Background Subtraction in Highly Dynamic Scenes

Vijay Mahadevan Nuno Vasconcelos
Department of Electrical and Computer Engineering
University of California, San Diego
vmahadev@ucsd.edu, nuno@ece.ucsd.edu

Abstract

A new algorithm is proposed for background subtraction in highly dynamic scenes. Background subtraction is equated to the dual problem of saliency detection: background points are those considered not salient by suitable comparison of object and background appearance and dynamics. Drawing inspiration from biological vision, saliency is defined locally, using center-surround computations that measure local feature contrast. A discriminant formulation is adopted, where the saliency of a location is the discriminant power of a set of features with respect to the binary classification problem which opposes center to surround. To account for both motion and appearance, and achieve robustness to highly dynamic backgrounds, these features are spatiotemporal patches, which are modeled as dynamic textures. The resulting background subtraction algorithm is fully unsupervised, requires no training stage to learn background parameters, and depends only on the relative disparity of motion between the center and surround regions. This makes it insensitive to camera motion. The algorithm is tested on challenging video sequences, and shown to outperform various state-of-the-art techniques for background subtraction.

1. Introduction

Natural scenes are usually composed of several dynamic entities. Objects of interest often move amidst complicated backgrounds that are themselves moving, e.g. swaying trees, moving water, waves and rain. Successful discrimination between the moving objects and the background motion presents a survival advantage, for example in terms of being able to identify potential predators or prey. Not surprisingly, biological visual systems have evolved to be extremely efficient in this task. In computer vision, background subtraction is useful for diverse applications. Algorithms that can produce reliable "figure-ground" segmentation are used as a pre-processing step for object and event detection, activity and gesture recognition, tracking, surveillance and video retrieval. As an example, in robotic path planning, an autonomous device could benefit from a background subtraction module to simplify the task of identifying objects that approach it.

Unlike biological vision, background subtraction has proven quite challenging for computer vision. After decades of research on this problem (see [20] for a review), there has been little progress in the development of methods that are robust and generic enough to handle the complexities of most natural dynamic scenes. For example, many of the state-of-the-art techniques [8, 16, 22] assume a static camera, and are unsuitable for video shot with hand-held cameras or from moving platforms (as in the robot example). The conventional approach to background subtraction in the presence of ego-motion is to first explicitly [17], or approximately [19], compensate for the camera motion, and then rely on stationary camera background subtraction techniques. Accurate compensation of ego-motion is, however, cumbersome and can be quite difficult when the background is itself dynamic.

Several popular methods also model the background explicitly, assuming a bootstrapping phase where the algorithm is presented with frames containing only the background [16, 22, 25]. We refer to these techniques as *implicitly supervised*, and to the initial phase as a *training* step for learning background parameters. This training must be repeated for each scene where the algorithms are deployed, but training information may not always available, and the background parameters may need to be continuously updated if the scene is dynamic. This is, once again, cumbersome and can sometimes be technically challenging. A further shortcoming is the use of several (often unjustified) assumptions on the motion characteristics of the foreground object. For instance, it is often assumed that the foreground moves in a consistent direction (temporal persistence) [2, 15, 24], with faster appearance changes than the background [20]. Such assumptions are not always valid, and are particularly questionable when there is egomotion (e.g. a camera that tracks a moving object).

To address these limitations, we propose a novel

paradigm for background subtraction. This paradigm is inspired by biological vision, where background subtraction is inherent to the task of deploying visual attention. This can be done in multiple ways but frequently relies on motion saliency mechanisms, which identify regions of the visual field where objects move differently from the background. We equate background subtraction to the problem of detecting salient motion, and propose a solution based on a generic hypothesis for biological salience, which is referred to as the *discriminant center-surround hypothesis*. Under this hypothesis, bottom-up saliency is formulated as the result of optimal discrimination between center and surround stimuli at each location of the visual field. Locations where the discrimination between the two can be performed with smallest expected probability of error are declared as most salient. Background subtraction is then equivalent to simply ignoring the locations declared as non-salient.

