Activity Recognition Using Dense Long-Duration Trajectories

Ju Sun, Yadong Mu, Shuicheng Yan, and Loong-Fah Cheong

NUS National University of Singapore

International Conference on Multimedia & Expo, 2010
Special Session on Event and Action Detection in Web Video

July 20, 2010

Outline

1. Representation for Activity Analysis
2. Extraction of Dense Trajectories
3. Description of Dense Trajectories
4. Preliminary Evaluations
5. Summary and Discussion

Outline

1. Representation for Activity Analysis
2. Extraction of Dense Trajectories
3. Description of Dense Trajectories
4. Preliminary Evaluations
5. Summary and Discussion

Activity Analysis – State of the Art

- Almost perfect performance on controlled video (KTH, Weizmann), but very poorly on less controlled (Hollywood – 2, TRECV challenge, general Internet video sequences)
- Learning-based recognition frameworks abound, but not features! [1]

[1] P Turaga, R Chellappa, VS Subrahmanian, O Udrea, "Machine recognition of human activities: A survey". TCSVT, 18(11):1473 – 1488, 2008.

Representation – State of the Art

- We'll focus on representation/features from raw video sequences.
- Most popular features people have been using
  - Laptev's STIP (HOG & HOF, Laptev & Lindeberg 2003)
  - Dollar's dense cuboids (Dollar *et al.* 2005)
  - Volumetric Features (such as Silhouettes, Shapes, *e.g.*, in Veeraraghavan *et al.*)
- And the less popular
  - Long-Duration Trajectories (but was popular in 1990's).

Representation – State of the Art

- We'll focus on representation/features from raw video sequences.
- Most popular features people have been using
  - Laptev's STIP (HOG & HOF, Laptev & Lindeberg 2003)
  - Dollar's dense cuboids (Dollar *et al.* 2005)
  - Volumetric Features (such as Silhouettes, Shapes, *e.g.*, in Veeraraghavan *et al.*)
- And the less popular
  - Long-Duration Trajectories (but was popular in 1990's).

Representation – State of the Art

- We'll focus on representation/features from raw video sequences.

- Most popular features people have been using
  - Laptev's STIP (HOG & HOF, Laptev & Lindeberg 2003)
  - Dollar's dense cuboids (Dollar *et al.* 2005)
  - Volumetric Features (such as Silhouettes, Shapes, *e.g.*, in Veeraraghavan *et al.*)

- And the less popular
  - Long-Duration Trajectories (but was popular in 1990's).

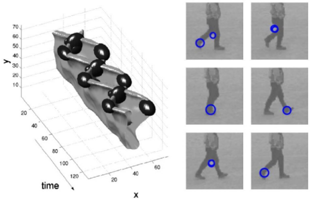
Representation – State of the Art

- We'll focus on representation/features from raw video sequences.
- Most popular features people have been using
  - Laptev's STIP (HOG & HOF, Laptev & Lindeberg 2003)
  - Dollar's dense cuboids (Dollar *et al.* 2005)
  - Volumetric Features (such as Silhouettes, Shapes, *e.g.*, in Veeraraghavan *et al.*)
- And the less popular
  - Long-Duration Trajectories (but was popular in 1990's).

Representation – State of the Art

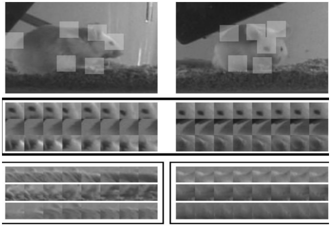
- We'll focus on representation/features from raw video sequences.
- Most popular features people have been using
  - Laptev's STIP (HOG & HOF, Laptev & Lindeberg 2003)
  - Dollar's dense cuboids (Dollar *et al.* 2005)
  - Volumetric Features (such as Silhouettes, Shapes, *e.g.*, in Veeraraghavan *et al.*)
- And the less popular
  - Long-Duration Trajectories (but was popular in 1990's).

