What’s going on?
Discovering Spatio-Temporal Dependencies in Dynamic Scenes

Daniel Kuettel, Michael D. Breitenstein, Luc Van Gool, Vittorio Ferrari

CALVIN group, Computer Vision Laboratory, ETH Zurich, Switzerland
Dynamic Scene Analysis

Setting
- single static camera
- observing dynamic scene
- *unsupervised*

Scene model
for activities with
- spatial dependencies
- *(long)* temporal dependencies
  “rules of the scene”
Applications

- surveillance (unusual event detection)
- video summarization
- information extraction
  - busy hours
  - tram schedule
Finding spatio-temporal rules

**approach overview**

- optical flow
  - fast and robust
  - scene-independent
  - focus on moving things

- find typical patterns of flow (activities)
  - activities can happen in parallel
  - cooccurrence
  - clustering

- find temporal dependencies of activities (rules)
  - markov models

References:

- [Hospedales09]
- [Li08]
- [Wang09]
- [Xiang08]
- [Zong04]
- [Hospedales09]
- [Li08]
- [Wang09]
Finding local rules

two stage approach

• optical flow
  – fast and robust
  – scene-independent
  – focus on moving things

• find typical patterns of flow (activities)  [Hospedales09]  [Li08]  [Wang09]
  – cooccurrence
  – activities can happen in parallel

• find temporal dependencies of activities
  – markov models
Finding global rules

*joint approach*

- optical flow
  - fast and robust
  - scene-independent
  - focus on moving things

- find typical patterns of flow (activities)
  - cooccurrence
  - activities can happen in parallel

- find temporal dependencies of activities
  - markov models
Finding local rules

two stage approach

- optical flow
  - fast and robust
  - scene-independent
  - focus on moving things

- find typical patterns of flow (activities)
  - cooccurrence
  - activities can happen in parallel

- find temporal dependencies of activities
  - markov models
Low level representation

- optical flow for all subsequent frame pairs
- threshold on flow magnitude
  - filter out noise
  - filter out pedestrians
- discretize over location / direction grid
- frame = bag of flow words

**flow** [Zach07]  **thresholded**  **flow words**
Low level representation

• frame = bag of flow words
  – 1/25 seconds too short to contain activities

• clip = 3 seconds of frames = bigger bag of flow
  – defining “at the same time”
  – can contain several activities

flow words  accumulating  accumulated
Finding local rules

two stage approach

- optical flow
  - fast and robust
  - scene-independent
  - focus on moving things

- find typical patterns of flow (activities)
  - cooccurrence
  - activities can happen in parallel

- find temporal dependencies of activities
  - markov models

[Wang09]
Model for Activities

- clip = mixture of activities
  - things can happen in parallel

- activity = pattern of flow
  - flow words of an activity cooccur in the same clip
  - activities reoccur over clips
  - cluster flow words based on cooccurrence in clips
Hierarchical Dirichlet Processes (HDP) [Teh06]

- All activities
- Mixture of activities for clip $t$
- Activity for word $i$ in clip $t$
- $i$th flow word in clip $t$
Activities found by HDP

\textit{all} activities

- continuous video of the scene
- overlaid with detected activities
- \textbf{strong activities} explain more flow than \textbf{weak activities}
Activities found by HDP

clips for activity 1

3 second clips of activity 1
Activities found by HDP

clips for activity 2

3 second clips of activity 2
Finding local rules

two stage approach

- optical flow
  - fast and robust
  - scene-independent
  - focus on moving things

- find typical patterns of flow (activities)
  - cooccurrence
  - activities can happen in parallel

- find temporal dependencies of activities
  - markov models

novel
Finding local rules

Two stage approach

3 example activities
(A) left turn start
(B) left turn end
(C) blocking lane

2 situations
(left turn not blocked
(left turn blocked

Diagram:

- $H$ → $G_0$ → $G_t$ → $\theta_{ti}$ → $x_{ti}$
- $\gamma$ → $G_0$
- $\alpha$ → $G_t$
Finding local rules

two stage approach

3 example activities
• (A) left turn start
• (B) left turn end
• (C) blocking lane

2 situations
- left turn not blocked
- left turn blocked
Finding local rules

**two stage approach**

- **rule = markov model**
  - user provided exemplary transition matrix

- **automatic search matching**
  - assign activities to states
  - estimates actual transition matrix
Finding local rules

*blocking rule*

- rule for 3 activities
  - the bus (b) is blocked by
    - tram left-right (a)
    - tram right-left (b)
Finding local rules

traffic light rule

- 4 activities (lanes) with traffic lights
- exemplary rule?
  - how many states?
  - what transition probabilities?
Finding global rules

joint approach

- optical flow
  - fast and robust
  - scene-independent
  - focus on moving things

- find typical patterns of flow (activities)
  - cooccurrence
  - activities can happen in parallel

- find temporal dependencies of activities
  - markov models
Finding global rules

*joint approach*

- jointly learn
  - activities
  - markov models
- inject temporal dependency into HDP
Finding global rules

*joint approach*

- jointly learn
  - activities
  - markov models
- inject temporal dependency into HDP

original HDP (expanded)
Finding global rules

**joint approach**

- jointly learn
  - activities
  - markov models
- inject temporal dependency into HDP

HDP with temporal dependency
Finding global rules

*joint approach*

- jointly learn
  - activities
  - markov models
- markov model (HMM)

HDP with Markov Model
Finding global rules

*joint approach*

- jointly learn
  - activities
  - markov models
- markov models (HMM)

HDP with Markov Models
Finding global rules

joint approach

• jointly learn
  – activities
  – markov models
• markov models (HMM)
  – infer # of HMM
  – infer # of states

dependent dirichlet process hidden markov model (DDP-HMM)
Joint Model (DDP-HMM) Summary

Input:
- clips with discretized flow
- clips in temporal order

Joint Learning:
- dependent Dirichlet processes
- Markov models
- Gibbs sampling

Output:
- Mixture of Markov models
- For each Markov model:
  - Number of states
  - Sequence of states for scene
  - Pattern flow for each state (activity)
Joint model result
traffic light controlled scene

- 4 states covering all lanes
- simple transition matrix
  - sequence = BD BD BD BD ...
  - sequence = BDCA BDCA ...

[Hospedales09]
Joint model result
traffic light controlled scene

- continuous video
- annotated with states and history
- 3x speed
Joint model result comparison [Hospedales09]

- States do not cover all lanes
- Up + down missing
- Turns incomplete
- No traffic light cycle

- Number of states fixed
- Limited to one Markov model
Joint model result
larger traffic light controlled scene
Joint model result
stochastic scene
Conclusion

• optical flow
  – no detection, independent of object classes

• find activities
  – flow patterns

• local rules involving few activities

• global rules involving all activities
  – automatically find # of markov models and # of states
  – activities and rules are learnt jointly in a single model

• What’s going on!
Thank you for your attention

Code will be available at http://www.vision.ee.ethz.ch/~calvin