DISCRIMINATIVE MODEL SELECTION FOR OBJECT MOTION RECOGNITION

Jacinto C. Nascimento [(a)] *Jorge S. Marques* [(a)] *Mário A. T. Figueiredo* [(b)]

[(a)]Instituto de Sistemas e Robótica [(b)]Instituto de Telecomunicações
Instituto Superior Técnico
1049-001 Lisboa, **Portugal**

ABSTRACT

A central issue in mixture-type models is the determination of a suitable number of components that best suits the observed data. In this paper, we address this issue in the context of trajectory classification based on mixtures of motion vector fields.

We adopt a discriminative criterion for choosing among alternative models for each class, based on the classification accuracy on a held out dataset. The key idea is that we make use of the knowledge that the obtained model is going to be used for a specific task: classification. Experiments with both synthetic and real data concerning pedestrian activity classification illustrate the performance of the adopted criterion.

1. INTRODUCTION

Model selection questions arise in many statistical inference and learning problems and are usually tackled by choosing a model that balances parsimony with adequacy to the observed data.

In this paper, we focus on model selection in the context of a problem of trajectory classification. Specifically, the trajectories of the objects are described by a set of vector motion fields, as recently proposed in [1]. Each trajectory can be split into a set of consecutive segments, each generated by a vector field. Switching among fields can occur at any point in the image, with switching probabilities depending on the object location via a field of (stochastic) switching matrices.

To illustrate the idea, consider two intersecting roads (or streets), as depicted in Fig. 1. The typical motion regime in each street is well modeled by a simple (smooth) vector field. At the intersection, pedestrians (or cars) have a considerable probability of changing (switching) direction, whereas far from it, this switching probability is low (or zero).

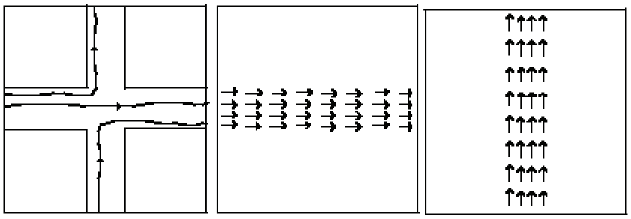

Fig. 1. Crossroad example. Set of trajectories (left) and two vector fields modeling to the trajectories (center and right).

This approach is flexible enough to represent a wide variety of trajectories and allows modeling space-varying behaviors without resorting to non-linear dynamical models, which are infamously hard to estimate from training data. Moreover, it was also shown that the model (*i.e.*, the motion and switching fields) can be efficiently estimated via an expectation-maximization (EM) algorithm.

The classical overfitting/underfitting tradeoff naturally arises in this context: an arbitrarily large number of motion fields allows fitting the training data (set of observed training trajectories) arbitrarily well. Of course, a model with this overfit to the data will not generalize well, and will not lead to a classifier with good performance. On the other hand, a too small number of fields will not be able to adequately capture the underlying complexity of the scene. This also has a negative impact on the performance of any classifier based on such model. Choosing the right number of fields is clearly a model selection problem.

The literature on statistical model selection is too vast to be comprehensively reviewed in this paper. There are several families of approaches, such as variational Bayesian methods [6, 7, 8] and information-theoretic criteria, such as the minimum description length (MDL), the minumum message length (MML), and the Akaike information criterion (AIC) [2, 3, ?]. All these criteria are generative, in the sense that they assume that the goal is to obtain a model that strikes a balance between explaining well the observed data (formally, having a high likelihood) and not being overly complex. When generative models are used in the context

This work was partially supported by Fundação para a Ciência e Tecnologia, Portuguese Ministry of Science, Technology, and Higher Education, under the scope of project PTDC/EEA-CRO/098550/2008 (ARGUS).

of classification tasks (namely, as class-conditional densities), there is no guarantee that a generative model selection criterion yields models with the best possible (among the alternatives) classification performance.

