GraDED: A graph based parametric dictionary learning algorithm for event detection

Tamal Batabyal, Rituparna Sarkar, Scott T. Acton

Virginia Image and Video Analysis (VIVA)

C.L. Brown Department of
Electrical and Computer Engineering

University of Virginia, Charlottesville, VA 22904

2017.09.15

Virginia Image and Video Analysis

UNIVERSITY of VIRGINIA
BICENTENNIAL

Events (Fire at a street corner)

Events (Car accidents → fire)

VIVA

UNIVERSITY of VIRGINIA
BICENTENNIAL

Event detection

Event detection from videos

Temporal localization

(In how many frames the event happened?)

+

Spatial localization

(location of the event in each frame)

Challenges

- Dynamic background videos

  (Car-mounted camera, hand-held camera)

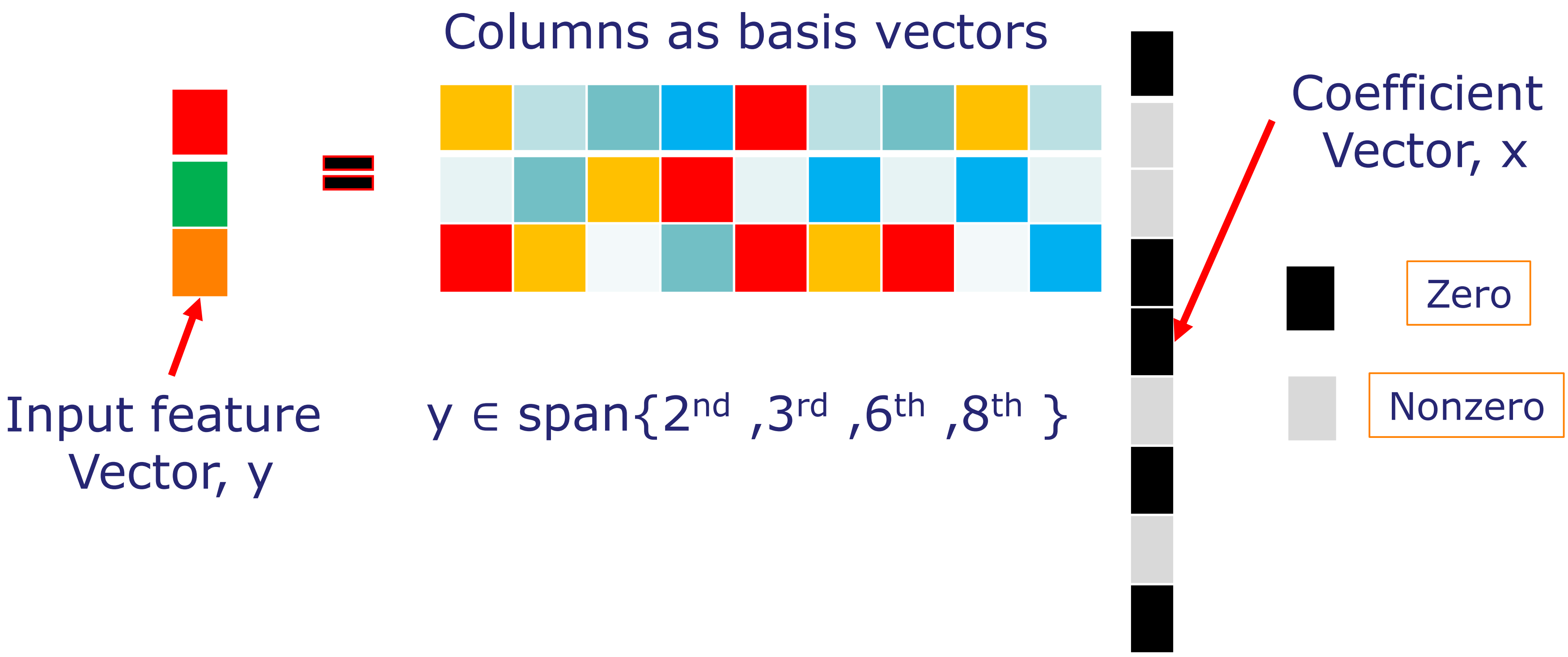
- Illumination/Intensity variation

- Camera jitter

  (Gait motion, motion of cars on uneven surface)

