A Probabilistic Background Model for Tracking

J. Rittscher[1], J. Kato[1], S. Joga[2], and A. Blake[3]

[1] University of Oxford, Department of Engineering Science, Parks Road, Oxford OX1 PJ3, UK
[2] École Nationale Supérieure des Télécommunications, 46 rue Barrault, 75634 Paris Cedex 13, France
[3] Microsoft Research ,1 Guildhall Street, Cambridge, CB2 3NH, UK

Abstract. A new probabilistic background model based on a Hidden Markov Model is presented. The hidden states of the model enable discrimination between foreground, background *and* shadow. This model functions as a low level process for a car tracker. A particle filter is employed as a stochastic filter for the car tracker. The use of a particle filter allows the incorporation of the information from the low level process via importance sampling. A novel observation density for the particle filter which models the statistical dependence of neighboring pixels based on a Markov random field is presented. The effectiveness of both the low level process and the observation likelihood are demonstrated.

1 Introduction

The main requirement of a vision system used in automatic surveillance is robustness to different lighting conditions. Lighting situations which cast large shadows are particularly troublesome (see figure 1) because discrimination between foreground and background is then difficult. As simple background subtraction or inter-frame differencing schemes are known to perform poorly a number of researchers have addressed the problem of finding a probabilistic background model [6,17,10,13,20]. Haritaoglu *et al.* [6] only learn the minimal and maximal grey-value intensity for every pixel location. The special case of a video camera mounted on a pan-tilt head is investigated in [17]. Here a Gaussian mixture model is learnt. Paragios and Deriche [13] demonstrate that a background foreground/segmentation based on likelihood ratios can be elegantly incorporated into a PDE Level Set approach. In order to acquire training data for these methods it is necessary to observe a static background without any foreground objects. Toyama *et al.* [20] address the problem of background maintenance by using a multi-layered approach. The intensity distribution over time is modelled as an autoregressive process of order 30. This seems to be an unnecessarily complex model for a background process. None of the above models are able to discriminate between background, foreground, and shadow regions. In the present paper we propose a probabilistic background model based on a Hidden Markov Model (HMM). This model has two advantages. Firstly it is no longer necessary to select training data. The different hidden states allow the learning of distributions for foreground and background areas from a mixed sequence. By adding a third state it is possible to extend the model so that it can discriminate shadow regions. The background model is introduced in section 2.

In addition to the low level process it is necessary to build a high level process that can track the vehicles. Probabilistic trackers based on a particle filters [7] are known to

D. Vernon (Ed.): ECCV 2000, LNCS 1843, pp. 336–350, 2000.

be robust and can be extended to tracking multiple objects [11]. The benefit of using a particle filter is that the tracker can recover from failures [7]. But very importantly the use of a particle filter also allows a way to utilise the information of the low level process modelled by the HMM. The propagated distribution for the previous time-step $t-1$ is effectively used as a prior for time t. It is very difficult to fuse two sources of prior information. However, importance sampling, as introduced in [8], can be used to incorporate the information obtained from the low level process. Instead of applying the original algorithm an importance sampling scheme which is linear in time [16] is used here. The importance function itself is generated by fitting a rectangle with parameters X_I to the pixels which are classified as foreground pixels (see figure 5) and using a normal distribution with fixed variance and mean X_I as the importance function. The remaining challenge is to build an observation likelihood for the particle filter which takes account of spatial dependencies of neighbouring pixels. The construction of this observation likelihood is discussed in in section 3. We demonstrate that by employing a Markov random field it is possible to model these statistical dependencies.

Such a car tracking system has to be able to compete with existing traffic monitoring systems. Beymer *et al.* [2] built an very robust car tracker. Their tracking approach is based on feature points and works in most illumination conditions. The disadvantage of the system is that it is necessary to run a complex grouping algorithm in order to solve the data association problem. The use of additional algorithms would be necessary to extract information about the shape of the cars. By modelling cars as rectangular regions it would be possible to infer about their size and allow classification into basic categories. Koller *et al.* [10] as well as Ferrier [4] *et al.* already demonstrated applications of contour tracking to traffic surveillance. [10] extracts a contour extraction from features computed from inter-frame difference images as well as the grey value intensity images themselves. In the case of extreme lighting conditions as shown in figure 1 this system is likely to get distracted. Approaches which model vehicles as three dimensional wire frame objects [18,12,15] are of course less sensitive to extreme lighting conditions. The main drawback of modelling vehicles as three dimensional objects is that the tracking is computationally expensive. The challenge is to design a robust real-time system which allows the extraction shape information.

2 A Probabilistic Background Model

In addition to being able to discriminate between background and foreground it is also necessary to detect shadows. Figure 2 clearly shows that the grey-value distributions of the shadow differs significantly from the intensity distributions in the foreground and background regions. This is the motivation for treating the shadow region separately. Since all three distributions have a large overlap it is of course not possible to construct a background model which is purely based on intensity values. However another source of information is available: the temporal continuity. Once a pixel is inferred to be in a foreground region it is expected to be within a foreground region for some time. An suitable model to impose such temporal continuity constraints is the Hidden Markov Model (HMM) [14]. The grey-value intensities over time for one specific pixel location is to be modelled as a single HMM, independent of the neighbouring pixels. This is of course an

Fig. 1. A traffic surveillance example. *This is a typical camera image from a traffic surveillance camera. Notice that especially for dark coloured cars intensity differences between foreground and background are small. In order to track the cars robustly it is necessary to detect the shadows as well as the cars.*

unrealistic independence assumption. The spatial dependencies of neighbouring pixel locations will be modelled by the higher level process (see section 3). The reader should note that the specific traffic surveillance situation (see figure 1) is particularly suited to investigate this class of model because the speed of the cars does not vary greatly. It is therefore possible to learn parameters which will determine the expected duration a pixel belongs to a foreground, shadow or background region.

