Quantitative Imaging Laboratory

Change Detection in Dynamic Scenes using Local Adaptive Transform

Hakan Haberdar and Shishir K. Shah

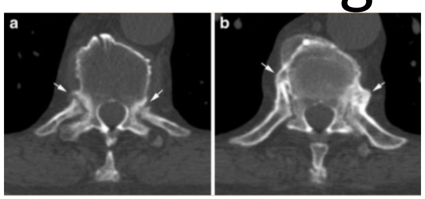
British Machine Vision Conference 2013
September 10, 2013

Motivation

- Identifying changes of interest in videos is a problem in various application domains

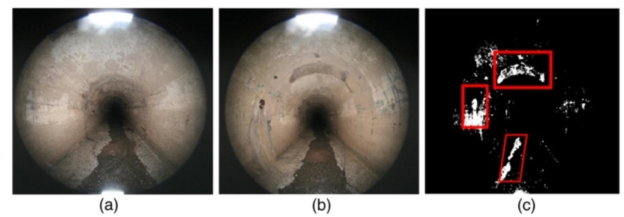
- Video Surveillance

- Medical Diagnosis



[2]

- Condition Assessment



[1]

- Remote Sensing

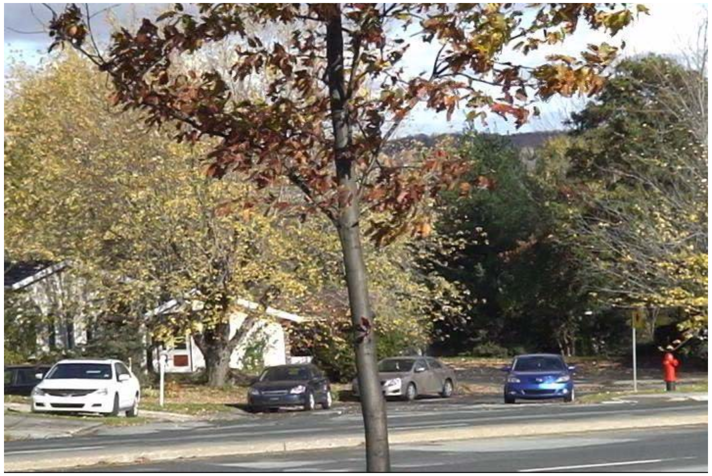
[3]

- Driver Assistance Systems

[1] W. Guo, L. Soibelman, J.H. Garrett Jr., Automated defect detection for sewer pipeline inspection and condition assessment, Automation in Construction, 2009
[2] http://openi.nlm.nih.gov/detailedresult.php?img=3259357_13244_2010_61_Fig15_HTML&req=4
[3] http://huettichs.wordpress.com/2011/09/01/synthesis-summary-of-urban-remote-sensing-at-dlr/

Challenges

- What are changes and are they all interesting?
- To establish a clear distinction between what is relevant and what is not is a very challenging task.
- Illumination variations
  - Conditions of the data acquisition
- Several altering elements in the background may cause false alarms:
  - Shimmering or wavy water
  - Swaying trees
  - Water fountain



Green Rectangle: An example of a relevant change
Red Arrows: Examples of altering elements in the background, not a type of desired change

Ordinary Change vs. Relevant Change

- We categorize the change into two main classes:
  - Ordinary Change: "Recurrent elements and changes pertaining to the dynamic background".

The entire frame region, depicted as **the blue region**, is considered as a region of ordinary change.

  - Relevant Change: "Alterations that do not conform to the expected pattern of ordinary change".

The boat and the crew are considered as a region of relevant change.

- If we can model the ordinary change patterns, we can subsequently employ the model for the detection of relevant changes.

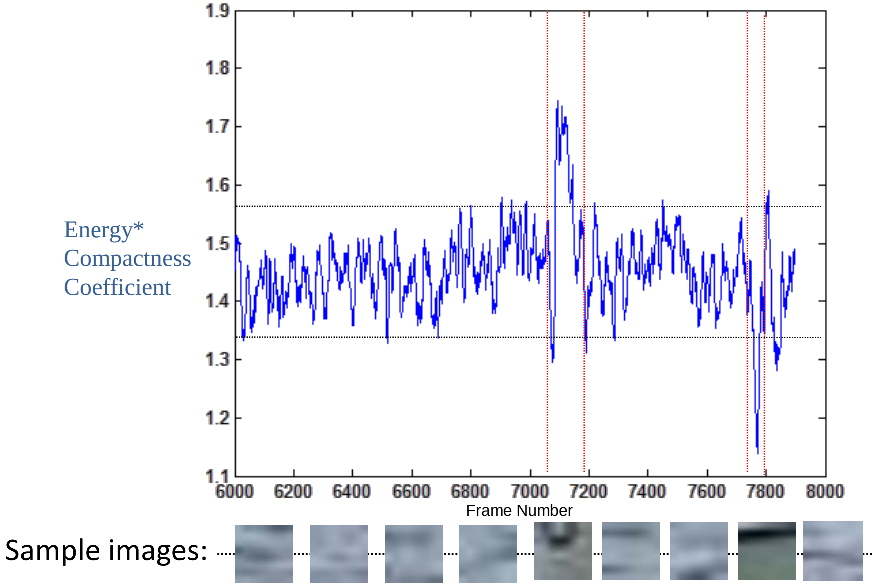
The Blue Region: Ordinary Change.

