An Ontology Framework for Automated Visual Surveillance System

Faranak Sobhani, Nur Farhan Kahar, Qianni Zhang
Multimedia Vision Research Group
School of Electronic Engineering and Computer Science
Queen Mary University of London
Mile End Road, London E1 4NS
Email: f.sobhani@qmul.ac.uk, n.f.kahar@qmul.ac.uk, qianni.zhang@qmul.ac.uk

***Abstract*—This paper presents analysis and development of a forensic domain ontology to support an automated visual surveillance system. The proposed domain ontology is built on a specific use case based on the severe riots that swept across major UK cities with devastating effects during the summer 2011. The proposed ontology aims at facilitating the description of activities, entities, relationships, resources and consequences of the event. The study exploits 3.07 TB data provided by the Londons Metropolitan Police (Scotland Yard) as a part of European LASIE project[1]. The data has been analyzed and used to guarantee adherence to a real-world application scenario. A 'top-down development' approach to the ontology design has been taken. The ontology is also used to demonstrate how high level reasoning can be incorporated into an automatop-ted forensic system. Thus, the designed ontology is also the base for future development of knowledge inference as response to domain specific queries.**

Keywords: forensic computing, ontology, visual surveillance

I. Introduction

Nowadays, automated visual surveillance is regarded as a critical enabling technology to support security, crime detection and prevention. Clearly the world is becoming increasingly digitized and digital recording of day-to-day events is becoming pervasive. Accordingly, an important objective of video surveillance is the automatic understanding of CCTV footage. As a consequence, the requirements for automated visual surveillance focusing on the prediction of unexpected and dangerous events to enable timely prevention and intervention, are of growing relevance. Currently, it is possible to create well defined specific algorithms to perform analytic tasks with regards to specific topics including pedestrian, baggages, logo and intrusion detection and tracking [1], [2]. However, the results of these algorithms largely depend upon the training and testing databases used for development. Unfortunately, in many real-world applications such approaches break down. The underlying model shortcomings lead to very low performance. More generic approaches in the context of visual information retrieval [3] try to overcome known shortcomings by combining several descriptors [4], [5], [6] exploiting crowdsourcing for semantic inference [7], using relevance feedback [8], implementing description hierarchies [9] or using biologically inspired approaches for key-word inference [10]. Nevertheless the main challenges in dealing with real-world scenarios remain open. Thus, the development of techniques that exploit available knowledge and semantic relationships is critical. Here, approaches based on the design of an ontological knowledge description aligned to the scenario or use case of concern are key. Ontologies have been widely used in many domains to formally represent the domain knowledge. They can be used to perform automated reasoning, knowledge inference and answer domain specific queries. Some of the groundwork has been presented in previous papers with the aim to define the basic concepts and create a new approach to the study of related fields. Heum Park et al. [11] developed Cyber Forensic Ontology for cyber investigation in cyber space. A small scale digital device forensics ontology was presented by Harill and Mislan [12] to provide law enforcement with the appropriate knowledge regarding the devices found in the Small Scale Digital Device (SSDD) domain. In [13] Kahvedzic and Kechadi proposed a framework for modelling, analysis and reuse of digital forensic knowledge to describe an investigation at different level of detail.

The focus of this paper is on the development of an advanced intelligent forensic retrieval system. Due to fact that Ontological knowledge representation is regarded as the main important step in this development, we present analysis and development of a forensic domain ontology for an intelligent visual surveillance system. The designed Ontology is capable to share the common understanding of the domain structure among forensic investigators. which is demonstrating the uniqueness of the designed Ontology as a solid design for addressing difficulties in handling, investigating and filtering a large amount of data in the forensic computing context daily. The ontology is applied to a relevant use case drawn from a real-world event (the UK riots 2011). Which is able to describe activities, entities, relationships, resources and consequences of the event. We propose an ontology design based on 'top-down development' approach. The rest of the paper is structured as follows. Section II describes high-level semantic reasoning in visual surveillance and in Section III we explain on data source and meta data generation. Scenario setting for case study is presented in Section IV with details on prototype implementation, meta data model, ontological relations and reasoning. Finally, conclusions are drawn in Section V.

