

# Non-linear Parametric Bayesian Regression for Robust Background Subtraction

Federico Tombari

federico.tombari@unibo.it

Luigi Di stefano

luigi.distefano@unibo.it

Alessandro Lanza

alanza@arces.unibo.it

Stefano Mattoccia

stefano.mattoccia@unibo.it

DEIS - ARCES, University of Bologna, Bologna, Italy

[www.vision.deis.unibo.it](http://www.vision.deis.unibo.it)

## Abstract

*This paper presents a novel statistical method for background subtraction aimed at robustness with regards to common disturbance factors such as sudden illumination changes, variations of the camera parameters, noise. The proposed approach relies on a novel non-linear parametric model for the local effect of disturbance factors on a neighbourhood of pixel intensities. Assuming additive gaussian noise, we also propose Bayesian estimation of model parameters by means of a maximum-a-posteriori regression and a statistical change detection test. Experimental results demonstrate that the proposed approach is state-of-the-art in sequences where disturbance factors yield linear as well as non-linear intensity transformations.*

## 1. Introduction

Many computer vision applications, such as video surveillance and traffic monitoring, rely on change detection as a first processing step upon which to build higher level capabilities, such as target tracking and classification. Given two or more images taken by a stationary camera at different times, the goal of change detection is to identify the pixels in one image sensing a *significant* change with respect to the others. Image changes are regarded as significant if they can be ascribed to variations in the 3-d geometry of the monitored scene. We call *disturbance factors* the causes of image changes not related to a varying geometry. Main disturbance factors are illumination changes, camera gain and exposure variations, noise. Hence, a robust change detector should be able to discriminate between the image change patterns induced, respectively, by scene changes and by disturbance factors in the chosen feature space.

A variety of algorithms dealing with this pattern classification problem have been proposed [10], with those

based on background subtraction being definitely the most widely adopted. A first class of algorithms detect changes by comparing the current pixel intensity with a pixel-wise time-adaptive statistical model of the background appearance. Different models have been proposed, ranging from a single Gaussian [13] to mixtures of Gaussians [12] and kernel-based non-parametric models [4]. Though different background model maintenance procedures have been proposed as well, pixel-wise time-adaptive approaches are inherently not robust against disturbance factors yielding sudden image changes. In fact, temporal consistency of pixel intensities can be violated in an undistinguishable manner as a consequence of scene changes and of disturbance factors. A second class of algorithms relies on a priori modelling the effect of disturbance factors within a spatially extended support, i.e. a neighbourhood of pixels in the background and in the current frame. This allows for handling seamlessly slow as well as sudden spurious image changes. The algorithm proposed in this paper belongs to this second class, which can be further split into two major sub-classes based on the a priori model adopted for disturbance factors.

The first sub-class (e.g. [3, 6, 9, 11]) assumes a linear transformation of intensities, so that change detection consists in a linearity test. In [11] the test is carried out by simply thresholding the variance of the inter-frame pixel-by-pixel intensity ratios computed in the considered neighbourhood. In [3] the authors regard change detection as a linear algebra problem and, accordingly, propose a linear-dependence test based on the concept of Wronskian determinant. In [6] a statistical linearity test is proposed based on the well known total least squares linear regression technique. By assuming additive, zero-mean, i.i.d gaussian noise, the sum of square residuals between the sensed intensities and the maximum likelihood estimated transformation is  $\chi^2$ -distributed, thus allowing for a significant linearity test. Finally, in [9] the author adopts an affine image formation model but shows how the model offset, which

is assumed to be a characteristic of the imaging device, can be computed once for all through a calibration process. The compared images are thus transformed by subtracting the computed offset from each pixel intensity. This renders linear the virtual image formation model, so that scene changes can be detected by a linearity test. The author proposes a statistical test which is very similar to that in [6]. The only difference is that noise variance is allowed to be different in the two images.

In order to account for the non-linearities possibly arising in the image formation process, a second sub-class of algorithms (e.g. [1, 7, 14, 15]) relies on the more general assumption that disturbance factors yield locally order-preserving intensity transformations. Most of these algorithms are image transform methods, actually proposed for visual correspondence purposes. The two images under comparison are transformed by replacing the intensity of each pixel with a feature vector that encodes the rankings between the intensities measured in a surrounding neighbourhood. Criteria are then developed to measure scene changes based on variations in the rankings. In [15] the authors propose two algorithms based, respectively, on the *rank transform* and the *census transform*. In [1] images are transformed by replacing each pixel intensity with the rank permutation of intensities in the surrounding neighbourhood. A novel distance between rank permutations is proposed to carry out a robust order-consistency test. Although [1] and [15] are both inherently robust to locally order-preserving transformations of intensities, they can not deal effectively with noise. This problem has been dealt with implicitly in [7]. In fact, each order flip between pairs of pixel intensities contributes to the proposed order-inconsistency measure according to an increasing function of the absolute difference of the two intensities. Finally, an explicit model of noise is exploited in [14] to carry out a statistical order-consistency test. Once the noise distribution has been inferred, a statistical change measure is computed based on the probability of an order flip between noiseless intensities given the measured noisy intensities.

