

An Ontological Approach to Semantic Video Analysis for Violence Identification

T. Perperis¹ S. Tsekeridou²

¹University of Athens, Department of Informatics and Telecommunications,
Greece, (<http://www.di.uoa.gr>)

²Athens Information Technology,
Greece, (AIT, <http://www.ait.edu.gr>)

AIT Research Seminar 2011
AIT, Athens, Greece, 15 February, 2011

Outline

- 1 Introduction
- 2 Existing Approaches
- 3 Ontological Methodology for Violence Identification
- 4 Conclusions

Motivation

Motivation

- Exponential growth of Multimedia Content
- Uncontrollable dissemination of Objectionable Content

Common users and Industry:

- Demand for Intelligent, Human Like methods to:
 - ① Automatically Search and Classify video data
 - ② Automatically Detect and Annotate dangerous content
 - ③ Filter out this content, thus enabling high level parental control

Outline

- 1 Introduction
- 2 Existing Approaches**
- 3 Ontological Methodology for Violence Identification
- 4 Conclusions

Outline

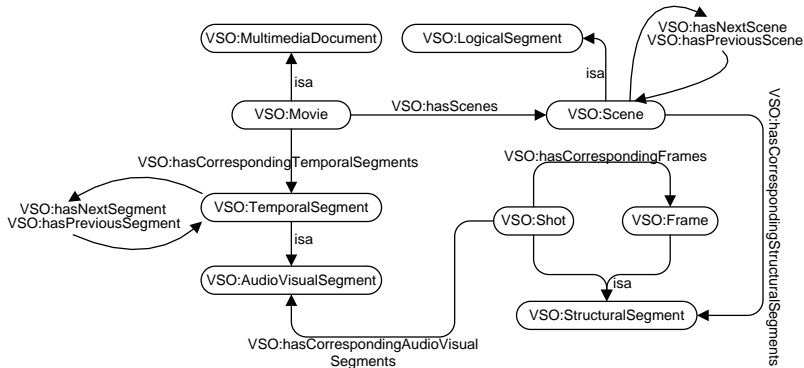
- 1 Introduction
- 2 Existing Approaches
- 3 Ontological Methodology for Violence Identification**
- 4 Conclusions

Ontological Methodology

Goals

- 1 Automatically detect any violence hidden in video data
- 2 Automatically annotate them accordingly
- 3 Enable content filtering for parental control
- 4 Not to devise high quality low level analysis processes
- 5 Combine existing single modality low to mid-level semantics detectors with ontologies and reasoning
- 6 Open/Extendable framework

Video Structure Ontology



Visual Analysis I¹

Activity Detection

- No Activity
- Normal Activity
- High Activity

Gunshot Detection

- No Gunshot
- Gunshot

Person Detection

- Haar Based Face Detection
- RGB Skin Detection

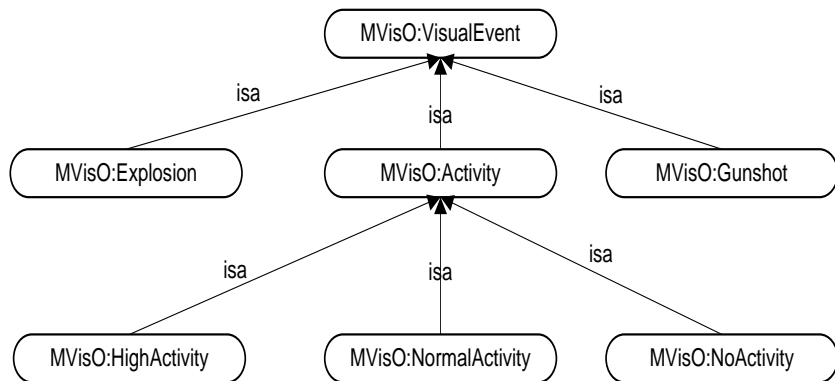
¹[Makris, 10]

Visual Ontology I

Include the hierarchical definition of:

- Detected visual objects possibly related with violence (or not)
- Detected visual events possibly related with violence (or not)
- An extensive range of various visual events and objects to provide attachment points for other concept detectors

Visual Ontology III - Events



Audio Analysis / Semantics

Audio Analysis I²

Detected Non Violent Classes

- 1 Music
- 2 Speech
- 3 Others1 (Sounds of low energy - Smooth environmental sounds like background noise, rain, wind)
- 4 Others2 (Sounds of abrupt changes in signal energy - Sharp environmental sounds like thunder, door closing etc.)

