International Journal of Information Technology and Knowledge Management
July-December 2009, Volume 2, No. 2, pp. 231-236

MODELING AND SUBTRACTION OF SCENES EXHIBITING CONSTANT DYNAMIC CONDUCT IN VIDEO SEQUENCE

Vivek Sharma[*], D. K. Swami[**], R. C. Jain[***]

Identifying moving objects is a critical task for many computer vision applications; it provides a categorization of the pixels into either foreground or background. A common approach used to achieve such categorization is removing background. There are many background removal algorithms in the literature, most of them pass through four major steps, which are pre-processing, background modeling, foreground detection and data validation. An important constituent of vision systems is background modeling. Existing work in background modeling has mostly addressed scenes that consist of stationary structures. But if the scene exhibits constant dynamic behavior in time, such an hypothesis is dishonored and detection performance deteriorates. We propose a method for the modeling and subtraction of scenes exhibiting constant dynamic behavior in time.

1. Introduction

Existing methods for background modeling may be classified as either predictive or non-predictive. Predictive methods model the scene as a time series and develop a dynamical model to recover the current input based on past observations. Predictive mechanisms of varying complexity have been considered in the literature. Several authors [16, 17] have used a Kalman-filter based approach for modeling the dynamics of the state at a particular pixel. Recent methods are based on more complicated models. In [7], an autoregressive model was proposed to capture the properties of dynamic scenes. In [12], this idea is extended by using multiple Gaussians to model the scene and develop a fast approximate method for updating the parameters of the model incrementally. Such an approach is capable of dealing with multiple hypothesis for the background and can be useful in scenes such as waving trees, beaches, escalators, rain or snow. The mixture-of-Gaussians method is quite popular and was to be the basis for a large number of related techniques [15, 13]. In [10], a statistical characterization of the error associated with this algorithm is studied. When the density function is more complex and cannot be modeled parametrically, a non-parametric approach able to handle arbitrary densities is more suitable. Such an approach was used in [8] where the use of Gaussian kernels for modeling the density at a particular pixel was proposed. Existing methods can effectively describe scenes that have a smooth behavior and limited variation. Consequently, they are able to cope with gradually evolving scenes. However, one can claim that their performance deteriorates when the scene to be described is dynamic and exhibits non-stationary properties in time. Examples of such scenes include ocean waves, waving trees, rain, moving clouds, etc. Most of the dynamic scenes exhibit persistent motion characteristics. Therefore, a natural approach to model their behavior is via optical flow. Combining such flow information with standard intensity information, we present a method for background-foreground differentiation that is able to detect objects that differ from the background in either motion or intensity properties.

[*] VNS Institute of Technology Bhopal, R.G.T.U. Bhopal (M.P.), INDIA. *E-mail: sharma.vivek95@yahoo.in*

[**] VNS Institute of Technology Bhopal, R.G.T.U. Bhopal (M.P.), INDIA. *E-mail: dksvns@ymail.com*

[***] S.A.T.I., Vidisha, R.G.T.U. Bhopal (Madhya Pradesh) INDIA *E-mail: jcr_dr@yahoo.com*

2. Estimation of Background Density

We assume that flow measurements [18, 14] and their uncertainties are available. Then, we propose a theoretical agenda to obtain an estimate of the probability distribution of the observed data in a higher dimensional space. Several methods parametric and nonparametric-can be considered for determining this probability distribution. A mixture of multivariate Gaussians can be considered to approximate this distribution. The parameters of the model, *i.e.* the mean and the covariance matrix of the Gaussians, can be estimated and updated in a manner similar to [12]. Care has to be exercised, however, in dealing with the uncertainties in the correct manner. A more suitable approach refers to a non-parametric method. One can claim that such method has the characteristic of being able to deal with the uncertainties in an accurate manner. On the other hand, such a method is computationally expensive. Let x_1, x_2, ..., x_n be a set of d-dimensional points in R_d and H be a symmetric positive definite $d \times d$ matrix (called the bandwidth matrix). Let K: $\mathbb{R}^d \rightarrow \mathbb{R}^1$ be a kernel satisfying certain conditions that will be defined later. Then the multivariate fixed bandwidth kernel estimator is defined as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(x - x_i)) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\|H\|^{1/2}} K(H^{-1/2}(x - x_i)) \quad (1)$$