[1]www.lasie-project.eu

978-1-4673-6870-4/15/$31.00

II. High-Level semantic reasoning in visual surveillance

The attention on the forensic use in an intelligent visual surveillance is getting increased in both real time observation and post incident investigation. Hence, the existence of an automated visual surveillance with the capability of prediction of unexpected and dangerous events for enabling timely prevention and intervention is necessary. However,the high-level information analysis is difficult particularly in the complex plot [14]. Thus, the critical requirements for achieving the high-level semantic reasoning capabilities which is regarded as our focus for the future development are defined below.

- **Meta-data representation:** For increasing the potential utilisation of the system for the upcoming query and reasoning process, all the existing semantic understanding and intermediate result needs to be explored. Generally, in the less complex plot, manual annotation could be used for recognizing the meaning of different concept. Whereas in the process of more complex plot and concepts, automatic detection algorithm (i.e. human , logo, ..) is necessary. Due to fact that visual surveillance domain mostly deals with complex plot, flexible machine-readable tools such as a meta data representation model for archiving the processing result is required. The ontological knowledge description on the possible meta data is a key factor in provision of machine representation. By providing the specific and shareable semantic to terminology enables the collection of meaningful subset of information from unstructured data set. Ontologies have been defined as an important tool for the semantic information representation and automated process of information in variety applications such as data mining, retrieval and automated discovery in a knowledge management scene. The main focus of ontology is on knowledge sharing and formal representation. Ontologies have been regarded as the appropriate tool for domain knowledge representation because of several advantages. Their most important property is that they provide a formal framework for supporting explicit, shareable, machine-processable semantics definition of domain knowledge, and they enable the derivation of implicit knowledge through automated inference. In particular, an ontology is a formal specification of a shared conceptualization of a domain of interest [15] and form an important part of the emerging semantic web, in which ontologies allow to organize contents through formal semantics. Web Ontology Language (OWL) and Semantic Web Rule Language (SWRL) have been proposed by the World Wide Web Consortium (W3C) as language standards for representing ontologies and rules, respectively. SPARQL Protocol and RDF Query Language has been approved as W3C recommendation as query language for the Semantic Web technologies [16]. Thus, ontologies are suitable for expressing video content semantics and may aid in filtering mass amounts of data.
- **Semantic knowledge representation** The system needs an efficient approach for describing the contextual knowledge about scenes while interfacing the complex plot at reasoning time. Unlike the state model based approach, here the reasoning objective has not been defined at the development time. Thus a declared problem could not be directly solved by the recognized set of steps. As a result, the more flexible semantic inference mechanism allowing the modularity of each knowledge segment is of quite importance in our development.
- **Reasoning under uncertainty** Development of the sophisticated representation and reasoning capabilities for rules, logic and proof layer of semantic web is the next vital step in semantic web construction. Constructing knowledge representation formalism with the capability to deal with uncertainty has recently attracted much attention [17]. In spite of the fact that the ontology has an undeniable success, classical ontologies are not appropriate enough for dealing with vagueness in knowledge, as a result several research efforts are going in this direction. Particularly, some researches are going on integration of rules with ontologies while there is a large concentration on the handling uncertainty and vagueness in the semantic web focusing on probabilistic uncertainty and adding probabilistic and fuzzy vagueness for integration of rules with uncertainty.

III. Metadata Generation

For the metadata source, we have processed 3.07 TB data of CCTV surveillance footage, human operator cameras and devices provided by the Scotland Yard as a part of European LASIE project. *The LASIE project aims to design and implement an open and expandable framework that will significantly increase the efficiency of current investigation, by providing an automated initial analysis of the vast amounts of heterogeneous forensic data that analysts have to cope with. The framework will be able to handle forensic data that has been acquired from a variety of different sources including CCTV surveillance content, confiscated desktops and hard disks, mobile devices, Internet, social networks, handwritten and calligraphic documents.*

In the first step, 3.07 TB of raw video data provided by Met Police is analyzed. The 3.07 TB sample video footage includes several types of video files such as;

- MP4 Video File (.mp4)
- Quick Time Video File (.mov)
- Media Player Classic (.ssf)
- Media Player Classic

