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

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 (Fig. 1).

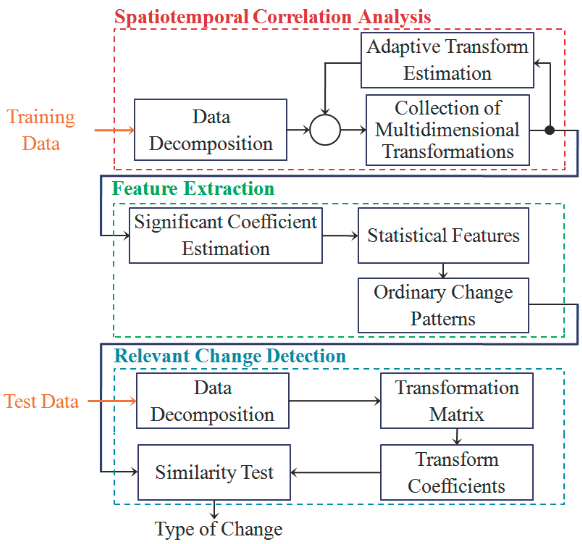

Figure 1: Our algorithm consists of 3 main blocks. Given a video containing only ordinary changes, our goal is to find representations where the spatiotemporal features of the ordinary changes can be captured. Then, we use the training examples to extract spatiotemporal signatures of the ordinary change patterns. Finally, we estimate the existence of relevant change in a test input by interpolating from the training samples.

In dynamic environments, 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 on the dynamic background of the scene. 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 [6], 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 proposed framework makes use of a collection of orthogonal linear transforms to capture spatiotemporal signatures of local ordinary change patterns and subsequently employ them in the detection of relevant changes. We employ three orthogonal linear transforms as the base transforms: i) discrete cosine transform (DCT) [2], ii) Walsh-Hadamard transform (WHT) [1], and iii) Slant transform (ST) [5]. 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. Our goal is to estimate the most suitable transform for each local ordinary change pattern.

Given a set of frames containing only the ordinary changes, 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. Every stack is composed of $8 \times 8 \times 8$ blocks called as *cubes*. To estimate the most suitable transform to extract the spatial signature of a cube, we define the compactness metric ξ_s as follows:

$$\xi_s = \sum_{i=1}^{N} \left(\frac{1}{N} - \mathfrak{p}_i \right)^2, \text{ and } \xi_s \quad [0, 1 - \frac{1}{N}] \tag{1}$$

where $\mathfrak{p}_i$ represents a pixel in a cube $\mathbb{C}$ and N is the number of the pixels in the cube $\mathbb{C}$. The transform having the largest compactness coefficient value is considered as the most suitable for the cube. Given a test cube, we construct a maximum likelihood model by using the bivariate inequality of Lal [4] and interpolating from the training instances:

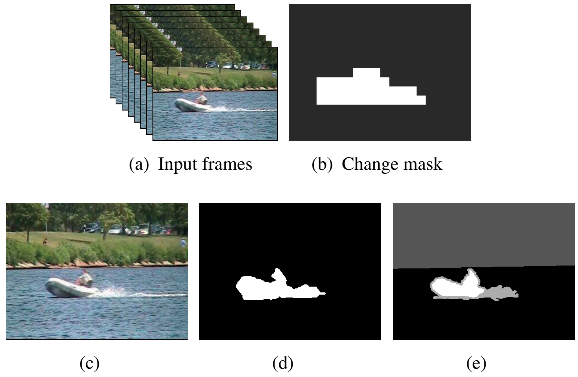

Figure 2: In (b), we present the binary change mask estimated for the 8 consecutive frames given in (a). The change mask is a two-dimensional projection of spatiotemporal changes in these 8 frames. In (c), (d), and (e), we present an input image, relevant changes detected, and the ground truth. The gray levels in (e) are 0:*ordinary change*, 255:*relevant change*, 85:outside region of interest, and 170:unknown motion [3]. Our goal is to detect pixels labeled as *relevant change*.

$$P(\lambda_{L_X} < X < \lambda_{U_X}, \lambda_{L_Y} < Y < \lambda_{U_Y}) \geq P_{XY}, \text{ and} \tag{2}$$

$$P_{XY} = 1 - \frac{1}{2k_X^2 k_Y^2}\left(k_X^2 + k_Y^2 + \sqrt{(k_X^2 + k_Y^2)^2 - 4\rho^2 k_X^2 k_Y^2}\right), \tag{3}$$

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. 2 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. At the frame level, this corresponds to a two-dimensional projection of spatiotemporal changes within the stack of 8 consecutive frames (Fig. 2 (b)). This summary image is called *binary change mask*, where 1 and 0 indicate the relevant and ordinary change, respectively. We apply two-dimensional version of the framework to obtain the relevant change at the pixel level.

The use of this framework is demonstrated in a variety of videos with highly dynamic backgrounds including lakes, pools, and roads. The major limitation of our method is that estimating base transforms requires a set of frames without relevant changes. 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.

[1] N. Ahmed and K.R. Rao. Walsh-Hadamard transform. In *Orthogonal Transforms for DSP*, pages 99–152. Springer, 1975.

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

[3] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection.net: A new change detection benchmark dataset. In *CVPRW 2012*, pages 1–8, 2012.

[4] D.N. Lal. A note on a form of tchebycheff's inequality for two or more variables. *The Indian J. of Statistics*, 15(3):317–320, 1955.

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

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