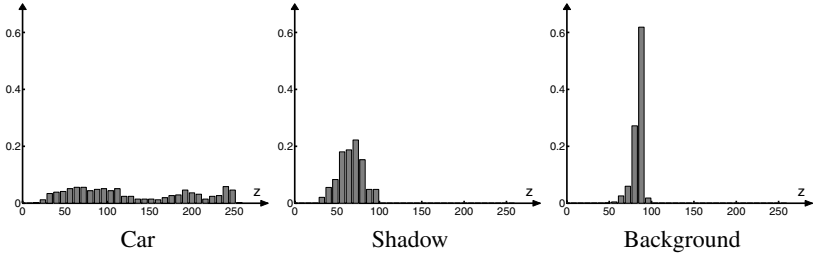

Fig. 2. Intensity histograms of the different regions. *Intensity values for single pixel positions were collected from a 30 seconds long video sequence and manually classified into the regions: foreground, shadow or background. The intensity histograms of the different regions clearly show a large amount of overlap. A method which is purely based on grey-value intensities is therefore inadequate for this problem.*

The model parameters of the HMM with N states are the initial state distribution $\pi = \{\pi_i\}$, the state transition probability distribution $A = \{a_{i,j}\}$, and the emission or observation probability for each state $p_f(z), p_b(z)$ and $p_s(z)$. The set of parameters defining the HMM model will be abbreviated as $\omega := (A, \pi, p_f, p_s, p_b)$. Standard texts

include [14,9]. Based on the intensity histograms of figure 2 the emission models of the background and shadow regions are modelled as Gaussian densities. Since very little about the distribution of the colours of vehicles is known, the observation probability of the foreground region is taken to be uniform. Hence

$$p_f(z) = \frac{1}{256}, \quad p_s(z) = \frac{1}{\sqrt{2\pi\sigma_s^2}} e^{-\frac{(z-\mu_s)^2}{2\sigma_s^2}}, \quad \text{and} \quad p_b(z) = \frac{1}{\sqrt{2\pi\sigma_b^2}} e^{-\frac{(z-\mu_b)^2}{2\sigma_b^2}}. \tag{1}$$

It is of course possible to employ more complex emission models. In section 2.2 it will be shown that is in fact necessary to use a more complex model for the observations.

2.1 Parameter Learning

For a given training sequence the model parameters are estimated by using a maximum likelihood approach. Because the model has hidden parameters an expectation maximisation (EM) type approach is used. In this particular case the Baum Welch algorithm [9] is applied as a learning algorithm. Because EM-type algorithms are not guaranteed to find the global maximum and are very sensitive to initialisation it is necessary to explain how the initialisation is done. In order to find an initialisation method the following time

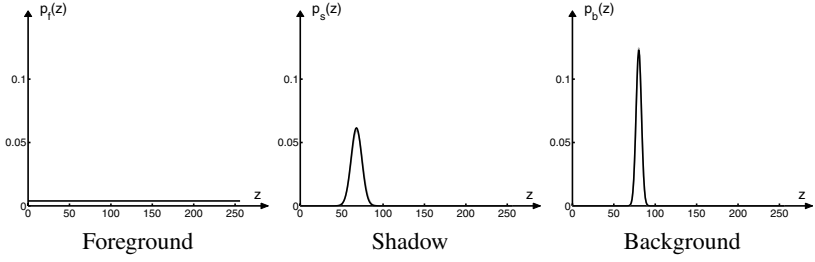

Fig. 3. Learnt emission models. *Shown is a set of emission models for one pixel location. The distributions p_f, p_s and p_b model the intensity distributions for all three states foreground, shadow and background. It should be noted that the emission models can vary between pixel locations.*

constants are defined: τ_b - the typical time duration a pixel belongs to the background, and τ_s, τ_f the typical duration for shadow and foreground. Let λ_b, λ_s, and λ_f be the proportion of the time spent in background, shadow and foreground, with $\lambda_f + \lambda_s + \lambda_b = 1$. All these parameters are determined empirically. Using these definitions an intuitive transition matrix can be chosen as

$$A = \begin{pmatrix} 1-\tau_b^{-1} & \tau_b^{-1}\Lambda_{sf} & \tau_b^{-1}\Lambda_{fs} \\ \tau_s^{-1}\Lambda_{bf} & 1-\tau_s^{-1} & \tau_s^{-1}\Lambda_{bs} \\ \tau_f^{-1}\Lambda_{bs} & \tau_s^{-1}\Lambda_{sb} & 1-\tau_f^{-1} \end{pmatrix}, \tag{2}$$

where $\Lambda_{ij} = \lambda_i/(\lambda_i + \lambda_j)$. The initial state distribution π is chosen to be $\pi = \{\lambda_b, \lambda_s, \lambda_f\}$. The mean of the observation density for the background state μ_b can be

estimated to be the mode of the intensities at a given pixel since $\lambda_b \gg \lambda_s$ and $\lambda_b \gg \lambda_f$. The variance σ_b^2 is determined empirically. The initial parameters of the observation density for the shadow region are based on the assumption that the shadow is darker than the background, i.e.

