Metadata description of thermal videos for rescue operations

András Hajdu, Charalambos Giamas, Nicholas Vretos, Ioannis Pitas
Department of Informatics, University of Thessaloniki, Box 451, 54124 Thessaloniki, Greece
Email: pitas@aiia.csd.auth.gr
Tel.:+30-231-099-6361

Abstract— **In this paper we present a metadata scheme for the content description of thermal videos captured during rescue operations. The formal description is provided in terms of an XML Schema, thus the content description files have XML format. The schema was designed to provide highly detailed description possibilities for the movie scene. The scheme description, which is implemented in full compatibility with the MPEG-7 standard, provides fast and simple integration capabilities for database applications, and supports video/image indexing and retrieval for further use. It has been already tested that the proposed schema can be conceptualized in ontology services, which have other input besides video analysis.**

I. Introduction

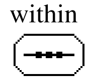
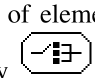
In one of our currently developed system, our main aim is to extract useful information from videos recorded by a thermal camera in a fire/rescue scenario. The main point of the application of thermal cameras instead of (or parallel to) normal ones, is that we can extract some useful data only from the infrared light domain. For example, in case of a firefighting operation, thermal cameras have the ability to capture visual data obscured by smoke or in complete darkness.

This video processing support is just one component of the complex mobile system SHARE [1] which has been designed for rescue forces with integrating multiple modes of interaction. Ontology services play an important role in the SHARE philosophy to establish connection between the different input channels (map data [2], deployment structure, audio/visual/text messages) regarding the rescue scenario. In this way, it becomes possible for the rescue team to access operative information based on a very wide basis through a retrieval client [3]. The information found by thermal video analysis is embedded into the ontology as a subset of its hierarchical data structure. The visual information extracted from the thermal videos are converted to metadata and populated to the ontology.

For video content metadata description, MPEG-7 has emerged as the standard [4], [5], [6], [7], [8] for multimedia data content description. For this reason, we decided to implement semantics of thermal videos in terms of an XML Schema in a fully MPEG-7 compatible way. To achieve larger flexibility, several description schemes were proposed to improve MPEG-7 performance in terms of semantic content description [9], [10]. We also provide a number of classes to extend the MPEG-7 standard to tune our description for rescue operations. Regarding to a firefighting operation, these classes include data to describe hotspots (fire), and object/human appearance. Temporal hotspot observation is essential to measure up the tactical effectiveness of the operation, while human appearance analysis is very helpful in tracking victim or personnel movement. Accordingly, besides having data for the physical details of a video, we define classes for hotspot and human occurrence. As for the hierarchical segmentation of the videos, we consider logical segments (clips), which can be further subdivided into segments with object appearances, inside which the actual processing is done at frame (instance) level.

The structure of the paper is as follows. In section II we explain the hierarchical partitioning of videos together with the classes defined to describe the basic identification and physical data. Section III presents two examples where the presented schema is used to describe fire and human appearance. Finally, in section IV we conclude with some possible future completions of the proposed schema to provide better support for a fire/rescue scenario.

Some general remarks for the figures in the paper, where we followed a popular way for the presentation of the schema:

- all the variable types for the field components have some standard MPEG7 type (e.g. integer, date, time, string),
- components bounded by solid boxes are obligatory (must be filled in),
- components bounded by dashed boxes are optional,
- number of allowed occurrences are written as "minoccur..maxoccur" under the respective components; $1..\infty$ stands for the unbounded condition,
- within a class, sequence of elements are represented by , while choices by .

II. Segmentation and basic metadata for thermal videos

The ThermalVideo class is the root class of the hierarchy, as it is depicted in Figure 1. It contains the classes for a thermal video file, and provides several representations for the civilian, firefighter, fire, and object appearances.

The ThermalVideoInfo class contains complementary data for the thermal video, like title, provider, scenario of rescue, date/time of recording, camera type used, and the physical description of the video in terms of its format, image size and length, as shown in Figure 2.

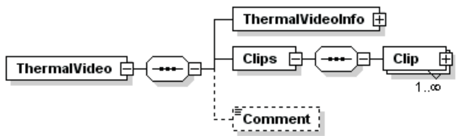


Fig. 1. The root (ThermalVideo) class of the hierarchy.

