SURVEILLANCE AND SPORTS VIDEO ANALYSIS

PROFESSOR DAVID MARSHALL

SCHOOL OF COMPUTER SCIENCE & INFORMATICS

CARDIFF UNIVERSITY

Partners in Crime:
Prof. Simon Moore (Dental School, Violence & Society Research Group Applied Clinical Research & Public Health)
Prof. Paul Rosin (Computer Science & *Informatics*)
Iwan Griffiths (Swansea Univ)
Sportsviz (www.sportsviz.com)
Rhys Long, Rhodri Bown (Welsh Rugby Union)
Daniel Howe, Paul McCarthy (South Wales Police)
PhD Students: Kaelon Lloyd, Aled Owen,

TGM Turing Gateway to Mathematics

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

SETTING THE SCENE

Two Areas:

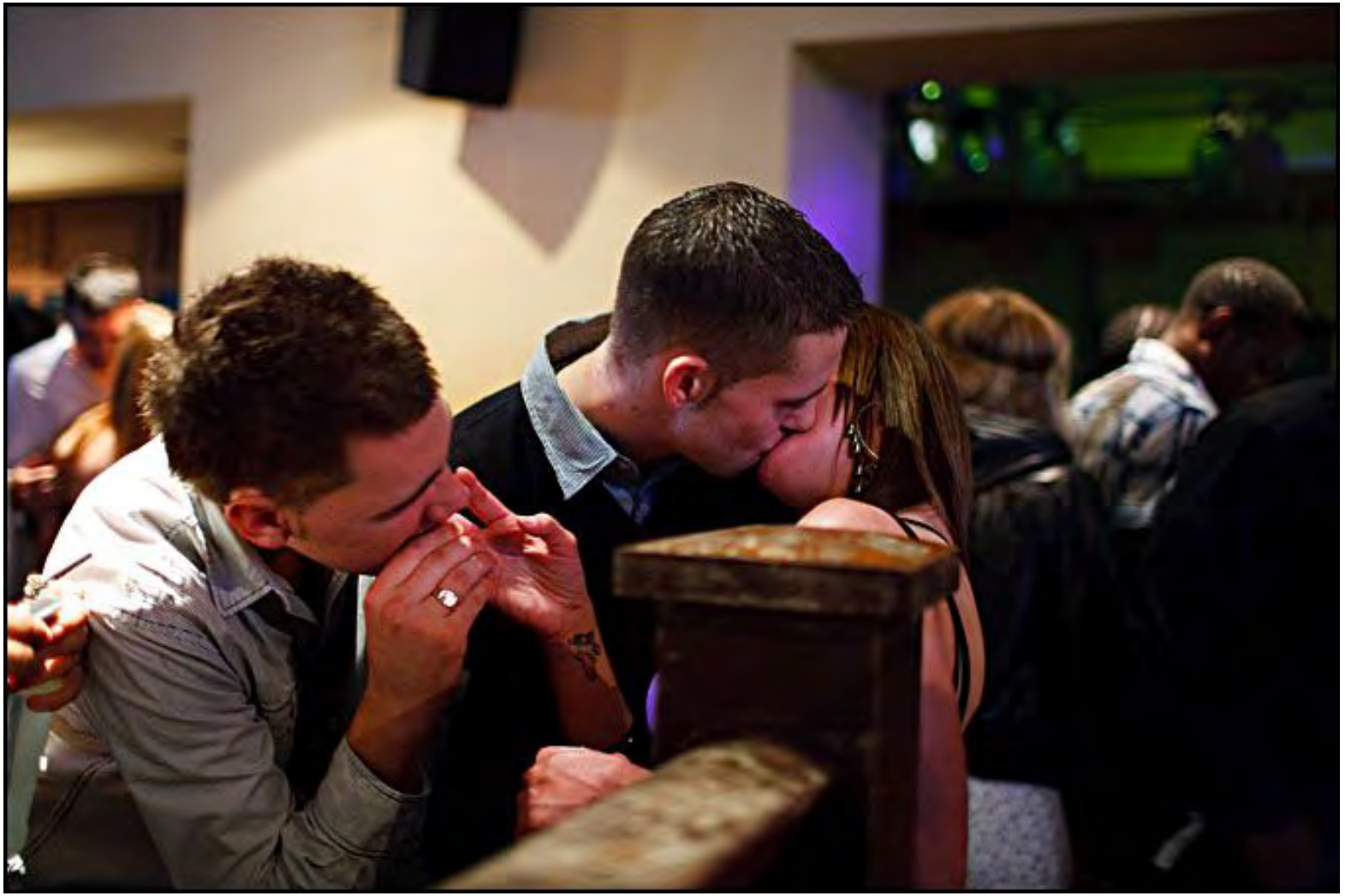
- **CCTV Surveillance Video**
  - **Night Time Economy**
  - Difficult in practice – Tough Conditions
  - Machine Learning/ Mining labelled video to recognise complex activities

- **Sports Video Analysis**
  - Machine Learning to the rescue
  - Mining labelled video to recognise complex activities
  - **Rugby Union Video Analysis**

