Management of Mobile Objects Location for Video Content Filtering

Franck Jeveme Panta, Mahmoud Qodseya, André Péninou, Florence Sèdes

Management of Mobile Objects Location for Video Content Filtering

Franck Jeveme Panta
IRIT
Université Paul Sabatier
Toulouse, France
franck.panta@irit.fr

Mahmoud Qodseya
IRIT
Université Paul Sabatier
Toulouse, France
mahmoud.qodseya@irit.fr

André Péninou
IRIT
Université Paul Sabatier
Toulouse, France
andre.peninou@irit.fr

Florence Sèdes
IRIT
Université Paul Sabatier
Toulouse, France
florence.sedes@irit.fr

ABSTRACT

The use of mobile devices and the development of geo-positioning technologies make applications that use location-based services very attractive and useful. These applications are composed of sensors that generate various and heterogeneous spatio-temporal data. Exploiting this spatio-temporal data to support video surveillance systems remains a relevant purpose for video content filtering. Since the data processed in such a context are heterogeneous (indoor and outdoor environment, various position types and reference systems, various data format), interoperability and management of these data remains a problem to be solved.

In this paper, we define an approach that integrates camera location and field of view metadata, mobile objects trajectories, and metadata from video content analysis algorithms (e.g. detection and movement of mobile objects) to address problems related to processing of huge amount of data (big data) generated by CCTV systems. We propose a new generic trajectory based query in order to handle trajectory segment's heterogeneity for both environments (indoor and outdoor). The proposed data model integrate multi-source metadata and enable to handle interoperability issue of data. Our querying mechanism enable to automatically retrieve video segments that could contain relevant information for the CCTV operator (suspects, trajectories, etc.).

We provide an experimental evaluation demonstrating the utility of our approach in a real-world case. Results show that the proposed approach enhances the efficiency of investigators by reducing the search space, as the operator will analyze only the relevant data, therefore he needs less time for video processing (video reviewing).

https://doi.org/10.1145/3282353.3282368

KEYWORDS

Mobiles objects, Spatio-temporal (meta)data, spatial query, trajectory, CCTV system, interoperability, indoor/outdoor location

ACM Reference Format:
Franck Jeveme Panta, Mahmoud Qodseya, André Péninou, and Florence Sèdes. 2018. Management of Mobile Objects Location for Video Content Filtering. In *16th International Conference on Advances in Mobile Computing and Multimedia (MoMM '18), November 19–21, 2018, Yogyakarta, Indonesia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3282353.3282368

1 INTRODUCTION

The number of cameras deployed for the purpose of protection of citizens in urban areas or in important sites grows dramatically. These cameras generate a massive amount of video data to be analyzed in order to solve crimes. Existing methods of automatic video analysis are challenged by three major problems: (i) CCTV cameras are numerous and largely geographically dispatched, which increases the complexity of gathering and managing of the relationships between them (cameras); (ii) Videos are recorded continuously and therefore generate a huge amount of heterogeneous data, making the organization and storage very complex. (iii) Analysis, content annotations and content information retrieval are challenging, time-consuming and unfortunately still often not satisfactory.

In this study, we are focusing on spatio-temporal metadata related to mobile objects, metadata related to surveillance cameras, data from mobile object trajectories, and metadata from video content algorithms in order to address issues related to video retrieval. Mobile objects in our context are persons (sometimes having devices that can be located like smartphones), cars (sometimes with integrated cameras) or any other objects equipped with location sensors (e.g. a mobile robot). Location management of these mobile objects is a step

of our approach (reconstruction of target object trajectories), therefore we are also focusing on location-based services.

Location-Based Services (LBS) are widely used for many applications like surveillance, detection, navigation, etc. Such applications are based on object positioning and provide users with geo-located data via spatio-temporal queries. Applications use in most cases location models based on GPS sensors that are widely embedded (cars, smartphones, etc.) and other location sensors deployed (wifi, RFID, ultra wide bande, etc.). The metadata generated by these sensors are heterogeneous and vary according to the environment (indoor, outdoor): (i) **indoor environment**. Positions are generated by different types of location sensors, expressed according to different references systems and can be either geometric (coordinates relative to a reference system as map of a building) or symbolic (more semantic description related to points of interest, parts of a building, etc.). For instance, location based on wireless gives geometric positions relative to 2 dimensional (2D) coordinate system while cellular network and RFID technologies give symbolic positions (e.g. an area) ; (ii) **outdoor environment**. The objects movement can be constrained for example by a road network (e.g., the cars movement follows the streets of a city) or a transport network (e.g., the buses movement follows predefined lines)[5]. So, positions are geometric or symbolic and can be expressed according to the following reference systems: geodesic system, road network, transport network. Such heterogeneity of spatio-temporal (meta)data related to objects, or sensors cause a problem of interoperability of location-based applications. It becomes difficult to track objects in different environments, or objects located by different sensors, since metadata generated are heterogeneous.

We propose the search and filtering of videos based on the modeling of spatio-temporal metadata from location sensors and mobile objects, and metadata related to video content. For example: based on a trajectory constructed using locations of a person and a time interval, we can retrieve videos that may have filmed a scene of interest, then use video content features obtained via automatic image processing algorithms to filter or rank videos according to their relevance (images processing algorithms are out of the scope of this paper). In this context, interoperability problem has to be tackled at different levels: (i) interoperability issues of spatio-temporal metadata from sensors and mobile objects (described above); (ii) interoperability issues for CCTV systems : CCTV cameras record continuously and therefore generate a huge amount of heterogeneous data. Such a heterogeneity of data is due to the different contextual installation (indoor, outdoor) and camera specificities (manufacturers, data formats etc.); (iii) interoperability of all this multi-source metadata is one of the goals of this paper.

