False Discovery Rate Approach to Unsupervised Image Change Detection

Vladimir A. Krylov, Gabriele Moser, Sebastiano B. Serpico, Josiane Zerubia

False Discovery Rate Approach to Unsupervised Image Change Detection

Vladimir A. Krylov, Gabriele Moser, *Senior Member, IEEE,* Sebastiano B. Serpico, *Fellow, IEEE,*
and Josiane Zerubia, *Fellow, IEEE*

Abstract

In this paper we address the problem of unsupervised change detection on two or more coregistered images of the same object or scene at several time instants. We propose a novel empirical-Bayesian approach that is based on a false discovery rate formulation for statistical inference on local patch-based samples. This alternative error metric allows to efficiently adjust the family-wise error rate in case of the considered large-scale testing problem. The designed change detector operates in an unsupervised manner under the assumption of the limited amount of changes in the analyzed imagery. The detection is based in the use of various statistical features, which enable the detector to address application-specific detection problems provided an appropriate *ad hoc* feature choice. In particular, we demonstrate the use of the rank-based statistics: Wilcoxon and Cramér-von Mises for image pairs, and multisample Levene statistic for short image sequences. The experiments with remotely sensed radar, dermatological, and still camera surveillance imagery demonstrate accurate performance and flexibility of the proposed method.

Index Terms

Change detection, statistical hypothesis testing, false discovery rate, synthetic aperture radar images, dermatological images, still camera images.

I. INTRODUCTION

Change detection is one of the fundamental image processing problems and multiple detection, monitoring and tracking applications rely on its accurate and timely performance. The central problem of change detection is to process a sequence of images representing the same object or location at different time instants in order to determine image areas that have undergone some *significant* changes. The most common applications include camera surveillance, object tracking, medical follow-ups, monitoring performed on remotely sensed imagery, etc. It

is immediate that in most cases a very particular application-specific kind of changes is sought for. In fact even the considered types of image sequences are very different: from static camera videos with tens of frames per second to two satellite images obtained with a time gap of several months and significant variation in acquisition conditions, like incidence angle and weather effects. When a specific change detection application is considered the first thing to determine is the type of significant changes that is to be detected. This is directly related to the considered class of *disturbance factors*, i.e., the reasons that bring to *non-significant* changes that should be ignored by the detector. These might include acquisition noise, coregistration artefacts, illumination changes, etc. Naturally, there is an appreciable amount of research work addressing various change detection problems [1], [2]: most popular techniques include the background subtraction [3], [4], [5], [6], and local patch approaches based on pixel ordering [7], [8], [9].

In this paper we focus on the classic change detection scenario when changes are evaluated from two or several *coregistered*[1] images obtained at different time instants. The central assumption is that the number of changed pixels should be small relative to the number of unchanged pixels, allowing the algorithm to identify areas exhibiting abnormal statistical behavior as changes. Such a scenario arises when limited changes are sought within limited quantities of data, e.g., in surveillance, remote sensing and medical image processing problems. To address this problem we develop a generic statistical approach, that takes pixel-level detection decisions based on statistical hypothesis tests performed on grouped image samples. Since the construction of representative temporal samples is unfeasible, we consider local patch samples obtained with a square sliding window. The resulting spatial samples are processed with the appropriate test statistics, that are referred hereafter as *features*. The large choice of such features allows to take into account various disturbance factors, and, thus, address particular change scenarios.

The developed empirical-Bayesian approach belongs to the class of large-scale inference problems, where thousands of individual inferences (tests) are performed simultaneously (one per every image pixel). This formulation leads to a difficult problem of adjusting the family-wise error (FWE) [10], [11], which consists in designing a specific multistep procedure to control the amount of the so-called type I errors, or, in other words, false positives. We address this problem with a false discovery rate (FDR) approach [12], [11], a fairly recent statistical method that replaces the classical significance testing by an error metric that controls the proportion of false positives among all detections. This metric is appropriate for testing a large number of hypotheses when the standard statistical testing protocol is too conservative. The FDR-based techniques bring to more powerful testing procedures [12], [13], and more robust to the patch-based sample dependences [14].

The contribution of this paper is in the development of a novel statistically sound FDR-based approach to image change detection. This approach allows to adapt a wide variety of the statistical tests existing in the literature to the problem of patch-based image change detection. We demonstrate the ability of the developed approach to produce accurate application-specific detection results by an appropriate choice of the employed feature. Specifically, in case of image pairs we show the use of Wilcoxon and Cramér-von Mises statistics, that verify the equality of first order moments and the equality of the distributions, respectively. In case of multiple images we propose the use

---

[1]This term is used in broad sense and, depending on the particular dataset, might imply calibration, rescaling, as well as the coregistration in the strict sense, which means the transformation of the input images such that the pixel-to-pixel level correspondence is established.

of the Levene multisample statistic, that tests the hypothesis of the equality of the sample variances. For validation purposes we consider (*i*) remotely sensed synthetic aperture radar (SAR) imagery, where the main disturbance factor is given by the speckle noise, (*ii*) dermatological images with approximate coregistration and varying illumination conditions, and (*iii*) still camera images with disturbance factors of varying illumination and shadows. A preliminary short version of this study was presented in [15].

The paper is organized as follows: In Section II we briefly review relevant previous work on statistical image change detection. In Section III we formulate the FDR-approach and propose the approaches to null distribution and empirical density estimation. In Section IV we propose a novel feature-based change detector. In Section V we present several statistical features for two and multiple images scenarios. We report the experimental results on three types of image pairs in Section VI, and a semi-synthetic experiment on an image sequence in Section VII. The paper is concluded in Section VIII.

II. RELATED WORK

For a comprehensive overview of change detection methods we refer the reader to surveys performed in [1], [2], [5], [6], [16]. The brief selection of methods presented in this section is intended to give some insight into the recent developments in statistical image change detection. Note that we leave aside such wide areas of change detection techniques as segmentation and object-based detection, machine learning methods [2].

We first present several relevant FDR techniques that have been employed in the image processing applications. In [17] the FDR-analysis is performed on single-band astronomical images to identify specific (source) areas based on the p-values estimated via an *a priori* assumption of standard normal distribution for pixel intensities. In [18] the processing is done on long temporal sequences (100 frames) from surveillance cameras with the null assumption of normality of temporal samples at each location. In [19] and [20] the authors apply FDR-techniques to the MRI/fMRI imagery to identify brain areas (voxels) reacting to visual stimuli and subject to the ageing effect, respectively. In the former, the FDR-testing is performed on the iteratively thresholded output of the wavelet-based statistical parametric mapping (with the underlying normalized Student's test) of the input. In [20] the authors applied FDR-inference to the z-test p-values to identify the departures from zero of the voxel means (under normality assumption) in the simulated study, and to the Wald-like statistic on the coefficient of the age factor in the linear regression model of the voxel intensities in case of the real data. The authors have also proposed to integrate a second thresholding rule based on an isotropic MRF term to promote spatial regularity of the detection. In [21] FDR is used to find video volumes similar to a given video query by resorting to the trace-based distances with the corresponding p-values recovered from the Student's t-statistic.

We now review several general classes of change detection methods. The first wide class is given by background subtraction approaches [3], [7], [4], [5], [6], [16]. These methods analyze the temporal consistency of image pixels in large image sequences. The second large class of methods rely on the spatial support by processing the local information at two time instants. Typically, the order statistics are used due to the fact that pixel orderings are invariant to monotonic transforms of the data, and also allow more generality with respect to different noise models. The disturbance factors effect has been modelled as affine [7] or more general order-preserving transformations [8],

[9]. The central disturbance factors in these papers are the illumination changes and camera adjustments. The approach developed in [8] tests the rank-consistency under an explicit noise model. Some contour-based change detectors have been developed [22], [23]. Although capable of operating on only two images (of which one is considered as a background), these methods require estimates (variances) of the acquisition/camera noise, that are typically obtained from available stack of previously acquired frames.

In this paper we focus on a scenario when only several images are available for the change detection analysis. This represents a different problem for which many of the background-based methods, e.g., [3], [5], [6], [16], cannot be applied. Some methods, like [7], [24], [9], can be employed in two frames case given the noise variance estimates. For comparisons we will use methods proposed by Lanza et al. [9], that verifies local order consistency and formulates detection as a maximum likelihood nonparametric isotonic regression problem, and Xie et. al. [24], that relies on an explicit statistical imaging model and likelihood ratio test of order consistency.

Considering SAR change detection problem addressed in the experimental validation, most of the existing methods build on explicit parametric models [25]. Typically, due to the multiplicative nature of the speckle, SAR change detection methods rely on image ratio modeling, e.g., generalized Gaussian [26], or using lognormal- or Nakagami-ratio distributions [25]. The detection maps are typically obtained via minimum-error thresholding algorithms [25], [26]. Further methods applied to remote sensing data are based on fuzzy clustering [27], likelihood ratio testing [28], stochastic geometry [29], and multi-layer Markovian models [30]. For comparisons we will employ explicit statistical models developed by Moser et al. [25], that is based on lognormal-ratio modeling, and by Inglada and Mercier [31], that builds on thresholding of cumulant-based Kullback-Leibler distance between local patch-based samples and can be applied to arbitrary types of imagery.

III. FALSE DISCOVERY RATE approach

When addressing the change detection problem as a set of pixel-based decisions on a stack of images, we face a so-called *large-scale simultaneous hypothesis testing* problem [12], [11]. This type of problems is characterized by thousands of simultaneous tests that are performed on data samples of the same nature. Operating on two coregistered images, we calculate a value of a selected statistical feature $\mathfrak{F}$ at each pixel location. In this paper we employ the statistic features, for which distributions $D_{\mathfrak{F}}$ are known in a no-change scenario under the classical sample independence assumption. Working with local patch samples, we encounter the problem of dependence in the values of collected features, at least, at neighbouring locations.

Furthermore, the correction to the individual test significance level in a multi-test is a sophisticated matter [12]. Specifically, it is well known that when addressing a multiple inference problem, an adjustment needs to be made to preserve the significance level of the resulting test. The classical remedy given by the Bonferroni correction [10] consists in setting a new significance level of each individual test: If it is desired that the significance level of the group-test with n hypotheses should be α, then the corrected individual test's significance level should be set to α/n. Naturally, such a level of significance is overly conservative and impossible to obtain in practice (considering $n \geqslant 10^6$ pixels), which inevitably results in zero-detection. Other common methods imposing FWE-control include Tukey-Cramer method, step-down procedure, Games-Howell method, see [32] for more detail. Another common

technique to obtain empirical p-values, given by non-parametric bootstrap [10], cannot be employed here since the total amount of tests (equal to the image size) is far greater than the size of the samples employed for each test (local patch size), see [11].

