Statistical processing for spatiotemporal event localization in video

Alexia Briassouli

Introduction

- Event detection and recognition are attracting significant research attention in many areas

- Events are characterized by their spatial and temporal extent (i.e. where and when they occur)

- Statistical processing of the motion features can provide spatial and temporal localization of activities

- Further event processing can follow, e.g. recognition, classification etc.

Introduction

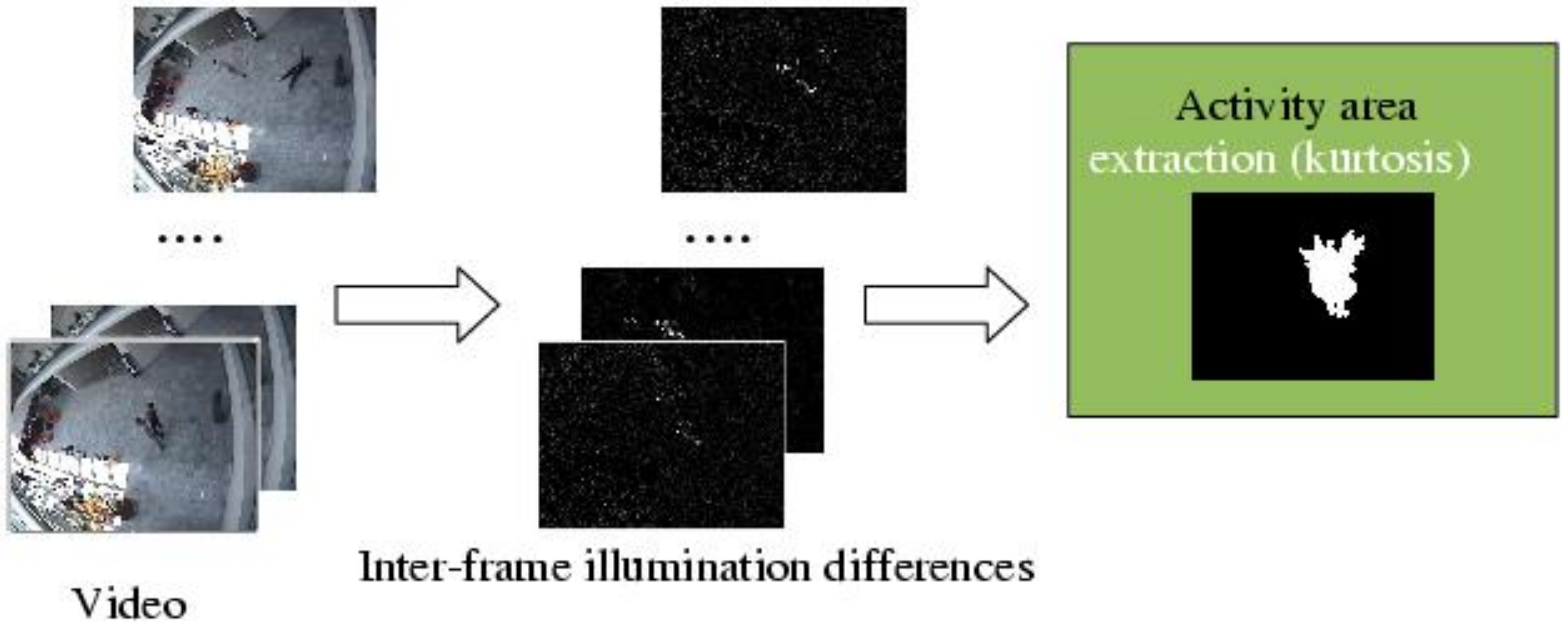
- We present a principled approach to detecting regions of motion and times of change in activities.

- Areas of activity are found by processing the kurtosis of illumination variations. This leads to binary masks called "Activity Areas".

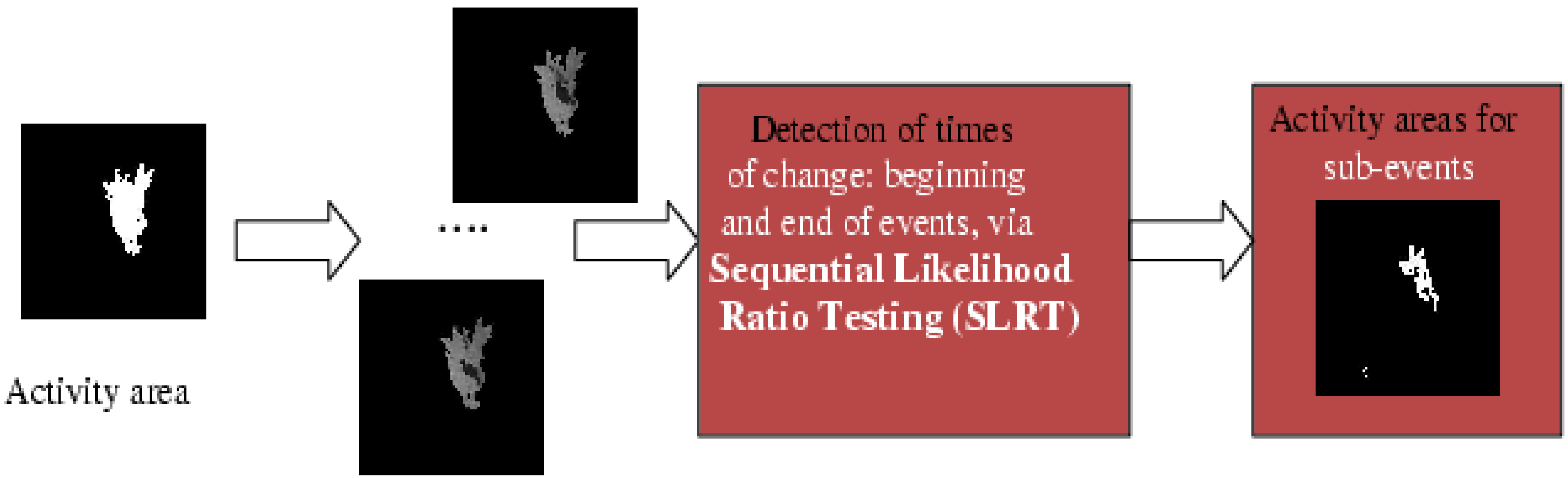
- Times of change in activities are found by applying sequential likelihood ratio testing.

- Activity areas for subsequences corresponding to different activities are extracted.

Spatial activity Localization



Temporal activity Localization



Temporal processing of data in activity area

Kurtosis-based spatial localization

- Inter-frame illumination variations are caused by measurement noise and/or pixel activity.

$$H_0 : v_k^0(r) = z_k(r)$$

$$H_1 : v_k^1(r) = u_k(r) + z_k(r)$$

- Displacement $= u_k(r)$
- Measurement noise = $z_k(r)$

- The measurement noise is assumed to follow a Gaussian distribution. This is a common assumption in practice and holds for a large number of data samples.

- The Kurtosis is a measure of a variable's Gaussianity: it is zero for Gaussian variables.

