CANDELA - Storage, Analysis and Retrieval of Video Content in Distributed Systems

E. Jaspers[1], R. Wijnhoven[1], R. Albers[1], J. Nesvadba[2], J. Lukkien[3], A. Sinitsyn[2], X. Desurmont[4], P. Pietarila[5], J. Palo[6], and R. Truyen[7]

[1] Bosch Security Systems, Eindhoven, The Netherlands
[2] Philips Research, Eindhoven, The Netherlands
[3] Eindhoven Technical University, Eindhoven, The Netherlands
[4] Multitel, Mons, Belgium
[5] VTT, Oulu, Finland
[6] Solid, Oulu, Finland
[7] Philips Medical Systems, Eindhoven, The Netherlands

Abstract. Although many different types of technologies for information systems have evolved over the last decades (such as databases, video systems, the Internet and mobile telecommunication), the integration of these technologies is just in its infancy and has the potential to introduce "intelligent" systems. The CANDELA project, which is part of the European ITEA program, focuses on the integration of video content analysis in combination with networked delivery and storage technologies. To unleash the full potential of such integration, adaptive video-content analysis and retrieval techniques are being explored by developing several pilot applications. . . .

1 Introduction

After the introduction of Digital Video Broadcasting, video enhancement and interactive video enabled the user to interact with the video delivery system. As a next step, there exist a growing desire for content retrieval, applying search queries that are natural for humans. Because the type of queries that are natural for humans depend on the application domain, the analysis and retrieval functionality should be adaptive to the application. For example, in a home video application one might search for a specific genre, whereas in a surveillance application suspicious behavior is searched. This requires understanding of the content and implies video content analysis (VCA) techniques like segmentation into video objects, metadata generation for large databases of video content and the use of search and presentation devices.

Currently, the development of digital video analysis is mainly focused on state-of-the-art video compression (MPEG-4/H.264), describing the video content (MPEG-7), and standardizing a framework to enable interoperability (MPEG-21). All these standards are very much related to the scope of CANDELA, but do not address analysis algorithms or the adaptivity to the application domain. Even though several VCA algorithms have been proposed and a standard for describing the content is available, complete system solutions remain proprietary. For example, how can we detect a pulmonary embolism in the huge amount of pictorial data from a medical CT scanner? How can we detect and identify a shoplifter in a warehouse without manually observing hundreds of security cameras? How can we retrieve information about our favorite holiday place on a mobile device by applying abstract search queries on huge databases?

In addition to these application-specific algorithmic and standardization issues, the architecture of these VCA systems is of importance. First, a VCA system will consist of several components both logically and physically. Video acquisition, coding and storage, computing the mentioned metadata and visualizing the combination of data and metadata are among the functionalities that can be recognized. It is to be expected that these functionalities are partitioned across networked components for reasons of geographical distribution, cost and performance. Coding and displaying have to be adapted to the capabilities of the system at hand. Even detailed functionalities like complicated VCA algorithms can be distributed across low-cost components leading to distributed processing. This is why the networked delivery and the distribution is an integral part of CANDELA architectures.

In this paper, we will describe a complete retrieval system, but focus on its adaptive aspects. Firstly, Section 2 will describe the system architecture with particular emphasis on the impact of the distributed aspects and corresponding adaptivity, e.g. in the resource allocation. Secondly, the paper will elaborate on the application-specific content analysis in Section 3. Thirdly, Section 4 shows how the retrieval application differs per application domain. One of the subsections explains how ontologies can be exploited to introduce human-like reasoning into the system, i.e. provide the relation between human-natural search queries and the content descriptors. Section 5 will finalize with some conclusions.

2 System architecture and adaptivity

Fig. 1a shows the general system architecture, which was already identified by Petrovic and Jonker [1], comprising the integration of content analysis, storage, querying and searching. In Fig. 1b the CANDELA system architecture is depicted which is largely in accordance with Fig. 1a. The figure indicates both the information streams through the system as well as the relevant system components. The system is highly distributed, even more as some of the components themselves have a distributed realization. The distribution brings several concerns about the functionality, which are addressed within the architecture as well. These are described in more detail in subsequent sections.