This *strictly local* approach to background subtraction has various advantages over the traditional *global* procedures. First, there is no need to train or maintain a *global* model of the background. As the latter changes, so do the surround windows at all locations of the visual field. Thus, the local saliency measures are automatically adapted to variations in the background, and there is no need to keep track of, or update, a global model. Second, background modeling is considerably simplified. While, globally, a dynamic background is rarely homogeneous (e.g. different trees have different motion), the assumption of spatial homogeneity is usually accurate locally. This enables the use of much simpler probabilistic models (e.g. unimodal distributions vs. mixtures) which are easier to learn and update. Third, because discriminant saliency compares the center and surround regions, it depends only on the *relative disparity* between their motion characteristics, and therefore is invariant to camera motion. Finally, discriminant saliency can be adapted to various problems by simply modifying the features and probabilistic models used to discriminate between center and surround. For example, motion features can be complemented with depth measurements, if range sensors are available, and different types of models can be chosen to account for different background dynamics. In this work, we choose dynamic texture [7] models, due to their versatility in modeling complex moving patterns, ability to replicate the motion of natural scenes, and the rich statistical formulations they lend themselves to.

Overall, the main contributions of this work are threefold. First, the proposed algorithm is completely unsupervised and does not require initial training with 'background-only' frames. In effect, it is a *bottom-up* approach that can adapt to any situation. Second, due to its *locally discriminant* nature, the algorithm is insensitive to egomotion, and applicable to video shot with moving cameras. Third, by relying on dynamic textures as models for the video, it accounts for joint saliency in motion and appearance in a principled manner, and is robust enough to handle backgrounds of complex dynamics. Experimental results on sequences with such dynamics show that the proposed algorithm outperforms the current state-of-the-art in background subtraction.

The paper is organized as follows. The discriminant saliency architecture is presented in Section 2. Dynamic texture models and their use in motion saliency are discussed in Section 3. Experimental evaluation and results form Section 4.

2. Discriminant Center-Surround Saliency

We use local measurements of motion contrast as the central source of information for the motion saliency detector now proposed. To produce a quantitative measure of saliency we rely on the principle of *discriminant saliency* [9, 10]. This is a generic saliency principle, applicable to a broad set of problems. For example, different specifications of its components have been used to define top-down [9] and bottom-up saliency for static images [10]. Here we consider bottom-up motion saliency, using a center-surround architecture and motion models which are suitable for dynamic scenes.

2.1. Mathematical Formulation

Discriminant saliency is defined with respect to two classes of stimuli: the class of *stimuli of interest*, and the *background* or null hypothesis, consisting of stimuli that are not salient. The locations of the visual field that can be classified, with lowest expected probability of error, as containing stimuli of interest are denoted as salient. This is accomplished by setting up a binary classification problem which opposes the stimuli of interest to the null hypothesis. The saliency of each location in the visual field is then equated to the discriminant power (expected classification error) of the visual features extracted from that location to differentiate the two classes.

Formally, let $\mathcal{V}$ be a d dimensional dataset ($d = 2$ for static images, $d = 3$ for video) indexed by location vector $l \in L \subset \mathbb{R}^D$ and consider the responses to visual stimuli of a predefined set of features $\mathbf{Y}$ (e.g. raw pixel values, Gabor or Fourier features), computed from $\mathcal{V}$ at all locations $l \in L$. A classification problem opposing two classes, of class label $C(l) \in \{0, 1\}$, is posed at location l. Two windows are defined: a neighborhood $\mathcal{W}_l^1$ of l which is denoted as *center*, and a surrounding annular window $\mathcal{W}_l^0$ which is denoted as the *surround*. The union of the two windows is denoted the *total* window, $\mathcal{W}_l = \mathcal{W}_l^0 \cup \mathcal{W}_l^1$.

Let $\mathbf{y}$ be the vector of feature responses at location $j \in L$. Features in the center are drawn from the class of interest (or alternate hypothesis) $C(l) = 1$, with prob-

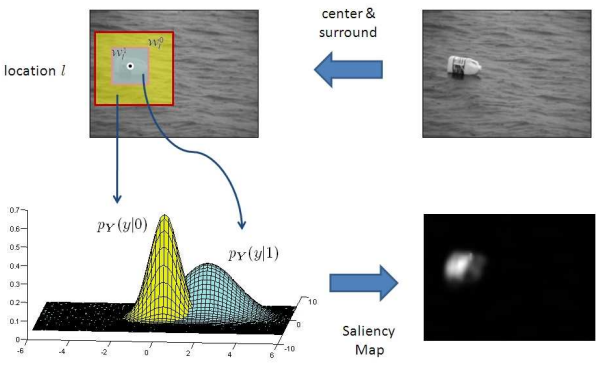

Figure 1. Illustration of discriminant center-surround saliency.