It is worth observing that the set of image changes assumed to arise from disturbance factors by the algorithms of the first sub-class is contained into the set assumed by those of the second sub-class. Clearly, the larger such a set the higher is robustness to disturbance factors. As shown in [14], the larger set assumed by order-preserving algorithms is required to achieve adequate robustness in many practical applications. On the other hand, a larger set implies a lower sensitivity, for true scene changes being more easily misclassified as disturbance factors.

In this paper we propose a novel background subtraction algorithm belonging to the second class which models spurious image changes through a set containing the smaller

one of linear algorithms and being contained into the larger one of order-preserving algorithms. In this manner, our approach holds the potential for achieving a better tradeoff between robustness and sensitivity. In particular, we assume that disturbance factors yield locally non-linear parametric transformations and, accordingly, carry out a regression in order to estimate parameters and derive a statistically sound change detection test.

## 2. Problem definition and formalization

Let us consider two grey level images<sup>1</sup> (i.e. background and current frame) taken at different times by a stationary camera and, for a generic pixel, the surrounding neighbourhood on the basis of which the changed/unchanged classification is to be performed. Let us denote as

$$X = (x_1, \dots, x_N) \quad \text{and} \quad Y = (y_1, \dots, y_N) \quad (1)$$

the neighbourhood intensities measured, respectively, in the background and in the current frame, with pixels taken in lexicographical order as shown in Figure 1. Scene changes occurring in the central pixel are detected by exploiting the local intensity information contained in  $X$  and  $Y$ .

$x_1$	$x_2$	$x_3$	$y_1$	$y_2$	$y_3$
$x_4$	$x_5$	$x_6$	$y_4$	$y_5$	$y_6$
$x_7$	$x_8$	$x_9$	$y_7$	$y_8$	$y_9$

Figure 1. Notations adopted for the background (on the left) and the current frame (on the right) neighbourhood intensities.

To formalize the problem within a classification framework, let the two  $N$ -dimensional intensity vectors  $X$  and  $Y$  be concatenated to form the  $2N$ -dimensional feature vector

$$\mathbf{f} = (X \parallel Y) = (x_1, \dots, x_N, y_1, \dots, y_N) \in F \quad (2)$$

where  $F = [0, L]^{2N} \subset \mathbb{Z}^{2N}$  is the (discrete) space of observable feature vectors and  $L$  is the number of measurable grey levels (e.g.,  $L = 256$  for 8-bit grey level images). Based on the observed feature vector  $\mathbf{f}$ , the pixel must be classified into one of the two mutually exclusive classes:

- $c = C$ : a scene change occurs;
- $c = U$ : only disturbance factors are acting.

The Bayesian binary classification rule is the following:

$$\begin{aligned}
 p(U|\mathbf{f}) & \underset{C}{\overset{U}{\gtrless}} T \cdot p(C|\mathbf{f}) \\
 & \Downarrow \\
 \frac{p(\mathbf{f}|U)}{p(\mathbf{f}|C)} & \underset{C}{\overset{U}{\gtrless}} T \cdot \frac{p(C)}{p(U)} = T \cdot \frac{p(C)}{1 - p(C)}
 \end{aligned} \quad (3)$$

<sup>1</sup>the proposed approach can be extended straightforwardly to colour images by applying it to each channel independently.