A GOOD NIGHT OUT IN CARDIFF ☺

A GOOD NIGHT OUT IN CARDIFF ☺

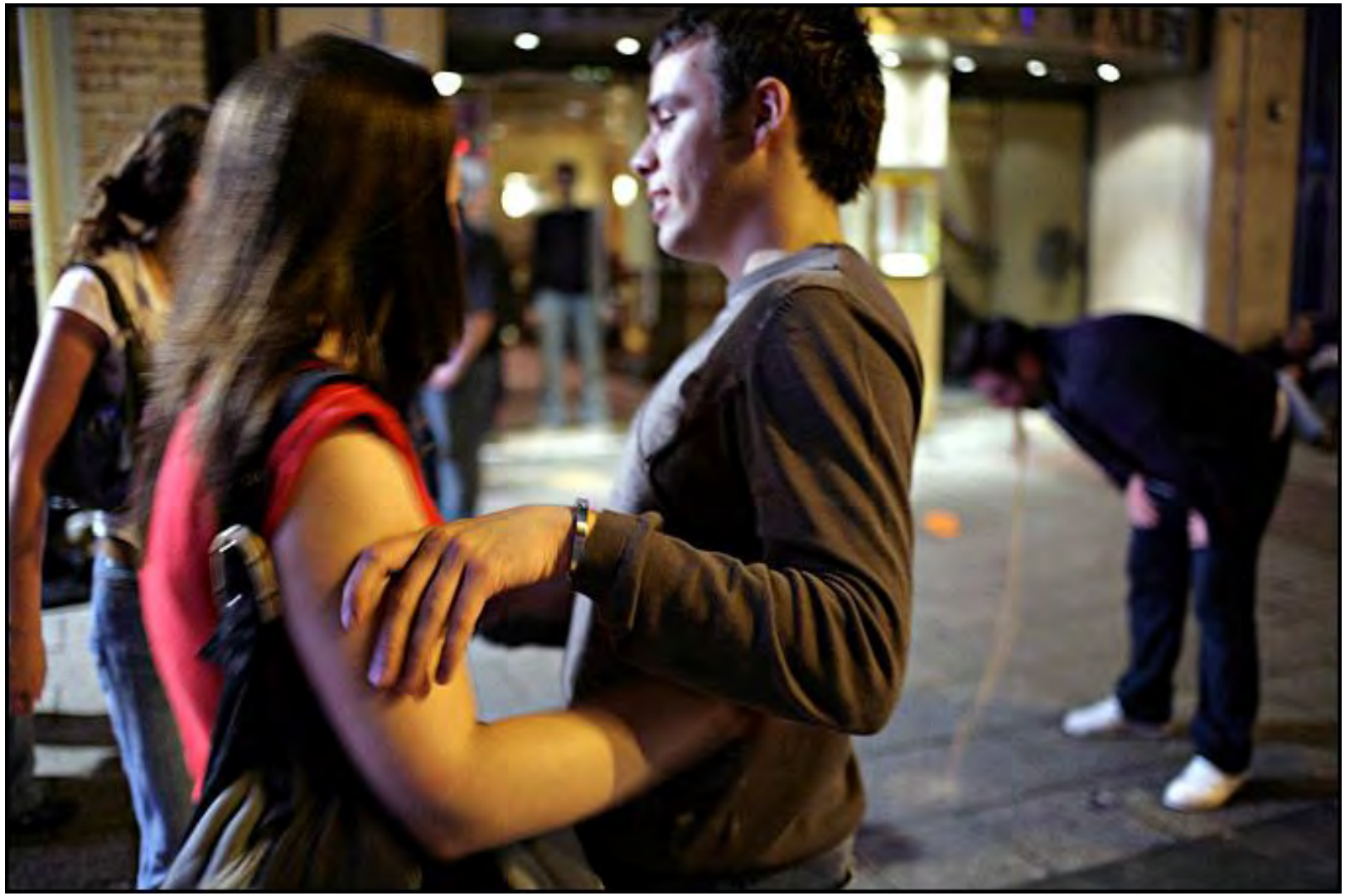
A GOOD NIGHT OUT IN CARDIFF ☺



A GOOD/BAD NIGHT OUT IN CARDIFF ☺/☹

A BAD NIGHT OUT IN CARDIFF ☹

A BAD NIGHT OUT IN CARDIFF ☹



A VERY BAD NIGHT OUT IN CARDIFF ☹

A GOOD/BAD NIGHT OUT IN CARDIFF?

AUSTERITY BITES

BBC | News | Sport | Weather | iPlayer | TV | Radio | More

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education

UK | England | N. Ireland | Scotland | Alba | Wales | Cymru

CCTV: Too many cameras useless, warns surveillance watchdog Tony Porter

26 January 2015 | UK

BBC | News | Sport | Weather | iPlayer | TV | Radio | More

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education

Wales | Wales Politics | North West | North East | Mid | South West | South East

Lampeter council abandons CCTV surveillance project

29 April 2014 | Mid Wales

Ceredigion will have no CCTV provision to tackle crime this year after Lampeter council abandoned a plan to continue the service in the town.

BBC | News | Sport | Weather | iPlayer | TV | Radio | More

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education

Wales | Wales Politics | North West | North East | Mid | South West | South East

Review of CCTV service in Dyfed-Powys in the offing

19 March 2014 | Mid Wales

Three councils have been asked to help fund a review of CCTV provision to tackle crime in the Dyfed-Powys area.

The area's police and crime commissioner has written to Ceredigion, Carmarthenshire and Pembrokeshire

PROBLEM: DETECTING VIOLENT INCIDENTS IN SURVEILLANCE VIDEO

VIDEO SURVEILLANCE

- Video surveillance is a remote observation method
- **Passive**
  - Recorded footage is archived for later use
- **Active**
  - Real-time observation used to identify illegal or disorderly behaviour
  - Can be used to direct ground units to a scene of interest

VIDEO SURVEILLANCE

- The number of installed cameras is between 1.85 and 4.20 million in Britain alone
  - Each person falls into view of a camera an estimated 70 times a day
  - The amount of data captured daily <u>cannot feasibly be analysed manually</u>
- Gill et al (2005) found that the installation of CCTV systems results in a reduced crime rate
- Violence increased after CCTV was installed[1][2][3]
  - We can see more, therefore identify more!

[1] M. Gill, A. Spriggs, *Assessing the impact of CCTV: 2005 Home Office Research Study 292. London: Home Office.*

[2] C Florence, J Shepherd, I Brennan, T Simon, *Effectiveness of anonymised information sharing and use in health service, police, and local government partnership for preventing violence related injury: experimental study and time series analysis*

[3] V. Sivarajasingam, J.P. Shepherd, K. Matthews; Effect of urban closed circuit television on assault injury and violence prevention. 2003

IMPORTANCE OF SURVEILLANCE

- 52 violence related emergency department visits per week in Cardiff in 2006[2]
- Research showed that emergency treatment due to street violence decreased when CCTV was adopted[3] [2]
- The sooner violence is identified, the less serious the potential damage will be. [3][2]
  - Failing to identify violence can be lethal!

WHY AUTOMATE?

- Too much data
  - It is <u>not feasible</u> to manually evaluate all recorded data
- Humans are flawed
  - Require Sleep
  - Lose Focus
  - Blink
  - Incompetence

www.metro.co.uk/news/889691-plain-clothes-police-office
proxies musicnoteslib SomaFM Read Later Instapaper Delicious Delicious ME Horde Core Admin Wiki SIMS home svr LearnCentral To Papers Sheetmusicscore
Plain clothes police officer chases himself after CCTV burglar bungle | Metro News

Plain clothes police officer chases himself after CCTV burglar bungle

Hayden Smith Tuesday 7 Feb 2012 6:56 pm

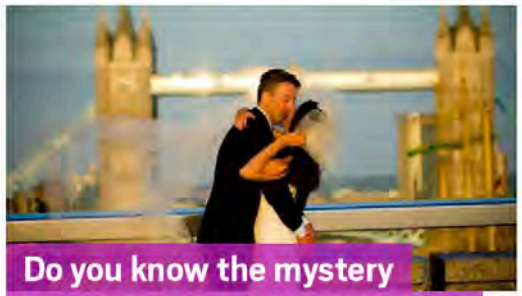
122

A plain clothes police officer chased himself for 20min after he was mistaken on CCTV for a burglary suspect.

The PC was working after hours in a market town that had suffered a spate of break-ins when a CCTV operator reported a man âacting suspiciouslyâ in a side road.

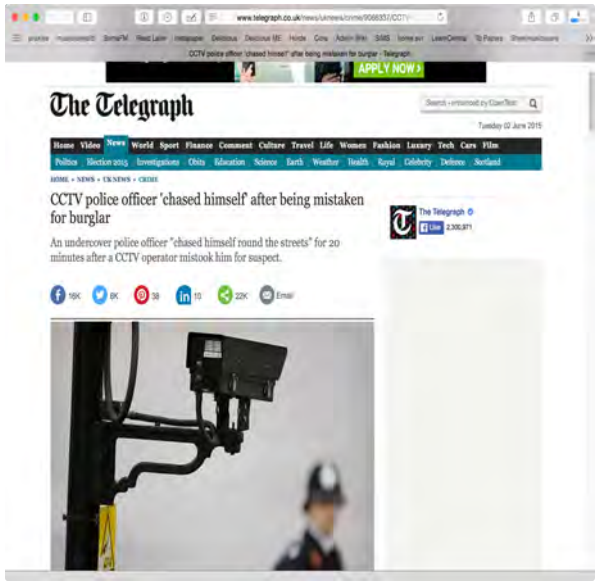
Unaware he had been mistaken as the suspect, the

MUST READ


Do you know the mystery

A plain clothes police officer chased himself for 20min after he was mistaken on CCTV for a burglary suspect.

- A CCTV operator reported a man 'acting suspiciously' in a side road.
- Unaware he had been mistaken as the suspect, the probationary Sussex police officer began to chase the elusive figure.
- As the CCTV operator, oblivious to his mistake, tracked the suspect's position, the plain clothes officer seemed to find himself in the same street.
- But despite being told he was 'on the heels' of his prey by the operator, the energetic officer was always one step behind the shadowy suspect.
- About 20min into the pursuit, a sergeant came into the control room and realised the suspect and the officer were the same person.



WHY AUTOMATE?