ability density $p(\mathbf{y}|1)$. Features in the surround are drawn from the null hypothesis $C(l) = 0$, with probability density $p(\mathbf{y}|0)$. An illustration of the classification problem involving center and surround for static images is shown in Figure 1. The saliency of location l, $S(l)$, is the extent to which the features $\mathbf{Y}$ can discriminate between *center* and *surround*. This is quantified by the mutual information between features, $\mathbf{Y}$, and class label, C,

$$S(l) = I_l(\mathbf{Y}; C) = \sum_c \int p(\mathbf{y}, c) \log \frac{p(\mathbf{y}, c)}{p(\mathbf{y})p(c)} d\mathbf{y}. \quad (1)$$

which can also be written as

$$S(l) \quad = \quad \sum_c p(c) \mathrm{KL}\left(p(\mathbf{y}|c) \,\|p(\mathbf{y})\right) \quad (2)$$

where $\mathrm{KL}\left(p \,\|q\right)$ represents the Kullback-Leibler divergence between two densities p and q. This mutual information is an approximation to the probability of correct classification (one minus the Bayes error rate) of the classification problem [23]. Hence, a large $S(l)$ implies that center and surround have a large disparity of feature responses, i.e. large *local feature contrast*.

2.2. Modeling spatio-temporal stimulus statistics

The discriminant saliency measure in (1) is defined in a generic sense, and the does not depend on the type of stimulus or feature set $\mathbf{Y}$. In [11] it was shown that for static saliency, under the common assumption of generalized Gaussian feature statistics [12], discriminant saliency can be mapped into a biologically plausible neural architecture which replicates the computations of the standard model of V1 [3]. In this work, we consider the problem of motion saliency, showing that by using suitable models of spatio-temporal stimulus statistics, the formulation can compute saliency in highly dynamic scenes.

In particular, we adopt the dynamic texture (DT) model of [7], due to its ability to account for spatial and temporal characteristics of the visual stimulus in an elegant unified stochastic framework. A DT is an autoregressive generative model that represents the appearance of the stimulus $\mathbf{y}_t \in \mathbb{R}^m$ (the two-dimensional image stimulus is first converted into a column vector of length m), observed at time t, as a linear function of a hidden state process $\mathbf{x}_t \in \mathbb{R}^n$ subject to Gaussian observation noise. The state and appearance processes form a linear dynamical system (LDS)

$$\begin{aligned} \mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{v}_t \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{w}_t \end{aligned} \quad (3)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the state transition matrix, $\mathbf{C} \in \mathbb{R}^{m \times n}$ the observation matrix, and $\mathbf{v}_t \sim_{iid} \mathcal{N}(0, \mathbf{Q})$ and $\mathbf{w}_t \sim_{iid} \mathcal{N}(0, \mathbf{R})$ are Gaussian state and observation noise processes, respectively. The initial condition is assumed to be distributed as $\mathbf{x}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{S}_1)$, and the model is parameterized by $\boldsymbol{\Theta} = (\mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\mu}_1, \mathbf{S}_1)$. The hidden state space sequence $\mathbf{x}_t$ is a first order Markov chain that encodes stimulus dynamics, while $\mathbf{y}_t$ is a linear combination of prototypical basis functions (the columns of $\mathbf{C}$) and encodes the appearance component of the stimulus at time t.

3. Background subtraction

In this work, background subtraction is formulated as the complement of saliency detection. Recall that we define saliency with respect to the expected probability of error of the classification problem which opposes the stimulus at a location to that in its surround. In particular, locations of minimal saliency are those where the distinction between stimulus and surround has *lowest confidence*. This provides a natural, objective, definition of *background* based on *strictly local* computations: background points are those of lowest center-surround saliency. We next present a background subtraction algorithm based on this definition.

We start with the estimation of the DT parameters $\boldsymbol{\Theta}$. Given center and surround regions, they could in principle be learned by maximum likelihood (using expectation-maximization [21], or N4SID [18]). However, due to the high dimensionality of video sequences, these solutions are too complex for motion saliency. A suboptimal alternative, that works well in practice, is to learn the spatial and temporal parameters separately [5, 7].

