Learning a factorized segmental representation of far-field tracking data

Chris Stauffer
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

There are many useful observable characteristics of the state of a tracked object. These characteristics could include normalized size, normalized speed, normalized direction, object color, position, and object shape among other characteristics. Although these characteristics are by no means completely independent of each other, it is desirable to determine a separate, compact description of each of each of these aspects. Using this compact factored description, different aspects of individual sequences can be estimated and described without overwhelming computational or storage costs. In this work, we describe Factored Latent Analysis (FLA) and its application to deriving factored models for segmenting sequences in each of K separate characteristics. This method exploits temporally local statistics within each of the latent aspects and their interdependencies to derive a model that allows segmentation of each of the observed characteristics. This method is data driven and unsupervised. Activity classification results for multiple challenging environments are shown.

1. Introduction

By simple observation of an object moving through an environment, most humans can effectively describe the instantaneous characteristics of the tracked object using very few bits of information. For instance, a person walking west on the school's front sidewalk or a car stopping at the second toll both are very compact descriptions that imply particular sizes, velocities, directions, and locations. Using this low-bandwidth description, it is possible to describe entire sequences in a compact manner. For instance, the person walked away from the car park, stopped, and ran back towards the car park.

In many scenes, an effective description of large numbers of complete tracking sequences should be both factored and segmental. With a **factored** representation, one can describe whether a person is stopped, walking, or running independently of the direction they are moving, or one can describe the lane of traffic independently of the vehicle class. With a **segmental** representation, one can describe a person that runs to an ATM, stops, and walks off towards the east.

Most previous work in *learning* effective descriptions of tracked objects was not factored. Different aspects of the description were either treated completely independently or thrown into the same feature space. The first approach doesn't exploit the information in the joint observations. The second approach uses the joint observations but doesn't provide an effective description of the individual aspects of the description.

Also, previous work in *learning* to describe tracked object sequences is often not segmental. Many approaches are only capable of describing the entire sequence, not each individual action in the sequence. For most tracking sequences in interesting environments, a single description is not sufficient for the entire sequence as the object may change direction, shape, speed, or move from one region to another.

The goal of this work is to use observations of object state over time to automatically generate an unsupervised description of the activity of the objects in the scene. This system is capable of compactly describing in what state an object is and what it is doing throughout an entire tracking sequence. With very little supervision, these latent classes can be associated with certain labels associated with high-level human concepts. Building such an intermediate representation that accurately represents the observations for a particular environment is extremely useful for unsupervised and semi-supervised description.

1.1. Previous Work

Our primary goal is to determine a set of latent class models for each of a number of different observable characteristics. Graphical models including latent class models have become increasingly popular over the last few years. Thomas Hofmann's aspect model[2] was capable of inducing latent word and document models from occurrence statistics of words within documents. Latent Dirichlet Allocation[1] added the ability to represent the process by which documents are drawn as *mixtures* of latent word classes as determined by a Dirichlet prior.

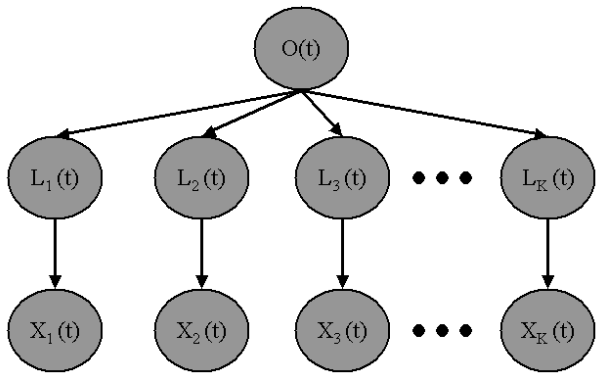


Figure 1: The graphical model used in Factored Latent Analysis (FLA).

Our model and estimation is different in two respects. First, we explicitly factored the observations as shown in Figure 1. This results in a separate latent class model for each of a number of different descriptions allowing direction, speed, size, and location to have completely independent descriptions. Second, we estimate the model using pairwise observation statistics. Stauffer and Grimson [6] built hierarchical models using pairwise joint co-occurrence of observations of a single type. This work is similar, but uses all the pairwise observation between different characteristics to induce latent class models that are consistent with those statistics across multiple aspects.

Many varieties of data-driven perceptual data mining[5] techniques have been applied to activity analysis. Johnson and Hogg [3] clustered trajectories in a scene into 400 representative clusters allowing generalization and prediction, but not compact description. Stauffer and Grimson [6] used a hierarchal clustering technique to develop a binary tree classification hierarchy, which is somewhat more compact but also not segmental. Because these approaches describe entire paths holistically it is unclear how they can be adapted to describe compound activities. Makris and Ellis[4] also clustered entire trajectories into independent paths, but they extracted common features of paths to describe way points enabling some segmentation of the entire sequence but still a limited description of each segment.