- Research has shown that when presented with multiple video feeds human effectiveness drops dramatically[4]

[4] G. Voorthuijsen, H. Hoof, M. Klima, K. Roubik, M. Bernas, P. Pata, *CCTV Effectiveness Study,* Security Technology, 2005. CCST '05. 39th Annual 2005 International Carnahan Conference on, 2005

WHY AUTOMATE?

- Norris and Armstrong observed multiple surveillance operating rooms
  - Operators would invite friends into the work environment
  - Internal conflicts between active operators on what to report to authorities
  - The job is too strenuous when undertaken in 8 and 12 hours shifts
- Operators applied knowledge derived from personal experience to focus the observation process
  - This behaviour is encouraged, but often led to racial discrimination

[5] C. Norris, G. Armstrong; The Maximum Surveillance Society: The Rise of CCTV (1999)

WHY AUTOMATE?

- Computers can process data faster than humans
- Computers do not lose focus
- Computers do not discriminate

COMPUTATIONAL CHALLENGES

- Footage is poor quality
  - Low frame rates
  - Low resolution
- Environments are constantly changing
  - Illumination
  - Weather
- Urban environments contain dense crowds

COMPUTATIONAL CHALLENGES

DETECTION METHODS

- **Violent Flows**: Motion texture descriptor that uses optical flow fields.
- **Motion Binary Pattern**: Motion texture descriptor that estimates motion intensity using local changes in pixel intensity.
- **LBP+TOF**: Local Binary Pattern Pattern, Dynamic Textures.
- **STIP+HOG+HOF**: An interest point detector with HOG and HOF descriptors.
- **OUR METHOD: Grey Level Co-Occurrence Texture Measures** with *_EDGE CARDINALITY AND PIXEL INTENSITY DIFFERENCE (GEP)*
    - State-of-the-Art methods **failed** at detecting violence in real-world data
    - A different approach was needed!

VIOLENT FLOWS

- Measures change in Optical Flow Magnitudes between Frames

$$m_{x,y,t} = \sqrt{u_{x,y,t}^2 + v_{x,y,t}^2}$$

- A Mean Magnitude Map:

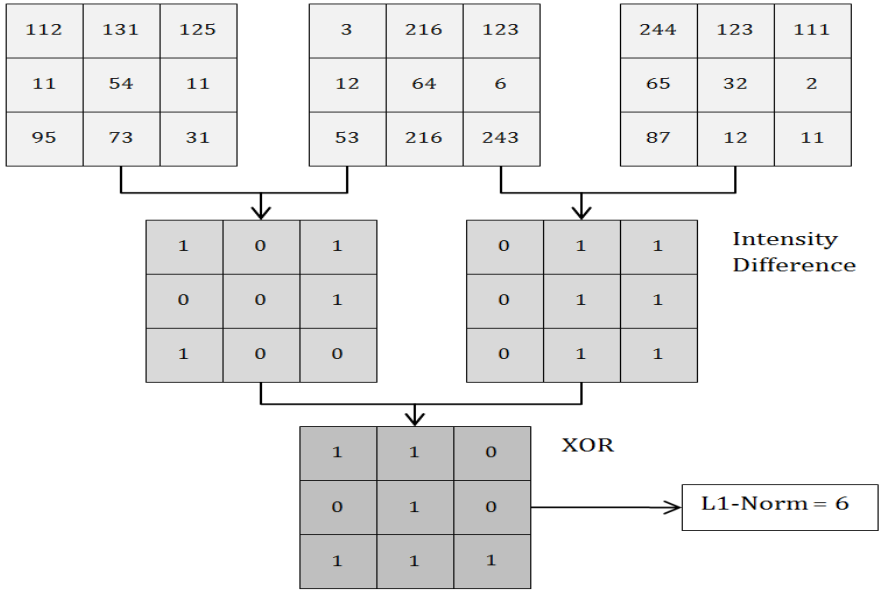
$$b_{x,y,t} = \begin{cases} 1 & \text{if } |m_{x,y,t} - m_{x,y,t-1}| > \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\bar{b}_{x,y,t} = \frac{1}{T} \sum_t b_{x,y,t}$$

Hassner et al., .Violent flows: Real-time detection of violent crowd behavior. CVPR Workshop, 2012

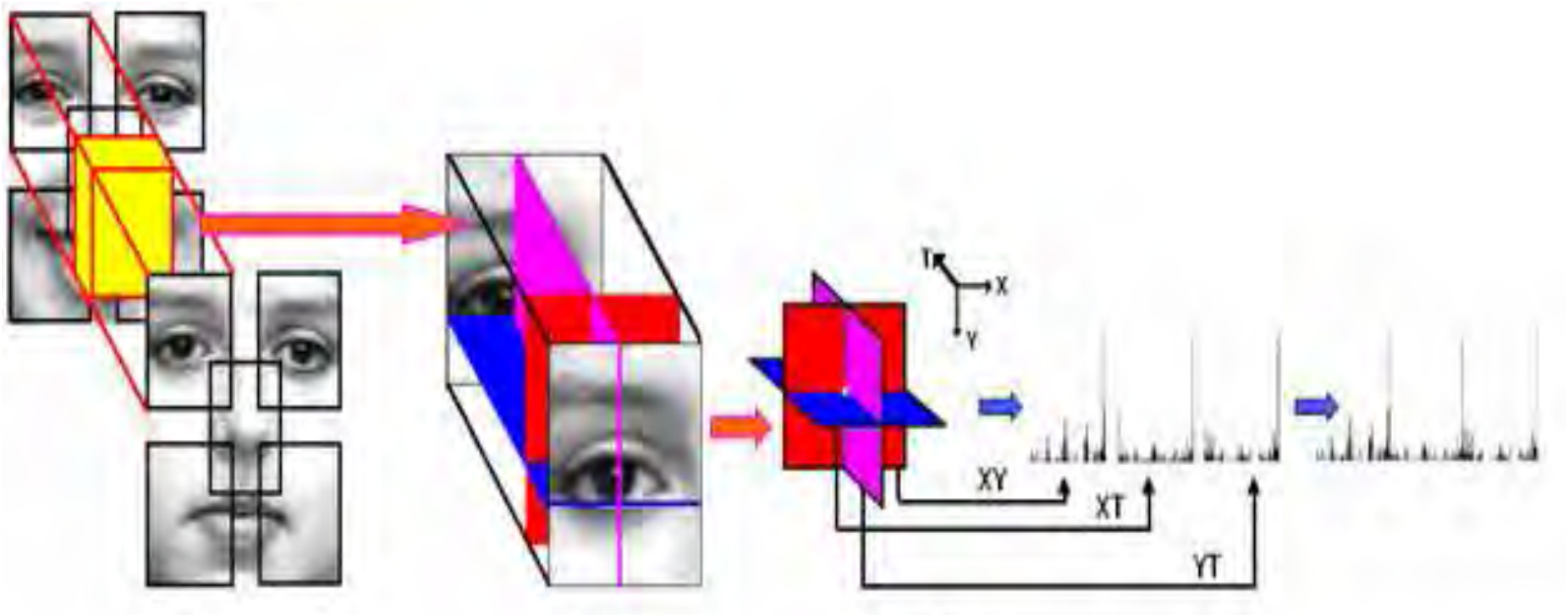
MODIFIED MOTION BINARY PATTERN

- Assumption: motion can be detected from the change in pixel intensities.
- MBP requires three frames to describe motion.
- Bag Of Words from sequence of Frames



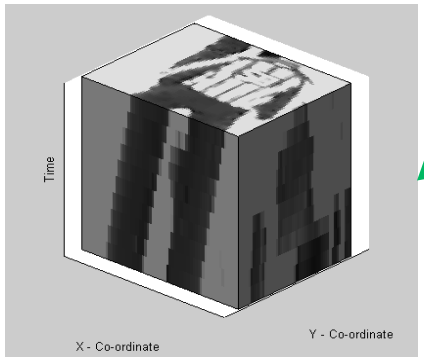
Baumann et al. Motion binary patterns for action Recognition.
3rd International Conference on Pattern Recognition Applications and Methods , 2014.

