Change Detection in Dynamic Scenes using Local Adaptive Transform

Hakan Haberdar
hhaberdar@uh.edu,www.haberdar.org
Shishir K. Shah
sshah@central.uh.edu,www.qil.uh.edu

Quantitative Imaging Lab
University of Houston
Houston, Texas, U.S.A

Abstract

In this paper, we propose a framework that can be used for detecting relevant changes in highly dynamic scenes, where the background has several changing elements. To establish a clear distinction between what is relevant and what is not is a very challenging task. Therefore, we first categorize the changes into two main classes called *ordinary change*s and *relevant change*s. Detected changes are considered as irrelevant if they are recurrent elements and changes pertaining on the dynamic background of the scene. The proposed framework makes use of a set of orthogonal linear transforms to capture spatiotemporal signatures of local ordinary change patterns and subsequently employ them in the detection of relevant changes. The use of this framework is demonstrated in a variety of videos with highly dynamic backgrounds including lakes, pools, and roads. Compared to existing methods reported on the same test videos, the proposed framework detects the relevant changes more accurately.

1 Introduction

Given a set of the same type of data, the detection of changes between two samples in the set is a problem of interest in variety of applications such as machine monitoring [25], medical diagnosis and treatment [34], video surveillance [10], and remote sensing [32]. The definition of what should be considered as *change* is usually domain specific. Furthermore, when there is more than one type of change, the change detection problem becomes cumbersome. In this paper, our focus is on finding regions of relevant change in videos acquired in dynamic outdoor environments, where there are many changing elements (e.g., shimmering water or blowing trees) in the foreground and background that may cause false alarms.

A wide variety of algorithms such as significance testing, predictive models, and background modeling have been proposed for image change detection [28]. Several algorithms have been dedicated to background modeling for identifying foreground objects [23]. The most common approach is to build a model for illumination changes and minor variations using Gaussian mixture models [15]. These methods are usually built on pixel-based structures and work well in detecting changes that can be modeled as independent events. However, they are unable to model complicated change patterns that may be related in space and/or time. We propose a change detection algorithm based on significance testing using spatiotemporal features. Our method shares foundations with video compression techniques [2, 22]. These methods employ only a single model for the entire image due to the

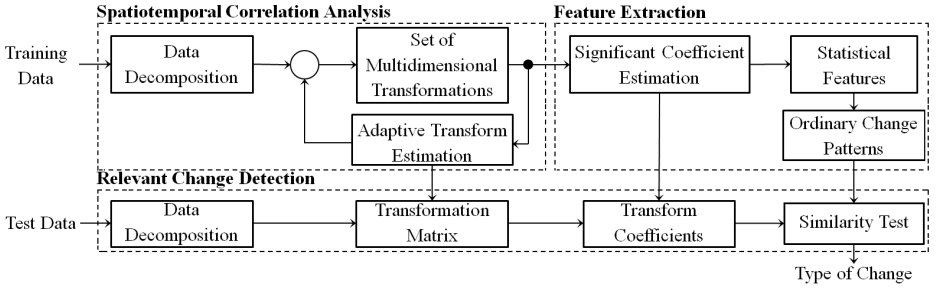


Figure 1: Our algorithm consists of three main blocks. First, given a video containing only ordinary changes, we are interested in finding representations where the spatiotemporal features of the ordinary changes can be captured. Then, we use the training examples to extract spatiotemporal signatures of the ordinary change patterns. Finally, we estimate the existence of the relevant change in given test input by interpolating from the training samples.

constraints imposed by the quantization and coding steps. Another related approach is to detect changes by analyzing the video directly in the compressed domain [8, 35]. However, these methods operate on previously quantized transformed values depending on the transformation methods. Its accuracy depends on the specific parameters of the compression settings. A common drawback of previous methods is the fact that the reported results are usually obtained on a few self-acquired videos, where the backgrounds have limited alterations [13, 18, 21]. Change detection in the presence of highly dynamic background elements is still a challenging problem, and no current methods have shown to be effective.

In this paper, we propose a framework (Fig. 1) that can be used for detecting relevant changes in videos with highly dynamic scenes, where the background has several altering elements. In such scenarios, there are almost always changes in the scene. To establish a clear distinction between what is relevant change and what is not, we first categorize the change into two main classes; namely, *ordinary change* and *relevant change*. Changes are considered as irrelevant if they are recurrent elements and changes pertaining to the dynamic background of the scene. On the other hand, an alteration that does not conform to the expected pattern of ordinary change is defined as the relevant change. We need to distinguish ordinary changes from relevant changes in order to avoid false alarms. Pixels, which belong to regions of the ordinary change, are typically correlated in space and/or time among a set of consecutive frames. This correlation stems from the repetitive nature of ordinary change patterns and induces spatiotemporal signatures [31, 33], which are specific to the ordinary change patterns. We propose that one can make use of the spatiotemporal signatures to discriminate ordinary changes from relevant changes. The image pixel space is usually not considered suitable for capturing spatiotemporal features. Instead, if we can transform a set of frames containing ordinary changes to another representation space where the pixels are decorrelated, we can capture spatiotemporal signatures. This will allow us to learn within the framework to recognize ordinary change patterns. Then, when a change unrelated to ordinary change patterns occurs, the framework can label it as a relevant change. Due to the amount of the data in video processing, a chosen transform should be fast and simple to implement such as linear transforms [29]. Orthogonal linear transforms redistribute the energy stored in the input data, decorrelate it, and provide compact representations [5]. Accordingly, we propose to use orthogonal linear transforms to exploit spatiotemporal signatures