$$\mu_s = \frac{\mu_f + 2\sigma_b}{2}, \qquad \text{and} \qquad \sigma_s = \frac{\mu_s}{2} \ . \tag{3}$$

This ensures that $\mu_s < \mu_b$ in case $\mu_b > 2\sigma_b$, i.e. the background intensities are not as low as intensities in the shadow areas. At each iteration of the Baum Welch algorithm, the backward and forward variables are rescaled for reasons of numerical stability [9]. It is not necessary to learn a transition probability distribution A for every pixel. By learning one transition probability distribution for an observation window the complexity of the learning is reduced considerably. A set of learnt emission models are shown in figure 3. The corresponding transition probability distribution is of the form

$$A = \begin{pmatrix} 0.986 & 0.012 & 0.001 \\ 0.013 & 0.884 & 0.101 \\ 0.033 & 0.025 & 0.941 \end{pmatrix} \ , \tag{4}$$

A close inspection of these transition probabilities reveals that during learning dark cars are mistaken for shadows. As a consequence the expected duration for being in a foreground state is unrealistically short. For the particular lighting situation (see figure 1) it is possible to solve the problem by adding the constraint $a_{fs} = 0$. This implies that the transition probability from foreground to shadow should be zero. Of course this constraint cannot be applied in the general case. It is therefore necessary to find a more general solution. As a result the parameters of the observation density for the shadow change. Especially the variance σ_S is now smaller $\sigma_S = 41.95$ instead of 44.97. The corresponding transition matrix A is

$$A = \begin{pmatrix} 0.980 & 0.015 & 0.003 \\ 0.013 & 0.897 & 0.891 \\ 0.047 & 0.000 & 0.952 \end{pmatrix} \ , \tag{5}$$

notice that the values of a_{ff} is increased.

2.2 Two Observations Improve the Model

Initial experiments show that by using only one observation, dark cars are not detected sufficiently well (see figure 4). In order to make the method more robust, it is desirable to reduce the amount of overlap of the observation densities. In particular it is necessary to reduce the ambiguity between dark foreground regions and shadows. These ambiguities can be reduced by introducing a second observation. To be precise the responses of two different filters will be used. The HMM is no longer modelled for every pixel but for sites on a lattice such that the filter supports of the different sites do not overlap. As a first observation a simple 3×3 average is used. It can be observed that background and shadow regions are more homogeneous than foreground regions. It would therefore make sense to introduce a second observation which measures the intensity variation in

a small neighbourhood each pixel. In order to test this approach a simple 3×3 Sobel filter mask is used as a second observation. It is possible to show empirically that for this specific data, the responses of the Sobel filter and the mean intensity response at a pixel are uncorrelated. Hence the two observations are considered to be independent. The comparison shown in figure 4 shows that the use of two observations greatly improves the detection of dark cars. Whereas the choice of the average filter is justified the chosen Sobel filter is by no means optimal. A filter which implies computing a higher order derivative of the image data as for example a Laplace filter or even a spatio temporal filter might be a much better alternative.

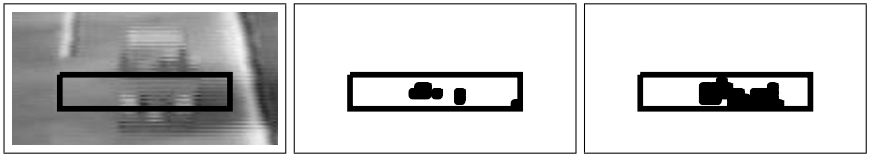

Fig. 4. Using two observations improves the model. *For each time step t every pixel is classified to be in a foreground, background, or shadow region. For visualisation purposes the pixels for which the forward probability $p(z_t, z_{t-1}, Y_t = f|\omega)$ is greater than the forward probability for the alternative states are marked in black. The image on the left shows the raw data. The black box indicates the area in which the model is tested. The two images on the right show the sets of pixels which are classified as foreground pixels. It shows that the classification based on two observations (right) is superior to the method based on only one measurement (middle).*

2.3 Practical Results

In order to test the performance of the model the forward probabilities $p(z_t, z_{t-1}, Y_t|\omega)$ are evaluated for the three different states $Y_t \in \{f, b, s\}$ for each time-step t. The discrete state Y_t for which the forward probability is maximal is taken as a discrete label. By determining discrete labels this classification method discards information which could be used by a higher level process. But for now this should be sufficient to discuss the results obtained with the method. Two typical results are shown in figure 5. A movie which demonstrates the performance of this process can be found in the version of this paper on our web site (http://www.robots.ox.ac.uk/∼vdg). The interior of the car is not detected perfectly. But there is clearly enough information to detect the boundaries of the vehicle. In order to illustrate the importance of the state transition probability the matrix A was altered by hand. The results are presented in figure 6 and display clearly that the transition probability plays an important role. The effect is of course most evident when the discrimination based on measurements alone is ambiguous.