where  $p(U|\mathbf{f})$  and  $p(C|\mathbf{f})$ ,  $p(U)$  and  $p(C)$ ,  $p(\mathbf{f}|U)$  and  $p(\mathbf{f}|C)$  are, respectively, the class posteriors, the class priors and the likelihoods,  $T$  is a threshold allowing for tuning the desired sensitivity-specificity tradeoff. The prior  $p(C)$  is the probability we would assign to a given pixel being sensing a scene change in absence of the measurement  $\mathbf{f}$ . Though in principle such a prior might be derived by exploiting information other than  $\mathbf{f}$  (e.g. predicting foreground objects position and extent by temporal tracking), this is not considered in this paper and, hence, we set  $p(C) = p(U) = 0.5$ . Moreover, a statistical characterization of the likelihood  $p(\mathbf{f}|C)$  is generally unfeasible. In fact,  $p(\mathbf{f}|C)$  represents the probability of measuring the feature vector  $\mathbf{f}$  given that a foreground object is covering the considered neighbourhood. Hence, unless some knowledge is assumed or learned off-line about foreground objects appearance (e.g. colour and shape), scene changes do not yield statistically predictable patterns in the adopted intensity-based feature space. On the other hand, the likelihood  $p(\mathbf{f}|U)$ , that is the probability of observing  $\mathbf{f}$  given that only disturbance factors are acting in the considered neighbourhood, can be modelled due to the well defined and characterizable physical phenomena underlying the image formation process. A learning phase may be necessary, in general, only to estimate the model parameters (e.g. camera noise variance). As a consequence, scene changes can only be detected by a-contrario testing the hypothesis that the observed feature vectors are due to disturbance factors. To this purpose, the possible effects of disturbance factors on a neighbourhood of intensities have to be modelled.

### 3. Modelling of disturbance factors effect

The main disturbance factors yielding changes of pixel intensities over time are:

- d1) imaging process noise;
- d2) adjustments of camera parameters (e.g. auto-exposure, auto-gain);
- d3) illumination changes.

Coherently with (2), let us denote the ideal noiseless feature vector as:

$$\tilde{\mathbf{f}} = (\tilde{X} \parallel \tilde{Y}) = (\tilde{x}_1, \dots, \tilde{x}_N, \tilde{y}_1, \dots, \tilde{y}_N) \in \tilde{F} \quad (4)$$

where  $\tilde{F} = \mathbb{R}^{2N}$  is the (continuous) noiseless feature space. Firstly, we assume that noise affecting pixel intensities is additive, zero-mean, i.i.d. Gaussian with variance  $\sigma_n^2$ . Moreover, we assume that the background is estimated once for all from a quite long initialization sequence (e.g. temporal averaging of tens of frames) so that noise affecting the inferred background intensities can be neglected. In other words, for each pixel the background feature half-vector

$X = \tilde{X}$  is not only constant along time but also deterministic, so that for the on-line statistical background subtraction procedure only noise affecting the current frame intensities has to be considered. Based on the previous assumptions, the probability of sensing the feature vector  $\mathbf{f}$  given its noiseless counterpart  $\tilde{\mathbf{f}}$  is given by:

$$p(\mathbf{f}|\tilde{\mathbf{f}}) = p(Y|\tilde{Y}) = \prod_{i=1}^N \mathcal{N}(y_i|\tilde{y}_i, \sigma_n^2) \quad (5)$$

As far as d2) and d3) are concerned, we assume that noiseless intensities within a neighbourhood of pixels can change due to illumination changes and variations of camera parameters according to the following non-linear parametric transformation:

$$\tilde{y}_i = \alpha \cdot \tilde{x}_i^\gamma + \beta, \quad i = 1, \dots, N \quad (6)$$

It is worth pointing out here that, as mentioned in Section 1, the subset of  $\tilde{F}$  defined by such a transformation contains that defined by the linear models adopted in [3, 6, 9, 11] and is contained into that associated with the order-preserving model of [1, 7, 14, 15]. To the best of our knowledge, no other background subtraction algorithm based on a non-linear parametric model of the local effects of disturbance factors has been proposed in literature.

### 4. Proposed background subtraction algorithm

This Section proposes a novel algorithm for background subtraction based on the non-linear model presented in the previous Section. More precisely, in a Bayesian framework, we try to estimate the most probable configuration of the three parameters  $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$  given the two neighbourhoods  $X = (x_1, \dots, x_N)$ ,  $Y = (y_1, \dots, y_N)$ , relative, respectively, to the background and the current frame:

$$\begin{aligned} (\bar{\alpha}, \bar{\beta}, \bar{\gamma})_{MAP} &= \underset{(\alpha, \beta, \gamma) \in \mathbb{R}^3}{\operatorname{argmax}} p(\alpha, \beta, \gamma|X, Y) = \\ &= \underset{(\alpha, \beta, \gamma) \in \mathbb{R}^3}{\operatorname{argmax}} p(X, Y|\alpha, \beta, \gamma) \cdot p(\alpha, \beta, \gamma) \end{aligned} \quad (7)$$

