Towards a Forensic Event Ontology to Assist Video Surveillance-based Vandalism Detection⋆

Faranak Sobhani[1] and Umberto Straccia[2]

[1] Queen Mary University of London, UK
[2] ISTI-CNR, Italy

Abstract. The detection and representation of events is a critical element in automated surveillance systems. We present here an ontology for representing complex semantic events to assist video surveillance-based vandalism detection. The ontology contains the definition of a rich and articulated event vocabulary that is aimed at aiding forensic analysis to objectively identify and represent complex events. Our ontology has then been applied in the context of London Riots, which took place in 2011. We report also on the experiments conducted to support the classification of complex criminal events from video data.

1 Introduction

In the context of vandalism and terrorist activities, video surveillance forms an integral part of any incident investigation and, thus, there is a critical need for developing an "automated video surveillance system" with the capability of detecting complex events to aid the forensic investigators in solving the criminal cases. As an example, in the aftermath of the London riots in August 2011 police had to scour through more than 200,000 hours of CCTV videos to identify suspects. Around 5,000 offenders were found by trawling through the footage, after a process that took more than five months.

With the aim to develop an open and expandable video analysis framework equipped with tools for analysing, recognising, extracting and classifying events in video, which can be used for searching during investigations with unpredictable characteristics, or exploring normative (or abnormal) behaviours, several efforts for standardising event representation from surveillance footage have been made [9,10,11,22,23,28,30,37].

While various approaches have relied on offering foundational support for the domain ontology extension, to the best of our knowledge, a systematic ontology for standardising the event vocabulary for forensic analysis and an application of it has not been presented in the literature so far.

In this paper, we present an OWL 2 [25] ontology for the semantic retrieval of complex events to aid video surveillance-based vandalism detection. Specifically, the ontology is a derivative of the DOLCE foundational ontology [7] aimed to

---

represent events that forensic analysts commonly encounter to aid in the investigation of criminal activities. The systematic categorisation of a large number of events aligned with the philosophical and linguistic theories enables the ontology for interoperability between surveillance systems. We also report on the experiments we conducted with the developed ontology to support the (semi-)automatic classification of complex criminal events from semantically annotated video data.

Our work significantly extends the preliminary works [12,31]. The work [12] is an embryonal work investigating about the use of an ontology for automated visual surveillance systems, which then has been then further developed in [31]. While our work shares with [31] some basic principles in the development of the ontology, here the level of details is now higher (e.g., the Endurant class (see Section 3.2) and its sub-classes have not been addressed in [31]) and various ontological errors have been revised. Additionally, and more importantly, in our work experiments have been conducted for criminal event classification based on London 2011 riots videos. Furthermore, but less related, is [32] in which the technical challenges facing researchers in developing computer vision techniques to process street-scene videos are addressed. The work focusses on standard image processing methods and does not deal with ontologies in any way.

The remainder of the paper is organised as follows. Related work is addressed in Section 2. Section 3 presents a detailed description of the forensic ontology about complex criminal events. In Section 4 we discuss how to use the ontology to assist video surveillance-based vandalism detection. In Section 5 we conduct some experiments with our ontology based on CCTV footage of London riots from 2011, and finally, Section 6 concludes.

2 Related Work

In [23], the *Event Recognition Language* (ERL) is presented, which can describe hierarchical representation of complex spatiotemporal and logical events. The proposed event structure consists of primitive, single-thread, and multi thread events. Another event representation ontology, called $CASE^E$, is based on natural language representation and is proposed in [11] and then extended in [10]. Subsequently, in [9,22] a *Video Event Representation Language* (VERL) was proposed for describing an ontology of events and the companion language called *Video Event Markup Language* (VEML), which is a representation language for describing events in video sequences based on OWL [21]. In [30], event detection is performed using a set of rules using the SWRL language [24].

The *Event Model E* [37] has been developed based on an analysis and abstraction of events in various domains such as research publications, personal media [1], meetings [13], enterprise collaboration [14] and sports [26]. The framework provides a generic structure for the definition of events and is extensible to the requirements ontology of events in the most different concrete applications and domains.

In [28] a formal model of events is presented, called *Event-Model-F*. The model is based on the foundational ontology DOLCE+DnS Ultralite (DUL) and provides comprehensive support to represent time and space, objects and persons, as well as mereological, casual, and correlative relationships between events. In addition, the Event-Model-F provides a flexible means for event composition, modelling event causality and event correlation, and representing different interpretations of the same event. The Event-Model-F is developed following the pattern-oriented approach of DUL, is modularised in different ontologies, and can be easily extended by domain specific ontologies.

While the above-mentioned approaches essentially provide frameworks for the representation of events, none of them address the problem of formalising forensic events in terms of a standard representation language such as OWL 2[3] and, importantly, none have been applied and tested so far in a real use case, which are the topics of the following sections.

3   A Forensic Event Ontology

In the following, we present an OWL 2 ontology to support to some extent the semantic retrieval of complex events to aid automatic or semi-automatic video surveillance-based vandalism detection. The idea is to develop an ontology that not only conveys a shared vocabulary, but some inferences based on it may assist a human being to support the video analysis by hinting to videos that may be more relevant than others in the detection of criminal events.

3.1   The Role of a Foundation Ontology

To facilitate the elimination of the terminological ambiguity and the understanding and interoperability among people and machines [19], it is common practice to consider a so-called *foundational ontology*. Let us note that several efforts have been taken by researchers in defining the foundational ontologies, such as BFO,[4] SUMO,[5] UFO[6] and DOLCE,[7] to name a few. As DOLCE ontology offers a cognitive bias with the ontological categories underlying natural language and human common sense, the same is selected for our proposed extension. We recall that the DOLCE foundational ontology encompasses Endurant and Perdurant entities. Endurant entities are ever-present at any time as opposed to Perdurant entities that extended in time by accumulating different temporal parts. A more thorough explanation on the DOLCE events conceptualisation can be found e.g. in [7].