Laptev's (sparse) STIP & Dollar's Dense Cuboids



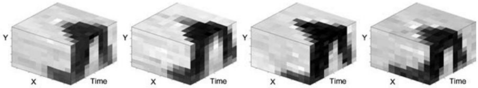
(a) (b)

Sparse Corner Points



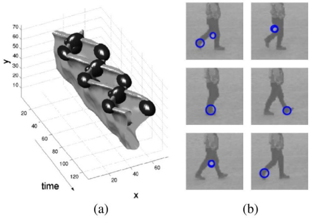
Dense Strong-Response Regions

$\Downarrow$

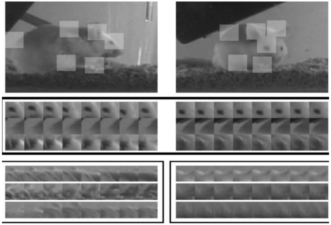


Description (Statistics of optic flows, spatial gradients, pixel values, etc)

Laptev's (sparse) STIP & Dollar's Dense Cuboids

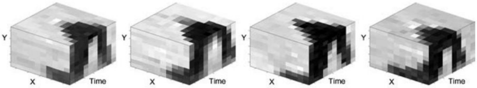


Sparse Corner Points



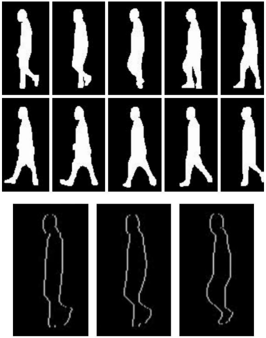
Dense Strong-Response Regions

⇓



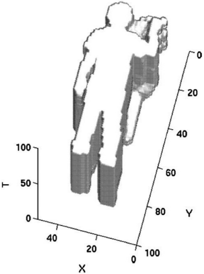
Description (Statistics of optic flows, spatial gradients, pixel values, etc)

Volumetric Features



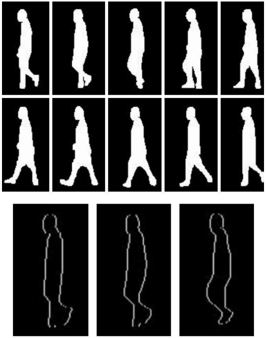
Motion Shapes/Silhouettes

Temporal Template: Motion Energy Image/Motion History Image



Spatio-Temporal Template: motion volumes

Volumetric Features



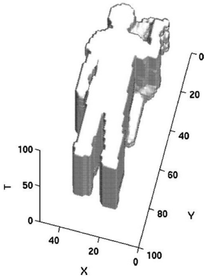
Motion Shapes/Silhouettes

Temporal Template:

Motion Energy

Image/Motion History

Image

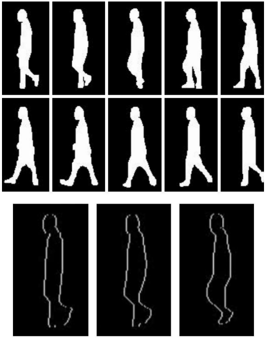


Spatio-Temporal

Template: motion

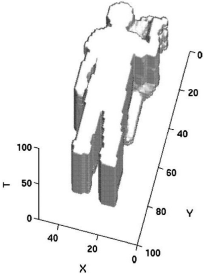
volumes

Volumetric Features



Motion Shapes/Silhouettes

Temporal Template: Motion Energy Image/Motion History Image



Spatio-Temporal Template: motion volumes

Long-Duration Motion Trajectories as Features

Distinct in video are the motion features, as compared to a single frame. Hence the interest to extract and encode long-duration *Motion Trajectories*.

- Lots of previous work on visual tracking, e.g., KLT
- Visual tracking is normally sparse and unreliable

Target: efficient long-duration (moderately) dense motion estimation to improve upon simple tracking, and application in action/activity analysis

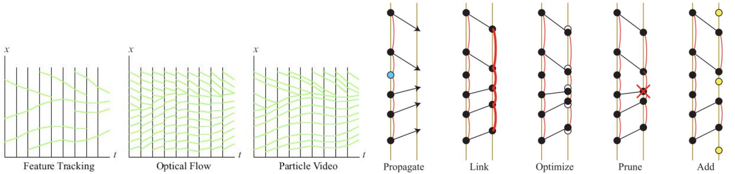
Long-Duration Motion Trajectories as Features

Distinct in video are the motion features, as compared to a single frame. Hence the interest to extract and encode long-duration *Motion Trajectories*.