LBP+TOF: LOCAL BINARY PATTERN PATTERNS, DYNAMIC TEXTURES



G. Zhao and M. Pietainen. Dynamic texture recognition using local binary patterns with an application to facial expressions PAMI, 29(6), 2007

SPACE TIME INTEREST POINT (STIP) WITH HOG + HOF



Top Left: Image,

Top Right: STIP points ("3D" Harris)

Bottom Left: Extracted Volume around interest point

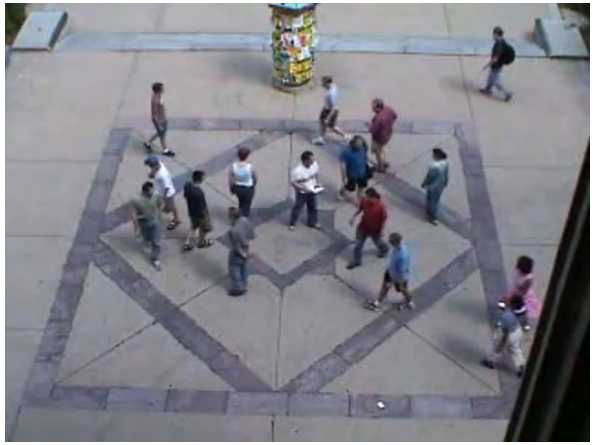
HOG + HOF Features from Volume.

Laptev and Lindeberg. Space-time interest points. ICCV 2003.

GREY LEVEL CO-OCCURRENCE TEXTURE MEASURES WITH *EDGE CARDINALITY AND PIXEL INTENSITY DIFFERENCE (GEP)*

- Feature Vector:
  - GLCM Texture Energy
  - GLCM Texture Contrast
  - ***EDGE CARDINALITY*** – *CANNEY EDGE DETECTOR: NUMBER OF EDGES.*
  - ***PIXEL INTENSITY DIFFERENCE*** – *NORMALISED PIXEL DIFFERENCE IN ADJACENT FRAMES*

TEXTURE

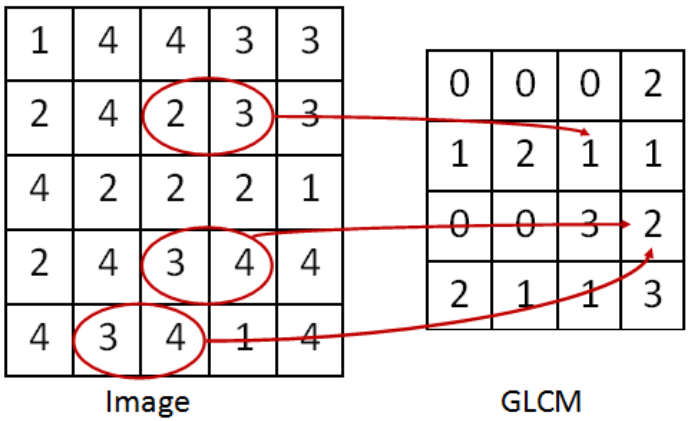


- Marana et al (1998) treated crowd density estimation as a measure of texture statistics
  - Texture is well suited for describing the random nature of patterns that appear from self-occluding crowds

[6] A.N. Marana, S.A. Velastin, L.F. Costa, R.A. Lotufo; Automatic Estimation of Crowd Density Using Texture; 1998

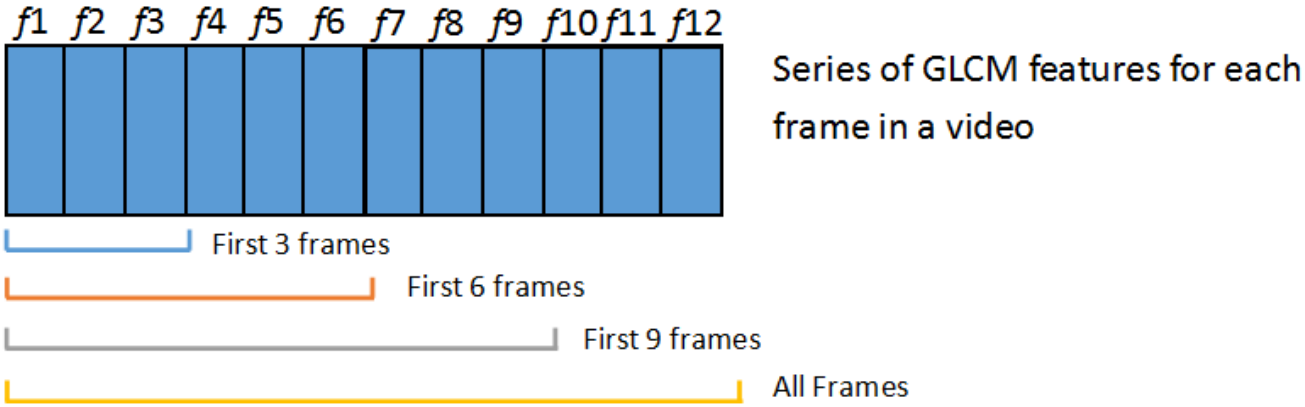
GREY LEVEL CO-OCCURRENCE MATRIX (GLCM)

- The proposed method treats a video as a series of texture measurements
- Texture is computed using a Grey Level Co-occurrence Matrix (GLCM) from which meaningful statistics are derived
- Images have pixels that have a range of values typically from [0 - 255]
- A GLCM is a frequency matrix of the number of times a pair of pixels are depicted in an image



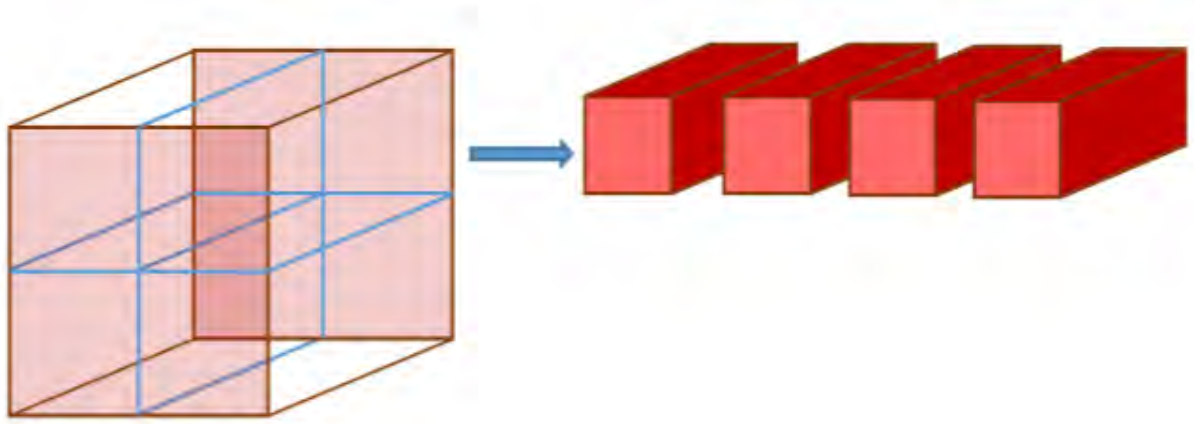
GEP FEATURES

- From the GLCM four measures are calculated that describe the visual texture
  - Energy, Correlation, Contrast and Homogeneity
  - We have these four values for every frame in a video
- We calculate the mean and standard deviation of each of these four measures across a series of frames, this describes how each visual feature changes
- We want to do this over various frame ranges of different length so to capture small and large changes in motion