This paper describes a method for learning an effective factored segmental description of activity in scenes directly from the tracking data using observed pairwise joint co-occurrences. Section 2 describes the statistical model. Section 3 discusses the estimation technique advocated by the graphical model. Section 4 describes an alternative approximate technique chosen for this work because of computational viability. Section 5 shows this model applied in multiple environments. Section 6 and Section 7 discuss future work and conclusions that can be drawn from this work.

2. Factored Latent Analysis

Even humans cannot guess the entire internal state of a tracked object O at a particular time t. Objects like people and cars have internal state that will never be observable, such as intention, mood, and distraction. The only aspects of the state of an object which can easily be modelled are those aspects that are directly reflected in the observations of an object, such as height, size, shape, speed, direction, location or color. This section describes our generative model of observed characteristics of an object.

At time t, the object state is $O(t)$ and the observed characteristics of the object are $X(t)$. $X(t)$ is a vector of K observations, $\{X_1(t), X_2(t), ..., X_K(t)\}$. Our goal is to compactly represent

$$p(X(t), O(t)) = p(X(t)|O(t))p(O(t)) \quad (1)$$

where $p(O(t))$ is the probably of being in a particular state and $p(X(t)|O(t))$ is the conditional likelihood of producing a particular set of observations. In this work we assume that $X(t)$ and $O(t)$ reflect only aspects of the object which are observable.

Figure 1 shows the graphical model used in Factored Latent Analysis. In the model, the observed state $X(t)$ is factored into K separate variables, $(X_1(t), X_2(t), ..., X_K(t))$, and each variable is represented by its own latent variable. We assume that the i^{th} observation $X_i(t)$ can be drawn independently from a distribution conditioned on the latent variable L_i. This is the latent class conditional output distribution $p(X_i|L_i)$. An instance of this variable can have one of k_i possible values, each representing a latent class.

For instance, observations of normalized size $X_s(t)$ ideally would be a single value with additive noise. In an environment with people, cars, and trucks, a coarse approximation would be three latent size classes, or $l_s \in \{0, 1, 2\}$. Conditioned on its latent class, an object should exhibit a particular size with additive measurement noise. In reality, there may be multiple classes of pedestrian and vehicle sizes, which may suggest a hierarchical approach. Observations of normalized velocity $X_v(t)$ in a particular environment may be well-represented by three latent classes: stopped, walking, jogging. Determining the "correct" number of classes for each latent variable k_i is a difficult problem that is not fully addressed in this work.

Our approximation to $p(O(t))$ is a multinomial distribution over every possible combination or latent class labels, or $p(l_1, l_2, ..., l_K)$. For instance, a particular scene may have a high likelihood of observing small slow objects moving in any direction and large fast objects moving in one of two primary directions and a very low likelihood of observing all other possible object states.

Thus, based on the independence assumptions in this

model, equation 1 becomes

$$p(x_1, x_2, ..., x_K) = \sum_{(l_1,l_2,...l_K)\in\mathbf{L}} p(l_1, l_2, ..., l_K) \prod_{i=1}^{K} p(x_i|l_i), \quad (2)$$

where $\mathbf{L}$ is the set of all possible combinations of latent aspect assignments. Although many of the possible combinations will tend to have zero probability of occurring, the number of potential combinations is

$$|\mathbf{L}| = \prod_{i=1}^{K} k_i, \quad (3)$$

where k_i is the number of latent classes for each type of observations. The number of latent classes for each observation, k_i, is the primary means by which the capacity of this model is limited. In the examples in this work, this number has been chosen, but Section 6 discusses how this part of the estimation could be automated.

3. Estimation Techniques

Given the form of the model and a large set of observations, it is possible to estimate the maximum likelihood parameters of the model. We raise two potential estimation techniques. Subsection 3.1 describes how the entire joint latent model could be estimated directly from the data. The potential advantages and pitfalls of this technique are outlined. Due to the type and quantity of data in the problem we are considering, a computationally feasible alternative is introduced in Section 4. This alternative is an approximate technique for estimating the model from aggregate observation statistics between pairs of different observation types.

3.1. Complete EM estimation

The obvious approach to estimating the parameters of the model is a straight-forward EM implementation. The hidden variable is the latent class assignment. The E-step involves computing the likelihood over each latent variable given model of $p(l_1, ..., l_K)$ and $p(x_i|l_i)$.