Due to the above reasons the direct use of the p-values obtained feature distribution $D_{\mathfrak{F}}$ is unfeasible. To address the problem, we suggest to capitalize on the knowledge of a significant amount of feature observations– one per each pixel of the considered image stack– to learn more about the actual $D_{\mathfrak{F}}$ distribution. This becomes possible since we assume that the majority of at least 75% of the pixels have not undergone any significant change. To learn from this large set of non-changed data we adopt the local FDR approach [12], [11] to image change detection.

We begin by introducing the notion of z-scores. This allows the approach to be formulated in a unified way regardless of the employed form of $D_{\mathfrak{F}}$ and assume that the transformed feature values (z-scores) are normally distributed. Let $D_{\mathfrak{F}}$ denote the cumulative distribution function (CDF) of the feature $\mathfrak{F}$, then we can define a z-score of the i-th image pixel as:

$$z_i = \Phi^{-1}(D_{\mathfrak{F}}(\mathfrak{f}_i)), \tag{1}$$

where Φ^{-1} is the inverse of a standard normal CDF, and $\mathfrak{f}_i$ are the observed values of the feature $\mathfrak{F}$ (centered in i-th pixel). Note that if $D_{\mathfrak{F}}$ is non-parametric then the calculation of z-scores can be performed directly from the observed features as defined in (1) without resorting to any additional parameter estimation procedure. Indeed, most of the standard hypothesis testing procedures [10] lead to non-parametric (asymptotic) distributions for their statistics. In this paper we consider only non-parametric $D_{\mathfrak{F}}$. It is immediate that in case of a normal distributed feature the z-scores are identical to the feature values.

We now formulate the testing problem in a Bayesian form. Suppose that the values of the feature statistic belong to two classes corresponding to unchanged (we will denote it the *null* hypothesis, $H_0^{\mathfrak{F}}$) and changed (alternative $H_1^{\mathfrak{F}}$), with prior probabilities p_0 and $1 - p_0$. We define the considered feature densities $f_0(z)$ and $f_1(z)$ depending on the class:

$$f_0(z) \text{ density if } \textit{unchanged } (H_0^{\mathfrak{F}}), \text{with probability } p_0,$$

$$f_1(z) \text{ density if } \textit{changed } (H_1^{\mathfrak{F}}), \text{with probability } 1 - p_0.$$

The complete feature density is, then, a mixture $f(z) = p_0 f_0(z) + (1 - p_0)f_1(z)$. Thus, according to the Bayes theorem we obtain the a posteriori probability

$$\text{Prob}\{\textit{unchanged} \mid z\} = p_0 f_0(z)/f(z). \tag{2}$$

The local false discovery rate is defined as:

$$l\text{FDR}(z) \equiv f_0(z)/f(z). \tag{3}$$

By ignoring the factor p_0 in (2), $l\text{FDR}(z)$ can be considered as an upper bound for the probability of observing the unchanged class for an observation with a given value of z-score. The standard $\text{FDR}(z)$ is a Bayesian-based formulation of a frequentist FDR quantity, and is expressed as an expectation of $l\text{FDR}(z)$ conditioned on $Z \leqslant z$ [11]. The local FDR is the probability that a particular test gives rise to a false positive. The advantage of $l\text{FDR}(z)$-control is that the

independence is not required of the z-scores, all that is needed is a reasonable estimate of their marginal distribution. In accordance with the considered change detection problem (assumption of relatively small amount of changes), we can assume that p_0 is sufficiently close to 1 (e.g., $p_0 = 0.8$ or $p_0 = 0.9$), so that $l\text{FDR}(z)$ is a slightly conservatively biased estimator [11] of (2). The FDR-based decisions in the following are taken by performing the thresholding and reporting z-scores with $l\text{FDR}(z) \leqslant \gamma$ as changed. This means that we label as changed the pixel locations that are the centres of local patches employed to produce the z-scores that are identified by the inequality above.

At this point it is important to underline the difference between a FDR-approach and more classical FWE-procedures. Both are designed to control the rejection of a true $H_0^{\widetilde{\mathfrak{F}}}$ hypothesis probability, however, whereas the FWE-methods control this quantity directly, the FDR-methods control the proportion of false positives among all rejected *null* hypotheses (discoveries). The advantage of the FDR-control is that it leads to fewer type II errors (false negatives) than controlling the FWE at the same level [12]. We stress furthermore that the proposed FDR-based procedure is different from the conventional thresholding (possibly with different thresholds applied at two tails) of the histogram of z-scores which constitutes the standard statistical test. Specifically, the thresholding of the FDR-function defined in (3) allows to consider pixels corresponding to some local z-value peak as changed, and at the same time leave less frequent z-values around in the unchanged class. From the statistical point of view that would mean that the z-values in the peak have been observed too often to consider them as following the *null* distribution. Thus, this particular z-value and its neighbourhood are considered characteristic of one of the change scenarios.

A. Null distribution estimation

We proceed to the empirical $\hat{D}_0^{\widetilde{\mathfrak{F}}}$ *null* distribution estimation based on the observed z-scores. This strategy allows accommodating the theoretical $D_0^{\widetilde{\mathfrak{F}}}$ distribution to the image patch samples, thus, addressing the dependence problems.

The most sensitive part in the considered methodological approach falls on the choice of the empirical *null* distribution $f_0(z)$ for the unchanged z-scores. If chosen perfectly, i.e., so that all the unchanged scores have $f_0(z) > 0$ and $f_1(z) = 0$ and the changed z-scores satisfy $f_0(z) = 0$ and $f_1(z) > 0$, then $\hat{p}_0$ is an unbiased estimate of the true percentage of the unchanged observations [11]. Unfortunately such a choice is impossible in practice and, therefore, an empirical procedure needs to be established to estimate the *null* distribution $f_0(z)$.

As has been observed in typical applicative problems [11], under the *null* hypothesis, the z-scores demonstrate a pronounced normal behavior even in cases, when substantial dependences are present in the data. Nevertheless, the classical standard $\mathcal{N}(0,1)$ does not reflect accurately the statistical behaviour of z-scores at *null* even in conventionally dependence-free large scale inference problems [11]. Therefore, in this paper we propose to consider the normal distribution family for *null* distribution, but allow the distribution parameters (mean and variance) to deviate from the standard $\mathcal{N}(0,1)$ configuration in order to better reflect the dependence structure present in the data.

To arrive at the estimate of the *null* distribution that is affected by the changed pixels as little as possible we perform its estimation on the z-scores that lie around its distribution center and assume that the changed pixels report z-scores which lie far from the observed distribution center. In other words, the zero assumption is that the non-*null* density $f_1(z)$ is zero on a continuous interval containing the origin $z^* = 0$, which renders the mixture model

identifiable from the central histogram counts. This assumption holds if the chosen statistical feature is appropriate for the considered change detection problem, i.e., if it corresponds well to the significant changes. We employ the central matching approach [11] by assuming that near the central score z^* the representation $f(z) = p_0 f_0(z)$ holds, and that $f_0(z)$ follows the normal distribution $\mathcal{N}(\delta_0, \sigma_0^2)$. Under independence assumption z^* should be set to zero, in practice we set it equal to the median of z-values. By taking the logarithm of $f(z)$, we obtain

$$\log f(z) = -\frac{1}{2\sigma_0^2} z^2 + \frac{\delta_0}{\sigma_0^2} z + \left[\log p_0 - \frac{\delta_0^2}{2\sigma_0^2} - \frac{1}{2} \log(2\pi\sigma_0^2) \right].$$

The central matching strategy consists in estimating $f_0(z)$ by assuming that $\log f(z)$ is quadratic near z^*, and, therefore, can be presented as:

$$\log f(z) = \beta_2 z^2 + \beta_1 z + \beta_0. \tag{4}$$

We can thus obtain the estimates

$$(\hat{\delta}_0, \hat{\sigma}_0^2) = \left(-\frac{\beta_1}{2\beta_2}, -\frac{1}{2\beta_2} \right). \tag{5}$$

Using (5) we arrive at an estimate of the empirical null $f_0(z)$ as $\mathcal{N}(\hat{\delta}_0, \hat{\sigma}_0^2)$. A critical issue is the choice of the z-scores near z^*, that are used to estimate (β_1, β_2), and correspond only to unchanged observations. In this paper, we employ a conservative strategy by taking $\Delta = 50\%$ of the data for the fitting of (4). To define the interval of the feature histogram counts that are employed for $f_0(z)$ estimation, we adopt the following approach. We start by uniformly partitioning the range of z-values into K bins (intervals):

$$\mathcal{Z} = \bigcup_{k=1}^{K} \mathcal{Z}_k. \tag{6}$$

The bins are of equal widths, the number of z-scores in every bin is $y_k = \#\{z_i \in \mathcal{Z}_k\}$, and x_k are the centerpoints of $\mathcal{Z}_k$.

In order to construct the set of bins $\mathcal{Z}_{null}$ used for the *null* estimation we first initialize it with the bin that contains the median z^*. We then proceed by iteratively adding the adjacent bin (to the left or right from $\mathcal{Z}_{null}$) that has the higher y_k to the current set. This is iterated until the portion of scores inside the selected bins reaches 50% of the total:

$$\sum_{\mathcal{Z}_k \in \mathcal{Z}_{null}} y_k \geqslant 0.5 \sum_{k=1}^{K} y_k.$$

Then, the set of points (x_k, y_k) characterizing the bins in $\mathcal{Z}_{null}$ is fitted by the right hand side of (4) using a least square error approach [11]. Finally, we get the estimates $(\hat{\delta}_0, \hat{\sigma}_0^2)$ as defined in (5). A typical example of such an empirical fit with $f(z)$, $f_0(z)$ and $\mathcal{N}(0, 1)$ is presented in Fig. 1.

Note that the $f_0(z)$ estimates based on more than $\Delta = 50\%$ of observations might be more accurate, at the expense of potentially getting biased from the inclusion of changed z-scores, and, therefore, should not be considered as a default option. According to this strategy we define the above mentioned fraction of the changed pixels as limited by 25% of all pixels. This typically ensures that the central part of the appropriate feature histogram is well separated from the changed z-scores in the tail parts, and, thus, the change detection can be performed in an unsupervised manner. Furthermore, this assumption ensures the consistency of the obtained FDR-estimates under some technical assumptions. In particular, the Gaussianity of the central z-scores provides consistency of the least square estimate
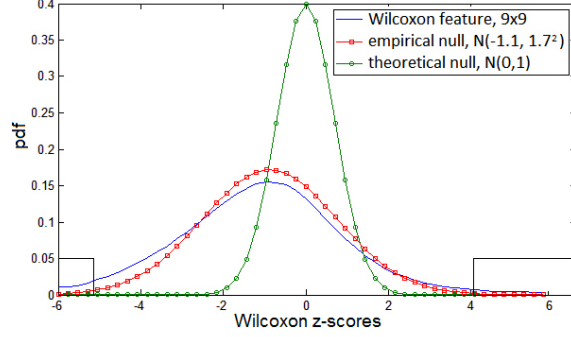
Fig. 1. The normalized histogram and its estimates for the Wilcoxon feature on the XSAR image pair. The z-scores in boxes correspond to the detection areas with FDR $\leqslant 0.1$.

and the continuity of the exp-transform preserves this property. The consequent demonstration of the consistency for the FDR-estimates follows the proof of Theorem 1 in [20].