Ordinary Change Patterns

- Ordinary change patterns are typically correlated in space and/or time among a set of consecutive frames.
- This correlation stems from the repetitive nature and induces spatiotemporal signatures specific to each local ordinary change pattern [1].

We present the energy compactness values of the spatiotemporal signature of the depicted local region .

Energy Compactness Values of the Spatiotemporal Signature of a 8x8x8 Local Region in a Frame Sequence



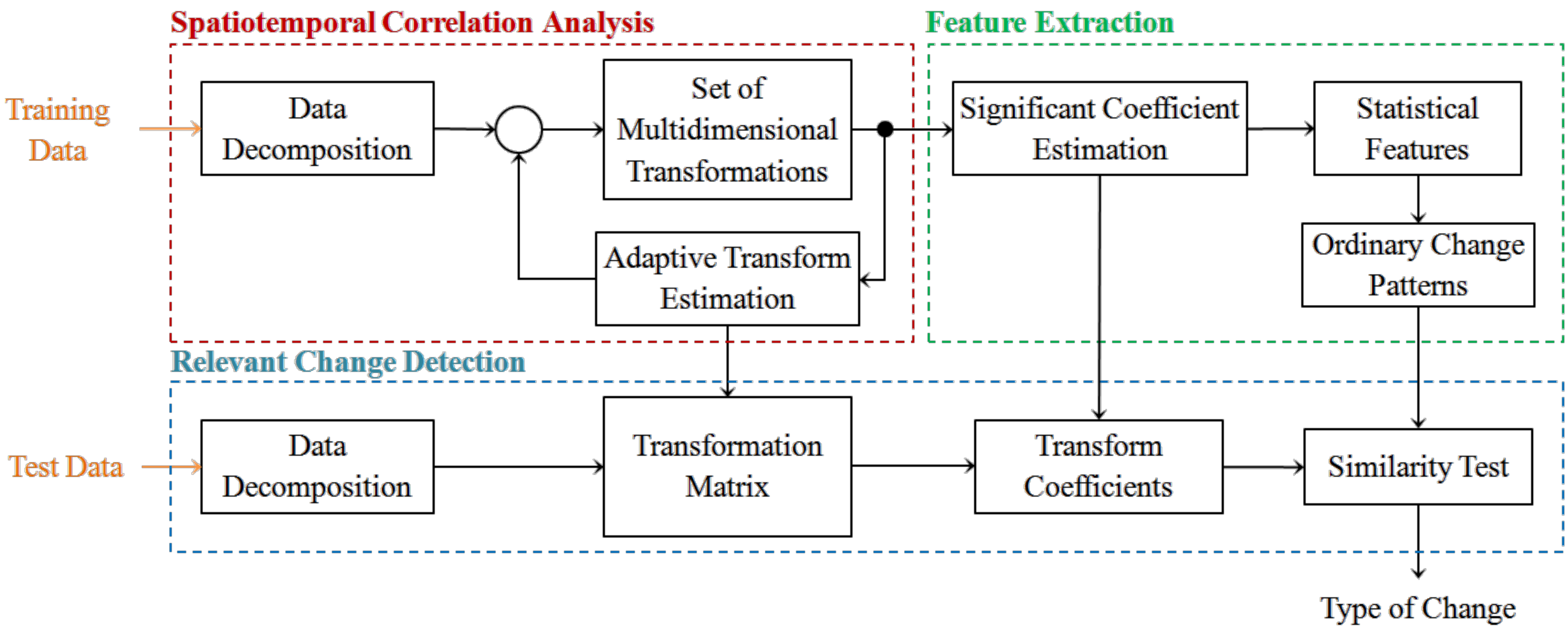
$$\xi_s = \sum_{i=1}^{N} \left(\frac{1}{N} - \dot{\omega}_i \right)^2$$

*Energy compactness coefficient, ξ_s, describes how the energy distributes in a given 8x8x8x data matrix, where N is the number of elements and $\dot{\omega}_i$ is a normalized element in the matrix.

[1] J. V. Stone. Object recognition using spatiotemporal signatures. Vision Research, 1998.

Proposed Framework

- We propose a framework that makes use of the spatiotemporal signatures to discriminate ordinary changes from relevant changes.

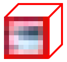


Spatiotemporal Signature Extraction

- Processing images one-by-one in the image pixel plane is not suitable for extracting spatiotemporal features[1]. Instead, we should capture spatiotemporal signatures in a three-dimensional transform space.
- Many approaches have been investigated to extraction of spatiotemporal signatures [2-3].
- We leverage linear transforms with an ability to decorrelate data and realize compact representations
  - A transform is considered as suitable for a local ordinary change pattern if the transform domain provides a compact representation of the local ordinary change pattern.

