Computing the Uncertainty of Motion Fields for Human Activity Analysis*

Jorge S. Marques António R. Moreira João M. Lemos

INESC-ID
Institute for Systems and Robotics (ISR/IST), LARSyS,
Instituto Superior Técnico, University of Lisboa, Portugal.

Abstract

The motion of pedestrians in outdoor scenes has been represented using motion fields that characterize typical pedestrian velocities at each position in the image. Several methods have been proposed to estimate the motion fields parameters from a set of pedestrians trajectories detected in the video signal. However, up to now no attempts have been made to evaluate the uncertainty associated to the estimates. In fact, the estimated fields are usually used as if they were accurate at all pedestrians positions and this is not a reasonable assumption. This paper is a first attempt to calculate the uncertainty of the estimates as a function of the pedestrian position. We consider two estimation frameworks (ridge regression and Bayesian inference) and discuss uncertainty estimates provided by both of them. Experiments under controlled condition are performed to evaluate the merits of these approaches.

Keywords: pedestrian motion, vector fields, uncertainty estimation, regression, Bayesian methods

1 Introduction

The analysis of human activities in video sequences has been thoroughly studied in the last decade [Poppe, 2010], [Turaga et al., 2008]. Many of these works are focused in indoor activities, *e.g.*, the activity of people inside a house [Lara & Labrador, 2013]. However, other works address human motion in wide spaces such as parks or university campi, in order to understand how people explore the space and what kind of activities are performed: which movements are typical and which are rare or abnormal [Aggarwal & Ryoo, 2011]. Many of these models can also be applied to characterize the motion of other objects in the scene (*e.g.*, vehicles [Melo et al., 2006], or animals [Beyan et al., 2013, Poiesi & Cavallaro, 2015]).

One way to characterize human motion consists of extracting the trajectories of pedestrians in a scene and clustering them [Hu et al., 2006, Ferreira et al., 2013, Pires & Figueiredo, 2017], or representing them as the output of dynamical models [Porikli, 2004, Nascimento et al., 2010]. A recent trend describes pedestrian trajectories using a bank of motion fields [Nascimento et al., 2013]. These models are equipped with a switching mechanism in order to allow the switching between different motion regimes. Model learning algorithms have been proposed to estimate the motion fields from video data. O However, up to now no attempt has been made to characterize the uncertainty associated to the motion field estimates.

This paper is a first attempt to address this problem. This question is an important issue since it will provide measures of confidence on the field estimates that may depend on the position of the pedestrian in the scene and the amount of information that was collected in the neighborhood of that position. In addition, this confidence measure may allow the design of new learning strategies targeted at regions in which the confidence is lower and needs to be enhanced.

*Work supported by FCT under contracts PTDC/EEIPRO/0426/2014, UID/CEC/50021/2013 and UID/EEA/50009/2013.

The paper is organized as follows. This section provides an introduction to the problem. Section 2 addresses the motion field model including the motion field representation. Section 3 discusses motion field estimation and uncertainty measurement. Section 4 describes some experimental results and section 5 concludes the paper.

2 Velocity model

For the sake of simplicity, we will consider that all the pedestrians trajectories are generated by a dynamical model with a single motion field

$$x_{t+1} = x_t + T(x_t) + u_t \,, \tag{1}$$

where $x_t \in [0,1]^2$ denotes the position of the pedestrian in the image at time t ($[0,1]^2$ stands for the image plane), $T(x) \in \mathbb{R}^2$ denotes the motion field at position x and u_t is a realization of a white noise process, with Gaussian distribution $u_t \sim N(0,\sigma^2 I)$. The motion field T is unknown and we want to estimate it from the pedestrian trajectories extracted from the video data.

We will also assume that we tracked S pedestrian trajectories, $(x_1^{(s)}, x_2^{(s)}, \ldots, x_{m^{(s)}}^{(s)})$, with $s = 1, \ldots, S$, and that these trajectories are associated to a set of *position/velocity* pairs

$$\mathcal{T} = \{(x_i, v_i), i = 1, \ldots, L\} \,,$$

where $x_i \in [0,1]^2$ stands for the ith position of the pedestrian and $v_i = x_{i+1} - x_i \in \mathbb{R}^2$ is the corresponding velocity (displacement). We have excluded from $\mathcal{T}$ the last position of each trajectory since we cannot compute the displacement in such cases.