Alternative estimates can be obtained by maximizing the pseudo-likelihood (PL) of $f(z)$, see in [11], [20]. This procedure allows to obtain robust estimates and exploits the whole set of z-scores. Computationally, the Expectation-Maximization-like procedure necessary for the PL evaluation is more demanding. Experimentally, we have observed similar performance of central matching and PL, with the exception of datasets featuring more changes (up to 25%), where PL reported slightly more biased estimates of p_0. For these two reasons we present the central matching procedure in this paper.

B. Empirical density estimation

To evaluate the FDR as defined in (3) we require a smooth estimate of the density $f(z)$ for the observed values $\mathfrak{f}_i$. One way to address this problem is by resorting to classical interpolation tools such as natural splines [11]. However, such tools do not take into account the normalization and non-negativity restrictions imposed on probability density functions. In this work, instead, we employ the approximation by a J-parametric exponential density function:

$$f(z) = \exp\left\{\sum_{j=0}^{J} \alpha_j z^j\right\}, \tag{7}$$

where the constant α_0 is required to ensure the normalization. The choice $J = 2$ makes (7) into the normal family. If J is large the estimate becomes almost non-parametric and tries to fit every z-value. A choice of $J^* = 7$ in (7) allows to achieve smooth estimates and does not impose too many restrictions on the density function shape.

To estimate the model parameters it is recommended to use the *Lindsey's method* that allows to obtain maximum likelihood estimates using a textbook Poisson regression. This involves the binning of the data as in (6), and builds on the multinomial representation of the binned exponential statistics in (7). As such, the problem is conveniently reduced to a standard Poisson generalized linear model as described in [11].

Algorithm 1: FDR-based change detector

---

1 calculate feature $\mathfrak{f}(i,j)$ for all pixels (i,j);

2 transform $\mathfrak{f}(i,j)$ into scores $z(i,j)$ as in (1);

3 estimate the *null* density $f_0(z)$, see Sec. III-A;

4 approximate the empirical density $f(z)$, see Sec. III-B;

5 calculate $\text{FDR}(z)$ as in (3);

6 report as *changed* the pixels with $\text{FDR}(z) < \gamma$.

IV. IMAGE CHANGE DETECTION

In this section we formulate a novel feature-based change detector that operates on two or more images. We first assume that the significant changes are present in a fraction of the images' pixels that is at most 25% of their total size. We also recall that to employ the FDR-approach we can only use the statistical features with a known distribution under a certain *null* no-change hypothesis. Typically, any statistic that is used to construct a statistical hypothesis test can be employed as a feature in the designed change detector.

In order to accommodate the sample-based approach to construct the pixel-based detection map we employ the local patches. Specifically, for a pixel at location (i,j), we consider the sample obtained from the values of the surrounding pixels in a $S \times S$ local window centered at (i,j). It is immediate that any sample constructed by taking pixel values in a local-window of a natural image cannot be considered independent in itself, because of the contextual dependences associated with the image (spatial dependence or *intra-sample dependence*). Moreover, we consider the change detection problem, in which we assume that the amount of significant changes cannot be large. As such, the two patch samples centered at pixel (i,j) on two coregistered images X and Y cannot be considered independent, i.e., for all $i = 1, \ldots, N$ observations at the same positions x_i and y_i are likely to be dependent (temporal dependence or *inter-sample dependence*). Therefore, when considering the two-sample hypothesis tests on local patch samples we have to deal with two kinds of dependences in samples.

For the problem of image change detection we propose to use the patch-based formulation of the FDR-based technique presented in Section III. The resulting empirical-Bayesian change detector allows to address: (*i*) the inference based on correlated individual tests (due to overlap of patch-based samples), and (*ii*) intra-sample dependence of patches. As has been analysed in [11], the FDR-formulation is robust to a certain level of individual tests' correlation for multiple inference. Considering the intra-sample dependence, we will assume that its level is moderate, so that the departures from statistic's distribution under the *null* hypothesis do not bring to appreciable non-normal behavior for the z-scores. Note that it is a common assumption for various applied problems, and that the normal statistical behavior of z-scores is commonly observed [11]. Also note that the majority of existing change detection approaches ignore these dependence issues by working under various independence assumptions, see [7], [8], [9], [24], [25], [31].

The outline of the FDR-based change detector is presented in Algorithm 1. The choice of γ for the FDR-based procedure is similar to that of the significance level in classical inference. Finally, the local patch size S should

be chosen manually with respect to the scale at which the significant changes manifest, which constitutes the same strategy as in [7], [8], [9], [24]. The computational complexity of the proposed algorithm is $O(S^2 A \log S)$: linear with respect to the number of the input pixels A and $S^2 \log S^2$ - to the patch size S (due to the use of order statistics that rely on sample ordering with merge / block sorting). In this respect the algorithm is similar to [9] and [24], and faster than that in [31].

V. STATISTICAL FEATURES

The developed change detector constitutes a general approach that can be adopted for specific applications by an appropriate choice of statistical features. These should be the multisample statistics with a known distribution under the no-change hypothesis.

In this paper we focus on the use of non-parametric statistics $\mathfrak{F}$ that are aimed at change detection from multitemporal images and allow a sufficient level of generality without restrictive distribution assumptions. Such generality is important when a single change detector is employed to reveal changes on imagery affected by several distinct disturbance factors. In particular, encouraged by the positive results obtained in [8], [9], [24] we employ the order (rank) based statistics in the role of features. The rank transform is widely used in statistics to convert parametric tests into non-parametric by replacing the observations by their ranks [32], [10]. The strict specific distribution assumptions in such cases are replaced by a very mild assumption of the similar distribution form. Note, however, that the proposed technique allows the use of parametric statistics as well when there is a strong evidence that the analysed data follow a particular statistical model.

We consider the use of several features to address disturbance factors given by coregistration artefacts, acquisition noise and illumination variation. Namely, we use (*i*) the two-sample Wilcoxon paired statistic [10], that is decisive with respect to the first order sample changes, (*ii*) the two-sample Cramér-von Mises statistic [10] for arbitrary type of distribution changes, and (*iii*) the Levene statistic on the rank-transformed multisample data [10], [33] that detects sample variance changes. These features are common in various applications thanks to their simplicity and robustness [10].

A. Two-sample Wilcoxon statistic

Given two independent random variables $\mathcal{X}$ and $\mathcal{Y}$, the paired Wilcoxon statistic [10] verifies the *null* hypothesis $\{H_0^W : P(\mathcal{X} < \mathcal{Y}) = 1/2\}$, which characterizes the comparative central symmetry of two observed populations. The information given by this statistical feature characterizes the first order moment changes [10].

The two-sample paired Wilcoxon statistic is based on the observed absolute differences $L_i = |X_i - Y_i|$, where X_i and Y_i are the observed samples of $\mathcal{X}$ and $\mathcal{Y}$, respectively, $i = 1, \ldots, N$. In the employed sliding window-approach $N = S^2$, where S is the local patch size. The values of L_i are transformed into the order statistic $L_{(1)} < \cdots < L_{(N)}$. Then the ranks of $L_{(i)}$ that correspond exclusively to the positive differences $X_i - Y_i$ are summed to get W_+. The standardized Wilcoxon statistic is defined as:

$$W = \left(W^+ - \frac{N(N+1)}{4}\right)\left(\frac{N(N+1)(2N+1)}{24}\right)^{-0.5} \tag{8}$$

This statistic converges in distribution to the standard normal $\mathcal{N}(0,1)$ [10] when $N \to \infty$ under H_0^W with the independence assumptions. Thus, for large samples, the critical values for the Wilcoxon test can be drawn from the normal approximation, that is classically employed when $N > 20$ [10]. The rejection region of the test is two-sided, i.e., H_0^W is rejected when $|W| > q$, where q is a relevant quantile of the standard normal distribution. Since $W \sim N(0,1)$, the statistic values $\mathfrak{f} = W$ can be considered directly as z-scores, following the definition in (1).

The paired Wilcoxon feature formulation (considering of differences $|X_i - Y_i|$ instead of X_i and Y_i values) allows to compensate the inter-sample (temporal) dependence problem. The considered hypotheses H_0^W and H_1^W render the feature sensitive to the illumination changes (linear shifts), which, however, can be an important source of significant changes, e.g., when the central disturbance factor is given by zero-mean additive noise.

B. Two-sample Cramér-von Mises statistic

The two-sample Cramér-von Mises (CvM) statistic [10] is designed to verify a general *null* hypothesis: $\{H_0^{\text{CvM}} : D_{\mathcal{X}} \equiv D_{\mathcal{Y}}\}$, where $D_{\mathcal{X}}$ and $D_{\mathcal{Y}}$ are the CDFs of the random variables $\mathcal{X}$ and $\mathcal{Y}$ observed by samples X and Y of size N, respectively. The CvM test can detect any kind of distribution deviations given sufficient sample sizes. Thus, compared to the Wilcoxon feature, the CvM feature can detect more complex changes like fluctuations in higher order moments. Note that similar results can be obtained with the other general two-sample goodness-of-fit features, e.g. Kolmogorov-Smirnov and Anderson-Darling statistics [10].

To calculate the two-sample CvM statistic, the observed samples X_i and Y_j are initially ordered into $X_{(i)} : X_{(1)} < \cdots < X_{(N)}$ and $Y_{(j)} : Y_{(1)} < \cdots < Y_{(N)}$. The ordered samples are concatenated into the sample L of size $2N$, which is then ordered as well to obtain $L_{(1)} < \cdots < L_{(2N)}$. Then,

$$T = \frac{\sum_{i=1}^{N}\left[(r(i) - i)^2 + (s(i) - i)^2 \right]}{2N^2} - \frac{4N^2 - 1}{12N} \tag{9}$$

defines the standardized CvM statistic, where $r(i)$, $s(i)$, $i = 1, \ldots, N$, are the observations' ranks in the ordered samples: $L_{(r(i))} = X_{(i)}$, and $L_{(s(i))} = Y_{(i)}$. T follows asymptotically the Cramér-von Mises distribution [10], [34]. However, the expression for the asymptotic CDF requires numerical approximations (since it involves the modified Bessel function of the third kind [35]), and the convergence of the distribution (as N increases) is relatively slow, especially so in the tail areas [34]. For these reasons we will employ the exact CDFs for the sample size N as evaluated in [34]. The rejection region of the CvM test is right sided. The z-scores can be estimated as defined in (1), and $\mathfrak{f} = T$.