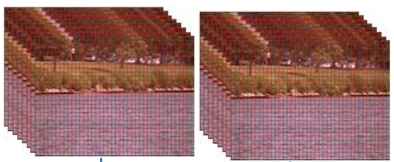
[1] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. Digital image processing using MATLAB, volume 2. Gatesmark Publishing Tennessee, 2009.
[2] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In VSPET Surveillance, 2005
[3] N. U. Ahmed and K. R. Rao. Orthogonal Transforms for Digital Signal Processing. Springer-Verlag 1975

Data Decomposition

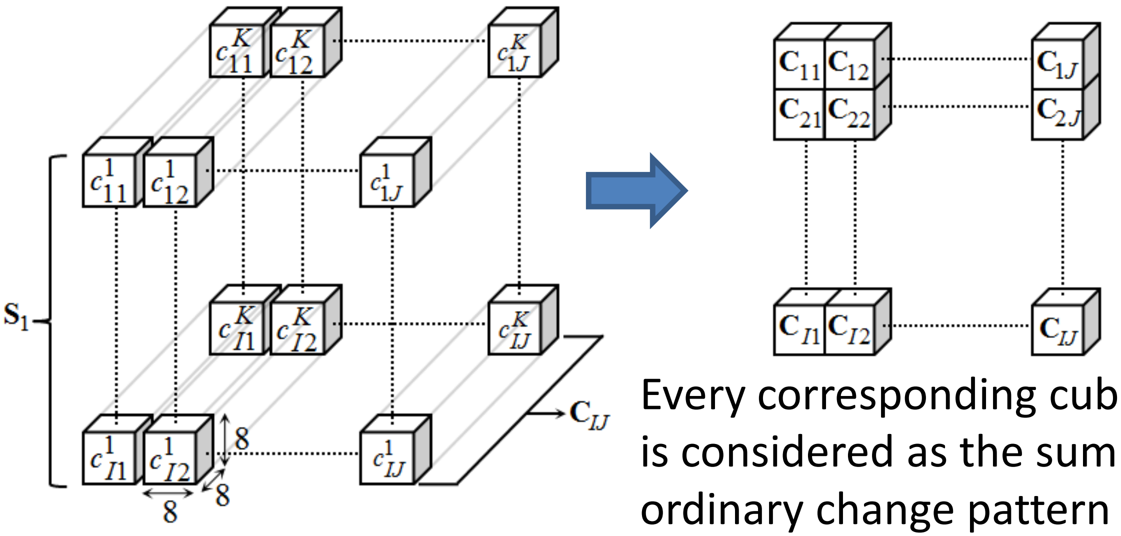
- Given a sequence of frames including only ordinary change patterns
- Divide each frame into 8 by 8 pixels regions in order to improve the localized correlation
- Group each 8 consecutive frames to form a stack.
- Each stack is composed of 8x8x8 cubes: 
- We perform a further grouping and collect the corresponding cube elements in different stacks and form the corresponding cube sets.

$$Video = \{ F_1, \ldots, F_8, \; F_9, \ldots, F_{18}, \ldots \}$$

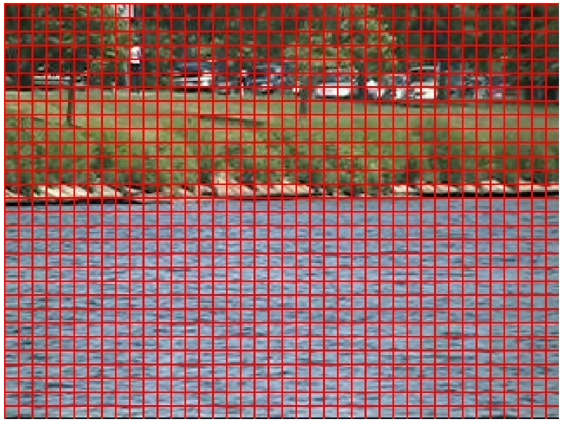
One stack:



$$Video = \{ \underbrace{F_1, \ldots, F_8,}_{S_1} \; \underbrace{F_9, \ldots, F_{18},}_{S_2} \; \ldots \} \quad \ldots$$



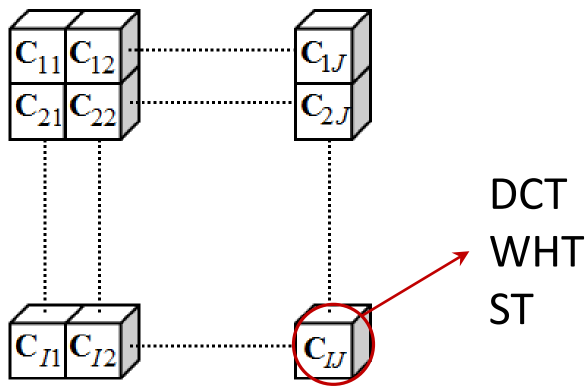
Every corresponding cube set C_{ij} is considered as the summary of ordinary change pattern in the local region at i and j.



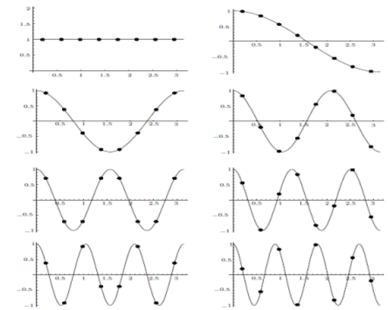
Pixel Grid

Linear Transforms

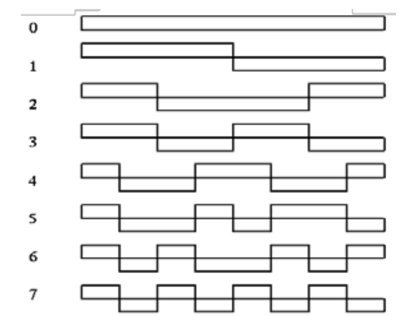
- Linear transforms
  - Discrete cosine transform (DCT)
  - Walsh-Hadamard transform (WHT)
  - Slant transform (ST)
- Why?
  - DCT, WHT, and ST have distinct basis vectors.
- We need to estimate a suitable transform for each corresponding cube set.