The raw video footage provided by Scotland Yard, is varied in terms of video quality, image format and the media player. In the video sample currently used by the Met Police, several media players were identified for displaying purposes. However, some video footage could only be displayed using their own specific players and was not entirely compatible with standard decoders. Moreover, several types of corruption have been identified in the entire video source. As research algorithms are applied on the high quality videos, all the readable videos have been separated from the initial sample for further processing. After further analysis 0.907 TB of readable

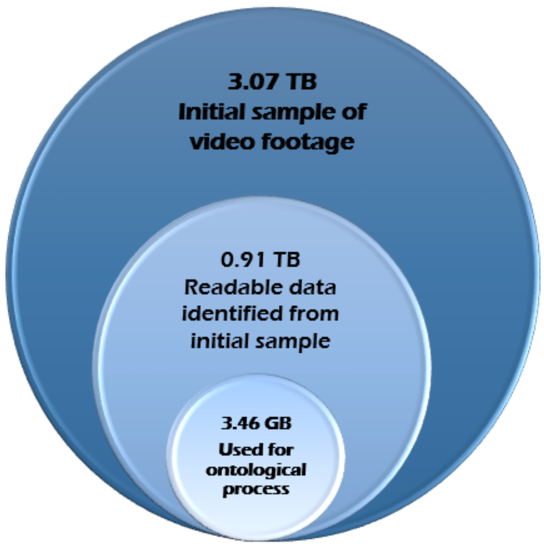

Fig. 1: Data set process

data has been identified for research usage. Due to fact that the presented work in this paper aims to focus on the analysis and development of a very small domain ontology, 3.46 GB data has been selected as the main source of our domain from the 0.907 TB for manual annotation and being used for ontological analysis. Figure 1, illustrated the process of data selection.

IV. Scenario Setting for Case Study

In this paper, we develop a comprehensive forensic domain ontology and apply a test case scenario of a sports riot event using the ontology. The event sources are obtained from CCTV footage and video footage captured using hand held devices.

The riot event happens on the 8th August 2011 in St Andrews, England in a football match between two local teams, Rovers and City. It is known that there has been a history of rivalry and animosity between the two sets of fans. The end of season match created further tension, as the losing side would be relegated from the Premier Division. The weather was hot and humid.The match took place in front of 40,000 fans at 3pm and was extremely tense. The score was 3-3 until extra time. Anger erupted when a last minute penalty was awarded to Rovers. The goal was scored resulting in a 4-3 victory for Rovers, which as a consequence resulted in humiliation for City and relegation. At this point missiles were thrown between the rival fans. Police attempts to segregate the crowd as it spilled out of the ground were initially successful, but due to the hot weather, the fans made for the town center, using a variety of public houses used by the one set of fans or the other

Based on manual annotation that has been done on 3.46 GB of the selected CCTV and video footage, a few activities have been identified during the riot. Some examples are:

- Loitering around the town center
- Smashing vehicles (buses) in the middle of the road

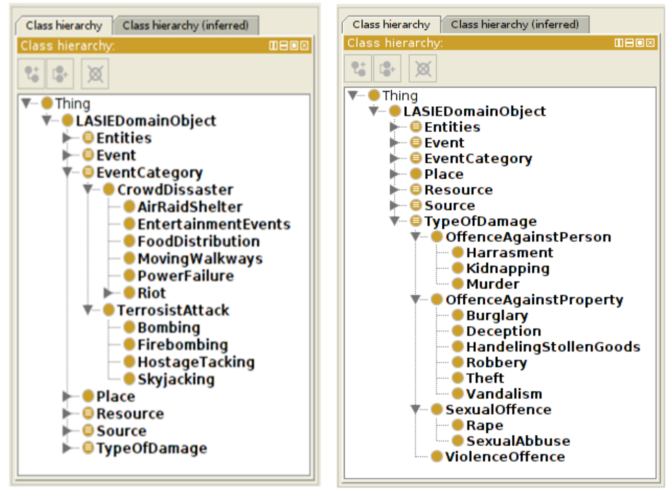

Fig. 3: Forensic domain Ontology with subclass of EventCategory and TypeOfDamage

- Attack the police and getting involve in a fight
- Potential argument with police officers
- Burning cars in protest to display their dissatisfaction
- Smashed a London black cab windows and moved the car to the middle of the road
- Entering private properties
- Throwing stones to the police
- Smash up shop premises and looted