Detected Violent Classes

- 1 Gunshots
- 2 Screams
- 3 Fights

²[Giannakopoulos, 07]

Audio Analysis II

Employed Features

	Frame Feature	Sequence Statistic
1	Spectrogram	σ^2
2	Chroma 1	μ
3	Chroma 2	<i>median</i>
4	Energy Entropy	<i>max</i>
5	MFCC 2	σ^2
6	MFCC 1	<i>max</i>
7	ZCR	μ
8	Sp. RollOff	<i>median</i>
9	Zero Pitch Ratio	—
10	MFCC 1	<i>max</i> / μ
11	Spectrogram	<i>max</i>
12	MFCC 3	<i>median</i>

Audio Analysis III

One Vs All Strategy

- Split the main problem into K ($K = 6$) binary sub-problems (e.g. gunshot vs. no gunshot)
- Randomly split the 12-D feature vector into 3, 4-D feature sub vectors
- Train 3 kNN binary classifiers for each subproblem
- Feed the results of the 3 kNN classifiers into a Bayesian Network and produce the probability estimation for the subproblem
- 7 probability estimations are produced
- The winner is the class with the highest probability

Audio Ontology

Audio Ontology

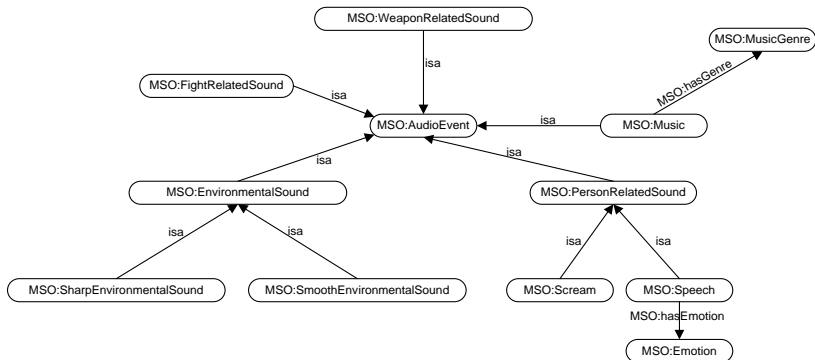
Defines Taxonomy of:

- 1 Audio Events (e.g. MSO:Gunshot, MSO:Screams, MSO:explosions, MSO:Speech, MSO:Music)
- 2 Type of background music (MSO:Pop, MSO:Rock, MSO:Classic etc)
- 3 Emotional Speech (e.g. MSO:Anger, MSO:Fear, MSO:Cry)

Contains:

- A much broader set of audio semantics for future extension

Audio Ontology



Scene Detection

Scene Detection

Shot Clustering

- Create a 10-D vector composed of Audio And visual classes probability estimations
- Feed a MCL algorithm that uses Random Walks to create clusters
- Instantiate VSO:Cluster class in the VS Ontology
- Use SQWRL to retrieve and instatiate Scenes (i.e. consecutive shots of the same cluster)

Scene Detection

Shot Clustering

- Create a 10-D vector composed of Audio And visual classes probability estimations
- Feed a MCL algorithm that uses Random Walks to create clusters
- Instantiate VSO:Cluster class in the VS Ontology
- Use SQWRL to retrieve and instatiate Scenes (i.e. consecutive shots of the same cluster)

Harmful Content Domain Ontology Definition

Harmful Content Ontology Definition I

Harmful Content Ontology

- First attempt to conceptualize violence and pornography in an organized way
- Synthesis of psychologists' violence definition and extensive investigation of movies depicting violent content
- Generic representation for use by psychologist, pedagogist, police
- Non violence classes implementation (scenery, dialogue, action) due to open world reasoning in OWL

