Data-driven Crowd Analysis in Videos

Mikel Rodriguez Josef Sivic Ivan Laptev Jean-Yves Audibert

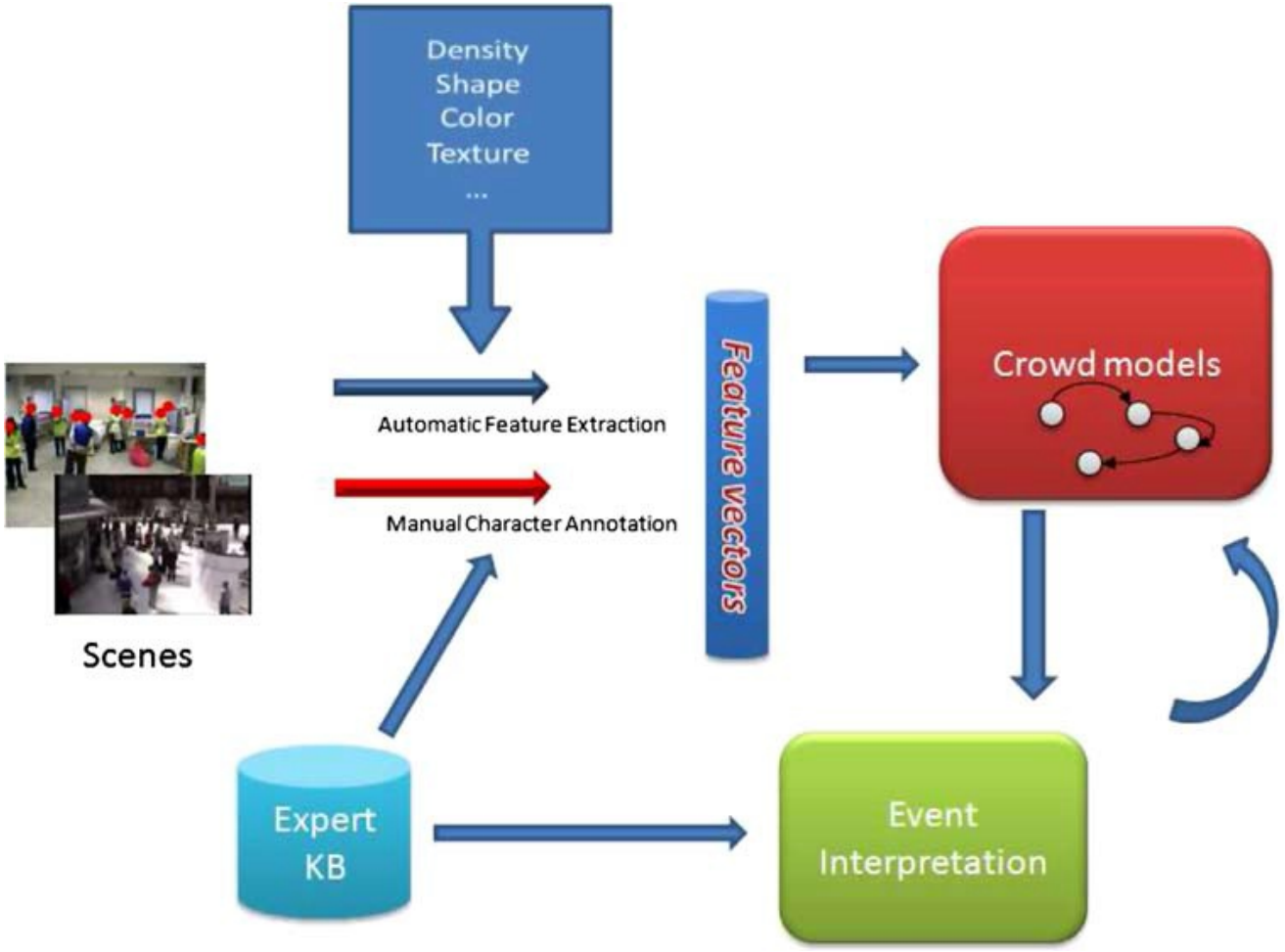
WILLOW project

Presented by Ezgi Mercan

Crowd Analysis

Crowd analysis: a survey, Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., Xu, L., Machine Vision and Applications, Vol 19, No 5-6, p. 345-357, DOI: 10.1007/s00138-008-0132-4.