3 The Car Tracker

The remaining challenge is to build a robust car tracker. Probabilistic trackers based on a particle filters [7] are known to be robust and can be extended to tracking multiple

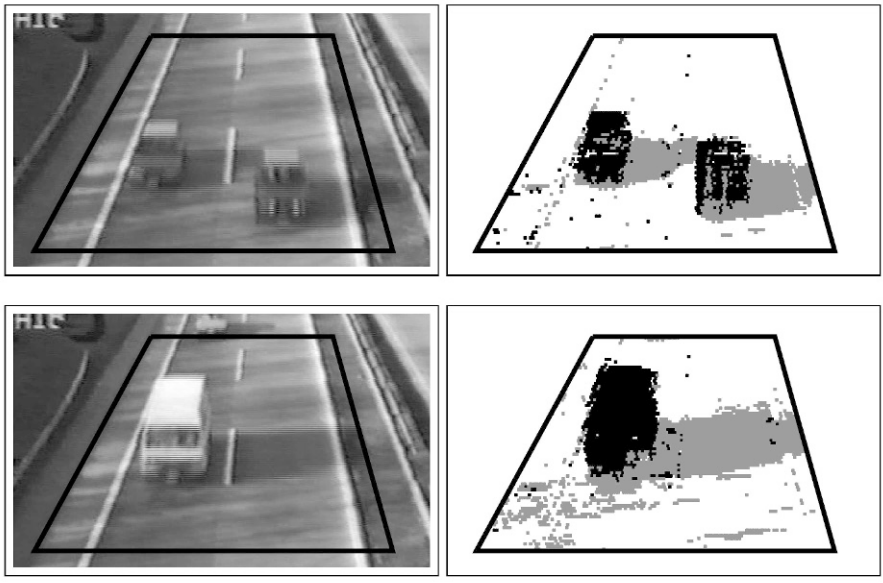

Fig. 5. Results of the background modelling. *The discrete label Y_t for which the forward probability $p(z_t, z_{t-1}, Y_t|\omega)$ is maximal is used as a discrete label for visualisation (see text). Foreground pixels are marked in black, shadow pixels in grey, and background pixels in white. It should be noted that even for dark coloured cars the results are respectable. The labels will then be used by a higher level process to locate the vehicles.*

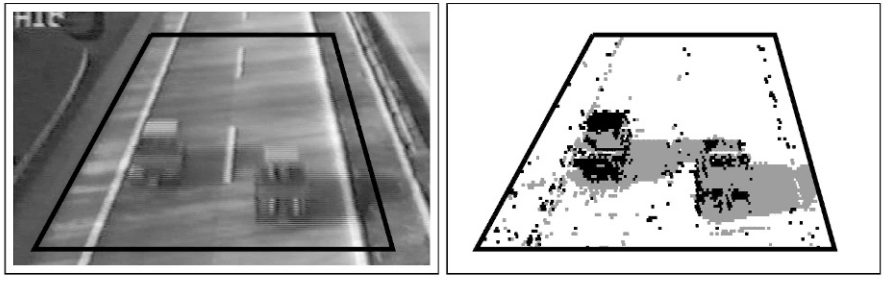

Fig. 6. Importance of the temporal continuity constraint. *Like in figure 5 the pixels are assigned a discrete label Y_t as which forward probability $p(z_t, z_{t-1}, Y_t|\omega)$ is maximal. In this experiment the transition probability of a model which uses two observations was altered such that all $a_{ij} = 1/3$ in order to explore the importance of the temporal continuity constraint. Each pixel is classified (see text) as foreground (in black), shadow (in grey) or background (white). A comparison with the images shown in figure 5 shows that these results are clearly worse. Obviously the transition probability A plays a crucial role.*

objects [11]. In order to build such a tracker it is necessary to model the observation likelihood

$$p(Z|X;\vartheta) \tag{6}$$

for a set of measurements Z and a hypothesis X. The parameters of the model are denoted by ϑ. For the present purpose it is sufficient to model the outlines of the cars as a perspectively distorted rectangle which will be parameterised by the state vector X. In order to track cars robustly it is not sufficient to take edge measurements as in [7]. [19] showed that detection of the background aids finding the foreground object. The problem is that in this case the measurements Z cannot be assumed to be independent (also see [19]). These conditions lead us to model the likelihood (6) as a conditioned Markov random field (MRF) (see for example [5,21]). In Gibbs form an MRF can be written as

$$P(Z|X;\vartheta) = \frac{\exp(-H^{\vartheta}(Z,X))}{\sum_{Z'\in\mathcal{Z}} \exp(-H^{\vartheta}(Z',X))} \ . \tag{7}$$

The denominator of the fraction is known as the partition function of the MRF. The difficulty is now to find a model which is tractable yet still captures the spatial dependence of neighbouring measurements.

