Distributed Vision Processing in Smart Camera Networks

Andrea Cavallaro

Part 1/4
Filtering and annotation

Queen Mary, University of London

andrea.cavallaro ((at)) elec.qmul.ac.uk
http://www.elec.qmul.ac.uk/staffinfo/andrea

Acknowledgements

Emilio Maggio
Murtaza Taj
Gabin Kayumbi
Nadeem Anjum
Matteo Bregonzio
Stefan Karlsson
Fahad Daniyal
Huiyu Zhou
Example: why multiple cameras?

APIDIS dataset

Example: why multiple cameras?

CREDS Dataset courtesy of RATP
Smart cameras

- **Smart cameras**
  - suited for applications requiring several sensors → *distributed vision* (inspection/surveillance/video production)
  - sample the environment independently/collaboratively and can be asynchronous/synchronous

- **Applications**
  - quality assurance (defects, missing parts...)
  - automated surveillance (intruders, fire/smoke)
  - biometrics/access control (face, fingerprint, iris)
  - non-contact measurements
  - code reading, part sorting, part identification
  - robot guidance/automated picking

Framework

- input video
- pre-filtering
- change detection
- post-processing
- event detection
- 3D analysis
- tracking and classification
- a priori information and info from other cameras
Framework

input video → pre-filtering → object detection → post-processing

symbols → event detection → 3D analysis → tracking and classification

a priori information and info from other cameras
Information filtering

- **Objective**
  - To extract relevant information from a scene (images)
    - without the need for an external processing unit
    - with the help of other cameras (sensors)
  - The format of the output format can be different from the signal that was captured by the sensor (*transmoding*).
  - Processing in the camera is nothing new!
    - coding, balancing, etc.
    - smart cameras were sold in the 80s! (industrial inspection)

- **Object detection** (i.e., *how to find relevant information*)
  - motion classification
  - change detection
  - object classification

- **Object tracking** (i.e., *how to propagate relevant information*)
  - over time (in the same camera)
  - across cameras
  - despite occlusions
  - despite multiple simultaneous objects
  - despite local and global changes in illumination

- **Object and scene description** (*compact and unambiguous*)
  - communication
  - storage
Visual content description

- **Algorithms**
  - statistical signal processing
  - pattern recognition
  - foreground/background segmentation
  - feature point extraction
  - learning algorithms
  - target tracking
  - change detection
  - color segmentation
  - temporal filtering
  - collaborative filtering
  - ..
From visual information to knowledge

- **Objective** → to extract the **main content message**
  - e.g., automatic video object segmentation
  - classification of the pixels in the video into two classes
    - *foreground* + *background pixels*
  - decompose each frame into sets of mutually exclusive and jointly exhaustive segments

- **Use a priori information**
  - Application dependent
  - Human abstraction

- **Alternative**
  - probabilistic description of the observations, supported by time and/or multi-camera integration

A priori information

- **Examples**
  - Template matching (shape)
  - Extraction of road sings and text (geometry)
  - Face detection (colour)
  - Moving object segmentation (change)
    - Sport broadcasting
    - Video surveillance
Transmoding

- « MPEG-7 » camera
  - to describe a scene in terms of objects and of their properties
  - uses video analysis (incl. tracking)
  - extracts and transmits only the relevant information
MPEG-7 camera applications

- Virtual display
  - Virtual objects (e.g., blobs) follow the movement of the persons

Applications
- Privacy: only the behavior of the persons are transmitted
- Extract various statistics without revealing identity of people
- Checking intentions in surveillance

![Real scene](image1.png) ![Moving objects](image2.png) ![Virtual objects](image3.png)
Privacy-preserving smart cameras

Input video → rendering covering identities

Privacy-preserving smart cameras

- Privacy cam: coverage monitoring
- Personal data: local management
- Behavioural data: high-level reasoning
- Surveillance operator: automatic/interactive decision
- Separate authority: alert

High-level reasoning & multi-camera management
Annotation formats

• For evaluation
  • ViPER: Video Performance Evaluation Resource
  • De facto standard for generating ground truth (XML format)
  • Evaluation metrics available
  • http://viper-toolkit.sourceforge.net/

• For description/search
  • MPEG Video surveillance MAF
    (surveillance metadata application format)
    • Interoperability
    • Digital rights management support
  • ISO/IEC CD 23000-10 Video Surveillance MAF
Frontal faces

Right profile faces

Left profile faces

Framework

input video → pre-filtering → object detection → post-processing

symbols

event detection → 3D analysis → tracking and classification

a priori information and info from other cameras

a priori information and info from other cameras

a priori information and info from other cameras

a priori information and info from other cameras

a priori information and info from other cameras

a priori information and info from other cameras
References

• Semantic video analysis for adaptive content delivery and automatic description
  A. Cavallaro, O. Steiger, T. Ebrahimi

• Privacy in video surveillance
  A. Cavallaro

• Shadow-aware object-based video processing
  A. Cavallaro, E. Salvador, T. Ebrahimi

• Cast shadow segmentation using invariant colour features
  E. Salvador, A. Cavallaro, T. Ebrahimi
  Computer Vision and Image Understanding, vol. 95, n.2, August 2004, pp. 238-259

• Perceptual semantics
  A. Cavallaro, S. Winkler