As highlighted in the right-hand term of (7), this formulation aims at maximizing the *posteriori* probability given a *likelihood* distribution and some *priors* on the three parameters (*Maximum A Posteriori* Bayesian estimation). The use of this *a priori* knowledge is useful to constrain the variability of the parameters that have to be estimated. In particular, assuming the noise model presented in Section 2, the likelihood distribution can be defined as:

$$\begin{aligned} p(X, Y|\alpha, \beta, \gamma) &= \\ &= \frac{1}{(\sqrt{2\pi}\sigma_n)^N} \exp \left( -\frac{\sum_{i=1}^N (y_i - \alpha x_i^\gamma - \beta)^2}{2\sigma_n^2} \right) \end{aligned} \quad (8)$$

For what concerns prior distributions, we assume independence between the parameters to be estimated, that is:

$$p(\alpha, \beta, \gamma) = p(\alpha) \cdot p(\beta) \cdot p(\gamma) \quad (9)$$

and we propose to make an explicit assumption only over the distribution of parameter  $\beta$ . This is motivated by two reasons. The first one is that, without some more specific a priori knowledge on the nature of the disturbance factors that affect images, one can only assume  $\alpha, \gamma \in \mathbb{R}^+$ , since any physically meaningful model for d2) and d3) should take the form of a non-decreasing function. The second reason is that the use of priors on these two parameters would yield to a mathematical problem which is hard to solve. For this reason, as it will be shown in the following, a limitation on their variability will be imposed *a posteriori* as a *hard constraint*, i.e. once the parameter estimation has been performed. On the contrary, we empirically found that the prior on parameter  $\beta$  can be modelled effectively by a zero-mean Gaussian distribution:

$$p(\beta) = \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp\left(-\frac{\beta^2}{2\sigma_\beta^2}\right) \quad (10)$$

Plugging (8) and (10) in (7), and using the log-likelihood formulation, we obtain the following expression for the cost function to be minimized:

$$\begin{aligned} (\bar{\alpha}, \bar{\beta}, \bar{\gamma})_{MAP} &= \\ &= \underset{(\alpha, \beta, \gamma) \in \mathbb{R}^3}{\operatorname{argmin}} \left( \frac{\sum_{i=1}^N (y_i - \alpha x_i^\gamma - \beta)^2}{\sigma_n^2} + \frac{\beta^2}{\sigma_\beta^2} \right) \end{aligned} \quad (11)$$

It is interesting to note that this formulation can be seen as a *Tikhonov Regularization*, with term  $\frac{\beta^2}{\sigma_\beta^2}$  representing a one-dimensional regularization term. The term

$$E(\alpha, \beta, \gamma) = \sum_{i=1}^N (y_i - \alpha x_i^\gamma - \beta)^2 \quad (12)$$

can be seen as the vertical regression error computed between the estimated model fitted on the points and the points themselves. From an algorithmic point of view, the variances  $\sigma_n^2, \sigma_\beta^2$  are two parameters needed to solve the problem. Actually, the problem depends only from their ratio:

$$\phi = \frac{\sigma_n^2}{\sigma_\beta^2} \quad (13)$$

Hence, we can rewrite (11) as:

$$(\bar{\alpha}, \bar{\beta}, \bar{\gamma})_{MAP} = \underset{(\alpha, \beta, \gamma) \in \mathbb{R}^3}{\operatorname{argmin}} (E(\alpha, \beta, \gamma) + \phi \cdot \beta^2) \quad (14)$$

In order to solve this problem, we need to compute the values of parameters  $\alpha, \beta, \gamma$  that minimize (14). While  $\bar{\alpha}$  and  $\bar{\beta}$  can be computed in closed-form, taking the derivative of (14) with respect to  $\gamma$  would yield a transcendent relation requiring an iterative solution method, which unfortunately would hardly be efficient enough given the typical *real-time* requirements of background subtraction applications.

Hence, we propose an approximated, two-pass approach to solve (14). In the first stage,  $\alpha$  and  $\beta$  are optimally estimated assuming  $\gamma = 1$ . This hypothesis relies on the assumption, often verified in practice, that typical non-linear disturbance factors yield values of  $\gamma$  not far from 1. Hence, computation of the partial derivatives of the term to be minimized

$$\begin{aligned} \frac{\partial (E(\alpha, \beta, \gamma = 1) + \phi \cdot \beta^2)}{\partial \alpha} &= 0 \\ \frac{\partial (E(\alpha, \beta, \gamma = 1) + \phi \cdot \beta^2)}{\partial \beta} &= 0 \end{aligned} \quad (15)$$

yields to the following closed-form solution for  $\bar{\beta}$ :

$$\bar{\beta} = \frac{\sum_{i=1}^N x_i y_i \cdot \sum_{i=1}^N x_i - \sum_{i=1}^N y_i \cdot \sum_{i=1}^N x_i^2}{\left(\sum_{i=1}^N x_i\right)^2 - \sum_{i=1}^N x_i^2 \cdot (N + \phi)} \quad (16)$$

and, consequently, for  $\bar{\alpha}$ :

$$\bar{\alpha} = \frac{\sum_{i=1}^N x_i y_i - \bar{\beta} \cdot \sum_{i=1}^N x_i}{\sum_{i=1}^N x_i^2} \quad (17)$$