- No actual object to track (Fire, no shape prior)

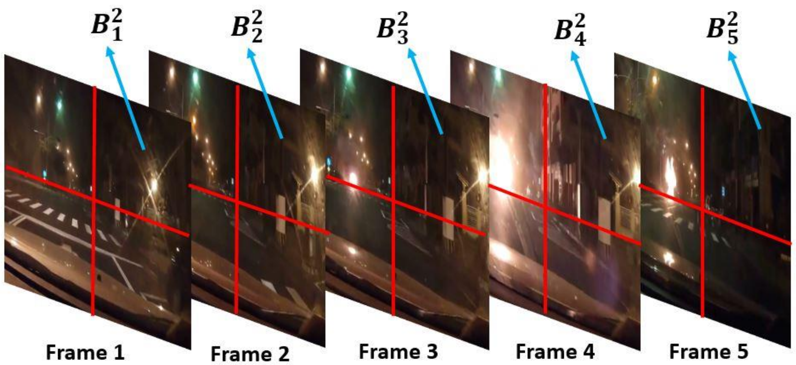
Subspace based approach



Subspaces ⇒ noise removal, minor illumination variation removal, low-dimensional representation

Block-based dictionary

$$(D^*, X^*) = \min_{D,X} \|Y - DX\|_F^2 \text{ s.t. } \|X\|_0 \leq T.$$



In our case,

$$\mathrm{D} = [D_1\ D_2\ \cdots\ D_P]$$

P = total number of partitions of each frame

Number of blocks, P = 4
Number of frames, K = 5

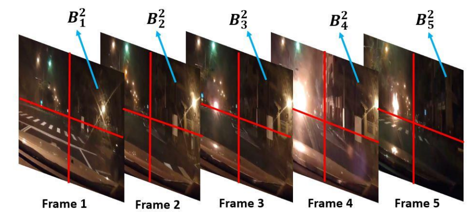
$B^i = i^{th}$ feature sub-volume

$$B^i = [B_1^i\ B_2^i\ \cdots B_K^i]^T \in R^{K \times S}$$

B_j^i = S-dimensional feature (i^{th} block, j^{th} frame)

Feature = HOG features with a cell size 16×16

Block-based dictionary



$B^i = i^{th}$ feature sub-volume

$B^i = \left[B_1^i \, B_2^i \, \cdots B_K^i\right]^T \in R^{K \times S}$

$\mathbf{B^i}$

Feature dimension

Frame

Event dynamics

Illumination variation

Camera jitter

Camera motion

Clutter

$\mathbf{D_i} =$

$\mathbf{C^i_{event}} \cup \mathbf{C^i_{illn}} \cup \mathbf{C^i_{jitter}} \cup \mathbf{C^i_{motion}} \cup \mathbf{C^i_{clutter}}$

$\mathbf{C^i_{event}}$ **=** A subset of columns of $\mathbf{D_i}$ that spans the event in the i^{th} subvolume