The M-step involves maximizing the model parameters given the latent likelihoods. The exact form of the model will dictate the complexity of both steps. For instance, $p(x_i|l_i)$ could be a multinomial distribution, a Gaussian distribution, or a mixture of Gaussians and $p(l_1, ..., l_K)$ could be a multinomial joint distribution with a uniform prior, a Dirichlet prior, or any other alternative.

Using this approach, one can estimate the full joint distribution over the latent classes $p(l_1, l_2, ..., l_K)$ using all the data while simultaneously estimating the class conditional output distributions $p(x_i|l_i)$. Unfortunately, the computational complexity of this procedure is at least $O(Nk_1k_2...k_K)$, where N is the number of observation vectors observed. For a nominal number of latent classes (approximately 3 or 4) over a nominal number of observation types (3 or 4) over the number of observations in a single hour of tracking, a single EM iteration can take as much as 1 day. If the observations x_i are kept continuous and are more complex than scalar values like velocity (for example, silhouettes or color histograms), each EM step could take weeks. For this reason, we present the alternative in the next section.

4. Estimation from pairwise observation statistics

This section describes our approach to approximating the Factored Latent Analysis model as applied to the problem of activity analysis. It describes: quantization of the (potentially) continuous output observations; estimation of pairwise joint observation statistics; estimating the latent class likelihoods and mixing probabilities from the pairwise statistics; and classification using the model.

4.1. Quantization of observation spaces

Our first approximation is to discretize the output observations for each type of observation. This can be done by uniformly binning the output space or using vector quantization to more compactly represent the output space. Thus, potentially continuous observations, x_i, can be represented by a discrete set of prototypical values $\dot{x}_i$.

Figure 2 shows a scene with overlayed tracking data. For this scene, we will make discretized measurements of size, velocity, direction, and position. In most cases, we will simply partition the measurements from their min to max value into equal sized bins. Using 64 bins to represent sizes, velocities, and directions is sufficent, whereas position is a two dimensional space requiring more bins.

In theory, this quantization is not necessary, but it makes it computationally feasible to collect estimates of joint occurrences of observations of type i and type j, or $\hat{p}(x_i, x_j)$. The primary advantage is that the discretized joint occurrence estimate $\hat{p}(\dot{x}_i, \dot{x}_j)$ will remain a constant size regardless of the number of observations. Because there are potentially tens of thousands of tracked objects per hour and each tracked object can contain hundreds of valid temporal windows, this approximation is required to make this application computationally feasible.

4.2. Estimating pairwise joint observations

If one was provided segmented, labeled data for a particular observation type, the class conditional probabilities $p(x_i|l_i)$ could be estimated directly. Unfortunately, our data is neither segmented nor labeled for any of the types of observations.

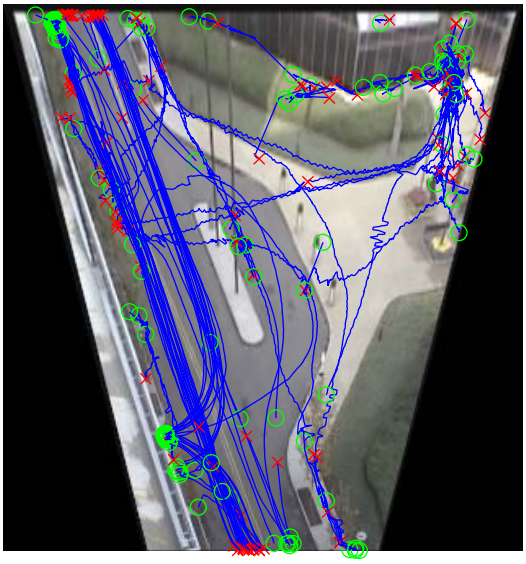


Figure 2: This figure shows the traffic in a particular environment after rectification. Green circles denote beginnings of track and red x's denote ends of tracks. A large portion of the traffic is on the roads in either primary direction. But pedestrians are often seen crossing the road and moving to or away from the office building on the right.

Because we lack an oracle to segment the sequences for each observation type, we make an assumption that the observations within a temporal Gaussian weighted window are of the same latent class. This assumption is an integral part of this learning technique. For object activity, we use a window of one second duration. Within such a window an object's size, velocity, direction of travel and location are assumed to be drawn independently from a single latent class for each observation type. For instance, over a single second a particular object could be described as a small, fast-moving object travelling north on the highway. This corresponds to one possible N-tuple joint latent state assignment in the set $\mathbf{L}$.

If the temporal window size is too small, a randomly chosen window would exhibit little within class variability in the observations. If the temporal window is too large, many randomly sampled windows would include observations drawn from multiple latent classes.