- *Resource limitations in network and storage* - Communication between the various components of the system is via networks which have varying capabilities. Particularly, the last step in content transport to and from terminals (for acquisition, display and retrieval) is expected to be wireless. This can

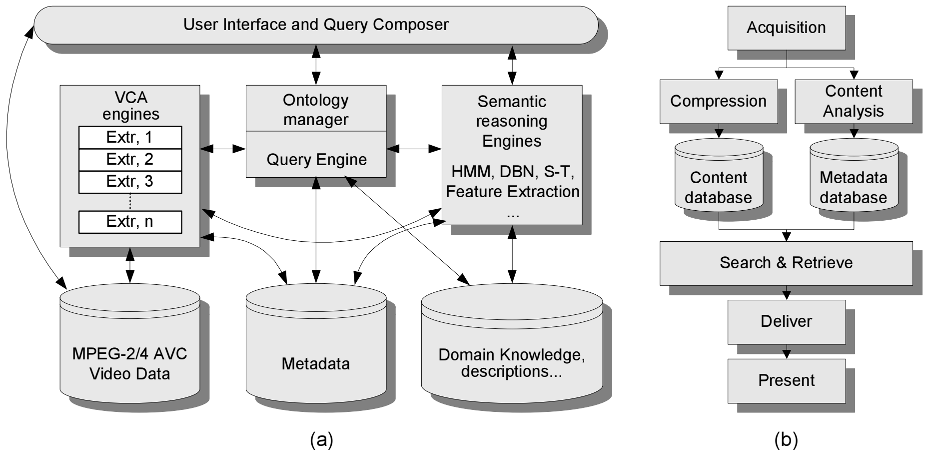

Fig. 1. General system architecture from Petrovic *et al.* (a) and from CANDELA (b).

be either long-distance wireless networks like the cellular system or short-distance networks, like wireless LAN technologies or even Bluetooth. Bandwidth in these networks is expected to be limited and variable; scalable video coding and dedicated transmission protocols are required to guarantee real-time video streaming. Coding is also important for the storage requirements.

– *Content adaptation to terminal and user* - Content adaptation to terminal and user - Content will be retrieved through different types of terminals. In particular, terminal capabilities will vary greatly from powerfull laptops to small-display telephones. In addition, the specific preferences of a user as summarized in her profile are taken into account. This requires content adaptation.
– *Interoperability and reliability in distributed processing* - As mentioned in the introduction, the components in Fig.1b may have a distributed realization. This holds in particular for the database. In addition, the functions of the system (like the VCA algorithms) may be realized through the cooperation of several devices. This requires these devices to be interoperable. In both cases, system reliability is a special concern.

2.1 Resource limitations in network and storage

To reduce the storage requirements and communications requirements, video processing is adopted. To achieve a high fidelity content analysis, the encoding and decoding stages should limit the loss of visual quality. Selection of the video coding format and the transmission method is based on requirements from the application, the terminal characteristics (e.g. camera/display resolution and computational capacity), the network connection, and the available storage capacity. Both non-scalable and scalable coding technologies have been studied. As a result, H.264/AVC and MPEG-4 simple profile (SP) are identified as the most important non-scalable coding standards, whereas AVC scalable video coding (SVC) looks very promising for the scalable solution. The non-scalable coding technologies do not support real adaptive behaviour on the fly. The non-scalable encoder typically is initialized to produce a fixed frame rate and frame size. Only the bit rate can be adjusted during the encoding process. MPEG-4 SP is found to be the best solution for low computational capability terminals. H.264/AVC coding requires much more computations and therefore hardware accelerators are often needed. Software solutions can be used for small frame sizes while the computational load of encoding can be reduced by limiting the coding features (tools) that may be exploited according to the standard (see Fig. 2a). Scalable coding enables true adaptation to the network conditions by changing frame rate, size and quality. The MPEG-4 fine granularity scalability (FGS) includes tools for adaptation, but the nature of the scalable parts in the video stream leads to considerable bandwidth overhead. A more promising solution is offered by the the MPEG-21/SVC, which shows a superior rate-distortion behaviour compared to MPEG-4 FGS (see Fig. 2b). However, the computational requirements might be too high for mobile terminals.