Then, in the second stage, for each pair  $(x_i, y_i)$  a  $\gamma$  term can be easily estimated as follows:

$$\gamma_i = \log_{x_i} \frac{y_i - \bar{\beta}}{\bar{\alpha}} \quad (18)$$

Given the set  $\Gamma = \{\gamma_1, \dots, \gamma_N\}$ ,  $\bar{\gamma}$  is computed by means of a robust estimator  $G(\cdot)$ :

$$\bar{\gamma} = G(\Gamma) \quad (19)$$

In our experiments presented in the next Section,  $G(\cdot)$  is the median operator.

Once  $\bar{\alpha}, \bar{\beta}, \bar{\gamma}$  have been estimated, the current pixel in the change mask  $M$  is classified as *C (Changed)* or *U (Unchanged)* based on the regression error between the data and the model predictions:

$$M = \begin{cases} C & \text{if } E(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) + \phi \cdot \beta^2 > \tau \vee \bar{\alpha}, \bar{\gamma} < 0 \\ U & \text{otherwise} \end{cases} \quad (20)$$



where  $\tau$  is a fixed threshold and  $\vee$  denotes the logical OR operator. As previously mentioned, at this stage we also deploy fixed constraints on the values of  $\bar{\alpha}$  and  $\bar{\gamma}$ , which were not modelled by means of prior distributions. In particular, all configurations reporting negative values of either  $\bar{\alpha}$  or  $\bar{\gamma}$  denotes a change since they lead to models that can not be induced by the considered disturbance factors.

## 5. Experimental results

This Section presents experimental results aimed at comparing the performance of our proposal to that of two state-of-the-art approaches [9, 14] belonging to the second class outlined in Section 1, namely algorithms that model a priori the effect of disturbance factors over a neighborhood of pixel intensities. Analogously to ours, the considered algorithms, hereinafter referred to respectively as Ohta and Xie, rely on a statistically sound test derived from the a priori disturbance factors model. Recalling again Section 1, each of the two is representative of a different sub-class of robust change detectors, i.e. those invariant, respectively, to linear [9] and to order-preserving [14] transformations.