Along with CvM, we propose the use of a modified CvM (mCvM) feature. It is constructed by preliminary subtracting sample medians from both samples before constructing T, see eq. (9). This allows to tune the statistic to fluctuations of higher order moments, ignoring the first moment changes, for example, constant illumination shifts. The distribution and critical region are the same as for CvM. Various further statistical features can be employed for detecting changes of higher order moments [10].

The non-paired CvM and mCvM construction allows to address the coregistration disturbance factor, since T is insensitive to the displacements of observations within large patches. The mCvM feature can be employed to address the illumination changes disturbance factor.

C. Multisample Levene statistic

In the case of change detection from a sequence of K images, consider K random variables $\mathcal{X}_1, \ldots, \mathcal{X}_K$ modeling the intensities in the same patch in K images. These variables are observed by samples of size N each: $(X_{k1}, \ldots, X_{kN}) \sim \mathcal{X}_k$, $k = 1, \ldots, K$. In this scenario there are two conceptual ways to validate the hypothesis that a certain characteristic of the samples is unchanged. The first consists in running an *omnibus* test that verifies the equality of all K sample characteristics. The term omnibus relates to the test formulation admitting multiple alternatives. Specifically, the null assumption of equality can be violated by any pair of samples which, when formulated explicitely, is equivalent to multiple distinct alternative hypotheses. The omnibus testing procedure is relatively fast, but in case when the null hypothesis is rejected the omnibus test does not specify in which pair(s) the changes occur. The second way is more computationally intensive and requires running all the pairwise tests. These two options will be explored in this section in case of the Levene statistic evaluating the hypothesis of equal population variances. Note, that this statistic can be readily modified to test the equivalence of the sample means [32], and allows further adjustments to target higher order statistical characteristics [36].

1) Omnibus Levene statistic: The null hypothesis is that the population variances σ_k^2 are not changing: $\{H_0^{\text{Levene}} : \sigma_1^2 = \ldots = \sigma_K^2\}$. The omnibus alternative stipulates that there exist two distinct populations $\mathcal{X}_u$ and $\mathcal{X}_v$ such that $\sigma_u^2 \neq \sigma_v^2$. Note, that this test can also be employed in the two-sample scenario. The Levene test [32], [37] is formulated as a repeated measures analysis of variance (ANOVA) test on the absolute deviation of the scores X_{kn} from their corresponding sample medians $\bar{X}_k$: Levene $\equiv$ ANOVA$\left(|X_{kn} - \bar{X}_k|\right)$.

The classical formulation of the Levene test assumes the normality of the variables under scrutiny. One solution to relax this tight condition and tune the test to the non-parametric scenario exlored in this paper is provided by the concept of rank transformation [10]. This approach consists in performing the test on ranks rather than on the initial observations themselves, i.e.:

$$\text{ANOVA}\left(|R_{kn} - \bar{X}_k^R|\right), \tag{10}$$

with R_{kn} denoting the rank of X_{kn} in the pooled sample constructed by merging obervations from $\mathcal{X}_1, \ldots, \mathcal{X}_K$ with KN observations. $\bar{R}_k$ is the median rank in the k-th sample. Whenever ties occur, the average ranks are used. Such a strategy constitutes a statistically valid and efficient testing procedure for multisample problems [37], [38].[2]

As an omnibus test, (10) has slightly less power than single-alternative tests and is more sensitive to outliers [32]. Therefore, to reduce the impact of noise and outliers on the detection results, we will take measures to render the test more robust by clipping $[\omega N]$ observations in each of the input rank samples. Here ω is a predefined clipping percentage which is typically set to small values like 0.05, 0.1, see in [32], and $[x]$ denotes the integer part of x.

The Levene statistic for the dependent samples is given by

$$F = \frac{Q_c/(K-1)}{Q_e/\left((N - [\omega N] - 1)(K-1)\right)}, \tag{11}$$

---

[2]In particular, compared to the textbook non-parametric Friedman test [10], which is the multisample counterpart the sign test, Levene test has the following advantages: (*i*) higher values of the asymptotic relative power [38], and (*ii*) more sensitive *p*-values due to the ranking procedure performed on KN observations, and, thus, $(KN)!$ possible combinations, which is substantially larger number than $(N)!^K$ combinations available for the Friedman test.

Algorithm 2: Pairwise post-hoc testing

---

1 perform $C = K(K-1)/2$ pairwise Levene tests to obtain the corresponding p-values $p_1, \ldots, p_C$;

2 order the results so that $p_{(1)} \geqslant \ldots \geqslant p_{(C)}$;

3 set $k = 0$;

4 increment k by 1;

5 **if** $k > C$, **stop**;

6 **if** $p_{(k)} \leqslant \gamma(C - k + 1)/C$, reject hypotheses reporting $p_{(k)}, \ldots, p_{(C)}$ and **stop**, **else goto** step 4.

(a) XSAR image 1   (b) XSAR image 2

(c) Ln-GKIT [25]   (d) thresholded KL [31], 9×9

(e) FDR-based CvM, 9×9 (f) FDR-based Wilcoxon, 13×13
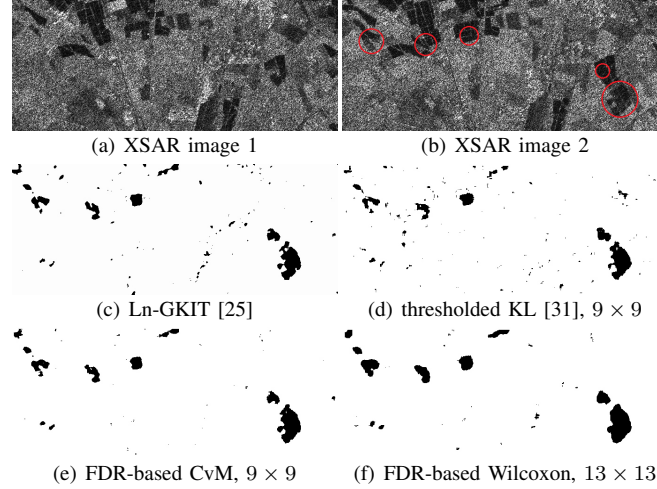
Fig. 2. (a),(b) 700×300 pixel XSAR image pair of Pavia, Italy, the largest changed areas are highlighted by red circles in (b). Various detection results (white - no change, black - change) in (c)-(f).

where the estimation of quantities Q_c and Q_e from the ranks R_{kn} is presented in Appendix A.

The critical values for the statistic F are taken from the Fisher CDF $\mathcal{F}_{v1,v2}$ with degrees of freedom v_1 and v_2 whose evaluation is also reported in Appendix A. The rejection region is right sided. The CDF of the Fisher distribution is expressed as an incomplete beta function, whose values can be easily obtained numerically [35] in order to produce the z-scores.

2) Pairwise Levene testing: Once the Levene test rejects the hypothesis of constant variance, one could ask a question which pairs of samples (images) in the sequence demonstrate meaningful changes and which not. To answer this question a panel of pairwise comparisons is required: if the omnibus test considered K populations, then $K(K-1)/2$ pairwise tests need to be performed. This procedure is commonly referred to as *post-hoc* testing [32] because it relies on the result of the first test. To adjust the FWE in such multiple testing procedure we suggest to employ the Benjamini-Hochberg procedure [12], which has more power than the classical Bonferroni correction. This brings to a frequentist FDR inference procedure presented in Algorithm 2. The aim here is to ensure than the FDR is less or equal than γ. The rejected hypotheses correspond to pairs of images that present statistically meaningful differences in the variance parameter. Note, that in order to obtain the FWE-correction, an alternative

TABLE I

XSAR PAIR DETECTION RESULTS: FALSE POSITIVE RATES (FPR), TRUE POSITIVE RATES (TPR), FALSE DISCOVERY PROPORTIONS (FDP),

TOTAL AMOUNTS OF DETECTION AND COMPUTATION TIMES (SECONDS)

Method	FPR	TPR	FDP	Detection	Time
Ln-GKIT [25]	0.58%	90.36%	9.78%	3.24%	1.08
KL-based [31]	0.40%	86.67%	7.12%	3.10%	2.03
FDR CvM	0.08%	94.76%	1.33%	3.01%	1.54
FDR Wilcoxon	0.28%	98.80%	4.55%	3.69%	2.84

strategy could consist in pooling all the pairwise tests from all pixels rejected by the omnibus Levene FDR-analysis, and performing a second round of FDR-detection targeting image pairs according to Algorithm 2.

VI. EXPERIMENTS WITH IMAGE PAIRS

In this section we validate the proposed FDR-based approach on remotely sensed SAR, dermatological and still camera image pairs. The threshold for the FDR, which is the only conceptual parameter of the model and needs to be selected *a priori*, is set to $\gamma = 0.1$, see discussion in [11]. However, examples of receiver operator characteristic (ROC) curves of the proposed method as a function of the threshold γ are also reported. The central matching is performed on $\Delta = 50\%$ of the data z-scores, and the partitioning with $K = 75$ provides good smoothness of the estimation. For the benchmark methods [9] and [24] the variance levels have been set in supervised manner in remote sensing and dermatological applications, estimated from 100 frames for the still camera application, and the thresholds (for Fig. 6) were set manually by choosing the best test performance. The experiments were performed on a C++ implementation with CPU-parallelization on an Intel Core-i7 2GHz, 8Gb RAM, Win7 system.