3.1. Probability Distributions

Using the learned model parameters, we can compute probability distributions over the DT. Since the states of a DT form a Markov process with Gaussian conditional probability of $\mathbf{x}_t$ given $\mathbf{x}_{t-1}$, and the initial state conditions are Gaussian, the density of the state sequence, $\mathbf{x}_{1:\tau} = [\mathbf{x}_1^T \dots \mathbf{x}_\tau^T]^T$ is also Gaussian [5]:

$$p(\mathbf{x}_{1:\tau}) = G(\mathbf{x}_{1:\tau}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (4)$$

where $\boldsymbol{\mu} = \left[\begin{array}{ccc} \boldsymbol{\mu}_1^T & \cdots & \boldsymbol{\mu}_\tau^T \end{array} \right]^T$ and the covariance is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{S}_1 & (\mathbf{A}\mathbf{S}_1)^T & \cdots & (\mathbf{A}^{\tau-1}\mathbf{S}_1)^T \\ \mathbf{A}\mathbf{S}_1 & \mathbf{S}_2 & \cdots & (\mathbf{A}^{\tau-2}\mathbf{S}_2)^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{\tau-1}\mathbf{S}_1 & \mathbf{A}^{\tau-2}\mathbf{S}_2 & \cdots & \mathbf{S}_\tau \end{bmatrix}. \quad (5)$$

Similarly, the image sequence $\mathbf{y}_{1:\tau}$ is distributed as

$$p(\mathbf{y}_{1:\tau}) = G(\mathbf{y}_{1:\tau}, \boldsymbol{\gamma}, \boldsymbol{\Phi}) \quad (6)$$

where $\boldsymbol{\gamma} = \boldsymbol{\mathcal{C}}\boldsymbol{\mu}$ and $\boldsymbol{\Phi} = \boldsymbol{\mathcal{C}}\boldsymbol{\Sigma}\boldsymbol{\mathcal{C}}^T + \boldsymbol{\mathcal{R}}$, and $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{R}}$ are block diagonal matrices formed from $\mathbf{C}$ and $\mathbf{R}$ respectively:

$$\boldsymbol{\mathcal{C}} = \begin{bmatrix} \mathbf{C} & 0 & \cdots & 0 \\ 0 & \mathbf{C} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{C} \end{bmatrix}, \boldsymbol{\mathcal{R}} = \begin{bmatrix} \mathbf{R} & 0 & \cdots & 0 \\ 0 & \mathbf{R} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{R} \end{bmatrix}.$$

For a given location l, the densities of (6) can be estimated from a collection of spatio-temporal patches extracted from the center and surround windows. The computation of $S(l)$, with (2), requires the evaluation of the KL divergence between DTs. Let $p_0(\mathbf{y}_{1:\tau})$ and $p_1(\mathbf{y}_{1:\tau})$ be the probabilities of a sequence of τ frames under two DTs parameterized by $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$, respectively. For Gaussian p_0 and p_1, the KL divergence has the closed-form [6]:

$$\begin{aligned} &\mathrm{KL}\left(p_0 \,\|p_1\right) \quad (7) \\ &= \frac{1}{2}\left[\log\frac{|\boldsymbol{\Phi}_1|}{|\boldsymbol{\Phi}_0|} + \mathrm{tr}\left(\boldsymbol{\Phi}_1^{-1}\boldsymbol{\Phi}_0\right) + \|\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}_1\|^2_{\boldsymbol{\Phi}_1} - m\tau\right] \end{aligned}$$

where m is the number of pixels in each frame. Direct evaluation of the KL is computationally intractable, since the expression depends on $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Phi}_1$, which are very large covariance matrices. An efficient recursive procedure is, however, available [4].

3.2. Background subtraction algorithm

Background pixels are identified by computing the saliency map $S(l)$ at each location l. Center and surround windows are centered at the location, and a collection of spatio-temporal patches extracted from each window. DT parameters are then learned, from the center, surround, and total windows, to obtain the densities $p(\mathbf{y}_{1:\tau}|1)$, $p(\mathbf{y}_{1:\tau}|0)$ and $p(\mathbf{y}_{1:\tau})$, respectively. $S(l)$ is finally computed with (2), using the efficient implementation of (7) given in [4]. The procedure is summarized in Algorithm 1, and illustrated in Figure 2. All locations whose saliency is below a threshold are assigned to the background.