Since the main goal of our model, a mixture of motion fields, is trajectory classification, we proposed to adopt a discriminative criterion for model selection, following the rationale in [9]. The key idea is to make use of the knowledge that the obtained models are going to be used for a specific task: classification. That is, rather than basing the model selection of the generative properties of the model, it is their classification performance that is used as the model selection criterion. To compute the discriminative criterion, we split the available data into disjoint *training* and *selection* sets. We then estimate a collection of models from the training set, using the standard maximum likelihood (ML) criterion via the EM algorithm introduced in [1], and finally select the one achieving the highest classification accuracy on the *selection* set.

The rest of the paper is organized as follows: Section 2 reviews the adopted generative motion model; Section 3 describes how this model is used for trajectory classification; the adopted discriminative model selection criterion in described in Section 4; Section 5 reports experimental results and Section 6 draws some conclusions and points to future work.

2. TRAJECTORY MODEL AND ITS ESTIMATION

A trajectory is simply a length-n sequence of positions of the center of mass of the person (or object) in the image plane, $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, with $\mathbf{x}_t \in \mathbb{R}^2$. This sequence of positions is provided by some tracking algorithm (although this is by no means a trivial task, in the context of this paper we consider it a solved problem).

We model each trajectory has having been generated by a set of K vector (motion) fields $\mathcal{T} = \{\mathbf{T}_1, \dots, \mathbf{T}_K\}$, with $\mathbf{T}_k : \mathbb{R}^2 \to \mathbb{R}^2$, for $k \in \{1, \dots, K\}$. The velocity vector at point $\mathbf{x} \in \mathbb{R}^2$ of the k-th field is denoted as $\mathbf{T}_k(\mathbf{x})$. At time instant t, one of the velocity fields is *active*, *i.e.*, is driving the motion, thus the trajectory is generated according to

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{T}_{k_{t-1}}(\mathbf{x}_{t-1}) + \mathbf{w}_t, \quad t = 2, ..., n, \qquad (1)$$

where $k_t \in \{1, ..., K\}$ is the label of the active field at time t, $\mathbf{w}_t \sim \mathcal{N}(0, \sigma^2_{k_t}\mathbf{I})$ is white Gaussian noise with zero mean and variance $\sigma^2_{k_t}$. We assume that the initial position $\mathbf{x}_1$ follows a known distribution $p(\mathbf{x}_1)$.

Furthermore, we assume that label sequence $k = (k_1, ..., k_n)$ is generated by a Markov chain with space varying transition probabilities, *i.e.*, the next active field depends on the current active field as well as on the position of the object on the image. Formally, model switching is characterized by the transition probabilities,

$$P(k_t = j | k_{t-1} = i, \mathbf{x}_{t-1}) = B_{ij}(\mathbf{x}_{t-1}), \qquad (2)$$

where $B_{ij}(\mathbf{x})$ is the probability of switching from the field i to the field j at position $\mathbf{x}$. Therefore, $\mathbf{B}(\mathbf{x}) = \{B_{ij}(\mathbf{x})\}$ is a field of stochastic matrices which verify the following properties at each position $\mathbf{x}$:

$$B_{ij}(\mathbf{x}) \geq 0, \qquad \sum_{p=1}^{K} B_{ip}(\mathbf{x}) = 1, \qquad \forall i, j \ . \qquad (3)$$

The joint probability function associated to the pair of sequences $\{\mathbf{x}, k\}$ is given by

$$p(\mathbf{x}, k) = p(\mathbf{x}_1, k_1) \prod_{t=2}^{L} p(\mathbf{x}_t | k_t, \mathbf{x}_{t-1}) p(k_t | \mathbf{x}_{t-1}, k_{t-1}), \qquad (4)$$

where

$$p(\mathbf{x}_t | k_t, \mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t | \mathbf{x}_{t-1} - T_{k_t}(\mathbf{x}_{t-1}), \sigma^2_{k_t}\mathbf{I}\right),$$

with $\mathcal{N}(\mathbf{v}|\boldsymbol{\mu}, \mathbf{C})$ denoting a Gaussian density of mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$, computed at $\mathbf{v}$, and $P(k_t | \mathbf{x}_{t-1}, k_{t-1})$ is as given in (2).