---

[3] We recall that the relationship to our previous work [12,31,32] has been addressed in the introductory section.
[4] http://ifomis.uni-saarland.de/bfo/
[5] http://www.adampease.org/OP/
[6] https://oxygen.informatik.tu-cottbus.de/drupal7/ufo/
[7] http://www.loa.istc.cnr.it/old/Papers/DOLCE2.1-FOL.pdf

3.2 A Forensic Complex Event Ontology

Our complex event classes extend DOLCE's Perdurant class. To assign the action classes into respective categories, we follow a four-way classification of action-verbs: namely, into State, Process, Achievement and Accomplishment using event properties such as telic, stage and cumulative (see [27,35,36]). The distinction between these concepts are derived from the event properties as illustrated in Table 1, which we summarise below.

Table 1. Classification of Event Types.

State	-telic	-stage	cumulative
Process	-telic	+stage	-
Achievement	+telic	-stage	not cumulative
Accomplishment	+telic	+stage	not cumulative

- **State [-telic,-stage]** This action category represents a long, non-dynamic event in which every instance is the same: there cannot be any distinction made between the stages. States are cumulative and homogenous in nature.
- **Process [-telic, +stage]** The action category, like State, is atelic, but unlike State, the action undertaken are dynamic. The actions appear progressively and thus can be split into a set of stages for analysis.
- **Accomplishment [+telic, +stage]** Accomplishments are telic and cumulative activities, and thus behave differently from both State and Process. The performed action can be analysed in stages and in this way, they are similar to Process. Intuitively, an accomplishment is an activity which moves toward a finishing point as it has variously been called in the literature. Accomplishment is also cumulative activity.
- **Achievement [+telic, -stage]** Achievements are similar to Accomplishment in their telicity. They are also not cumulative with respect to contiguous events. Achievements do not go on or progress, because they are near instantaneous, and are over as soon as they have begun.

Forensic Perdurant Entities. Perdurant entities extend in time by accumulating different temporal parts and some of their proper temporal parts may be not present. To this end, Perdurant entities are divided into the classes Event and Stative, classified according to their temporal characteristics.

The axiom sets below provide a subset of our formal extension of the Perdurant vocabulary.

$$
\begin{aligned}
\text{Perdurant} &\sqsubseteq \text{SpatioTemporalParticular} \\
\text{Perdurant} &\sqsubseteq \exists \text{participant.Endurant} \\
\text{Fighting} &\sqsubseteq \exists \text{participant.GroupOfPeople} \\
\text{Perdurant} &\sqsubseteq \neg \text{Endurant} \\
\text{Kicking} &\sqsubseteq \neg \text{Vehicle}
\end{aligned}
$$

$$
\begin{aligned}
\text{State} &\sqsubseteq \text{Stative} \\
\text{MetaLevelEvent} &\sqsubseteq \text{State} \\
\text{Accusing} &\sqsubseteq \text{MetaLevelEvent} \\
\text{Believing} &\sqsubseteq \text{MetaLevelEvent} \\
\text{PsycologicalAggression} &\sqsubseteq \text{State} \\
\text{Blaming} &\sqsubseteq \text{PsycologicalAggression} \\
\text{Bullying} &\sqsubseteq \text{PsycologicalAggression}
\end{aligned}
$$

$$
\begin{aligned}
\text{Process} &\sqsubseteq \text{Stative} \\
\text{Action} &\sqsubseteq \text{Process} \\
\text{Gesture} &\sqsubseteq \text{Process} \\
\text{PhysicalAggresion} &\sqsubseteq \text{Process} \\
\text{ActivePhysicalAggresion} &\sqsubseteq \text{PhysicalAggresion}
\end{aligned}
$$
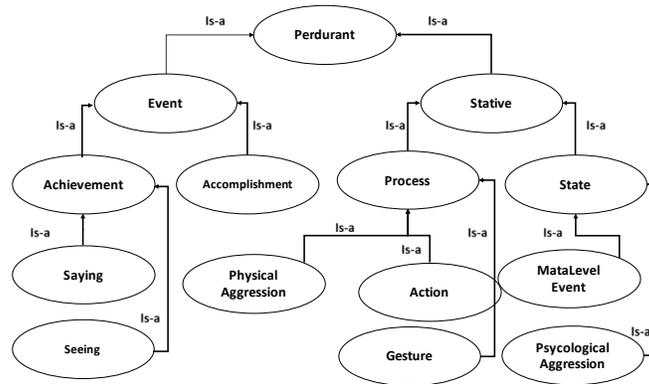
Fig. 1. The Perdurant class hierarchy for forensic events descriptions.

$$Accomplishment \sqsubseteq Event$$
$$CriminalEvent \sqsubseteq Accomplishment$$
$$EventCategory \sqsubseteq Accomplishment$$
$$Crimecategory \sqsubseteq Stative$$

$$Achievement \sqsubseteq Event$$
$$Saying \sqsubseteq Achievement$$
$$Seeing \sqsubseteq Achievement \ .$$

An excerpt of the forensic ontology is shown in Figure 1.