- Lots of previous work on visual tracking, e.g., KLT
- Visual tracking is normally sparse and unreliable

Target: efficient long-duration (moderately) dense motion estimation to improve upon simple tracking, and application in action/activity analysis

Particle Video



- Particle video algorithm[2] depends upon (expensive) variational optic flow calculations
- Optic flow estimation is less reliable than keypoint tracking.

[2] P Sand and S Teller, "Particle video: Long-range motion estimation using point trajectories". IJCV, 80(1):72 – 91, 2008.

Outline

1. Representation for Activity Analysis
2. Extraction of Dense Trajectories
3. Description of Dense Trajectories
4. Preliminary Evaluations
5. Summary and Discussion

Initial tracking

$\Downarrow$

Triangulation

$\Downarrow$

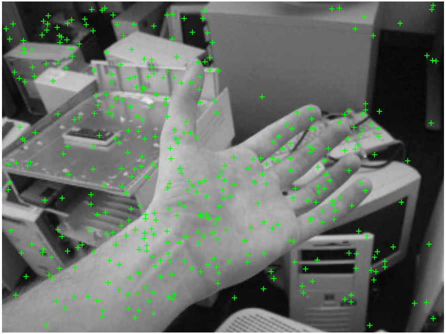
Trajectory Optimization

$\Downarrow$

Tracker Maintenance

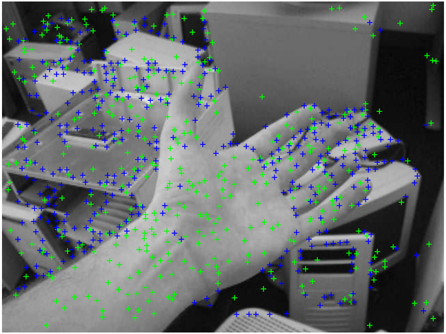
Initial Tracking

- SIFT Trackers
  - Blobs and Corners
- KLT Trackers
  - Corners
- Random Trackers
  - Less Structured Regions
  - Connectivity & Density Adaptiveness



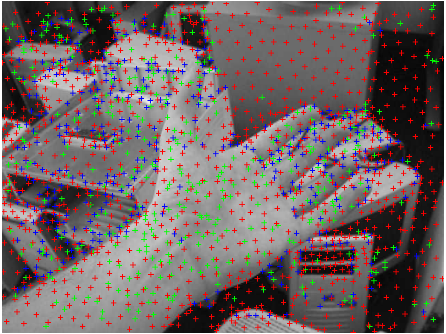
Initial Tracking

- SIFT Trackers
  - Blobs and Corners
- KLT Trackers
  - Corners
- Random Trackers
  - Less Structured Regions
  - Connectivity & Density Adaptiveness



Initial Tracking

- SIFT Trackers
  - Blobs and Corners
- KLT Trackers
  - Corners
- Random Trackers
  - Less Structured Regions
  - Connectivity & Density Adaptiveness



Random Trackers

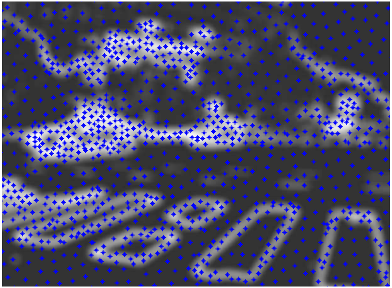
- Structure-Aware Random Tracker (local gradient variance)
- Blue-Noise Importance Sampling (avoiding regular patterns)

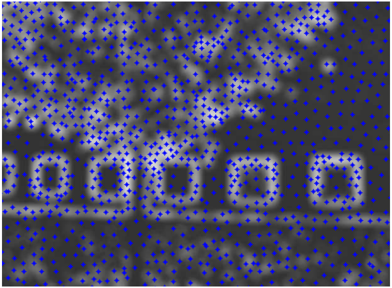
Random Trackers

- Structure-Aware Random Tracker (local gradient variance)
- Blue-Noise Importance Sampling (avoiding regular patterns)