of local ordinary change patterns. Prediction of the optimal orthogonal linear transforms for different ordinary changes patterns is not a trivial process. In terms of energy compaction, Karhunen-Lo'eve transform (KLT) has the best efficiency; however, KLT has high computational complexity [29]. Instead, we propose to estimate a suitable transform for a local ordinary change pattern from a collection of linear transformations having complementary orthogonal basis vectors. Our approach is built up on a data decomposition model generating local three dimensional blocks. Therefore, the estimated change mask may suggest if there is a relevant change within the block, but we need to examine each frame region in the block to obtain individual pixels belonging to the regions of change in each frame. This may cause blocking artifacts. In order to compensate for these artifacts, we apply Markov random field regularization [17]. To evaluate the performance of the proposed method, experiments are performed using the test videos with highly dynamic backgrounds provided by *ChangeDetection.net* [9]. The quantitative comparison of the detection results from the proposed framework to other methods demonstrates improved accuracy.

2 Change Detection Framework

The proposed framework employs spatiotemporal features to detect the relevant changes. This requires three dimensional block-based processing and extends the change detection problem from comparing *two regions* to comparing *two sets of consecutive regions*.

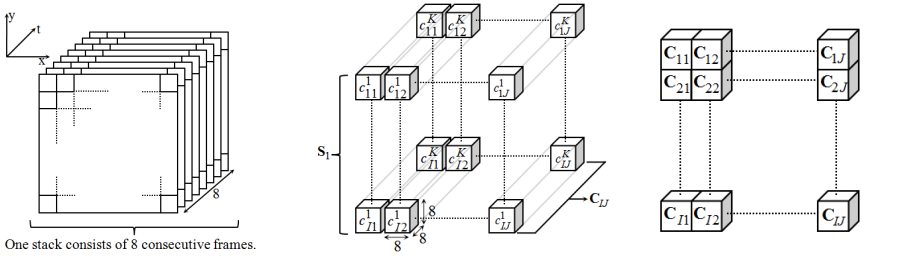
2.1 Spatiotemporal Correlation Analysis

Ordinary change regions are typically correlated in space and/or time among consecutive frames. This correlation induces spatiotemporal signatures specific to ordinary change patterns. Our goal is to find representation spaces that we can capture spatiotemporal signatures.

2.1.1 Data Decomposition

The data decomposition is needed to divide a frame sequence into subblocks such that local spatiotemporal signatures can be extracted. Let $\mathbf{V}$ denote a sequence of frames, with $\mathbf{V} = \{F_1,\ldots,F_t,F_{t+1},\ldots,F_\mathfrak{T}\}$. Let $\mathbf{V}_o$ be a subset of $\mathbf{V}$, including the frames F_τ for $\tau = 1,\ldots,t$. We are given that the subset $\mathbf{V}_o$ contains only ordinary changes. This is a reasonable assumption for the change detection problem, where two states of an entity are under investigation. The rest of $\mathbf{V}$ may contain ordinary changes, relevant changes, or both. Following the same approach in data compression techniques [29], we divide every frame into regions of 8 by 8 pixels in order to improve the localized correlation. Then, 8 consecutive frames are grouped to form a stack as shown in Fig. 2. Let $\S$ denote the set of stacks, with $\S = \{\mathbf{S}_k\}_{k=1}^{K}$, where $K = t/8$. Every stack is composed of $8\times 8\times 8$ blocks called as *cube*s (Fig. 2 (b)). Cubes in the stack $\mathbf{S}_k$ are denoted as c_{ij}^k where $i = 1,\ldots,I$, $j = 1,\ldots,J$, I, and J are the number of the cubes in vertical and horizontal directions, respectively. After the decomposition, a set of t frames turns into a set of K stacks, each of which contains $I*J$ cubes. We anticipate that spatiotemporal signature of each cube in a stack may be unique. We perform a further grouping for corresponding cubes. Cubes in different stacks are defined as *corresponding cubes* if $u = p$ and $v = r$ for $c_{uv}^{\kappa_1}$ and $c_{pr}^{\kappa_2}$, where $\kappa_1 \neq \kappa_2$. We collect corresponding cubes in sets denoted by $\mathbf{C}_{ij}$ for each i and j as shown in the example in Fig. 2 (c). Namely, the set $\mathbf{V}_o$ becomes an $I\times J$ grid of corresponding cube sets. Each $\mathbf{C}_{ij}$ is considered as a summary of

ordinary change patterns in that local region. We expect the cubes in a corresponding cube set to share similar spatiotemporal signatures. We now need to estimate a suitable transform for each corresponding cube set to exploit the spatiotemporal signatures.