The concept State offers representation for MetaLevelEvent which encompasses abstract human events such as Accusing, Believing and Liking among others. As previously stated, the concept State represents a collection of events which are exhibited by a human that is time-consuming, non-dynamic, cumulative and homogenous. The other sub-class of State is PsychologicalAggression which characterises the human actions such as Blaming, Decrying, Harassing and so forth. The concept Process includes several human action categories that represent dynamic events which can be split into several intermediate stages for analysis. For the purposes of clarity, the concept Process offers three sub-concepts namely Action, Gesture and PhysicalAggression. The Action class incorporates different event such as Dancing, Greeting, Hugging among other concepts defined. The concept Gesture formalises the different interest points related to human gestures. In order to eliminate the ambiguity traditionally present in human gestures across cross-cultural impact, the action performed during the gesture is captured and represented in the ontology and, thus, enabling the removal of subjectivity from the concept definition. The final sub-class of the Process class includes the concept PhysicalAggression and formalises human conflicting actions.

By and large, the human actions categorised into State and Process represent the microscopic movements of humans.

From the automatic surveillance viewpoint, these microscopic events may be extracted from media items. In contrast, the event representations formalised by means of the concepts Achievement and Accomplishment offer a rich combination of human events that allow for the construction of complex events with or without the combination of microscopic features. For instance, the concept
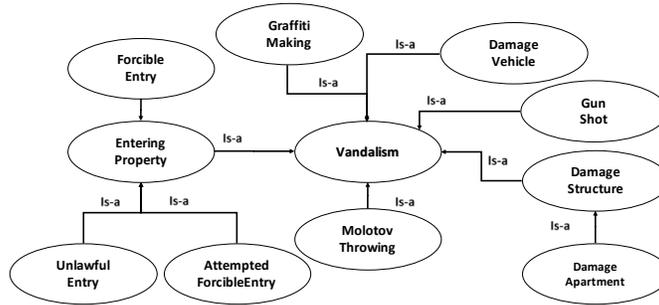
Fig. 2. The concept hierarchy of Vandalism, direct subclass of CrimeAgainstProperty. The latter is a subclass of class Accomplishment.
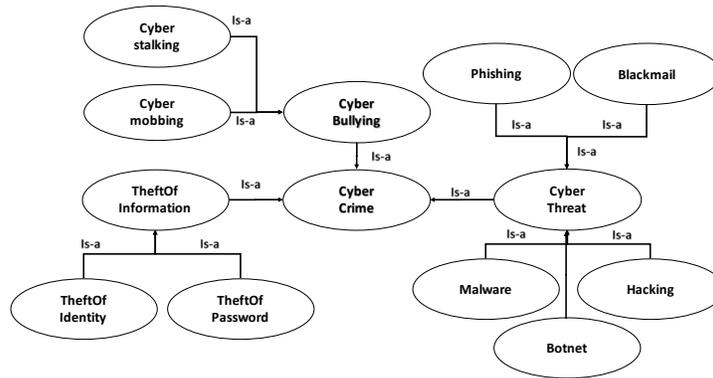

Fig. 3. The concept hierarchy of CyberCrime.

hierarchy for Vandalism is illustrated in Figure 2, while the concept hierarchy for CyberCrime is shown in Figure 3 instead.

Forensic Endurant Entities. DOLCE is based on fundamental distinction among Endurant and Perdurant entities. The difference between Endurant and Perdurant entities is related to their behaviour in time. Endurant are wholly present at any time they are present. Philosophers believe that endurant are entities that are in time while lacking, however, temporal parts [19]. Therefore, the proposed vocabulary structure of all possible forensic entities also extends on Endurants entities.
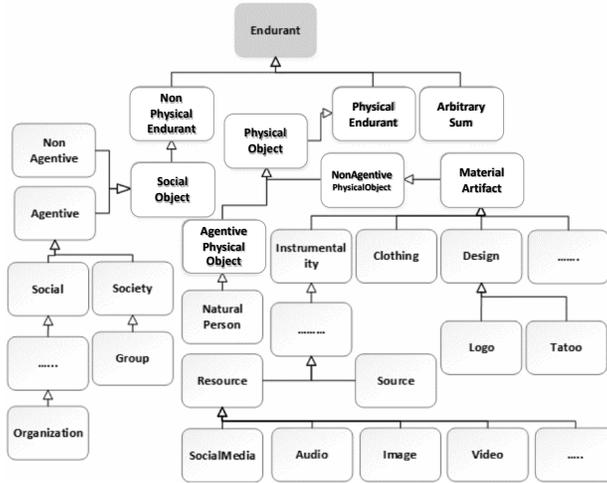
Fig. 4. Excerpt of the Endurant concept hierarchy in the forensic ontology.

Axiom set (1) describes a subset formalization of the Endurant vocabulary and an excerpt of the forensic extension of the ontology structure shown in Figure 4.

$$
\begin{aligned}
\text{Endurant} &\sqsubseteq \text{SpatioTemporalParticular} \\
\text{Endurant} &\sqsubseteq \exists \text{participantIn.Perdurant} \\
\text{participantIn} &= \text{participant}^{-} \\
\text{NonPhysicalEndurant} &\sqsubseteq \text{Endurant} \\
\text{PhysicalEndurant} &\sqsubseteq \text{Endurant} \\
\text{ArbitrarySum} &\sqsubseteq \text{Endurant} .
\end{aligned}
\tag{1}
$$

4 Assisting Video Surveillance-based Vandalism Detection

We next show how the so far developed ontology is expected to be used to assist video surveillance-based vandalism detection.