In order to tackle interoperability issues, we propose a generic data model allowing efficient management and interoperability of spatio-temporal metadata from location sensors and mobile objects, and CCTV systems metadata. Metadata modelling takes into account the metadata dictionary described in ISO 22311/IEC 79, standard that aims to facilitate interoperability of CCTV systems. The proposed data model is used by a robust querying mechanism based on metadata for automatic retrieval of relevant video segments from a CCTV system during research of evidence in accident/crime event.

To sum up, the main contributions of this paper are summarized as follows:

- We provide a generic and scalable data model for multi-source metadata management and interoperability.
- We define a new generic trajectory based query in order to handle trajectory segment's heterogeneity for both environments (indoor and outdoor). Trajectory based query is related to some suspect or victim;
- We propose a method to process queries and retrieve video segments that could contain relevant information in the context of an investigation;
- We conduct experimental evaluations demonstrating the usability of our approach in a real-world case.

The rest of the paper is organized as follows. Section 2 reviews the related works; Section 3 presents a description of our approach; Section 4 shows an real-case experimentation that we performed to evaluate our methods for relevant video retrieval and content filtering. Finally, in section 5, we discuss and conclude with suggestions for future research.

2 RELATED WORK

During the last few years, many approaches for video metadata processing have been proposed. These approaches usually aim to describe and facilitate the extraction of relevant metadata related to videos, and use these metadata to simplify the videos management processing. An effective and efficient method for constructing and extracting descriptive metadata for Web videos is presented in [4]. The authors proposed a metadata model that can help the users to search and personalize video data in an efficient ways. Proposed model describe technical metadata, web metadata and descriptive metadata, but this model is only appropriate for the classification of web videos. Metadata modeling for distributed multimedia document management and interoperability of video surveillance systems has been addressed in [13, 16]. The authors propose a metadata format associated with any type of multimedia content. The proposed format of metadata enables multiple systems to exchange data and to use these exchanged data without additional processing. Data exchange is not highly valued in our approach, but it's used in the background.

There are also few works that focus on spatio-temporal metadata and metadata related to the camera (position, fields of view), but the goals are not similar to ours. In [21], the authors present a spatio-temporal extension named STOC (a PL/SQL package) of Oracle Spatial. Moving regions are represented as geometries (SDO GEOMETRY) that move over time. The use case presented is a traffic information management system that answers questions such as: "which vehicles have crossed a given region?". In [8], an approach to annotating images based on camera location and orientation is presented. The originality lies in the fact that between the location and the optical characteristics of the camera (viewing angle), the proposed system (TagPix) computes a distance between the user and different objects located in the visibility area of the camera in order to choose the most relevant tag. The main

similarities with our approach lie in the computation of the field of view and distance seen by the camera without having access to the content. TagPix aims at annotating photos so does not consider the mobility of objects and cameras nor trajectory queries. These approaches also differ in types of spatial query the system can respond to (e.g., position queries [1, 2], range queries [14, 20], nearest neighbor queries [12], predictive queries [11]).

There were few reported researches on a posteriori video analysis for investigative purposes via CCTV metadata. Analytical tools and initiatives for standardization with the aim of assisting the users of large-scale video surveillance systems and police forces for the purposes of a posteriori investigation has been presented in [17]. A framework for fast forensic video event retrieval using geospatial computing is proposed in [9]. The authors described video event analysis and retrieval system using geospatial computing techniques. Their approach consist in transforming and fusing video analysis and tracking results into map coordinates and saving them in spatial database. Then various user-defined video events queries can be directly executed in the geospatial framework through a geobrowser. Similarly, a framework using geographic information to retrieve videos from the web is presented in [10]. The authors describe and use geographic metadata related to video, such as video location, camera field of view, and trajectories for video retrieval. But interesting features such as those related to content are not taken into account. Our approach combines the above-mentioned metadata with others such as context metadata (e. g., indoor/outdoor), video content metadata (e. g., objects detection) extracted through video content analysis algorithms.

The basis of our methodology has been established in [7], which aims to help human video surveillance operators in the research of relevant video sequences by offering them a set of cameras that could have filmed a desired scene.The researches in [7] are applied to outdoor environments. Settings in an indoor environment are different, thus we proposed an appropriate approach in [19]. Recently, we improved the approach for indoor environment, and performed experiments that show the effectiveness of the method in a real-world case [18]. The novelties in the present paper are: continuity of the mobility of objects (indoor-outdoor/outdoor-indoor trajectories), and video content filtering using metadata from automatic image processing algorithms.

3 PROPOSED APPROACH

This section develops our approach for relevant video retrieval and video content filtering with as application the research of evidence during investigations (crimes/incidents). The main goal is to tackle problems (e.g. time-consumption) related to the huge amount of data (big data) generated by CCTV cameras. In addition to this main goal, there is a need to handle the interoperability issues of multi-source data related to our method. An overview of the proposed approach is presented in the following steps:

- *Step 1:* modeling of different types of metadata, namely (i) metadata related to camera (position, field of view); (ii) spatio-temporal metadata from location sensors (timestamp, object ID, location) , and metadata from content analysis algorithms (object detection/movement).
- *Step 2:* definition of a generic trajectory based query, including "approximate" locations of target objects (victims/suspects) and the time interval of an incident or crime.
- *Step 3:* definition of a querying mechanism for relevant video segments retrieval. Modeled metadata (step 1) such as camera positions and field of view , and defined trajectory query (step 2) are required here.
- *Step 4:* video content filtering based on metadata from content analysis algorithms. Inputs of this step are the video segments retrieved in step 3.