4. Experiments

To evaluate background subtraction performance, we tested it on two sequences with object(s) of interest moving in extremely dynamic backgrounds. The sequences were collected from the Internet, and representative frames are shown in panel (a) of Figures 3 - 4. In both cases, the background is non-stationary and complex. Frames in Figure 3(a), depict two people skiing in a heavy snowfall, while those of Figure 4(a) show a surfer riding a wave. The lower frequency sweeping wave is interspersed with high frequency components due to turbulent wakes (created by the surfer, and crest of the sweeping wave) creating significant challenges for background subtraction.

4.1. Comparison to previous methods

To compare the performance of the proposed algorithm (denoted in short as DiscSal) with existing methods, we selected four representatives of the current state of the art in background subtraction - the modified Gaussian mixture model (GMM) of [1, 25], the non-parametric kernel density estimator (KDE) of [8], the linear dynamical model of Monnet et al. [16], and the "surprise" model proposed by Itti and Baldi [13, 14]. The original implementation of Monnet et al. [16] is not publicly available, and the algorithm requires explicit training with background frames. Since no training data was available for the sequences considered, we implemented an adaptive version, where the auto-regressive model parameters were estimated from the 20 frames preceding the location under consideration.

The sequences were converted to grayscale, and saliency maps computed at subsampled locations of the video, using a grid scaled down by a factor of 4 spatially and 2 temporally. At each grid location, the center window occupied 16×16 pixels and spanned 11 frames - 5 past frames , the current frame and 5 frames in the future ($n_c = 16, \tau = 11$). The surround window was, in both cases, set to six times the size of the center. DTs with a 10-dimensional state space, patch dimension $n_p = 8$, and temporal dimension $\tau = 11$, were learned using overlapping $8 \times 8 \times 11$ patches from the center and surround windows.

Saliency maps obtained with DiscSal, Surprise, KDE, Monnet, and GMM are shown in panels (b)-(f), respectively, of Figures 3-4. The proposed algorithms clearly outperform all other methods, detecting the foreground motion and almost entirely ignoring the complex moving background. For all other methods, foreground detection is very noisy, and does not adapt well to the fast background dynamics, sometimes missing the foreground objects completely.

Acknowledgments

This research was supported by NSF awards IIS-0448609, and IIS-0534985. The authors thank Prof. Ahmed Elgammal for providing the code used in [8], Antoni Chan, and Dashan Gao for useful discussions.