Let us assume that the motion field $T : [0,1]^2 \to \mathbb{R}^2$ to be estimated is defined on a $n \times n$ grid of nodes regularly distributed in the image, as proposed in [Nascimento et al., 2013]. The parameters to be estimated are the velocities $t^j \in \mathbb{R}^2$, $j = 1, \ldots, n^2$ at the grid nodes. If x denotes an arbitrary point in the image plane, the motion field at x will be obtained by interpolating the grid velocities

$$T(x) = \sum_{j=1}^{n^2} t^j \phi_j(x),$$

where $\phi_j(x)$ is the bilinear interpolation function associated to the j-th node (see [Nascimento et al., 2013] for details). This equation can be easily written in a simpler way, by using matrix notation

$$T(x) = \left[\begin{array}{cc|c|cc} \phi_1(x) & 0 & \cdots & \phi_{n^2}(x) & 0 \\ 0 & \phi_1(x) & \cdots & 0 & \phi_{n^2}(x) \end{array} \right] \left[\begin{array}{c} t^1 \\ \vdots \\ t^{n^2} \end{array} \right] = \phi(x) T, \tag{2}$$

where T is a vector obtained by concatenating the velocities associated to all the grid nodes. If we assume that each velocity measured in the image, v_i, is corrupted by a random measurement noise, u_i, we can write

$$v_i = \phi(x_i) T + u_i \qquad i = 1, \ldots, L,$$

where, for the sake of simplicity, u_i is assumed to be a Gaussian random vector with zero mean and isotropic covariance matrix: $u_i \sim N(0,\sigma^2 I)$. The last equation can be written in a more elegant way if we aggregate all the observations into a single vector

$$\left[\begin{array}{c} v_1 \\ \vdots \\ v_L \end{array} \right] = \left[\begin{array}{c} \phi(x_1) \\ \vdots \\ \phi(x_L) \end{array} \right] T + \left[\begin{array}{c} u_1 \\ \vdots \\ u_L \end{array} \right],$$

$$v = \Phi T + u \,,$$

where $v, u \in \mathbb{R}^{2L\times 1}, \Phi \in \mathbb{R}^{2L\times 2n^2}, T \in \mathbb{R}^{2n^2\times 1}$. The conditional distribution of v, given T, is

$$v|T \sim N(\Phi T, \sigma^2 I) \,;$$

it is remarked that matrix Φ has a number of lines that equals the double of the number of data points.

 ISBN 978-0-9934207-2-6

3 Field and uncertainty estimation

3.1 Least squares and ridge regression

The simplest approach to estimate the field parameters is based on the minimization of a squared error (least squares) criterion

$$E_{\text{ls}} = \| v - \Phi T \|^2 .$$

The least squares estimate of the grid node velocities is $\hat{T}_{\text{ls}} = (\Phi^T \Phi)^{-1} \Phi^T v$ and the covariance matrix associated with this estimate is [Hastie et al., 2009]

$$\text{Cov}\{\hat{T}_{\text{ls}}\} = \sigma^2 (\Phi^T \Phi)^{-1} .$$

Both expressions require that the matrix $\Phi^T \Phi$ is non-singular. This is usually not true in this problem since there are regions in the image that are never crossed by a trajectory and for which there is no information available.

An alternative to this approach is ridge regression which includes a regularization term that attracts the motion estimates towards zero [Hastie et al., 2009]. The estimation of T given the observations v can be achieved by minimizing

$$E_{\text{ridge}} = \| v - \Phi T \|^2 + \lambda \| T \|^2 ,$$

where $\| \cdot \|$ denotes the ℓ_2 norm. In this case, the ridge estimate for the vector field is $\hat{T}_{\text{ridge}} = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T v$ and the uncertainty is

$$\text{Cov}\{\hat{T}_{\text{ridge}}\} = \sigma^2 (\Phi^T \Phi + \lambda I)^{-1} \Phi^T \Phi (\Phi^T \Phi + \lambda I)^{-1},$$

which can be easily computed. It is stressed that the inverse exists even if no trajectories are observed in the vicinity of a grid node or a set of grid nodes.