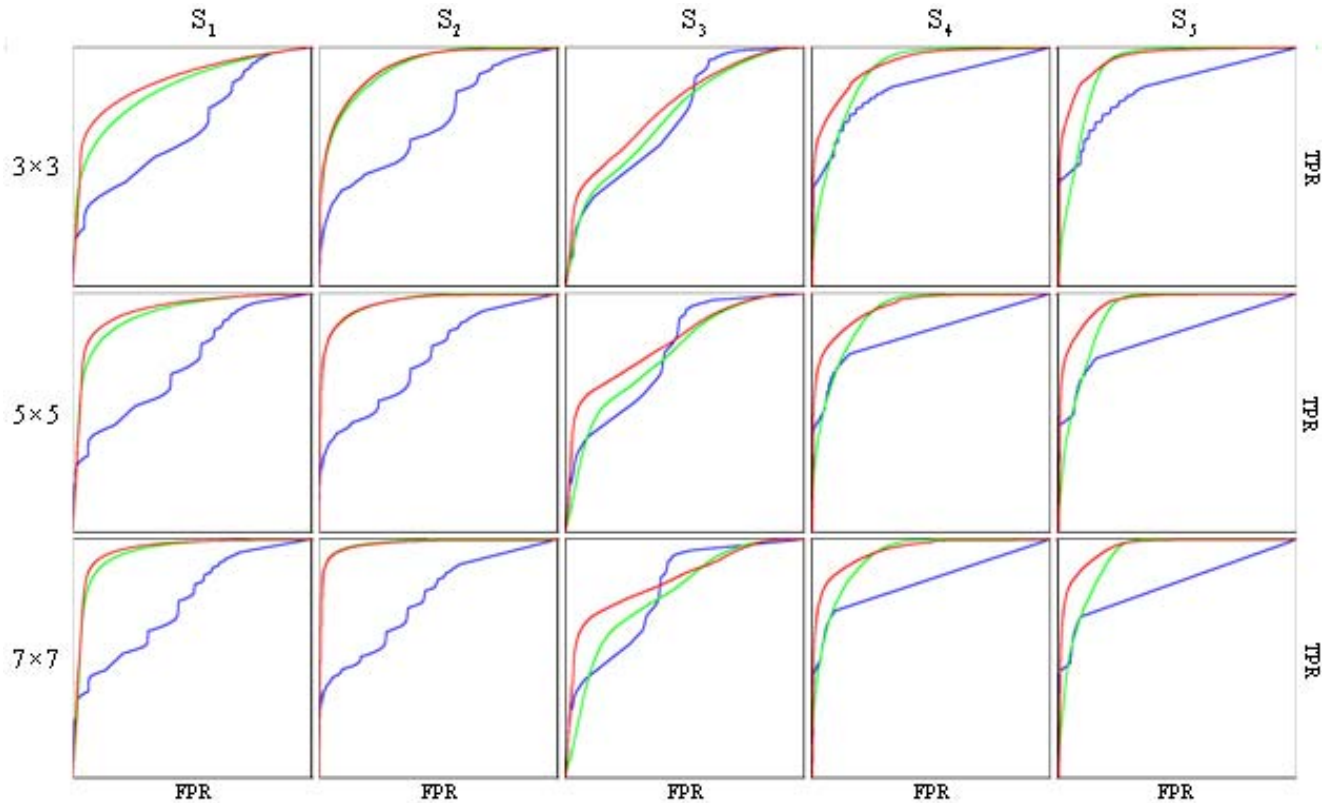
Experimental results have been obtained on five test sequences  $S_1, \dots, S_5$  characterized by sudden and strong photometric changes due to disturbance factors. Some frames of these sequences are shown in the second row of Figure 3. In particular,  $S_1, S_3$  and  $S_4, S_5$  are, respectively, real indoor and outdoor sequences we have acquired by means of different imaging devices, namely a very high quality camera for  $S_1$ , a webcam for  $S_3$ , a standard surveillance camera for  $S_4, S_5$ . In  $S_1$  a person moves around a room illuminated by a single artificial light source and illumination changes are induced by varying the power of the source. In  $S_3$  an object is put into a cluttered indoor scene lit by sun light coming from the outside. Along the sequence, an artificial light source is turned on and the camera gain and exposure vary due to both manual and automatic adjustments. An outdoor parking area is crossed by some people, cars and motorbikes in  $S_4$ , by a crowd of people in  $S_5$ . Apart from some weak illumination changes, photometric variations are induced by manually adjusting the camera gain and exposure. Finally,  $S_2$  is a synthetic sequence available on the web for benchmarking purposes [8] in which moving people as well as photometric changes have been artificially superimposed on an outdoor background. Some compression artifacts are also present, particularly in correspondence of the textured parts of the scene. To allow for quantitative performance evaluation, ground truth images have been obtained, respectively, by manual labelling for  $S_1, S_3, S_4, S_5$ , from [8] for  $S_2$ . Computation of the *comparagrams* [5], i.e. of the joint histograms of background versus frame intensities, for the frames of each sequence highlighted the presence of predominant linear intensity transformations in  $S_1, S_2$ , that of significant

non-linear transformations in  $S_3, S_4, S_5$ . Some sample comparagrams are shown in the bottom row of Figure 3.

For each sequence, the background has been inferred by averaging an initial sequence of frames free of moving objects. Then, algorithms have been run using square neighbourhoods of increasing size ( $3 \times 3, 5 \times 5, 7 \times 7$ ) and quantitative performance measures have been obtained by computing the true positive rate (TPR) versus false positive rate (FPR) receiver operating characteristic (ROC) curve. Figure 2, top, shows the ROC curves obtained for each considered algorithm, sequence and neighbourhood size. Moreover, from each curve we have also extracted a well-known scalar measure of performance, the area under the curve (AUC), which represents the probability for the algorithm to assign a randomly chosen changed pixel a higher change score than a randomly chosen unchanged pixel [2]. This measure is reported in the table shown in Figure 2, bottom, with best performance denoted in boldface.