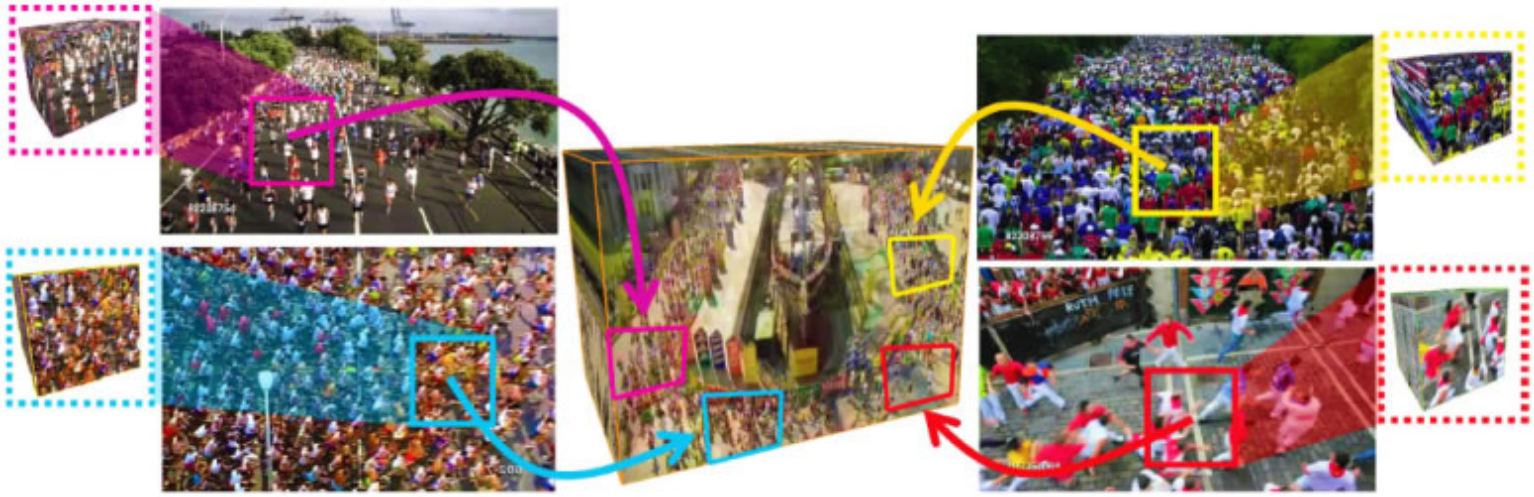
Crowd Analysis



Crowd analysis: a survey, Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., Xu, L., Machine Vision and Applications, Vol 19, No 5-6, p. 345-357, DOI: 10.1007/s00138-008-0132-4.

Data-driven Crowd Analysis

- Any given video can be thought as being a mixture of previously observed videos.



Learning Motion Patterns

database

Global Matching

similar scenes

test video

find similar scenes

query for similar scenes

Local Matching

find similar cells

query for similar cells

cell of interest

view of Method

similar cells

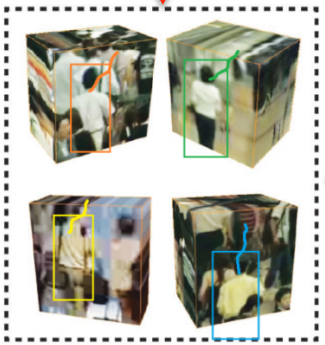
test video

individual to track

get motion patterns

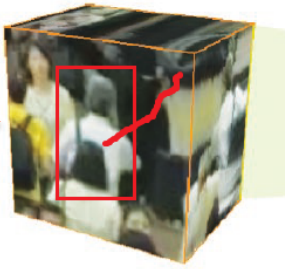
Tracking using Motion Patterns

similar cells



motion patterns of similar cells

predict motion



Learning Motion Patterns

database

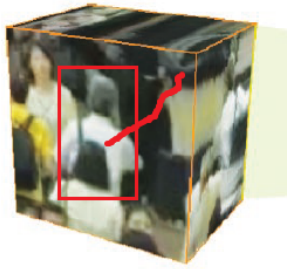
Global Matching

similar scenes

test video

Local Matching

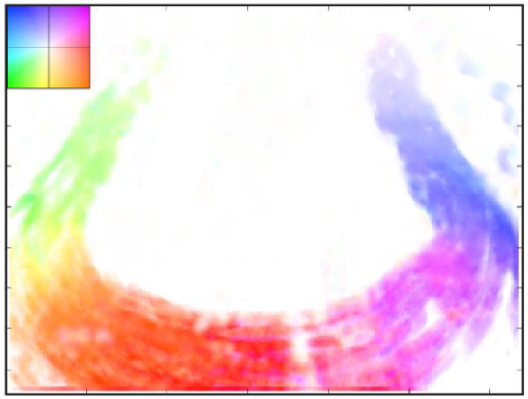
Tracking using Motion Patterns



Learning Motion Patterns

Low-level Representation: Dense Optical Flow

- For each pixel in each frame, calculate average optical flow.
- Combine the optical flow vectors into a global motion field for a temporal window.
  - temporal window ω = 60 frames
  - spatial window 20 pixel x 20 pixel