A. Prototype Implementation

We used an approach called ”top-down development process” [15] in designing our ontology. The process starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts. In this project, we start with creating classes for the generic concepts of forensic data which are *entities*, *event*, *event category*, *place*, *resource*, *source* and *types of damage*. We then specialize each class by creating its subclasses. The graphical representation of ontological metadata model is illustrated in Fig. 1.

Clearly, an ontology defines various objects within a program's domain. It can be said that an ontology defines the world as a program depicts it. To accurately model a domain, it is important to answer several questions regarding the ontology like:

- What kind of event categories happening?
- What event are in the process ?
- Who is involved in the event?
- Where is the location of the event?
- What information can be obtained from the event?
- What are the consequences of the event?
- What sources are available for this event?

The competency questions help to determine the scope of the ontology. Judging from the list of questions, the ontology will include information on the kind of event happening, people involved, location, effect and data origin.

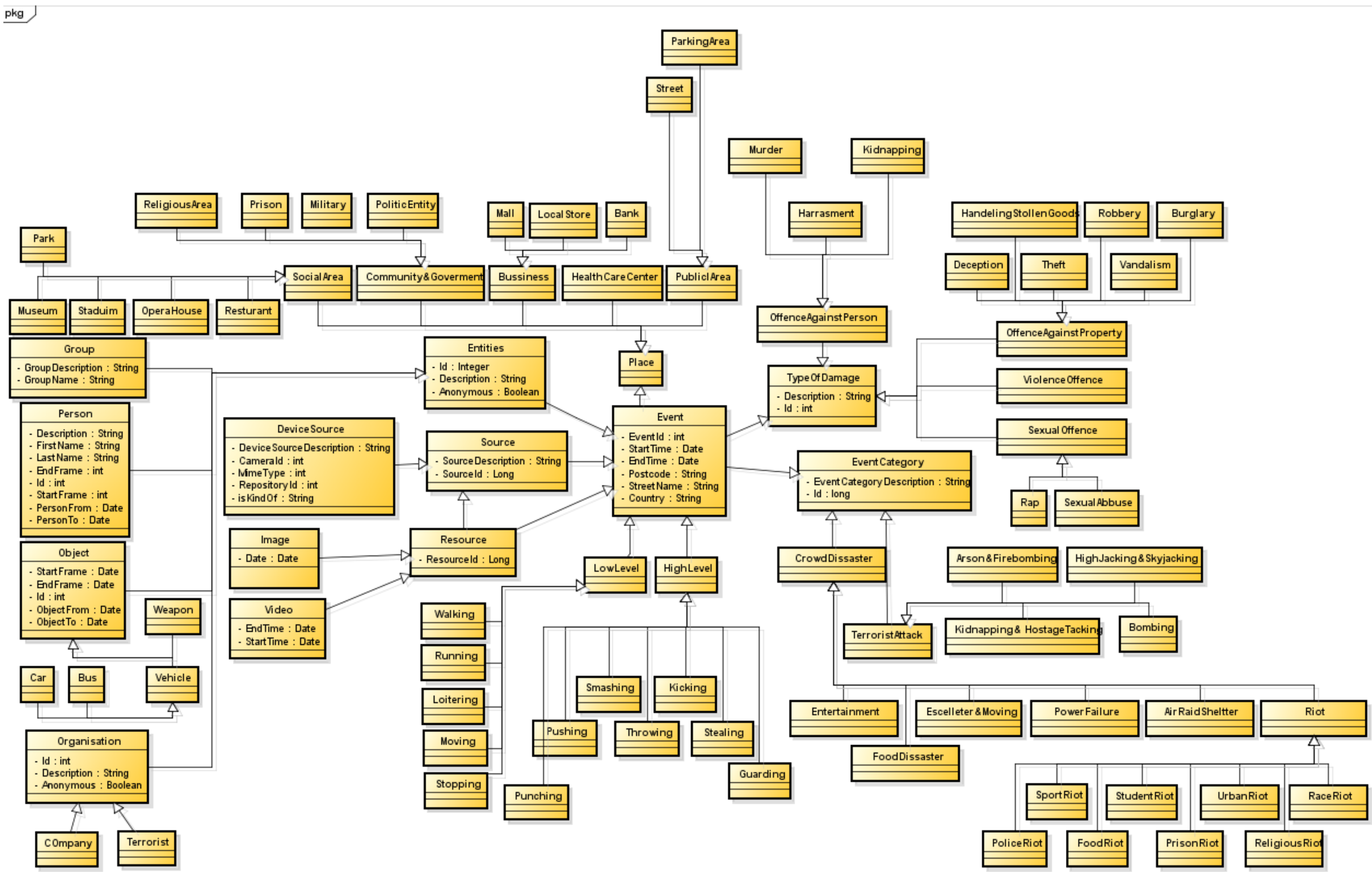

Fig. 2: Graphical representation of ontological metadata model

B. Metadata Model

OWL is chosen as the ontology representation language for its rich representation for the semantics and support for many tools. In this paper, Protege is used to develop the ontology. Protege is a free open-source platform that provides tools to construct domain models and knowledge-based applications with ontology. It was originally developed to support knowledge acquisition for specialized medical expert systems, but today it has many plugins for various features, from making constraints on attribute values to exporting ontology in different formats (CLIPS, OWL, RDF, RDFS and HTML) [18] and importing concepts from other tools into Protege.