The final question is: what is the uncertainty associated to the velocity estimates at an arbitrary position $x \in [0,1]^2$? This can be answered using (2). The velocity estimate is

$$\hat{T}(x) = \phi(x) \hat{T},$$

and the covariance matrix is

$$\Sigma(x) = \phi(x) \text{Cov}\{\hat{T}\} \phi(x)^T . \tag{3}$$

Both can be computed for each point in the image. If we want to characterize the uncertainty by a single number, the trace of matrix $\Sigma(x)$ can be adopted.

3.2 Bayesian inference

Until now T is assumed to be an unknown deterministic variable. Let us assume now that T is a random vector with Gaussian *prior* distribution $T \sim N(T_0, P_0)$. In this case, the *a posteriori* distribution of T is also Gaussian

$$T | v \sim N(\hat{T}, P) ,$$

with mean vector, $\hat{T}$, and covariance matrix, P, given by the Kalman filter equations[Arulampalam et al., 2002]

$$\hat{T} = T_0 + K(v - \Phi T_0) , \tag{4}$$

$$P = (I - K\Phi) P_0 , \tag{5}$$

$$K = P_0 \Phi^T \left(\Phi P_0 \Phi^T + \sigma^2 I \right)^{-1} . \tag{6}$$

In this case, P is the covariance matrix of the state vector, T, after making the observations v, and describes the uncertainty of the motion field at the grid nodes. It should be stressed, however, that these equations involve the inversion of a matrix that may have a large dimension (twice the number of observations) which makes them more complex then the ones obtained for the ridge regression.

 ISBN 978-0-9934207-2-6

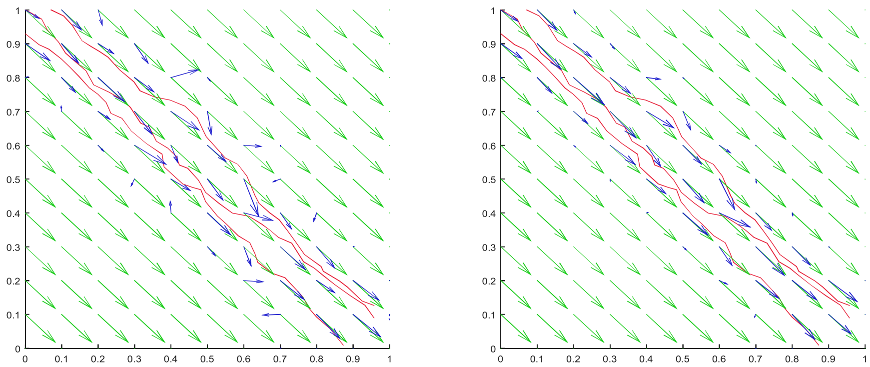

Figure 1: True motion field (green), synthetic trajectories (red), and estimated field (blue) for ridge regression (left) and Bayesian method (right)

4 Experimental results

We performed experiments with synthetic trajectories generated by equation (1) and with pedestrian trajectories extracted from a video signal. In the first case, we know the true motion field used to generate the trajectories. We can therefore compare the motion estimates with the ground truth information. We have also studied the effect of trajectory parameters (*e.g.*, number of trajectories, density) on the model uncertainty. In a third experiment we considered pedestrian trajectories extracted from a video signal, for which the true motion field is not known. All these experiments are carried out using a regular grid of 11×11 nodes, using ridge regression and Bayesian approach (Kalman filtering).

4.1 Bayesian vs ridge regression uncertainty

In the first experiment, we consider a uniform motion field and generated three random trajectories using the dynamic model (1). Then, we estimated the vector field from these trajectories using ridge regression and the Bayesian method under Gaussian hypothesis. Figure 1 shows the vector field used in these experiments (green), three trajectory realizations generated by the model (red) and the field estimates (blue) obtained by the ridge regression and Bayesian method.

We repeated this experiment N times ($N = 100$) and compared the velocity estimates, $\hat{t}_i^k$, obtained at each experiment k and grid node i, with the ground truth velocity vector, t_i^{gt}. The mean square error at the $i-th$ node,

$$E_i = \frac{1}{N} \sum_{k=1}^{N} \| \hat{t}_i^k - t_i^{gt} \|^2 , \tag{7}$$

is a measure of uncertainty. Figure 2 (1st row) shows the mean square error associated with both estimation methods.