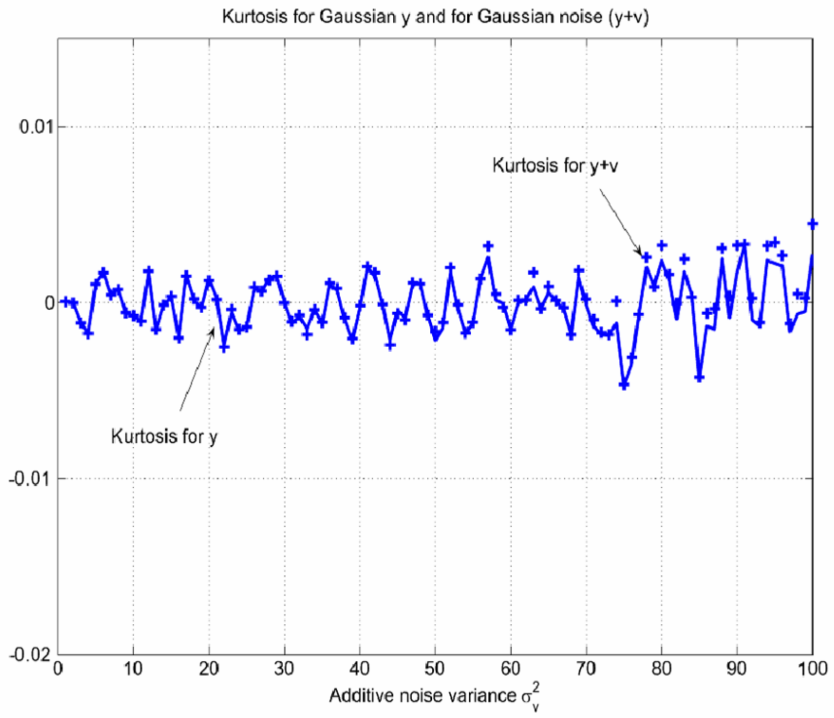
$$kurt(y) = E[y^4] - 3(E[y^2])^2$$

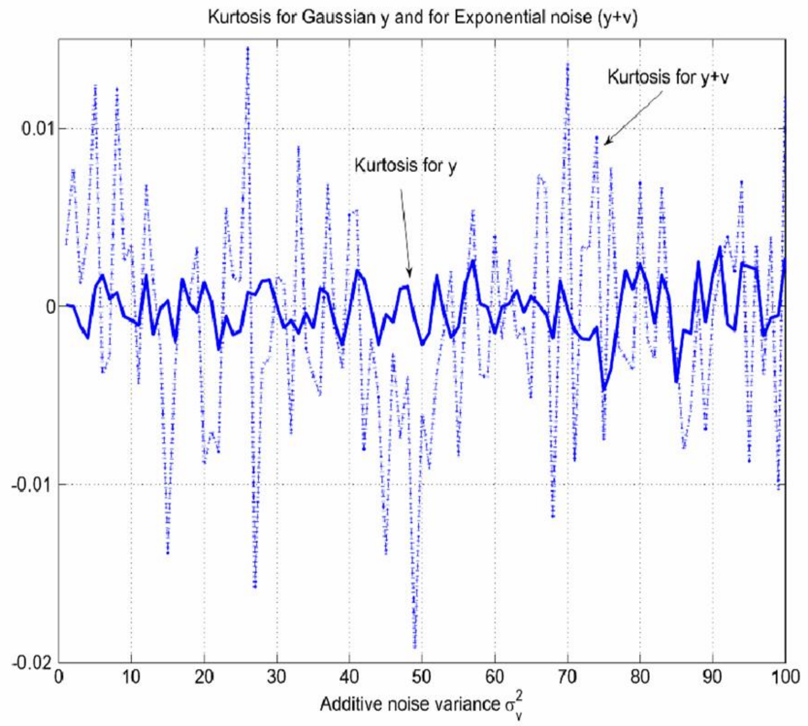
- In our case, the kurtosis of a series of flow measurements will be zero for pixels whose flow is caused by (Gaussian) noise.

- In practice, the data is not perfectly Gaussian and the kurtosis is not precisely zero.

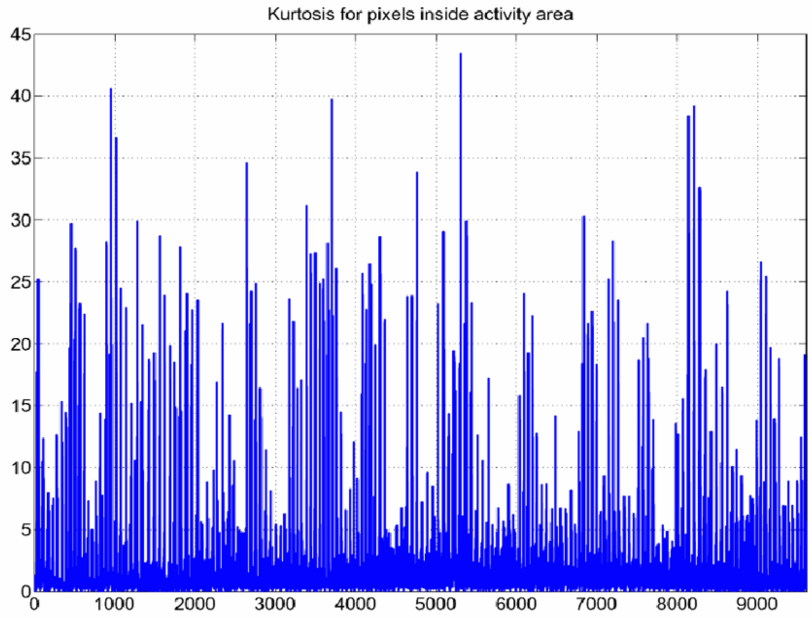
- However: the *kurtosis is sensitive to outliers*, so its values are much higher when the data contains an actual displacement (i.e. an outlier) and not only measurement noise.

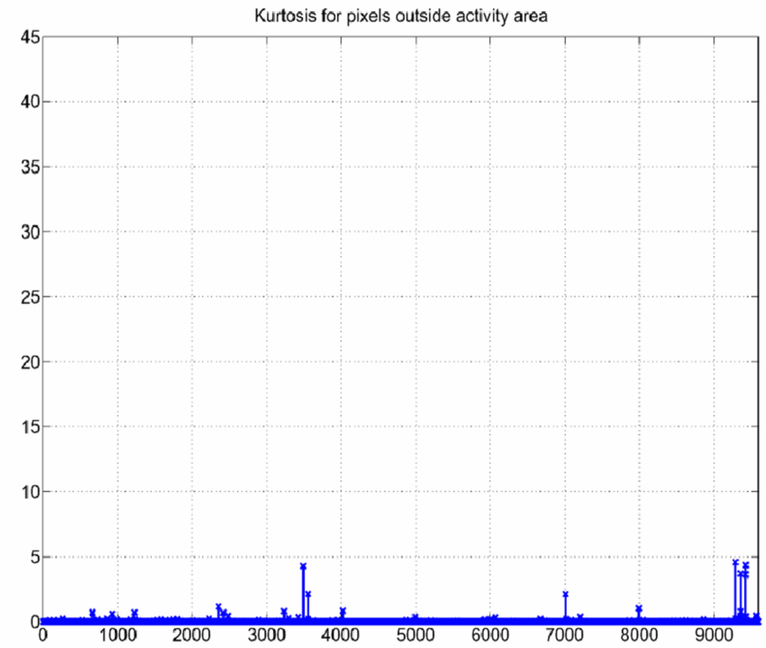
Example of kurtosis values with Gaussian vs. Exponential noise





Example of kurtosis values of static vs. active pixels





• Application of the kurtosis to the optical flow measurements over several frames leads to binary masks, the "Activity areas".