SPATIAL BINNING

- In video, different areas of the frame will depict different actions; so we split the frame spatially
- Then we apply the steps mentioned previously to each sub-video



GEP: TEXTURE + OPTICAL FLOW (OF)

If robust optical flow fields can be computed

- Compute GEP GLCM Features on OF instead of grey level intensities
  - Dense optical flow fields via Lucas-Kanade [19] approach.
- Add to GLCM Texture Features

DEMONSTRATION

DEMONSTRATION

OTHER DATA SETS

Violent Flows Data

- **UMN :** 11 separate video samples Normal v. Panic
- **Web Crowd Abnormality :** 20 videos depicting either normal or abnormal crowd behaviour – Panic, Clash or Fight

DRAWBACKS

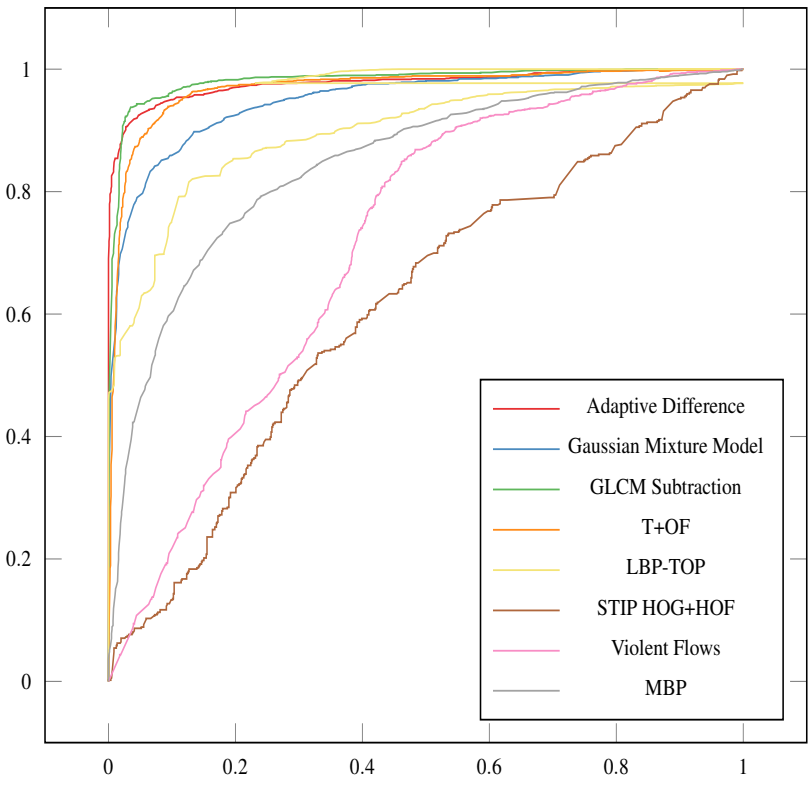
- Camera movement causes issues
  - Operator invoked movement
  - Automated movement routine
- The adopted method is not view independent
- Initial research shows promise but the dataset provided was limited
  - Small number of violent scenes

RESULTS ON CARDIFF DATA

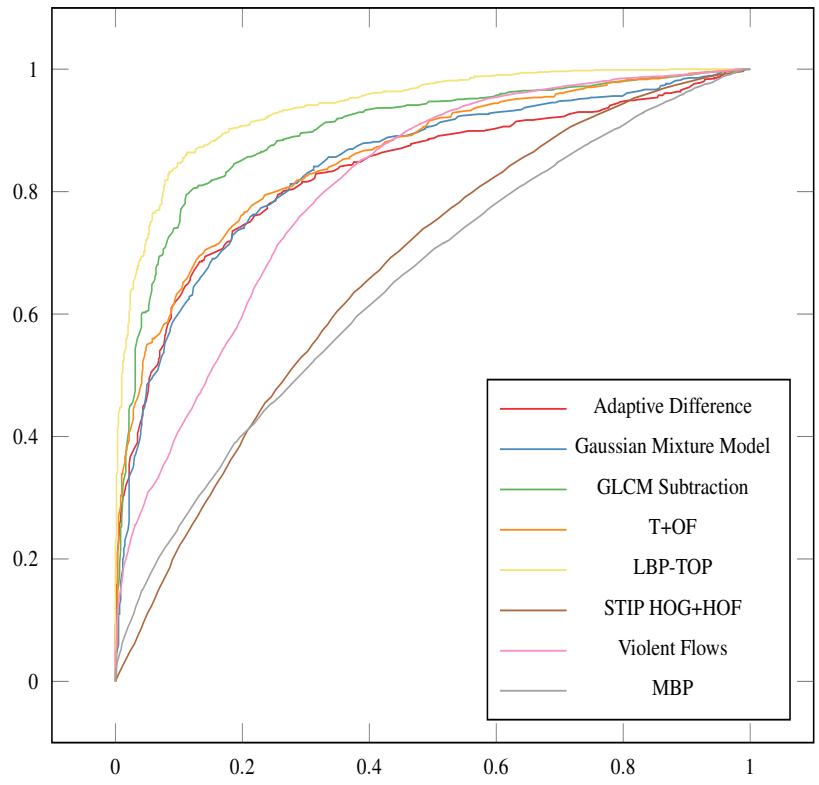
Method	TP	TN	Accuracy
GEP (Adaptive Difference)	71.58	97.82	84.7
GEP (GMM)	60.38	97.34	79.1
GEP(Subtraction)	**83.94**	97.89	**90.88**
GEP-OF (T+OF)	70	97.34	83.67
LBP-TOP	58.13	**99.68***	78.91
STIP HOG+HOF	46.51	96.42	1.47
Violent Flows	56.29	86.12	71.2
MBP	56.49	89.27	72.88