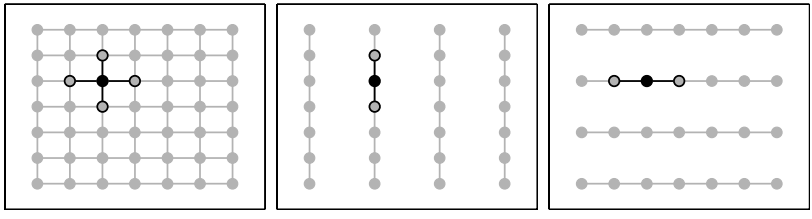

Fig. 7. Neighbourhood structure of the MRF. *The set of sites on a lattice S is marked by circles. The neighbourhood structure at one particular site s (marked as a filled black circle) is different in each case. The neighbours $r \in \delta(s)$ of the site s are marked by black circles which are filled grey. The set of cliques are indicated by lines connecting neighbouring sites.*

3.1 Modelling the Observation Likelihood

As mentioned in the previous section, one difficult problem is to find an energy function H for which the likelihood $P(Z|X;\vartheta)$ can be evaluated efficiently. The energy function H will depend on a lattice S and a corresponding *neighbourhood system* $\delta := \{\delta(s) : s \in S\}$ (see figure 7). The set of cliques will be denoted by $\mathcal{C}$. In order to take the distribution of the measurement z at a given site and the statistical dependence of measurements at neigbouring sites into account we let the energy function

$$H_A^{\vartheta}(Z,X) = \sum_{s\in A_X} g_A(z_s) + \sum_{(s,r)\in\mathcal{C}\cap A_X^2} \vartheta_A \cdot (z_s - z_r)^2 \ , \tag{8}$$

where A_X denotes an area which is either in the foreground or background, i.e. $X \in \{B, F\}$. Since the function g_A models the distribution of the measurement at a given site it would be ideal if one could make use of the emission models which were learnt for the different states of the HMM (see section 2.1). But as it will be shown later the energy function needs to be translational invariant (13) so therefore g_A cannot depend on a particular site s. And in order to compute the partition function efficiently (section 3.3) it is necessary that the functions g_f and g_b are normal distributions. The foreground distribution g_f is therefore chosen to be a normal distribution with a large variance. The background distribution g_f is taken to be the normal with mean μ_f and variance σ_f such that it approximates the the mixture of the background and shadow emission models (1) learnt by the HMM.

The set of sites which belong to a given area A_X depends of course on the hypothesis X. Because the partition function depends also on X it will be necessary to evaluate it for every hypothesis X. It turns out that if the lattice S is two dimensional, the partition function is too expensive to compute. In the following it is explained that it is not possible to approximate the observation likelihood (7). It is therefore necessary to find a simpler model. It is known that under certain conditions the pseudolikelihood function [1,21], defined as

$$\prod_{s \in S} p(z_s | z_{S \setminus s}; \vartheta) \tag{9}$$

can be used for parameter estimation instead of the Maximum Likelihood approach based on the MRF (7). It can be shown [21] that estimators obtained by maximizing the pseudolikehood can compete in terms of statistical properties with maximum likelihood estimators. Although some authors state that when the variables are weakly correlated, the pseudolikehood is a good approximation to the likelihood [3] it seems to be an open problem under which conditions precisely it can be used as an approximation to the likelihood function. In section 3.2 it will also become clear why the pseudolikehood method cannot be used to estimate X. An alternative is to restrict the MRF to measurements on scan lines taken out of the image. This will simplify the model considerably. The observation likelihoods of the different scan lines will be treated as independent. Based on the grid in figure 11 it is possible to formulate a random field for each of the horizontal $\{h_i\}$ and vertical lines $\{v_i\}$. The likelihood is now of the following form:

$$p(Z|X;\vartheta) = \prod_l \frac{\exp(-(H_B^\vartheta + H_F^\vartheta)(Z,X))}{\sum_{Z \in \mathcal{Z}} \exp(-(H_B^\vartheta + H_F^\vartheta)(Z,X))} \ , \tag{10}$$

where $\{l\}$ is the set of lines on the grid. The energies H_B^ϑ and H_F^ϑ are defined as in (8) except that the neighbourhood system has changed (see figure 7). The partition function for the set of lines can be written as

$$\sum_{Z \in \mathcal{Z}} \exp(-(H_B^\vartheta + H_F^\vartheta)(Z,X)) = \prod_i \sum_{Z \in \mathcal{Z}_i} \exp(-H_{A(i)}^\vartheta(Z,X)) \tag{11}$$

where for every $i \neq j$ one has $\mathcal{Z}_i \cap \mathcal{Z}_j = \emptyset$. So $\mathcal{Z}$ is union of mutually disjoint sets $\mathcal{Z}_i$. Therefore it is now possible to compute the partition function because it only depends on line segments which are entirely in the foreground or background.