3.1 Metadata modeling

Metadata is defined as structured information that describes, locates and facilitates the retrieval, use and management of a resource. One contribution in this paper is the design of a generic data model that enables the management of heterogeneous (meta)data from video surveillance systems.

3.1.1 Metadata related to camera. Camera field of view and geolocation are two key elements of the proposed approach. A camera located at a given position, with a specific orientation and installation can capture a given area. The field of view represents the area of the scene shot by the camera. Figure 1 illustrates a camera field of view. Depending on where the

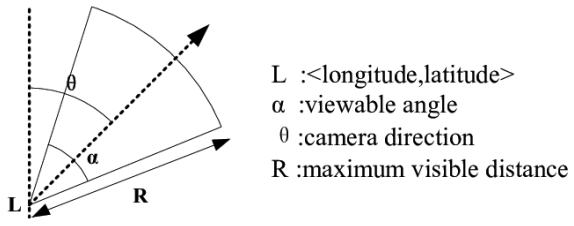

Figure 1: 2D camera field of view (FOV).

camera is deployed, on a fixed place (e.g. in a street, a subway, a room) or on a mobile object (e.g. in a bus), it is called a fixed or mobile camera. The captured scenes can change with possible camera rotations (Pan Tilt Zoom camera). Our data model handles all these requirements. Figure 2 shows the relationships between the camera, the field of view and the location, specifying that a camera can have several positions with variable fields of view at different times. Other data such as the observed scene (e.g. building entrance) and the image quality of the camera are represented in the model. This data model can be considered as the implementation of sensors metadata described in the standard ISO 22311 (which describes the operational requirements for CCTV systems)

3.1.2 Metadata from location sensors. In the specific case of indoor environments, many sensors are installed in order to locate devices attached to objects (people, robots, etc.). The

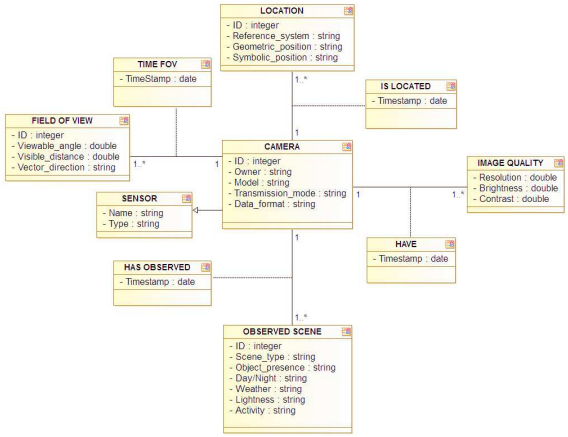

Figure 2: Metadata related to the camera.

(meta)data generated by these sensors can be used to reconstruct object trajectories and can be combined with metadata related to cameras installed for identification purposes. Based on the geolocation sensor used (Wifi, ICCARD, Ultra-wide band, Radio Frequency Identification, etc.)[3, 13] the locations generated are geometric (e.g. 2D or 3D coordinates, latitude/longitude) or symbolic (more semantic description related to points of interest, addresses, etc.) with regards to different reference systems. We have implemented the data model shown in Figure 3, which handles all these heterogeneous data. A Reader connected to a location sensor can record the different locations of the many devices over time.

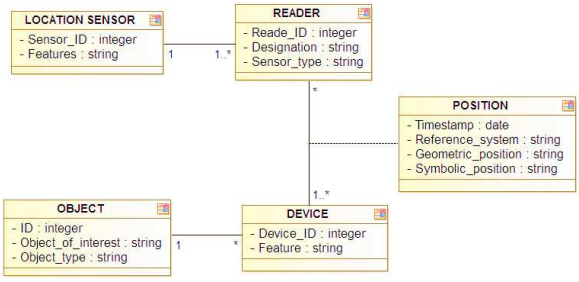

Figure 3: Metadata from location sensors.

3.1.3 Metadata from content analysis algorithms. Metadata from content analysis algorithms are considered in the proposed approach. We are focused on features such as the appearance of objects (people, trains, cars, etc.) in videos. The features are extracted by frame, using YOLO[1]: a real-time object detection algorithm; but any other algorithm may be used instead. A data model for these features is presented in Figure 4. This model handles video elements such as: video segments, frames and events. The relationships between the classes "VIDEO","SEGMENT","FRAME" and "EVENT" are defined as follows: a video contains at least one segment, and the segment contains at least one frame; An event is an action involving content items at a particular place and over a particular time interval (e.g. suspicious hooded persons leaving a building and entering a car parked in an inappropriate location). A video can contain more than one event. The event time interval can be larger than the video segment, so the event is directly related to the video in the model. Object detection is done frame by frame. A frame can contain several objects. The model is able to handle the presence of objects in different frames, also it is able to indicate if there is a movement in the frame comparing with the previous one. Low-level characteristics such as entropy, brightness, contrast, etc. are taken into account in the model and will enable us to filter video content in future work ("negative" filtering).

All metadata described in the previous sections are aggregated in a generic data model shown in Figure 5. It enables to integrate all the previously modeled metadata and then to handle the interoperability of heterogeneous data related to CCTV systems (one of the goals of this paper).

