Bayesian Loop for Synergistic Change Detection and Tracking

Samuele Salti, Alessandro Lanza, and Luigi Di Stefano

Computer Vision Lab
ARCES-DEIS, University of Bologna
Bologna, Italy
{samuele.salti,alessandro.lanza,luigi.distefano}@unibo.it

Abstract. In this paper we investigate Bayesian visual tracking based on change detection. Although in many proposals change detection is key for tracking, little attention has been paid to sound modeling of the interaction between the change detector and the tracker. In this work, we develop a principled framework whereby both processes can virtuously influence each other according to a Bayesian loop: change detection provides a completely specified observation likelihood to the tracker and the tracker provides an informative prior to the change detector.

1 Introduction and Related Work

Recursive Bayesian Estimation (RBE) [1] casts visual tracking as a Bayesian inference problem in state space given noisy observation of the hidden state. Bayesian reasoning has been used also to solve the problem of Change Detection (CD) in image sequences [2], and CD is at the root of many proposals in visual tracking. Nonetheless, interaction between the change detection and tracking modules is usually modeled heuristically. This negatively affects the quality of the information flowing between the two computational levels, as well as the soundness of proposals. Furthermore, the interaction can be highly influenced by heuristically hand-tuned parameters, such as CD thresholds. Hence, a first original contribution of this paper is a theoretically grounded and almost parameters-free approach to provide an observation likelihood to the RBE tracker from the posterior obtained by a Bayesian Change Detection (BCD).

Recently, Cognitive Feedback has emerged as an interesting and effective proposal in Computer Vision [3]. The idea is to let not only low-level vision modules feed high-level ones, but also the latter influence the former. This creates a closure loop, reminiscent of effects found in psychophysics. This concept has not been deployed for the problem of visual tracking yet. Nevertheless, it fits surprisingly well in the case of BCD, where priors can well model the *stimuli* coming from RBE. Hence, the second original contribution of this paper deals with investigating on using Cognitive Feedback to create priors from the state of an RBE tracker in the case of visual tracking based on change detection.

The third novel contribution deals with exploiting the synergy between the previously presented approaches, so as to obtain a fully Bayesian tracking system.

R. Koch et al. (Eds.): ACCV 2010 Workshops, Part I, LNCS 6468, pp. 43–53, 2011.

As a preliminary investigation into this novel approach, in this paper we have conducted the theoretical analysis and the experimental validation only for the simpler case of single-target tracking.

As for related work, a classical work on blob tracking based on background subtraction is W4 [4]. In this system the output of the change detector is thresholded and a connected component analysis is carried out to identify moving regions (blobs). However, the interaction between tracking and change detection is limited, tracking is not formalized in the context of RBE, CD depends on hard thresholds, no probabilistic reasoning is carried out to derive a new measure from the CD output or to update the object position. [5] and [6] are examples of blob trackers based on change detection where the RBE framework is used in the form of the Kalman filter. Yet, the use of this powerful framework is impoverished by the absence of a truly probabilistic treatment of the CD output. In practice, covariance matrices defining measurement and process uncertainties are constant, and the filter evolves toward its steady-state regardless of the quality of the measures obtained from change detection. [7] is one of the most famous attempt to integrate RBE in the form of a particle filter with a statistical treatment of background (and foreground) models. The main limitations are the use of a calibrated camera with reference to the ground plane and the use of a foreground model learned off-line. While the former can be reasonable,the use of foreground models is always troublesome in practice, given the high intra-class variability of human appearances. Moreover, no cognitive feedback is provided from the Particle Filter to influence the change detection.

2 Models and Assumptions

Recursive Bayesian Estimation [1] allows for hidden state estimation from noisy measures in discrete-time systems. From a statistical point of view, the problem of estimating the state translates into the problem of estimating a degree of belief in its possible values, *i.e.* its PDF, given all the available information, *i.e.* the initial state and all the measurements up to a given moment. The solution is seeked recursively: given the PDF of the state at time $k-1$ conditioned on all previous measurements, $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$, and the availability of a new measurement, $\mathbf{z}_k$, a new estimate for the PDF at time k is computed.