In this paper, the forensic domain ontology is divided into seven classes as explained in section A. These classes are selected based on important terminology in forensic domain describing an event, which can occur in a wide variety of venues and different circumstances. In the design, the *event category* class is classified into two sub classes which are *crowd disaster* and *terrorist attack*. These sub classes are represented further by various incidents which results in chaos, injuries and deaths. Types of damage that might occur as a consequence of the event are also defined in the ontology. These details can be seen in Fig. 2.

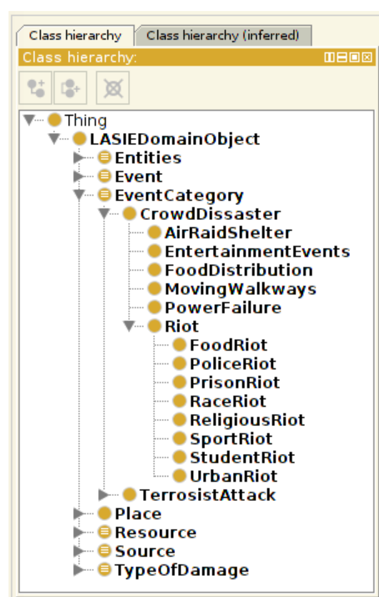

Fig. 4: Riot subclass

To test the ontology, we define some events based upon the London riots aftermath happened on Monday 8 August 2011. Riot is a form of civil disorder commonly characterized by a group lashing out in a violent public disturbance against authority, property or people. It often occurs in reaction to a perceived grievance or out of dissent. Riots typically involve vandalism and the destruction of property, public or private. In this work, we define *riot* as a subclass of *crowd disaster* and divided it into several types as in Fig. 3.

Today, there is a growing interest among the computer vision community in vision-based human action and activity recognition with applications to visual surveillance and video retrieval. Generally, the main goal of human activity and action recognition system is to recognize an action or goal of one or several agents from a series of observations on each particular agent and a given context. Many authors in the vision community have discussed that it makes sense to define a hierarchy of different action complexities as movements, activities and actions. Movements are defined as an action primitive, out of which activities and actions are compose [19]. Therefore, in the ontology design, *events* are defined based on level of complexity. Accordingly, in the designed ontology an action is considered as a sequence of primitive actions that fulfill a function or simple purpose such as jumping, walking, or kicking; while activity is composed of sequences of actions over space and time, such as people smashing bus windows or people burning a car. In this paper, *loitering*, *moving*, *running*, *stopping* and *walking* are classified as low-level events and *guarding*, *kicking*, *punching*, *pushing*, *smashing*, *stealing* and *throwing* are grouped as a high-level events.