This comparison is possible since we know the true value of the vector fields at the grid nodes (ground truth) but cannot be computed when we use real data since we do not have the true motion field in that case.

To circumvent this difficulty, we use the predicted variance associated to the ridge regression and Bayesian estimators (see (3,6)), displayed in Figure 2 (2nd line). They have different structures. The variance of the Bayesian estimates is high in all the regions that have not been visited by a trajectory. In such regions, there is no information available concerning pedestrian motion and the uncertainty is therefore very high since we do not know the direction of the pedestrian motion. On the contrary, the uncertainty is low in regions visited by trajectories, as expected.

The ridge regression performs, however, in a different way. Regions far from the trajectories have zero variance. This is counter intuitive but it is correct since the ridge regression assigns a zero value to the velocity estimates whenever there is no information. The variance of the estimates is therefore equal to zero in such

 ISBN 978-0-9934207-2-6

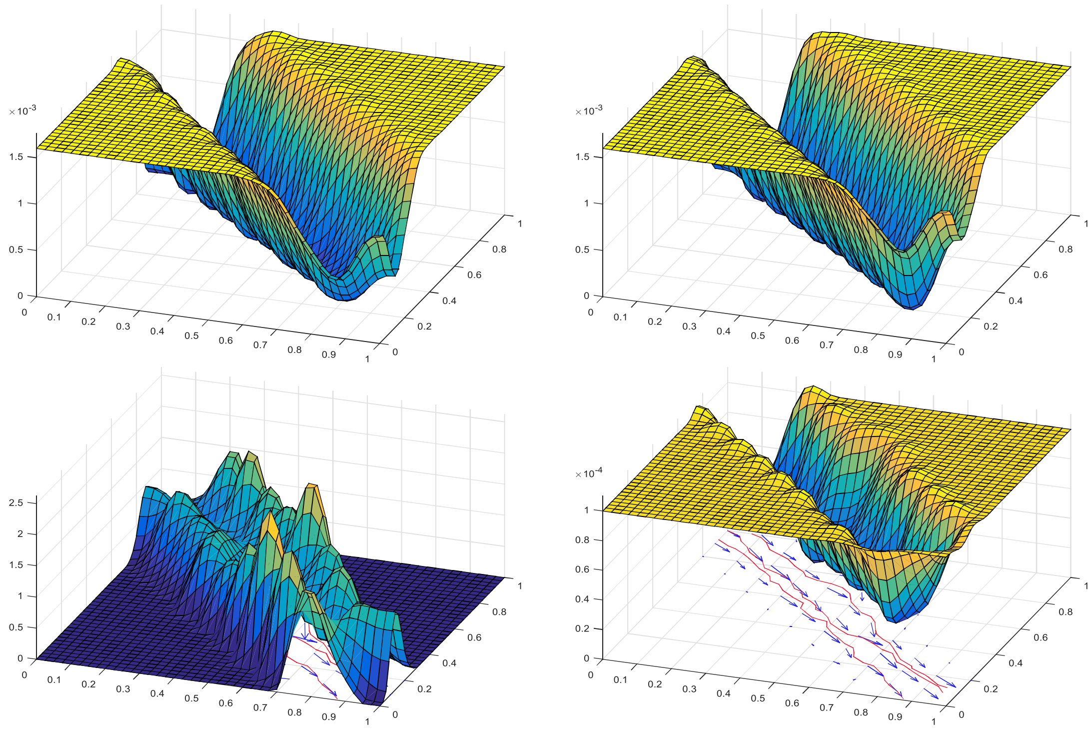

Figure 2: Mean square error estimation of the node velocities by Monte Carlo (1st line) and predicted variance (2nd line), for ridge regression (left) and Bayesian method (right). The Monte Carlo results (1st line) require the ground truth velocities while the variances (2nd line) do not.

cases. This does not mean that there is no error but simply that there is a strong bias associated to the motion estimate at each grid node and zero variance.

We conclude that the results obtained with the Bayesian approach are informative and a reliable measure of the uncertainty. On the contrary, the variance obtained for the ridge regression estimator does not provide a good quality measure because the uncertainty is dominated by a strong bias that cannot be computed from real data. Therefore, we will use the Bayesian approach from now on

4.2 Trajectory parameters

The motion field uncertainty is related to the amount of observations available in the vicinity of each image point. Taking the previous experiment as a reference (3 lines generated by a uniform field) we will i) change the number of trajectories, keeping the same average distance between them and ii) increase the density of trajectories. Typical examples are shown in Figs. 3,4.