$\hat{p}(\dot{x}_i, \dot{x}_j)$, is an estimate of the likelihood that a pair of observations of type i and type j would be drawn from a short temporal window. Given a large set of randomly sampled temporal windows that are primarily uniform in their underlying joint latent class, $\hat{p}(\dot{x}_i, \dot{x}_j)$ is a frequentist approximation to the true joint occurrence probability.

Figure 3 shows $\hat{p}(\dot{x}_s, \dot{x}_v)$ for a scene containing pedestrians and vehicles, where x_s is an observation of object size and x_v is an observation of object velocity. In this case,

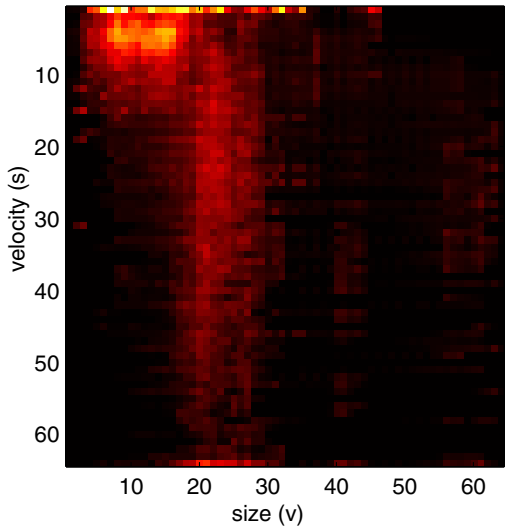


Figure 3: An example joint occurrence over size and velocity, or $\hat{p}(\dot{x}_s, \dot{x}_v)$, estimated from a scene containing pedestrians and vehicles. Brighter indicate higher likelihood of cooccurrences within a window of observations drawn from a single joint latent class.

"size" is estimated as the square root of the number of pixels and "velocity" refers to the speed an object is travelling in normalized coordinates after rectification. As is evident in the figure, there are two main types of objects: slow-moving small objects and fast-moving large objects. Although, it is interesting to note that both the small objects and the large objects can be observed in a stopped or loitering state.

This estimate can be accumulated in an online fashion, as follows

$$\hat{p}(\dot{x}_i, \dot{x}_j) = \sum_{n=1}^{N} w_n p_n(\dot{x}_i, \dot{x}_j), \tag{4}$$

where $p_n(\dot{x}_i, \dot{x}_j)$ is the joint cooccurrence within a particular window and w_n is a weight for that window. Thus, the aggregate estimate and the sum of the weights are sufficient statistics. The weight can be used to decrease the affect of extremely redundant observed states. For instance, the weight for the hundreds of redundant observations of a car that stops in the scene for two minutes can be decreased to lessen their total effect on $\hat{p}(\dot{x}_i, \dot{x}_j)$.

4.3. Estimating pairwise latent class conditional and prior distributions

Given an estimate of $\hat{p}(\dot{x}_i, \dot{x}_j)$ for pairs of observations of type i and type j and a number of latent size classes k_s and number of latent velocity classes k_v, it is possible to estimate the model parameters that best approximate those joint statistics.

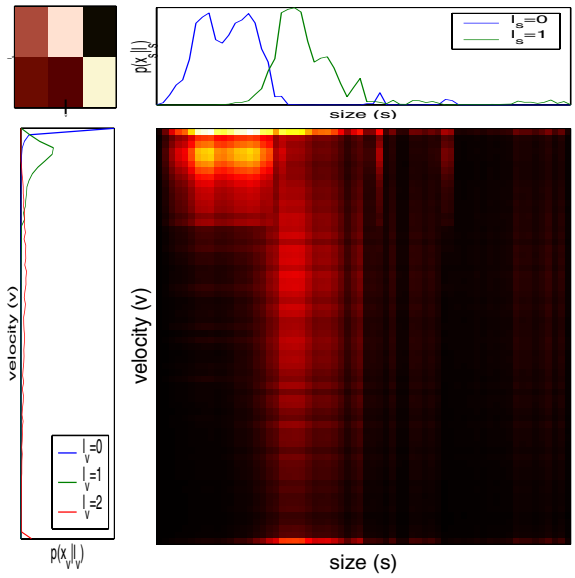


Figure 4: This figure shows: the latent pair likelihoods, $\tilde{p}(l_s, l_v)$, on the upper left; the conditional likelihood models for each latent object size class, $\tilde{p}(\dot{x}_s|l_s)$, on the upper right; the conditional likelihood for each latent object velocity class, $\tilde{p}(\dot{x}_v|l_v)$, on the lower left; and the models implied joint statistics as outlined in Equation 5 on the lower right.