Given a training set of S observed trajectories $\mathcal{X} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(S)}\}$, where $\mathbf{x}^{(s)} = (\mathbf{x}^{(s)}_1, \dots, \mathbf{x}^{(s)}_{L_s})$ is the s-th trajectory, we have shown in [1] that the ML estimate of the model (*i.e.*, the set of motion fields $\mathcal{T}$, the switching field $\mathbf{B}$, and the noise variances $\boldsymbol{\sigma} = (\sigma^2_1, ..., \sigma^2_K)$) can be obtained efficiently via an EM algorithm. That EM algorithm results from treating the label sequences $\mathcal{K} = (k^{(1)}, ..., k^{(S)})$ as missing data, and yields a maximum marginal likelihood (MML) estimate,

$$\widehat{\theta} = \arg\max_{\theta} \sum_{s=1}^{S} \log p(\mathbf{x}^{(s)} | \theta),$$

where $p(\mathbf{x}^{(s)}|\theta)$ is the marginal

$$p(\mathbf{x}^{(s)}|\theta) = \sum_{k^{(s)} \in \{1, \dots, K\}^{L_s}} p(\mathbf{x}^{(s)}, k^{(s)} | \theta)$$

and $\theta = (\mathcal{T}, \mathbf{B}, \boldsymbol{\sigma})$.

3. TRAJECTORY CLASSIFICATION

Each trajectory class, say $j \in \{1, ..., J\}$, is represented by a model of the type described in the previous section, that is, by $\theta^{(j)} = (\mathcal{T}^{(j)}, \mathbf{B}^{(j)}, \boldsymbol{\sigma}^{(j)})$. Given a set of training trajectories from each of the classes, these models may be estimated via EM, as referred in the previous section. These generative models can then be used as class-conditional densities to obtain the *maximum a posteriori* (MAP) classifier, following the standard Bayesian classification paradigm. Formally, the predicted class $\widehat{j}(\mathbf{x})$ for some new observed trajectory $\mathbf{x}$ is given by

$$\widehat{j}(\mathbf{x}) = \arg \max_{j \in \{1,\ldots,J\}} \left\{ \log p(\mathbf{x}|\theta^{(j)}) + \log P(j) \right\},$$

where $P(j)$ is the *a priori* probability of class j, herein assumed to be simply $P(j) = 1/J$.

4. DISCRIMINATIVE MODEL SELECTION

A fundamental question that we have left aside up to this moment is: how should we select the number of motion fields used to model each activity class? That's the question to which we provide an answer in this section.

Let $\mathbf{K} = (K_1, ..., K_J)$, referred to as a *model configuration*, denote the numbers of fields used in the model of each of the activity classes. We assume that these numbers satisfy $1 \leq K_j \leq M$, for $j \in \{1, ..., J\}$, where M is the maximum number of vector fields allowed in each class-conditional model.

Assume that in addition to the training set from which the generative model of each class is estimated, we have a so-called *selection set*, with trajectories from all the classes: $\mathcal{D} = (\mathcal{D}^{(1)}, ..., \mathcal{D}^{(J)})$, where $\mathcal{D}^{(j)} = \{\mathbf{x}^{(j,1)}, \ldots, \mathbf{x}^{(j,D_j)}\}$ denotes a set of D_j trajectories from class j.