The ROC curves in Figure 2 show that, overall, the proposed algorithm (red curves) tends to achieve a better trade-off between sensitivity (TPR) and robustness (1-FPR) compared to the other considered algorithms (Ohta shown in green, Xie in blue). In particular, Figure 2 shows that, compared to Xie, our algorithm generally yields more true positives (thus, less false negatives) given the same amount of false positives, with exceptions found only in  $S_3$  and for relatively high FPRs. This is motivated by the fact that both algorithms rely on a non-linear model that includes the intensity transformations yielded by disturbance factors in the considered sequences, but the proposed novel parametric model, being less general, allows for a reduced number of false negatives. In addition, unlike Ohta, the proposed algorithm is capable of handling non-linear intensity transformations. In particular, we can note that our algorithm and Ohta exhibit almost equivalent performances in  $S_1$  and  $S_2$ , where disturbance factors yield linear transformations. Instead, in the other three sequences, which exhibit non-linear transformations, our algorithm clearly outperforms Ohta as shown by the ROC curves and the AUC values.

Finally, some qualitative results are also presented in Figure 3. In particular, for each sequence we show, from top to bottom, the inferred background, one sample frame, the corresponding ground truth mask, the binary change mask yielded by the proposed algorithm and the associated comparagram. The binary change masks are obtained by choosing, for each sequence, the threshold value that yields a TPR equal to 60% and by using a  $7 \times 7$  neighbourhood size. These masks indicate notable robustness of the proposed approach towards false positives that may arise due to the significant intensity changes yielded by the disturbance factors affecting the considered frames, as can be judged by visually comparing corresponding background-frame pairs and also shown by the associated comparagrams.



	AUC			AUC			AUC			AUC			AUC		
	Ohta	Xie	Our	Ohta	Xie	Our	Ohta	Xie	Our	Ohta	Xie	Our	Ohta	Xie	Our
3×3	.830	.664	<b>.861</b>	.914	.689	<b>.918</b>	.717	.712	<b>.756</b>	.899	.834	<b>.927</b>	.918	.827	<b>.952</b>
5×5	.911	.709	<b>.927</b>	<b>.963</b>	.733	.961	.751	.757	<b>.804</b>	.915	.829	<b>.942</b>	.920	.818	<b>.960</b>
7×7	.940	.735	<b>.951</b>	<b>.982</b>	.752	.978	.773	.784	<b>.832</b>	.926	.821	<b>.949</b>	.924	.808	<b>.962</b>

Figure 2. ROC curves (top) and AUC values (bottom) reported by the evaluated algorithms.

## 6. Conclusions

We have proposed a novel Bayesian approach to solve the problem of background subtraction in presence of typical disturbance factors such as sudden illumination changes, variations of camera parameters, noise. On one hand, the use of a non-linear model allows for handling non-linear intensity relations that can not be modelled correctly by methods relying on linear transformations (e.g. [9]). On the other hand, the proposed parametric non-linear model is more restrictive in terms of the transformations allowed by the model compared to order-preserving approaches (e.g. [14]). This yields better sensitivity at comparable levels of robustness. Overall, the proposed approach outperforms state-of-the-art methods on challenging sequences characterized by disturbance factors yielding locally non-linear, as well as linear, transformations of pixel intensities.

## References

- [1] D. N. Bhat and S. K. Nayar. Ordinal measures for image correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(4):415–423, Apr. 1998. **2, 3**
- [2] A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, July 1997. **5**
- [3] E. Durucan and T. Ebrahimi. Change detection and background extraction by linear algebra. *Proc. IEEE*, 89(10):1368–1381, Oct. 2001. **1, 3**
- [4] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *Proc. IEEE Int’l Conf. Computer Vision*, Sept. 1999. **1**
- [5] S. Mann. Comparometric equations with practical applications in quantigraphic image processing. *IEEE Trans. Image Process.*, 9(8):1389–1406, Aug. 2000. **5**