People and objects involved in the riot scene are defined in *entities* class. *Entities* includes *person*, *group*, *organization* and *object*. Locations of the event are also defined in the ontology design. Another important classes that are presented in the ontology is the *source* and *resource* class. *Source* denotes source of data from various devices such as CCTV, hand held devices or mobile phones, while *resource* are types of content such as image or video produced by the device.

C. Ontological Relations

Relationships (also known as relations) between objects in an ontology specify how objects are related to other objects. Typically, a relation of a particular type (or class) specifies in what sense the object is related to the other object in the ontology. Much of the power of ontologies comes from the ability to describe relations. The set of relations describes the semantics of the domain. Resource Description Framework (RDF), is a language standards for representing ontologies. It allows ontology developers to define statements about things (or resources) in the form of *subject-predicate-object* expressions (known as RDF-triples). In short, we can say that a *subject* is the resource in question, an *object* is the target or the value of the triple and *predicate* is used to denote relations between the subject and the object. Individual instances are the most specific concepts represented in an ontology. Individuals are created as class instances. In this paper, 12 individuals are defined in *event* class and eight individuals are defined in *person* class. A total of 31 individuals have been defined in different classes of the ontology. As mentioned earlier, a statement describing things is represented in the form of a triple in an ontology. For example, in a statement *Person A has damaged Vehicle 1*, *person A* is defined as a subject, *has damaged* is a predicate and *Vehicle 1* is an object. In Protege, predicate is defined in Object Properties tab. Another features called Data Properties describes relationships between an individual and data value in an ontology. This allows more individual information to be included in the ontology structure. Table 1 shows general ontological relations that has been identified in the riot events and Fig. 4 explains the events in graphical forms.

Observe that the pictures in Fig 4 have been catoonized to obey privacy regulations governing the use of the very sensitive data provided by Scotland Yard. Though the original footage cannot be published, the images in Fig. 4 closely represent anonymized versions of the real footage.

Ontological Relations		
Subject	Relation	Object
Person A	is	Walking
Person A	is	Smashing
Person A	hasDamaged	Vehicle-1
Person B	is	Walking
Person C	is	Pushing
Person D	is	Pushing
Person E	is	Kicking
Person F	is	Loitering
Person G	is	Throwing
Group B	is	Loitering
Group A	is	Running
Group A	is	Guarding
Group B	is	Smashing
Group B	is	Moving
Group B	hasDamaged	Vehicle-3

TABLE I: General Ontological Relations

D. Reasoning

Many applications developed for the semantic purpose requires some kind of reasoning capability. This is because the intermediate metadata comes with uncertainty which affects the acceptable accuracy and robustness for a semantic complex query. Providing sound and complete reasoning services are essential for many of these applications to function properly. Pellet [20] is the first sound and complete OWL-DL reasoner with extensive support for reasoning with individuals (including nominal support and conjunctive query), user-defined datatypes, and debugging support for ontologies. A practical OWL reasoner provides at least the standard set of Description Logic inference services, such as consistency checking. This ensures that an ontology does not contain any contradictory facts. An OWL consistency checker takes a document as input, and returns one word being Consistent, Inconsistent, or Unknown. In this paper, we apply Pellet reasoner to check our ontology design consistency. Tableaux reasoner [21] is the core of the system that checks the consistency of a knowledge base. The output of the process shows that there are no inconsistency issues in our ontology design where all the facts and axioms that has been defined in the ontology are satisfactory.

V. Conclusion

Visual surveillance is proved to be a non-cost nor time effective process. This process usually requires a considerable amount of manual process done by investigators. The current available technologies for visual surveillance systems are restricted to basic query engines, lacking enhanced and exploratory investigation techniques. Proposing a more effective