4.1 Annotating Media Objects, viz. Surveillance Videos

Given surveillance videos and any media in general, we need a method to annotate them by using the terminology provided by our ontology. This gives rise to a set of facts that, together with the inferred facts, may support a more effective automatic or, more likely, semi-automatic retrieval of relevant information, such as e.g. vandalic acts. Specifically, the inferred information may suggest a user look at some e.g. video sequences or video still images, rather than to others first.

The general model we are inspired on is based on [20]. Conceptually, according to [20], a media object *o* (e.g., an image region, a video sequence, a piece

of text, etc.) is annotated with one (or more) entities t of the ontology (see e.g. Figure 5).
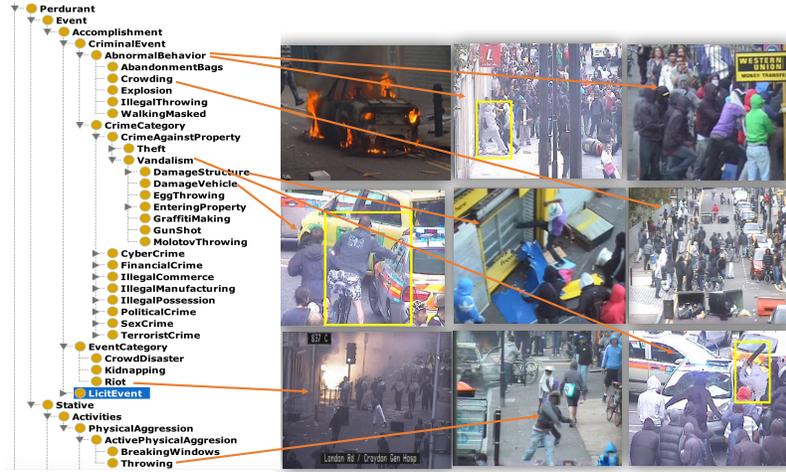
Fig. 5. Examples of still image annotations from the London Riots 2011 of events as per Table 2.

For instance, stating that an image object o *is about* a DamageVehicle can be represented conceptually via the DL expression

$$(\exists \mathsf{isAbout.DamageVehicle})(o) \ .$$

As specified in [20], such an annotation may come manually from a user or, if, available, from an image classifier. In the latter case, it may annotate the image automatically, or, semi-automatically by suggesting to a human *annotator*, which are the most relevant entities of the ontology that may be used for a specific media object o. Note, however, that, the above methodology just illustrates the concept. In our case, for the sake of ease the annotation, we may not enforce the use of the object property isAbout (see Example 2 later on). Generally, we will annotate a Resource with Perdurants and Endurants: thus, if an image is annotated with e.g. a perdurant that is a damaged vehicle, then this means that the image *is about* a damaged vehicle.

We recall that Resources (and Sources) are modelled as follows:

$$
\begin{aligned}
\mathsf{Source} \ &\sqsubseteq \ \mathsf{Endurant} \sqcap \exists \mathsf{has.Resource} \\
&\sqcap \exists \mathsf{hasCameraId.string} \\
&\sqcap \exists \mathsf{hasLatitude.string} \\
&\sqcap \exists \mathsf{hasLongitute.string} \\
&\sqcap \exists \mathsf{hasLocationName.string} \\
\mathsf{Resource} \ &\sqsubseteq \ \mathsf{Endurant} \sqcap \exists \mathsf{has.Perdurant} \\
\mathsf{has} \ &= \ \mathsf{isFrom}^- \\
\mathsf{has} \circ \mathsf{has} \ &\sqsubseteq \ \mathsf{has} \ .
\end{aligned}
$$

Note that in the last role inclusion axiom, $\circ$ is role composition and, thus, has $\circ$ has $\sqsubseteq$ has dictates that the property has is transitive, while with has =

isFrom⁻ we say that isFrom is the inverse of has. Therefore, isFrom is transitive as well.

The following example illustrates the mechanism of image of annotation together with a meaningful inference.

Example 1. Consider the following DL axioms resulting from annotating images of a video (video6) registered by a camera (cameraC004):

$$
\begin{aligned}
&\mathsf{participateIn(personA, throwing5)} \ , \mathsf{Throwing(throwing5)} \\
&\mathsf{NaturalPerson(personA)} \ , \mathsf{Throwing} \sqsubseteq \mathsf{ActivePhysicalAggression} \\
&\mathsf{ActivePhysicalAggression} \sqsubseteq \mathsf{PhysicalAggression} \ , \mathsf{PhysicalAggression} \sqsubseteq \mathsf{Process} \\
&\mathsf{isFrom(throwing5, endurant6)} \ , \mathsf{Resource(endurant6)} \\
&\mathsf{hasVideoId(endurant6, video6)} \ , \mathsf{Source(endurant7)} \\
&\mathsf{hasCameraId(endurant7, cameraC004)} \ , \mathsf{has(endurant7, endurant6)} \ .
\end{aligned}
$$

Now, as isFrom is transitive, we may infer:

$$\mathsf{isFrom(throwing5, endurant7)} \ .$$

Then, it is not difficult to see that we finally infer

$$
\begin{aligned}
&\exists \mathsf{paticipateIn.(PhysicalAggression} \sqcap \\
&\qquad\qquad \exists \mathsf{isFrom.(Source} \ \sqcap \ \exists \mathsf{hasCameraID.\{cameraC004\}))\,(PersonA)} \ ,
\end{aligned}
$$

which can be read as:

"A person (PersonA) participated in a physical aggression that has been registered by camera C004".

4.2   Modelling GCIs for Vandalism Event Detection

As we are focusing on forensic domain and dealing with variety of concepts aiming at aiding forensic analysis, to objectively identify and represent complex events, we next show that a (manually build) *General Concept Inclusion* (GCI) axiom may help to classify high-level events in terms of a composition of some lower level events. The following are such GCI examples:

Fig. 6. Example of DamageVehicle and DamageStructure scenes in CCTV.