(a) (b) (c)

Figure 2: Illustration of data decomposition. Each frame is divided into 8 by 8 regions, and each 8 consecutive frames are stacked as in (a). Stacking is performed across the frames. $\mathbf{S}_1$ denotes the first stack. A stack is composed of $8 \times 8 \times 8$ blocks, called *cube*s. In (b), cubes in the stack $\mathbf{S}_k$ are denoted by c_{ij}^k, where i=1,...,I, j=1,...,J. K is the total number of the stacks. In (c), we present corresponding cube sets. A corresponding cube set is composed of corresponding cubes in the different stacks (e.g., $\mathbf{C}_{IJ} = \{c_{IJ}^1, \ldots, c_{IJ}^K\}$). We expect the cubes in a corresponding cube set to share similar spatiotemporal signatures.

2.1.2 Base Transforms

Let us define an $I \times J$ matrix $\mathbf{T}$ of transforms. An element T_{ij} of $\mathbf{T}$ represents the transform that is suitable for $\mathbf{C}_{ij}$. A suitable transform is defined as the one, where transformed values are independent of one another and the energy is compacted on a few transformed values regardless of their relative locations. Estimating $\mathbf{T}$ is not a straightforward procedure. Orthogonal linear transforms redistribute the energy stored in the input and provide compact representations. Therefore, we employ three orthogonal linear transforms as the base transforms: i) discrete cosine transform (DCT) [3], ii) Walsh-Hadamard transform (WHT) [4], and iii) Slant transform (ST) [26]. DCT, WHT, and ST are used together because they have complementary basis vectors that enable the framework to capture different types of ordinary change patterns. DCT is a sinusoidal transform that is widely used in applications requiring compact representations [2, 19]. WHT is a non-sinusoidal transform having basis vectors that are rectangular or square waves; therefore, it can represent patterns with sharp discontinuities more accurately using fewer values than DCT. ST has been successfully applied to image coding applications [7, 27]. The basis vectors of ST are derived from sawtooth waveforms and considered as a good complement to WHT [1]. Depending on the ordinary change patterns in each set, elements of $\mathbf{T}$ are assigned to one of these three base transforms.

2.1.3 Adaptive Transform Estimation for the Ordinary Change Patterns

We call a transform *compact* if the energy of transformed values does not uniformly scatter in the representation space. Let D be a set of N real valued numbers, with $D = \{d_1, \ldots, d_N\}$. Our goal is to estimate the most suitable transform available for D from the collection of base transforms. In our setting, D refers to a cube in a corresponding cube set. Out of the

base transforms provided, the most suitable transform for D is the one having transformed values where the energy is least scattered. Let Ω denote the set of transformed values of D, with $\Omega = \{\omega_1, \ldots, \omega_N\}$. Let E be the total energy stored in Ω, with $\mathrm{E} = \sum_{\omega_i \in \Omega} \omega_i^2$. In terms of energy scattering, the worst case is a uniformly distributed energy across the transformed values. For such a case, $\omega_i^2 = \frac{\mathrm{E}}{N}$ for each i. Let us normalize Ω based on the energy stored in each transformed value so that E is going to be 1, and accordingly ω_i^2, $\omega_i \in \Omega$ will be $\frac{1}{N}$. We can use this extreme case to define a coefficient that describes how compact the energy in Ω would be. Let ξ_s denote the compactness coefficient and $\grave{\Omega}$ denote the set of normalized transformed values, with $\grave{\Omega} = \{\grave{\omega}_1, \ldots, \grave{\omega}_N\}$. We define ξ_s as follows:

$$\xi_s = \sum_{i=1}^{N} \left(\frac{1}{N} - \grave{\omega}_i \right)^2, \text{ and } \xi_s \in [0, 1 - \frac{1}{N}]. \tag{1}$$

If a transform can compact all the energy of the input in one single transformed value, the transform can be considered as the most suitable one for the input. In such a case, only one element of $\grave{\Omega}$ will be 1, while the rest is zero. Accordingly, the value of ξ_s would be $1 - \frac{1}{N}$. On the other hand, when the energy is distributed uniformly, ξ_s becomes zero. We can estimate the most suitable transform available for a corresponding cube set $\mathbf{C}_{ij}$ in two steps. First, we compute compactness coefficients ξ_s^{DCT}, ξ_s^{WHT}, and ξ_s^{ST} for the base transforms for each cube c_{ij}^k in $\mathbf{C}_{ij}$. Then, the transform having the largest compactness coefficient value is considered as the most suitable for c_{ij}^k. This results in K transforms for $\mathbf{C}_{ij}$. Then, the transform that is the most common amongst estimated K transforms is assigned to T_{ij}.