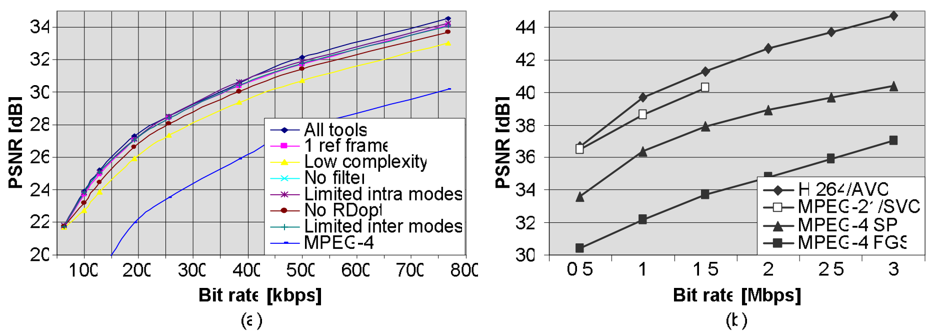

Fig. 2. Rate-distortion curves for different MPEG-4 AVC tools (a) and for scalable vs. non-scalable codecs (b).

2.2 Content adaptation to terminal and user

The vast choice of user terminal types has created a new challenge for information service providers. Typically, the multimedia-oriented web services provide transportation, taking the different capabilities of the user terminals and network connections into account. The user terminals can vary from mobile phones to personal communicators, set-top boxes and personal computers. Three issues are important in order to provide appropriate content and layout to users. Firstly, the system needs to be aware of the user terminal capabilities and configuration. Secondly, it has to know the current state of the network connections. These two are used to select the right scalable coding and transport protocols. Thirdly, it should know the personal details of the user. It is important this type of information is handled automatically, without the need for manual interference. The web page content is automatically adjusted for transmission to a user client, depending on the current state. This minimizes the need for content creation on the author side as the same content can be used on a variety of user terminals. In CANDELA systems, the user interfaces for searching and browsing the video storage are generated dynamically by adapting the content according to the user terminal profile. This feature is provided by using XSLT stylesheets to adapt the MPEG-7/XML content descriptors. This approach gives a possibility to have suitable web content and layouts for numerous type of user terminals.

2.3 Interoperability and reliability of VCA applications

Database architecture

Both metadata and video data have to be stored in the database, requiring the ability to download and stream data from and to the database (DB). From a logical point of view, this DB should be a single entity, however physically it should be distributed. This offers scalability of the system in data volumes and in terms of number of users.

The underlying DB technology is based on the Solid BoostEngine: a relational database server with SmartFlow data distribution capabilities [2]. Essentially, the DB is a small-footprint relational database that provides typical functionality, such as SQL transactions, multi-user capabilities, and automatic data recovery. The server is embeddable in multiple HW/OS environments allowing cross-platform data distribution and supports API standards like ODBC and JDBC.

On top of the server there is a schema-dependent metadata abstraction layer. This layer contains an extendable set of object-oriented interfaces to support metadata in different formats, i.e. UPnP DIDLite and MPEG-7. The underlying object-oriented data model performs the mapping onto the relational model, thereby allowing access to the data both via high-level object-oriented and directly via more flexible relational (SQL) APIs.

The Solid DB management system (DBMS) has two separate storage methods: one for in-memory and another more traditional for on-disk based storage. Both methods can be used simultaneously to manage both *regular data*, which fits to limited-length database table columns, and *blob data*, which comprise large binary chunks of data. Because of this separation, the binary data can be handled more efficiently than alternatively storing them in regular files. Moreover, it is beneficial to keep all data within the database: only one API is needed to access all data; it enables to combine access to all content and metadata in the same queries; all data can be treated transactionally; all data of the system is behind the access control mechanism of the DBMS; and all data can be distributed across the databases of the system and various devices using a unified data-synchronization mechanism.

Distributed VCA processing

For at least one of the applications domains, viz., the home multimedia, there is a natural trend towards distributed processing. There is a rapid growth of devices with storage and VCA capabilities aboard such as voice recognition or elementary video analysis. The concept of ambient intelligence, in which devices share their assets to realize a user-friendly and adaptive environment leads to the requirement of cooperation of all these devices. VCA and other applications are then realized through the cooperation of these local processing and storage capabilities, Therefore, we have developed a distributed platform for real-time VCA processing in the context of home networks. The focus in this design was on the separation between basic signal and video processing functions on the one hand and their composition into an application (e.g., commercial or genre detection, face recognition) on the other hand. The basic functions are available as services on the network in the way a standard like UPnP (Universal Plug 'n Play) defines them.