where

$$K_H(x) = \|H\|^{-1/2} K\left(H^{-1/2}x\right).$$

The matrix H is the smoothness parameter and specifies the "width" of the kernel around each sample point x_i. A well-behaved kernel K must satisfy the following conditions:

$$\int_{\mathbb{R}^d} K(w)dw = 1, \quad \int_{\mathbb{R}^d} wK(w)dw = 0, \quad \int_{\mathbb{R}^d} ww^T K(w)dw = I_d \quad (2)$$

The first condition accounts for the fact that the sum of the kernel function over the whole region is unity. The second equation imposes the constraint that the means of the marginal kernels $\{K_{i(wi)}, i = 1, ..., d\}$ are all zero. Last but not the least, the third term states that the marginal kernels are all pair wise uncorrelated and that each has unit variance. The simplest approach would be to use a fixed bandwidth matrix H for all the samples. Although such an approach is a reasonable compromise between complexity and the quality of approximation, the use of variable bandwidth can usually lead to an improvement in the accuracy of the estimated density. Smaller bandwidth is more appropriate in regions of high density since a larger number of samples enables a more accurate estimation of the density in these regions. On the other hand, a larger bandwidth is more appropriate in low density areas where few sample points are available. It is possible to consider a bandwidth function that adapts to the point of estimation, as well as to the observed data points and the shape of the underlying density[24]. In the literature, two simplified versions have been studied. The first varies the bandwidth at each estimation point and is referred to as the balloon estimator. The second varies the bandwidth for each data point and is referred to as the sample-point estimator. Thus, for the balloon estimator,

$$\hat{f}_B(x) = \frac{1}{n}\sum_{i=1}^{n} K_{H(x)}(x - x_i)) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\|H(x)\|^{1/2}} K(H(x)^{-1/2}(x - x_i))$$

Where $H_{(x)}$ is the smoothing matrix for the estimation point x. For each point at which the density is to be estimated, kernels of the same size and orientation are centered at each data point. The density estimate is computed by taking the average of the heights of the kernels at the estimation point. A popular choice for the bandwidth function in this case is to restrict the kernel to be spherically symmetric that further simplifies the approximation. Then, only one independent smoothing parameter remains $h_k(x)$ which is typically estimated as the distance from x to the k^{th} nearest data point. Such an estimator suffers from several disadvantages - discontinuities, bias problems and integration to infinity. An alternate strategy is to have the bandwidth matrix be a function of the sample points. Such estimator is called the sample-point estimator [24]:

$$\hat{f}_S(x) = \frac{1}{n}\sum_{i=1}^{n} K_{H(x_i)}(x - x_i)) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\|H(x_i)\|^{1/2}} K(H(x_i)^{-1/2}(x - x_i))$$

The sample-point estimator still places a kernel at each data point. These kernels each have their own size and orientation regardless of where the density is to be estimated. This type of estimator was introduced by [4] who suggest using

$$H(x_i) = h(x_i)I$$

where $h(x_i)$ is the distance from xi to the k^{th} nearest data point. Asymptotically, this is equivalent to choosing $h(x_i) \propto f(x_i)^{-1/d}$ where d is the dimension of the data. A popular choice for the bandwidth function, suggested by [1], is to use $h(x_i) \propto f(x_i)^{-1/2}$ and, in practice, to use a pilot estimate of the density to calibrate the bandwidth function. In this paper, we introduce a hybrid density estimator where the bandwidth is a function not only of the sample point but also of the estimation point x. The particular property of the data that will be addressed is the existence of the uncertainty estimates of not only the sample points, but also the estimation point x.