We assume a rectangular model for the tracked object, as done in many proposals such as *i.e.* [8]. Hence, the state of the RBE tracker, $\mathbf{x}_k$, comprises at least four variables

$$\mathbf{x}_k = \left\{ i_k^b, j_k^b, w_k, h_k, \dots \right\} \tag{1}$$

where (i_k^b, j_k^b) are the coordinates of the barycenter of the rectangle and w_k and h_k its dimensions. These variables define the position and size at frame k of the tracked object. Of course, the state internally used by the tracker can beneficially include other cinematic variables (velocity,acceleration,...). Yet, change detection can only provide a measure and benefit from a prior of the position and size of the object. Hence, other variables are not used in the reminder of the paper, though they can be used internally by the RBE filter, and are used in our implementation (Sec. 5).

In Bayesian change detection each pixel of the image is modeled as a categorical Bernoulli-distributed random variable, c_{ij}, with the two possible realizations $c_{ij} = \mathcal{C}$ and $c_{ij} = \mathcal{U}$ indicating the event of pixel (i, j) being changed or unchanged, respectively.

In the following we refer to the matrix $\mathbf{c} = [c_{ij}]$ of all these random variables as the *change mask* and to the matrix $\mathbf{p} = [p(c_{ij} = \mathcal{C})]$ of probabilities defining the Bernoulli distribution of these variables as *change map.* The change mask and the change map assume values, respectively, in the $(w \times h)$-dimensional spaces $\Theta = \{\mathcal{C}, \mathcal{U}\}^{w \times h}$ and $\Omega = [0, 1]^{w \times h}$, with w and h denoting image width and height, respectively. The output of a Bayesian change detector is the posterior change map given the current frame f_k and background model b_k, *i.e.* the value of the Bernoulli distribution parameter for every pixel in the image given the frame and the background:

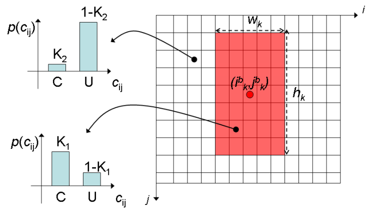


Fig. 1. Model for the change map given a bounding box

$$p(c_{ij} = \mathcal{C} | f_k, b_k) = \frac{p(f_k, b_k | c_{ij} = \mathcal{C}) p(c_{ij} = \mathcal{C})}{p(f_k, b_k)} \tag{2}$$

Clearly, either a non-informative prior is used, such as a uniform prior, or this information has to flow in from an external module. We assume that the categorical random variables c_{ij} comprising the posterior change mask are independent, *i.e.* they are conditionally independent given f_k, b_k.

All the information that can flow from the RBE filter to the BCD and vice versa is in principle represented in every frame by the joint probability density function $p(\mathbf{x}_k, \mathbf{c})$ of the state vector and the change mask. Both information flows can be formalized and realized as its marginalization:

$$p(c_{ij}) = \int_{\mathbb{R}^4} \sum_{\mathbf{c}^{ij} \in \Theta^{ij}} p(\mathbf{x}_k, c_{ij}, \mathbf{c}^{ij})\, d\mathbf{x}_k \quad (3) \qquad\qquad p(\mathbf{x}_k) = \sum_{\mathbf{c} \in \Theta} p(\mathbf{x}_k, \mathbf{c}) \quad (4)$$

where $\mathbf{c}^{ij}$ denotes the change mask without the (i, j)-*th* element, taking values inside the space $\Theta^{ij} = \{\mathcal{C}, \mathcal{U}\}^{w \times h - 1}$. The PDF computed with (3) defines an informative prior for the BCD algorithm, and the estimation of the state obtained with (4) can then be used as the PDF of a new measure by the RBE tracker, *i.e.* as $p(\mathbf{z}_k | \mathbf{x}_k)$. We detail in Sec. 3 and Sec. 4 the solutions for (3) and (4).