DamageVehicle:

Perdurant ⊓
∃participant.(Vehicle ⊓
∃participantIn.(BreakingDoor ⊔ BreakingWindows))
⊑ DamageVehicle .

"If an event involves a vehicle that is subject of a breaking door or breaking windows then the event is about a damaged vehicle" (see Figure 6).

DamageStructure:

Perdurant ⊓
∃participant.(Structure ⊓
∃participantIin.Kicking) ⊑ DamageStructure .

"If an event involves a structure that is subject of kicking, then the event is about a damaged structure" (see Figure 6).

The following example illustrates the use of such GCIs.

Example 2. Suppose we have an image classifier that is able to provide us with the following facts. Specifically, assume it is able to identify vehicles and breaking windows:

participant(Perdurant2, Endurant1), Vehicle(Endurant1) , BreakingWindows(Perdurant2) .

From these facts and the GCI about DamageVehicle, we may infer that the image is about a damaged vehicle, *i.e.* we may infer

DamageVehicle(Perdurant2) .

The following set of GCIs illustrates instead how one may have multiple GCIs to classify a single event, such as those for Vandalism (see, e.g. Figure 7).[8]

Fig. 7. Example of Vandalism scenes in CCTV videos.

Perdurant ⊓ ∃part.(Crowding ⊓ DamageStructure) ⊑ Vandalism
Perdurant ⊓ ∃part.(Crowding ⊓ DamageVehicle) ⊑ Vandalism
Perdurant ⊓ ∃part.(Explosion ⊓ Throwing) ⊑ Vandalism .

Note that in the example above, we assume that events (perdurant) may be complex in the sense that they may compose by multiple sub-events (parts). So, e.g. in the last GCI, we roughly state

---

[8] Recall that all these GCIs provide sufficient conditions to be an instance of Vandalism, but no necessary condition.

> *"If a (complex) event involves both throwing and an explosion (two sub-events) then the event is about vandalism".*

Following our previous examples, we next are going to formulate another kind of background knowledge. Our main focus in this example is on recognizing high-level events, which occur in the same location (same street in our modelling). In order to model this scenario, we may use the *Semantic Web Rule Language* (SWRL) to model the locatedSameAs role and then use it in GCIs. The SWRL rule is:

> *"Two perdurants that occur in the same street occur in the same place.*

$$\text{Perdurant}(?p1), \text{Perdurant}(?p2), \text{hasLocationName}(?p1, ?l1),$$
$$\text{hasLocationName}(?p2, ?l2), \text{SameAs}(?l1, ?l2) \rightarrow \text{locatedSameAs}(?p1, ?p2) \ .$$

Fig. 8. Examples of events that happen in the same location (locatedSameAs) from CCTV.

The following axioms illustrate how to use the previously defined relation (few examples captured from our data set by these rules are illustrated in Figure 8).

$$\text{Perdurant} \sqcap$$
$$\exists \text{part.}(\text{Crowding} \ \sqcap \exists \text{locatedSameAs.Explosion}) \sqsubseteq \text{Vandalism}$$

$$\text{Perdurant} \sqcap$$
$$\exists \text{part.}(\text{Crowding} \ \sqcap \exists \text{locatedSameAs.DamageStructure}) \sqsubseteq \text{Vandalism}$$

$$\text{Perdurant} \sqcap$$
$$\exists \text{part.}(\text{Crowding} \ \sqcap \exists \text{locatedSameAs.Throwing}) \sqsubseteq \text{Vandalism}$$

$$\text{Perdurant} \sqcap$$
$$\exists \text{part.}(\text{DamageStructure} \ \sqcap \exists \text{locatedSameAs.Throwing}) \sqsubseteq \text{Vandalism} \ .$$

5 Experiments

We conducted two experiments with our ontology, which we are going to describe in the following.[9]

In the first case, we evaluated the classification effectiveness of manually built GCIs to identify crime events, while in the second case we drop the manual-built GCIs and, try to learn such GCIs instead automatically from examples and compare their effectiveness with respect to the manually built ones.

---

[9] The ontologies used in the experiments and experimental results can be found at http://www.umbertostraccia.it/cs/ftp/ForensicOntology.zip.

5.1 Classification via Manually Built GCIs

Roughly, we have considered a number of crime videos, annotated them manually and then checked whether the manually built GCIs, as described in Section 4.2, were able to determine crime events correctly.

Setup. Specifically, we considered our ontology and around 3.07 TB of video data about the London riot 2011,[10] of which 929 (GB) is in a non-proprietary format. We considered 140 videos (however, the videos cannot be made publicly available). Within these videos, all the available CCTV cameras (35 CCTV) along with their features such as latitude, longitude, start time, end time and street name, have been annotated manually according to our methodology described in Section 4 and included into our ontology. We have also calculated all the geographic distances between each camera. The resulting ontology contains 1800 created individuals of which, e.g. 106 are of type Event.

Table 2. Criminal event classes considered.

