukhatme. Relaxation on a mesh: a formalism for generalized localization. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1055–1060, Wailea, USA, October 2001.
- [11] Yuri Ivanov, Chris Stauffer, Aaron Bobick, and W.E.L. Grimson. Video surveillance of interactions. In Second IEEE Workshop on Visual Surveillance, pages 82–90, Fort Collins, Colorado, USA,, June 26 1999.
- [12] J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.
- [13] Nuria Oliver, Barbara Rosario, and Alex Pentland. Graphical models for recognizing human interactions. In *Advances in Neural Information Processing Systems*, pages 924–930, Denver, USA, December 1-3, 1998.
- [14] Anand Panangadan, Maja J. Matarić, and Gaurav Sukhatme. Modeling human interactions in indoor environments. In *Third International Joint Conference on Autonomous Agents and Multi Agent Systems* (AAMAS), 2004. (To Appear).
- [15] D.J. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring high-level behavior from low-level sensors. In *Proceedings of the Fifth International Conference on Ubiquitous Computing (UBICOMP)*, pages 73–89, Seattle, USA, October 12-15, 2003.
- [16] Helen Yan and Maja J. Matarić. General spatial features for analysis of multi-robot and human activities from raw position data. In *Proceedings of* the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 2770–2775, EPFL, Switzerland, September 30 - October 4, 2002.