Our contribution

$$\mathbf{D_i} = U_i^T \chi_i$$

Temporal dynamics of the event

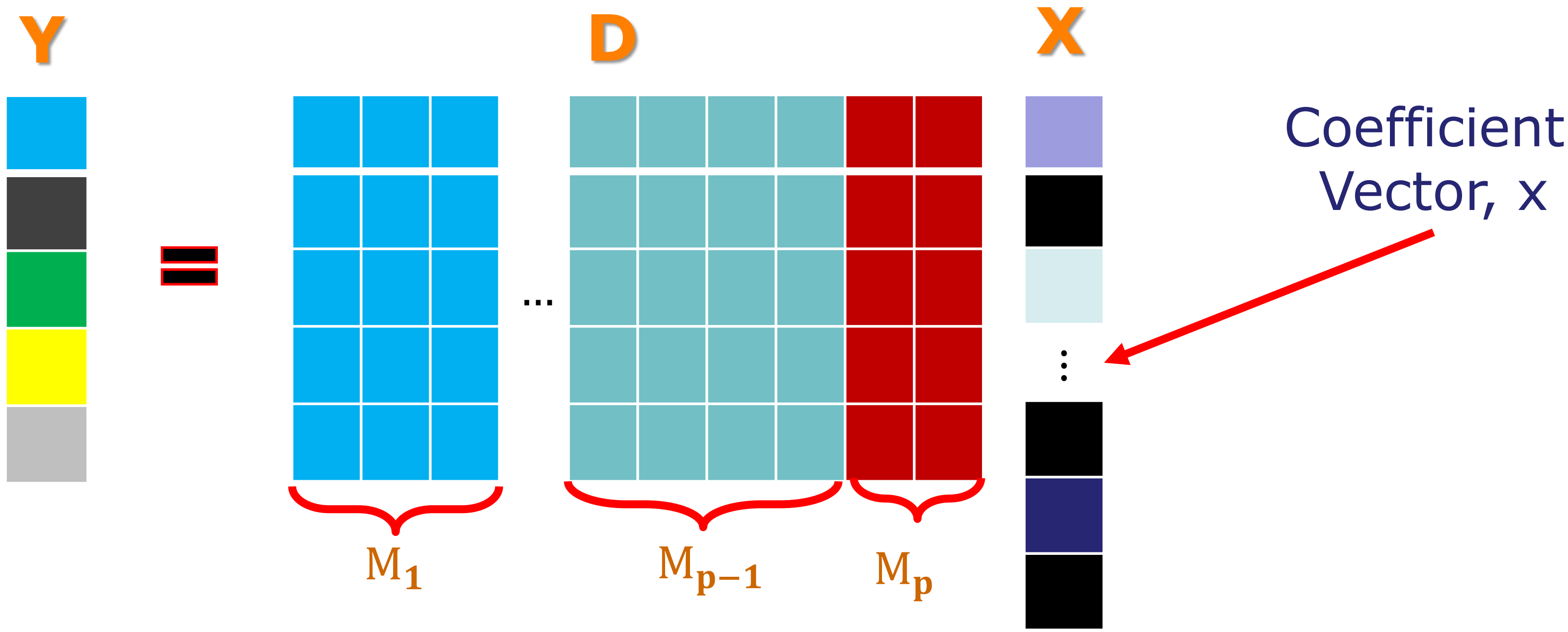
(Graph between consecutive frames of i^{th} subvolume)

Noise removal

(PCA, sparse approximation of coefficients)