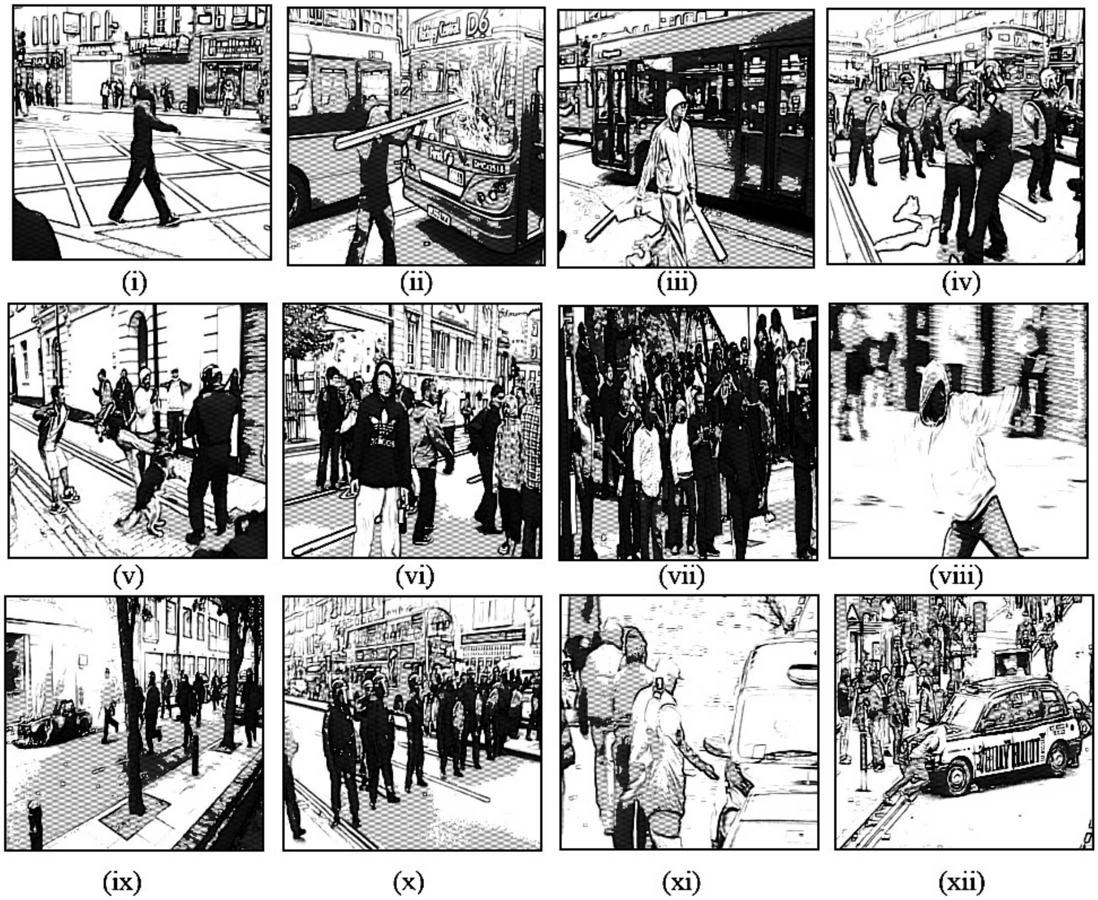

Fig. 5: General ontology relation in graphical form. (i) and (ii) Person A is walking and smashing Vehicle 1, (iii) Person B is walking, (iv) Person C is pushing Person D, (v) Person E is kicking, (vi) and (vii) Person F and Group B is loitering, (viii) Person G is throwing, (ix) and (x) Group A is running and guarding, (xi) and (xii) Group B is smashing and moving Vehicle B.

approach by automating the visual surveillance investigation process is believed to be a necessity.

Through out the paper the focus was put to the development of an advanced intelligent forensic retrieval system. it has has been proved that an ontological knowledge description is the vital requirement at the earliest stage of this system. Accordingly the paper covered an analysis and development of a forensic domain ontology for an intelligent visual surveillance system. A hierarchical ontology framework was created for forensic domain purposes based on the data related to the London riots aftermath incident in 2011. With this ontology, common understanding of the domain structure can be shared among forensic investigators. This demonstrates the uniqueness of ontology as a solid design to address handling, searching and filtering mass amounts of data in a forensic computing context. It can be concluded that the use of ontologies will indeed aid investigators in filtering mass amounts of data and thus also in completing investigations within the required time frame in a cost effective manner.