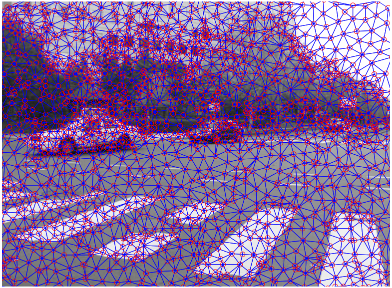
Random Trackers

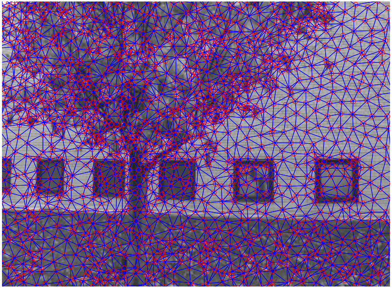
- Structure-Aware Random Tracker (local gradient variance)
- Blue-Noise Importance Sampling (avoiding regular patterns)





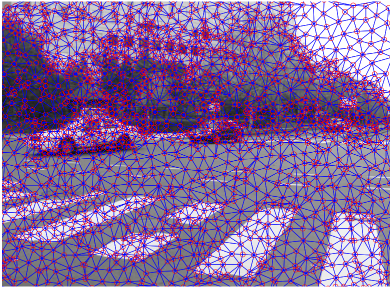
Delaunay Triangulation





Trajectory Optimization

- Goal: to determine and refine the motion vectors $\mathbf{v}$ (from frame t to $t+1$) for each tracker
- The motion vectors for random trackers are initialized with the respective optic flows using the Lucas-Kanade method.



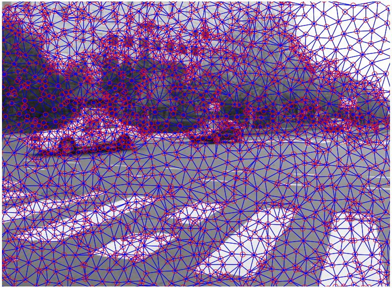
- Quadratic Programming (QP)

$$\min. \sum_{(\mathbf{v}_i,\mathbf{v}_j)\in\mathcal{D}} \omega_{ij}\|\mathbf{v}_i-\mathbf{v}_j\|^2$$

$$\text{subj.} \quad \|\mathbf{v}_i-\widetilde{\mathbf{v}}_i\|_2 \le \eta_i\|\widetilde{\mathbf{v}}_i\|_2, \forall i.$$

Trajectory Optimization

- Goal: to determine and refine the motion vectors $\mathbf{v}$ (from frame t to $t+1$) for each tracker
- The motion vectors for random trackers are initialized with the respective optic flows using the Lucas-Kanade method.



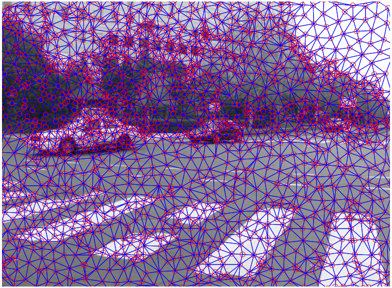
- Quadratic Programming (QP)

$$\begin{array}{ll} \min. & \sum_{(\mathbf{v}_i,\mathbf{v}_j)\in\mathcal{D}} \omega_{ij}\|\mathbf{v}_i - \mathbf{v}_j\|^2 \\ \text{subj.} & \|\mathbf{v}_i - \widetilde{\mathbf{v}}_i\|_2 \leq \eta_i\|\widetilde{\mathbf{v}}_i\|_2, \forall i. \end{array}$$

Trajectory Optimization

$$\min. \sum_{(\mathbf{v}_i,\mathbf{v}_j)\in\mathcal{D}} \omega_{ij}\|\mathbf{v}_i-\mathbf{v}_j\|^2$$

$$\text{subj.} \quad \|\mathbf{v}_i-\widetilde{\mathbf{v}}_i\|_2 \le \eta_i\|\widetilde{\mathbf{v}}_i\|_2, \forall i.$$



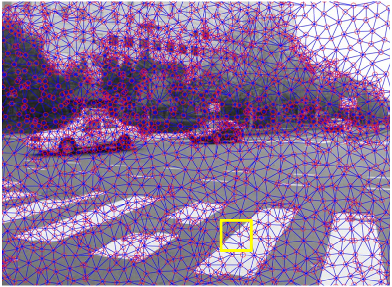
- $\mathcal{D}$ denotes the Delaunay system
- ω_{ij} is the adaptive weight