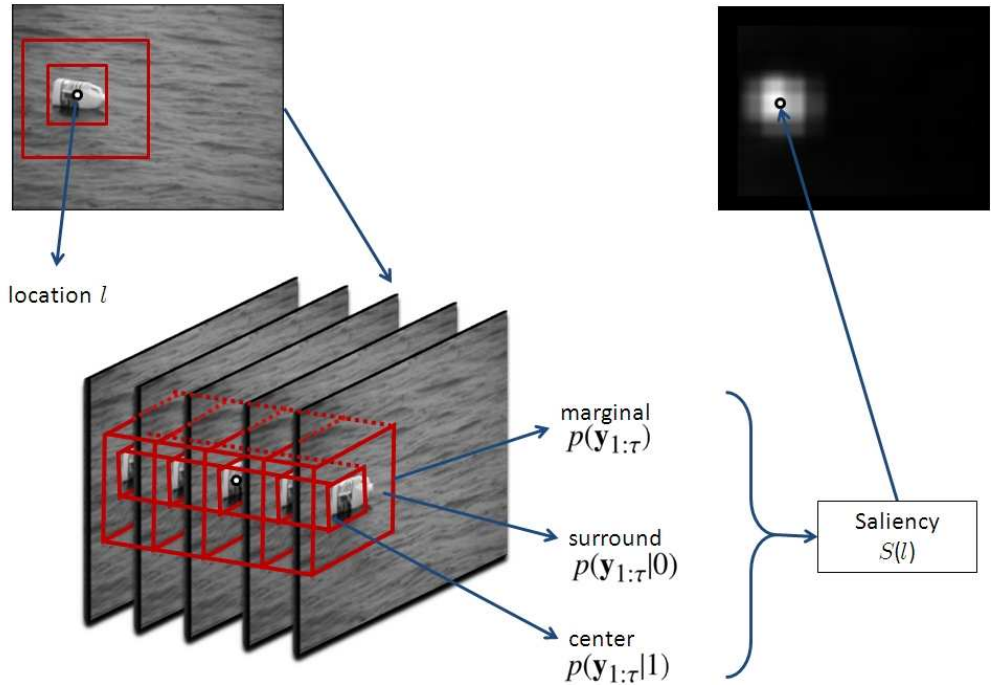

Figure 2. Illustration of the center and surround windows for every location l in the video clip. Using conditional distributions learned from the center and surround window, and the marginal distribution learned from the total window, the saliency measure $S(l)$ is computed using (2).

Algorithm 1 Computing Discriminant Center Surround Motion Saliency

1: **Input:** Given video $\mathcal{V}$ indexed by location vector $l \in L \subset \mathbb{R}^3$, state-space dimension n, center window size n_c, patch size n_p, temporal window τ.
2: **for** $l \in L$ **do**
3: Identify center $\mathcal{W}_l^1$ and surround $\mathcal{W}_l^0$.
4: list all overlapping patches $\{\mathbf{y}_{1:\tau}\}$ of size $n_p \times n_p \times \tau$ in $\mathcal{W}_l^1$ and $\mathcal{W}_l^0$
5: Learn dynamic texture parameters for surround, center and total windows.
6: Compute the class conditional probability density for surround $p(\mathbf{y}_{1:\tau}|0)$ and center $p(\mathbf{y}_{1:\tau}|1)$ and marginal density $p(\mathbf{y}_{1:\tau})$ using (6).
7: Compute the mutual information, $S(l)$, between class-conditional and marginal densities (2), using the efficient implementation of (7) given in [4].
8: **end for**
9: **Output:** Saliency map for $S(l), l \in L$

References

[1] http://staff.science.uva.nl/ zivkovic/download.html.

[2] A. Bugeau and P. Perez. Detection and segmentation of moving objects in highly dynamic scenes. In *CVPR*, 2007.

[3] M. Carandini, J. Demb, V. Mante, D. Tolhurst, Y. Dan, B. Olshausen, J. Gallant, and N. Rust. Do we know what the early visual system does? *J. Neurosci.*, 25, 2005.

[4] A. B. Chan and N. Vasconcelos. Efficient computation of the kl divergence between dynamic textures. Technical Report SVCL-TR-2004-02, Dept. of ECE, UCSD, 2004.

[5] A. B. Chan and N. Vasconcelos. Probabilistic kernels for the classification of auto-regressive visual processes. In *CVPR*, volume 1, pages 846–851, 2005.

[6] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons Inc., New York, 1991.

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

[8] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *ECCV*, pages 751–757, 2000.

[9] D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In *Proc. NIPS*, Vancouver, Canada, 2005.

[10] D. Gao and N. Vasconcelos. Bottom-up saliency is a discriminant process. In *ICCV*, 2007.

[11] D. Gao and N. Vasconcelos. V1 is an optimal saliency detector. In *Computational Cognitive Neuroscience Conference*