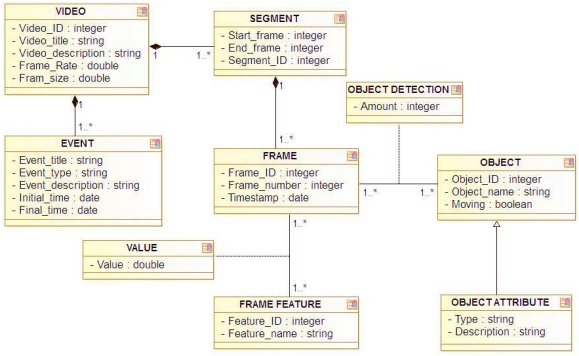

Figure 4: Metadata from content analysis algorithms.

3.2 Generic trajectory based query definition

Relevant information for an investigation may include: the location (or sequence of locations), date and time interval of the incident, and any information that allows identification of the suspect. The spatio-temporal information required for a query during an investigation are defined in [7] as outdoor hybrid trajectory based query. For indoor context, [19] used an indoor hybrid trajectory based query. In this paper we define a new generic trajectory based query in order to handle trajectory segments heterogeneity for both environments (indoor and outdoor); an example is shown in Figure 6. This trajectory based query is constructed by investigators based on facts, testimonies, etc. It is composed of two main parts: a spatial part and a temporal one. The spatial part can contain indoor and/or outdoor sub-parts, each of them consists of a sequence of segments, each segment consisting of a reference system identifier and a sequence of positions (geometric or symbolic) expressed relatively to the corresponding reference system (e.g. Floor8 and ICCARD are the reference systems for indoor environment, WGS84 - geodetic system and RRTLSE -

[1] https://pjreddie.com/darknet/yolo/

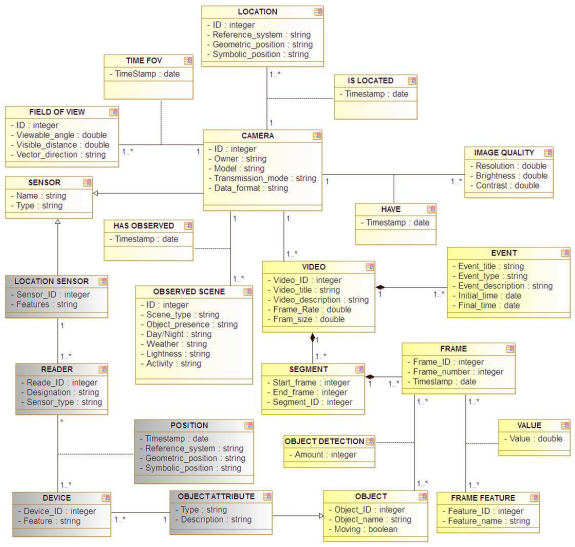

Figure 5: Generic data model.

Toulouse road network, are the reference systems for outdoor environment). The temporal part is an interval of time $[t_1, t_2]$. This hybrid trajectory will constitute the entry point of our querying framework.

```
{
    "query": {
        "spatial": [
            "indoor":[
                {
                "Floor8": [ [85.12254,425.3325],
                           [92.4178,468.58037] ]
                },
                {
                "ICCARD": ["0000110015010005"]
                }
            ]
            "outdoor":[
                {
                "WGS84": [ [43.12254,1.0124],
                           [43.4178,1.02045] ]
                },
                {
                "RRTLSE": [ "Route de Narbonne" : 116,
                            "Route de Narbonne" : 118
                          ]
                }
            ]
        ],
        "temporal":
            {
            "start": "14-02-2018 05:00:00",
            "end": "14-02-2018 05:50:00"
            }
    }
}
```

Spatial part

Temporal part

Figure 6: Example of generic trajectory based query.

3.3 Querying mechanism: trajectory based query intersection with camera fields of view

The main goal is to deliver to investigators the video sequences that may contain interesting images for their investigation (suspects, trajectories, etc.). To do this, it is necessary to search for cameras whose field of view (which can be variable) intersected the trajectories of the query in the given interval $[t_1, t_2]$. We used the *hasSeen* operator defined in [6, 7] as follows: given a spatial trajectory composed of segments $t_r = (u_1, ..., u_n)$ and the time interval $[t_1, t_2]$, $hasSeen(t_r, t_1, t_2)$ returns the set of cameras $c_i (1 \leq i \leq m)$ that captured at least one segment $u_k (1 \leq k \leq n)$ and a video sequence between two moments t^i_{start} and t^i_{end} within the interval $[t_1, t_2]$ ($t_1 \leq t^i_{start} \leq t^i_{end} \leq t_2$).

$$hasSeen : u_1, \ldots, u_n, [t_1, t_2] \implies Z$$

$$Z = \begin{cases} c_1 : t^1_{start} \rightarrow t^1_{end}, \ u_k (1 \leq k \leq n) \\ c_2 : t^2_{start} \rightarrow t^2_{end}, \ u_k (1 \leq k \leq n) \\ \ldots \\ c_m : t^k_{start} \rightarrow t^m_{end}, u_k (1 \leq k \leq n) \end{cases}$$

Two algorithms have been developed for the camera selection process: one for fixed camera and the other for mobile camera. Both algorithms proceeds in two steps: the candidate selection step (purely spatial filtering) and the results refining step (temporal filtering):

- The filtering step applies an algorithm corresponding to a “Region Query" type similar to the one presented in [15]. It allows to select for each segment of trajectory, the cameras located at a distance less than or equal to the maximum visibility distance of all existing cameras in the database. This avoids the evaluation of the spatial intersection (expensive operation) for fields of view of the cameras which are located at a distance that makes it impossible to shot the query segments.
- For the previously selected cameras, the refining step calculates the geometries of the field of view during the time interval of the query, and selects those whose geometries “intersect" the trajectories segments and calculates the time interval $[t_a , t_b]$ for each of them.