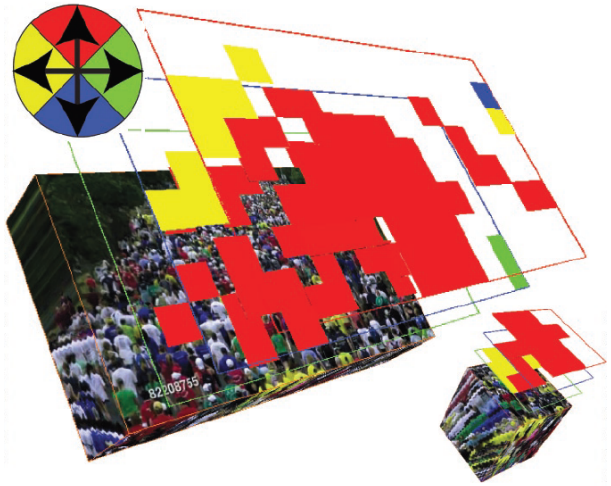


An iterative image registration technique with an application to stereo vision. B. Lucas and T. Kanade. In IJCAI, volume 3, pages 674–679, 1981.

Learning Motion Patterns

Mid-level Representation: Correlated Topic Model

- CTM captures spatial dependencies of different behaviors in the same scene.
- Video(720x480)=> 10 sec clips
  => 36x24 cells(20x20)
- Optical flow is quantized into directions
  => $\{V_0, V_{up}, V_{down}, V_{left}, V_{right}\}$
- Motion word dictionary is constructed
- Behavior is (hidden) topic from which motion words are generated.



A correlated topic model of science. D. Blei and J. Lafferty. AAS, 1(1):17–35, 2007

Learning Motion Patterns

database

Global Matching

similar scenes

test video

Local Matching

similar cells

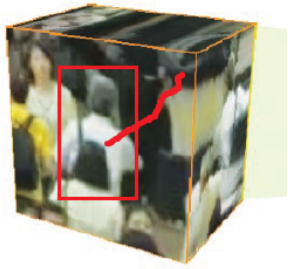
test video

individual to track

Tracking using Motion Patterns

similar cells

motion patterns of similar cells



Global Crowded Scene Matching

- Gist scene descriptor is used to retrieve similar scenes from the database.
- Global matching provides semantically similar scenes.

Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope, Oliva, A., Torralba, A., International Journal of Computer Vision 42(3), 145-175, 2001.

Learning Motion Patterns

database

Global Matching

similar scenes

test video

Local Matching

similar cells

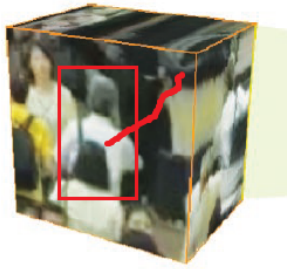
test video

individual to track

Tracking using Motion Patterns

similar cells

motion patterns of similar cells



Local Crowd Patch Matching

- HOG3D is used to retrieve similar patches from the selected scenes.
- HOG3D demonstrates good performance in action recognition.

A Spatio-Temporal Descriptor Based on 3D-Gradients, Kläser, A., Marszałek, M., Schmid, C., British Machine Vision Conference - sep 2008

Learning Motion Patterns

database

Global Matching

similar scenes

test video

Local Matching

similar cells

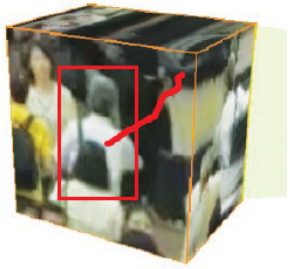
test video

individual to track

Tracking using Motion Patterns

similar cells

motion patterns of similar cells



Tracking using Motion Patterns

Prediction of system

Prediction by Kalman filter

Using:
- Optical Flow
- CTM

Learnt from:
- Test video
- Database of videos

Tracker position for person at location *O*

$$P_O = K + \lambda S$$

Experiments

- Data: Downloaded from video web sites using text queries like “crosswalk”, “political rally”, “festival”, “marathon”.
- 2 types of experiment:
  1. Tracking Typical Crowd Behavior
  2. Tracking Rare and Abrupt Events

Experiments

- Test videos are manually annotated to measure the error in pixels.
  - Blue = Typical crowd behavior
  - Red = Rare events

Experiments

- Error = # of pixels between the positions of tracker and individual in each frame
  - Yellow = ground truth
  - Red = tracking results