Given the model's estimates of the mixing likelihoods $\tilde{p}(l_s, l_v)$, and conditional output likelihoods for $\tilde{p}(\dot{x}_s|l_s)$ and $\tilde{p}(\dot{x}_v|l_v)$ the likelihood of observing a particular pair of observations from the model is simply

$$\tilde{p}(\dot{x}_i, \dot{x}_j) = \sum_{(l_s,l_v)\in \mathbf{L_{sv}}} \tilde{p}(l_s, l_v)\tilde{p}(\dot{x}_s|l_s)\tilde{p}(\dot{x}_v|l_v), \qquad (5)$$

where $\mathbf{L_{sv}}$ is the set of all pair of latent size and velocity assignments.

Given random initial estimates of $\tilde{p}(l_s, l_v)$, $\tilde{p}(\dot{x}_s|l_s)$, and $p(\dot{x}_v|l_v)$, the model parameters that maximize the likelihood of the data can be iteratively estimated using an EM procedure. This equates to minimizing the KL divergence between the joint statistics of the model $\tilde{p}(\dot{x}_i, \dot{x}_j)$ and the observed joint statistics $\hat{p}(\dot{x}_i, \dot{x}_j)$.

Figure 4 shows the maximum likelihood model for the $\hat{p}(\dot{x}_i, \dot{x}_j)$ shown in Figure 3 after convergence. It is evident by comparing the two joints, that the model is able to effectively approximate the joint statistics with only two latent size classes and three latent velocity classes. The two latent size classes correspond to pedestrians and vehicles. The bimodal pedestrian class results from individual pedestrians and pairs of connected pedestrians. The three latent velocity classes roughly correspond to loitering, walking, and driving through the scene.

As is evident in the mixing matrix, $\tilde{p}(l_s, l_v)$, on the upper left, most pedestrians walk, most vehicles drive, but

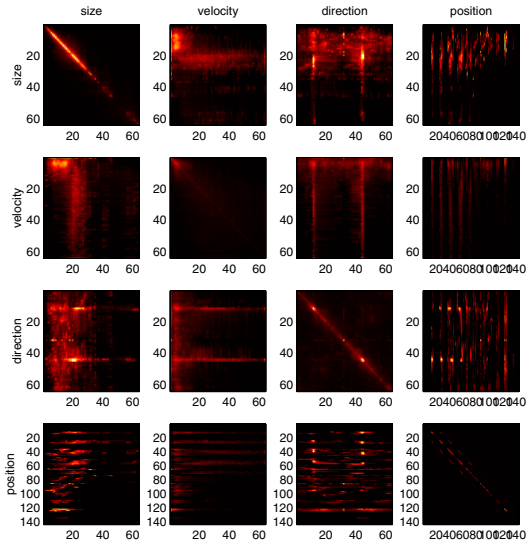


Figure 5: This figure shows 16 matrices corresponding to each possible $\hat{p}(\dot{x}_i, \dot{x}_j)$ for the scene shown in Figure 2 given a one second equivalency window.

both classes are sometimes observed loitering for periods of time. Also, the loitering distribution captures the variance that is observed in loitering objects, because even a parked car will will exhibit some measurement noise in the observed location. This is an important feature when classifying sequences.

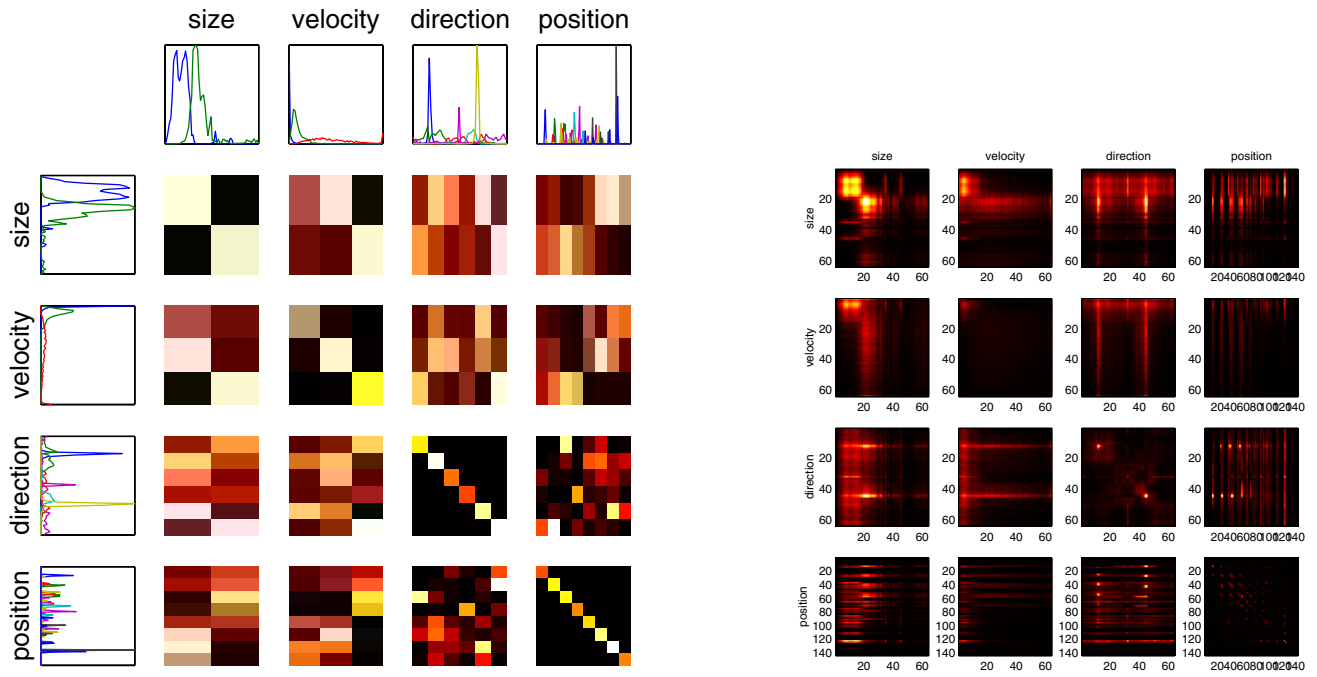
While this process could be done independently for each possible pair of observations, it would result in multiple latent class models for each independent factor. E.g., two latent class models for size given velocity; two latent class models for size given direction; and so on. The next subsection describes how just K sets of latent models can be estimated that are consistent with K^2 pairwise joint estimates.