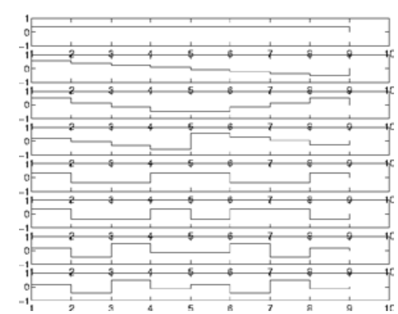
- If a transform can compact the energy of the input in few transformed values, the transform can be considered as the most suitable one.



DCT: Sinusoidal waveforms



WHT: Rectangular waveforms



SL: Sawtooth waveforms

Base Transform Estimation

- We use a energy compaction criterion, called compactness coefficient:

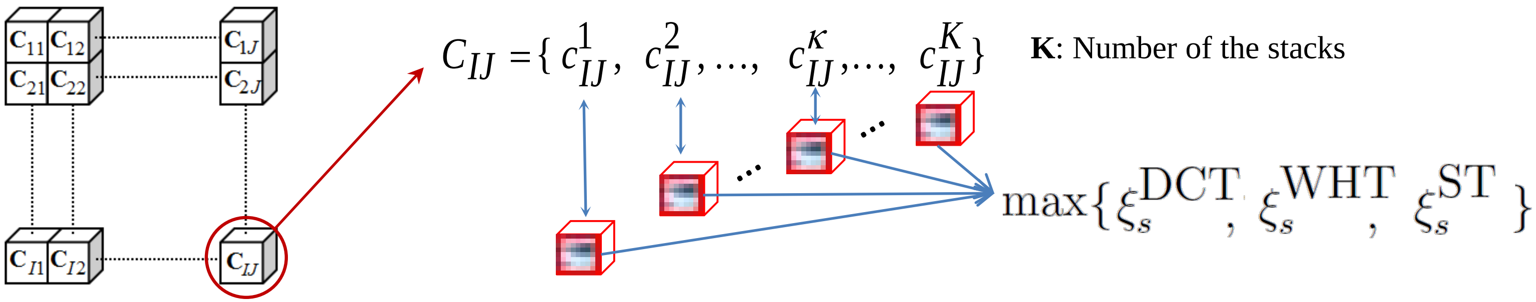
$$\boxed{\xi_s \triangleq \sum_{i=1}^{N}\left(\frac{1}{N}-\grave{\omega}_i\right)^2} \qquad \xi_s \in \left[0, 1-\frac{1}{N}\right]$$

$\{\omega_1, \ldots, \omega_N\}$: Transform coefficients

$\grave{\omega}_i = \frac{\omega_i}{E}$: Normalized transform coefficients

E: Total energy of the coefficients

- Given a corresponding cube set, we compute compactness coefficients of each cube in the corresponding cube set for DCT, WHT, and ST.
  - Transform with the largest compactness coefficient value is chosen as the most suitable transform for the cube set



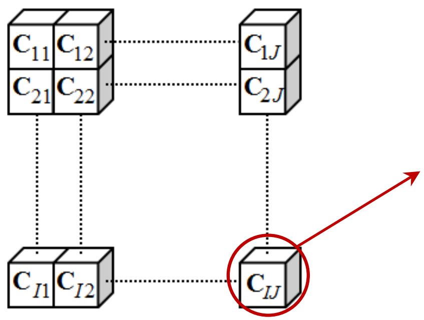
- Do we need all 8x8x8 transform coefficient?
  - Coefficients may be specific to the change pattern
  - Significant Coefficient Subset

Spatiotemporal Signature : {Estimated transform, Significant Coefficient Subset}

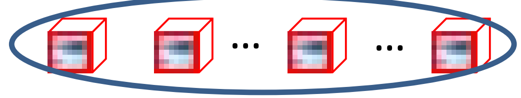
Signature Modeling with Statistical Properties

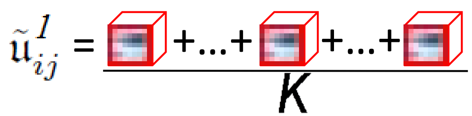
- Mean: $$\tilde{u}^{l}_{ij} = \frac{1}{K} \square \, L(x^{l}_{ij}, k)$$

$\mathbf{x}^{\mathbf{l}}_{\mathbf{ij}}$: Index of the significant $\mathbf{l}$ coefficient of the corresponding cube set $\mathbf{C}_{\mathbf{ij}}$, $\boldsymbol{L(x^{l}_{ij},\kappa)}$: value of the significant coefficient of the cube element $\mathbf{c}^{\kappa}_{\mathbf{ij}}$ in $\mathbf{C}_{\mathbf{ij}}$.