• Activity areas have been proven to be more robust to noise than "Motion Energy Images", defined as:

binarized interframe differences

Motion Energy Images

10

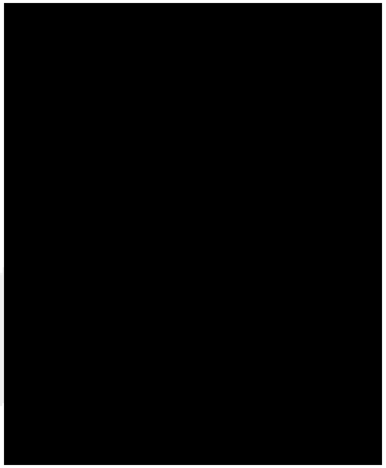
MEIs, Activity Areas

Activity

MEI

11

MEIs, Activity Areas for noisy data



Activity

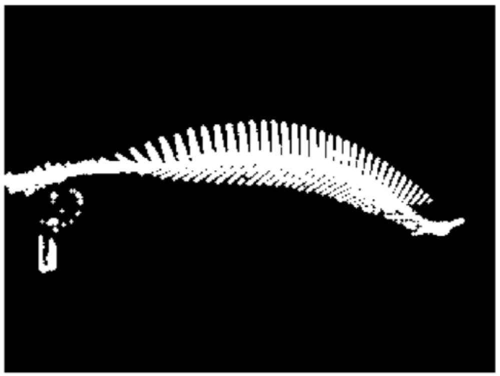
MEI

12

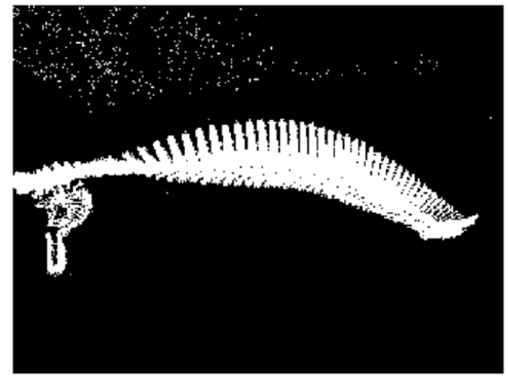
MEIs, Activity Areas for noisy data

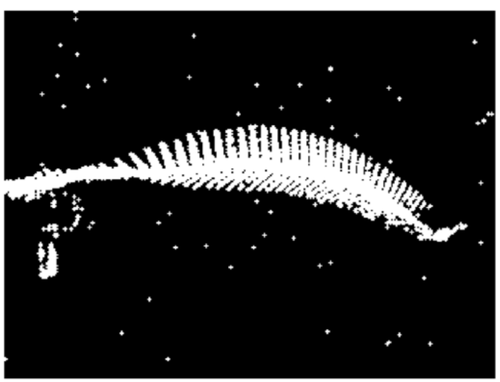
Activity area

MEI

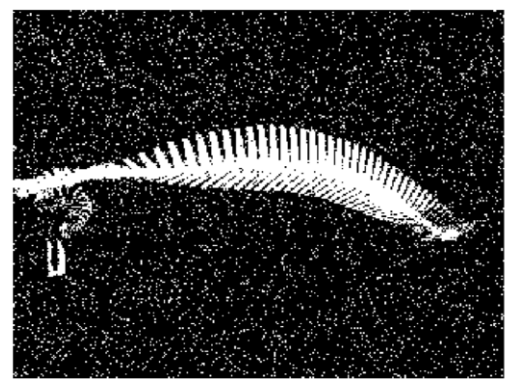


Noiseles
s
data





Data with
additive
noise



13

MEIs, Activity Areas for noisy data

Video frame

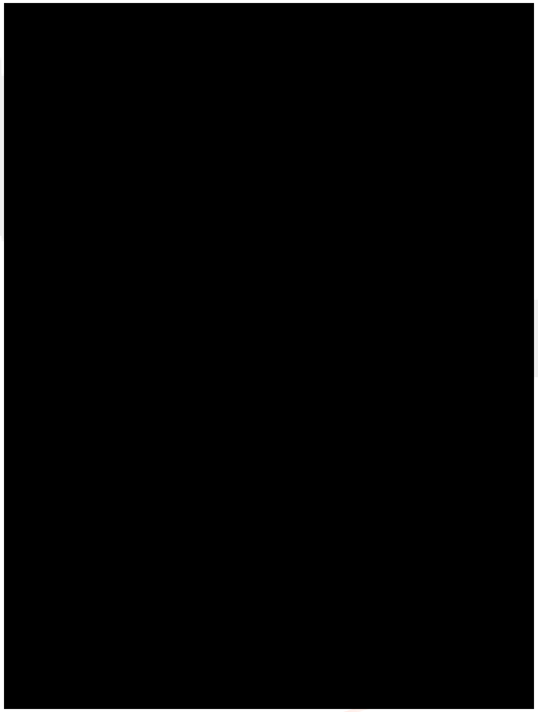
Activity

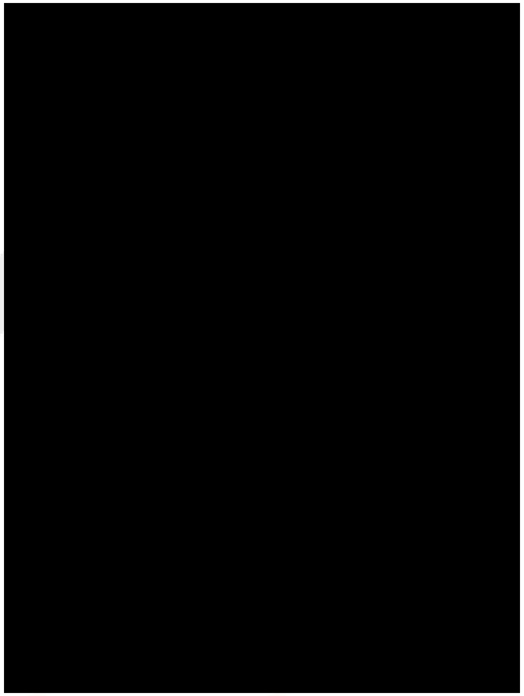
Data with additive noise

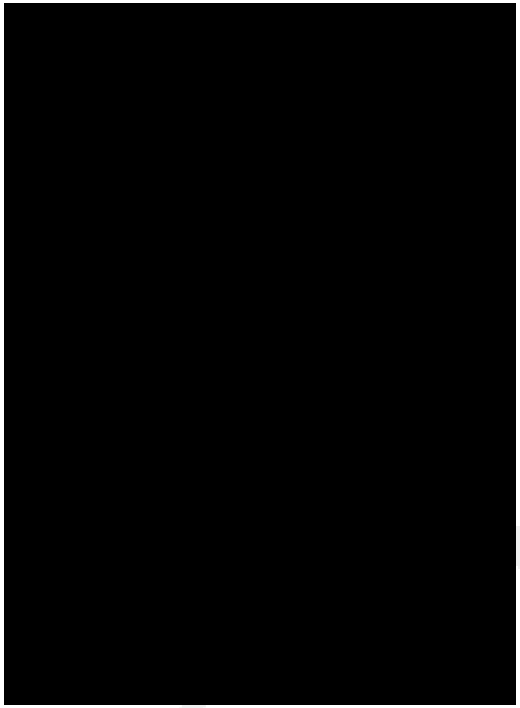
MEI

14

Activity Areas obtained sequentially over several frames







15

Activity Areas obtained sequentially over several frames

Sequential Change Detection for temporal event localization

- Changes in the activity/event taking place are often reflected in changes in the motion present in a scene.

- The motion (optical flow field) follows a different statistical distribution before and after a change.