Let $\{x_i\}^n{}_{i=1}$ be a set of measurements in d-dimensional space such that each x_i has associated with it a mean μ_i (in $\mathbb{R}^d$) and a $d \times d$ covariance matrix Σ_i. Also, let x (with mean μx and covariance Σ_x) be the current measurement whose probability is to be estimated. We define the multivariate hybrid density estimator as:

$$\hat{f}_H(x) = \frac{1}{n}\sum_{i=1}^{n} K_{H(x,x_i)}(\mu - \mu_i)) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\|H(x,x_i)\|^{1/2}} K(H(x,x_i)^{-1/2}(\mu - \mu_i)) \quad (2)$$

where the bandwidth matrix $H(x, xi)$ is a function of both the estimation measurement x and the sample measurement x_i. Chen *et. al.* [5] have suggested using $H_i = x^2_{\gamma,p}\Sigma_i$ for Epanechnikov kernels in the absence of error measurements. Expanding this idea, we propose the use of $H(x, x_i) = \Sigma x_i + \Sigma x$ as a possible bandwidth matrix for the Normal kernel. Thus, the density estimator becomes.

$$\hat{f}_H(x) = \frac{1}{n(2\pi)^{d/2}} \sum_{i=1}^{n} \frac{1}{\left\|\Sigma_{x_i} + \Sigma_x\right\|^{1/2}} \exp\left(-\frac{1}{2}(\mu - \mu_i)^T (\Sigma_{x_i} + \Sigma_x)^{-1}(\mu - \mu_i)\right) \quad (3)$$

This particular choice for the bandwidth function has a simple but meaningful mathematical foundation. Suppose $x1$ and $x2$ are two normally distributed random variables with means $\{\mu_i\}$ and covariance matrices $\{\Sigma_i\}$, *i.e.* $x_i \sim N(\mu_i, \Sigma_i)$, $i = 1, 2$. It is well-known that if x_1 and x_2 are independent, the distribution of $(x_1 - x_2)$ is $N(\mu_1 - \mu_2, \Sigma_1 + \Sigma_2)$. Thus, the probability that $x_1 = x_2$ or $x_1 - x_2 = O$ is

$$p(x_1 = x_2) = \frac{1}{(2\pi)^{d/2}\left\|\Sigma_1 + \Sigma_2\right\|^{1/2}} \exp\left(-\frac{1}{2}(\mu_1 - \mu_2)^T (\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2)\right)$$

Thus, Equation 3 can be thought of as the average of the probabilities that the estimation measurement is equal to the sample measurement, calculated over all the sample measurements. The choice for the bandwidth matrix can also be justified by the fact that the directions in which there is more uncertainty are given proportionately less weightage. Such uncertainty can be either in the estimation measurement or the sample measurements. Experimentally, the results obtained using these criteria were satisfactory when compared with the fixed bandwidth estimator or the balloon/sample-point estimators.

3. CATEGORIZATION

Once an appropriate mechanism for density approximation is built, the next step is to determine a categorization mechanism for the observed data. Categorization may be performed by thresholding on the probability of a new observation to belong to the background. However, two observations need to be taken into account:

- The threshold should be adaptive and determined based on the uncertainty or spread of the background distribution at a particular pixel (called entropy in in-formation theory).
- Any available prior information about the foreground distribution should be utilized.

Figure 1: Adaptive Thresholds for (a) The Ocean Sequence, (b) Traffic Sequence. Notice that the Thresholds are Higher in Regions of Low Variability and Low in Regions of High Variability.

More formally, guarantying a false-alarm rate of less than α_f requires that the threshold T should be set such that:

$$\int_{\hat{f}(x)<T} \hat{f}(x)dx < \alpha_f \quad (4)$$

Furthermore, if $f_o(x)$ is the foreground distribution, guaranteeing a miss probability of αmleads to the following condition on T:

$$\int_{\hat{f}_o(x)<T} \hat{f}_o(x)dx < \alpha_m \quad (5)$$

Meeting both constraints simultaneously could be impossible, therefore a compromise is generally required. Furthermore, the foreground distribution is generally unknown weakening the use of the second constraint. Determination of the threshold according to Equation (4) involves the inversion of complex integrals of clipped distributions. Such solution is feasible only for simple distributions like the Gaussian [10]. In the presence of more complex underlying densities, a statistical approximation is more suitable. We propose the use of sampling to get an