4.4. Exploiting K^2 pairwise joint observations

In general, Factored Latent Analysis includes K sets of latent classes $\tilde{p}(\dot{x}_i|l_i)$, one for each type of latent observation, and a model for drawing latent class combinations $\tilde{p}(l_i, l_j)$. The goal of our method is to estimate the K sets of latent class models for each independent factor and the pairwise mixing proportions that best approximate *all* the pairwise observations. One can trivially estimate K^2 pairwise joint observation estimates[1], each of which corresponds to a possible pair of observation types.

Figure 5 shows all pairwise joint observations for the scene shown in Figure 2. It is obvious that there is significant structure in these pairwise joint statistics. For instance,

[1]Since $\hat{p}(\dot{x}_i, \dot{x}_j)$ is redundant with $\hat{p}(\dot{x}_j, \dot{x}_i)$, only $\frac{K(K+1)}{2}$ pairs need to be stored.



(a) (b)

Figure 6: This figure shows latent classes and their mixing ratios (a) and the resulting $\tilde{p}(\dot{x}_i, \dot{x}_j)$.

small objects tend to loiter or move slow. Large objects tend to move faster in one of two primary directions (north and south). Because of the 2D nature of position, it is difficult to interpret the last row and column without performing classification (see Section 5).

Given a complete set of K latent class conditional output distributions $\tilde{p}(\dot{x}_i|l_i)$ and K^2 pairwise latent marginal distributions $\tilde{p}(l_i, l_j)$, every marginal pairwise output distribution $\tilde{p}(\dot{x}_i, \dot{x}_j)$ can be estimated. By minimizing the KL divergence between all of the pairwise output distributions and pairwise output estimates the likelihood of the data given the model can be maximized. The criterion is

$$\tilde{\theta}^* = \underset{\tilde{\theta}}{argmin} \sum_{i,j} d(\hat{p}(\dot{x}_i, \dot{x}_j)||\tilde{p}(\dot{x}_i, \dot{x}_j)), \qquad (6)$$

where $\tilde{\theta}$ is the parameters of the K sets of k_i multinomials over $|\dot{x}_i|$ and the K^2 k_i by k_j latent joint distributions. This equates to maximizing the likelihood of *all* pairwise measurements.

Figure 6 (a) shows the model parameters $\tilde{\theta}$ and Figure (b) shows the corresponding $\tilde{p}(\dot{x}_i, \dot{x}_j)$ for the scene shown in Figure 2 after convergence. This model was automatically estimated given two, three, six, and eight latent classes respectively for size, velocity, direction, and position.

Not surprisingly, many of the latent class conditional distributions have intuitive interpretations. As in the previous subsection, the latent size classes roughly correspond to pedestrians and vehicles even though the distribution of pedestrian sizes is bimodal. This is due to the fact that pedestrians and pairs of pedestrians exhibit similar speeds, directions, and locations. The same velocity classes also occur. Using only the diagonal elements $\tilde{p}(\dot{x}_i, \dot{x}_i)$ of Figure 6(a), the likelihood of each latent class is evident. There are two common directions corresponding to the road directions. These directions are taken by both pedestrians and vehicles, but there are many directions that are only taken primarily by pedestrians. These results will be discussed further in Section 5 along with other examples.