The platform naturally supports the mobility of services through dynamic and late binding. However, this also means that applications must be able to deal with services that disappear. More generally, the distribution leads to an increased number of points of failure. Particular attention has been paid to reliability through monitoring and signalling faulty behavior and subsequently taking corrective action, for example, by restarting a service on a different device. This is in fact similar to the reliability concerns in the distributed data base.

3 Video content analysis

To provide the metadata for content-based video retrieval, annotation of the video content is required. In CANDELA, we aim at automatically analysing the video content to facilitate this annotation. The metadata is then stored in a multimedia database associated with the video data. The following describes application-specific VCA, addressed within the CANDELA project.

3.1 Medical systems

The majority of colorectal cancers start as benign polyps that take many years to develop. The detection and removal of these polyps drastically reduces the number of cancer related fatalities. A CT scanner can be used to create three-dimensional images of the colon and detect polyps in an early stage. A typical CT volume consists of 500-1000 slices of each 512x512 pixels, resulting in 250-500 MByte of data. Radiology departments are challenged to cope with this data explosion while at the same time improving the quality of diagnosis. Methods were developed to assist visualization, extraction and navigation of the colon structure, but are still not able to perform a fully automated analysis. Related research in this field can be found in [3] to [6].

Three steps can be identified in the automated polyp detection schemes. First, to reduce computation time and limit the search space the colon wall is extracted. Then, for each voxel (3D pixel) in the colon wall, features are calculated that describe the local shape. Curvature of the colon surface is a powerful feature to distinguish the polyps from flat or ridge-like structures. In a next step, polyp-like features are clustered to obtain polyp candidates [7]. In this step the scheme yields a high false positive rate (ten to a few hundred per dataset), whereas the sensitivity is 90% or better. To reduce the false-positives, further content analysis is performed. The resulting features are subsequently used by a statistical classifier that is manually trained with a set of expert annotated polyps [8][9].

3.2 Surveillance

Video content analysis in the developed surveillance application comprises the detection of moving objects to provide automatic recognition of abnormal events (e.g. abandoned luggage at an airport) [10]. Objects can be detected by taking the difference between input frames and the scenes background [11]. By tracking objects over multiple video frames, trajectories and a simple description of events can be created.

One of the critical research areas that is gaining more attention is the validation of such VCA systems [12]. Quite some work on validation has been done, however, standardization is required for viable validation and benchmarking. To stimulate the research in this area, validation sequences from CANDELA have been made publicly available.

Basically, the video content analysis (VCA) modules in the system analyze incoming video frames and segments the images into a static background and moving foreground objects. In addition, the objects are tracked over time, giving each object a unique object Id. Summarizing, for each frame and for each object the VCA modules output the location, the bounding box and an Id. The following subsection explain how higher-level analysis is applied to increase the semantic level of the video descriptions.

Metadata analysis
Figure 3a shows an example of how the above-mentioned VCA outputs the results over the lifetime of an object. This format is not is suitable for storage and retrieval. Therefore, the system contains separate Metadata Analysis modules (MDAs) that convert the frame-based descriptions from the VCA into object-based descriptions to remove redundancy and to provide a data format that is more suitable for retrieval (see Subsection 4.1). Notice that human reasoning is more object-oriented and hence search queries of this type are preferred. The conversion reduces the set of trajectory points by Piecewise Linear Approximation (PLA) [13] and the Piecewise Aggregate Approximation (PAA) [14]. This results in a more compact description of the trajectory. The conversion is applied for each new object that is detected in the scene. The location of the object is defined by the center of the bounding box (see the example in Figure 3a and 3b). When the conversion algorithm decides that the location of the object at the current frame is relevant, it is stored into the database. The relevance is determined by two criteria. Firstly, the maximum distance between two location