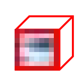


$$C_{IJ} = \{ c^{1}_{IJ},\ c^{2}_{IJ}, \ldots,\ c^{\kappa}_{IJ}, \ldots,\ c^{K}_{IJ} \}$$





- Compute the deviation d^{κ}_{ij} of each cube element c^{κ}_{ij} from its mean:



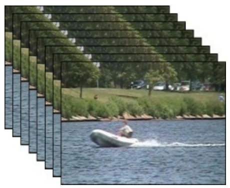
$$d^{\kappa}_{ij} = | \ c^{\kappa}_{ij} - \tilde{u}^{l}_{ij} \ | \text{ for } \kappa = 1, \ldots, K$$

- We then compute μ^{κ}_{ij} and σ^{κ}_{ij} of each d^{κ}_{ij} for κ=1,…,K
- Using the significant coefficient subset, μ^{κ}_{ij}, σ^{κ}_{ij} , and d^{κ}_{ij}, we construct a model for the relevant change detection.

Relevant Change Detection

- Given a set of test frames $V^{test}=\{F^{test}_1,\ldots, F^{test}_8\}$, we stack/decompose the frames and extract the cube elements for each ij using the significant subsets:

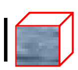
V^{test} =



stack and decompose → cube elements c_{ij}^{test} for each location ij

- We compute the deviation of each cube element c^{test}_{ij} in the test stack from the training cube samples c^{κ}_{ij} for $\kappa=1,\ldots,K$ in the corresponding cube set C_{ij}:

$$C_{ij} = \{ c_{ij}^{1}, \ \ldots, \ c_{ij}^{\kappa}, \ldots, \ c_{ij}^{K} \}$$

$d^{\kappa,\, test}_{ij} = |$  $- c^{test}_{ij}\,|$

c_{ij}^{κ}

- We compute $\mu^{\kappa,test}_{ij}$ and $\sigma^{\kappa,test}_{ij}$ of each $d^{\kappa,\, test}_{ij}$ for $\kappa=1,\ldots,K$
- We can now detect relevant change through a significance test

Relevant Change Detection

- The null hypothesis H_0 is characterized using the training samples which are assumed to have only ordinary change patterns.
- For H_0, let X, 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 [1]:

$$P(\lambda_{L_X} < X < \lambda_{U_X}, \lambda_{L_Y} < Y < \lambda_{U_Y} | \mathcal{H}_0) \geq P_{XY}, \text{ and}$$

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

where $\lambda_{L_X} + \lambda_{U_X} = 2\mu_X$, $\lambda_{L_Y} + \lambda_{U_Y} = 2\mu_Y$,

$$k_X = (\lambda_{U_X} - \lambda_{L_X})/2\sigma_X, \text{ and } k_Y = (\lambda_{U_Y} - \lambda_{L_Y})/2\sigma_Y.$$

- P_{XY} gives a lower bound for the joint probability in the interval [λ_{Lx}, λ_{Ux}] around μ^X and the interval [λ_{Ly}, λ_{Uy}] around μ^Y for X and Y .
- We propose that if X and Y are dependent events, we expect P_{XY} to be large for the same interval [$\lambda_{Lx}= \lambda_{Ly}$, $\lambda_{Ux}= \lambda_{Uy}$] around μ^X and μ^Y.
- Accordingly, we define a symmetric interval

$$\lambda_{L_X} = \lambda_{L_Y} = (\mu_X + \mu_Y)/2 - 2 * (\sigma_X + \sigma_Y)$$
$$\lambda_{U_X} = \lambda_{U_Y} = (\mu_X + \mu_Y)/2 + 2 * (\sigma_X + \sigma_Y)$$

- We can use the value of P_{XY} to estimate the likelihood of X and Y to be dependent random events.

[1] D. Lal. A note on a form of Tchebycheff's inequality for two or more variables. The Indian Journal of Statistics, 1955.

Relevant Change Detection

- In our change detection setting:

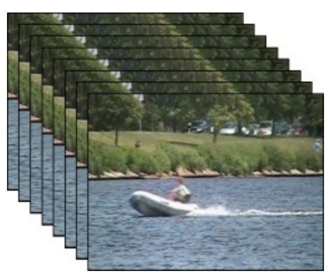
$X: d^{\kappa,\ test}_{ij}$ $Y: d^{\kappa}_{ij}$	$\mu^X: \mu^{\kappa,test}_{ij}$ $\mu^Y: \mu^{\kappa}_{ij}$	$\sigma^X: \mu^{\kappa,test}_{ij}$ $\sigma^Y: \mu^{\kappa}_{ij}$

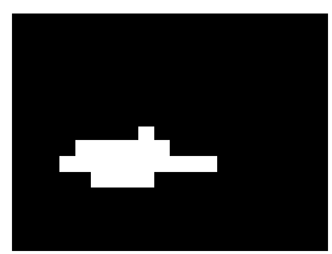
P^{κ}_{XY}

- If $d^{\kappa,\ test}_{ij}$ is found to be independent from d^{κ}_{ij}, one can deduce that there is a relevant change in the test cube c^{test}_{ij}.
- We can compute a joint probability P^{κ}_{XY} given training samples $\kappa=1,\ldots,K$. Then, we can compute P_{XY}:

$$P_{XY} = \frac{1}{K}\sum_{\kappa=1}^{K} P^{\kappa}_{XY}$$

- This process is repeated for each cube. At the frame level, this corresponds to a two dimensional projection of spatiotemporal changes within the stack of 8 consecutive frames:



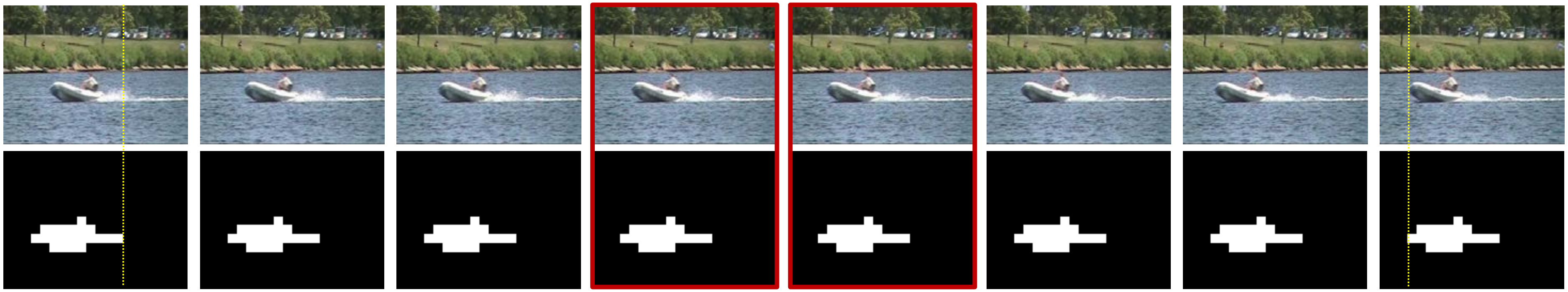


Binary Change Mask

[1] D. Lal. A note on a form of Tchebycheff's inequality for two or more variables. The Indian Journal of Statistics, 1955.

Relevant Change Detection

- Use the binary change mask to analyze the mid-frames of the test stack to avoid large blocking artifacts:



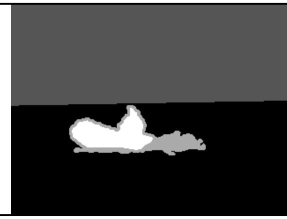
- For smoothing the binary mask, we apply spatial regularization.

1	2	3
8		4
7	6	5

Experiments

- We obtained 6 test videos from the **dynamic background category** on ChangeDetection.net. The test videos were captured in outdoor scenes where the background has several altering elements (i.e., ordinary changes) that may cause false alarms.
- ChangeDetection.net provides a comprehensive set of annotated ground truth change areas to enable a precise quantitative evaluation:

Ground truth image:



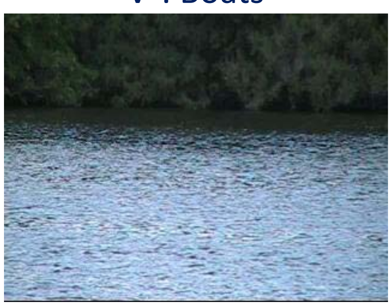
```
  0: Ordinary change
255: Relevant change
 85: outside region of interest
170: unknown motion
```

- We used **20 stacks** (160 frames) to extract spatiotemporal signatures of the ordinary change patterns:

V^1: Boats

V^3: Fall

V^5: Fountain02



V^2: Canoe

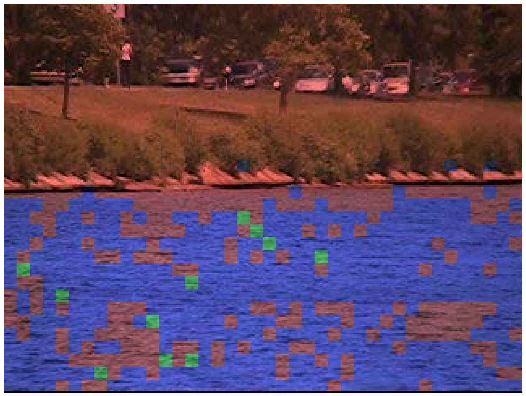
V^4: Fountain01

V^6: Overpass

- 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.

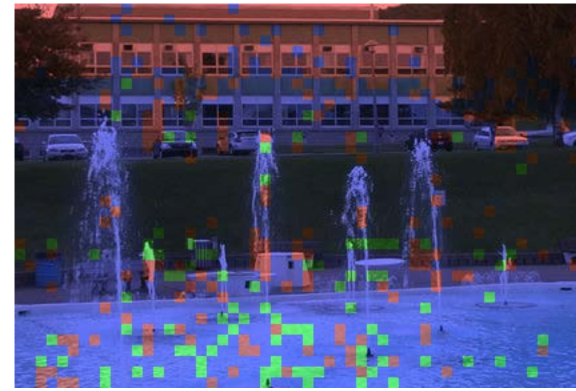
Base Transform Estimation*

- The type of the base transform chosen varies according to scene content.