3.2 Learning the Parameters of the Random Field

Learning the model parameters by a maximum likelihood method is computationally expensive [21]. And as mentioned above, maximising the pseudeolikelihood (9) with respect to ϑ leads to an effective estimator for ϑ. For reasons which will be apparent later we consider the pseudolikelihood for a observation window $T \subset S$ which is entirely in the foreground or background. That implies that the conditioning on the hypothesis X can be ignored for this analysis. The energy function of $p(z_s|z_{\delta(s)};\vartheta)$ is in this case equal to the neighbourhood potential. The logarithm of the pseudolikelihood for an observation window $T \subset S$ has the form

$$PL_T(Z;\vartheta) = \sum_{s\in T}\left[g(z_s) + \vartheta V_s(z_s z_{\delta(s)}) - \ln\sum_{z_s}\exp(-\vartheta V_s(z_s z_{\delta(s)}))\right] \ , \qquad (12)$$

where V is defined as $V_s := \sum_{r\in\delta(s)}(z_s - z_r)^2$. The neighbourhood potential must satisfy a special spatial homogeneity condition. The potential is *shift* or *translational invariant* if for all $s,t,u \in S$

$$t \in \delta(s) \longleftrightarrow t+u \in \delta(s+u) \qquad \text{and} \qquad V_{C+u}(z_{s-u}) = V_C(z_s) \ . \qquad (13)$$

Furthermore a parameter ϑ is said to be *identifiable* if for every $\vartheta' \in \theta$ there is a configuration Z such that

$$p(Z;\vartheta) \neq p(Z;\vartheta') \ . \qquad (14)$$

The maximum pseudolikelihood estimator for the observation window T maximises $PL_T(Z,\cdot)$. If the potential is translational invariant and the parameter ϑ is identifiable Winkler [21] (Theorem 14.3.1 on page 240) proves that this estimator is asymptotically consistent when the size of the observation window increases. Winkler also proves that that the log of the pseudolikelihood PL_T is concave. In the present setting it is of course necessary to learn the parameters for the foreground and background energies H_F^ϑ and H_B^ϑ separately. Since the PL_T is concave it is possible to use a standard gradient decent algorithm to find the maximum of the log pseudolikelihood. In order to compute the gradient of the log pseudolikelihood it is desirable that the potential only depends on the parameters linearly. The gradient of the log pseudolikelihood can be written as

$$\nabla PL_T(Z;\vartheta) = \sum_{s\in T}\left[V(z_s z_{\delta(s)}) - E(V(Z_s z_{\delta(s)})|z_{\delta(s)};\vartheta)\right] \ , \qquad (15)$$

where $E(V(Z_s z_{\delta(s)}))$ denotes the conditional expectation with respect to the distribution $p(z_s|z_{\delta(s)};\vartheta)$ on Z_s. The graphs of the pseudolikelihood can be found in figure 8.

3.3 Computing the Partition Function

The main reason for adapting a one dimensional model was the problem of computing the partition function of the observation likelihood (10). Due to equation (11) it is possible to to compute the partition function by precomputing

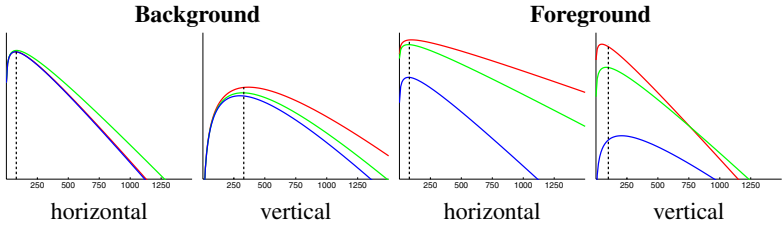

Fig. 8. Pseudolikelihood of training data. *The pseudolikelihood (12) is plotted for different values of ϑ. The distance between neighbouring sites d is set be $d = 4$ for horizontal and $d = 2$ for the vertical lines. Because we work on fields, d differs for horizontal and vertical lines. It should be noted that there is a difference between the models. The functions are concave, as expected.*

$$B_N := \sum_{Z \in \mathcal{Z}} -\exp(H_B^{\vartheta}(Z)) \quad \text{and} \quad F_N := \sum_{Z \in \mathcal{Z}} -\exp(H_B^{\vartheta}(Z)) \ , \tag{16}$$

where vector of measurements Z has length N. Rather than computing the value of the partition function for a particular hypothesis X it is desirable to compute a factor $\alpha(X)$ such that

$$\sum_{Z \in \mathcal{Z}} \exp(-(H_B^{\vartheta} + H_F^{\vartheta})(Z, X)) = \alpha(X) C \ , \tag{17}$$

where C is some constant. Now the problem of computing B_N and F_N needs to be addressed. The energy functions H_B^{ϑ} can be written as a quadratic form, i.e. $H_B^{\vartheta}(Z) = Z^t M Z$. The matrix M is of the form

$$\begin{pmatrix} (\lambda+\vartheta) & -\vartheta & 0 & \cdots & 0 \\ -\vartheta & (\lambda+2\vartheta) & -\vartheta & \cdots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & -\vartheta & (\lambda+2\vartheta) & -\vartheta \\ 0 & 0 & \cdots & -\vartheta & (\lambda+\vartheta) \end{pmatrix} \tag{18}$$

The matrix M is symmetric so it is possible to approximate B_N as

$$B_N = \sum_{Z \in \mathcal{Z}} \exp(-Z^t M Z) \approx \int_{\mathbf{R}^N} exp(-Z^t M Z)\, dZ = (2\pi)^{N/2} \det(M)^{-\frac{1}{2}} \ . \tag{19}$$

Since g_f and g_b are normal distributions this approximation holds for B_N as well as F_N.