- If a change takes place at an unknown time instant L, for N data samples, we have:

Sequential Change Detection for temporal event localization

- Sequential change detection techniques need to be employed to find:
  - *Unknown* moment of change k, for *unknown* distributions and before and after a change, respectively.
- Sequential change detection advantages:
  - Data is processed as it arrives, can provide online, real-time solutions
  - Change points are proven to be detected with a minimum delay.

- In practice, the data distributions are approximated from the available data.

Cumulative Sum (CUSUM) for Sequential Change Detection

- The CUSUM method is one of the most commonly used sequential change detection techniques in practice.
- For a data vector the log-likelihood ratio of the data at each frame k is estimated:
- The test statistic examined at each frame is:

Cumulative Sum (CUSUM) for Sequential Change Detection

• Changes are found when the test statistic becomes higher than a predefined threshold .

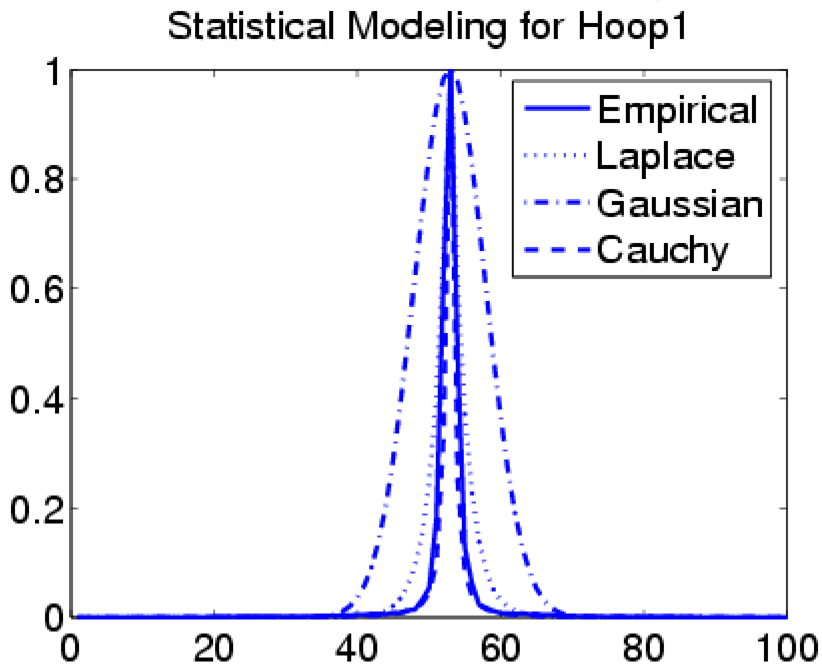
•

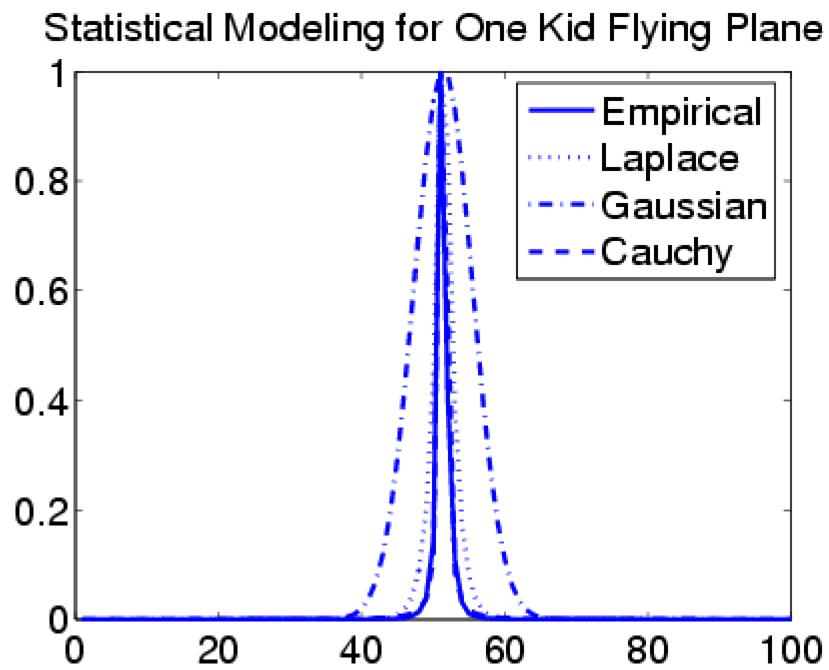
• The threshold is estimated empirically with training data in order to give the smallest detection delay, and its optimum value is found to be:

• are the mean and standard deviation of the test statistic values until frame k.

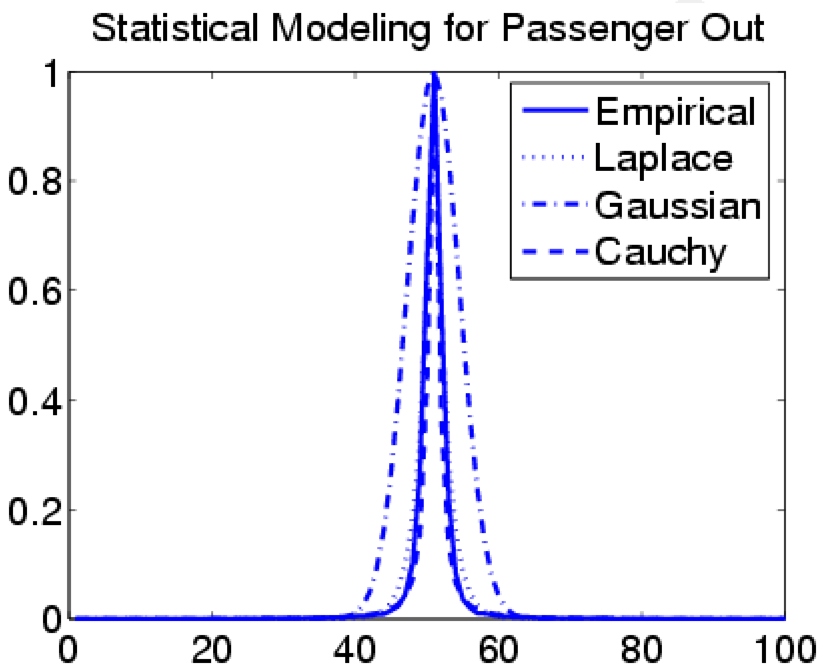
• Best values for c in [2,3] for human activity videos.

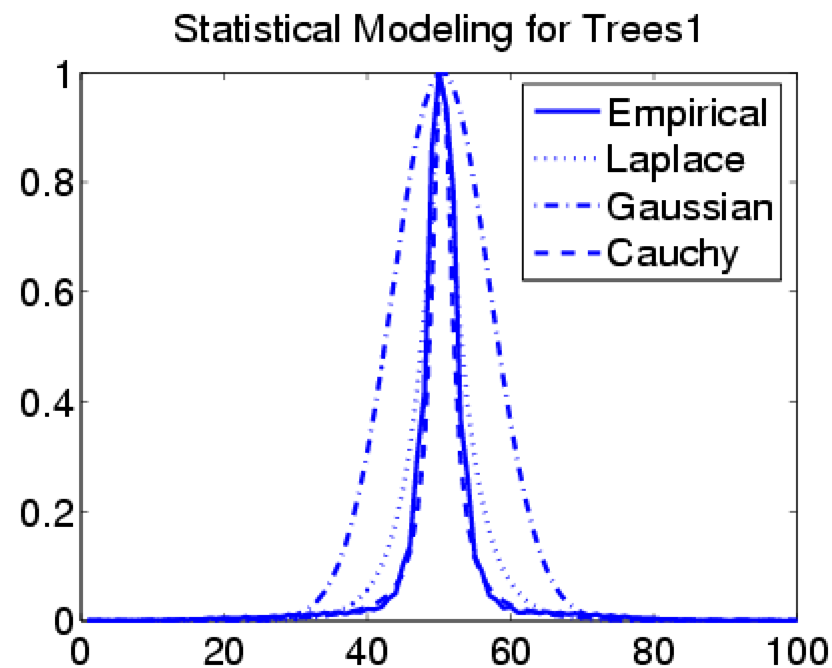
Data modeling





Data modeling





CUSUM test for Cauchy data

- For a N sample vector, the CUSUM test becomes:

- When compared to a threshold chosen based on training data, we obtain correct change detection.
- Separates the video into subequences of different activities.
- Activities in the subsequences can be correctly classified once the frame of change is found.