estimate of the false alarm rate for a given threshold. Samples are drawn from the learnt background distribution (estimated via kernels in the present work) and the density at these sample points is classified using the current threshold as background or foreground. These categorizations provide an estimate of the false alarm rate for the current thresh-old value. Such information can then be utilized for adjusting the threshold according to the desired false alarm rate. Since the "spread" of the distribution at a particular pixel is not expected to vary significantly over time, such thresh-old can be adjusted incrementally. Incremental adaptation of the threshold reduces the false alarms in the regions of high variation (e.g. waves, trees) while maintaining high detection rates in stationary areas.

4. UNCERTAINTIES AND FEATURE MEASUREMENT

Once the appropriate generic model for background subtraction is introduced, addressing the selection/estimation of the features is to be considered. As mentioned earlier, we utilize five features-two for optical flow and three for the intensity in the normalized color space. We have assumed that the uncertainties in their measurements are available. Here, we briefly describe methods that might be used for obtaining such measurements and their associated uncertainties.

4.1 Optical Flow

Several optical flow algorithms and their extensions [22, 18, 14, 6, 2] can be considered. The basic idea of this algorithm is to apply the optical flow constraint equation

$$\nabla^T g.f + gt = 0$$

where ∇g and gt are the spatial image gradient and temporal derivative, respectively, of the image at a given spatial location and time, and f is the two-dimensional velocity vector. Such equation puts only one constraint on the two parameters (aperture problem). Thus, a smoothness constraint on the field of velocity vectors is a common selection to address this limitation. If we assume locally constant velocity and combine linear constraints over local spatial regions, a sum-of-squares error function can be defined:

$$E(f) = \sum_i w_i \left[\nabla^T g(x_i,t) f + g_t(x_i,t)\right]$$

Minimizing this error function with respect to f yields:

$$f = -M^{-1}b$$

where

$$M = \sum \nabla g \nabla^T g = \begin{bmatrix} \sum g_x^2 & \sum g_x g_y \\ \sum g_x g_y & \sum g_y^2 \end{bmatrix},$$

$$b = \begin{bmatrix} \sum g_x g_t \\ \sum g_x g_t \end{bmatrix} \tag{6}$$

and all the summations are over a patch around the point. Define f as the optical flow, as the actual velocity field, and $n1$ as the random variable describing the difference between the two. Then:

$$\hat{f} = f + n_1$$

Similarly, let $\hat{g}_t$ be the actual temporal derivative, and g_t the measured derivative. Then:

$$g_t = \hat{g}_t + n_2$$

where n_2 is a random variable characterizing the uncertainty in this measurement relative to the true derivative. The uncertainty in the spatial derivatives is assumed to be much smaller than the uncertainty in the temporal derivatives. Under the assumption that $n1$ and $n2$ are governed by a normal distribution with covariance matrices $\Lambda_1 = \lambda_1 \boldsymbol{I}$ and $\Lambda 2 = \lambda_2$ (it is scalar), and the flow vector f has a zero-mean Normal prior distribution with covariance Λ_p, the covariance and mean of the optical flow vector may be estimated:

$$\Lambda_f = \left[\sum_i \frac{w_i \mathbf{M}_i}{\left(\lambda_1 \|\nabla g(x_i)\|^2 + \lambda_2\right)} + \Lambda_p^{-1}\right]^{-1}$$
$$\mu_f = -\Lambda_f \sum_i \frac{w_i \mathbf{b}_i}{(\lambda_1 \|\nabla g(x_i)\|^2 + \lambda_2)} \tag{7}$$

where w_i is a weighting function over the patch, with the points in the patch indexed by i, and Mi, and b_i are the same as matrices defined in Equation 6 but without the summation and evaluated at location xi. In order to handle significant displacements, a multi-scale approach is considered that uses the flow estimates from a higher scale to initialize the flow for a lower level. Towards the propagation of variance across scales, a kalman filter is used with the normally used time variable replaced by scale.