Harmful Content Domain Ontology Definition II

Violence Ontology Defines:

- Complex Semantics of extensive violent acts
- Inter-relation of medium and low level semantics

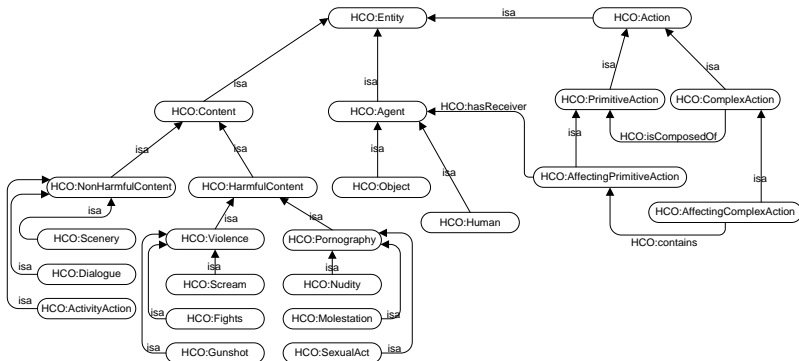
Construction of violent actions hierarchy

In a movie scene containing violence (e.g. torture, fight, war) a spectator can quickly grasp the form of violence (e.g. fighting without weapons), recognize a sequence of violent (e.g. punching, kicking), of generic (e.g. running, walking) and of consequence (e.g. falling, crawling, scream) actions.

Association Mechanisms

- Medium level classes ↔ Inferred multimodal actions
- Low level concepts ↔ Represented in visual and audio ontology

Harmful Content Domain Ontology



Up to Now

- Preprocessing
- Visual analysis
- Audio analysis
- Scene Detection

- Video Structure Ontology
- Visual Ontology
- Audio Ontology
- Video Structure Ontology

- Harmful Content Ontology
- Inferencing Procedure

Up to Now

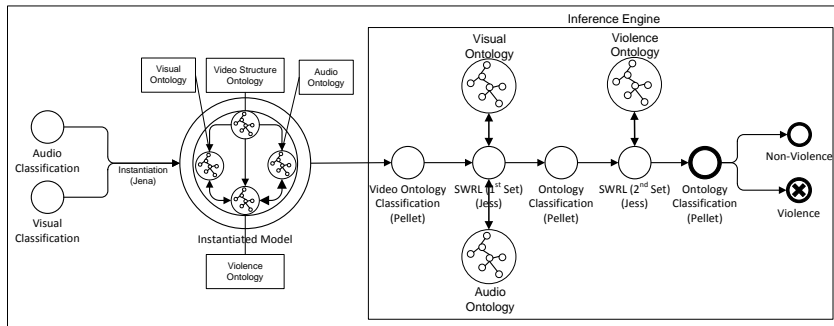
- Preprocessing
- Visual analysis
- Audio analysis
- Scene Detection

- Video Structure Ontology
- Visual Ontology
- Audio Ontology
- Video Structure Ontology

- Harmful Content Ontology
- Inferencing Procedure

Inferencing Procedure

Inferring Procedure



Discussion On Instantiation

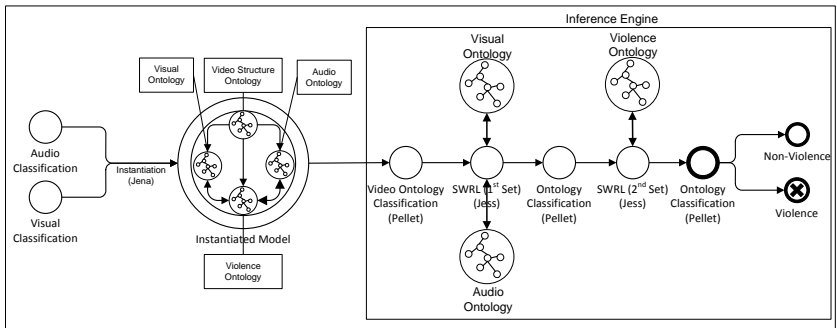
- Ontologies define the Terminological Box (TBox) of our knowledge base
- The Instantiated model - ABox:
 - Captures existing and extracted knowledge for the movie in question
 - Derives directly from the **Segmentation, Audio, Visual and Scene analysis**
 - Forms the basic facts in terms of individuals (shots, frames, events and objects)
 - Includes low level numerical values and accuracy probability results