As we shall see in next sections, to use the above equations we need a statistical model that links the two random vectors $\mathbf{x}_k$ and $\mathbf{c}$. In agreement with our rectangular model of the tracked object, as shown in Fig. 1 we assume

$$p(c_{ij} = \mathcal{C} | \mathbf{x}_k) = \begin{cases} K_1 & \text{if } (i, j) \in R(\mathbf{x}_k) \\ K_2 & \text{otherwise} \end{cases} \tag{5}$$

where $R(\mathbf{x}_k)$ is the rectangular region delimited by the bounding box defined by the state $\mathbf{x}_k$ and $0 \leq K_2 \leq K_1 \leq 1$ are two constant parameters specifying the probability that a pixel is changed inside and outside the bounding box, respectively. Moreover, we assume that the random variables c_{ij} are conditionally independent given a bounding box, *i.e.*

$$p(\mathbf{c}|\mathbf{x}_k) = \prod_{ij} p(c_{ij}|\mathbf{x}_k) \tag{6}$$

3 Cognitive Feedback

Given the assumptions of the previous section, we can obtain an exact solution for (3), *i.e.* , given the PDF of the state vector $p(\mathbf{x}_k)$, we can compute a prior $p(c_{ij})$ for each pixel of the frame that can then be beneficially used by the BCD algorithm. Starting from (3), we can rewrite it as

$$p(c_{ij}) = \int_{\mathbb{R}^4} \sum_{\mathbf{c}^{ij} \in \Theta^{ij}} p(\mathbf{x}_k, c_{ij}, \mathbf{c}^{ij})\, d\mathbf{x}_k = \int_{\mathbb{R}^4} p(\mathbf{x}_k, c_{ij})\, d\mathbf{x}_k = \int_{\mathbb{R}^4} p(c_{ij}|\mathbf{x}_k)p(\mathbf{x}_k)\, d\mathbf{x}_k \tag{7}$$

In the final marginalization we can recognize our model of the change map given a bounding box defined in (5) and the PDF of the state. Therefore, this equation provides a way to let the current estimation of the state computed by the RBE module influence the prior for the BCD algorithm, thereby realizing the Cognitive Feedback. In particular, as discussed above, we will use the prediction computed for the current frame using the motion model, *i.e.* $p(\mathbf{x}_k|\mathbf{z_{1:k-1}})$.

To solve (7) we have to span the space $\mathbb{R}^4$ of all possible bounding boxes $\mathbf{x}_k$. We partition $\mathbb{R}^4$ into the two complementary sub-spaces B_{ij} and $\bar{B}_{ij} = \mathbb{R}^4 \setminus B_{ij}$ of bounding boxes that contain or not the considered pixel (i, j), respectively. Given the assumed model (5), we obtain

$$\begin{aligned} p(c_{ij} = \mathcal{C}) &= \int_{\mathbb{R}^4} p(c_{ij}|\mathbf{x}_k)p(\mathbf{x}_k)\, d\mathbf{x}_k = K_1 \int_{B_{ij}} p(\mathbf{x}_k)\, d\mathbf{x}_k + K_2 \int_{\bar{B}_{ij}} p(\mathbf{x}_k)\, d\mathbf{x}_k \\ &= K_2 + (K_1 - K_2) \int_{B_{ij}} p(\mathbf{x}_k)\, d\mathbf{x}_k\ . \end{aligned} \tag{8}$$

Since, obviously, $I_{ij} = \int_{B_{ij}} p(\mathbf{x}_k)\, d\mathbf{x}_k$ varies in $[0, 1]$, it follows that $p(c_{ij} = \mathcal{C})$ varies in $[K_2, K_1]$: if no bounding box with non-zero probability contains the pixel, we expect a probability that the pixel is changed equal to K_2, if all the bounding boxes contain the pixel the probability is K_1, it is a weighted average otherwise.