Vandalism $(13, 57)$	Riot $(4, 21)$	AbnormalBehavior $(2, 80)$
Crowding $(1, 64)$	DamageStructure $(3, 9)$	DamageVehicle $(3, 16)$
Throwing $(1, 30)$		

Then, we considered criminal events occurring in the videos (specifically, we focused on vandalic events). For each class of events, we manually built one or more GCIs, as illustrated in Section 4.2. The list of crime events considered is reported in Table 2. In it, the first number in parenthesis reports the number of GCIs we built for each of them, while the second number indicates the number of event instances (individuals) we created during the manual video annotation process. So, for instance, for the event DamageStructure we have built 3 classification GCIs and we have created 9 instances of DamageStructure during the manual video annotation process. For further clarification, the 3 GCIs for DamageStructure are

$$
\begin{aligned}
&\text{Perdurant} \sqcap \\
&\exists\text{participant.}(\text{Structure} \sqcap \\
&\qquad\qquad \exists\text{participantIin.Kicking}) \sqsubseteq \text{DamageStructure} \\
&\text{Perdurant} \sqcap \\
&\exists\text{participant.}(\text{Structure} \sqcap \\
&\qquad\qquad \exists\text{participantIin.Beating}) \sqsubseteq \text{DamageStructure} \\
&\text{Perdurant} \sqcap \\
&\exists\text{participant.}(\text{Structure} \sqcap \\
&\qquad\qquad \exists\text{participantIin.BreakingWindows}) \sqsubseteq \text{DamageStructure} ,
\end{aligned}
$$

while, e.g., an instance of DamageStructure is the individual Kicking1, whose related information excerpt is:

$$
\begin{aligned}
&\text{Kicking(Kicking1), isFrom(Kicking1, 2bdf), Resource(2bdf), isFrom(2bdf, C004),} \\
&\text{has(2bdf, pr11), part(pr11, Kicking1), part(pr11, BreackingWindows3),} \\
&\text{BreackingWindows(BreackingWindows3), . . .}
\end{aligned}
$$

As a matter of general information, the global metric statistics of the so built ontology is reported in Table 3.

Table 3. Ontology Metrics.

Axioms	9889
Logical axiom count	7176
Class count	483
Object property count	148
Data property count	51
Individual count	1800
DL expressivity	SHIQ(D)

SubclassOf axioms count	532
EquivalentClasses axioms count	5
DisjointClasses axioms count	11
GCI count	38
Hidden GCI Count	5

SubObjectPropertyOf axioms count	93
InverseObjectProperties axioms count	20
TransitiveObjectProperty axioms count	5
SymmetricObjectProperty axioms count	2
ObjectPropertyDomain axioms count	19
ObjectPropertyRange axioms count	18

SubDataProperty axioms count	11
DataPropertyDomain axioms count	1
DataPropertyRange axioms count	5

ClassAssertion axioms count	1793
ObjectPropertyAssertion axioms count	2964
DataPropertyAssertion axioms count	1706

AnnotationAssertion axioms count	195

Evaluation. Let $\mathcal{O}$ be the built ontology from which we drop axioms stating explicitly that an individual is an instance of a crime event listed in Table 2. Please note that without the GCIs none of the crime events instances in $\mathcal{O}$ can be inferred to be instances of the crime events in Table 2.[11] Now, on $\mathcal{O}$ we run an OWL 2 reasoner that determines the instances of all crime event classes in the ontology.

To determine the classification effectiveness of the GCIs, we compute the so-called micro/macro averages of precision, recall and F1-score w.r.t. inferred data. The evaluation result of the first test is shown in Table 4.

Table 4. Results for the experiment on classification of manually build GCIs .

Event	TP	FP	FN	TN	$|C|$	$|trueC|$	$Precision_C$	$Recall_C$	$F1_C$
Vandalism	42	0	15	168	42	57	1.00	0.74	0.85
DamageVehicle	11	0	5	209	11	16	1.00	0.69	0.81
DamageStructure	9	0	0	216	9	9	0.89	0.89	0.89
Crowding	60	1	4	160	61	64	0.98	0.94	0.96
Throwing	30	0	0	195	30	30	1.00	1.00	1.00
Riot	5	0	16	204	5	21	1.00	0.24	0.38
AbnormalBehaviour	70	22	10	123	92	80	0.76	0.88	0.81

	$Precision_{micro}$	$Recall_{micro}$	$F1_{micro}$	$Precision_{macro}$	$Recall_{macro}$	$F1_{macro}$
	0.91	0.82	0.86	0.96	0.78	0.86

5.2 Classification via Automatically Learned GCIs

In the second experiment, we apply a concept learning approach to replace the manually built GCIs describing the crime events listed in Table 2. To this end,

---

[11] Roughly, crime events are subclasses of the Event class, while crime event instances are instances of the class Stative (see Figure 3).

the DL-Learner[12] system was used to learn descriptions of the criminal events in Table 2, based on existing instances of these classes.

Setup. Let now $\mathcal{O}$ be the ontology as in Section 5.1, but from which we also drop additionally the manually created GCIs for the crime event listed in Table 2. On it we used the CELOE algorithm [6,15] with its default settings to generate suggestion definitions (inclusion axioms) for each target class C.

Specifically, we used a K-fold cross style validation method [8], which divides the available crime event instances into a K disjoint subsets. That is, we split each target class C into K disjoint subsets $C_1, \ldots, C_K$ of equal size. In our experiment, K is the number of instances of C and, thus, each C_i has size one For each C_i, the *training set* is $(\bigcup_{i=1}^{K} C_i) \setminus C_i$ and is denoted as $Trainset_i$. Then, for each C_i we run CELOE on the training set $Trainset_i$ and generated at most 10 class expressions of the form $D_j \sqsubseteq C_i$, out of which we have chosen the best solution (denoted $D_{C_i} \sqsubseteq C_i$ or GCI_i). If the best solution is not unique, we select the first listed one.