A. Remotely sensed images

We first present the experiments with speckled remotely sensed SAR imagery. Speckle is a noise-like phenomenon that appears as a result of mutual interference of wavefronts. It is characteristic of coherent imaging, in particular, in radar, sonar, lidar imagery [39]. Here we consider a flood monitoring sequence of two radar images in an agricultural area near Pavia, Italy, see Fig. 2, acquired by the shuttle-born XSAR sensor on (a) April 16 and (b) April 18, 1994 (approx. 15 m res., four-looks). The performance was evaluated using a manually prepared ground truth map, the main changed areas are highlighted in Fig. 2(b). It is immediate, that the results obtained by the proposed method, with (e) CvM and (f) Wilcoxon features, are accurate and that they are less noisy than that of (d) the cumulant-based Kullback-Leibler (KL) thresholding [31] and (c) SAR-specific lognormal distribution-based generalized Kittler-Illingworth thresholding (Ln-GKIT) [25] methods. The numerical accuracies and computation times are reported in Table I, and the ROC curves presented in Fig. 3(a). The FDR is the expectation of the false discovery proportion (FDP) quantity presented in Table I. A ROC curve consists of only one point for Ln-GKIT method, since there is no input parameter; and for the proposed FDR-based methods, the γ parameter has

(a) ROCs, XSAR in Fig. 2   (b) ROCs, still camera in Fig. 6(a)

(c) ROCs. still camera in Fig. 6(c)   (d) ROCs, still camera in Fig. 6(d)
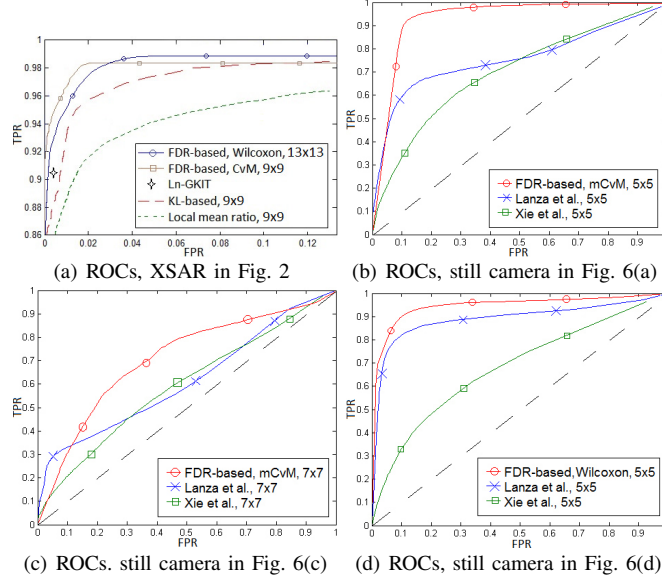
Fig. 3. ROCs for the employed detectors on XSAR and still camera datasets.

been considered as a variable in order to construct its ROC curve. We observe that the employed features report more accurate results and are stable to the speckle disturbance factor with a slight oversmoothing on borders as a consequence of the employed sliding window approach.

Another remotely sensed SAR image pair, obtained by the space-born COSMO-SkyMed (CSK) sensor, depicts the port area of Port-au-Prince, Haiti, (a) before (April 2009) and (b) after (January 2010) the earthquake of January 2010 (StripMap acquisition mode, 2.5 m resolution, single-look), see Fig. 4. The purpose of this experiment is to compare visually the quality of the detection in a complex urban context. We report the benchmark results by (c) KL-based detection with thresholding at 2%, (d) Ln-GKIT [25], and (e) Lanza et al. method [9]. In (f) we demonstrate the results of the classical inference procedure ($\alpha = 0.01$ for individual tests) based on p-values of the Wilcoxon test, ignoring the intra-sample dependence. We then present the results obtained by the proposed FDR-based detector with: (g) the Wilcoxon feature, and (h) the mCvM feature. In the absence of ground-truth data we qualitatively observe that the results obtained with these different features are substantially different. The mCvM result is a more conservative version (with less detection) of the SAR-specific approach result reported in (d), both highlighting the changes in urban areas. Meanwhile the Wilcoxon result in (g) is more similar to general change detector result reported in (c). The visual examination suggests less noisy detection by the proposed FDR-based method as compared to methods in (c)-(e).

B. Dermatological images

We next validate the designed change detector in the application to hyper-pigmentation clinical treatment progress detection, see Fig. 5. Initial acquisitions were provided by one of the world leading companies in dermatology. The images were obtained by an optical sensor and evaluated by medical experts to identify the degree of significant change (one number per image pair). The images were coregistered and the zones of interest were defined (inside the inner contours): Solely the pixels inside these zones are tested. To compensate for the strong illumination changes the

(a) CSK SAR image 1
(b) CSK SAR image 2
(c) thresholded KL [31], 7×7
(d) Ln-GKIT [25]

(e) Lanza et al. [9], 7×7
(f) standard Wilcoxon, 7×7
(g) FDR-based Wilcoxon, 7×7
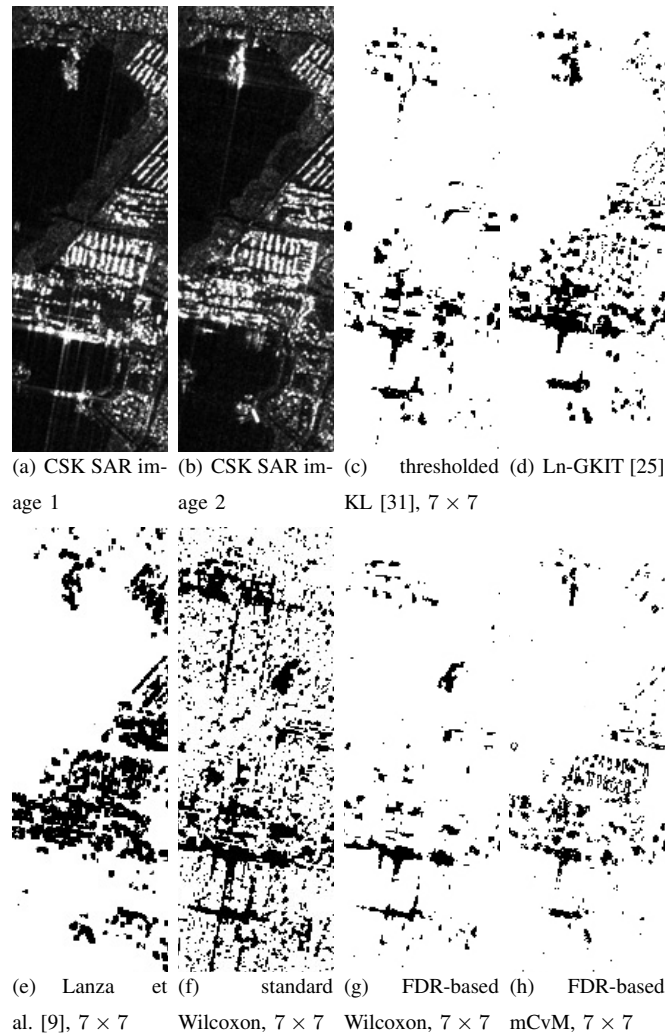(h) FDR-based mCvM, 7×7

Fig. 4. 650×200 pixel COSMO-SkyMed 2.5m resolution StripMap image pair (©ASI) of Port-au-Prince, Haiti [(a),(b)], and various detection results (white - no change, black - change) in (c)-(h).

reference zones (between contours) were selected to normalize the observation values in test zones. We perform the FDR-based CvM feature experiments with different patch sizes to analyse the contribution of this parameter: 5×5 (top row) and 9×9 (bottom row). We have observed that the total percentage of detection varies only slightly (within 2%) with patch sizes in range $5 \div 11$. The local mean ratio, that constitutes a simple visual indicator of the changes, comparisons with methods [24] and [9], as well as KL-based thresholded detection are presented in third, fourth and fifth columns, respectively. No dermatology-specific methods were available for comparisons. The results obtained with the proposed FDR-based detection method using the CvM feature give 5.35% (top) and 21.02% (bottom) of detection. This is considered a relevant result, since the expert's estimates assign the same level of severity of hyper-pigmentation (no significant change) in the top row image pair, and an appreciable improvement (significant change) in bottom image pair. The proposed method has been validated on a total of 8 hyper-pigmentation image sequences, reporting automatic results in agreement with the experts' evaluations. Note that the FDR-based method gives a

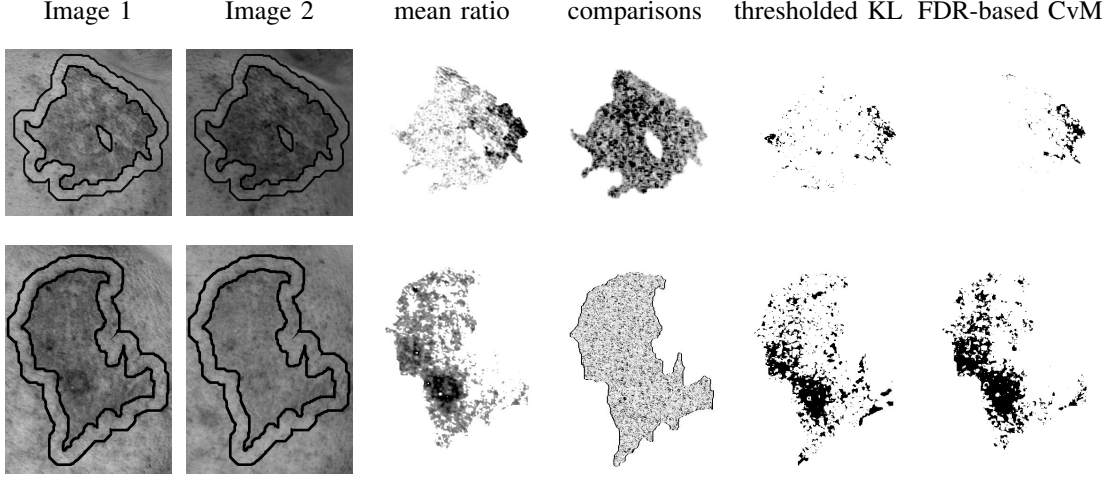Image 1 Image 2 mean ratio comparisons thresholded KL FDR-based CvM

Fig. 5. Skin hyper-pigmentation evolution detection on two dermatological image pairs (©INRIA) 400×400 (top) and 400×240 (bottom) pixels. Processing performed with 5×5 (top) and 9×9 (bottom) local patches. The comparisons are obtained with Lanza et al. [9] (top), and Xie et al. [24] (bottom) methods.

statistically meaningful thresholding procedure whereas the benchmark methods [24], [9] only produce profiles. Furthermore, for this specific quantitative application the control of false detection given by FDR is a strong advantage.

C. Still camera images

Here we present the application of the developed method to a classical still camera surveillance application [16], [9], [24] on the images from the *changedetection.net* database [16], see Fig. 6. The detection results are compared with that of the benchmark order-based methods proposed by Xie et al. [24] and Lanza et al. [9]. The mCvM feature (a)-(c), with $\gamma = 0.05$, gives a contouring detection, similar to the benchmark methods. In case of the mCvM feature, this effect is due to its construction that is sensitive to all changes (since Cramér-von-Mises test is derived to validate the general goodness-of-fit hypothesis), except for that of the first moment, which is subtracted from the individual samples. Consequently, when one homogeneous area is changed into another characterized by a distinct mean value but similar shape of the distribution (variance, and other higher order characteristics), the change is detected only along the border, which is the effect of patch-based formulation. Furthermore, the greater the considered patch size, the wider border is detected along the areas with an abrupt mean-change.