By defining new variables i_L, j_T, i_R, j_B to represent the current bounding box, more suitable for the next computations, as

$$\mathbf{A} = \begin{bmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{bmatrix}\ ,\ \begin{bmatrix} i_L \\ i_R \end{bmatrix} = \mathbf{A} \begin{bmatrix} w_k \\ i_k^b \end{bmatrix}\ ,\ \begin{bmatrix} j_T \\ j_B \end{bmatrix} = \mathbf{A} \begin{bmatrix} h_k \\ j_k^b \end{bmatrix} \tag{9}$$

and assuming the newly defined random variables to be independent, the integral of the previous equation becomes

$$
\begin{aligned}
I_{ij} &= \underset{\left.\begin{matrix} i_L \le i \le i_R \\ j_T \le j \le j_B \end{matrix}\right\} B_{ij}}{\iiiint} p(i_L)p(i_R)p(j_T)p(j_B)\, di_L di_R dj_T dj_B \\
&= \int_{-\infty}^{i} p(i_L)di_L \int_{i}^{+\infty} p(i_R)di_R \int_{-\infty}^{j} p(j_T)dj_T \int_{j}^{+\infty} p(j_B)dj_B \\
&= F_{i_L}(i)\left(1 - F_{i_R}(i)\right) F_{j_T}(j)\left(1 - F_{j_B}(j)\right) \qquad (10)
\end{aligned}
$$

where F_x stands for the CDF of the random variable x.

This reasoning holds for any distribution $p(\mathbf{x}_k)$ we might have on the state vector. If, for instance, we use a particle filter as RBE tracker, we can compute an approximation of the CDF from the approximation of the PDF provided by the weighted particles, after having propagated them according to the motion model and having marginalized them accordingly. In the case of the Kalman Filter all the PDFs are Gaussians, hence we can define all the factors of the product in (10) in terms of the standard Gaussian CDF, $\Phi(\cdot)$

$$
I_{ij} = \Phi\left(\frac{i - \mu_L}{\sigma_L}\right) \Phi\left(\frac{\mu_R - i}{\sigma_R}\right) \Phi\left(\frac{j - \mu_T}{\sigma_T}\right) \Phi\left(\frac{\mu_B - j}{\sigma_B}\right) \qquad (11)
$$

where μ_x and σ_x stand for the mean and the standard deviation of the random variable x. The factors of the product in (11) can be computed efficiently with only 4 searches in a pre-computed Look-Up Table of the standard $\Phi(\cdot)$ values.

4 Reasoning Probabilistically on Change Maps

Given the change map $\mathbf{p} = [p(c_{ij} = \mathcal{C})]$ obtained by the BCD algorithm, we aim at computing the probability density function $p(\mathbf{x}_k)$ of the current state of the RBE filter, to use it as the observation likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$. To this purpose, from the marginalization in (4) we obtain:

$$
p(\mathbf{x}_k) = \sum_{\mathbf{c} \in \Theta} p(\mathbf{x}_k, \mathbf{c}) = \sum_{\mathbf{c} \in \Theta} p(\mathbf{x}_k|\mathbf{c})p(\mathbf{c}) = \sum_{\mathbf{c} \in \Theta} p(\mathbf{x}_k|\mathbf{c}) \prod_{ij} p(c_{ij}) \qquad (12)
$$

where the last equality follows from the assumption of independence among the categorical random variables c_{ij} comprising the posterior change map computed by BCD. To use (12), we need an expression for the conditional probability $p(\mathbf{x}_k|\mathbf{c})$ of the state given a change mask, based on the assumed model (5), (6) for the conditional probability $p(\mathbf{c}|\mathbf{x}_k)$ of the change mask given a state. Informally speaking, we need to find the inverse of the model (5), (6). By Bayes rule, eq. (6) and independence of the variables c_{ij}:

$$
p(\mathbf{x}_k|\mathbf{c}) = p^*(\mathbf{x}_k) \frac{p(\mathbf{c}|\mathbf{x}_k)}{p^*(\mathbf{c})} = p^*(\mathbf{x}_k) \prod_{i,j} \frac{p(c_{ij}|\mathbf{x}_k)}{p^*(c_{ij})} \ . \qquad (13)
$$