4.2 Normalized Color Representation

Suppose R, G and B are the RGB values observed at a pixel. Then, the normalized features are defined as:

$$r = R/S,\ g = G/S,\ I = S/3 \tag{8}$$

where $S = R + G + B$. The advantage of such transformation is that, under certain conditions, it is invariant to a change in illumination. However, such transformation introduces heteroscedastic (point-dependent) noise in the data that needs to be modeled correctly. Assuming that the sensor noise (in RGB space) is normally distributed with a diagonal covariance matrix having diagonal terms σ, it is not too

difficult to show [11] that the uncertainties in the (a) (b) (c) (d)

Normalized features is:

$$\sum_{r,g} = \frac{\sigma^2}{S^2} \begin{pmatrix} \left(1 - \frac{2R}{S} + \frac{3R^2}{S^2}\right) & \left(-\frac{R+G}{S} + \frac{3RG}{S^2}\right) \\ \left(-\frac{R+G}{S} + \frac{3RG}{S^2}\right) & \left(1 - \frac{2G}{S} + \frac{3G^2}{S^2}\right) \end{pmatrix}$$

4.3 Combining the Features

The covariance σ_i for an observation x_i (in 5D space) may be estimated from the covariances of the components the normalized color and optical flow. Assuming that the intensity and optical flow features are uncorrelated (which may not be true in general), an expression for the covariance matrix may be derived:

$$\sum_i = \begin{bmatrix} \Sigma_{r,g} & 0 & 0 \\ 0 & \sigma_i & 0 \\ 0 & 0 & \Lambda_f \end{bmatrix} \quad (9)$$

where Λ_f is obtained as in Equation 7 and boldface *O*'s represent appropriate zero matrices.

5. RESULTS AND CONCLUSION

In order to validate the proposed technique, two different types of scenes were considered. The first is the challenging scene of the ocean front. Such scene involves wave motion, blowing grass, long-term changes due to tides, global illumination changes, shadows etc. Our algorithm was able to detect events of interest in the land and simulated events on the ocean front with extremely low false alarm rate as shown in Figure 3. The algorithm was able to detect simulated objects having almost no visual difference from the ocean if they were moving in a pattern that was different from the ocean [Figure 3(i)-(l)]. A typical traffic surveillance scenario was considered next where the challenge was due to the vigorous motion of the trees and bushes Figure 2].

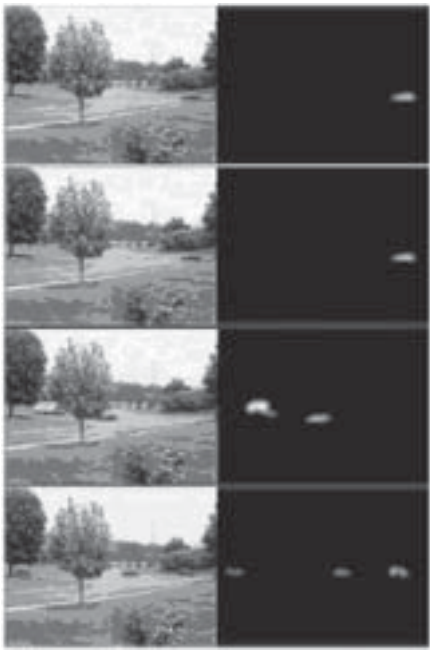

Figure 2: Detection for a Traffic Sequence Consisting of Waving Trees.