*LBP-TOP is **NOT** real-time

COMPARATIVE RESULTS: ROC CURVES

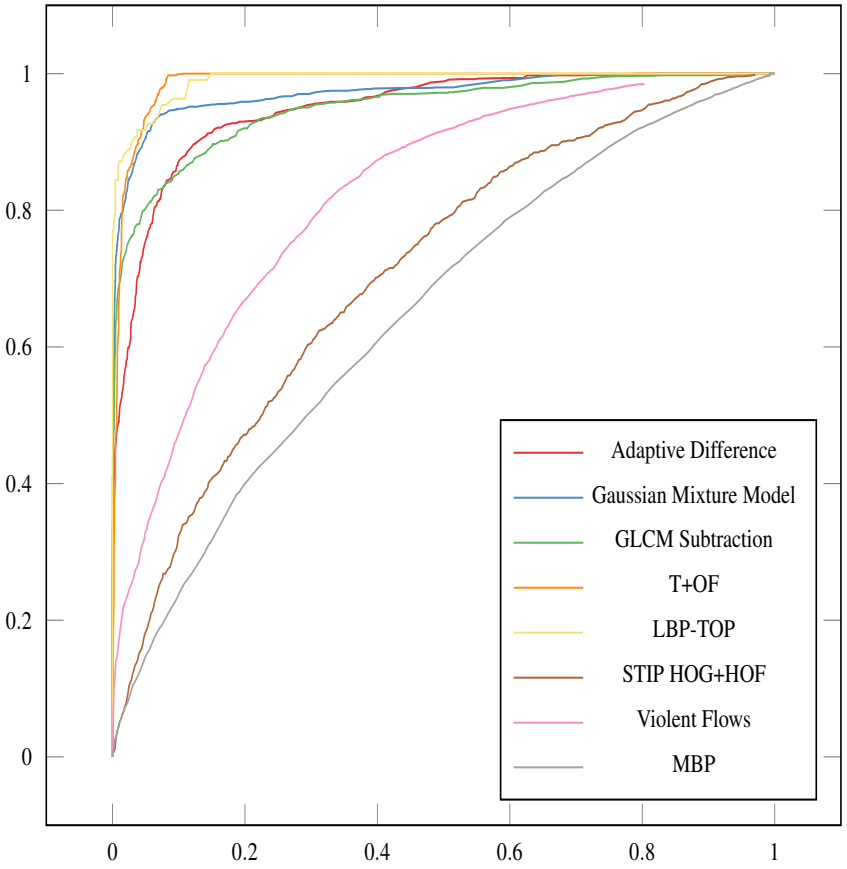


Cardiff

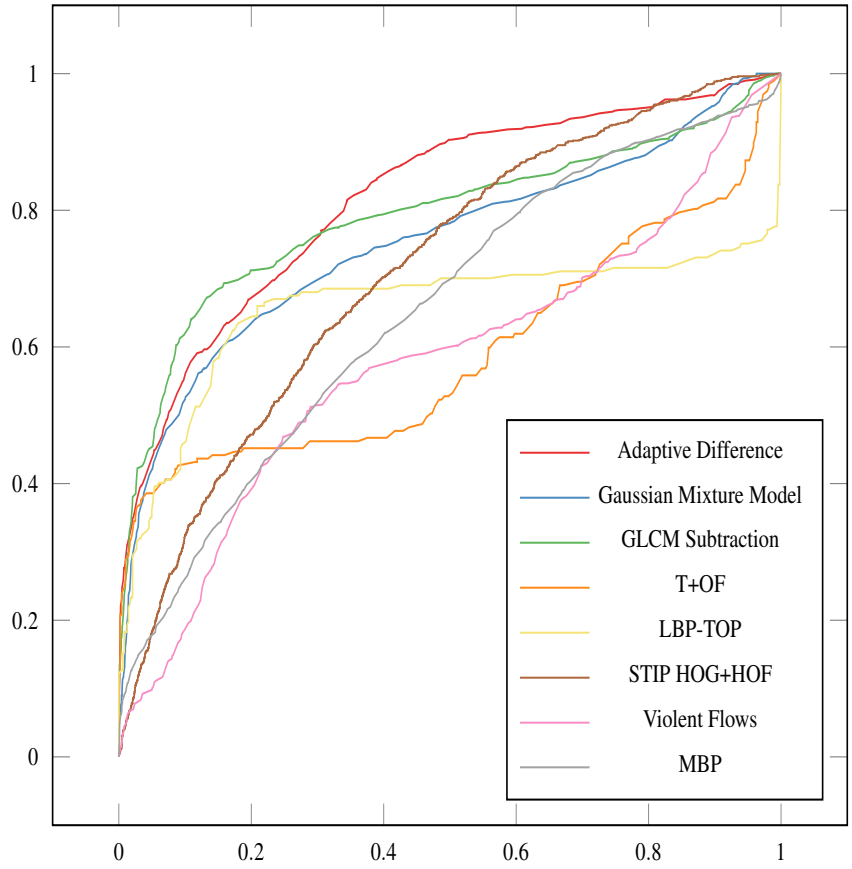


Violent Flows

COMPARATIVE RESULTS: ROC CURVES



UMN



Web Abnormality

COMPARATIVE RESULTS: AREA UNDER RECEIVER OPERATOR CHARACTERISTICS CURVE FOR EACH DATASET

Method	Real World	Violent Flows	UMN	Web
GEP (Adaptive Difference)	0.9712	0.8367	0.9472	**0.8002**
GEP (GMM)	0.9501	0.8362	0.9719	0.7503
GEP(Subtraction)	**0.9838**	0.8963	0.9445	0.7982
GEP-OF	0.9502	0.8611	**0.9845**	0.5837
LBP-TOP	0.8998	**0.9386***	0.9724	0.6882
STIP HOG+HOF	0.6136	0.6723	0.8614	0.6061
Violent Flows	0.7443	0.8367	0.8212	0.5929
MBP	0.843	0.6523	0.6518	0.6555

CONCLUSION

- Video surveillance provides a means of effectively identifying scenes of misconduct and criminal behaviour
- Observation allows human operators to manually manage law enforcement assets
- The human observer is not able to witness every scene recorded, an automated method is suggested to aid this process
- A method based on visual texture analysis that is capable of classifying between scenes of violence and non-violence in real world environments.

FUTURE WORK

- Aim to develop a more comprehensive model of crowd dynamics in order to analyse correlations between crowd behaviour and various events
- Provide useful information that can be used for effective police asset deployment
  - Example: Based on predicted crowd behaviour, map out the most efficient route for an officer to take in order to reach a place of interest
  - GPS data on Police Assets

SPORTS VIDEO ANALYSIS

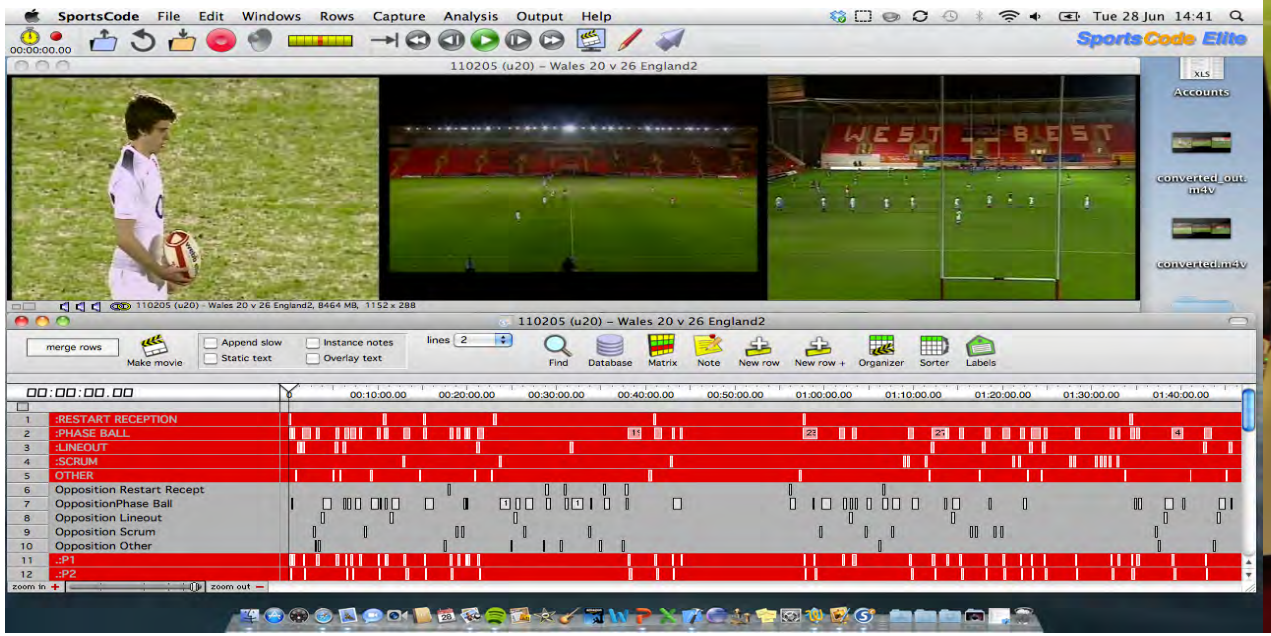
- Sportsviz project – WRU data
- Massive resource of labelled video data
- Machine Learn and mine data
- **Automatic Video Labelling**
- **Event detection/modelling/analysis**