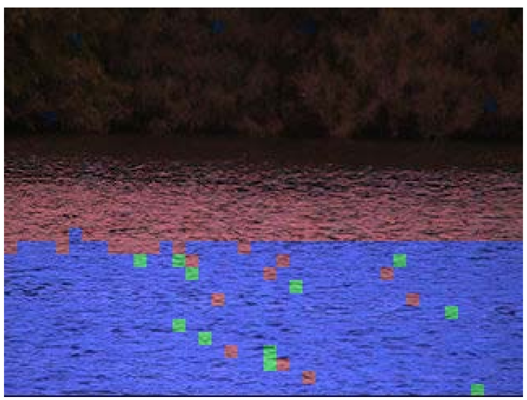


V^1: Boats

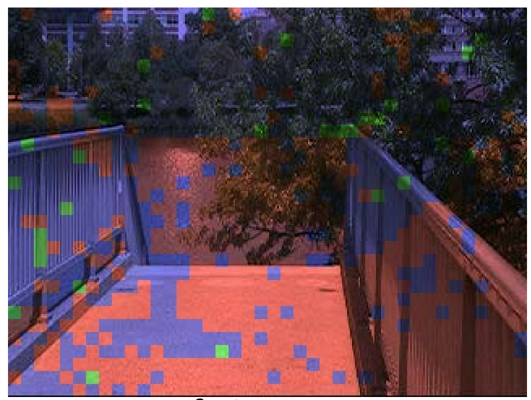
V^3: Fall



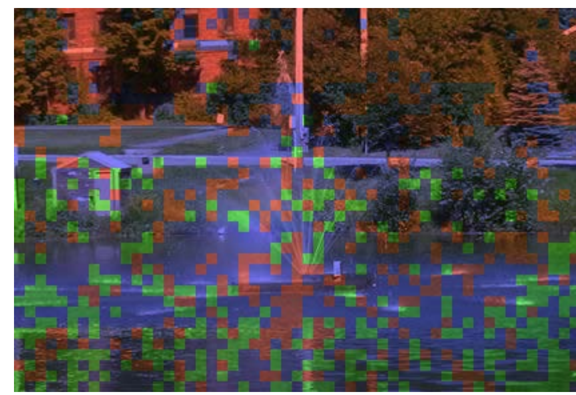
V^4: Fountaion1



V^2: Canoe



V^6: Overpass



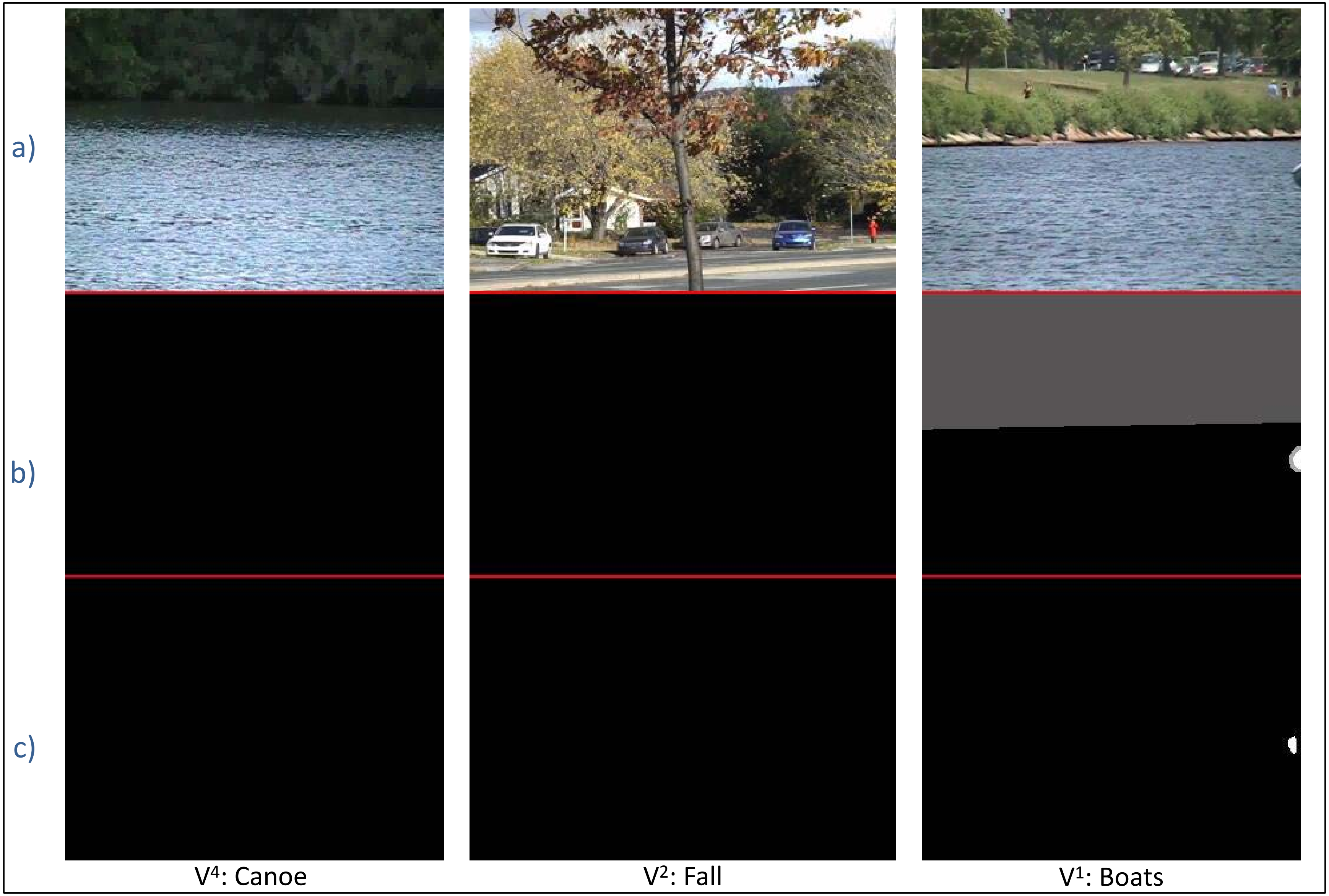
V^5: Fountaion2

	Estimated Base Transform (%)		
	DCT	**WHT**	**SLT**
V^8	63.75	2.16	34.09
V^9	61.00	1.66	37.34

	Estimated Base Transform (%)		
	DCT	**WHT**	**SLT**
V^{10}	55.48	8.92	35.60
V^{13}	52.58	5.16	42.26

	Estimated Base Transform (%)		
	DCT	**WHT**	**SLT**
V^{11}	27.00	8.44	64.56
V^{12}	36.00	14.81	49.19

Visual Change Detection Results



a) Input frame sequence b) Ground truth c) Our result

In ground truth, 0: Ordinary change, 255: Relevant change, 85: outside region of interest 170: unknown motion.

Quantitative Change Detection Results

- 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.
- TP: If the change detection method labels p_{rc} as relevant change, this case is called true positive. If not, false negative (FN)
- TN: If the change detection method labels p_{oc} as ordinary change, this case is called true negative. If not, false positive (FP)
- Specificity: TN/(TN+FP) and Accuracy: (TP+TN)/(TP+TN+FP+FN)

	V^1	V^2	V^3	V^4	V^5	V^6	
Number of Test Frames	6,100	390	785	1,000	2,001	3,001	**Average**
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

- We compare our method to the top-three methods under the dynamic background category on ChangeDetection.net (ranking results retrieved on **June 2013**).

Method	**boats**		**canoe**		**fountain01**		**fountain02**		**overpass**		**fall**		**Average**	