The discriminative performance of a given model configuration $\mathbf{K}$ is obtained naturally on the selection set as

$$\mathcal{M}(\mathbf{K}, \mathcal{D}) = \sum_{j=1}^{J} \sum_{l=1}^{D_j} \delta\Big(j, \arg \max_{r \in \{1,\ldots,J\}} p(\mathbf{x}^{(j,l)}|\boldsymbol{\theta}_{\mathbf{K}}^{(r)})\Big), \tag{5}$$

where $p(\mathbf{x}^{(j,l)}|\boldsymbol{\theta}_{\mathbf{K}}^{(r)})$ is the likelihood of the trajectory $\mathbf{x}^{(j,l)}$ under the model for class r in a model configuration $\mathbf{K}$, and $\delta(a, b) = 1$, if $a = b$, and zero otherwise.

The discriminative model selection criterion thus consists in finding the model configuration that maximizes the function $\mathcal{M}(\mathbf{K}, \mathcal{D})$, that is

$$\widehat{\mathbf{K}} = \arg \max_{\mathbf{K} \in \{1,\ldots,M\}^J} \mathcal{M}(\mathbf{K}, \mathcal{D}).$$

If M and J are not too large, it is feasible to find this maximum by exhaustive search over all M^J possible configurations. In some cases, this number may be reduced by considering some model restrictions; for example, one may assume that all the classes share the same number of models, thus $\mathbf{K} = K(1, ..., 1)$, and the search space becomes simply $\{1, ..., M\}$.

5. EXPERIMENTS

5.1. Synthetic data

We first consider a synthetic experiment with two trajectory classes ($J = 2$), generated according to the model described in Section 2. Class 1 simulates a "left-turn" and is well described by two uniform motion fields, one pointing right and the other pointing up; switching between the two fields may occur with probability 0.5 near the center of the image and with probability zero elsewhere. Class 2 corresponds to walking in an approximately straight line from left to right; this is well modeled by a single right pointing uniform field (of course, there's no switching as there is only one field). Fig. 2 shows examples of trajectories from these two synthetic classes.

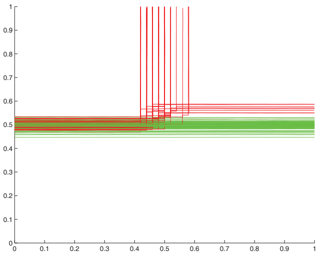

Fig. 2. Synthetic trajectories from the two activities classes described in the text; class 1 in red; class 2 in green.

Both the training and the selection sets each contains 100 trajectories (50 per class). We consider both the fully unrestricted case, where $\mathbf{K} = (K_1, K_2) \in \{1, 2, 3\}^2$ (9 possible configurations) and the restricted case, where $\mathbf{K} = K(1, 1)$, with $K \in \{1, ..., 5\}$ (5 possible configurations). Fig. 3 shows the classification accuracies on the selection set for the unrestricted case (left) and restricted case (right). It is clear from these plots that any model configuration with 2 or more models in class 1 achieves the maximum accuracy, thus the reasonable choice is $\mathbf{K} = (2, 1)$.

Fig. 3. Classification accuracies for the unrestricted (left) and restricted (right) model configurations.

5.2. Real Data

For the real data experiments, we use the UCSD dataset [10], which contain 189 labeled trajectories. There are two different activities: "walking towards the camera" (class 1) and "walking away from the camera" (class 2). Fig. 4 shows examples of trajectories of both classes (in red and green).

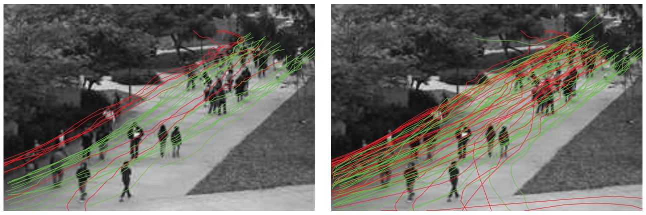

Fig. 4. Examples of trajectories of the two classes: “walking towards camera” (red), “walking away from the camera” (green).