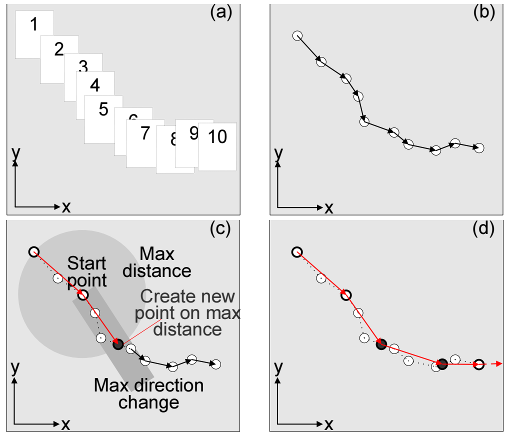

Fig. 3. Bounding box for each video frame (a), location for each video frame (b), filtering locations over time (c), original and filtered locations (d).

points from the conversion and secondly, the maximum deviation of the direction in which the object is moving. Both criteria are visualized in Figure 3c. When the conversion engine decides that any of these two criteria are exceeded, a new trajectory point is generated and stored. To compute the location of this new point, interpolating between the current and previous frame is applied. This algorithm is continued until the object has disappeared. The results from the conversion are shown in Figure 3d.

Perspective transformation

As mentioned before, each object is described by its location and the bounding box, denoted in the number of pixels. However, units of pixels are not desired due to the perspective distortion that is introduced by the 2-D image acquisition of the 3-D world. Note for example that the object size decreases when the object moves further away from the camera. Therefore, in order to use the location and bounding box information for intuitive search queries, the pixels coordinates are transformed to real-world size coordinates. This requires manually or automatic calibration of the camera [15] [16], i.e. the height of the camera and the distance to two points in the scene have to be determined (see Figure 4). Subsequently, perspective transformation can be applied to compute the real-world sizes of detected objects.

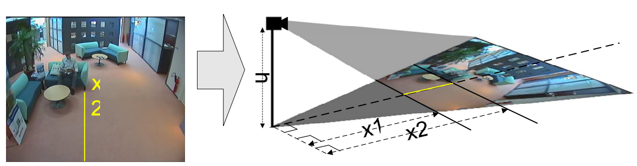

Fig. 4. Pixels to meters, using perspective transform.

Classification

After deriving the real-world sizes, the metadata processing applies straightforward classification by using *a priori* knowledge on the typical sizes of persons, cars, trucks, and smaller objects like rabbits or suitcases. Moreover, the trajectory coordinates are used to determine the speed of the objects. This enables a more accurate classification, since it is unlikely that for example a person walks at a speed of 40 km/h. Hence, the combination of these properties enables some mean of higher level reasoning.

3.3 Home multimedia

Terabytes of storage capacities, millions of MIPS (Million Instructions Per Second) of processing power and gigabits of bandwidth become a reality in consumer homes [17]. The resulting explosive growth of digital multimedia assets in the inter- and broadband connected home creates the need for advanced methods of browsing, navigating, retrieval and content management. Content descriptors [18] enable these advanced methods and extend their capabilities. In CANDELA several VCA solutions have been elaborated and benchmarked, such as a multimodal scene segmentation algorithm [19] [20]. Moreover, face detection, face localization and identification have been researched to enable semantic meaningful high-level search queries on consumer devices [17].

In addition to the VCA algorithms, also advanced distributed-VCA architectures, protocols and technologies based on a network of legacy CE devices have been investigated in CANDELA, utilizing distributed processing power as available in In-Home networks. [21].

4 Retrieval application

Content-based video retrieval is in essence the paradigm of web search engines extended to multimedia data. Basically, video data is annotated with metadata to describe its semantic contents, thereby allowing to search and retrieve videos containing the types of actions, objects, behavior, scenery etc. To enable the system to search for these properties, metadata from the VCA needs to be processed to match the users search queries. Moreover, to enable the system to truly "understand" the content, ontologies are provided. Basically, the ontologies mainly define the relation between descriptors and search request. These ontologies offer the ability to make the queries adaptive to the search request by automatically learning more relations.

4.1 Surveillance

From a user point of view, it is desirable to search through the video database without any additional expert knowledge. For the surveillance application, we have amongst others looked at search queries that are related to the trajectories (motion paths) of the objects [22][23]. Typically, such trajectory data is conveyed by the VCA algorithm on a frame-basis. However, as explained in Subsection 3.2, this format is not suitable for efficient storage in a database (DB) nor for matching the trajectory with a query request. Let us explain by example. At the Graphical User Interface (GUI) side, a user is able to sketch a trajectory in the plane of the camera-view to search for all objects with a similar trajectory pattern. However, a fundamental problem is that all trajectory points of all objects in the database have to be examined to find the correct matches. Therefore, a fast and accurate method is required to reduces this computational burden.