Activity subsequence separation

24

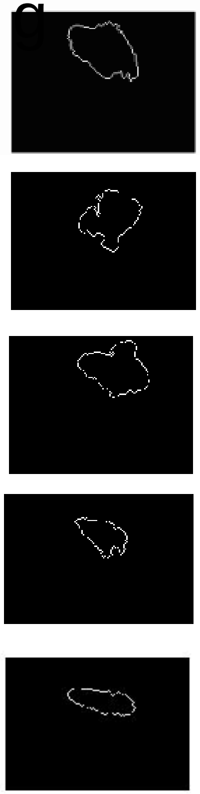
Activity subsequence separation



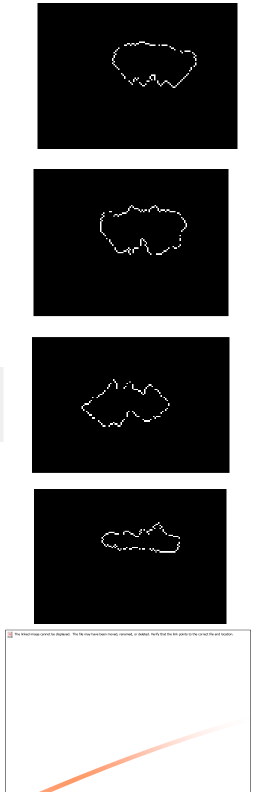



Activity subsequence recognition

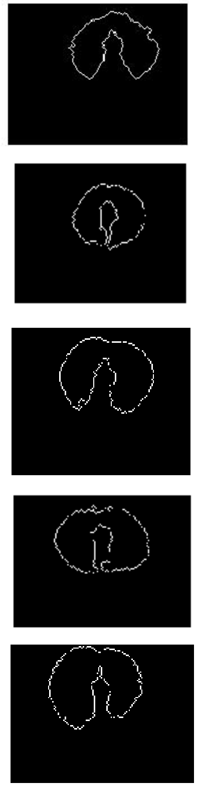
boxin
g



clappin



wavin



- Estimation of Fourier descriptors of activity area outlines
- Euclidean distance estimated between the Fds for these activities.

RECOGNITION FOR BOXING, HANDCLAPPING, HANDWAVING (%)

Activity	Box	Handclap	Handwave
Box	75.49	24.51	0
Handclap	17.39	79.45	3.16
Handwave	0	12.85	87.15

Sequential change detection for varying scene illumination

- Data is processed sequentially in time.
- This allows to check for global scene illumination changes.

- Illumination induced changes can be eliminated – the method is robust to global illumination variations.

- False alarm detections are eliminated based on motion information.

False alarm elimination

Global illumination changes

- Ratio of average illumination at frame k:

- For there is a global illumination change.

Motion false alarms

- Ratio of average motion magnitudes at frame k:

- For , or when the sign of the average motion changes, the detected change is a false alarm.

Experiments

- Changes detected at frames 18, 32:

Experiments

- Changes detected at frames 14, 46, 88:

Experiments

- Changes detected at frames 16, 61, 84:



Experiments with varying illumination

- Person enters, exits:


- Person appears from behind plant, leaves:


32

Experiments with varying illumination

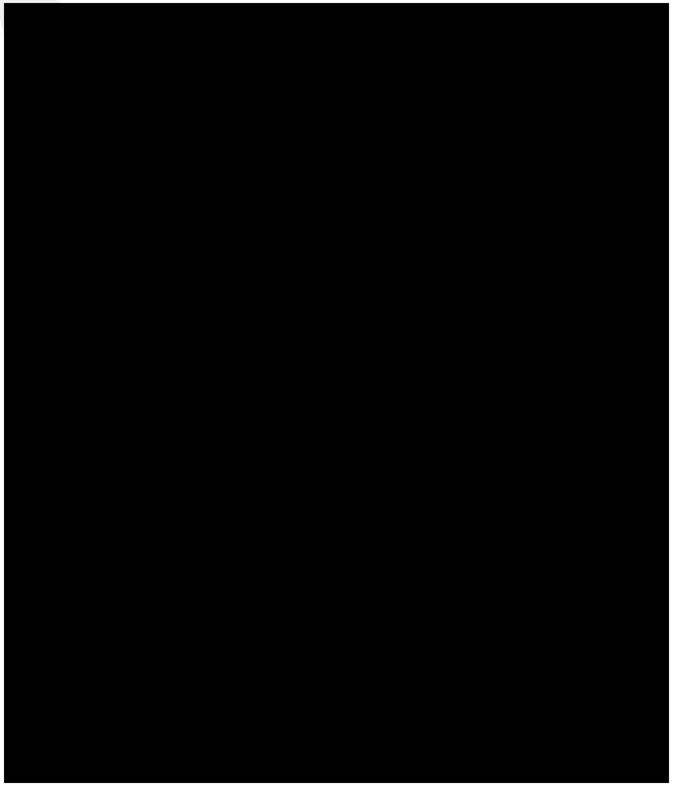
- Person enters, walks behind plant:


- No change is detected at illumination changes:



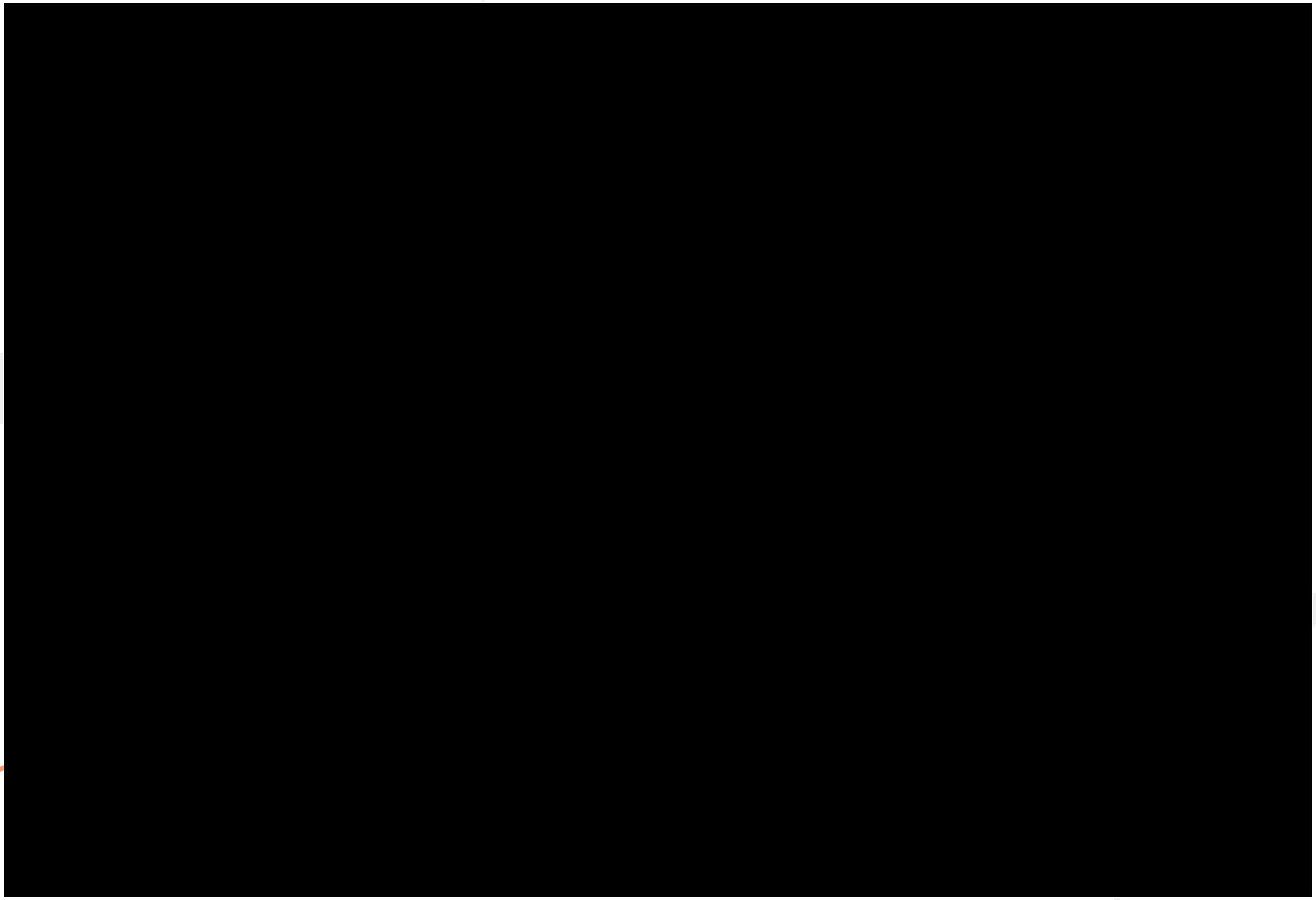

Experiments with varying illumination

- Hoop with varying illumination, changes found at correct change points, not when illumination changes.



Experiments with varying illumination

- Kid throws plane under varying illumination



35

Experiments with varying illumination

• Changes at frame 17 when plane is thrown, 41 when it changes trajectory, 90 when it lands. Change is not detected at 60 when illum. changes.

Change detection in temporal textures videos

- Temporal textures videos contain highly non-rigid motions, e.g. water flowing, tree leaves fluttering, crowds of people walking, traffic.
- Difficult to analyze their motion using traditional motion estimation techniques, e.g. optical flow.
- Work on temporal textures has focused mostly on their modeling. The SpatioTemporal AutoRegressive model (STAR) has received much attention for the representation of temporal textures.
- This work presents a non-parametric approach to modeling the motion in temporal textures, rather than their appearance.

Change detection in temporal textures videos

- The modeling of motion in temporal textures allows the detection of changes between different temporal texture subsequences based on motion rather than appearance features.

38

Mathematical model for temporal texture sequences

- Consider that each frame consists of M moving "objects" (leaves, cars in traffic, people in a crowd) that are undergo a random displacement of from frame to frame:

- The FT of the frame illumination values is:

Mathematical model for temporal texture sequences

- The displacements are considered to follow the same random distribution:

- The characteristic function of a random distribution is defined as its FT:

Mathematical model for temporal texture sequences

- The expected value of the FT of frame k is:

- The motion characteristic function can be found from the video FT

41

Mathematical model for temporal texture sequences

- The characteristic function provides a complete description of the random motion's statistics:
- All moments (mean, variance, HOS) and the data pdf can be derived from the characteristic function.

- The motion distribution can be derived from its characteristic function.

Mathematical model for temporal texture sequences

- Sequential change detection can be applied to find changes between one motion distribution and another, i.e. between two successive temporal textures.
- In practice, we need to approximate the expected value of the frame FTs.
- Theoretically we need many instantiations of the video, but in practice this is not possible.
- Assumptions:
  - Ergodic data
  - Motion distribution does not change significantly in a window of frames.

Mathematical model for temporal texture sequences

- Approximation of expected values of Fts:

- So the characteristic function is given by:

- and the motion distributions are extracted from it

44

Sequential change detection for temporal texture sequences

• Initial ("baseline") data distribution is approximated from the first frames.

• Current data distribution is approximated from

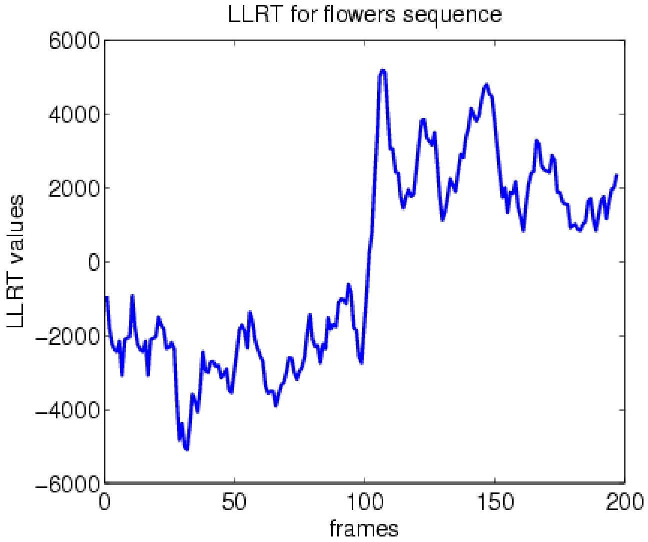
{f_1(x_n)}

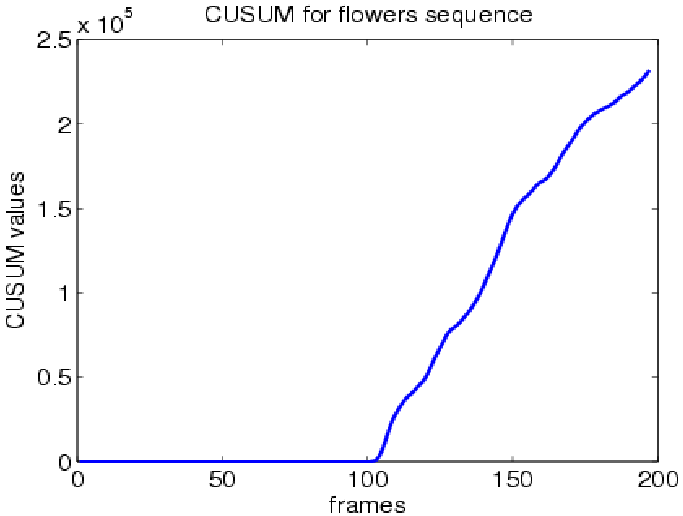
frames neighboring the current frame k.