For this experiment, we used 10 trajectories per class for training and the remaining trajectories as the selection set. Fig. 5 shows the classification accuracy for the unrestricted model configurations. It is clear that all the configurations with 1 model in class 1 achieve the highest accuracy, thus the reasonable choice is the simplest of these model configurations: $\mathbf{K} = (1, 1)$. Since the restricted set (in this case) can not achieve higher accuracy, we don’t include the corresponding results. As expected, each activity in the UCSD dataset requires a single motion vector field to be well classified. Fig. 6 depicts the estimated motion fields for the two classes.

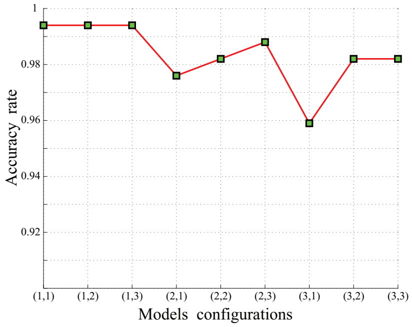

Fig. 5. Classification accuracy with supervised (left) and unsupervised (right) schemes.

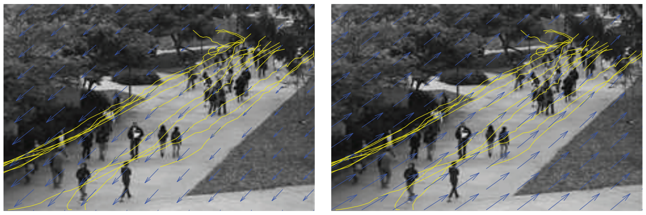

Fig. 6. Estimates of the motion fields (blue arrows) and some unlabeled trajectories (yellow lines).

6. CONCLUSIONS

We have presented a discriminative method for model selection in the context of trajectory classification using a recently introduce generative motion model. This motion model is based on a mixture of vector fields, so the model selection issue (choosing the number of fields) is of central importance. We have presented simple proof-of-concept experiments using both synthetic and real data. The results reported suggest that the discriminative model selection criterion is capable of selecting the model configuration leading to the highest classification accuracy.

Future work will include more exhaustive testing on more challenging problems. Another direction for future research will consider the problem of searching for the best model configuration when the number of classes and/or the maximum number of model is too large to allow exhaustive search.

7. REFERENCES

[1] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques, “Vector fields estimation for motion in natural images,” in *Proc. of the IEEE Int. Conf. on Imag. Proc.*, 2009. 1, 2

[2] A. D. Lanterman, “Schwarz, wallace, and rissanen: Intertwining themes in theories of model selection,” 2000. 1

[3] M. H. Hansen and B. Yu, “Model selection and the principle of minimum description length,” *Journal of the American Statistical Association*, vol. 96, pp. 746–774, June 2001. [Online]. Available: http://ideas.repec.org/a/bes/jnlasa/v96y2001mjunep746-774.html 1

[4] A. Bartoli, “Maximizing the predictivity of smooth deformable image warps through cross-validation,” *J. Math. Imaging Vis.*, vol. 31, no. 2-3, 2008.

[5] M. A. T. Figueiredo and A. K. Jain, “Unsupervised selection of finite mixture models,” in *Proc. of the IEEE Int. Conf. on Pattern Recognition*, 2000.

[6] C. M. Bishop, J. M. Winn, and C. C. Nh, “Non-linear bayesian image modelling,” in *In Proc. of the Sixth European Conference on Computer Vision*. Springer-Verlag, 2000, pp. 3–17. 1

[7] Z. Ghahramani and M. J. Beal, “Variational inference for bayesian mixtures of factor analysers,” in *In Advances in Neural Information Processing Systems 12*. MIT Press, 2000, pp. 449–455. 1

[8] A. Corduneanu and C. M. Bishop, “Variational bayesian model selection for mixture distributions,” in *Artificial Intell. and Statistics*, 2001, p. 2734. 1

[9] B. Thiesson and C. Meek, “Discriminative model selection for density models,” 2003. 2

[10] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, “Privacy preserving crowd monitoring: counting people without people models for tracking,” *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, June 2008. 3