1st Experiment

Tracking typical crowd behavior

Ground Truth
Track
No Prior
98671532

Batch-mode tracking

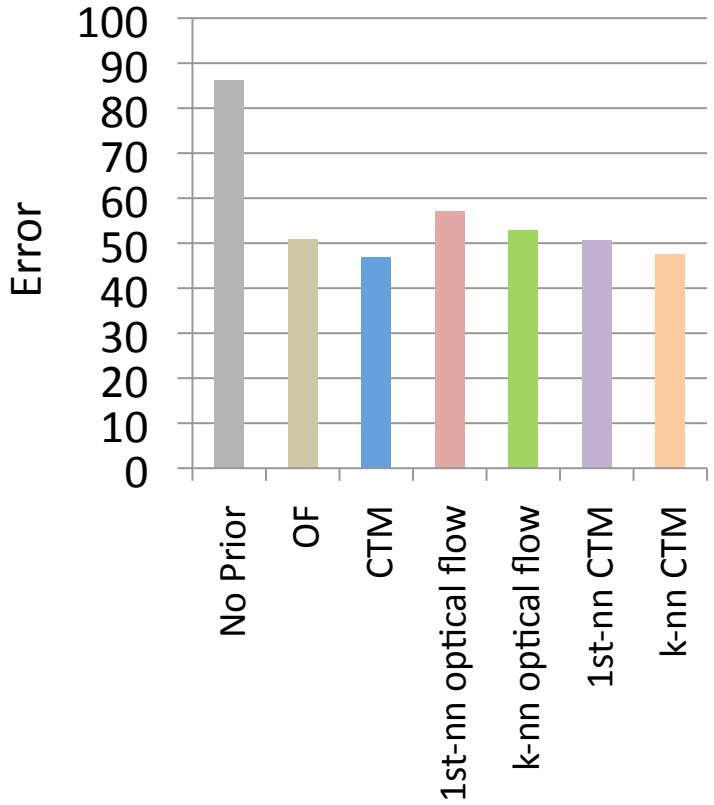
Training and testing video are from the same scene

Proposed data-driven tracking

Motion priors are transferred from the database of crowd videos

Results for tracking typical crowd behavior

		Error
No prior		86.24
Learned on test video	OF	50.93
	CTM	46.93
Learned on database	1st-nn OF	57.06
	3-nn OF	52.76
	1st-nn CTM	50.59
	3-nn CTM	47.47



Error is measured in pixels.

2nd Experiment

Tracking rare events

Ground Truth
Track
No motion prior

Ground Truth
Batch-mode tracking

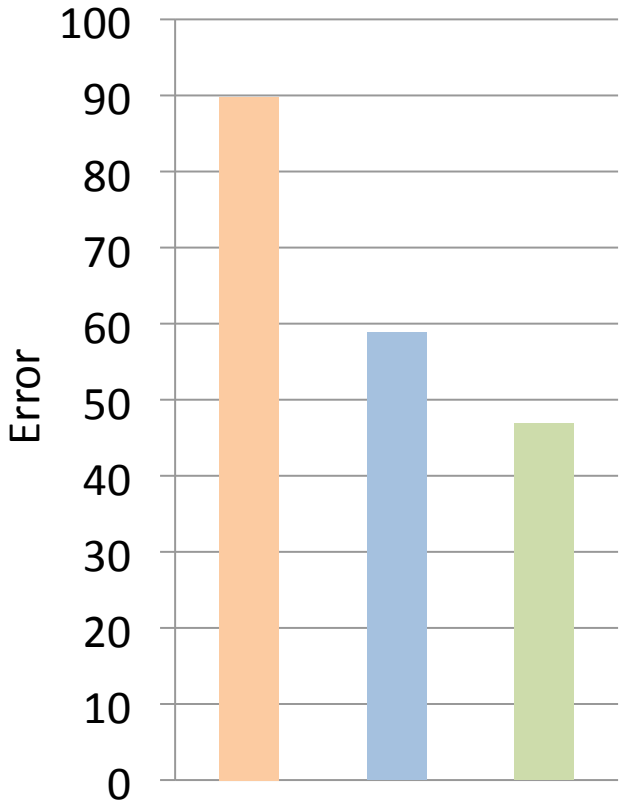
Matched crowd patches
Data-driven tracking

Results for tracking rare events

- Red
  Ground Truth
- Yellow
  Batch mode
- Green
  Data-driven

Results for tracking rare events

		Error
No prior		89.8
Learned on test video	CTM	58.82
Learned on database	k-nn CTM	46.88



Error is measured in pixels.

Discussion

- There is no information about the performance of ‘just optical flow’ method on tracking rare events (2nd experiments).
- There is no information about the speed of the system although it is intended to be an online system.

Resources

- Website: http://www.di.ens.fr/willow/research/datadriven/index.html
- Database will be public soon.