The result is a set of triplets: $R = \{r = (c_i, u_k, [t_a, t_b])\}$, $c_i \in SetOfCamera$, $u_k \in tr$, $t_1 \leq t_a$, $t_b \leq t_2$.

3.4 Video content filtering

Once the relevant videos segments are retrieved, some of them may contain no object (people, vehicle, etc.) nor movement. Removing these useless sequences will reduce processing time for investigators.Thus, based on the video content features extracted and modeled in section 3.1.3, we have implemented a content metadata-based query to retrieve video sequences that contain moving objects.

We have defined a $videoOfInterest(VoI)$ operator as follows: given a set of cameras $c_i (1 \leq i \leq m)$ and each related interval of time $[t_a, t_b]$ returned by the *hasSeen* operator ($R = \{r = (c_i, u_k, [t_a, t_b])\}$), $videoOfInterest(\{c_i, [t^i_a, t^i_b]\})$ returns each c_i with a set of interval of time $[t^{i,k}_{start}, t^{i,k}_{end}]$, $t_a \leq t^{i,k}_{start} \leq t^{i,k}_{end} \leq t_b$. Each interval of time $[t^{i,k}_{start}, t^{i,k}_{end}]$ represents a video sequence in which there are objects and

movement.

$$VoF : \{c_i, [t_a^i, t_b^i]\} = \begin{cases} c_1 : t_{start}^{1,1} \rightarrow t_{end}^{1,1}, \ldots, t_{start}^{1,n} \rightarrow t_{end}^{1,n} \\ c_2 : t_{start}^{2,1} \rightarrow t_{end}^{2,1}, \ldots, t_{start}^{2,n} \rightarrow t_{end}^{2,n} \\ \ldots \\ c_m : t_{start}^{m,1} \rightarrow t_{end}^{m,1}, \ldots, t_{start}^{m,n} \rightarrow t_{end}^{m,n} \end{cases}$$

The result is the following set: $V = \{v = (c_i, \{[t_s^{i,k}, t_e^{i,k}]\})\}$, $c_i \in R,\ t_a \leq t_s^{i,k},\ t_e^{i,k} \leq t_b$.

4 EXPERIMENTS AND RESULTS DISCUSSION

4.1 Prototype architecture

Figure 7 illustrates the prototype architecture that we have developed and that implements the data model and the operators described in the previous sections.

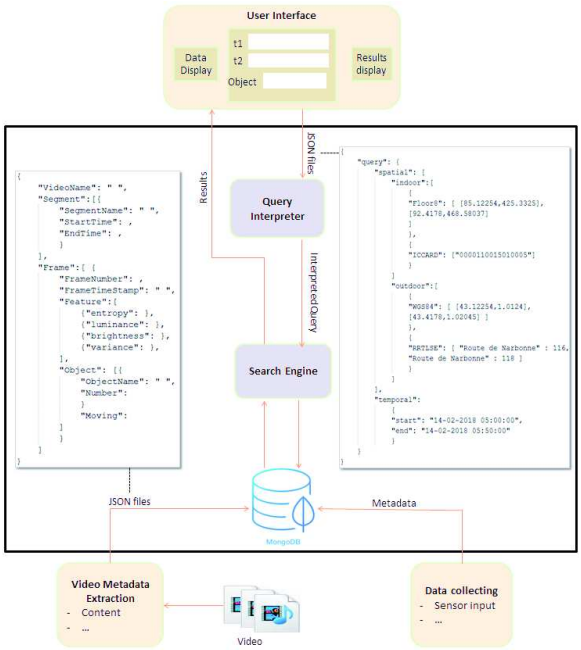

Figure 7: Architecture of the implementation of our method for relevant video segments retrieval.

The main modules of our prototype are:

- Query interpreter: interprets (transforms) the generic trajectory query submitted by the user into a spatio-temporal query;
- Search Engine: implements the research operators (*hasSeen*, *videoOfInterest*) defined in the previous sections;
- Database-MongoDB: contains metadata collections based on the generic proposed model.

The following external modules are used:

- User interface: can be used to build the generic trajectory query. It also enables data and results visualization;
- Video Metadata Extraction: is used to extract data related to the video (object detection, movement, etc.);
- Data Collecting: is used to collect spatio-temporal data and sensor-related data (e.g. data related to the camera field of view).

4.2 Dataset and experiments

In order to have data sets for our research, we have shot realistic scenarios that include scripted actions, like driving a car, walking, entering or leaving a building, or holding an item in hand. In addition to ordinary actions, some suspicious behaviors were present. We have simulated a video surveillance system composed of 24 fixed cameras installed on the university campus. We will focus on a scenario with 10 videos, each video duration is 5 minutes 40 seconds. In this scenario, indoor context is not taken into account, since the cameras were not installed in the building. However, indoor context can be addressed in our approach. Cameras installed outdoors are located on the map as shown on Figure 8.

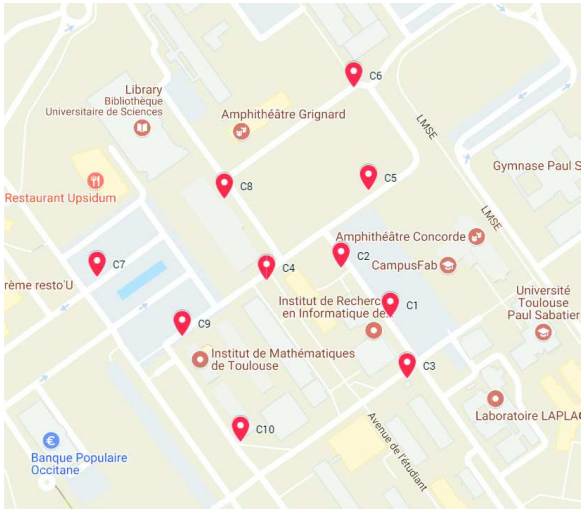

Figure 8: Cameras locations.