The first example (Fig. 3) shows that the variance map changes (the valley gets wider) when the number of lines increases, as expected. The valley is made of points which are close to a trajectory where *close* means that the distance from the point to the trajectory is smaller than the grid step.

The second experiment keeps the width constant but increases the density of trajectories. In this case then valley of the variance map becomes deeper, as expected. This example shows that the uncertainty behaves in a simple and predictable way in this case.

4.3 Video surveillance

The main issue that remains to be tested is how does the uncertainty analysis behave with real data. The next experiment involves the estimation of a motion field from video data, in an outdoor scene (see Fig. 5). First we

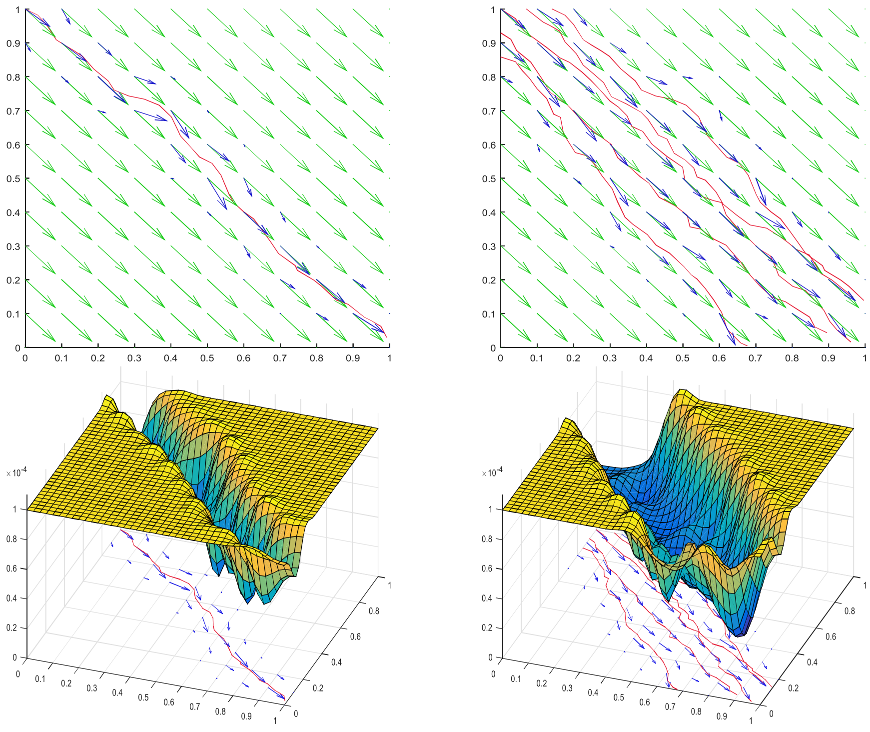

Figure 3: Estimated fields (1st line) and field variances (2nd line) obtained by the Bayesian method: 1 trajectory (left) and 5 equally spaced trajectories (right).