Three different challenges can be distinguished. Firstly, the definition of the *data representation* to model trajectory data for efficient indexing in the database. This was discussed in Subsection 3.2. Secondly, which *database indexing structure* is used that provides fast searching, without scanning the whole database? Thirdly, we need to define the *matching model*: Which metric is going to be used as a distance (quality) measure between trajectories? As a requirement, the chosen data structure should support different types of queries: retrieval of parts (sub-trajectories) of the trajectory data that match with the sketched query; retrieval of objects that crosses a user drawable line and; retrieval of objects in a selected interesting area. The following will describe the indexing and matching challenges separately.

Database indexing structure
After studying several database indexing structures that can store spatial data, R-tree variants [24] seem to fit our requirements best. Many geographical information systems (GIS), already extensively use R-trees for similar applications [25]. Spatial indexing is done by grouping the approximate trajectory representations into sub-trails and representing each of them with a minimum bounding rectangle (MBR). For our application, a special variant of R-trees (R*-tree) is used to store the MBRs in an hierarchical way (see the left part of Figure 5). After the user sketches a trajectory to search for, a window is placed over the

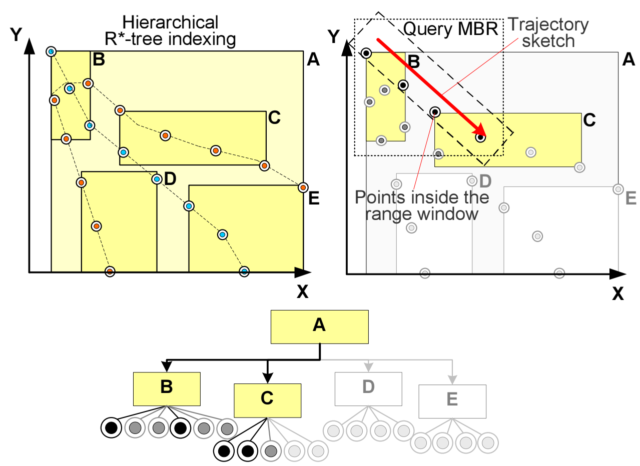

Fig. 5. Hierarchical storage of trajectory data.

drawn line segments to define a distance range. This range defines the area of the trajectory points to search for. After this process, the hierarchical R*-tree filled with trajectory data is traversed for each query MBR (see Figure 5 for an example where two trajectories are present in the R-tree).

Matching model
For matching the sketched line(s) with the trajectories in the database, two different metrics are adopted: the Euclidean point-to-line distance and the directional difference between the queried line and the stored line segments. If the sketch trajectory query contains more than one MBR, the matching is first applied to each MBR. To enable a Google-like ranking of the retrieved objects and to provide preliminary results for fast feedback to the user, the ranked results from each MBR query are combined into one global result set. Therefore, for each two MBR query result sets, a rank-join algorithm is executed that joins the trajectory points from the two sets. Finally, one large result set, ranked in the order of similarity, is left that contains all trajectories that match with the user sketch. The ranking phase in the rank-join algorithm is adaptive to the size of the MBR and its number of processed points [26].

4.2 Ontologies

A common definition of ontology in information technology is the mean to provide a shared and common understanding of a domain that can be communicated between people and heterogeneous, widely-spread application systems [27]. Obviously, such a shared understanding would be extremely helpful for an effective annotation and retrieval of video content in CANDELA as well. This requires the capture of concepts to describe all aspects of all CANDELA users, which are all different. Obviously, such universal ontology is not feasible, but general categories of concepts can be identified, such as: home, work, hobbies, family etc. In addition, one may consider an additional set of descriptions: who is acting; where is the action happening; what type of action; etc.