The scenario is the following: a suspect car (C) with two men inside (D the driver and P the passenger) arrives and parks in front of the main building. P gets off of C and enters the building. One minute later, a woman complains to D about his bad parking. D quickly goes away (driving C) and stops. Approximately one minute later, P leaves the main building holding a packet, and runs away. P meets D a little further, gets in C, and quickly leave the university campus.

We are looking for the car(C) and the two people (D *and* P) in the car.

4.3 Results and discussion

Based on our approach, we defined a trajectory based query (Figure 9) which is an approximation of the suspect trajectory car and a time interval $[21-07-2017\ 10:51:09;\ 21-07-2017\ 10:56:48]$ of the incident. Interpretation of this query is the trajectory (pink poly-line) displayed on Figure 10

```
{
    "query": {
        "spatial": [
            "outdoor":[
                {
                "WGS84":
                    [
                        [1.4686578962046042, 43.561584215006405],
                        [1.468262674736252, 43.56196841813949],
                        [1.4681406342260743, 43.561842083509354],
                        [1.468187572883835, 43.56193051777826 ],
                        [1.4674085029782873, 43.56285104036154],
                        [1.4668626734437566, 43.562645992408854],
                        [1.4662739285649877, 43.56335053814065],
                        [1.4667835482778173, 43.56360028547715],
                        [1.4678121754350286, 43.56412018419902],
                        [1.4680428454103094, 43.56421055865928],
                        [1.4683620282830816, 43.564505003540916],
                        [1.4689221978933347, 43.564801727085325]
                    ]
                }
            ]
        ],
        "temporal":
            {
            "start": "21-07-2017 10:51:09",
            "end": "21-07-2017 10:56:48"
            }
    }
}
```


Figure 9: Trajectory based Query.

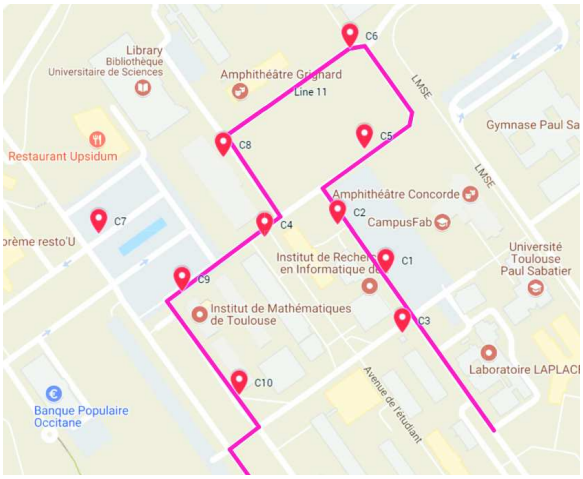

Figure 10: Query interpretation trajectory.

The filtering (spatial) step of our algorithm returned 9 cameras (C1, C2, C3, C4, C5, C6, C8, C9, C10). Since the cameras used do not have a variable field of view, they always filmed in a specific direction (Figure 11), it is not possible to reduce the time intervals returned by the cameras. Therefore, the refining step returned the time interval of the query ($[21-07-2017\ 10:51:09;\ 21-07-2017\ 10:56:48]$) for each camera obtained in the previous step.

As expected by the last step of our approach, we applied content filtering based on metadata from the content analysis algorithms. So that, video segments in which there are no objects or movement have been eliminated. Object detection, including target objects (person and car) at some locations, is shown in (Figure 12). The results are a set of relevant video segments for each camera having filmed the trajectory query, with the aim of helping the investigators in their task. The results are shown in Figure 13. We present for each camera the

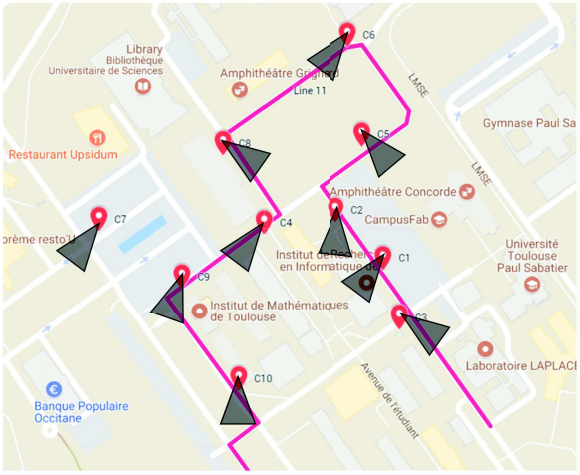

Figure 11: FOV of installed cameras.

video time to watch for the investigation. The different results concern the following three cases: (1)"Manual": video time to watch without filtering (all video); (2)"Proposed Approach": video time to watch after applying our method; (3)"Ground Truth": video time really to be watched because they always contain suspicious persons/objects.The "Ground Truth" has been done manually, by viewing and annotating all the videos of the scenario.