• The test statistic is the log-likelihood ratio:

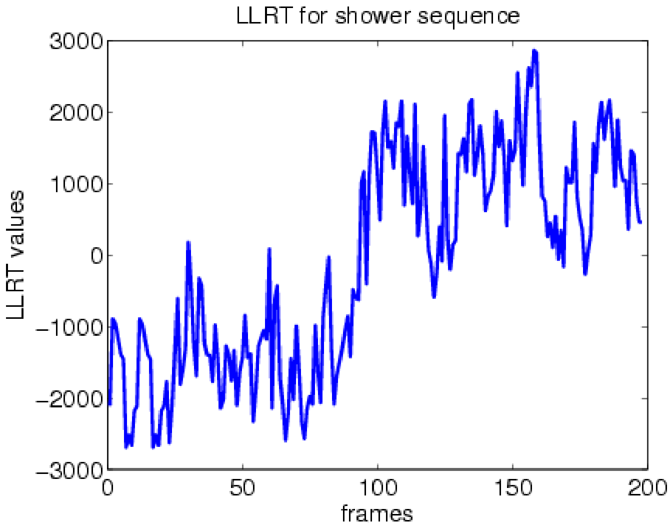
• and the CUSUM test is, as before:

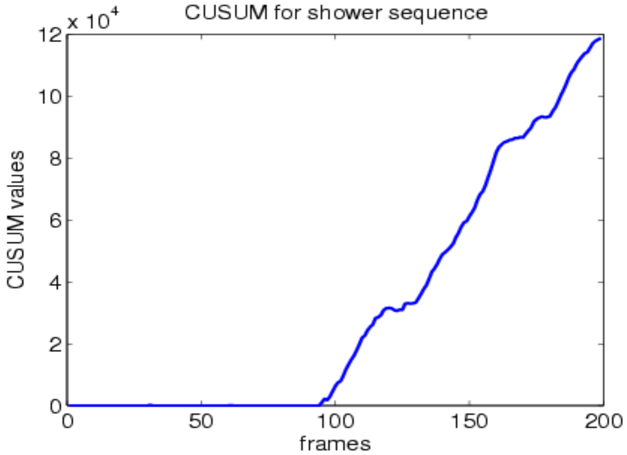
Experimental Results



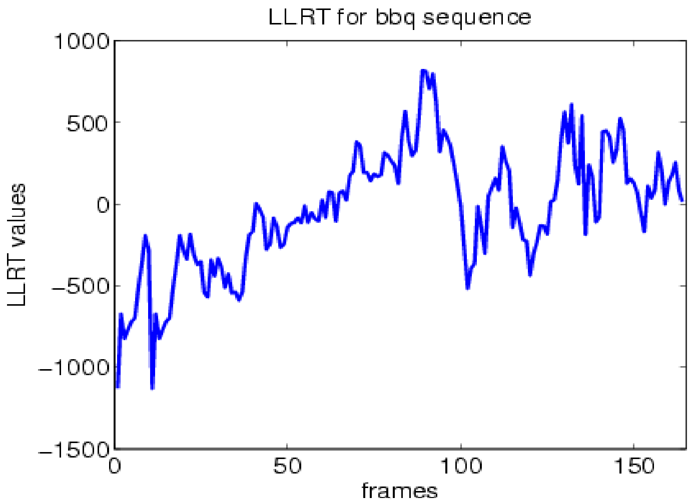


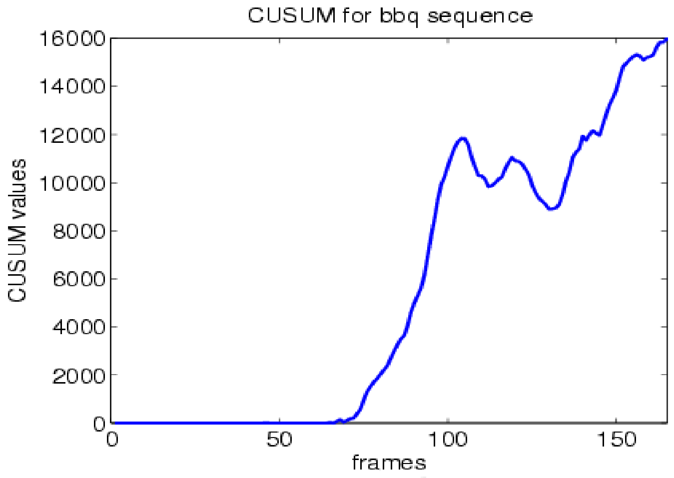
Experimental Results





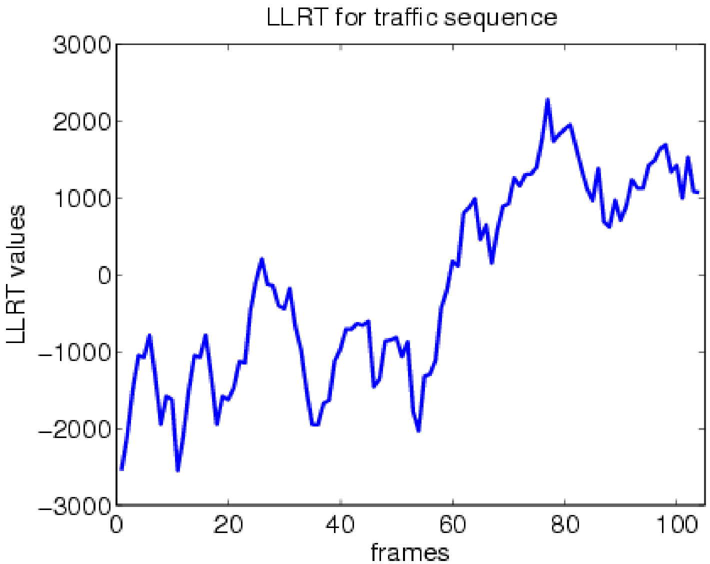
Experimental Results

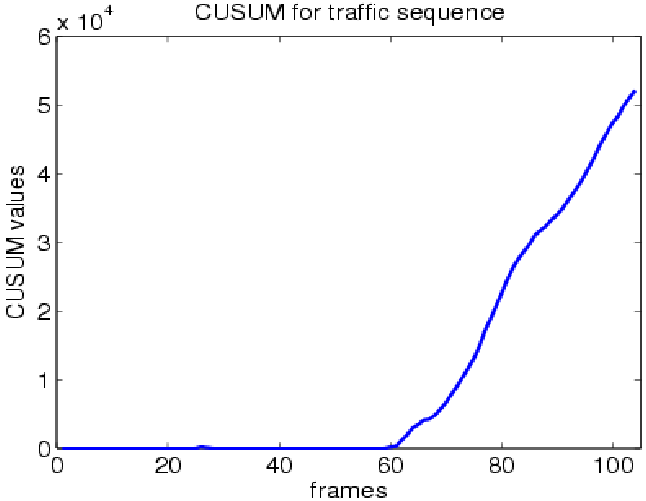




48

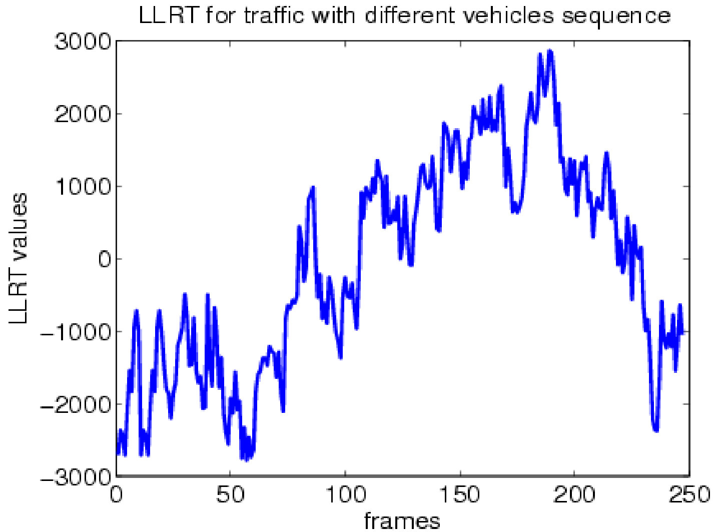
Experimental Results

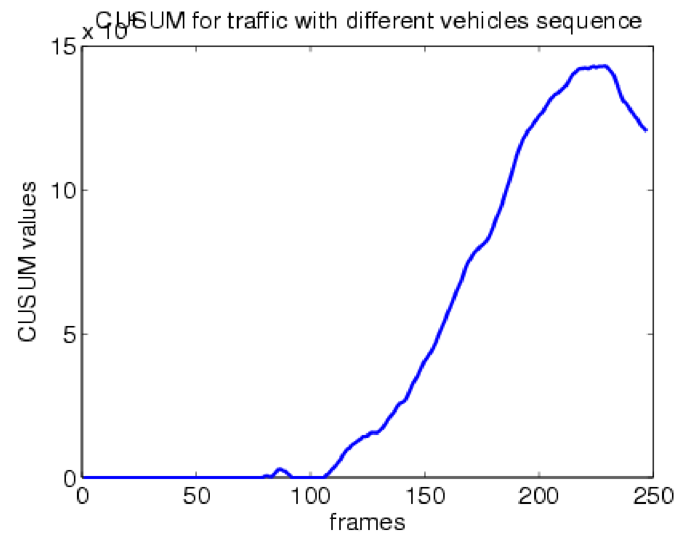




49

Experimental Results





50

Experimental Results

51

Experimental Results

Videos	Real Change	Det. change	Shot change det	Motion before change	Motion after change
Flowers	100	104	100	Motion in all directions	Directed motion
Shrubs	80	85	84	Faster leaf motion	Slower leaf motion
Forest Fire	130	131	130	Fire spread over many trees	Fire more concentrated
Shower	90	93	90	High water flow	Medium water flow
Waves	100	106	99	Waves flowing	Waves flowing faster
BBQ	90	106	3	BBQ flame to the left	flame to the right

Videos	Real Change	Det. change	Shot change det	Motion before change	Motion after change
Traffic 1	53	54	3	Light traffic	Heavy traffic
Traffic 2	140	142	3	Trucks	Cars
Walking	20, 35, 57	18, 37, 59	3	Kids exit, enter, exit	Walking
Crowd	130, 290	140, 300	476	Crowd meets, separates	Walk separately
Pedestrians	70, 210	200	297	Pedestrians meet, stop crossing	No pedestrians

Conclusions

• Sequential change detection can be used for the detection of changes in a variety of applications.
• Allows the processing of data online, as it arrives.
• This can lead to robustness to illumination changes due to its temporally local nature.
• Issues remain with the online estimation of data pdfs and the testing threshold.
•However these issues concern the area of sequential change detection in general.
•Alternative methods such as Sequential Probability Ratio Testing (SPRT) or non-sequential techniques can be used, but they still require prior knowledge of data distributions

53

Related journal publications

- A. Briassouli, I. Kompatsiaris, "Robust Temporal Activity Templates Using Higher Order Statistics", IEEE Transactions on Image Processing, Vol. 18, Issue 12, pp. 2756-2768.

- A. Briassouli, V. Tsiminaki, I.Kompatsiaris, "Human Motion Analysis via Statistical Motion Processing and SequentialChange Detection," EURASIP Journal on Image and Video Processing, vol. 2009, Article ID 652050, 16 pages, 2009. doi: 10.1155/2009/652050

- G. Th. Papadopoulos, A. Briassouli, V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "Statistical Motion Information Extraction and Representation for Semantic Video Analysis", IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 10, pp. 1513-1528, October 2009.

- A. Briassouli, V. Mezaris, I. Kompatsiaris, "Combination of Accumulated Motion and Color Segmentation for Human Activity Analysis", EURASIP Journal on Image and Video Processing, vol. 2008, Article ID 735141, 2008.

- Alexia Briassouli, Vasileios Mezaris, Ioannis Kompatsiaris, "Color aided motion-segmentation and object tracking for video sequences semantic analysis", International Journal of Imaging Systems and Technology (IJIST), Special Issue on Applied Color Image Processing, Volume 17, Issue 3, pp 174-189, 2007.