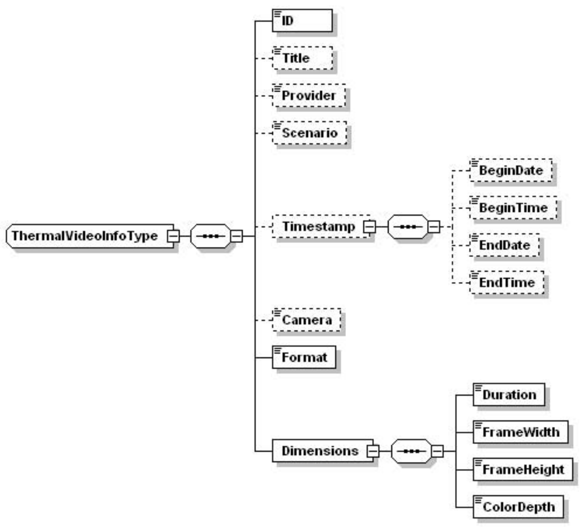


Fig. 2. The ThermalVideoInfo class for basic metadata of thermal videos.

The whole video represented by the ThermalVideo class can be divided into (even overlapping) video segments represented by the Clips class. A clip is a movie which contains civilian, firefighter, fire and object activity, as shown in Figure 3.

Similarly to ThermalVideo the Clip contains a ClipInfo class, as well, with complementary data (title, provider, scene of rescue, date/time of recording, video length). See Figure 4 for this class.

A clip is subdivided into smaller video segments for the description of possible appearances of humans or interesting objects. Thus we consider Appearance classes for civilian/firefighter/fire/object to describe the corresponding entity.

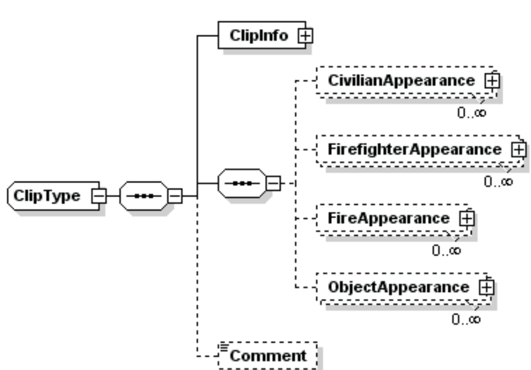


Fig. 3. The Clip class to describe segments of the whole video for specific appearances.

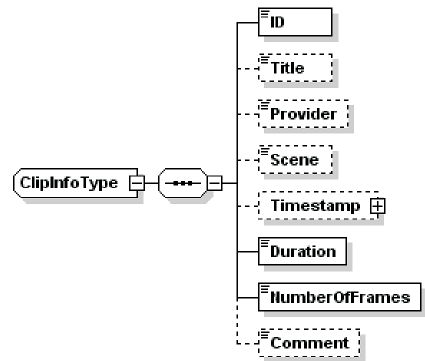


Fig. 4. The ClipInfo class for basic metadata of clips.

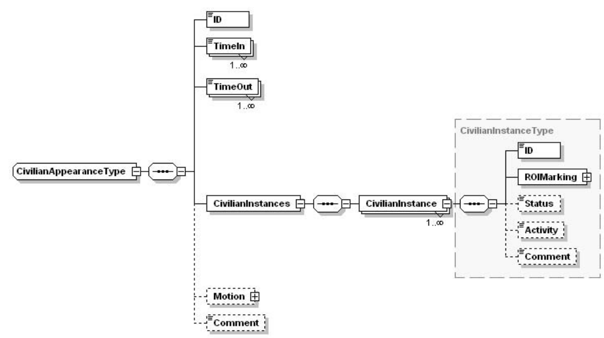


Fig. 5. A CivilianAppearance class to describe the occurrences and activities of an observed object within a clip.

These appearance classes are quite similar, see e.g. Figure 5 for the definition of the CivilianApperance class.

Finally, as the last step of the logical division of the thermal video, the Appearance classes are described through a sequence of instances at frame level. Thus Instance classes contain descriptions for the spatial position of the observed entity within the frame, together with its status or activity, as can be observed in Figure 5.

The FirefighterInstance class is just like the CivilianInstance one, where the differentiation between firefighters and civilians is achieved by image processing methods. Regarding fire description, we measure up only the status of the fire, as depicted in Figure 6.

The ObjectInstance class, shown in Figure 7, is slightly more detailed than the ones assigned for human behavior, as we have a separate Type field here to show the exact type of

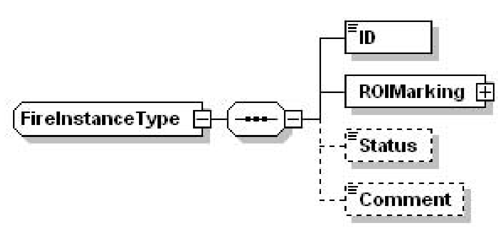


Fig. 6. The FireInstance class for a frame level description of a fire.