Without filtering (Manual), investigators have to watch all the videos (5 minutes 40 seconds for each video). Consequently, the value of the blue bar is the same (340 seconds) for all cameras in the Figure 13. Let's focus on the cameras "C7", "C3", "C8" and "C9":

- Camera "C7": video time to watch for "Ground Truth" is 0 because the suspects do not appear in the video recorded by this camera. This value is the same for "Proposed Approach" because our camera selection algorithm did not select camera "C7", since its field of view did not intersect the trajectory query. This is an example of the effectiveness of our camera selection method.
- Camera "C3": The results show that the difference between the video time to watch for "manual" (340 seconds) and "Proposed Approach" (250 seconds) is less significant. This can be explained by the continuous presence of moving objects/persons in the video generated by camera "C3". To overcome this issue, it is necessary to embed our model with new video content metadata specific to the targeted objects/persons (ex: car color, T-Shirt color, etc.), which will be the subject of future work.
- Camera "C8" and "C9": filtering based on the objects detection and movement gives a result approximating the ground truth result. That means, there are less objects and/or movement in these videos.

With our approach,the needed time to check (watch) all selected cameras is approximately 19 minutes. On the other

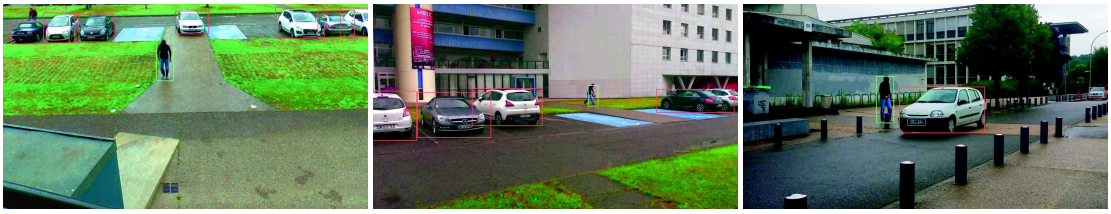

Figure 12: Content filtering based on objects/movement detection

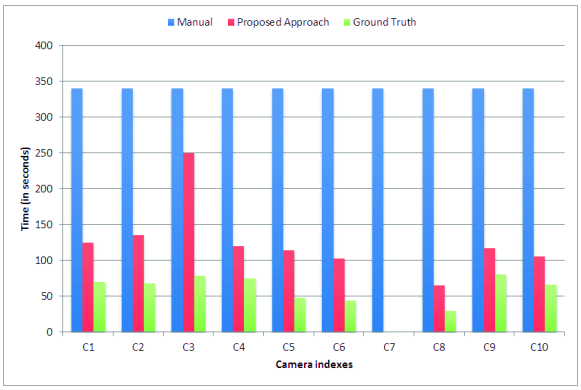

Figure 13: Video time to watch per camera.

hand, the needed time in the manual analysis (without prototype) is approximately 57 minutes. Thus, in this scenario, results show that the developed prototype drastically reduced the search space and processing time for investigators.

For verification purposes, we reviewed all the videos and extracted all the relevant time intervals to the investigation described in the scenario. We found that our approach returns all relevant sequences. So for this experience, our approach enables the reduction of video time to watch without losing any relevant content.

We computed Precision and Recall to evaluate the degree of accuracy and comprehensiveness of our resulting set. We denoted P and R, the precision and recall measures related to the total time of video segments retrieved by our method. In our context, $P < 1$ means that the resulting set contains non-relevant time interval, and $R < 1$ implies that some relevant time interval have been ignored.

We computed P and R as follows:

$$P = \frac{\sum_i |PA(i) \cap GT(i)|}{\sum_i |PA(i)|} = 0.51,$$
$$R = \frac{\sum_i |PA(i) \cap GT(i)|}{\sum_i |GT(i)|} = 1$$

where $PA(i)$ and $GT(i)$ are respectively the total time of video segment retrieved by *Proposed Approach* and by *Ground Truth*, regarding to the camera index i.

In addition to the precision and recall we compute the time ratio (TR) between the total reviewing time of our proposed approach with regards to the total video(s) time as follows:

$$TR = \frac{\sum_i PA(i)}{\sum_i VT(i)} = 0.33$$

where $PA(i)$ is the total time of video segment retrieved by *Proposed Approach* and $VT(i)$ is the original *Video Time*, regarding to the camera index i.

We can observe that our method provided the expected Recall (1), since in the forensic field, no evidence should be lost. The precision shows that the video time to watch for evidence retrieval is approx doubled comparing to the ground truth. However time ratio (TR) shows that we reduce the reviewing time approx by 70% comparing with the total video(s) time.

We are aware that the completeness of the results from a single dataset may not imply the same for the general case. Our approach consists in video retrieval and is applied to the research of a posteriori evidence in videos. In the literature, few works on the videos retrieval are applied to the search for a posteriori evidence, so it is difficult to compare our method to these works. That is why we implemented a "real scenario" to compare our results to the ground truth. Further experiments on a large amount of dataset are needed to evaluate the global performances of the proposed method.

5 CONCLUSION AND FUTURE WORK

In the context of over-expansion of information generated by CCTV systems, reducing research space and time is a frequent and challenging issue in video analysis for investigation purposes. In this paper, we proposed a generic data model for CCTV management and a method to automatically retrieve video segments that could contain relevant information for investigators. The proposed approach consists in using CCTV metadata-based queries to retrieve relevant video segments. Algorithms running in two steps (filtering / refining) have been proposed and enable to obtain results that are pairs composed of the selected camera and the associated time interval. Experimental results show that our approach is efficient in providing and a rapid response with relevant content to CCTV operators.

We are currently working in collaboration with industry (Thales Communications & Security SA) and technical scientific police (PTS) to ensure that our approach is applicable and effective in the real-world case. The next step of our work is to extend this approach, in FILTER2 French ANR project and VICTORIA H2020 European project. A research direction is to enhance video retrieval by ranking video segment relevance.