2.2 Feature Extraction

Let us call the transformed values as *transform coefficients*. With a suitable transform, majority of the coefficients tend to have small values. Our goal is to find a *significant subset* of transform coefficients for each corresponding cube set. A significant subset should contain a small number of coefficients that contribute to the most of the energy.

2.2.1 Significant Transform Coefficients

Several studies using orthogonal transforms assume that the same set of transform coefficients can be neglected for all types of data [16, 30]. The accuracy of this assumption relies on the properties of the input, and it may cause loss of distinctive features. Instead, we propose to estimate a significant subset based on the energy of each coefficient throughout the cube set. Let Ω_{ij}^k be the set of transform coefficients of a cube c_{ij}^k in $\mathbf{C}_{ij}$, with $\Omega_{ij}^k = \{\omega_{ij}^{k,s}\}_{s=1}^N$, where N would be 2^9 for a $8 \times 8 \times 8$ cube. We first compute Ω_{ij}^k for all the cubes $k = 1, \ldots, K$. Second, transform coefficients in every cube are normalized to carry the unit energy, and we store them in the set $\grave{\Omega}_{ij}^k = \{\grave{\omega}_{ij}^{k,s}\}_{s=1}^N$. Transform coefficients are defined as *corresponding coefficients* if $s_1 = s_2$, $u = p$, and $v = r$ for $\omega_{uv}^{\kappa_1,s_1}$ and $\omega_{pr}^{\kappa_2,s_2}$, where $\kappa_1 \neq \kappa_2$. Let us define a parameter ς_{ij}^s which describes the significance (or average energy) of the corresponding coefficients $\omega_{ij}^{k,s}$ in $\mathbf{C}_{ij}$. ς_{ij}^s is computed for each s as follows: $\varsigma_{ij}^s = \frac{1}{K} \sum_{k=1}^K \grave{\omega}_{ij}^{k,s}$. This results in $\varsigma_{ij}^s \in [0,1]$ values, where $\sum_{s=1}^N \varsigma_{ij}^s = 1$. Finally, we use an iterative forward selection algorithm [8] to form one significant subset for each corresponding cube set. We start with no coefficients and add them one by one based on ς_{ij}^s values, at each step adding the one that stores the most energy, until any further addition does not increase

the total energy in the subset or increases it only slightly. This generates a significant subset $\mathfrak{C}_{ij} = \{x_{ij}^l\}_{l=1}^{L}$,where $L << N$. Elements of $\mathfrak{C}_{ij}$ represent coordinates of the coefficients.

2.2.2 Statistical Features

The advantage of using statistical features compared to strategies assuming a priori parametric distribution is that we can distinguish fluctuations due to the fact that the assumed model may not be valid over the whole input space. A corresponding cube set $\mathbf{C}_{ij}$ is specified by the estimated base transform $T_{ij} \in \mathbf{T}$ and the significant subset $\mathfrak{C}_{ij}$. Let us define a function $\mathfrak{L}(l,k)$ that maps coordinates in $\mathfrak{C}_{ij}$ to actual coefficient values in the cubes of $\mathbf{C}_{ij}$: $\mathfrak{L}(x_{ij}^l,k) \rightarrow \omega_{ij}^{k,s}$ for $l = 1,\ldots,L$. One should note that the distribution of each significant coefficient may be different, and the estimation of each unique distribution is not a trivial process. Instead, we construct a maximum likelihood model by interpolating from the training instances. Let $\mathfrak{M}$ be a $I \times J$ matrix of the number of significant coefficients, with $\mathfrak{M} = \{m_{ij}\}_{i=1,j=1}^{I,J}$. m_{ij} is the number of significant coefficients in the significant subset $\mathfrak{C}_{ij}$ for the corresponding cube set $\mathbf{C}_{ij}$. Let $\tilde{\mathfrak{u}}$ denote a vector parameter called *unbiased mean*. $\tilde{\mathfrak{u}}_{ij}^k \in \mathbb{R}^{m_{ij}}$ is defined for the cube k in $\mathbf{C}_{ij}$. An element $\tilde{\mathfrak{u}}_{ij}^{k,l}$ of $\tilde{\mathfrak{u}}_{ij}^k$ is calculated as follows:

$$\tilde{\mathfrak{u}}_{ij}^{k,l} = \frac{1}{K-1}\sum_{\kappa=1}^{K}\mathfrak{L}(x_{ij}^l,\kappa) \text{ and } \kappa \neq k. \tag{2}$$

We compute $\tilde{\mathfrak{u}}_{ij}^{k,l}$ for $l = 1,\ldots,m_{ij}$. We represent each cube in $\mathbf{C}_{ij}$ using values of the coefficients in the significant subset. Let $\mathcal{C}_{ij}^k$ be a m_{ij}–dimensional vector of significant coefficient values for the cube c_{ij}^k. We define a deviation vector $\mathfrak{d}_{ij}^k \in \mathbb{R}^{m_{ij}}$, which describes the deviation of $\mathcal{C}_{ij}^k$ from its unbiased mean $\tilde{\mathfrak{u}}_{ij}^k$ as follows: $\mathfrak{d}_{ij}^k = |\mathcal{C}_{ij}^k - \tilde{\mathfrak{u}}_{ij}^k|$. We calculate standard deviation σ_{ij}^k and mean μ_{ij}^k of the elements of $\mathfrak{d}_{ij}^k$. Values of $\mathcal{C}_{ij}^k$, $\mathfrak{d}_{ij}^k$, σ_{ij}^k, and μ_{ij}^k for $k = 1,\ldots,K$ will be used to construct a maximum likelihood model for the relevant change detection.