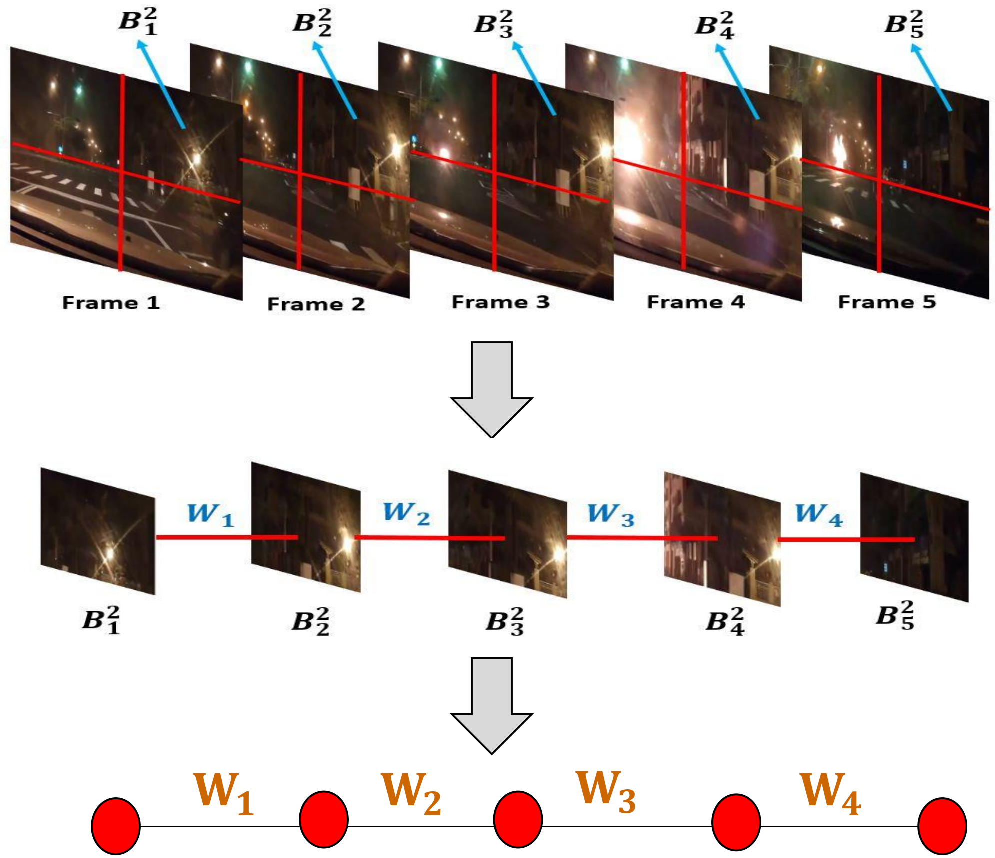
PCA of B^{i}

χ_i = first M_i eigenvectors taken from PCA of B^i



Dimension of X, $\mathrm{M} = \sum_{i=1}^{P} M_i$

Block graph



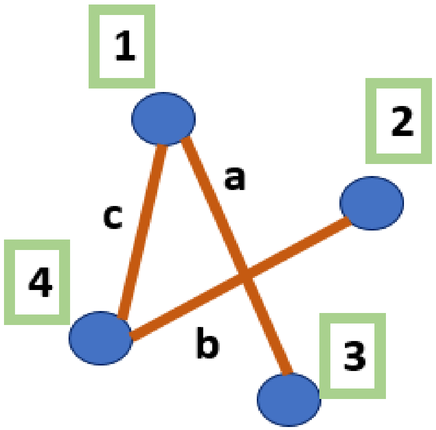
Number of blocks = P

Number of frames = K

Number of free parameters = Number of weights in all P graphs = P(K-1)

Linear graph of 2nd blocks

Graph basics



Adjacency Matrix, $A = \begin{bmatrix} 0 & 0 & a & c \\ 0 & 0 & 0 & b \\ a & 0 & 0 & 0 \\ c & b & 0 & 0 \end{bmatrix}$

Degree Matrix, $D = \begin{bmatrix} a+c & 0 & 0 & \\ 0 & b & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & b+c \end{bmatrix}$

Laplacian Matrix, $L = D - A = \begin{bmatrix} a+c & 0 & -a & -c \\ 0 & b & 0 & -b \\ 0 & 0 & a & -a \\ -c & -b & 0 & b+c \end{bmatrix}$

Normalized symmetric Laplacian, $\tilde{L} = D^{\frac{-1}{2}} L D^{\frac{-1}{2}}$

Eigenmatrix of graph Laplacian

$$\boxed{\mathbf{L_i} = \mathbf{U_i \Lambda_i U_i^T}}$$

It belongs to O(K),

Distance-preserving transformation/modulation

Precomputed and fixed during iteration

$$\mathbf{D_i} = \boldsymbol{U_i^T \chi_i}$$

Intra-block mutual coherence is kept intact :
$\mu_i\left(U_i^T \chi_i, U_i^T \chi_i\right) = \mu_i\left(\chi_i, U * U_i^T \chi_i\right) = \mu_i(\chi_i, \chi_i)$

Dictionary learning algorithm

$$(D^*, X^*) = \min_{D,X} \|Y - DX\|_F^2 \text{ s.t. } \|X\|_0 \leq T.$$

Update X:

Orthogonal matching pursuit

Update D:

Gradient descent

Alternating minimization

↓

Update D, keeping X fixed

↓

Update X, keeping D fixed

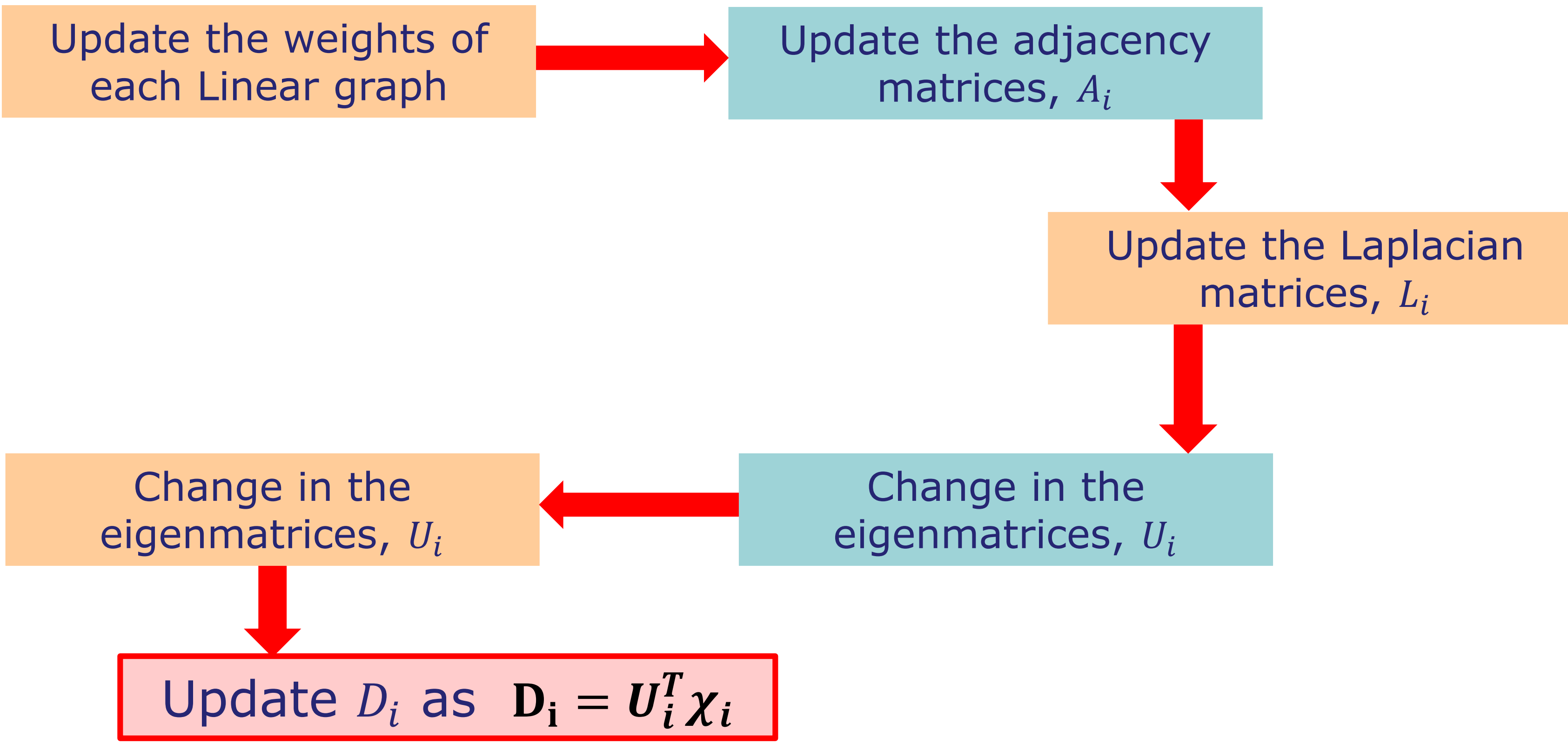
↓

Update D, keeping X fixed

↓

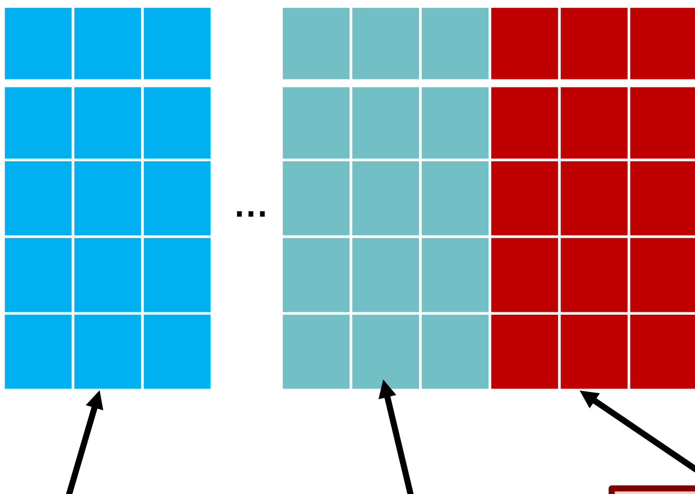
⋮

Flow chart for gradient descent



Spatial localization