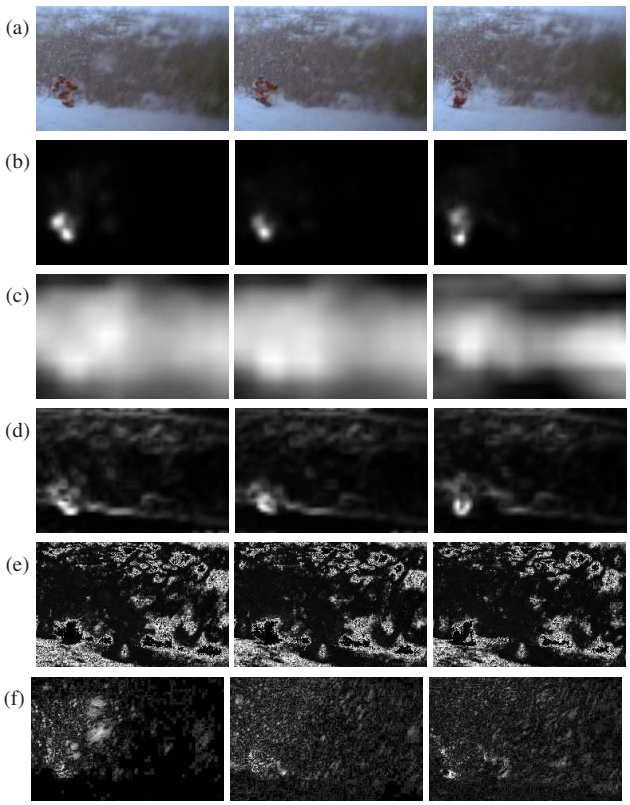

Figure 3. Results on "skiing": (a) original; (b) DiscSal; (c) surprise; and (d) Monnet et al. (e) KDE (f) GMM model.

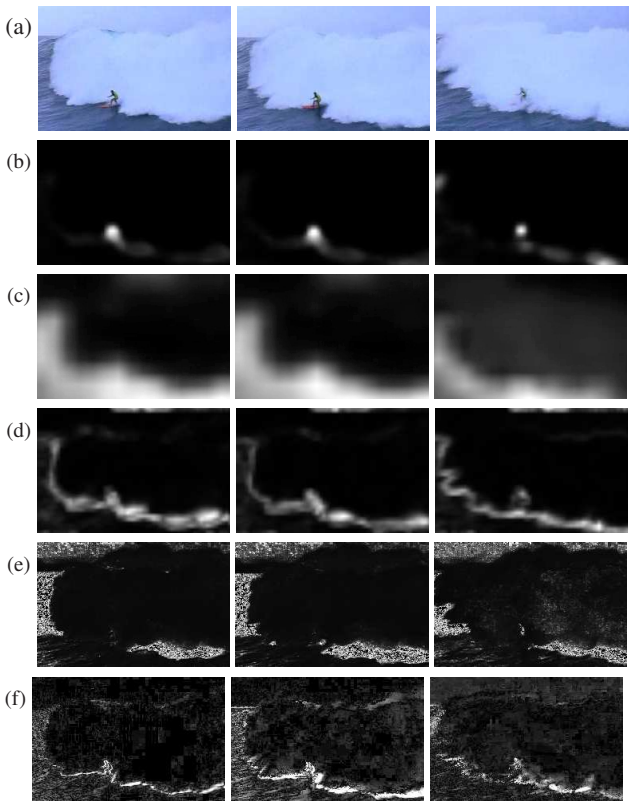

Figure 4. Results on "surf": (a) original; (b) DiscSal; (c) surprise; and (d) Monnet et al. (e) KDE (f) GMM model.

(CCNC), 2007.

[12] J. Huang and D. Mumford. Statistics of Natural Images and Models. In *CVPR*, pages 541–547, 1999.

[13] L. Itti. The ilab neuromorphic vision c++ toolkit: Free tools for the next generation of vision algorithms. *The Neuromorphic Engineer*, 1(1):10, Mar 2004.

[14] L. Itti and P. Baldi. A principled approach to detecting surprising events in video. In *CVPR*, pages 631–637, 2005.

[15] Y. Li. On incremental and robust subspace learning. *Pattern Recognition*, 37(7):1509–19, 2004.

[16] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh. Background modeling and subtraction of dynamic scenes. In *CVPR*, 2003.

[17] A. Murray, D. Basu. Motion tracking with an active camera. *IEEE Trans. PAMI*, 16(5):449–459, 1994.

[18] P. V. Overschee and B. D. Moor. N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30:75–93, 1994.

[19] Y. Ren, C. Chua, and Y. Ho. Motion detection with nonstationary background. *Machine Vision and Applications*, 13(5-6):332–343, 2003.

[20] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE PAMI*, 27(11):1778–92.

[21] R. Shumway and D. Stoffer. An approach to time series smoothing andforecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):433–467, 1982.

[22] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. CVPR*, 1999.

[23] N. Vasconcelos. Feature selection by maximum marginal diversity. In *Proc. NIPS*, Vancouver, Canada, 2002.

[24] L. Wixson. Detecting salient motion by accumulating directionally-consistent flow. *IEEE PAMI*, 22(8):774–780, 2000.

[25] Z. Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. In *ICPR*, 2004.