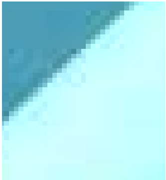
Trajectory Optimization

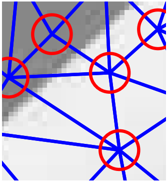
$$\min. \quad \sum_{(\mathbf{v}_i,\mathbf{v}_j)\in\mathcal{D}} \omega_{ij}\|\mathbf{v}_i - \mathbf{v}_j\|^2$$

$$\text{subj.} \quad \|\mathbf{v}_i - \widetilde{\mathbf{v}}_i\|_2 \leq \eta_i\|\widetilde{\mathbf{v}}_i\|_2, \forall i.$$

- $\mathcal{D}$ denotes the Delaunay system
- ω_{ij} is the adaptive weight





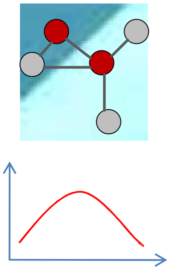


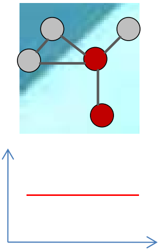
Trajectory Optimization

$$\min. \sum_{(\mathbf{v}_i,\mathbf{v}_j)\in\mathcal{D}} \omega_{ij}\|\mathbf{v}_i - \mathbf{v}_j\|^2$$

$$\text{subj.} \quad \|\mathbf{v}_i - \widetilde{\mathbf{v}}_i\|_2 \leq \eta_i\|\widetilde{\mathbf{v}}_i\|_2, \forall i.$$

- $\mathcal{D}$ denotes the Delaunay system
- ω_{ij} is the adaptive weight
- η_i is the relaxation coefficients





Tracker Addition and Removal

- Tracker Addition
  - Emerging Gaps (e.g., due to occlusions and motions)
  - Lost Trackers (due to appearance changes)
  - Shot Boundaries
- Tracker Removal
  - Mismatches
  - Image Borders
  - Too-Close Trackers
- Straightforward for SIFT and KLT trackers

Tracker Addition and Removal

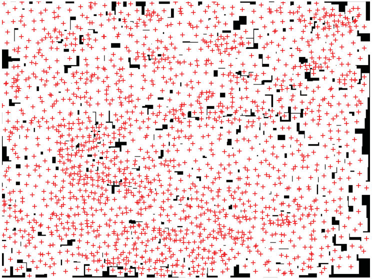
- Tracker Addition
  - Emerging Gaps (e.g., due to occlusions and motions)
  - Lost Trackers (due to appearance changes)
  - Shot Boundaries
- Tracker Removal
  - Mismatches
  - Image Borders
  - Too-Close Trackers
- Straightforward for SIFT and KLT trackers

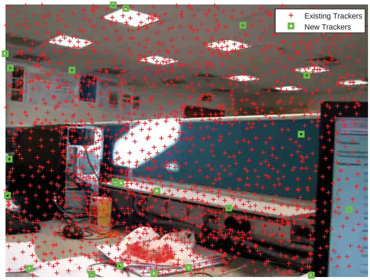
Tracker Addition and Removal

- Tracker Addition
  - Emerging Gaps (e.g., due to occlusions and motions)
  - Lost Trackers (due to appearance changes)
  - Shot Boundaries
- Tracker Removal
  - Mismatches
  - Image Borders
  - Too-Close Trackers
- Straightforward for SIFT and KLT trackers

Tracker Addition for Random Trackers

- Calculate the availability map and density map in F_{i+1}
- Sample candidate random trackers in F_{i+1}
- Select only trackers within the available region

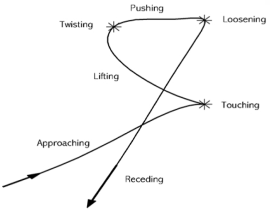




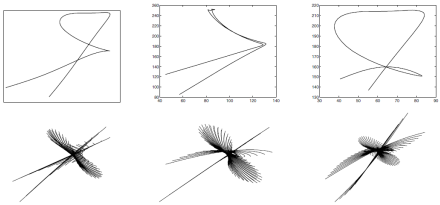
Outline

1. Representation for Activity Analysis
2. Extraction of Dense Trajectories
3. Description of Dense Trajectories
4. Preliminary Evaluations
5. Summary and Discussion