Dictionary, D



$\mathbf{D}_1 = U_1^T \chi_1$

$\mathbf{D}_{p-1} = U_{p-1}^T \chi_{p-1}$

$\mathbf{D}_p = U_p^T \chi_p$

Feature for block 1

Feature for block P-1

Feature for block P

Hierarchical clustering

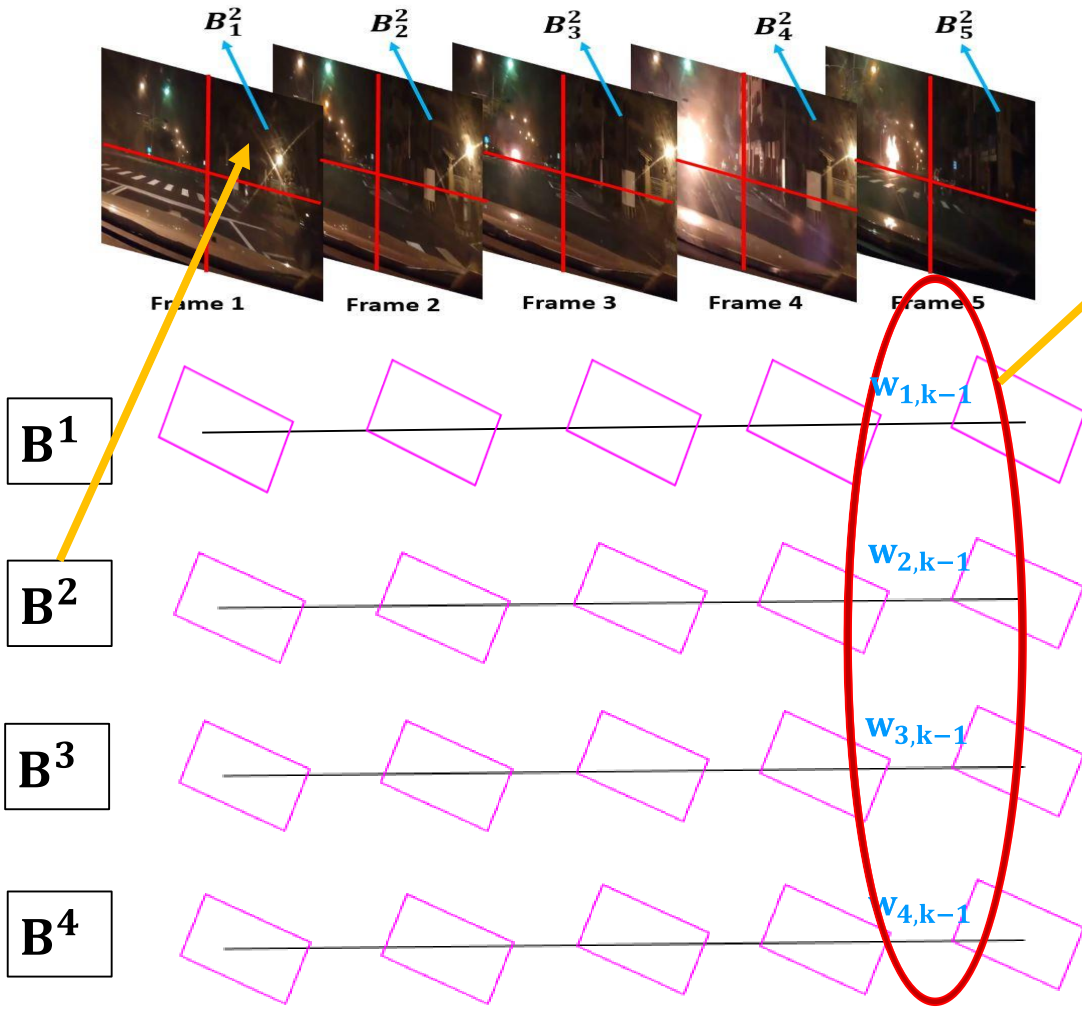
Total number of final clusters = 2.