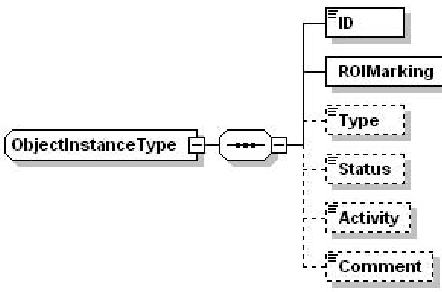


Fig. 7. The ObjectInstance class for a frame level description of an object.

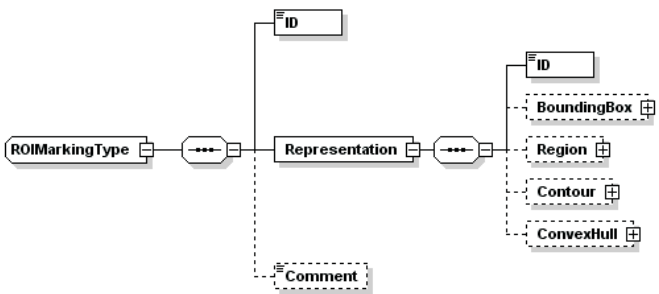


Fig. 8. The ROIMarking class used to spatially localize an object.

the object. The knowledge on the type of the object can be based on preliminary information or image processing.

The spatial location of the object is represented by the ROIMarking class, shown in Figure 8. The simple ROI marking of an object can be BoundingBox (rectangular), Contour (sequence of boundary points), ConvexHull (set of points of the convex hull).

III. Examples of use

In this section we present two examples, where the schema is used to describe the corresponding metadata. The first example refers to fire detection and tracking performed on the thermal video displayed in Figure 9. As for the second example, shown in Figure 10, we demonstrate how human (victim) detection is represented by the schema. The results of the video processing are also shown in the respective figures.

In our system, active contours (snakes) serve as the most precise basis for the image processing background of locating objects. The Contour ROI description class is a natural representation of snakes, as it was chosen in our current examples, as well. The slight difference between the two examples is that for fire detection the aim now is to give a temporal description, while for human detection is to simply decide whether we can locate a victim, or not.

As a result of metadata description, XML files are created, according to the schema provided. Both XML files contain some basic metadata (shown in Figure 11) for the respective thermal video. As we discussed above, the fire related XML contains a FireAppearance class consisting of multiple FireInstances subclasses, as shown in Figure 12. On the other hand, human detection is performed at a single frame level,

Fig. 9. Consecutive frames from a thermal video to detect fire presence.

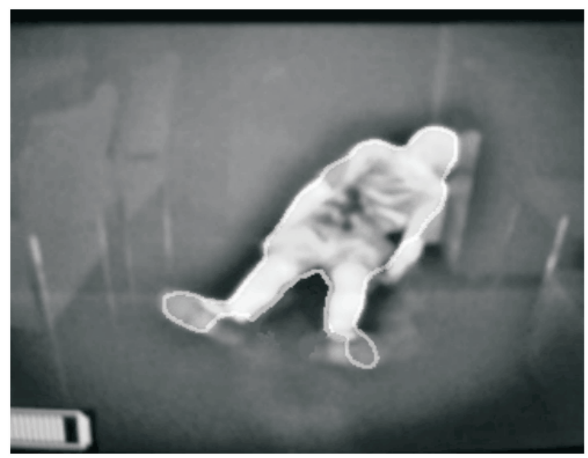


Fig. 10. Result of thermal video analysis for human detection.

```
<ThermalVideoInfo>
        <ID>20020809210000255_000005.avi</ID>
        <Format>avi</Format>
        <Timestamp>
                <BeginDate>09/08/2002</BeginDate>
                <BeginTime>21:00:00</BeginTime>
                <EndDate>09/08/2002</EndDate>
                <EndTime>21:00:00</EndTime>
        </Timestamp>
        <Dimensions>
                <Duration>0</Duration>
                <FrameWidth>317</FrameWidth>
                <FrameHeight>246</FrameHeight>
                <ColorDepth>1</ColorDepth>
        </Dimensions>
</ThermalVideoInfo>
```