The best-selected GCIs found by CELOE for each of the target classes in Table 2 are:

PhysicalAggression ⊓ ∃immediateRelation.Structure ⊑ DamageStructure
∃immediateRelation.Vehicle ⊑ DamageVehicle
∃immediateRelation.Vandalism ⊑ AbnormalBehavior
∃immediateRelation.Arm ⊑ Throwing
∃immediateRelation.Group ⊑ Crowding .

With the help of a reasoner, we then infer all instances in $\mathcal{O}$, that are not in $Trainset_i$, that are instances of the selected D_{C_i} and consider them as our *result set* (denoted $Resultset_i$).

Evaluation. To determine the classification effectiveness of the learned GCIs, *i.e.* of GCI_i, average precision, recall and F1 measures across the folds are computed. The evaluation results of the second test are shown in Table 5.

Table 5. Results for the experiment on classification using DL-Learner CELOE algorithm.

Event	$Precision_C$	$Recall_C$	$F1_C$
DamageVehicle	0.69	0.98	0.81
Damage Structure	1.00	1.00	1.00
Crowding	0.96	1.00	0.98
Throwing	0.86	0.99	0.92
AbnormalBehavior	0.69	0.99	0.81

$Precision_{micro}$	$Recall_{micro}$	$F1_{micro}$	$Precision_{macro}$	$Recall_{macro}$	$F1_{macro}$
0.753	0.964	0.845	0.599	0.709	0.649

Discussion. The results are generally promising. In the manually built GCI case, precision and F1 are reasonably good, though in one case (Riot) the recall and, thus, F1 is not satisfactory. For the learned GCI case, the individual measures are generally comparable to the manual ones.

---

[12] http://dl-learner.org/

Given that the learned GCIs are completely different than the manually built ones, it is surprising that both sets perform more or less the same. However, please note that DL-Learner was neither able to learn a GCI for Vandalism nor for Riot. This fact is reflected in the generally worse micro/macro precision, recall and F1 measures.

Eventually, we also merged the manually built GCIs and the learned ones together and tested them as in Section 5.1. The results in Table 6 show, however, that globally their effectiveness is as for the manual case (and does not improve).

Table 6. Results of merging manual and learned GCIs.

Event	$Precision_C$	$Recall_C$	$F1_C$
Vandalism	1.00	0.74	0.85
DamageVehicle	1.00	0.69	0.81
Damage Structure	0.89	0.89	0.89
Crowding	0.98	0.94	0.96
Throwing	1.00	1.00	1.00
Riot	1.00	0.24	0.38
AbnormalBehavior	0.76	0.89	0.82

$Precision_{micro}$	$Recall_{micro}$	$F1_{micro}$	$Precision_{macro}$	$Recall_{macro}$	$F1_{macro}$
0.90	0.82	0.86	0.95	0.77	0.85

6 Conclusions

In this work, we have proposed an extensive ontology for representing complex criminal events. The proposed ontology focuses on events that are often required by forensic analysts. In this context, the Perdurant, as defined in the DOLCE ontology as an occurrence in time, and the Endurant, defined in the DOLCE ontology as contentious in time, have both been extended to represent all forensics entities together with meaningful entities for video surveillance-based vandalism detection. The aim of the built ontology is to support the interoperability of the automated surveillance system.

To classify high-level events in terms of the composition of lower level events we focused on both manually built and automatically learned GCIs and have compared the evaluation results of both experiments. The results are generally promising and the effectiveness of machine derived definitions for high-level crime events is encouraging though needs further development.

In the future, we intend to deal with vague or imprecise knowledge and we would like to work on the problem of automatically learn fuzzy concept description [4,5,16,17,18,33,34] as most of the involved entities are fuzzy by nature.

References

1. Appan, P., Sundaram, H.: Networked multimedia event exploration. In: Proceedings of the 12th Annual ACM International Conference on Multimedia. pp. 40–47. ACM (2004)

2. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)

3. Baader, F., Horrocks, I., Sattler, U.: Description logics. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 21–43. International Handbooks on Information Systems, Springer Verlag (2009), `https://doi.org/10.1007/978-3-540-92673-3_1`

4. Bobillo, F., Straccia, U.: Fuzzy ontology representation using OWL 2. International Journal of Approximate Reasoning 52, 1073–1094 (2011)

5. Bobillo, F., Straccia, U.: The fuzzy ontology reasoner *fuzzyDL*. Knowledge-Based Systems 95, 12 – 34 (2016), `http://www.sciencedirect.com/science/article/pii/S0950705115004621`

6. Bühmann, L., Lehmann, J., Westphal, P.: DL-Learner framework for inductive learning on the semantic web. Web Semantics: Science, Services and Agents on the World Wide Web 39, 15–24 (2016)

7. Casati, R., Varzi, A.: Events. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Winter 2015 edn. (2015), `http://plato.stanford.edu/archives/win2015/entries/events/`

8. Forman, G., Scholz, M.: Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. ACM SIGKDD Explorations Newsletter 12(1), 49–57 (2010)

9. François, A.R.J., Nevatia, R., Hobbs, J.R., Bolles, R.C.: VERL: an ontology framework for representing and annotating video events. IEEE MultiMedia 12(4), 76–86 (2005), `https://doi.org/10.1109/MMUL.2005.87`

10. Hakeem, A., Shafique, K., Shah, M.: An object-based video coding framework for video sequences obtained from static cameras. In: Proceedings of the 13th Annual ACM International Conference on Multimedia. pp. 608–617. ACM (2005)

11. Hakeem, A., Sheikh, Y., Shah, M.: CASE^E: A hierarchical event representation for the analysis of videos. In: Proceedings of the 19th National Conference on Artificial Intelligence. pp. 263–268. AAAI Press (2004), `http://dl.acm.org/citation.cfm?id=1597148.1597192`