2.3 Relevant Change Detection

Let us recall the given frame set $\mathbf{V} = \{F_1,\ldots,F_t,F_{t+1},\ldots,F_{\mathfrak{T}}\}$. We used the subset $\mathbf{V}_o = \{F_1,\ldots,F_t\}$ to estimate spatiotemporal signatures of the ordinary change patterns. We will analyze the changes in the rest of the frames. Let us assume that we initially process the first 8 frames $\{F_{t+1},\ldots,F_{t+8}\}$. As described in the Sec. 2.1.1, we group them to form the stack $\mathbf{S}_{test}$ and and decompose the stack into the cubes c_{ij}^{test}. We compute the transform coefficients of c_{ij}^{test} using the base transform $T_{ij} \in \mathbf{T}$ estimated for the corresponding cube set $\mathbf{C}_{ij}$. The significant subset $\mathfrak{C}_{ij}$ and $m_{ij} \in \mathfrak{M}$ are used along with the mapping function $\mathfrak{L}(l,k)$ to construct a $m_{ij}-$dimensional descriptor $\mathcal{C}_{ij}^{test}$ for each i and j. We then compute the deviation of $\mathcal{C}_{ij}^{test}$ from the training samples $\mathcal{C}_{ij}^k$: $\mathfrak{d}_{ij}^{k,test} = |\mathcal{C}_{ij}^k - \mathcal{C}_{ij}^{test}|$, for $k = 1,\ldots,K$. We calculate standard deviation $\sigma_{ij}^{k,test}$ and mean $\mu_{ij}^{k,test}$ of the elements of $\mathfrak{d}_{ij}^{k,test}$. Let X and Y be two random variables with means μ_X, μ_Y, standard deviations σ_X, σ_Y, and correlation coefficient ρ_{XY}. The bivariate inequality of Lal [14] is given by

$$P(\lambda_{L_X} < X < \lambda_{U_X}, \lambda_{L_Y} < Y < \lambda_{U_Y}) \geq P_{XY}, \text{ and} \tag{3}$$

$$P_{XY} = 1 - \frac{1}{2k_X^2k_Y^2}(k_X^2 + k_Y^2 + \sqrt{(k_X^2+k_Y^2)^2 - 4\rho^2k_X^2k_Y^2}), \tag{4}$$

where $\lambda_{L_X}+\lambda_{U_X}=2\mu_X$, $\lambda_{L_Y}+\lambda_{U_Y}=2\mu_Y$, $k_X=(\lambda_{U_X}-\lambda_{L_X})/2\sigma_X$, and $k_Y=(\lambda_{U_Y}-\lambda_{L_Y})/2\sigma_Y$. Eq. 3 gives a lower bound for the joint probability of the interval $[\lambda_{L_X},\lambda_{U_X}]$ around μ_X and the interval $[\lambda_{L_Y},\lambda_{U_Y}]$ around μ_Y for the random variables X and Y. We propose that if X and Y are dependent events, we expect P_{XY} to be large for the same interval $[\lambda_{L_X}=\lambda_{L_Y},\lambda_{U_X}=\lambda_{U_Y}]$ around μ_X and μ_Y for X and Y. Accordingly, we define a symmetric interval $\lambda_{L_X}=\lambda_{L_Y}=(\mu_X+\mu_Y)/2-2*(\sigma_X+\sigma_Y)$ and $\lambda_{U_X}=\lambda_{U_Y}=(\mu_X+\mu_Y)/2+2*(\sigma_X+\sigma_Y)$ for X and Y. We can use the value of P_{XY} to estimate the likelihood of X and Y to be independent random events. In our change detection setting, if $\mathfrak{d}_{ij}^{k,test}$ is found to be independent from $\mathfrak{d}_{ij}^{k}$, we can conclude that there is a relevant change in c_{ij}^{k}.

Let elements of the deviation vector $\mathfrak{d}_{ij}^{k}$ represent the values of the random variable X with the mean μ_{ij}^{k} and the standard deviation σ_{ij}^{k}. Let elements of the deviation vector $\mathfrak{d}_{ij}^{k,test}$ represent the values of the random variable Y with the mean $\mu_{ij}^{k,test}$ and the standard deviation $\sigma_{ij}^{k,test}$. Using Eq. 4, we can compute a joint probability P_{XY}^{k} for training stack k. Because Eq. 3 provides a lower bound but not the actual probability, we can compute P_{XY} for all training samples by $P_{XY}=\frac{1}{K}\sum_{k=1}^{K}P_{XY}^{k}$.

