Activity Recognition
Using Dense Long-Duration Trajectories

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

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
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

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).
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).
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).
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).
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

Sparse Corner Points

Dense Strong-Response Regions

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 algorithm\(^2\) depends upon (expensive) variational optic flow calculations.

Optic flow estimation is less reliable than keypoint tracking.

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

⇓

Triangulation

⇓

Trajectory Optimization

⇓

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 $v$ (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)**

$$\min \sum_{(v_i, v_j) \in D} \omega_{ij} \|v_i - v_j\|^2$$

subj. $\|v_i - \tilde{v}_i\|_2 \leq \eta_i \|\tilde{v}_i\|_2, \forall i.$
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)

\[
\begin{align*}
\min & \quad \sum_{(\mathbf{v}_i, \mathbf{v}_j) \in D} \omega_{ij} \| \mathbf{v}_i - \mathbf{v}_j \|^2 \\
\text{subj.} & \quad \| \mathbf{v}_i - \tilde{\mathbf{v}}_i \|_2 \leq \eta_i \| \tilde{\mathbf{v}}_i \|_2, \forall i.
\end{align*}
\]
Trajectory Optimization

\[ \min \sum_{(v_i, v_j) \in D} \omega_{ij} \|v_i - v_j\|^2 \]

\[ \text{subj. } \|v_i - \tilde{v}_i\|_2 \leq \eta_i \|\tilde{v}_i\|_2, \forall i. \]

- \( D \) denotes the Delaunay system
- \( \omega_{ij} \) is the adaptive weight
Trajectory Optimization

\[
\begin{align*}
\min \quad & \sum_{(v_i, v_j) \in \mathcal{D}} \omega_{ij} \|v_i - v_j\|^2 \\
\text{subj.} \quad & \|v_i - \tilde{v}_i\|_2 \leq \eta_i \|\tilde{v}_i\|_2, \forall i.
\end{align*}
\]

- \(\mathcal{D}\) denotes the Delaunay system
- \(\omega_{ij}\) is the adaptive weight
Trajectory Optimization

\[
\begin{align*}
\text{min.} & \quad \sum_{(v_i, v_j) \in D} \omega_{ij} \|v_i - v_j\|^2 \\
\text{subj.} & \quad \|v_i - \tilde{v}_i\|_2 \leq \eta_i \|\tilde{v}_i\|_2, \forall i.
\end{align*}
\]

- \(D\) denotes the Delaunay system
- \(\omega_{ij}\) is the adaptive weight
- \(\eta_i\) 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
Description of Trajectories: Prior Approaches

Dynamic Instants

Star Diagrams

Sun, Mu, Yan, and Cheong

AR Using Dense Long-Duration Trajectories
Description of Trajectories: Our Prior Approach

(a) Displacement Quantization

(b) State Transition Diagram

(c) Occurrence Matrix

(d) Markov Stationary Distribution

Stationary distribution of Quantized Markov Chains
Temporal Quantization and Averaging

Different Statistics of Flow Fields as Motion Features

Spatio-Temporal Statistics

Temporal Statistics

Spatial Statistics
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)

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial Feature</strong></td>
<td>81.3 ± 9.51</td>
</tr>
<tr>
<td><strong>Temporal Feature</strong></td>
<td>83.1 ± 7.22</td>
</tr>
<tr>
<td><strong>Spatial&amp;Temporal Feature</strong></td>
<td>86.8 ± 10.7</td>
</tr>
<tr>
<td><strong>Method by STIP</strong></td>
<td>71.7 ± 16.7</td>
</tr>
</tbody>
</table>
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!