4.5. Applying the model to classification

While any state observation $X(t)$ could be classified into one of each type of latent class independently, it is important to remember what the class conditional likelihoods actually represent. They were calculated based on an assumption of temporal coherency, thus the latent class likelihoods for a particular time t are estimated from more than simply $X(t)$.

Currently, we simply assume the latent class posteriors are the weighted expectation of the independent posterior estimates over the observations temporal coherency window. This performs smoothing which greatly reduces the number of spurious detections from instantaneous tracking glitches and short-term changes. In video surveillance, such spurious measurements can result from object interactions, shadow effects, lighting effects, and other tracking failures

and are extremely common.

5. Results

While the model learned in the previous section is intuitively pleasing, this section will show classification of specific tracking sequences and show how these provide an effective description of tracking sequences.

Figure 7(a) shows a new set of tracked data from our previous scene. The latent classes from the unsupervised training are nearly identical to the example in the previous section despite differences in the number of objects or distributions of their behaviors. Figure 7 (b)-(e) show the maximum likelihood classification of the observation for hundreds of sequences. See the figure for a detailed description. These results illustrate how position latent classes tend to represent areas which contain similar types of object moving in similar ways. For example, in Figure 7(e) the blue class corresponds to an area where cars and pedestrians are present and both are likely to be in the "loitering" state.

Figure 8 and Figure 9 show results for two other environment with the same number of latent classes. See the figures for detailed description. Each example learned a set of classes for all four types of observations that are specific to a particular environment and input. In these two cases, the number of latent classes was arguably too large, but the clustering was still meaningful.

By a simple process of labeling these latent classes with textual descriptions, interesting compound queries can be constructed. E.g., "show me the pedestrians that crossed the road." "How many vehicles stopped in the loading zone?" "How many individuals who stopped in the scene started in the opposite direction?" "How much of the traffic on the road is vehicles?" "Did anything stop in the scene for more than 30 seconds?" These queries illustrate the importance of factored representation and the ability to segment effectively.

6. Future Work

We look forward to implementing an efficient estimation technique for estimation of the complete $p(l_1, l_2, ..., l_K)$ as described in Section 2. This would have a number of advantages. Rather than having estimates of all pairwise latent joints $\tilde{p}(l_i, l_j)$, we would have the full latent joint distribution, $\tilde{p}(l_1, l_2, ..., l_K)$.

Another major issue is choosing the "correct" number of latent classes for each observation type. We intend to implement a model selection criteria based on the Bayesian Information Criterion, which will incorporate a measure of the mutual information in the latent joint distribution. This will make splitting a latent class for one type of observation undesirable if it does not increase the information about other types of observations. For example with latent size, it will be more desirable to split vehicles into cars and trucks if they also exhibit different speed, direction, or location characteristics. Related to the issue of choosing the number of classes is learning class labels. In fact, labels given from an oracle could simply be another observation type.

Finally, there are many more characteristics that can be included in this estimation. This could include average color, color histograms, shape, sparsity, internal object deformation, source location, or sink location. Each of these characteristics may bear new information on object activity, class, or appearance.

7. Summary

We have presented a method for approximating the Factored Latent Analysis model using pairwise joint observation estimates $\hat{p}(\dot{x}_i, \dot{x}_j)$. This factored model enables a compact description of object properties, which is more useful than many previous automatically-generated unfactored descriptions. Our model also enables tracking sequences to be segmented into component parts effectively. This is essential in describing compound activities that occur in most environments.

This algorithm exhibited good performance across many environments by exploiting the pairwise observations and temporal coherence. We believe this work shows promise and intend to pursue further related avenues of research.

References

[1] D. M. Blei, A. Y Ng, and M. I. Jordan. Latent dirichlet allocation. In *Journal of Machine Learning Research*, volume 3(4), pages 993–1022, 2003.

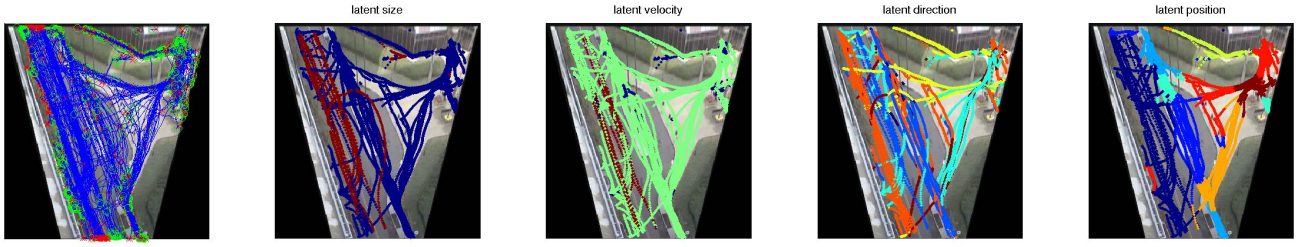
[2] Thomas Hofmann. Probabilistic latent semantic analysis. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99)*, 1999.