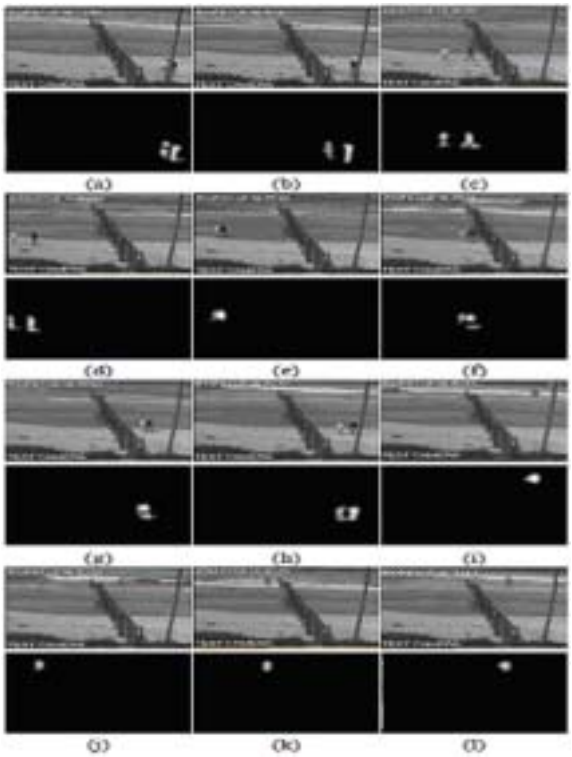

Figure 3: Results for the Ocean Sequence using the Proposed Algorithm. Figures (i) - (l),

REFERENCES

[1] I. Abramson. On Bandwidth Variation in Kernel Estimates-A Square Root Law. *The Annals of Statistics*, 10:1217–1223 (1982).

[2] P. Anandan. A Computational Agenda and An Algorithm for the Measurement of Visual Motion. *IJCV*, **2**(3) (1989) 283–310.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of Optical Flow Techniques. *IJCV*, **12**(1) (1994) 43–77.

[4] L. Breiman, W. Meisel, and E. Purcell. Variable Kernel Estimates of Multivariate Densities. *Technometrics*, **19** (1977) 135–144.

[5] H. Chen and P. Meer. Robust Computer Vision Through Kernel Density Estimation. In *ECCV*, pages I: 236–250, Copenhagen, Denmark, (2002).

[6] D. Comaniciu. Nonparametric Information Fusion for Motion Estimation. In *CVPR*, Madison, Wisconsin, (2003).

[7] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic Textures. *IJCV*, **51**(2) (2003) 91–109.

[8] A. Elgammal, D. Harwood, and L. Davis. Non-Parametric Model for Background Subtraction. In *ECCV*, pages II : 751–767, Dublin, Ireland (2000).

[9] N. Friedman and S. Russell. *Image Segmentation in Video Sequences:* A Probabilistic Approach. In Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI), (1997).

[10] X. Gao, T. E. Boult, F. Coetzee, and V. Ramesh. Error Analysis of Background Adaption. In *CVPR*, pages I : 503–510, Hilton Head Island, SC, (2000).

[11] M. Greiffenhagen, V. Ramesh, D. Comaniciu, and H. Niemann. Statistical Modeling and Performance Characterization of a Real-Time Dual Camera Surveillance System. In *CVPR*, pages II : 335–342, Hilton Head, SC (2000).

[12] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using Adaptive Tracking to Classify and Monitor Activities in a Site. In *CVPR*, Santa Barbara, CA, (1998).

[13] M. Harville. A Agenda for High-Level Feedback to Adaptive, Per-Pixel, Mixture-of-Gaussian Background Models. In *ECCV*, page III : 543 ff., Copenhagen, Denmark, (2002).

[14] B. K. P. Horn and B. G. Schunck. Determining Optical Flow. *Artificial Intelligence*, pages 17 : 185–203, (1981).

[15] O. Javed, K. Shafique, and M. Shah. A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information. In *MVC*, pages 22–27, Florida, (2002).

[16] Klaus-Peter Karmann and Achim Von Brandt. *V Cappellini (ed.), Time Varying Image Processing and Moving Object Recognition*, **2**, Chapter Moving Object Recognition using an Adaptive Background Memory. Elsevier, Amsterdam, The Netherlands, (1990).

[17] Dieter Koller, Joseph Weber, and Jitendra Malik. Robust Multiple Car Tracking with Occlusion Reasoning. In *ECCV*, pages 189–196, Stockholm, Sweden, (1994).