Event and no-event clusters

Linkage: Unweighted average distance

Feature similarity: mutual coherence

Temporal localization (P=4, K=5)



Feature vector between frames K-1 and K

Collection of all such feature vectors

2-means classification

Dataset

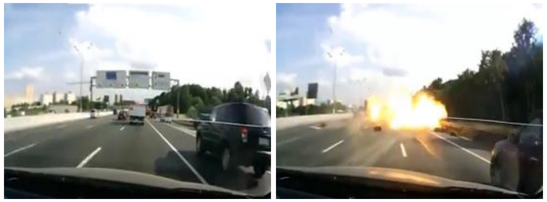
Disappearance of boat

Frames: 75
Frame dim: 378×281

Explosion at gas station

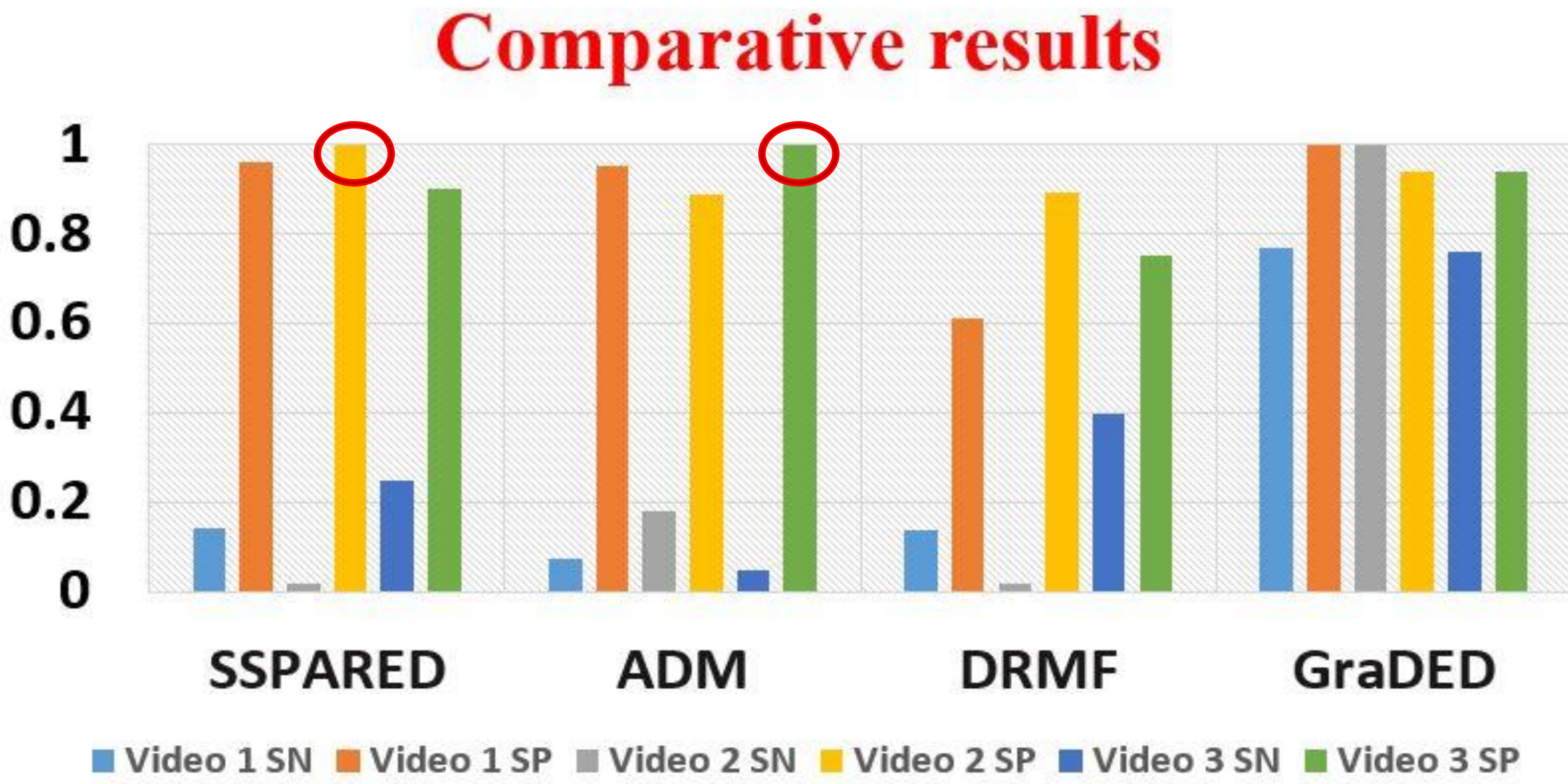
Frames: 69