[3] Neil Johnson and David Hogg. Learning the distribution of object trajectories for event recognition. In *British Machine Vision Conference*, volume 2, pages 583–592, September 1995.

[4] Dimitrios Makris and Tim Ellis. Finding paths in video sequences. In *British Machine Vision Conference*, volume 1, pages 253–272, Manchester, UK, September 10-13 2001.

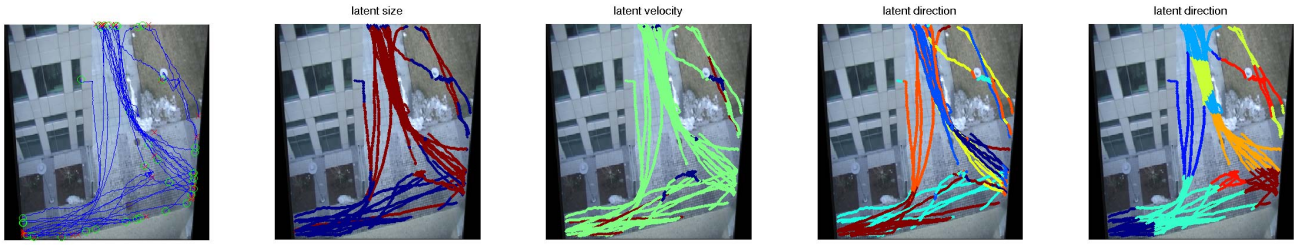
[5] Chris Stauffer. *Perceptual Data Mining*. Phd dissertation, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, June 2002.

[6] Chris Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition*, pages 246–252, 1999.



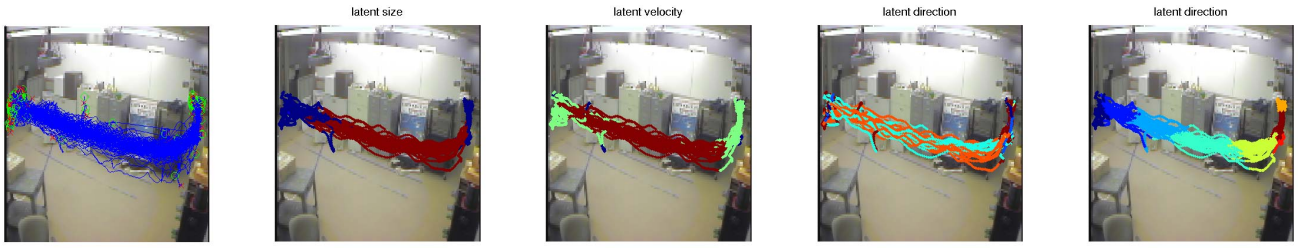
(a) (b) (c) (d) (e)

Figure 7: This figure shows a new set of tracked objects for the previously introduced scene as well as the classification of individual object states in each of the four latent class types. The latent size classes show that vehicles tend to be on the road or in the drop-off region. The latent velocity classes show that only the vehicles passing through on the north-south road travel faster than the nominal speed. Interestingly, there are some examples of pedestrian sized objects going the faster speed in the road region (bicycles and rollerbladers). Figure 7(d) shows north (orange) and south (blue) classes as well as traffic moving towards the office from the south (cyan) and from the north (yellow) and traffic moving away from the office (red). Figure 7(e) shows regions with similar characteristics. The road (dark blue) is a region unto itself. The sidewalk along the street on the north and south side of the scene (light blue) is split by either walking through the drop-off zone (blue) or through the angled sidewalks (orange and red).



(a) (b) (c) (d) (e)

Figure 8: This figure shows tracked data for pedestrians in a courtyard (a) and each observation classified in each of the four latent class models. The latent sizes primarily result from occluded and unoccluded pedestrians. Three pedestrians jogged away from the building and four pedestrians stopped. The stopping zones were clustered together (red) as were the major lanes of movement. There are north, south, east, west, northwest and southeast directions.



(a) (b) (c) (d) (e)

Figure 9: This figure shows tracked data for a indoor, mid-field scene with pedestrians. Again, the size classes correspond to occluded humans. The two major directions of travel (orange and cyan) correspond to most of the traffic. The red and blue regions were areas in the scene where pedestrians stopped.