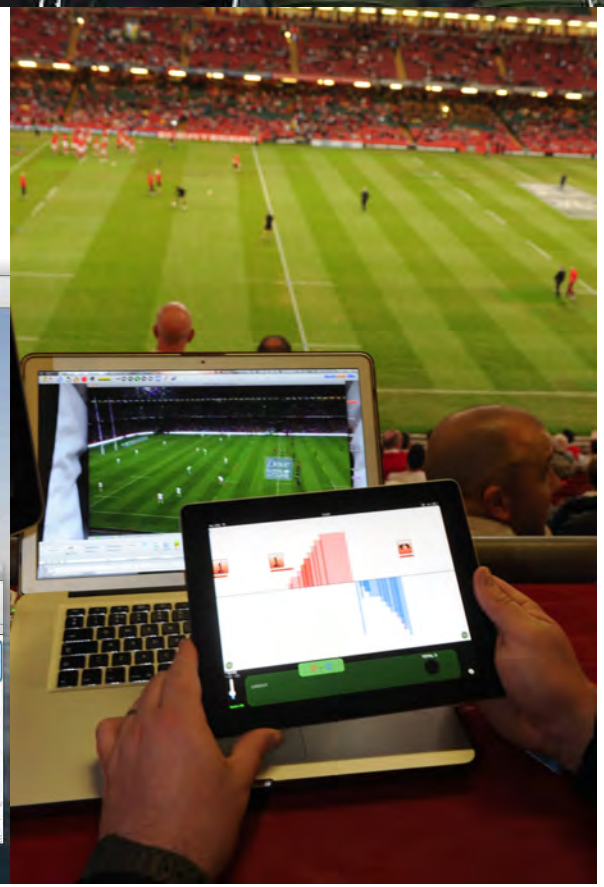
Labelled Data

Welsh Rugby Union record all matches

- Richly Labelled
  - Game and Player Stats
  - Live and offline

MANUAL ANNOTATION!





AUTOMATED RUGBY VIDEO ANALYSIS
DEMO: INPUT VIDEO

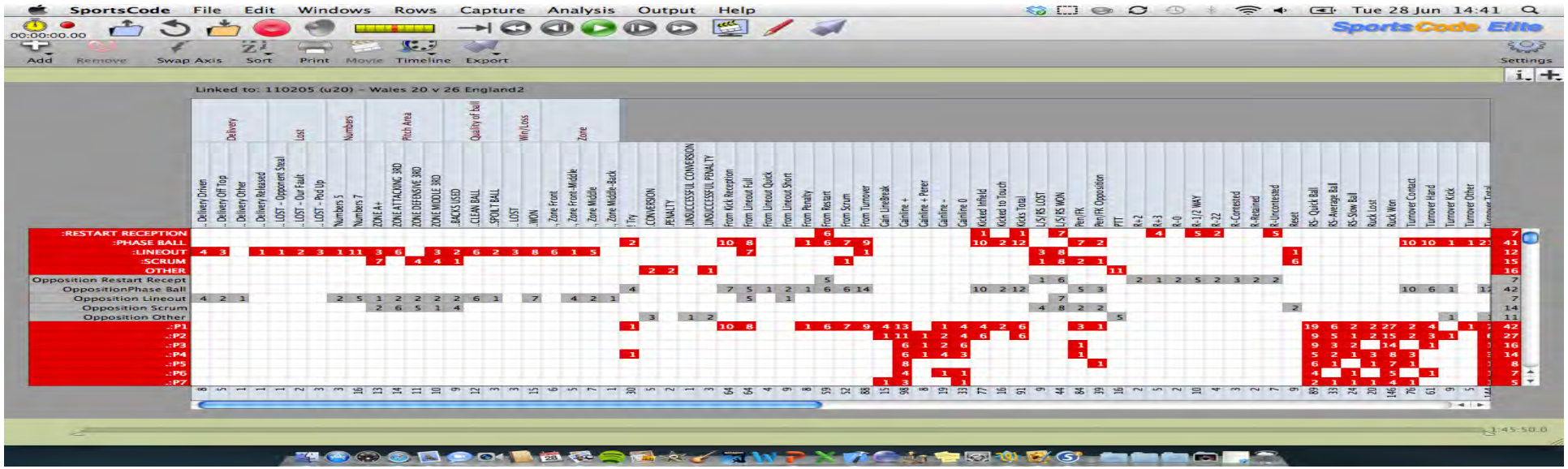
Can we recognise: **Scrum**, **Lineout, Ruck, Maul** or **Neither**?

MACHINE LEARNING: TRAINING DATA

200 Frames from Wales v Scotland each

- **Scrums, Lineouts, Rucks, Mauls**
- **"None":** random sample of non-scrum/lineout/ruck/maul frames

- WRU Annotated Video
  - Extract out XML Tags to retrieve Frames
- Train a Video Classifier



MACHINE LEARNING: TRAINING DATA: SCRUMS

MACHINE LEARNING: TRAINING DATA: LINEOUTS

MACHINE LEARNING: TRAINING DATA: "NONE"

TRAINING PREPROCESSING:
PITCH AND CROWD REMOVAL

MACHINE LEARNING: EXTRACT IMAGE FEATURES

- Bag of Words/Image Features (SIFT)

- Classifier: Learns correlation between Features and their locations and given class labels

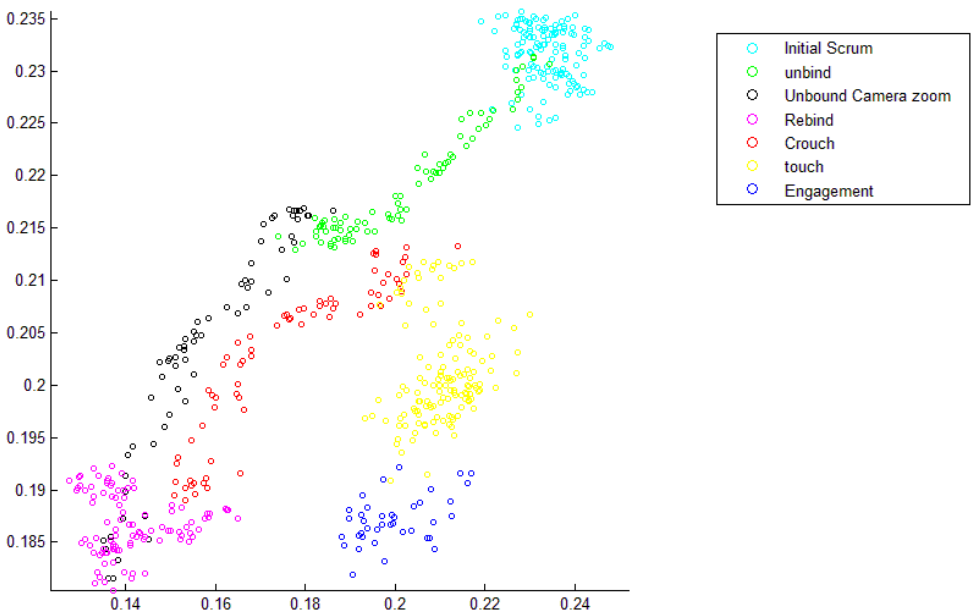
CLASSIFICATION: WALES V FRANCE LINEOUTS
COMPLETELY UNSEEN 100%!