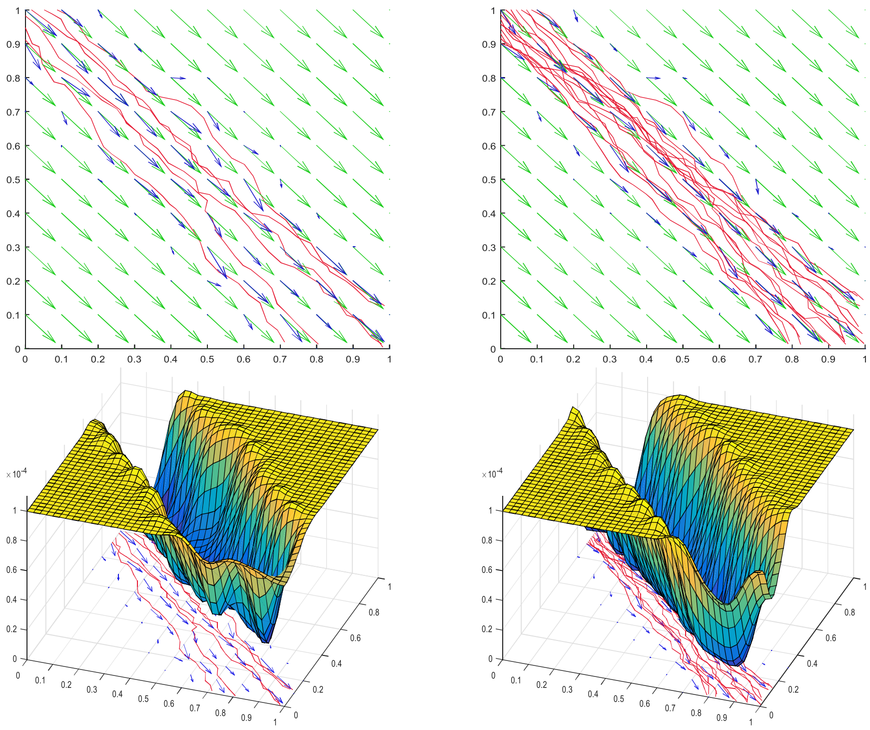

Figure 4: Estimated fields (1st line) and field variances (2nd line) obtained by the Bayesian method with different densities: 5 trajectory (left) and 15 trajectories (right).

 ISBN 978-0-9934207-2-6

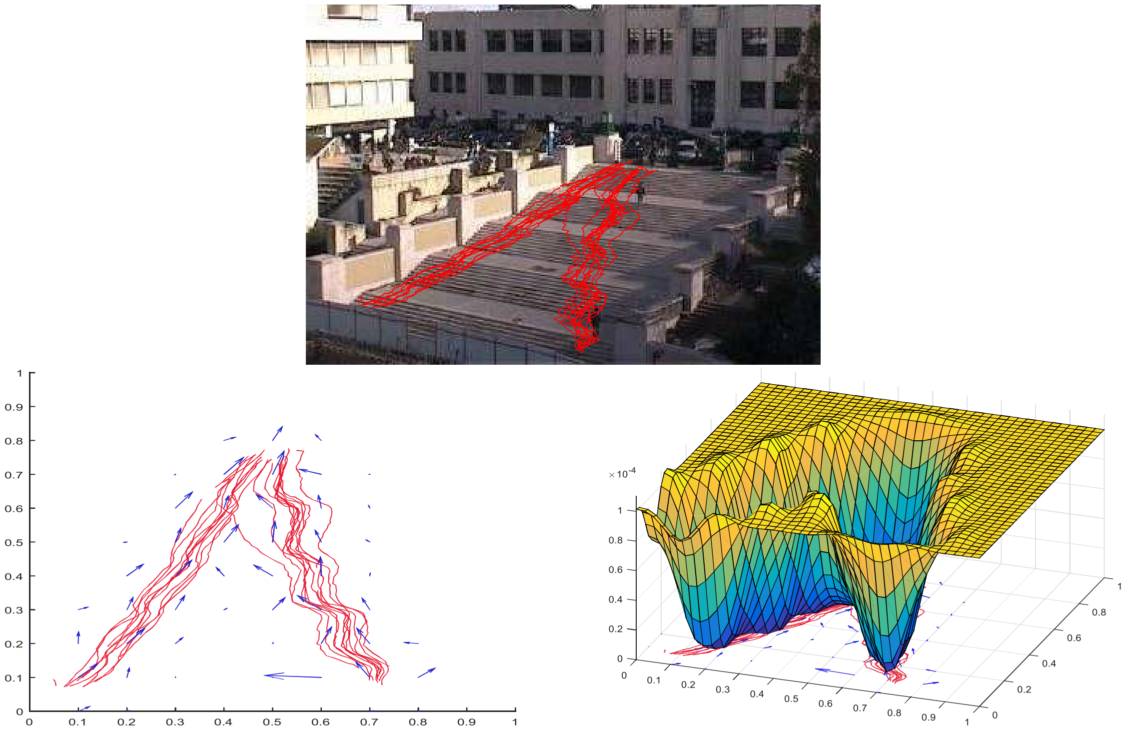

Figure 5: Staircase scene (1st line); 2nd line: pedestrians trajectories and estimated motion fields using the Bayesian method (left) and model variance (right).

extracted the pedestrians trajectories. The trajectories were then transformed by an homography to compensate for the perspective distortion. A motion field was estimated using the Kalman filter (6). Fig. 5 (2nd line left), shows the estimated field. There are some wrong estimates at nodes located far from the trajectories in which there is a large uncertainty. The variance map (Fig. 5 (2nd line right)) shows a small variance in regions close to the trajectories, as expected, and a large variance in regions that were not visited by trajectories.