It is worth pointing out that we have used the notation $p^*(\mathbf{x}_k)$ and $p^*(c_{ij})$ in (13) since here these probabilities must be interpreted differently than in (12): in (12) $p(\mathbf{x}_k)$ and $p(c_{ij})$ represent, respectively, the measurement and the change map of the current frame, whilst in (13) both must be interpreted as priors that form part of our model for $p(\mathbf{x}_k|\mathbf{c})$, which is independent of the current frame. Furthermore, using as prior on the state $p^*(\mathbf{x}_k)$ the prediction of the RBE filter, as done in the Cognitive Feedback section, would have created a strong coupling between the output of the sensor and the previous state of the filter, that does not fit the RBE framework, where measures depend only on the current state, and could easily lead the loop to diverge. Hence, we assume a uniform non-informative prior $p^*(\mathbf{x}_k) = \frac{1}{\alpha}$ for the state. Instead, the analysis conducted for the Cognitive Feedback is useful to expand each $p^*(c_{ij})$ in (13). Since we are assuming a uniform prior on an infinite domain for the state variables, *i.e.* a symmetric PDF with respect to $x = 0$, it turns out that its CDF is constant and equals to $\frac{1}{2}$:

$$CDF(x) = \frac{1}{\alpha}x + \frac{1}{2} \xrightarrow{\alpha \to +\infty} \frac{1}{2} \tag{14}$$

Hence, every $p^*(c_{ij})$ in (13) can be expressed using (8) and (10) as:

$$p^*(c_{ij} = \mathcal{C}) = K_2 + (K_1 - K_2)\left(\frac{1}{2}\right)^4 = K_C. \tag{15}$$

By plugging (13) in (12) and defining $K_U = p^*(c_{ij} = \mathcal{U}) = 1 - K_C$:

$$\alpha p(\mathbf{x}_k) = \prod_{i,j}\left(\frac{p(\mathcal{C}|\mathbf{x}_k)p(\mathcal{C})}{K_C} + \frac{p(\mathcal{U}|\mathbf{x}_k)p(\mathcal{U})}{K_U}\right) \tag{16}$$

where, for simplicity of notation, we use $\mathcal{C}$ and $\mathcal{U}$ for $c_{ij} = \mathcal{C}$ and $c_{ij} = \mathcal{U}$, respectively. Since we know that $p(\mathcal{U}) = 1 - p(\mathcal{C})$ and $p(\mathcal{U}|\mathbf{x}_k) = 1 - p(\mathcal{C}|\mathbf{x}_k)$, we obtain:

$$\frac{p(\mathbf{x}_k)}{\beta} = \prod_{i,j}\Big(p(\mathcal{C})\big(p(\mathcal{C}|\mathbf{x}_k) - K_C\big) + K_C\big(1 - p(\mathcal{C}|\mathbf{x}_k)\big)\Big) \tag{17}$$

with $\beta = 1/\alpha(K_C(1-K_C))^{w \times h}$. By substituting the model (5) for $p(\mathcal{C}|\mathbf{x}_k)$ and taking the logarithm of both sides to limit round-off errors, after some manipulations we get:

$$\gamma + \ln p(\mathbf{x}_k) = h(\mathbf{x}_k, \mathbf{p}) = \sum_{(i,j)\in R(\mathbf{x}_k)} \ln \frac{p(\mathcal{C})K_3 + K_4}{p(\mathcal{C})K_5 + K_6} \tag{18}$$

where $\gamma = -\ln\beta - \sum \ln\big(p(\mathcal{C})K_5 + K_6\big)$ and $h(\cdot)$ is a computable function of the state vector value $\mathbf{x}_k$ for which we want to calculate the probability density, of the change map $\mathbf{p}$ provided by the BCD algorithm, and of the constants

$$K_3 = K_1 - K_C \quad K_4 = K_C\,(1 - K_1) \quad K_5 = K_2 - K_C \quad K_6 = K_C\,(1 - K_2) \tag{19}$$

Hence, by letting $\mathbf{x}_k$ vary over the space of all possible bounding boxes, (18) allows us to compute, up to the additive constant γ, a non-parametric estimation

$h(\cdot)$ of the log-PDF of the current state vector of the RBE tracker. This holds independently of the PDF of the state.