CLASSIFICATION: WALES V ENGLAND: SCRUMS
COMPLETELY UNSEEN

CLASSIFICATION: WALES V ENGLAND LINEOUTSS
COMPLETELY UNSEEN

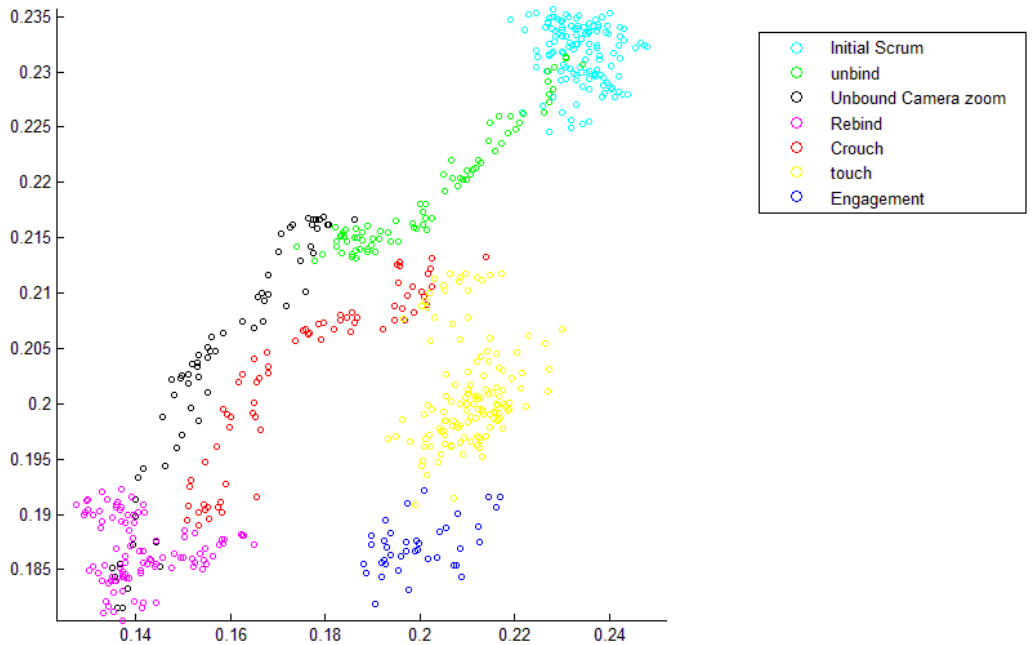
TEMPORAL MODELLING OF RUGBY EVENTS

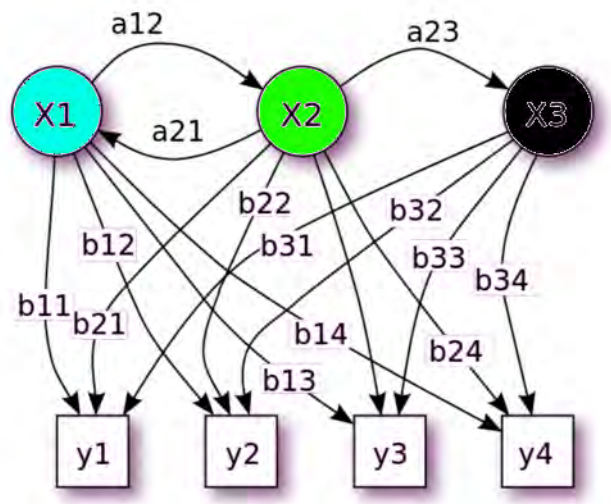
- Taking the existing training data, a scrum sequence was further classified.
- Cluster the SIFT Features.
- Trajectory through Cluster Space



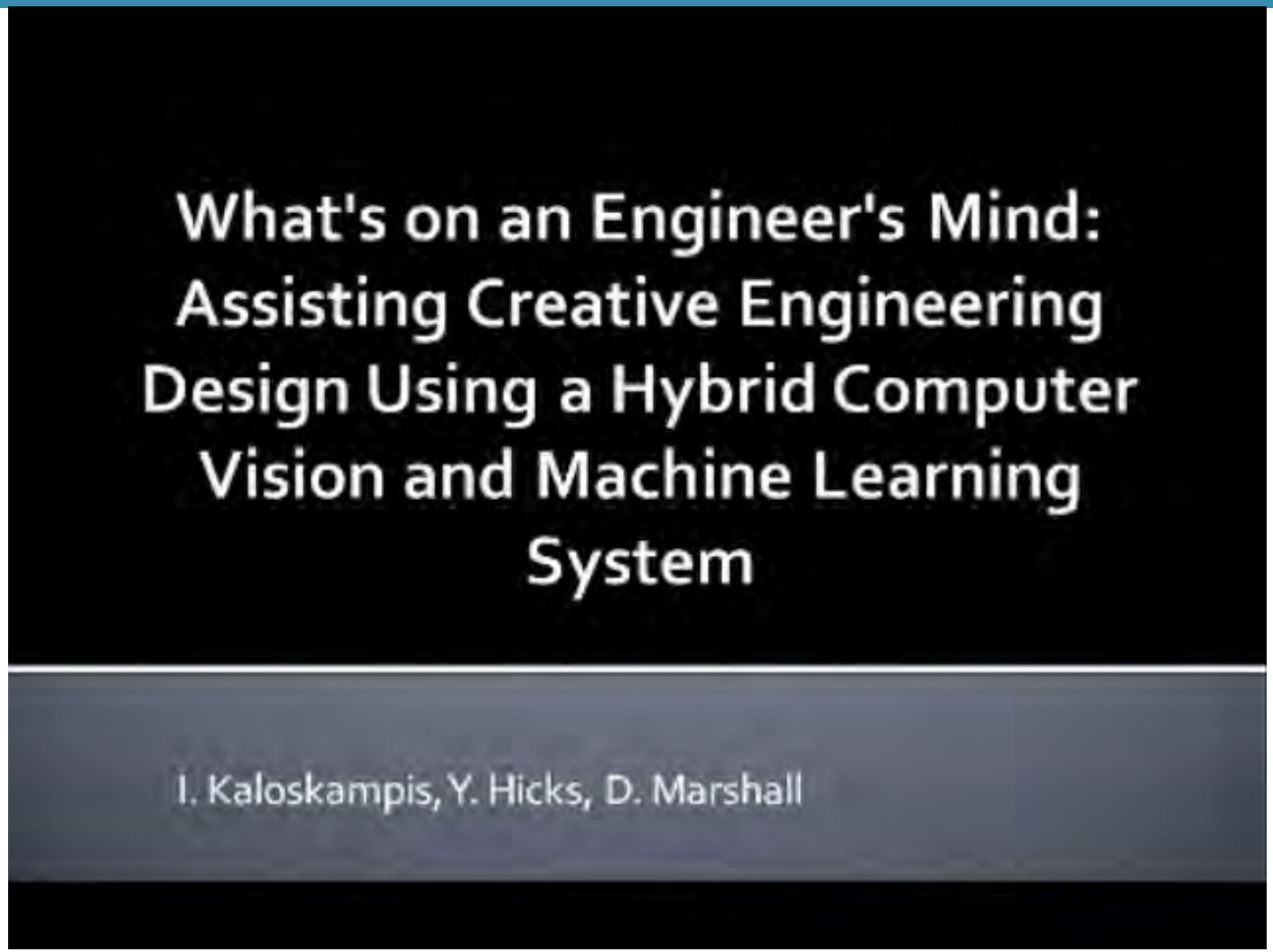
TEMPORAL MODELLING OF RUGBY EVENTS

- Identify a scrum and eventually other events in rugby using a temporal model?
  - (Hierarchical) Hidden Markov Modelling (HMM)?
  - Variable Length Markov Model (VLMM)?
- Issues: Variable Length and Variable Outcomes





ACTIVITY RECOGNITION: MORE ADVANCED TEMPORAL MODELLING OF VIDEO



I. Kaloskampis, Y. A. Hicks, and D. Marshall, "Automatic analysis of composite activities in video sequences using Key Action Discovery and hierarchical graphical models," ICCV'11 Workshop

RUGBY VIDEO SUMMARY

- Promising Start
  - State of **the Art Machine Learning** **does work!**

What's Next

- All views
- More training/testing
  - Input ore games
  - More classes/labels – tags from XML
- Position on Pitch detection and distance between plays
- Player detection/recognition/classification
  - Use audio/text labelling, if available?
- Full Temporal model of a game of Rugby

OVERALL SUMMARY

- Video Analysis still an active research problem
  - Challenging Scenarios need new approaches
  - Modern Machine Learning Approach Works
  - Advanced Temporal Analysis Under Investigation

ANY QUESTIONS?