Knowledge Discovery and Representation for Large Scale Exploitation of Forensic Data

Cesar Pantoja
c.pantoja@qmul.ac.uk

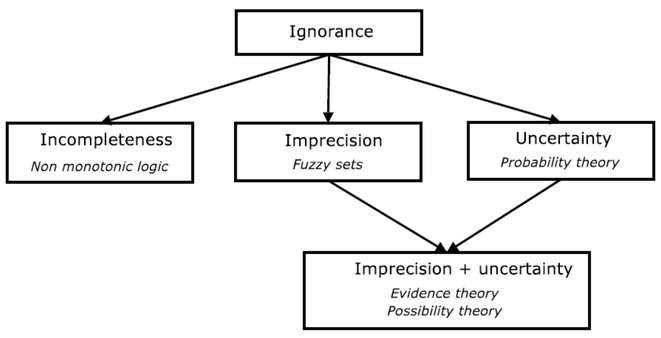
1. Introduction

- Large amount of surveillance/forensic information
- Manual browsing leads to a potential risk of information overload on the operators, missing key evidence
- Real world information is incomplete/imprecise/uncertain
- How can we help security forces to detect automatically information of interest?

2. What is “ignorance”?

- Can be defined of three types[1]:

- **Incompleteness:** Not all the data is known
- **Imprecision:** Data is available with an imprecise measurement
- **Uncertainty:** The data may be wrong



Varieties of ignorance[1]

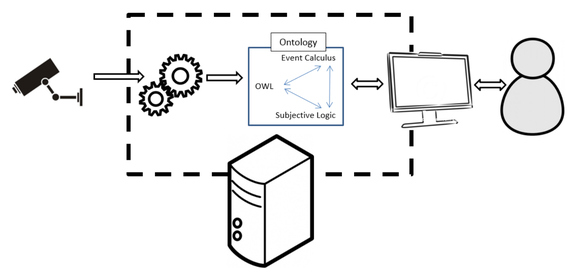
Reasoning under ignorance

- *Incompleteness* is handled using non monotonic logic (like default reasoning[2], abductive reasoning [3] and backtracking).
- *Imprecision* is dealt with using fuzzy sets
- *Uncertainty* using probability theory
- Evidence Theory and Possibility Theory have been proposed to reason with imprecision and uncertainty

3. Proposed Solution

Reasoning framework for surveillance and forensic scenarios, which copes with incomplete, uncertain and imprecise data.

- **Knowledge reasoning architecture:** A robust reasoning engine which can cope with this conditions, must be developed.
- **Knowledge representation framework:** This includes building the metadata (taxonomy) as well as the semantic framework.
- **Visual analysis system:** Visual analysis methods, which will populate the semantic framework, will be developed.
- **Visualization engine:** A visual interface will be developed, which will allow the user to input information, define rules and queries, and get results.



Conceptual architecture of the proposed framework

4. An Example

The complexity of the queries can be progressively increased. For example when trying to find the suspect of an event[4]:

- Atomic level of granularity for semantic queries; Compositional queries that can be seen as conjunction of atomic metadata (e.g. person, vehicle, etc).
  - *Q1 ‘scenes in which a person and a red vehicle are shown’*
  - *Q2 ‘scenes in which a person shown on a given photo is in’*

- Medium level of granularity for semantic queries; Compositional queries having relational semantics that may require more contextual information. (e.g. group of, passing by, carrying, talking to, witness of, etc).
  - *Q3 ‘group of four people passing by a telephone booth’*
  - *Q4 ‘scenes in which a person is carrying an oil box around the place of the reported incident’*
  - *Q5 ‘scenes one of selected person, among retrieved after Q4, is talking to somebody’*
  - *Q6 ‘probable witnesses of a selected person after Q4’*

- Large level of granularity for semantic queries; Queries that iterative and active collection of contextual information is required.
  - *Q7 ‘who is the suspect of the event?’*
  - *Q8 ‘who is the most probable witness of the suspect of the event?’*

Actual image of the London Bombings Suspects released by the Metropolitan Police[4]

5. References

[1] Nicolas Burrus and David Lesagne. Theory of Evidence (DRAFT). 2003.
[2] Seunghan Han, Bonjung Koo, and W. Stechele. Subjective logic based approach to modeling default reasoning for visual surveillance. In Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on, pages 112-119, 2010.
[3] Juan Gomez-Romero, Miguel A. Patricio, Jesus Garcia, and Jose M. Molina. Ontology-based context representation and reasoning for object tracking and scene interpretation in video. Expert Systems with Applications, 38(6):7494-7510, 2011.
[4] Seunghan Han, A. Hutter, and W. Stechele. Toward contextual forensic retrieval for visual surveillance: Challenges and an architectural approach. In Image Analysis for Multimedia Interactive Services (WIAMIS), 2009 10th International Workshop on, pages 201-204, 2009.