54

Related conference publications

• Alexia Briassouli, Ioannis Kompatsiaris, "Change Detection for Video", 2nd International Workshop in Sequential Methodologies (IWSM 2009), June 15-17 2009, Troyes, France.

•Alexia Briassouli, Ioannis Kompatsiaris, "Detection of Multiple Subevents in Space and Time for Video Analysis", WCVIM International Workshop on Computer Vision and its Application to Image Media Processing, January 13 2009, Tokyo, Japan.

•Alexia Briassouli, Ioannis Kompatsiaris, "Human Activity Localization via Sequential Change Detection", ACM Multimedia-MIR 2008, Oct 27-31 2008, Vancouver, Canada.

•Alexia Briassouli, Ioannis Kompatsiaris, "Statistical Processing of Video for Detection of Events in Space and Time", IEEE International Conference on Multimedia and Expo (ICME), June 23-26, 2008, Hannover, Germany.

•G. Th. Papadopoulos, A. Briassouli, V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "Semantic Video Analysis Based on Estimation and Representation of Higher-Order Motion Statistics", 3rd International Workshop on Semantic Media Adaptation and Personalization (SMAP '08), pp. 21-26, Prague, Czech Republic, December 2008.

•A. Briassouli, V. Mezaris, I. Kompatsiaris, "Video Segmentation and Semantics Extraction from the Fusion of Motion and Color Information", Proc. IEEE International Conference on Image Processing (ICIP 2007), September 2007, San Antonio, TX, USA, vol. III, pp. 365-368.

•A. Briassouli, V. Mezaris, I. Kompatsiaris, "Joint Motion and Color Statistical Video Processing for Motion Segmentation", Proc. IEEE International Conference on Multimedia & Expo (ICME 2007), July 2007, Beijing, China, pp. 2014-2017.

55

Brownies

BROWNIES

- 140 gr butter
- 280 grams sugar
- 82 grams unsweetened cocoa powder
- 1/4 teaspoon salt
- 1/2 teaspoon pure vanilla extract
- 2 large eggs, cold
- 66 grams flour

- Melt butter in a saucepan. Remove from heat when melted. Add
- Add cocoa, mix well. Add sugar, mix well. Add eggs, mix. Add flour, salt, mix.
- Add additional ingredients (cherries, chocolate chips, nuts).
- Pour in pan.
- Add cream cheese topping (see below). Bake at 180 C for 25-30 min.
- Cut into squares when cooled. EAT.

- CREAM CHEESE TOPPING
- 200g cream cheese, at room temperature
- 1 large egg yolk
- 5 tablespoons (75g) sugar
- 1/8 teaspoon vanilla extract

- Beat all ingredients until smooth. Put spoonfuls over brownie batter and swirl with a sharp knife. Bake at 180C for 25-30 min.

Blondies

BLONDIES

- 8 tablespoons butter, melted
- 1 cup brown sugar (= 1 cup white sugar + 1 tbsp molasses in the blender)
- 1 egg
- 1 teaspoon vanilla
- Pinch salt
- 1 cup flour

- Mix melted butter with brown sugar – beat until smooth. Beat in egg and then vanilla.
- Add salt, stir in flour. Mix in any additions (nuts, chocolate chips, whisky (!) etc).
- Pour into prepared pan. Bake at 180°C 20-25 minutes, or until set in the middle.
- Cut into squares when cooled. EAT.

Thank you for your attention!

Any questions?