Extracting human behaviour from video streams

Ioan Marius BILASCO
FOX-MIIRE, LIFL
Université Lille 1 – CNRS

{marius.bilasco, jean.martinet, chabane.djeraba}@lifl.fr
Introduction

- Profusion of video streams available
  - In home environments: Webcams (2D or 3D)
  - In private or public areas: Surveillance systems

- Video streams enclose human actions

- Extract knowledge about
  - Implicit user feedback (visual feedback, emotions ...)
  - Crowd analysis (dynamics, abnormal situation ...)

- Use regular camera and satisfy real-time constraints
Overview

- Introduction

- Single person behaviour
  - Gaze-tracking [A. Lablack]
  - Emotions and drowsiness [T. Danisman]

- Crowd behaviour [Y. Benabbas]
  - Emergency situations
  - Crowd dynamics

- Conclusions
(Single) Person behaviour

- **Gaze-tracking**
  - What the user is looking at?
  - What are the elements seen by a person in a scene?

- **Emotion recognition**
  - How a person perceives a digital content or a product?
  - What’s the current state of the user?

- **Drowsiness detection**
  - Is the person about to fall asleep?
Person behaviour – Gaze tracking

- STAR: Intrusive solutions (Tobii, Facelab, etc.)
- Non-intrusive and affordable solutions
  - Various distances, multi-person

- Head pose estimation + eye locations
  - joint work with U. Amsterdam
Gaze tracking

- Head pose estimation
  - KNN and Gabor wavelets
- Cylindrical head tracker
  - Three translation and three rotation parameters
- Adaptation of the visual field to the head pose
- Gaze point correction using eye center location
  - Displacement of the projected gaze point
Gaze tracking - demo
Developing a fully automatic, real-time emotion recognition system using seven facial feature points

- Eyebrows (2)
- Iris (2)
- Nose (1)
- Corners of mouth (2).

Automatic facial point detection using low-cost camera
We propose a combination of different methods:

- Face Detection
  - OpenCV, Viola-Jones
- Eye detection
  - Neural Network, Rowley
- Eyebrow & Nose Localization
  - Adaptive thresholding
- Mouth Corner Detection
  - Neural Network
- Emotion Detection
  - Decision rules
Person behaviour - Drowsiness

- Developing a proactive, real-time Drowsy Driver Detection System (DDDS) using eye blink patterns with non-intrusive low cost camera.
  - No calibration/no initialization for the system
  - Low computational time
  - Affordable cameras

- Horizontal symmetry feature of the eyes
Drowsy Driver Detection System

- Face & Eye Detection
- Eye region extraction
- Eye blink detection
  - Horizontal symmetry
  - Threshold
- Drowsiness Detection

<table>
<thead>
<tr>
<th>Drowsiness level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awake</td>
<td>Blink durations &lt; $T_{Drowsy}$.</td>
</tr>
<tr>
<td>Drowsy</td>
<td>Blink durations &gt; $T_{Drowsy}$ and Blink durations &lt; $T_{Sleeping}$</td>
</tr>
<tr>
<td>Sleeping</td>
<td>Blink durations &gt; $T_{Sleeping}$</td>
</tr>
</tbody>
</table>
Drowsiness - Experimental Results

- We used JZU Eye blink database (ICCV 2007)
  - Total of 80 videos (40 with glasses) at 30fps
  - Static and moving faces; dynamic backgrounds with different lighting conditions
  - We manually annotated each frame to create ground truth data by observing 11,750 frames.

Accuracy = 94.8%
Precision = 90.7%
Recall = 71.4%
False Positive Rate = 1%
« Crowd » behaviour

- **Scene description**:  
  - Extraction of most/less frequented areas.  
  - Extraction of trajectory classes.

- **Unusual event detection**:  
  - Detection of abnormal crowd situations.  
  - Abnormal situation logs (Indexing/Retrieval).
Methodology

1. Extraction of "Points of Interest"
2. Tracking of "Points of Interest"
3. Static and noise points removal
4. Grouping vectors by blocks
5. Block clustering
6. Group tracking
7. Direction model
8. Event detection
9. Magnitude model
10. Event report
Point of Interest Extraction and Tracking

- Extraction of KLT points of interest.
- Calculation of Optical flow vectors (Lukas-Kanade Optical Flow.)
Grouping vectors by blocks

The scene is divided into blocks.
Cluster the neighboring blocks that have a similar direction into a group.
Group Clustering and Tracking

- Cluster the neighboring blocks that have a similar direction into a group.
- Define an main orientation for each group.
- Each group is tracked over the next frames.
Directional statistics

- Mean angle at frame $f$:
  \[ \overline{X}_{0,f} = \arctan \left( \frac{\sum_{i=1}^{n_f} \sin(X_{i,f})}{\sum_{i=1}^{n_f} \cos(X_{i,f})} \right) \]
  \[ X_{i,f} : \text{orientation of the group } i \text{ at frame } f \]
  \[ n_f : \text{number of groups at frame } f \]

- Circular variance at frame $f$:
  \[ S_{0,f} = 1 - \frac{1}{n_f} \sum_{i=1}^{n_f} \cos \left( X_{i,f} - \overline{X}_{0,f} \right) \]

- For $S_{0,f} = 0$, the angles are identical
- For $S_{0,f} = 1$, the angles are perfectly opposite
Event detection

- If the circular variance exceeds a threshold, then the groups in the frame have different directions.
  - Merging:
    - The groups are near.
    - They head for the same destination.
  - Local dispersion:
    - The groups are oriented towards opposite directions.
    - The groups are near, short period of time
  - Splitting:
    - Same as local dispersion but the separation duration is longer
  - Evacuation:
    - Local dispersion + running.
Experiments

Split / local dispersion events
Experiments: abnormal situations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trajectory matching probability</td>
<td>0.0145290913 / 3 / 39</td>
</tr>
<tr>
<td>Motion area ratio</td>
<td>0.0825</td>
</tr>
<tr>
<td>Direction dispersion</td>
<td>0.375969</td>
</tr>
<tr>
<td>Magnitude Variance</td>
<td>0.0641966</td>
</tr>
<tr>
<td>Magnitude Mean</td>
<td>0.0658039</td>
</tr>
<tr>
<td># Histogram peaks</td>
<td>9</td>
</tr>
<tr>
<td>Time</td>
<td>0:0:0 / Metric: 0.00244502</td>
</tr>
</tbody>
</table>
Conclusions

- Extracting human behaviour from video streams
  - Person-centered
    - Gaze estimation
    - Emotion and drowsiness evaluation
  - Crowd-centered methods
    - Group detection and tracking
    - Crowd dynamics and abnormal event analysis

- We are also working on:
  - Multimedia indexing and retrieval
  - Multimedia semantics