The image pair (c) is obtained by manually varying the brightness of the images in order to demonstrate the robustness of the mCvM feature to the constant intensity shifts. The above mentioned bordering effect is observed both on significant changes (moving people), and along the illumination shift border. This effect is, however, mitigated by the (partial) detection of the inner parts of the moving shapes, and, as such, the bordering is less pronounced than what is obtained with both [24] and [9]. Interestingly, the method in [24] did not produce false positives along the intensity shift, instead failed to detect any significant changes inside this area. This might be due to the single level of variance set for the whole frame that resulted in the reduced sensitivity in the lower intensity area. The Wilcoxon (d) and CvM (e) features detect the inner-parts of the changed objects as well, yet

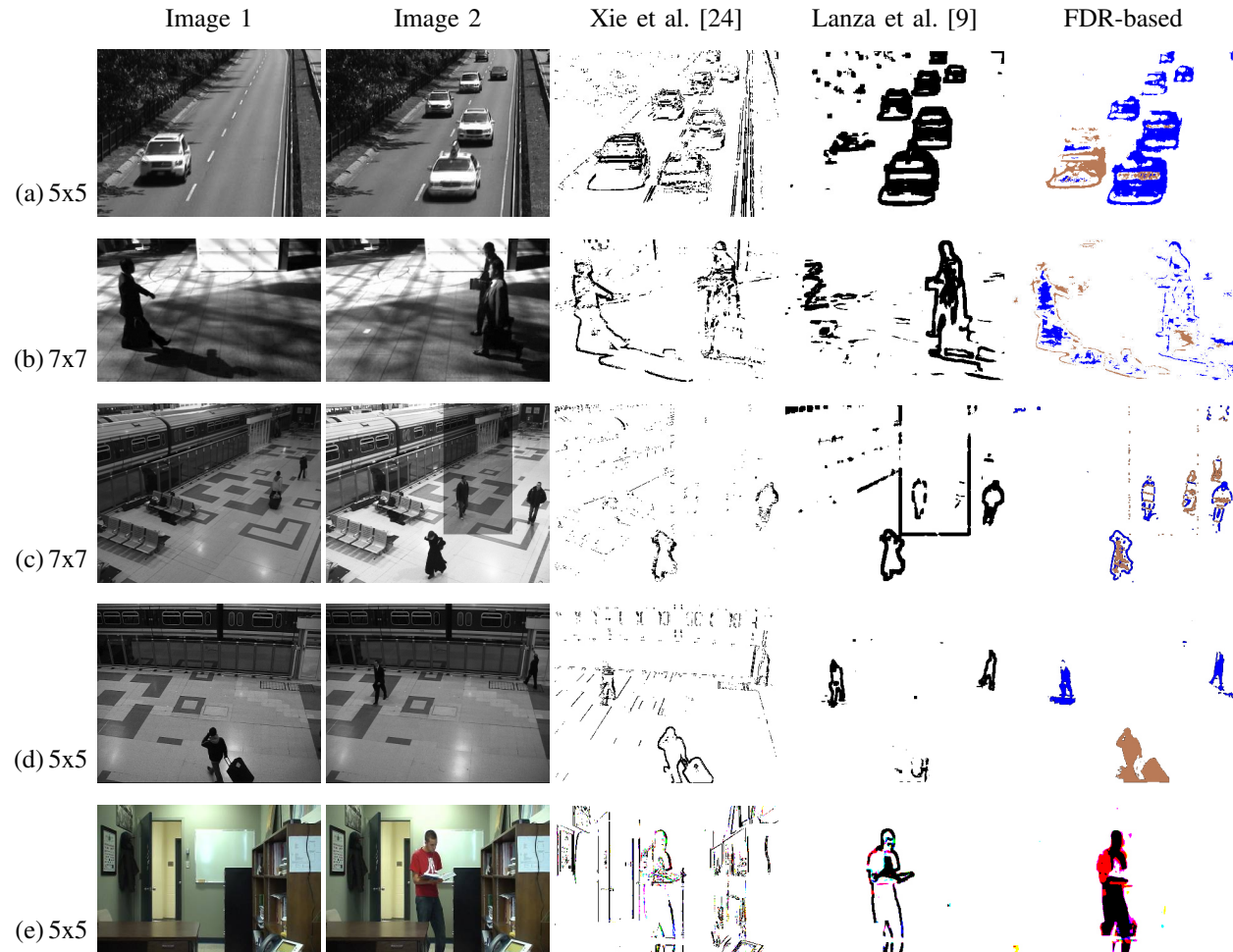| Image 1	Image 2	Xie et al. [24]	Lanza et al. [9]	FDR-based

(a) 5x5

(b) 7x7

(c) 7x7

(d) 5x5

(e) 5x5

Fig. 6. (a),(b),(e) 360×240 and (c),(d) 720×580 pixels still camera image pairs from the database [16], and the change detection results by methods [9], [24] and the FDR-based approach with (a)-(c) mCvM, (d) Wilcoxon and (e) CvM features.

demonstrate sensitivity to shadows in (d). The two colours in the detection results (a)-(d) are in order to distinguish when the changed areas become darker (brown) and brighter (blue); in (e) the inverted colours represent detection performed separately in R, G and B channels of the RGB input (this means that the black colour corresponds to the changes identified in all three bands). The reference maps were used to produce the ROC curves for image pairs (a), (c) and (d), see Fig. 3. The comparisons demonstrate that the considered features allow to obtain comparatively accurate results. These results demonstrate how the flexible choice of statistics allows to obtain substantially different detection. The calculation of the FDR-based map on the most computationally-intensive pair from (c) took 0.5 sec.

VII. EXPERIMENTS WITH MULTIPLE IMAGES

We present semi-synthetic experiments to validate the performance of the change detector in a three-image scenario. To this end we have manually constructed semi-synthetic 1000-by-1000 pixels SAR images from 6 zones, see Fig. 7: unchanged areas A and B that represent the vast majority of the pixels to ensure the fulfilment of the

(a) Ground truth   (b) Image 1   (c) Image 2   (d) Image 3

(e) Levene FDR on images 2 and 3   (f) combined pairwise Levene FDR   (g) p-values of omnibus FDR   (h) omnibus Levene FDR
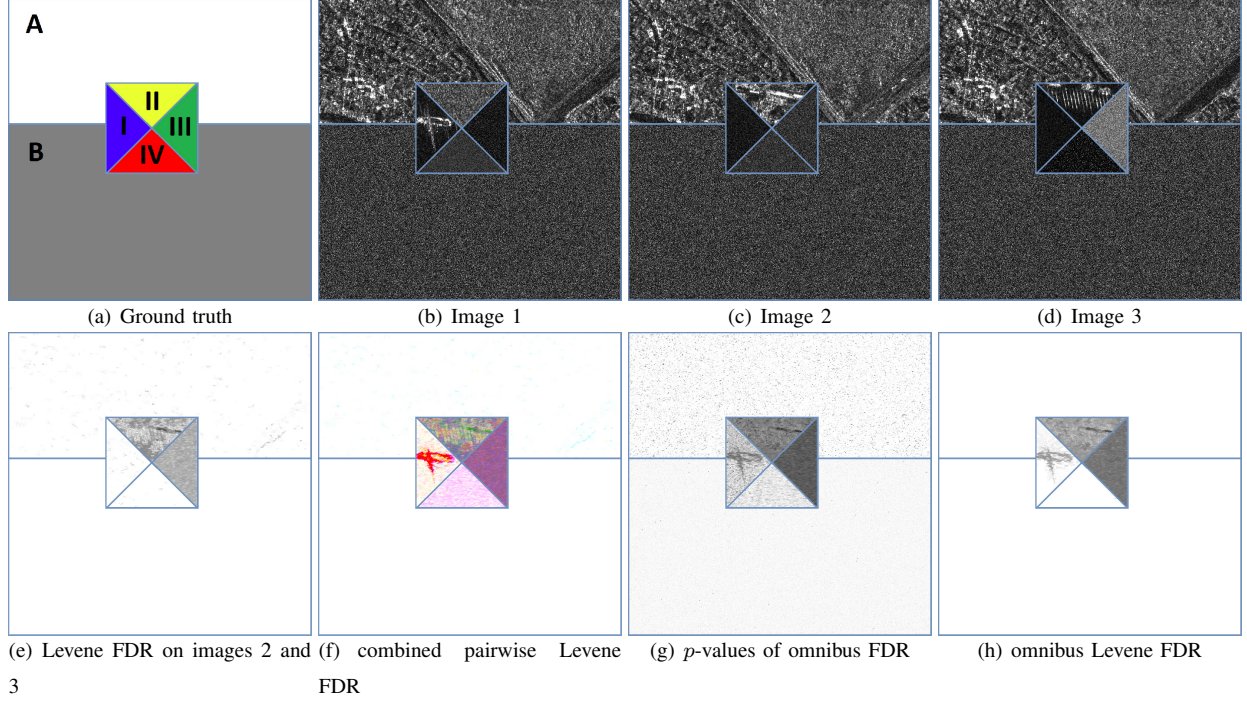
Fig. 7. Semi-synthetic SAR image sequence: (a) ground truth and (b)-(d) images, results of (e),(f) pairwise and (g),(h) omnibus Levene FDR-detection.

model's assumptions, and the central 300-by-300 square reporting four types of changes in zones I through IV. To produce these collages patches of the 1m resolution single-look COSMO-SkyMed SpotLight images of Haiti, Port-au-Prince (acquired on April 11, May 1 and August 17, all in 2011) were employed. We have produced synthetic SAR images using constant signal C and multiplicative Gamma-distributed speckle noise Γ with various shape k and scale θ parameters. This, in fact, constitutes one of the textbook models for SAR data affected by speckle [39]. Specifically, these synthetic areas report independent random amplitudes generated from $\mathcal{B}(C, k, \theta) = C \cdot \Gamma(k, \theta)$ by inverse transform sampling. The unchanged areas are: zone A that reports SAR images with no significant changes, and zone B with observations simulated as $\mathcal{B}(100, 1, 0.5)$ for all three images. The changed areas are: SAR zone I where the first image is significantly different from the second and the third that are substantially equivalent, SAR zone II with all images significantly different, simulated zone III with the first image $\mathcal{B}(100, 1, 0.2)$, second $\mathcal{B}(100, 2, 0.2)$, third $\mathcal{B}(100, 5, 0.2)$, and simulated zone IV with images generated as $\mathcal{B}(50, 4, 0.2)$, $\mathcal{B}(100, 1, 0.2)$ and $\mathcal{B}(150, 0.44, 0.2)$.

The experiments were performed with the Levene statistic both in the omnibus and the pairwise versions. The variance of the generated zones can be calculated as $\mathrm{Var}\mathcal{B} = C^2 k \theta^2$, and, as such, variances in zone IV are equal, and in zone III are 4, 8 and 20 for images 1, 2 and 3, respectively. Therefore, the Levene statistic-based analysis of these three images should reveal partial changes in zones I and II, entire zone III, and nothing in zones A, B and IV.