In the case of the Kalman Filter, the PDF of the state vector (i^b, j^b, w, h) is Gaussian. In such a case, the variables (i_L, j_T, i_R, j_B) are a linear combination of Gaussian Random Variables. Moreover, we are assuming that variables (i_L, j_T, i_R, j_B) are independent. Therefore, the variables (i_L, j_T, i_R, j_B) are jointly Gaussian and the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of the state variables are fully defined by the four means μ_L, μ_R, μ_T, μ_B and the four variances σ_L^2, σ_R^2, σ_T^2, σ_B^2 of (i_L, j_T, i_R, j_B). To estimate these eight parameters, let us substitute the expression of the Gaussian PDF for $p(\mathbf{x}_k)$ in the left-hand side of (18), thus obtaining:

$$\delta - \ln(\sigma_L \sigma_R \sigma_T \sigma_B) - \frac{(i_L - \mu_L)^2}{2\sigma_L^2} - \frac{(i_R - \mu_R)^2}{2\sigma_R^2} - \frac{(j_T - \mu_T)^2}{2\sigma_T^2} - \frac{(j_B - \mu_B)^2}{2\sigma_B^2} = h(\mathbf{x}_k, \mathbf{p}) \tag{20}$$

where $\delta = \gamma - 2\ln(2\pi)$. The eight parameters of the PDF and the additive constant δ might be estimated by imposing (20) for a number $N > 9$ of different bounding boxes and then solving numerically the obtained over-determined system of N non-linear equations in 9 unknowns. To avoid such a challenging problem, we propose an approximate procedure. First of all, an estimate $\widehat{\boldsymbol{\mu}}$ of the mean of the state vector $\boldsymbol{\mu} = (\mu_L, \mu_R, \mu_T, \mu_B)$ can be obtained by observing that, due to increasing monotonicity of logarithm, the mode of the computed log-PDF coincides with the mode of the PDF, and that, due to the Gaussianity assumption, the mode of the PDF coincides with its mean. Hence, we obtain an estimate $\widehat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ by searching for the bounding box maximizing $h(\cdot)$. Then, we impose that (20) is satisfied at the estimated mean point $\widehat{\boldsymbol{\mu}}$ and that all the variances are equal, *i.e.* $\sigma_L^2 = \sigma_R^2 = \sigma_T^2 = \sigma_B^2 = \sigma^2$, thus obtaining a functional relationship between the two remaining parameters δ and σ^2:

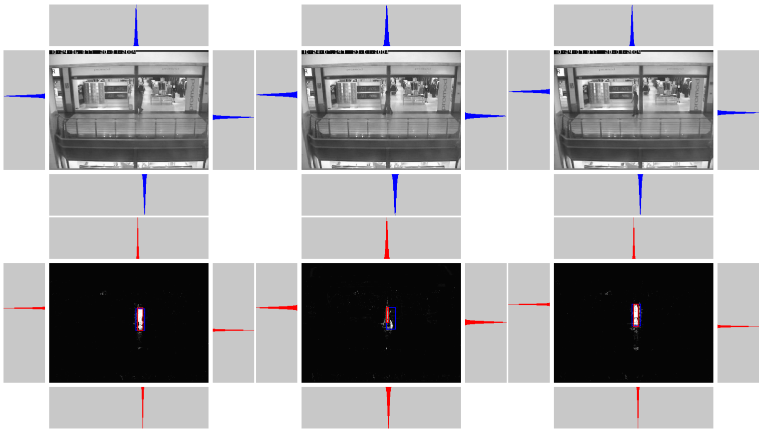
$$\delta = 2\ln\sigma^2 + h(\widehat{\boldsymbol{\mu}}, \mathbf{p}) \tag{21}$$

By substituting in (20) the above expression for δ and the estimated $\widehat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$, we can compute an estimate $\widehat{\sigma}^2(\mathbf{x})$ of the variance σ^2 by imposing (20) for whatever bounding box $\mathbf{x} \neq \widehat{\boldsymbol{\mu}}$. In particular, we obtain:

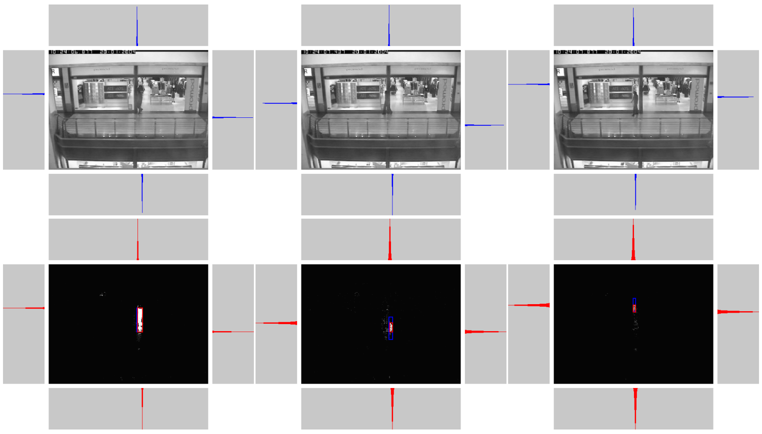
$$\widehat{\sigma}^2(\mathbf{x}) = \frac{1}{2} \frac{\|\widehat{\boldsymbol{\mu}} - \mathbf{x}\|_2^2}{h(\widehat{\boldsymbol{\mu}}, \mathbf{p}) - h(\mathbf{x}, \mathbf{p})} \tag{22}$$

To achieve a more robust estimate, we average $\widehat{\sigma}^2(\mathbf{x})$ over a neighborhood of the estimated mean bounding box $\widehat{\boldsymbol{\mu}}$. Finally, to obtain the means and covariance of the measurements for the Kalman Filter, we exploit the property of linear combinations of Gaussian variables:

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \end{bmatrix} \widehat{\boldsymbol{\mu}} \ , \ \boldsymbol{\Sigma} = \widehat{\sigma}^2 \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \end{bmatrix}^T \tag{23}$$



(a) Our proposal



(b) Constant Measurement Covariance Matrix

Fig. 2. The top row shows the frames and the next row the change maps. Along the sides of every picture we plot the marginal Gaussian probabilities of the four state variables $[i_L, i_R, j_T, j_B]$. Around the frames we report (in blue) the marginals of the Kalman prediction and around the change maps (in red) the marginals of the observation likelihood (Sec. 4). The means of the PDFs are drawn on the change maps.

5 Experimental Results

We have tested the proposed Bayesian loop on publicly available datasets with ground truth data: some videos from the CAVIAR[1]and ISSIA Soccer datasets [9]. We have used a Kalman Filter with constant velocity motion model as RBE tracker and the algorithm in [10] as BCD. The detection to initialize the tracker was done manually from the ground truth.

[1] `http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`

To illustrate the benefits of the probabilistic analysis of the change map, we discuss some frames of a difficult part of a CAVIAR sequence. In this video, the tracked subject wanders in and out of the shop, passing in front of a pillar similar in color to his clothes. Fig. 2(a) shows 3 frames of the sequence, respectively before, during and after the camouflage. During the camouflage the background subtraction correctly computes high probabilities that a pixel is changed only for the pixels lying outside the pillar. Our rectangular model cannot fit such output and selects only the portion of the person on the left of the column as the mean of the current PDF (red bounding box). A sensor based on change detection will always likely fail to handle camouflage. Yet, the RBE tracker is conceived to work with a noisy sensor, provided that it is possible to evaluate the uncertainty of its output. Thanks to the procedure of Sec. 4, our method exploits this trait of the framework: as can be seen in the middle pictures, the uncertainty of the measure during the camouflage increases and gets similar to the Kalman prediction uncertainty, for the rectangular model leaves out portions of the change map with high probabilities of being foreground. With this configuration, the correction step of the filter decreases the contribution to the final state estimation of the shrunk measure coming from BCD with respect to the predicted state, thus correctly tracking the target. Had a constant uncertainty model been used (see Fig. 2(b)), by the time of the camouflage the filter would have reached the steady state and would follow the measures with a constant amount of confidence in them. This leads to an incorrect prior for the change detection on the next frame and thus to divergence, as shown in the third frame of Fig. 2(b) where the system cannot recover from wrong measurements.