Frame dim: 632×342

Car accident and fire



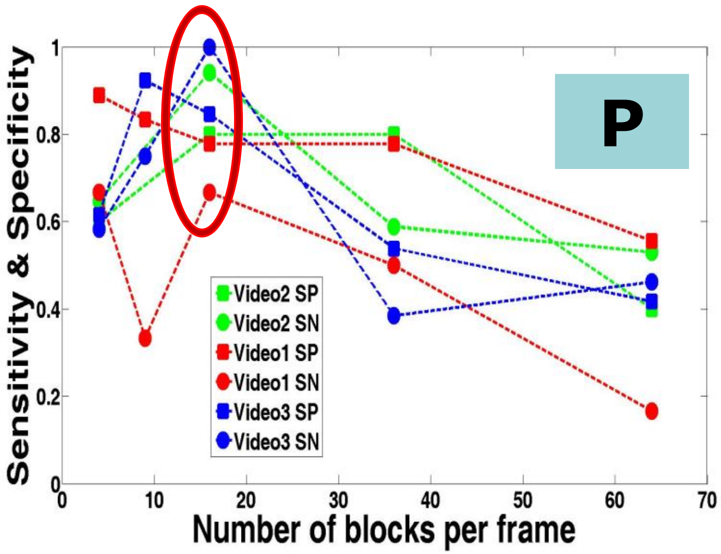
Frames: 77
Frame dim: 422×208

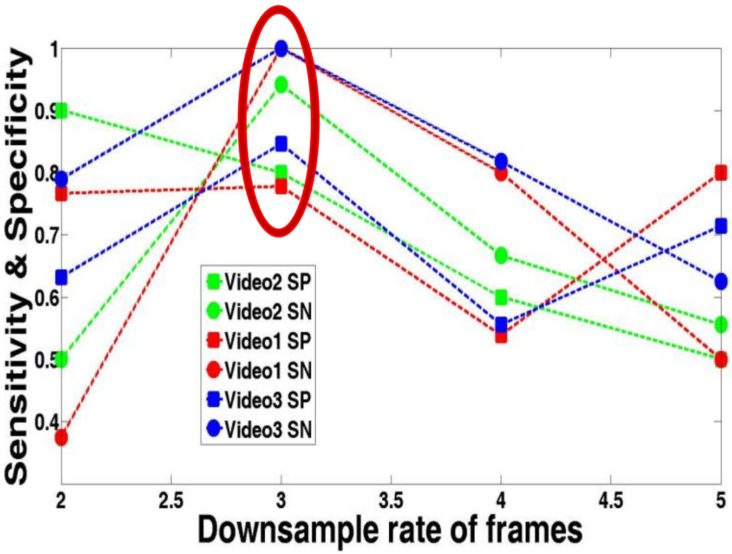
Comparison with state-of-the-arts

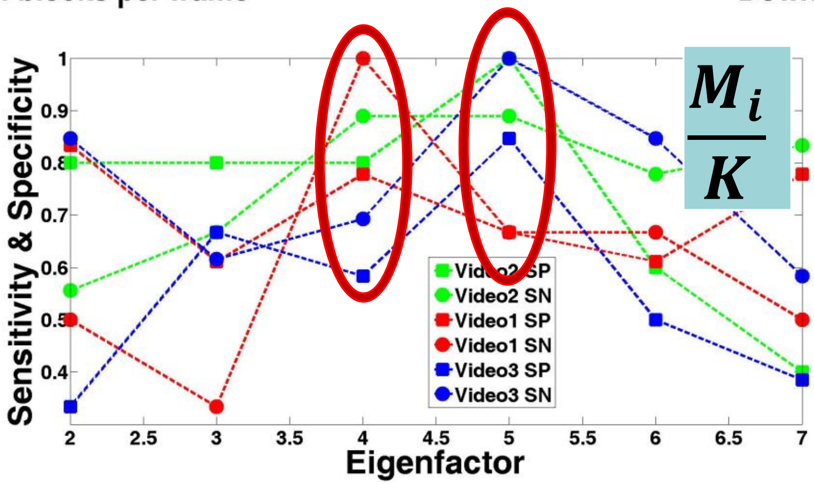


$$\textbf{SN} = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad \textbf{SP} = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

Result on parameter selection







Thank you

谢谢