3.4 Results

The observation likelihood $p(Z|X)$ as defined in (8) was tested on a set of single images. The results are summarised in figure 9. Whereas the results for horizontal and vertical

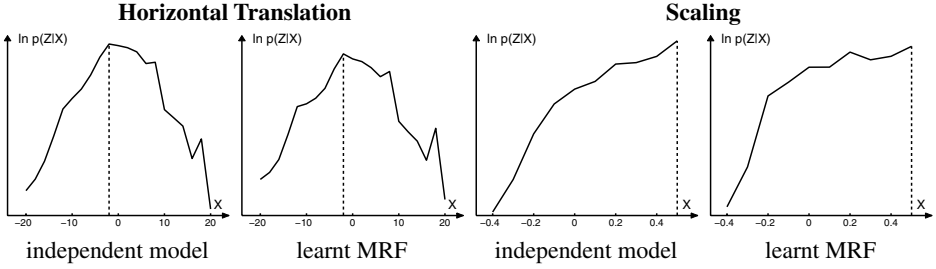

Fig. 9. Log-likelihood for horizontal translation and scaling. *The horizontal translation and scaling of the shape template is illustrated in figure 11. For both the horizontal translation and the scaling the log-likelihood for the independent model ($\vartheta = 0$) (left) and the MRF with the learnt parameter $vartheta$ (see figure 8). The parameters for the intensity distributions g_f and g_s are $\sigma_b^2 = 25, \mu_b = 102, \sigma_f^2 = 600, \mu_b = 128$. The results obtained for the scaling clearly need to be improved. See text for discussion.*

translation are good the results obtained for the scaling of the foreground window are poor. In order to test whether the MRF has any effect ϑ_F and ϑ_B are set to zero which is equivalent to assuming that two neighbouring measurements are independent. The graphs in figure 9 show that the modelling the statistical dependence of neighbouring measurement using the MRF does have an effect. As a first step to improve the model the neighbourhood structure was changed hoping that the interaction terms V_s (12) would have a greater effect. Now every pixel location on a scan lines is a site for the MRF. The resulting energy function is

$$H_A^\vartheta(Z, X) = \sum_{s \in A_X} g_A^s(z_s) + \sum_{\delta(s) \in A_X} \vartheta_A \cdot (z_s - z_{s+d})^2 \ . \tag{20}$$

Only the distance between neighbours depends on a predefined spacing d. The results of this improved method are shown in figures 10 and 11. The fact that the results obtained with the new observation likelihood (20) are better shows that the MRF is very sensitive to the chosen neighbourhood structure. This raises the question if there is any way to determine an optimal neighbourhood structure automatically. The hand-picked MRF we chose might not be the best after all.

A more ambitious step would be to construct a observation likelihood which makes use of the forward probabilities $p(z_t,\ z_{t-1},\ Y_t = \ f \mid \omega)$. This would complicate the computation of the partition function. But based on the encouraging results we obtained from the HMM (see figure 5) this could lead to a far more powerful model. It can be concluded that the MRF does the right thing but needs to be improved so it can be used in a tracker.

4 Conclusion

Both a new probabilistic background model as well as a observation likelihood for tracking cars are presented. Although the background model is particularly suited to the

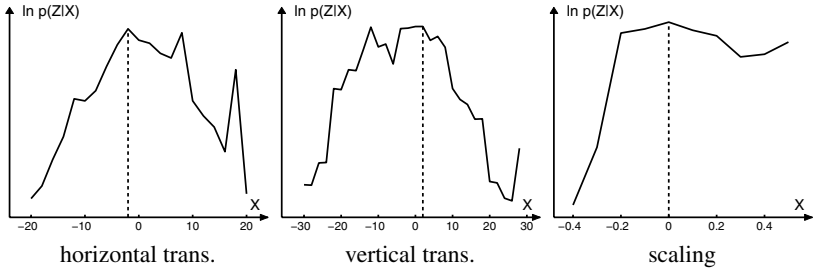

Fig. 10. Log-likelihoods for the improved model. *Similar to figure 9 the log-likelihoods are shown for horizontal and vertical translation as well as scaling using the improved model defined in (20). The model parameters itself are chosen as in figure 9. Although the maximum for the horizontal translation is not at zero figure 11 demonstrates that the most likely hypothesis leads to a correct localisation.*

traffic surveillance problem it can be used for a wide range of application domains. The results presented in figure 5 show that the use of this background model could lead to a robust tracker. The observation likelihood itself however still needs to be improved. The contribution this paper makes can be summarised as follows.

Probabilistic background model. Unlike many other background models the model presented here is capable of modelling shadow as well as foreground and background regions. Another considerable advantage of this model is that it is no longer necessary to select the training data. HMMs are a suitable model for this problem as they impose temporal continuity constraints. Although using two observation did improve the results significantly the choice of filters is not optimal. The results presented in figure 6 support the claim that it is crucial to model the transition probabilities correctly.

Car tracker. In order to build a robust car tracker it is necessary to model the inside of the vehicles as well as the background and the statistical dependence of neighbouring pixels. This is possible by modelling an observation density used in a particle filter which is based on an MRF. However it has to be noted that the MRF is very sensitive to the choice of the neighbourhood system. It remains an open problem which neighbourhood system is optimal. The formulation of the MRF based on scan-lines leads to a model which is computationally tractable. It should be noted that the presented observation likelihood is consistent with a Bayesian framework since the measurements do not depend on the hypothesised position of the vehicle. The use of importance sampling makes it possible to feed the information of the low level process into the car tracker in a consistent fashion.