Within CANDELA a set of top-level ontological classes are predefined that take above-stated descriptors into account (e.g., persons, places, objects, events), as well as several properties (e.g., name, location, sex). The latter are useful for interrelating and describing instances of these classes (e.g., a specific person or place). Subsequently, the classes and properties give a description of the video. The ontology manager is responsible for expanding the terms a user provides in a keyword search. Right before performing the keyword search on the metadata, the query engine consults the ontology manager to add the names of all classes, subclasses, and instances from the user's personal ontology. For example, if Anna supplies the term "child", the ontology manager consults her personal ontology and adds the terms "Sofia" and "daughter" to the query, because these are names of all subclasses and the instances of these subclasses of the class "child". The benefit of this is that the query will now also consider videos that Anna has annotated with the terms "Sofia" or "daughter" but not with "child" explicitly. At a higher stage, the ontology can include rules that will adapt queries to the context of the user. By using information about the place, the time or the personal data of the user, obtained through mobile terminals such as mobile phones, the ontology could make the query more efficient. For example, when

Anna is at work and queries the word "trip", unless otherwise stated, the result of the query will give priority to videos from working trips. More information can be found in [28].

5 Conclusion

Although many technologies exist for content-analysis, storage, retrieval and delivery, the integration of these technologies enable fast and efficient human interaction with distributed systems. In particularly, the high sematic level of interaction and the physically distributed means for storage and delivery are considered as novelties. This paper particularly highlights the adaptive aspects.

Concerning the system architecture, we distinguished the following adaptive aspects. Firstly, we studied scalable coding to adapt video transmission to network conditions. The study showed that conventional scalable coding such as MPEG-4 FGS results in considerable bandwidth overhead, whereas MPEG-21 SVC shows far more better results. However, it pays its price in computational requirements. Secondly, the adopted database technology is related to adaptive retrieval, since malfunctioning components are adaptively recovered. Although the database technology for distributed storage and delivery is generic, dedicated features were added such as: an MPEG-7 interface, the storage of arbitrary-size binary data (blobs), and streaming of multimedia content. Thirdly, we identified adaptive content delivery. User profiles and identification of terminal capabilities enable the system to adapt the content accordingly. Finally, we identified the service-based architecture as an adaptive aspect of the system architecture. Capabilities of the system are automatically adjusted to the availability of the audio-video components. To service a specific functionality, the corresponding components should be available. This offers quality of service, but also provides robust operation in case of malfunctions.

Although we consider the system architecture to be the main contribution to adaptive retrieval, the content analysis algorithms itself are adaptive by nature, since the processing highly depends on the content of the signal. By exploring several application domains we can conclude that general analysis technology is highly challenging. However, by gather several VCA modules in a service-based architecture, the system could automatically adapt to the application domain by requesting application-specific analysis services.

The last adaptive aspect concerns the retrieval itself. Besides the functionality to analyze content and store the results as metadata, a query engine is needed to access the desired content. This part of the system is able to considerably increases intelligence by adaptively updating the ontologies to the search queries.

Although the potential of CANDELA-like has been indicated, the development is still in its infancy. Hence, many research areas have been identified for future exploration.

Acknowledgements

Several authors are denoted on the paper. However, the work was performed by many more people. We would like to thank all members of the CANDELA project.