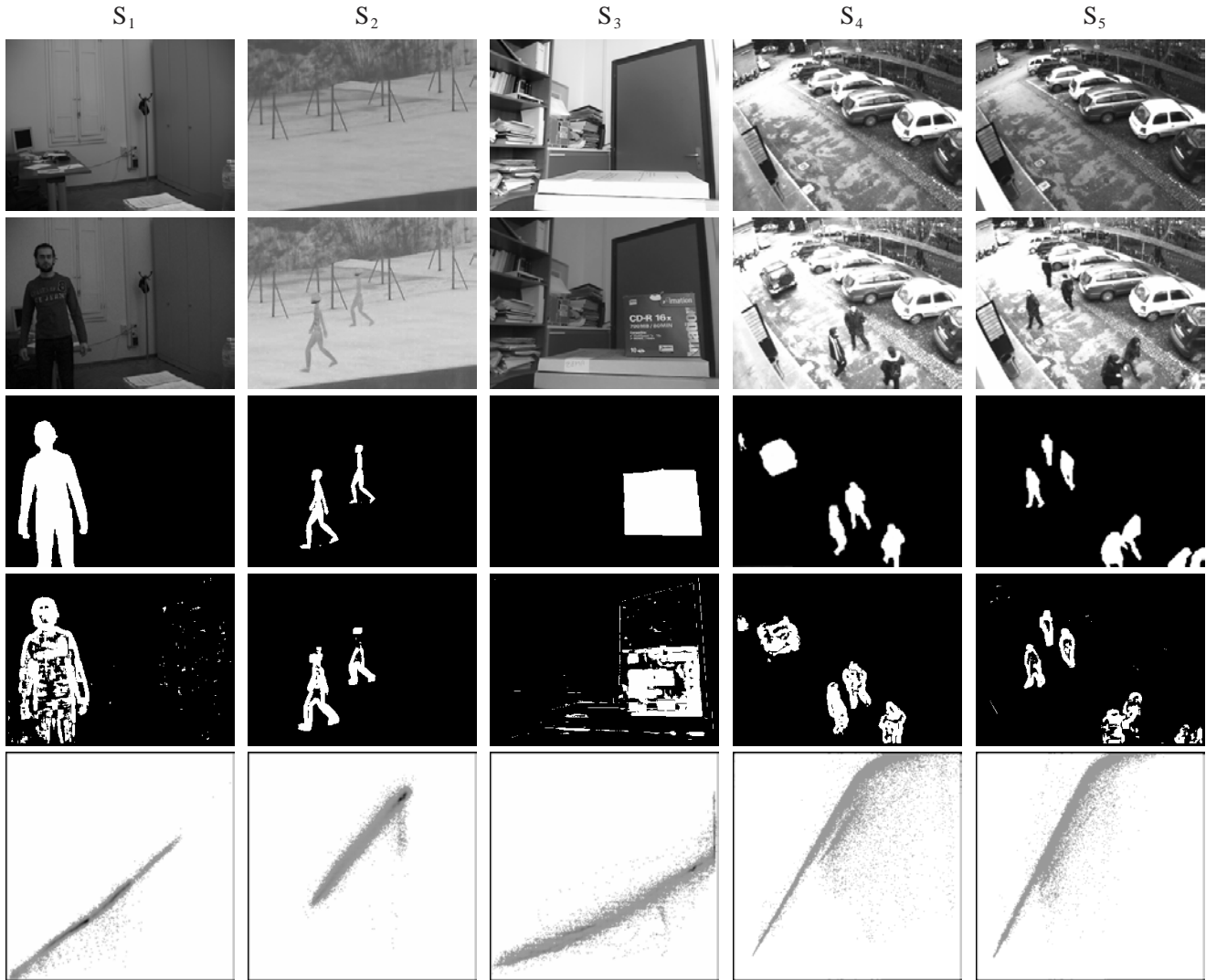


Figure 3. Qualitative results yielded by the proposed algorithm on the 5 test sequences. For each sequence, from top to bottom: the inferred background, one sample frame, the corresponding ground truth mask, the binary change mask and the associated comparison diagram.

- [6] R. Mester, T. Aach, and L. Dumbgen. Illumination-invariant change detection using a statistical colinearity criterion. In *Proc. 23rd DAGM Symposium Pattern Recognition*, pages 170–177, Sept. 2001. [1](#), [2](#), [3](#)
- [7] A. Mittal and V. Ramesh. An intensity-augmented ordinal measure for visual correspondence. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, volume 1, pages 849–856, June 2006. [2](#), [3](#)
- [8] MUSCLE Network of Excellence. Motion detection video sequences. [5](#)
- [9] N. Ohta. A statistical approach to background subtraction for surveillance systems. In *Proc. IEEE Int'l Conf. Computer Vision*, volume 2, pages 481–486, July 2001. [1](#), [3](#), [5](#), [6](#)
- [10] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *IEEE Trans. Image Process.*, 14(3):294–307, Mar. 2005. [1](#)
- [11] K. Skifstad and R. Jain. Illumination independent change detection for real world image sequences. *CVIP*, 46(3):387–399, 1989. [1](#), [3](#)
- [12] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, volume 2, pages 246–252, June 1999. [1](#)
- [13] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfunder: Real-time tracking of the human body. 19(7):780–785, July 1997. [1](#)
- [14] B. Xie, V. Ramesh, and T. Boult. Sudden illumination change detection using order consistency. *Image and Vision Computing*, 22(2):117–125, Feb. 2004. [2](#), [3](#), [5](#), [6](#)
- [15] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. European Conf. Computer Vision*, volume 2, pages 151–158, Sept. 1994. [2](#), [3](#)