Inference Engine Requirements

The inference engine should take under consideration:

- Intra- and Cross-modality spatial, temporal or spatio-temporal relationships
- Importance of each modality for identifying a concept or semantic event
- Cross-modality synchronicity relationships (simultaneous semantic instances in different modalities)
- Uncertainty of extracted medium level semantics
- Support reasoning with partial, imprecise information

Inferencing Lifecycle I



Inferring Lifecycle II

Steps:

- 1 Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

Infencing Lifecycle II

Steps:

- 1** Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

Inferring Lifecycle II

Steps:

- 1 Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

Inferring Lifecycle II

Steps:

- 1 Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

Inferring Lifecycle II

Steps:

- 1 Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

Inferring Lifecycle II

Steps:

- 1 Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

Inferring Lifecycle II

Steps:

- 1 Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

Inferring Lifecycle II

Steps:

- 1 Consistency check of the instantiated model and assertion of each individual's initial class
- 2 First SWRL Set Rules Application
- 3 Consistency checking and classification services on the implied model
- 4 Second SWRL Set Rules Application
- 5 Consistency checking and classification services are applied to:
 - Infer violent and non violent segments (children to parents)
 - Extract extended semantics (parents to children)

SWRL Example1

SWRL Example1: Identifying a non violent (Activity, Action)

Description	SWRL Rule
If	
?avs is an Audio Visual Segment	VSO:AudioVisualSegment(?avs) ∧
?ms is an individual of audio class Music	MSO:Music(?ms) ∧
?ha is an individual of visual class High-Activity	MVisO:HighActivity(?ha) ∧
In ?avs music is detected	VSO:hasAudioEvent(?avs,?ms) ∧
In ?avs high activity is detected	VSO:hasVisualEvent(?avs,?ha) ∧
Then	
?avs is an action segment	→ HCO:ActivityAction(?avs)

Inferring Higher Level Of Semantics

Identifying Person-On-Person-Fighting and Multiple-Person-Fighting

HCO:Fighting Subclasses Definition	Necessary and Sufficient Conditions
HCO:PersonOnPersonFighting	HCO:displaysObjects some MVisO:Face \wedge HCO:displaysObjects exactly 2
HCO:MultiplePersonFighting	HCO:displaysObjects some MVisO:Face \wedge HCO:displaysObjects min 3

Implementation

Implementation

- Matlab and the OpenCV library for audio/visual feature extraction and classification
- Protégé^a for the definition of ontologies and SWRL rules
- Pellet and Jess^b for ontology reasoning services and rules execution
- Jena semantic web framework^c for ontologies instantiation and synchronization of the knowledge generation lifecycle

^a<http://protege.stanford.edu>

^b<http://www.jessrules.com/>

^c<http://jena.sourceforge.net/index.html>

Results II

Segment based Violence vs Non-violence Detection Performance Measures

	Recall	Precision	F_1	Mean Accuracy
Audio-based	82.9%	38.9%	53%	61%
Visual-based	75.6%	34%	46.9%	54.8%
Ontology-based	91.2%	34.2%	50%	62.7%

Ontological Segment based Multiclass Inference Measures

	Recall	Precision	F_1	Mean Accuracy
Fights Inference	61.6%	68.2%	64.8%	64.9%
Screams Inference	41.4%	33.5%	37.1%	37.4%
Shots-Explosions Inference	63.3%	38.2%	47.6%	50.7%

Conclusions

Considering that

- 1 Extracted visual analysis clues are not at the desired level
- 2 Extracted audio and visual mid level clues are biased towards non-violence
- 3 Uncertain single modality results are treated as certain

We Conclude That

- 1 The attained results are really promising both for the binary and multiclass violence detection problem
- 2 The main advantage of using such an ontological approach still remains the higher level semantics extraction ability, using an unsupervised procedure and common sense reasoning