2.3.1 Change Detection at Pixel Resolution

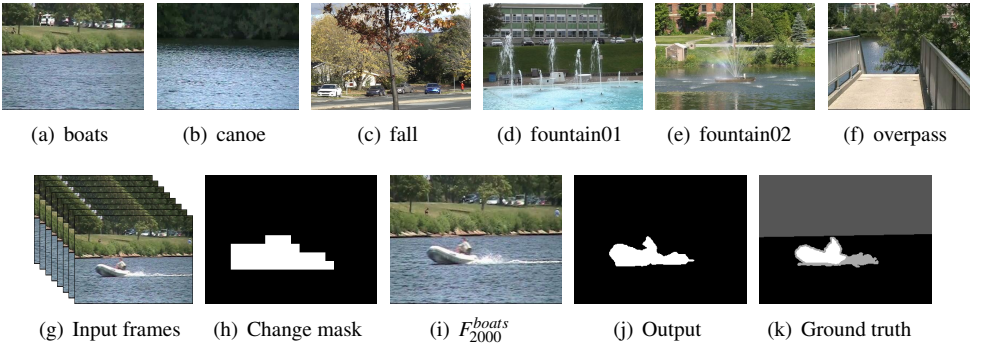
When a change having spatiotemporal signature different from ordinary change patterns is detected in a cube, the proposed method suggests that there may be relevant change within the regions comprising the cube. At the frame level, this corresponds to a two-dimensional projection of spatiotemporal changes within the stack of 8 consecutive frames (Fig. 3 (b)). This summary image is called *binary change mask*, where 1 and 0 indicate the relevant and ordinary change, respectively. We use the change mask to analyze the mid-frames in the stack to avoid large blocking artifacts. For pixel-based detection, we apply the two dimensional version of the estimated base transform within a window around pixels having the value of 1 in the change mask and follow the same approach explained above. The block-based nature of our approach may cause noise at the pixel level. To overcome this limitation, the resulting change mask is assumed to be a Markov random field. Each pixel is labeled as either relevant change or ordinary change based on the probability maximization achieved by the Markov Random Field regularization [24]. We repeat the process by sliding the frame stack in order to evaluate all frames.

3 Experiments

We obtained 6 test videos from *the dynamic background category* on *ChangeDetection.net* [9]. Videos contain scenes with highly varying elements in the background such as shimmering water, fountains, and blowing trees. The dataset includes a comprehensive set of annotated ground truth change areas to enable a precise quantitative evaluation. Videos are are named as *boats*, *canoe*, $fall$, $fountain01$, $fountaion02$, and $overpass$ (Fig 3 (a)-(f)).

3.1 Base Transform Estimation

The proposed method requires the estimation of suitable base transform for different types of ordinary change patterns. In Table 1, we present the ratio of regions modeled by different base transforms. The type of the estimated base transform can also be a good descriptor for



(a) boats (b) canoe (c) fall (d) fountain01 (e) fountain02 (f) overpass

(g) Input frames (h) Change mask (i) F_{2000}^{boats} (j) Output (k) Ground truth

Figure 3: Images (a)-(f): examples of the dynamic backgrounds in the video dataset. In (h), we present the binary change mask for the 8 consecutive frames in (g). The change mask is a two-dimensional projection of spatiotemporal changes in these 8 frames. In (i), (j), and (k), we present input F_{2000}^{boats}, relevant changes detected, and the ground truth. The gray levels in (k) are 0:*ordinary change*, 255:*relevant change*, 85:outside region of interest, and 170:unknown motion [9]. Our goal is to detect pixels labeled as *relevant change*.

Video	Dynamic Background Description	Estimated Base Transform (%)		
		DCT	WHT	SL
boats	shimmering water, blowing bushes	63.75	2.16	34.09
canoe	shimmering water, blowing trees	61.00	1.66	37.34
fall	blowing trees	55.48	8.92	35.60
*fountain*01	fountain, shimmering water	27.00	8.44	64.56
*fountain*02	fountain, shimmering water, blowing bushes	36.00	14.81	49.19
overpass	shimmering water, blowing trees	52.58	5.16	42.26

Table 1: DCT, WHT, and SL are the base transforms: discrete cosine, Walsh-Hadamard, and Slant. We present the results of the transform estimation for the backgrounds in the six test videos. For example, in video *boats* 63.75% of the frame region is modeled by DCT.

the scene content. For example, DCT is known to have strong energy compaction property when applied to natural images [29]. In the videos *fountain*01 and *fountain*02, there is a notable decrease in the overall use of DCT. This is because of the fountains that jet water into the air, causing artificial ordinary change patterns. This result stresses the importance of employing different base transformations with complementary basis vectors.