We present detection results in Fig. 7(e)-(h). Throughout this experiment we have employed sample clipping with $\omega = 0.05$, and a 9-by-9 local window. First, in (e) we report an FDR map of pairwise detection of changes between images 2 and 3, and in (f) a combined three-band FDR map of all pairwise comparisons in inverse colour (yellow

corresponds to image pair 1-2, sky blue - 2-3, magenta - 1-3, black - to changes in all three pairs). Note that here, as well as in (h), we perform soft FDR thresholding (i.e., do not replace the values below the FDR threshold) in order to demonstrate the contrast provided by the Levene statistic. In (g) we report the pre-FDR thresholding p-values collected from the null estimated using (5), and in (h) the FDR-based detection using the omnibus Levene statistic. The results in (e), (f) and (h) are post-processed by applying morphological reconstruction by opening [40]. It is employed with the smallest possible 3-by-3 filter window and serves to remove the isolated salt-and-pepper-like detected change pixels.

The analysis of the results demonstrates that the pairwise tests are more sensitive to changes, and produce results non-trivially different from those reported by the omnibus statistic. Indeed, the dark pixels in zone A in (e) correspond to small changes that are not distinguished by the omnibus statistic, see in (g). This omnibus decision is supported by the relative weakness of the change in the corresponding area identified by the other two pairwise tests. The same effect can be seen even better in zone IV in (f). Indeed, we observe a well visible change pattern in image pair 1-3, that is due to a strong change of distribution parameters used to generate the corresponding patches. This detection is erroneous, since the distribution variance is in fact the same. The reason for that is the sample size which needs to be larger in order to adequately adjust to strong variation in the distributions. The omnibus test did not suffer from the same mistake due to a generally higher level of tolerance to pairwise fluctuations. The performance in zones B and I-III is correct and rather comparable. Note, that here we have not relied on the post-hoc testing procedure presented in Algorithm 2, instead we have performed three pairwise and one omnibus Levene FDR detection procedures independently in order to evaluate their comparative performance.

Finally, we have empirically validated the stipulated above assumption of normality of z-scores with the dependencies present in the employed samples. We have observed Gaussian-like shape of the distribution even in the semi-synthetic experiment, which is less obvious since the images are comprised of areas of different nature– real and simulated SAR. Specifically, for the omnibus statistic the empirical null is estimated as $N(1.61, 2.14)$ with $Sk = 0.55$ and $Ku = 3.44$, where Sk is the skewness and Ku - the kurtosis of the data [32]. The latter two are typical characteristics of the normal distributions, that are exactly equal to 0 and 3, respectively, for perfectly Gaussian populations. In the pairwise analysis, the pair 2-3 demonstrated $N(1.24, 2.55)$, $Sk = 0.41$, $Ku = 2.85$, and essentially similar values for the other two pairs. The values of skewness and kurtosis are not too far from the Gaussian values, and typically improve with larger sample sizes (local window). Specifically, for the omnibus statistic $(Sk, Ku)_{\text{13-by-13}} = (0.48, 2.95)$ and $(Sk, Ku)_{\text{17-by-17}} = (0.25, 2.98)$. These values are typically worse for the omnibus version of the statistic since the dependence structure in the multi-image data is more complex.

VIII. DISCUSSION AND CONCLUSIONS

In this paper we have proposed a novel statistical approach to patch-based change detection on two or more coregistered images. This approach gives a unified statistical thresholding procedure to perform change detection based on statistical features that have a known distribution under the no-change hypothesis. The proposed FDR-formulation is based on the control of the proportion of false alarms among all detections. This efficient technique for large scale hypotheses testing allows to use the wide range of statistical tests developed in the state-of-the-art [32] by

adjusting to the dependence structure present in the images and the patch-based samples. The approach involves only a few parameters and is highly parallelizable. We have presented several rank-based statistical features that report accurate experimental results and the corresponding detectors positively compare with benchmark techniques in three different applications. Further features can be easily constructed to elaborate application-specific change detectors.

The developed change detection approach constitutes a flexible and statistically sound inference procedure that, nevertheless, has several limitations. First, we assume that the amount of changes present on images is limited (at most 25% of the total image size). This restriction is a price to pay for performing the change detection in an unsupervised manner with no learning data available (such as earlier frames). For the same reason the developed method underperforms in case of more complicated background motion or alteration scenario, e.g., waves or foliage motion for still camera imagery, where the temporal approaches [16], [3] are expected to perform well. This is due to the fact that in an unsupervised scenario analysing two or several images it is very challenging to tune the algorithm to disregard such false positives. The same limitation applies also to methods like [8], [9], [24]. Second, the considered statistical features should have a known distribution under the no-change hypothesis. The presented experiments demonstrate that these limitations are not critical for various applications.

We have investigated the use of FDR-based change detection in case of multiple input images. To this end we have employed an omnibus Levene statistic to test the equality of variances. This statistic can also be employed in the two-sample scenario, and can be easily modified to test the differences in the means by suppressing step 2 in Algorithm 3. The omnibus test can be followed by post-hoc pairwise Levene tests to identify the specific image pairs that report meaningful changes. It is important to mention that post-hoc testing corresponds to a two-step procedure that may not ensure a strict control over the type I error. Furthermore, it is possible to have a change reported by omnibus statistic and no significant change witnessed by any pairwise test. The opposite could also occur. The reason for that is the difference in the forms of the rejection regions specified by these two test procedures. Nevertheless, a strong advantage of the omnibus statistic is the severely reduced computational complexity, and for some applications the specification of the time of the change is not relevant.

In all of the performed experiments we have observed only minor deviations of z-scores from the normal distribution as a result of intra-sample dependence, see Fig. 1 as an example, which experimentally validates the assumptions of Section IV. The good results obtained with different types of images in diverse applications are interpreted as due to the generality of the proposed approach. Note that the developed method (as well as similar spatial approaches [8], [9], [24]) underperforms in case of complicated background motion, when the temporal approaches [3], [16] perform better in discriminating back- and foreground of the images.

The main directions for the future research include: *i*) the integration of the developed change detector into a multi-resolution approach to automatically select the patch sizes, and (*ii*) incorporation of the patches' dependence structure into the model, for instance, via hidden Markov models.

Algorithm 3: Multisample Levene statistic

---

1 replace X_{kn} by ranks R_{kn} in the pooled sample;

2 transform the sample to $O_{kn} = |R_{kn} - \bar{R}_k|$;

3 calculate the clipped sample Y_{kn} ;

4 calculate Q_c and Q_e, and the statistic value $\mathfrak{f} = F$;

5 calculate ε, and degrees of freedom v_1 and v_2;

6 estimate the p-value as $p = 1 - \mathcal{F}_{v_1,v_2}(\mathfrak{f})$.

APPENDIX A

LEVENE STATISTIC CALCULATION

The first step in the statistic's construction enables to compensate the non-normality of the samples. It consists in replacing the observations by their ranks. The Levene statistic is defined in (10), and we denote as $O_{kn} = |R_{kn} - \bar{R}_k|$ the transformed sample, where R_{kn} are the pooled ranks.

We next apply the clipping that allows to improve robustness of the statistic to ouliers. Specifically, let

$$Y_{kn} = \begin{cases} O_{kn}, & O_{kn} < O^R_{k(N-[\omega N])} \\ O_{k(N-[\omega N])}, & \text{otherwise} \end{cases},$$

where O^R_{kn} is the ordered version of O_{kn} such that values of O^R_{kn} in each group k are in ascending order, $[\omega N]$ is the integer part of ωN, and ω is the amount of clipping. The modulus transform employed in O_{kn} projects the "extreme" observations on the right side, and hence the clipping is also one-sided.

The Levene statistic for the dependent samples (11) requires

$$Q_c = ((N - [\omega N]) \sum_{k=1}^{K} (\bar{O}^T_k - \bar{O})^2,$$

$$Q_e = \sum_{k=1}^{K} \sum_{n=1}^{N} (Y_{kn} - \bar{Y}_{\cdot k} - \bar{Y}_{n \cdot} + \bar{Y})^2.$$

Here $\bar{O}^T_k$ denotes the mean of the ω-trimmed sample k (i.e. calculated by ignoring the clipped observations), $\bar{O} = \sum \bar{O}^T_k / K$, and averages $\bar{Y}_{\cdot k} = \sum_{n=1}^{N} Y_{kn}/N$, $\bar{Y}_{n \cdot} = \sum_{k=1}^{K} Y_{kn}/K$ and $\bar{Y} = \sum_n \sum_k Y_{kn}/(NK)$.

The critical values for the statistic F are taken from the Fisher CDF $\mathcal{F}_{v1,v2}$ with degrees of freedom $v_1 = \varepsilon(K-1)$ and $v_2 = \varepsilon(N - [\omega N] - 1)(K - 1)$. Compared to the textbook ANOVA [10], the ε-correction allows to relax the technical assumptions of sphericity of the variance and correlation structure of the K samples, see [32]. The quantity ε is estimated as

$$\varepsilon = \frac{N(K - 1)A/B - 2}{(K - 1)(N - 1 - (K - 1)A/B)},$$

with

$$A = \frac{K^2(\bar{\nu}_d - \bar{\nu})^2}{K-1},$$

$$B = \sum_{j=1}^{K} \sum_{k=1}^{K} \nu_{jk}^2 - 2K \sum_{j=1}^{K} \bar{\nu}_{j \cdot}^2 + K^2 \bar{\nu}^2.$$

The quantities $\nu_{jk} = \sum_{i=1}^{N}(Y_{ij} - \bar{Y}_{\cdot j})(Y_{ik} - \bar{Y}_{\cdot k})/(N-1)$ for $j = 1, \ldots, K$ and $k = 1, \ldots, K$ are the clipped sample covariances. The following averages are employed above: overall $\bar{\nu} = 1/K^2 \sum_j \sum_k \nu_{jk}$, diagonal $\bar{\nu}_d = 1/K \sum_j \nu_{jj}$, and row-wise $\bar{\nu}_{j\cdot} = 1/K \sum_k \nu_{jk}$.

The Levene statistic's z-values are calculated as $z = \Phi^{-1}(p)$, with the corresponding p-values computed through the Fisher distribution as $p = 1 - \mathcal{F}_{v_1, v_2}(\mathfrak{f})$. The complete process is presented in Algorithm 3.

ACKNOWLEDGMENT

REFERENCES

[1] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, 2005.

[2] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogram. Remote Sens.*, vol. 80, pp. 91–106, 2013.

[3] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Conf. on Computer Vision Pattern Recogn.*, 1999, pp. 2246–2252.

[4] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, 2011.

[5] T. S. Haines and T. Xiang, "Background subtraction with Dirichlet process mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 670–683, 2014.

[6] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, 2015.

[7] N. Ohta, "A statistical approach to background subtraction for surveillance systems," in *Int. Conf. Computer Vision*, vol. 2, 2001, pp. 481–486.

[8] M. Singh, V. Parameswaran, and V. Ramesh, "Order consistent change detection via fast statistical significance testing," in *IEEE Conf. Computer Vision Pattern Recogn.*, 2008.

[9] A. Lanza and L. Di Stefano, "Statistical change detection by the pool adjacent violators algorithm," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1894–1910, 2011.

[10] W. J. Conover, *Practical nonparametric statistics*, 3rd ed. New York: Wiley, 1999.