"Negative" filtering measures can be developed based on metadata or video characteristics in order to improve the information retrieval capability of our approach. One of the hypotheses of our work is that the diversity and the large amount of video content do not allow an exhaustive analysis. Therefore, relevant metadata in the context of CCTV must be used to reduce the search space and implicitly the processing time. A perpective of this work is to use such negative filtering measures to pre-filter irrelevant data for automatic video analyses algorithms, reducing even more the processing-time in real-time investigation cases.

REFERENCES

[1] Imad Afyouni, Cyril Ray, and Claramunt Christophe. 2012. Spatial models for context-aware indoor navigation systems: A survey. *Journal of Spatial Information Science* 1, 4 (May 2012), 85–123. https://doi.org/10.5311/JOSIS.2012.4.73

[2] Haidar AL-Khalidi, David Taniar, and Maytham Safar. 2013. Approximate algorithms for static and continuous range queries in mobile navigation. *Computing* 95, 10 (01 Oct 2013), 949–976. https://doi.org/10.1007/s00607-012-0219-7

[3] Abdulrahman Alarifi, AbdulMalik Al-Salman, Mansour Alsaleh, Ahmad Alnafessah, Suheer Al-Hadhrami, Mai A Al-Ammar, and Hend S Al-Khalifa. 2016. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors* 16, 5 (2016), 707.

[4] Siddu P Algur, Prashant Bhat, and Suraj Jain. 2014. Metadata Construction Model for Web Videos: A Domain Specific Approach. *International Journal of Engineering and Computer Science* 3, 12 (2014).

[5] Khaled Amriki and Pradeep K. Atrey. 2012. Bus surveillance: how many and where cameras should be placed. *Multimedia Tools and Applications* 71 (2012), 1051–1085.

[6] Dana Codreanu, Ana-Maria Manzat, and Florence Sedes. 2013. Mobile objects and sensors within a video surveillance system: Spatio-temporal model and queries. In *International Workshop on Information Management in Mobile Applications-IMMoA 2013*. pp–52.

[7] Dana Codreanu, Andre Peninou, and Florence Sedes. 2015. Video Spatio-Temporal Filtering Based on Cameras and Target Objects Trajectories–Videosurveillance Forensic Framework. In *Availability, Reliability and Security (ARES), 2015 10th International Conference on*. IEEE, 611–617.

[8] Hillol Debnath and Cristian Borcea. 2013. TagPix: Automatic Real-Time Landscape Photo Tagging for Smartphones. *2013 International Conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications* (2013), 176–184.

[9] Hongli Deng, Mun Wai Lee, Asaad Hakeem, Omar Javed, Weihong Yin, Li Yu, Andrew Scanlon, Zeeshan Rasheed, and Niels Haering. 2010. Fast forensic video event retrieval using geospatial computing. In *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*. ACM, 8.

[10] Zhigang Han, Caihui Cui, Yunfeng Kong, Fen Qin, and Pinde Fu. 2016. Video data model and retrieval service framework using geographic information. *Transactions in GIS* 20, 5 (2016), 701–717.

[11] Abdeltawab M. Hendawi and Mohamed F. Mokbel. 2012. Predictive spatio-temporal queries: a comprehensive survey and future directions. In *MobiGIS*.

[12] Sadegh Bafandeh Imandoust and Mohammad Bolandraftar. 2013. Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International Journal of Engineering Research and Applications* 3, 5 (2013), 605–610.

[13] Sébastien Laborie, Ana-Maria Manzat, and Florence Sedes. 2009. Managing and querying efficiently distributed semantic multimedia metadata collections. *IEEE MultiMedia* (2009).

[14] Jongtae Lim, Kyoungsoo Bok, and Jaesoo Yoo. 2016. Processing a Continuous Range Query in Mobile P2P Network Environments. In *Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory (EDB '16)*. ACM, New York, NY, USA, 102–105. https://doi.org/10.1145/3007818.3007834

[15] Xin Lin, Jianliang Xu, and Haibo Hu. 2013. Range-based skyline queries in mobile environments. *IEEE Transactions on Knowledge and Data Engineering* 25, 4 (2013), 835–849.

[16] Ana-Maria Manzat, Romulus Grigoras, and Florence Sèdes. 2010. Towards a user-aware enrichment of multimedia metadata. In *Workshop on Semantic Multimedia Database Technologies*.

[17] Denis Marraud, Benjamin Cépas, Sulzer Jean-Françoi, Christianne Mulat, and Florence Sédes. 2012. A Posteriori Analysis for Investigative Purposes. *Intelligent Video Surveillance Systems* (2012), 33–46.

[18] Franck Jeveme Panta, Geoffrey Roman-Jimenez, and Florence Sèdes. 2018. Modeling metadata of CCTV systems and Indoor Location Sensors for automatic filtering of relevant video content. In *12th International Conference on Research Challenges in Information Science, RCIS 2018, Nantes, France, May 29-31, 2018*. 1–9. https://doi.org/10.1109/RCIS.2018.8406677

[19] Franck Jeveme Panta and Florence Sèdes. 2016. Mobile objects in indoor environment: Trajectories reconstruction. In *Proceedings of the 14th International Conference on Advances in Mobile Computing and Multi Media*. ACM, 332–336.

[20] Guoqing Xiao, Kenli Li, Keqin Li, and Xu Zhou. 2015. Efficient top-(k, l) range query processing for uncertain data based on multicore architectures. *Distributed and Parallel Databases* 33, 3 (2015), 381–413.

[21] Lei Zhao, Peiquan Jin, Lanlan Zhang, Huaishuai Wang, and Sheng Lin. 2011. Developing an Oracle-Based Spatio-Temporal Information Management System. In *DASFAA Workshops*.