(*Ranking*)	**Re**	**Pr**	**Re**	**Pr**	**Re**	**Pr**	**Re**	**Pr**	**Re**	**Pr**	**Re**	**Pr**	**Re**	**Pr**
[1] (*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
[2] (*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
[3] (*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**

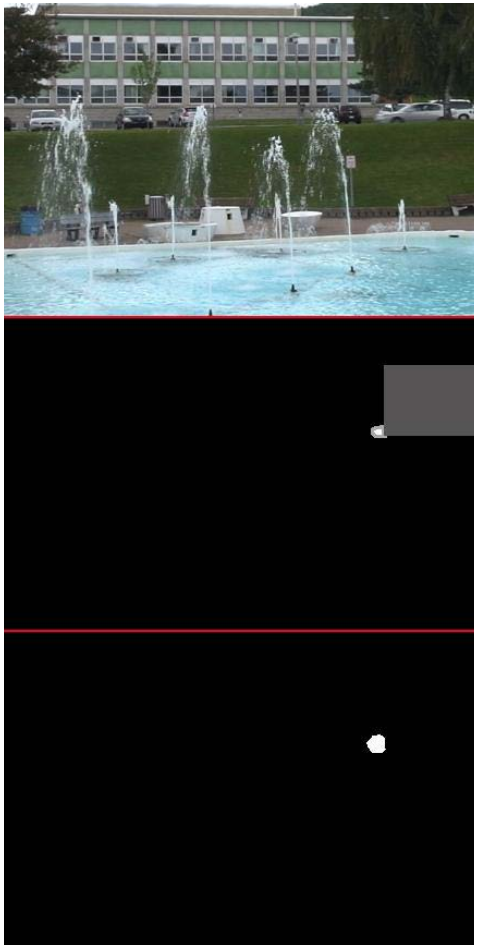
Re: Recall and Pr: Precision $Re = \frac{TP}{TP+FN}$ $Pr = \frac{TP}{TP+FP}$

[1] Tom SF Haines and Tao Xiang. Background subtraction with Dirichlet processes. ECCV 2012, pages 99–113. Springer, 2012.
[2] Ashutosh Morde, Xiang Ma, and Sadiye Guler. Learning a background model for change detection. CVPRW 2012, pages 15–20, 2012.
[3] Ismail, M., Hamed M., and Chilufya, C. Object segmentation using full-spectrum matching of albedo derived from colour images, 2011.

Limitations and Future Work

- *Limitations:*
  - The major limitation of our method is that estimating base transforms requires a set of video frames without relevant changes.
  - Another limitation arises from cube-based computations, which may cause blocking artifacts.
  - Not all types of ordinary change patterns can be modeled by the three transforms (i.e., DCT, WHT, and ST) and additional transforms should be considered.
- *Future Work:*
  - Dynamic update of transform estimation.
  - Extend the approach to model shadow regions.
  - Large scale testing.

a)
b)
c)



a) Input frame sequence
b) Ground truth
c) Our result

In ground truth, 0: Ordinary change, 255: Relevant change, 85: outside region of interest 170: unknown motion.

20

Questions