3.2 Quantitative Evaluation of the Relevant Change Detection

A precise validation of a change detection method requires ground truth at pixel resolution. Let p_{rc} denote a pixel in a region of relevant change, and let p_{oc} denote a pixel in a region of ordinary change. If a change detection method labels p_{rc} correctly, this case is called *true positive* (TP), and *false negative* (FN), otherwise. If a change detection method labels p_{oc} as ordinary change, this case is called *true negative* (TN), and *false positive* (FP), otherwise. For the entire test set, a joint probability value P_{XY} less than 0.33 is considered as an evidence that there is a relevant change. Table 2 shows change detection results at the pixel level.

	boats	canoe	fountain01	fountain02	overpass	fall	Average
Number of Test Frames	6,100	390	785	1,000	2,001	3,001	
Specificity (%)	99.961	99.742	99.495	99.952	99.996	99.962	99.833
Accuracy (%)	99.820	99.592	99.387	99.936	99.987	99.889	99.769

Table 2: The proposed method is able to identify ordinary changes with 99.833% specificity.

ChangeDetection.net uses seven metrics to rank different change detection methods. Let us here present the two of the metrics, Recall (Re) and Precision (Pr), to compare our method to the other methods under *the dynamic background category*. The details of all the metrics and the ranking are presented in [9]. Re and Pr are given by: $Re = \frac{TP}{TP+FN}$ and $Pr = \frac{TP}{TP+FP}$. We present the comparison of our method to the three methods having the highest ranking for *the dynamic background category* on *ChangeDetection.net* in Table 3.

Method	boats		canoe		fountain01		fountain02		overpass		fall		Average	
(*Ranking*)	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr
[ ] (*4.71*)	0.63	0.92	0.95	0.79	**0.99**	0.68	0.80	0.50	**0.96**	0.86	**0.99**	0.92	0.89	0.78
[ ] (*5.71*)	0.75	0.82	0.89	0.92	0.82	**0.90**	0.63	0.15	0.89	0.93	0.94	0.87	0.82	0.76
[ ] (*6.14*)	0.53	**0.97**	0.79	**0.99**	0.91	0.89	**0.86**	0.40	0.86	**0.98**	0.70	0.92	0.77	**0.86**
Ours (*2.14*)	**0.78**	0.93	**0.96**	0.93	0.93	0.77	0.81	**0.58**	**0.96**	**0.98**	0.95	**0.97**	**0.90**	**0.86**

Table 3: In this table, we compare Recall (Re) and Precision (Pr) values of the top-three methods under *the dynamic background category* on *ChangeDetection.net* to ours. On the far left, we provide the rankings of each method. The overall ranking of a method across seven metrics is computed by taking the average of its ranking for each metric. The overall ranking of our method is 2.14, and the proposed method outperforms other 23 methods demonstrated for *dynamic background category* (ranking results retrieved on June 2013).

4 Conclusion

We have presented a method for the detection of relevant changes in videos with highly varying elements in the scene background. In dynamic backgrounds, the distinction of relevant changes from ordinary changes requires exploiting spatial and temporal relationships. We used orthogonal linear transforms to capture spatiotemporal signatures of the local ordinary change patterns. Then, the framework employs these signatures in the detection of relevant changes. The major limitation of our method is that estimating base transforms requires a set of frames without relevant changes. This a common issue for data-driven approaches. Another limitation arises from cube-based computations, which may cause blocking artifacts. Compared to other methods demonstrated on the same test videos, our method shows significant improvement in change detection results.

5 Acknowledgments

This work was supported in part by the US Department of Justice 2009-MU-MU-K004. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of our sponsors.

References

[1] Sos Agaian, Khaled Tourshan, and Joseph P. Noonan. Generalized parametric Slant-Hadamard transform. *Signal Processing*, 84(8):1299 – 1306, 2004.

[2] A. Aggoun and M. Tabit. Data Compression of Integral Images for 3D TV. In *3DTV Conference, 2007*, pages 1–4, 2007.

[3] N. Ahmed, T. Natarajan, and K.R. Rao. Discrete cosine transform. *Computers, IEEE Transactions on*, C-23(1):90–93, 1974.

[4] Nasir Ahmed and KamisettyRamamohan Rao. Walsh-hadamard transform. In *Orthogonal Transforms for Digital Signal Processing*, pages 99–152. Springer Berlin Heidelberg, 1975.

[5] Nasir U. Ahmed and K. Ramamohan Rao. *Orthogonal Transforms for Digital Signal Processing*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1975.

[6] Ethem Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2nd edition, 2010.

[7] Rui Bao, Tianqi Zhang, Fangqing Tan, and Y.E. Wang. Semi-fragile watermarking algorithm of color image based on slant transform and channel coding. In *Image and Signal Processing (CISP), 2011 4th International Congress on*, volume 2, pages 1039–1043, 2011.