[11] B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge, UK: Cambridge University Press, 2010.

[12] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. R. Stat. Soc. Series B*, vol. 57, pp. 289–300, 1995.

[13] B. Efron, "Large-scale simultaneous hypothesis testing: The choice of a null hypothesis," *J. Am. Stat. Assoc.*, vol. 99, pp. 96–104, 2004.

[14] A. Farcomeni, "A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion," *Stat. Meth. Med. Res.*, vol. 17, no. 4, pp. 347–388, 2007.

[15] V. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "False discovery rate approach to image change detection," in *Proc. of IEEE ICIP*, 2013, pp. 3820–3824.

[16] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *IEEE Workshop on Change Detection, CVPR*, 2012.

[17] A. Hopkins, C. Miller, A. Connolly, C. Genovese, R. Nichol, and L. Wasserman, "A new source detection algorithm using the false-discovery rate," *Astronom. J.*, vol. 123, no. 2, p. 1086, 2002.

[18] J. M. McHugh, J. Konrad, V. Saligrama, P.-M. Jodoin, and D. Castanón, "Motion detection with false discovery rate control," in *Proc. of IEEE ICIP*, 2008, pp. 873–876.

[19] D. Van and M. Unser, "False discovery rate for wavelet-based statistical parametric mapping," *IEEE J. Sel. Top. Signal Process.*, vol. 2, no. 6, pp. 897–906, 2008.

[20] H. D. Nguyen, G. J. McLachlan, N. Cherbuin, and A. L. Janke, "False discovery rate control in magnetic resonance imaging studies via Markov random fields," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1735–1748, 2014.

[21] H. J. Seo and P. Milanfar, "Action recognition from one example," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 33, no. 5, pp. 867–882, 2011.

[22] M. Yokoyama and T. Poggio, "A contour-based moving object detection and tracking," in *Proc. of Int. Conf. Computer Vision*, 2005, pp. 271–276.

[23] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, Dec. 2006.

[24] B. Xie, V. Ramesh, and T. E. Boult, "Sudden illumination change detection using order consistency." *Image Vision Comput.*, pp. 117–125, 2004.

[25] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, 2006.

[26] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, 2005.

[27] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, April 2012.

[28] A. Lingg, E. Zelnio, F. Garber, and B. Rigling, "A sequential framework for image change detection," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2405–2413, May 2014.

[29] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, 2012.

[30] C. Benedek, T. Szirányi, Z. Kato, and J. Zerubia, "Detection of object motion regions in aerial image pairs with a multi-layer Markovian model," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2303–2315, 2009.

[31] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, 2007.

[32] R. R. Wilcox, *Applying contemporary statistical techniques*. Gulf Professional Publishing, 2003.

[33] W. J. Conover and R. L. Iman, "Rank transformations as a bridge between parametric and nonparametric statistics," *Amer. Statistician*, vol. 35, no. 3, pp. 124–129, 1981.

[34] Y. Xiao, A. Gordon, and A. Yakovlev, "A C++ program for the Cramér-von Mises two-sample test," *J. Stat. Softw.*, vol. 17, pp. 1–15, 2006.

[35] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1964.

[36] J. C. W. Rayner and D. J. Best, "Extended anova and rank transform procedures," *Aust. N. Z. J. Stat.*, vol. 55, no. 3, pp. 305–319, 2013.

[37] D. W. Nordstokke and B. D. Zumbo, "A new nonparametric Levene test for equal variances," *Psicológica*, vol. 31, no. 2, pp. 401–430, 2010.

[38] D. W. Zimmerman and B. D. Zumbo, "Relative power of the wilcoxon test, the friedman test, and repeated-measures anova on ranks," *J. Exp. Educ.*, vol. 62, no. 1, pp. 75–86, 1993.

[39] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, 2nd ed. NC, USA: SciTech, Raleigh, 2004.

[40] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, 2001.
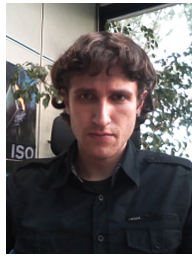
Vladimir A. Krylov received the "specialist" (M.Sc.) degree in applied mathematics and computer science, and the "candidate of physico-mathematical sciences" (Ph.D.) degree in statistics both from the Lomonosov Moscow State University, Moscow, Russia, in 2007 and 2011, respectively. In 2011–2012 he was a postdoctoral fellow with Ariana and Ayin research teams at INRIA, Sophia Antipolis, France; in 2012–2013 he was a research associate with the Dept. of Statistical Science at the University College London, London, UK. Since 2014 he is with the Dept. of Electrical, Electronic, Telecommunications Engineering and Naval Architecture (DITEN) at the University of Genoa, Italy.

His research interests are in the field of statistical signal processing and pattern recognition applied to medical and remote sensing imagery.

Gabriele Moser (S03M05SM14) received the Laurea (M.Sc. equivalent) degree in telecommunications engineering, and the Ph.D. degree in space sciences and engineering from the University of Genoa, Genoa, Italy, in 2001 and 2005, respectively. Since 2014, he has been an Associate Professor of Telecommunications at the University of Genoa. Since 2001, he has cooperated with the Image Processing and Pattern Recognition for Remote Sensing Laboratory, University of Genoa. Since 2013, he has been the Head of the Remote Sensing for Environment and Sustainability Laboratory at the Savona Campus of the University of Genoa. From January to March 2004, he was a Visiting Student with the Institut National de Recherche en Informatique et en Automatique, Sophia Antipolis, France. His research activity is focused on pattern recognition and image processing methodologies for remote sensing and energy applications. He has been an Area Editor of Pattern Recognition Letters (PRL) since 2015, and an Associate Editor of the IEEE Geoscience and Remote Sensing Letters since 2008. He was an Associate Editor of PRL from 2011 to 2015, and a Guest Editor of the September 2015 special issue of the IEEE Geoscience and Remote Sensing Magazine. He served as the Chairman of the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society from 2013 to 2015, and has been serving as IADF TC Co-Chair since 2015.He received the Best Paper Award at the 2010 IEEE Workshop on Hyperspectral Image and Signal Processing.

Sebastiano B. Serpico (M87-SM00-F09) received the Laurea (M.S.) degree in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Italy. Full professor of telecommunications at the Polytechnic School of the University of Genoa, he is the Head of the research group on Signal Processing and Recognition Methods and Systems of the Department of Electrical, Electronic, Telecommunications Engineering, and Naval Architecture of the University of Genoa. His current research interests include pattern recognition for remote sensing images and biomedical images. He is the Chairman of the Institute of Advanced Studies in Information and Communication Technologies (ISICT). He has been the project manager of numerous research projects and an evaluator of project proposals for various programs of the European Union, Italian Space Agency, Italian Ministry of Education and Research, etc. He is the author (or coauthor) of over 200 scientific articles published in journals and conference proceedings. He received the Best Paper Award at the 2010 IEEE Workshop on Hyperspectral Image and Signal Processing. He is an associate editor of the international journal IEEE Transactions on Geoscience and Remote Sensing (TGRS). He was a guest editor of two Special Issues of TGRS on the subject of the analysis of hyperspectral image data (July 2001 issue) and the subject Advances in techniques for the analysis of remote sensing data (March 2005 issue), and of the IEEE Proceedings journal on the subject of Remote Sensing of Natural Disasters (October 2012 issue). From 1998 to 2002, he was the chairman of the SPIE/EUROPTO series of conferences on Signal and Image Processing for Remote Sensing. He was Co-Chair of the IEEE International Geoscience and Remote Sensing Symposium in 2015 (Milan, Italy, July 2015). He is a Fellow of the IEEE.

Josiane Zerubia (S'78–M'81-SM'99-F'03) has been a permanent research scientist at INRIA since 1989 and director of research since July 1995. She was head of the PASTIS remote sensing laboratory (INRIA Sophia-Antipolis) from mid-1995 to 1997 and of the Ariana research group (INRIA/CNRS/University of Nice), which worked on inverse problems in remote sensing and biological imaging, from 1998 to 2011. Since January 2012, she has been head of Ayin research group (INRIA-SAM) dedicated to models of spatio-temporal structure for high resolution image processing with a focus on remote sensing and skincare imaging.

She has been professor at SUPAERO (ISAE) in Toulouse since 1999. Before that, she was with the Signal and Image Processing Institute of the University of Southern California (USC) in Los-Angeles as a postdoc. She also worked as a researcher for the LASSY (University of Nice/CNRS) from 1984 to 1988 and in the Research Laboratory of Hewlett Packard in France and in Palo-Alto (CA) from 1982 to 1984. She received the MSc degree from the Department of Electrical Engineering at ENSIEG, Grenoble, France in 1981, the Doctor of Engineering degree, her PhD and her Habilitation, in 1986, 1988, and 1994 respectively, all from the University of Nice Sophia-Antipolis, France.

She is a Fellow of the IEEE (2003-) and IEEE SP Society Distinguished Lecturer (2016-2017). She was a member of the IEEE IDMSP TC (SP Society) from 1997 till 2003, of the IEEE BISP TC (SP Society) from 2004 till 2012 and of the IVMSP TC (SP Society) from 2008 till 2013. She was associate editor of IEEE Trans. on IP from 1998 to 2002, area editor of IEEE Trans. on IP from 2003 to 2006, guest co-editor of a special issue of IEEE Trans. on PAMI in 2003, member of the editorial board of IJCV from 2004 till March 2013 and member-at-large of the Board of Governors of the IEEE SP Society from 2002 to 2004. She has also been a member of the editorial board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998, of the Foundation and Trends in Signal Processing since 2007 and member-at-large of the Board of Governors of the SFPT since September 2014. She has been associate editor of the on-line resource Earthzine (IEEE CEO and GEOSS) since 2006. She was co-chair of two workshops on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'01, Sophia Antipolis, France, and EMMCVPR'03, Lisbon, Portugal), co-chair of a workshop on Image Processing and Related Mathematical Fields (IPRM'02, Moscow, Russia), technical program chair of a workshop on Photogrammetry and Remote Sensing for Urban Areas (Marne La Valle, France, 2003), co-chair of the special sessions at IEEE ICASSP 2006 (Toulouse, France) and IEEE ISBI 2008 (Paris, France), publicity chair of IEEE ICIP 2011 (Brussels, Belgium), tutorial co-chair of IEEE ICIP 2014 (Paris, France), general co-chair of the workshop Earthvision at IEEE CVPR 2015 (Boston, USA) and a member of the organizing committee and plenary talk co-chair of IEEE-EURASIP EUSIPCO 2015 (Nice, France). She also organized and chaired an international workshop on Stochastic Geometry and Big Data at Sophia Antipolis, France, in November 2015.

Her main research interest is in image processing using probabilistic models. She also works on parameter estimation, statistical learning and optimization techniques.