This experiment shows that the uncertainty measures can be applied to pedestrian trajectories extracted from video data and produce acceptable results.

5 Conclusions

Several algorithms have been proposed to estimate motion fields from trajectories. However, the estimation of the uncertainty associated with the motion estimates is usually not addressed. This is an important issue since we must know what is the confidence associated to the velocity estimates in each position. It does not matter if we have velocity estimates in all image regions if the uncertainty is very high. Another important issue concerns data management. When we receive new trajectories we should be able to decide whether they should be accepted, in case they improve the motion field estimates, or if they should be rejected, in case there is already a lot of information in that region. This kind of decision can be taken, if we have uncertainty estimates. Furthermore, estimating the uncertainty provides a way to address the problem of finding the number and space distribution of trajectories that allows to estimate the field in the whole image with a maximum error bound.

This paper discusses two ways to estimate uncertainty. The one that achieves the best results is the Bayesian method, based on the Kalman update equations. The main conclusion can be loosely stated as follows: uncertainty is low at image points x such that the vicinity of x is crossed by a sufficient number of trajectories.

Future work should address the application of motion uncertainty in the management of new data, by defining criteria to accept new data for motion field update. The extension of these results to multiple motion

 ISBN 978-0-9934207-2-6

field models should also be considered and it is a challenging problem since the Gaussian hypothesis is no longer valid.

References

[Aggarwal & Ryoo, 2011] Aggarwal, J. & Ryoo, M. (2011). Human activity analysis: A review. *ACM Computing Surveys*.

[Arulampalam et al., 2002] Arulampalam, M., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, (pp. 174–188).

[Beyan et al., 2013] Beyan, C., Cigdem, & Fisher, R. (2013). Detection of abnormal fish trajectories using a clustering based hierarchical classifier. In *BMVC*.

[Ferreira et al., 2013] Ferreira, N., Silva, C., Klosowski, J., & Scheidegger, C. (2013). Vector field k-means: Clustering trajectories by fitting multiple vector fields. *Computer Graphics Forum*.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

[Hu et al., 2006] Hu, W., Xiao, X., Fu, Z., Xie, D., Tan, T., & Maybank, S. (2006). A system for learning statistical motion patterns. *IEEE transactions on pattern analysis and machine intelligence*, (pp. 1450–1464).

[Lara & Labrador, 2013] Lara, O. & Labrador, M. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, (pp. 1192–1209).

[Melo et al., 2006] Melo, J., Naftel, A., Bernardino, A., & Santos-Victor, J. (2006). Detection and classification of highway lanes using vehicle motion trajectories. *IEEE Transactions on intelligent transportation systems*, (pp. 188–200).

[Nascimento et al., 2010] Nascimento, J. C., Figueiredo, M. A. T., & Marques, J. S. (2010). Trajectory classification using switched dynamical hidden markov models. *IEEE Transactions on Image Processing*, (pp. 1338–1348).

[Nascimento et al., 2013] Nascimento, J. C., Figueiredo, M. A. T., & Marques, J. S. (2013). Activity recognition using mixture of vector fields. *IEEE Transactions on Image Processing*, (pp. 1712–1725).

[Pires & Figueiredo, 2017] Pires, T. & Figueiredo, M. (2017). Shape-based trajectory clustering. In *International Conference on Pattern Recognition Applications and Methods*.

[Poiesi & Cavallaro, 2015] Poiesi, F. & Cavallaro, A. (2015). Tracking multiple high-density homogeneous targets. *IEEE Transactions on Circuits and Systems for Video Technology*, (pp. 623–637).

[Poppe, 2010] Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing*, (pp. 976–990).

[Porikli, 2004] Porikli, F. (2004). Learning object trajectory patterns by spectral clustering. In *IEEE Int. Conf. Multimedia Expo*.

[Turaga et al., 2008] Turaga, P., Chellappa, R., Subrahmanian, V., & Udrea, O. (2008). Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, (pp. 1473–1488).