12. Henderson, C., Blasi, S.G., Sobhani, F., , Izquierdo, E.: On the impurity of street-scene video footage. In: 6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15). The Institution of Engineering and Technology (IET) (2015)

13. Jain, R., Kim, P., Li, Z.: Experiential meeting system. In: Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence. pp. 1–12. ACM (2003), `http://doi.acm.org/10.1145/982484.982486`

14. Kim, P., Podlaseck, M., Pingali, G.: Personal chronicling tools for enhancing information archival and collaboration in enterprises. In: Proceedings of the the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences. pp. 56–65. ACM (2004), `http://doi.acm.org/10.1145/1026653.1026662`

15. Lehmann, J., Auer, S., Bühmann, L., Tramp, S.: Class expression learning for ontology engineering. Web Semantics: Science, Services and Agents on the World Wide Web 9(1), 71–81 (2011)

16. Lisi, F.A., Straccia, U.: A logic-based computational method for the automated induction of fuzzy ontology axioms. Fundamenta Informaticae 124(4), 503–519 (2013)

17. Lisi, F.A., Straccia, U.: A foil-like method for learning under incompleteness and vagueness. In: 23rd International Conference on Inductive Logic Programming.

Lecture Notes in Artificial Intelligence, vol. 8812, pp. 123–139. Springer Verlag, Berlin (2014), revised Selected Papers

18. Lukasiewicz, T., Straccia, U.: Managing uncertainty and vagueness in description logics for the semantic web. Journal of Web Semantics 6, 291–308 (2008)

19. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: Wonderweb Deliverable D18, Ontology Library (final). ICT project 33052 (2003), `http://wonderweb.man.ac.uk/deliverables/documents/D18.pdf`

20. Meghini, C., Sebastiani, F., Straccia, U.: A model of multimedia information retrieval. Journal of the ACM 48(5), 909–970 (2001)

21. Motik, B., Grau, B.C., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C., et al.: OWL 2 Web Ontology Language Profiles. W3C recommendation 27, 61 (2009)

22. Nevatia, R., Hobbs, J.R., Bolles, B.: An ontology for video event representation. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. p. 119. IEEE Computer Society (2004), `https://doi.org/10.1109/CVPR.2004.301`

23. Nevatia, R., Zhao, T., Hongeng, S.: Hierarchical language-based representation of events in video streams. In: IEEE Conference on Computer Vision and Pattern Recognition. p. 39. IEEE Computer Society (2003), `https://doi.org/10.1109/CVPRW.2003.10038`

24. OWL, S.A.S.W.R.L.C., RuleML: `https://www.w3.org/Submission/SWRL/`. W3C (2004)

25. OWL 2 Web Ontology Language Document Overview: `http://www.w3.org/TR/2009/REC-owl2-overview-20091027/`. W3C (2009)

26. Pingali, G.S., Opalach, A., Jean, Y.D., Carlbom, I.B.: Instantly indexed multimedia databases of real world events. IEEE Transactions on Multimedia 4(2), 269–282 (2002)

27. Rothstein, S.: Chapter 1: Verb classes and aspectual classification. In: Structuring Events: A Study in the Semantics of Lexical Aspect, pp. 1–35. Wiley Online Library (2004)

28. Scherp, A., Franz, T., Saathoff, C., Staab, S.: F–A model of events based on the foundational ontology DOLCE+DnS ultralight. In: Proceedings of the Fifth International Conference on Knowledge Capture. pp. 137–144. ACM (2009), `http://doi.acm.org/10.1145/1597735.1597760`

29. Schmidt-Schauß, M., Smolka, G.: Attributive concept descriptions with complements. Artificial Intelligence 48, 1–26 (1991)

30. Snidaro, L., Belluz, M., Foresti, G.L.: Representing and recognizing complex events in surveillance applications. In: Fourth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 493–498. IEEE Computer Society (2007), `https://doi.org/10.1109/AVSS.2007.4425360`

31. Sobhani, F., Chandramouli, K., Zhang, Q., Izquierdo, E.: Formal representation of events in a surveillance domain ontology. In: 2016 IEEE International Conference on Image Processing. pp. 913–917. IEEE Computer Society (2016), `https://doi.org/10.1109/ICIP.2016.7532490`

32. Sobhani, F., Kahar, N.F., Zhang, Q.: An ontology framework for automated visual surveillance system. In: 13th International Workshop on Content-Based Multimedia Indexing. pp. 1–7. IEEE Computer Society (2015), `https://doi.org/10.1109/CBMI.2015.7153628`

33. Straccia, U.: Foundations of Fuzzy Logic and Semantic Web Languages. CRC Studies in Informatics Series, Chapman & Hall (2013)

34. Straccia, U., Mucci, M.: pFOIL-DL: Learning (fuzzy) $\mathcal{EL}$ concept descriptions from crisp OWL data using a probabilistic ensemble estimation. In: Proceedings of the

30th Annual ACM Symposium on Applied Computing (SAC-15). pp. 345–352. ACM, Salamanca, Spain (2015)

35. Vendler, Z.: Verbs and times. The Philosophical Review 62(2), 143–160 (1957)
36. Vendler, Z. (ed.): Linguistics in Philosophy. G - Reference, Information and Interdisciplinary Subjects Series, Cornell University Press (1967), `https://books.google.co.uk/books?id=OR1DAAAAIAAJ`
37. Westermann, U., Jain, R.: Toward a common event model for multimedia applications. IEEE Multimedia 14(1) (2007)