Description of Trajectories: Prior Approaches

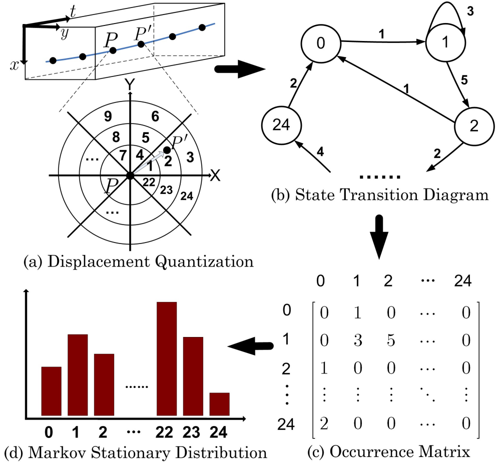


Dynamic Instants



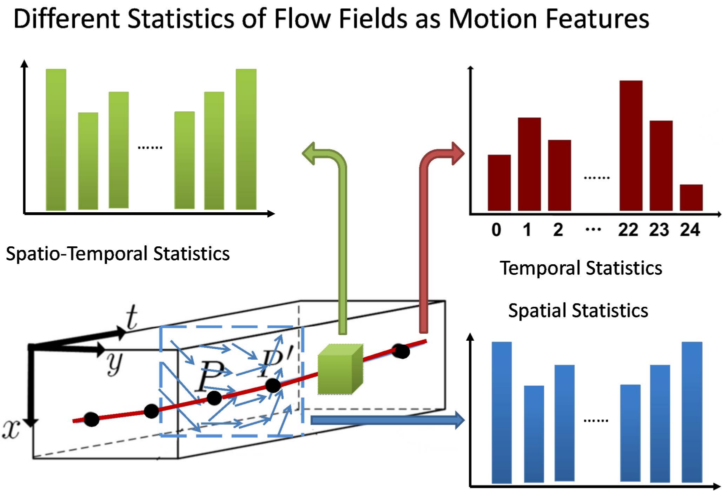
Star Diagrams

Description of Trajectories: Our Prior Approach



Stationary distribution of Quantized Markov Chains

Temporal Quantization and Averaging



Outline

1. Representation for Activity Analysis
2. Extraction of Dense Trajectories
3. Description of Dense Trajectories
4. Preliminary Evaluations
5. Summary and Discussion

Discrimination on KTH

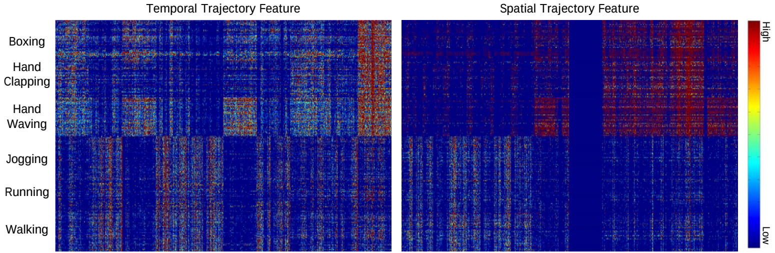


Figure: Visualization of codewords for all KTH video clips. Each row corresponds to a video clip, and those related to the same actions are stacked adjacently.

Performance Figures

- Holistic Bag-of-Features quantization and statistics
- Multiclass SVMs for classification with intersectional kernels
- Feature Combination by simple kernel averaging

Table: Performance of action recognition over all valid training/testing configurations. The first three methods are based on our proposed features. (accuracy±deviation)

Method	Average Accuracy (%)
Spatial Feature	81.3 ± 9.51
Temporal Feature	83.1 ± 7.22
Spatial&Temporal Feature	86.8 ± 10.7
Method by STIP	71.7 ± 16.7

Outline

1. Representation for Activity Analysis
2. Extraction of Dense Trajectories
3. Description of Dense Trajectories
4. Preliminary Evaluations
5. Summary and Discussion

Summary

- A proposal for extracting dense trajectories
- A simplistic representation
- Experimental validation

Challenges in activity analysis

- Representation
- Information Fusion
- ROI Search
- Large-scale problem solving

Thank you!