To demonstrate its capabilities and robustness, the complete system has been used to track people wondering in a shopping mall using three sequences from the CAVIAR dataset and soccer players during a match in the sixth sequence of the ISSIA dataset. Tracking results on these video are shown in the supplementary material . Our system does not require to set a threshold to classify the output of the change detection, only the model for $p(c_{ij} = \mathcal{C}|\mathbf{x}_k)$ must be set. We used $K_1 = 0.5$, allowing for unchanged pixels into our bounding box (approximation of the rectangular model) and $K_2 = 0.1$ to allow for a small amount of errors of the BCD out of the bounding box. To quantitatively evaluate the performance we use the mean ratio over a sequence between the intersection and the union of the ground truth bounding box with the estimated bounding box.

Table 1. Performance scores. $^{(*)}$ indicates loss of target.

Seq.	Full Loop	Partial Loop	Kalm+MS
CAV1	0.553	0.298	$0.208^{(*)}$
CAV2	0.474	0.382	$0.010^{(*)}$
CAV3	0.500	$0.055^{(*)}$	$0.012^{(*)}$
I_GK	0.457	$0.011^{(*)}$	0.581
I_PE	0.474	$0.012^{(*)}$	0.492

As for the CAVIAR dataset, the main difficulties are changes in appearance of the target due to lightening changes inside and outside the shop, shadows, camouflage, small size of the target and, for sequence 2, dramatic changes in target size onto the image plane (he walks inside the shop until barely disappears). Despite all these nuisances our system successfully tracks all the targets.

The ISSIA Soccer dataset is less challenging as far as color, lightening and size variations are concerned, and the players cast practically no shadow. Yet, it provides longer sequences and more dynamic targets. We used our system to track the goalkeeper and a player: the goalkeeper allows to test our system on a sequence 2500 frames long; the player shows rapid motion changes and unpredictable poses (he even falls to the ground kicking the ball). Our tracker was able to successfully track both targets throughout the whole sequence. Quantitative evaluation is reported in Table 1. To highlight the importance of the Bayesian loop, we have performed the same experiments without considering the full PDF estimated during the change map analysis, but just the mean and a constant covariance matrix (i.e. the same settings as in Fig. 2(b)): results achieved by our proposal are consistently better throughout all the sequences. We also compare our performance against a Mean Shift tracker used in conjunction with a Kalman Filter [8]. The CAVIAR sequences are too difficult for a tracker based on color histograms, because of the reasons discussed above: the tracker looses the target in all the sequences. On the ISSIA sequences, instead, it obtains slightly better performances than our proposal. We impute this to the use of gray levels in our tracker: for example, yellow parts of the tracked players get really similar to the green background. We are developing a color BCD to solve the problem.

6 Conclusions

A principled framework to model the interaction between Bayesian change detection and tracking have been presented. By modeling the interaction as marginalization of the joint probability of the tracker state and the change mask, it is possible to obtain analytical expressions for the PDFs of the tracker observation likelihood and the change detector prior. Benefits of such interaction have been discussed with experiments on publicly available datasets.

References

1. Maskell, S., Gordon, N.: A tutorial on particle filters for on-line nonlinear/non-gaussian bayesian tracking. IEEE Trans. on Signal Processing 50, 174–188 (2001)
2. Lanza, A., Di Stefano, L., Soffritti, L.: Bayesian order-consistency testing with class priors derivation for robust change detection. In: AVSS, pp. 460–465 (2009)
3. Thomas, A., Ferrari, V., Leibe, B., Tuytelaars, T., Gool, L.V.: Depth-from-recognition: Inferring metadata by cognitive feedback. In: 3DRR (2007)
4. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: Real-time surveillance of people and their activities. IEEE T-PAMI 22, 809–830 (2000)
5. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: Proc. of the CVPR 1999, pp. 246–252 (1999)
6. Harville, M., Li, D.: Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera. In: CVPR, vol. 2, pp. 398–405 (2004)
7. Isard, M., MacCormick, J.: Bramble: A bayesian multiple-blob tracker. In: Proc. of the ICCV 2001 (2001)

8. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. PAMI 25, 564–575 (2003)
9. D'Orazio, T., Leo, M., Mosca, N., Spagnolo, P., Mazzeo, P.: A semi-automatic system for ground truth generation of soccer video sequences. In: AVSS (2009)
10. Lanza, A., Di Stefano, L.: On-line learning of background appearance changes for robust background subtraction. In: IET Int. Conf. on Imaging for Crime Detection and Prevention (2009)