Fig. 11. Basic metadata of a thermal video described by the ThermalVideoInfo class.

```
<FireAppearance>
    <ID>1</ID>
    <TimeIn>23:59:59</TimeIn>
    <TimeOut>00:00:01</TimeOut>
    <FireInstances>
        <FireInstance>
        <FireInstance>
        <FireInstance>
        <FireInstance>
        .
        .
        .
        <FireInstance>
        <FireInstance>
        <FireInstance>
        <FireInstance>
    </FireInstances>
    <Comment/>
</FireAppearance>

<FireInstance>
    <ID>1</ID>
    <ROIMarking>
        <ID>1</ID>
        <Representation>
            <ID>1</ID>
            <Contour>
                <ComposingPoints>
                    (213,387)
                    (214,387)
                    (246,374)
                    (246,375)
                    .
                    .
                    .
                    (238,366)
                    (239,367)
                    (240,368)
                    (241,368)
                    (245,374)
                </ComposingPoints>
            </Contour>
        </Representation>
    </ROIMarking>
</FireInstance>
```


Fig. 12. A FireAppearance class containing multiple FireInstance subclasses.


```
<CivilianAppearance>
    <ID>1</ID>
    <TimeIn>21:00:00</TimeIn>
    <TimeOut>21:00:00</TimeOut>
    <CivilianInstances>
        <CivilianInstance>
            <ID>1</ID>
            <ROIMarking>
                <ID>1</ID>
                <Representation>
                    <ID>1</ID>
                    <Contour>
                        <ComposingPoints>
                            (245,70)
                            (245,69)
                            .
                            .
                            .
                            (238,366)
                            (245,70)
                        </ComposingPoints>
                    </Contour>
                </Representation>
            </ROIMarking>
        </CivilianInstance>
    </CivilianInstances>
    <Comment>Reliability=100%</Comment>
</CivilianAppearance>
```


Fig. 13. The corresponding CivilianAppearance class.

thus the related XML contains a single CivilianInstance class, presented in Figure 13.

IV. Future work

Though to provide a full-detailed 3D geometric description of the rescue scenario is outside the scope of the SHARE project, thermal video analysis looks meaningful to register the actual cam view to a pre-defined map. For example, in this way it would be possible to provide evacuation/escaping route planning for the personnel. Moreover, via a proper spatial registration of the thermal camera, it would be possible to make a correspondence to the geological coordinate system (provided by the map application) the operation ontology is based on. In this way, all the image data found on the videos could be mapped into the ontology coordinate system.

In case of a GPS support for the thermal camera, the camera coordinates can be easily found, but the determination of its orientation is still a challenging image processing task.

References

[1] http://www.ist-share.org

[2] L. Vande Velde, S. Chatzinotas, M. Larson, J. Löffler, G. Paliouras, ”Interactive 2D-3D digital maps for the support of emergency forces during rescue operations,” to appear in the Proceedings of the 12th World Congress on Intelligent Transportation Systems, 2005.

[3] J. Löffler, M. Larson: ”Robust keyword spotting and mobile information retrieval for emergency team communication in the EU Project SHARE,” *International Journal for Language Data Processing* 26, pp. 157-168, 2005.

[4] I.S.O, Information technology multimedia content description interface - part 1: Systems, no. ISO/IEC JTC 1/SC 29 N 4153, 2001.

[5] I.S.O, Information technology multimedia content description interface - part 2: Description definition language, no. ISO/IEC JTC 1/SC 29 N 4155, 2001.

[6] I.S.O, Information technology multimedia content description interface - part 3: Visual, no. ISO/IEC JTC 1/SC 29 N 4157, 2001.

[7] I.S.O, Information technology multimedia content description interface - part 4: Audio, no. ISO/IEC JTC 1/SC 29 N 4159, 2001.

[8] I.S.O, Information technology multimedia content description interface - part 5: Multimedia description schemes, no. ISO/IEC JTC 1/SC 29 N 4161, 2001.

[9] J.P. Eakins, ”Retrieval of still images by content,” *Lecture Notes in Computer Science* 1980, Springer-Verlag, Berlin Heidelberg New York, pp. 111-138, 2001.

[10] A. Vakali, M.S. Hacid, A. Elmagarmid, ”Mpeg-7 based description schemes for multi-level video content classification,” *Image and Vision Computing* 22, pp. 367-378, 2004.

Acknowledgement

Research was supported in part by the project SHARE: Mobile Support for Rescue Forces, Integrating Multiple Modes of Interaction, EU FP6 Information Society Technologies, Contract Number FP6-004218.