ACKNOWLEDGMENT

This work is supported in part by the LASIE project (http://www.lasie-project.eu/) with funding from the European Unions Seventh Framework Programme for research, technological development.

REFERENCES

[1] J. R. Cózar, N. Guil, J. M. González-Linares, E. L. Zapata, and E. Izquierdo, "Logotype detection to support semantic-based video annotation," *Signal Processing: Image Communication*, vol. 22, no. 7, pp. 669–679, 2007.

[2] T. Kliegr, K. Chandramouli, J. Nemrava, V. Svatek, and E. Izquierdo, "Combining image captions and visual analysis for image concept classification," in *Proceedings of the 9th International Workshop on Multimedia Data Mining: held in conjunction with the ACM SIGKDD 2008*. ACM, 2008, pp. 8–17.

[3] E. Izquierdo, J. R. Casas, R. Leonardi, P. Migliorati, N. E. O'Connor, I. Kompatsiaris, and M. Strintzis, "Advanced content-based semantic scene analysis and information retrieval: the schema project," 2003.

[4] Q. Zhang and E. Izquierdo, "combining low-level features for semantic inference in image retrieval," *Journal on Advances in Signal Processing, Accepted*, 2007.

[5] J. C. Caicedo and E. Izquierdo, "Combining low-level features for improved classification and retrieval of histology images," *Transactions on Mass-Data Analysis of Images and Signals*, vol. 2, no. 1, pp. 68–82, 2010.

[6] Q. Zhang and E. Izquierdo, "A multi-feature optimization approach to object-based image classification," in *Image and Video Retrieval*. Springer, 2006, pp. 310–319.

[7] L. Seneviratne and E. Izquierdo, "An interactive framework for image annotation through gaming," in *Proceedings of the international conference on Multimedia information retrieval*. ACM, 2010, pp. 517–526.

[8] K. Chandramouli, T. Kliegr, J. Nemrava, V. Svátek, and E. Izquierdo, "Query refinement and user relevance feedback for contextualized image retrieval," 2008.

[9] J. Calic and E. Izquierdo, "A multiresolution technique for video indexing and retrieval," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. I–952.

[10] K. Chandramouli and E. Izquierdo, "Image classification using self organising feature map and particle swarm optimisation," in *Proceedings of 3rd International Conference on Visual Information Engineering*, 2006, pp. 313–316.

[11] H. Park, S. Cho, and H.-C. Kwon, "Cyber forensics ontology for cyber criminal investigation," in *Forensics in Telecommunications, Information and Multimedia*. Springer, 2009, pp. 160–165.

[12] D. C. Harrill and R. P. Mislan, "A small scale digital device forensics ontology," *Small Scale Digital Device Forensics Journal*, vol. 1, no. 1, p. 242, 2007.

[13] D. Kahvedžić and T. Kechadi, "Dialog: A framework for modeling, analysis and reuse of digital forensic knowledge," *digital investigation*, vol. 6, pp. S23–S33, 2009.

[14] S. 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, 2009. WIAMIS'09. 10th Workshop on*. IEEE, 2009, pp. 201–204.

[15] J. Slay and F. Schulz, "Development of an ontology based forensic search mechanism: Proof of concept," *arXiv preprint arXiv:1407.8258*, 2014.

[16] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, and G. Serra, "Event detection and recognition for semantic annotation of video," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 279–302, 2011.

[17] T. Lukasiewicz and U. Straccia, "Managing uncertainty and vagueness in description logics for the semantic web," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, no. 4, pp. 291–308, 2008.

[18] J. Ćosić, Z. Ćosić, and M. Baća, "An ontological approach to study and manage digital chain of custody of digital evidence," *Journal of Information and Organizational Sciences*, vol. 35, no. 1, pp. 1–13, 2011.

[19] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 633–659, 2013.

[20] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: A practical owl-dl reasoner," *Web Semantics: science, services and agents on the World Wide Web*, vol. 5, no. 2, pp. 51–53, 2007.

[21] F. Baader and U. Sattler, "An overview of tableau algorithms for description logics," *Studia Logica*, vol. 69, no. 1, pp. 5–40, 2001.