References

1. M. Petkovic and W. Jonker, *Content-Based Video Retrieval, A Database Perspective*, Multimedia Systems and Applications, Vol. 25. Springer, 2003.
2. www.solidtech.com/library/AutonomicDataManagement.pdf.
3. R.E. van Gelder *et al.*, "Computed tomographic colonography compared with colonoscopy in patients at increased risk for colorectal cancer," *Gastroenterology*, vol. 27, no. 1, pp. 41–48, July 2004.
4. M.Medved *et al.*, "Segmentation and segment connection of obstructed colon," in *Proc. of the SPIE - Medical Imaging 2004*, Febr. 2004, vol. 5370.
5. A.H.de Vries *et al.*, "Feasibility of automatic prone-supine matching in ct-colonography: precision and practical use," in *Proc. of 5th Int. Symp. on Virtual Colonoscopy*, Oct. 2004.
6. J.Florie, "Automatic cleansing for ct colonography using a three material transition model," in *Proc. of 5th Int. Symp. on Virtual Colonoscopy*, Oct. 2004.
7. J. Nappi and H. Yoshida, "Feature-guided analysis for reduction of false positives in cad of polyps for computed tomographic colonography," *Med Phys*, vol. 30, no. 7, pp. 1592–1601, 2003.
8. A. K. Jerebko *et al.*, "Computer-assisted detection of colonic polyps with ct colonography using neural networks and binary classification trees," *Med Phys*, vol. 30, no. 1, pp. 52–60, 2003.
9. S. B. Gokturk *et al.*, "A statistical 3-d pattern processing method for computer-aided detection of polyps in ct colonography," *IEEE Trans Med Imaging*, vol. 20, no. 12, pp. 1251–1260, 2001.
10. P. Merkus *et al.*, "Candela - integrated storage, analysis and distribution og video content for intelligent information system," in *Proc. of the Euro. Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, EWIMT*, Nov. 2004, pp. 25–26.
11. X. Desurmont *et al.*, "Image analysis architectures and techniques for intelligent surveillance systems," *IEE Proc. - Vision, Image and Signal Processing*, 2005.
12. X. Desurmont *et al.*, "Performance evaluation of real-time video content analysis systems in the candela project," in *Proc. of the SPIE - Real-Time Imaging IX*, Jan. 2005.
13. C.B. Shim and J.W. Chang, "Spatiotemporal compression techniques for moving point objects," in *Advances in Database Technology - EDBT 2004: Proceedings of the 9th International Conference on Extending Database Technology.* 2004, pp. 765–782, Springer-Verlag.
14. E.J. Keogh and M.J. Pazzani, "A simple dimensionality reduction technique for fast similarity search in large time series databases," in *PADKK '00: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications.* 2000, pp. 122–133, Springer-Verlag.
15. S. Deng, Y. Yang, and X. Wang, "A calibration method using only one plane for 3d machine vision," in *16th International Symposium on Electronic Imaging, Storage and Retrieval Methods and Applications for Multimedia, SPIE 2004*, Jan. 2004.

16. J.R. Renno, J. Orwell, and G.A. Jones, "Learning surveillance tracking models for the self-calibrated ground plane," in *The 13th British Machine Vision Conference*, Sept. 2002, pp. 607 – 616.
17. J. Nesvadba *et al.*, "Face related features in consumer electronic (ce) device environments," in *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics*, Oct. 2004, pp. 641–648.
18. J. Nesvadba *et al.*, "Comparison of shot boundary detectors," in *Int. Conf. for Multimedia and Expo, ICME*, June 2005, submitted for publication.
19. J. Nesvadba *et al.*, "Low-level cross-media statistical approach for semantic partitioning of audio-visual content in a home multimedia environment," in *Proc. IEEE Int. Workshop on Systems, Signals and Image Processing*, September 2004, pp. 235–238.
20. European ITEA program, *Candela Deliverable D1.1B, State-of-the-art Report, Appendix: AV corpus*, March 2004, www.extra.research.philips.com/euprojects/candela/deliverables/candela-wp1-d1-signoff.pdf.
21. F.de Lange and J. Nesvadba, "A hardware/software framework for the rapid prototyping of multimedia analysis systems," in *Proc. of the Workshop of Image Analysis for Multimedia Interactive Systems*, April 2005, submitted for publication.
22. Y.Yanagisawa, J.Akahani, T.Satoh, "Shape-based similarity query for trajectory of mobile objects," in *Proc. Lecture Notes in Computer-Science on the 4th int. conf. on Mobile Data Management*, 2003, vol. 2574, pp. 63–77.
23. S. Satoh K. Aizawa, Y. Nakamura, "Sketchit: Basketball video retrieval using ball motion similarity," in *Advances in Multimedia Information Processing - PCM 2004: Proceedings of the 5th Pacific Rim Conference on Multimedia*, Tokyo, Japan, October 2004, vol. 3332, p. 256, Springer-Verlag GmbH.
24. Y. Manolopoulos *et al.*, "R-trees have grown everywhere," 2003.
25. R.K.V. Kothuri, S. Ravada, and D. Abugov, "Quadtree and r-tree indexes in oracle spatial: a comparison using gis data," in *SIGMOD '02: Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, New York, NY, USA, 2002, pp. 546–557, ACM Press.
26. I.F. Ilyas, W.G. Aref, and A.K. Elmagarmid, "Joining ranked inputs in practice," in *VLDB*, 2002, pp. 950–961.
27. D. Fensel and M.L. Brodie, *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, Springer- Verlag, 2001.
28. A.Sachinopoulou *et al.*, "Personal video retrieval and browsing for mobile users," in *Proc. of the SPIE - Multimedia on Mobile Devices*, Jan. 2005.