[8] Hyun Sung Chang and Kyeongok Kang. A compressed domain scheme for classifying block edge patterns. *Image Processing, IEEE Transactions on*, 14(2):145–151, 2005.

[9] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection.net: A new change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 1–8, 2012.

[10] Hakan Haberdar and Shishir K Shah. Disparity Map Refinement for Video Based Scene Change Detection Using a Mobile Stereo Camera Platform. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3890–3893. IEEE, 2010.

[11] Tom SF Haines and Tao Xiang. Background subtraction with dirichlet processes. In *Computer Vision–ECCV 2012*, pages 99–113. Springer, 2012.

[12] Mansour Moniri Ismail, Mohamed Hamed and Chibelushi Claude Chilufya. Object segmentation using full-spectrum matching of albedo derived from colour images, 12 2011.

[13] Seon Joo Kim, G. Doretto, J. Rittscher, P. Tu, N. Krahnstoever, and M. Pollefeys. A model change detection approach to dynamic scene modeling. In *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on*, pages 490–495, 2009.

[14] DN Lal. A note on a form of tchebycheff's inequality for two or more variables. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 15(3):317–320, 1955.

[15] Dar-Shyang Lee. Effective Gaussian Mixture Learning for Video Background Subtraction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):827–832, May 2005.

[16] Seong-Whan Lee, Young-Min Kim, and Sung Woo Choi. Fast scene change detection using direct feature extraction from MPEG compressed videos. *Multimedia, IEEE Transactions on*, 2(4):240–254, 2000.

[17] Stan Z Li. *Markov random field modeling in image analysis*. Springer, 2009.

[18] Andrew Lingg, E Zelnio, F Garber, and B Rigling. Image sequence change detection via sparse representations. In *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, pages 2028–2032. IEEE, 2010.

[19] Elmustapha Ait Lmaati, Ahmed El Oirrak, Mohammed Najib Kaddioui, Abdellah Ait Ouahman, and Mohammed Sadgal. 3d model retrieval based on 3d discrete cosine transform. *Int. Arab J. Inf. Technol.*, 7(3):264–270, 2010.

[20] Ashutosh Morde, Xiang Ma, and Sadiye Guler. Learning a background model for change detection. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 15–20. IEEE, 2012.

[21] Tae-Hyun Oh, Joon-Young Lee, and In So Kweon. Real-time motion detection based on discrete cosine transform. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2381–2384. IEEE, 2012.

[22] I.-M. Pao and Ming-Ting Sun. Modeling dct coefficients for fast video encoding. *Circuits and Systems for Video Technology, IEEE Transactions on*, 9(4):608–616, 1999.

[23] D.H. Parks and S.S. Fels. Evaluation of background subtraction algorithms with post-processing. In *Advanced Video and Signal Based Surveillance, 2008. AVSS '08. IEEE Fifth International Conference on*, pages 192–199, 2008.

[24] Rasmus R Paulsen and Klaus B Hilger. Shape modelling using markov random field restoration of point correspondences. In *Information Processing in Medical Imaging*, pages 1–12. Springer, 2003.

[25] Theodor D. Popescu. Blind separation of vibration signals and source change detection - Application to machine monitoring. *Applied Mathematical Modelling*, 34(11):3408 – 3421, 2010.

[26] W. Pratt, Wen-Hsiung Chen, and L. Welch. Slant transform image coding. *Communications, IEEE Transactions on*, 22(8):1075–1093, 1974.

[27] Yue Qin, Xiaolin Tian, and Shaowei Xia. Research on video watermark algorithm based on slant transform. In *Image and Signal Processing (CISP), 2011 4th International Congress on*, volume 1, pages 31–33, 2011.

[28] R.J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *Image Processing, IEEE Transactions on*, 14(3):294–307, 2005.

[29] David Salomon. *Data Compression: The Complete Reference*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[30] Ishwar K Sethi and Nilesh Patel. A statistical approach to scene change detection. In *Storage and retrieval for image and video databases*, volume 3, pages 329–339, 1995.

[31] James V Stone. Object recognition: View-specificity and motion-specificity. *Vision Research*, 39(24):4032–4044, 1999.

[32] Jan Verbesselt, Rob Hyndman, Achim Zeileis, and Darius Culvenor. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12):2970 – 2980, 2010.

[33] Quoc C Vuong and Michael J Tarr. Structural similarity and spatiotemporal noise effects on learning dynamic novel objects. 2006.

[34] Erol Yeniaras, NikhilV. Navkar, AhmetE. Sonmez, DipanJ. Shah, Zhigang Deng, and NikolaosV. Tsekos. Mr-based real time path planning for cardiac operations with transapical access. In *Medical Image Computing and Computer-Assisted Intervention*, volume 6891 of *LNCS*, pages 25–32. Springer Berlin Heidelberg, 2011.

[35] Wei Zeng, Jun Du, Wen Gao, and Qingming Huang. Robust moving object segmentation on h. 264/avc compressed video using the block-based mrf model. *Real-Time Imaging*, 11(4):290–299, 2005.