Future work. Since the illumination changes throughout the day it is necessary to derive a criterion when the the parameters of the background model need to be updated. It is furthermore necessary to investigate how the observation density can be improved.

Acknowledgements. We are grateful for the support of the EPSRC and the Royal Society (AB) and the EU (JR).

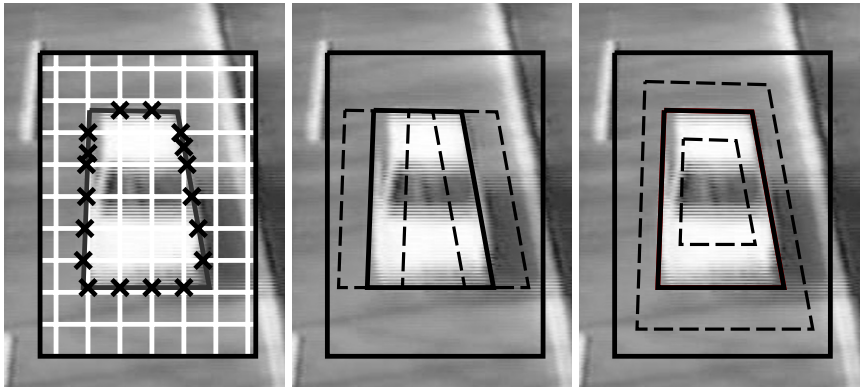

Fig. 11. Observation window and scan lines of the car tracker. *The right image illustrates the grid used by the algorithm. The observation window is marked in black. The measurements are taken on scan-lines (in white). The hypothesised position of the car is shown in dark grey. The other two images illustrate how well the improved model localises. The most likely hypothesis is shown as a solid black line. The dashed lines illustrate the minimal and maximal configurations of the variation. See figures 10 and 9 for the corresponding log-likelihood functions.*

References

[1] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc. B*, 36:192–236, 1974.

[2] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real-time computer vision system for measuring traffic pramameters. In *IEEE Conf. Computer Vision and Pattern Recognition, June 1997, Puerto Rico*, 1997.

[3] F. Divino, A. Frigessi, and P. J. Green. Penalized pseudolikelihood inference in spatial interaction models with covariates. *Scandinavian Journal of Statistics (to appear)*, 1998.

[4] N. Ferrier, S. Rowe, and A. Blake. Real-time traffic monitoring. In *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, pages 81–88, 1994.

[5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. PAMI*, 6:721–741, 1984.

[6] I. Haritaoglu, D. Harwood, and L. S. Davis. W4 - a real time system for detection and tracking people and their parts. In *Proc. Conf. Face and Gesture Recognition, Nara, Japan*, 1998.

[7] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *Proc. European Conf. on Computer Vision, Cambridge, UK*, 1996.

[8] M.A. Isard and A. Blake. ICondensation: Unifying low-level and high-level tracking in a stochastic framework. In *Proc. 5th European Conf. Computer Vision*, pages 893–908, 1998.

[9] B.-H. Juang and L.R. Rabiner. Mixture autoregressive hidden markov models for speech signals. *IEEE Trans. Acoustics, Speech, and Signal Processing*, December:1404–1413, 1985.

[10] D. Koller, J. Weber, and J. Malik. Robust multiple car tracking with occlusion reasoning. In *Proc. of ECCV 94, Stockholm, Sweden*, pages 189–196, 1994.

[11] J. MacCormick and A. Blake. A probablilistic exclusion principle for tracking multiple objects. In *Proc. 7th Int. Conf. on Computer Vision*, volume 2, pages 572 – 578, 1999.
[12] S.J. Maybank, A.D. Worrall, and G.D. Sullivan. Filter for car tracking based on acceleration and steering angle. In *Proc. 7th BMVC*, 1996.
[13] N. Paragios and R. Deriche. A PDE-based Level Set Approach for detection and tracking of moving objects. Technical Report 3173, INRIA Sophia Antipolis, 1997.
[14] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286, February 1989.
[15] P. Remagnino, A. Baumberg, T. Grove, D. Hogg, T. Tan, A. Worrall, and K. Baker. An integratedtraffic and pedestrian model-based vision system. In *Proc. of BMVC '97, Univ. of Essex*, volume 2, pages 380–398, 1997.
[16] J. Rittscher and A. Blake. Classification of human body motion. In *Proc. 7th Int. Conf. on Computer Vision*, pages 634–639, 1999.
[17] S.M. Rowe and A. Blake. Statistical background modelling for tracking with a virtual camera. In *Proc. British Machine Vision Conf.*, volume 2, pages 423–432, 1995.
[18] G.D Sullivan, K.D. Baker, and A.D. Worrall. Model-based vehicle detection and classification using orthographic approximations. In *Proc. of 7th BMVC*, pages 695–704, 1996.
[19] J. Sullivan, A. Blake, M. Isard, and J. MacCormick. Object localisation by Baysian correlation. In *Proc. 7th Int. Conf. on Computer Vision*, volume 2, pages 1068–1075, 1999.
[20] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. 7th Int. Conf. on Computer Vision*, pages 255–261, 1999.
[21] G. Winkler. *Image analysis, random fields and dynamics Monte Carlo methods: a mathematical introduction*. Spinger, 1995.