Authentication of MPEG-4 data: risks and solutions

Andreas Lang[a], Stefan Thiemert[b], Enrico Hauer[b], Huajian Liu[b], Fabien A. P. Petitcolas[c]
[a]University Otto-von-Guericke, Magdeburg, Germany, 39016
[b]Fraunhofer IPSI, Darmstadt, Germany, 64293
[c]Microsoft Research, Cambridge, UK, CB3 0FB

ABSTRACT

MPEG-4 is an international object-based standard that provides technological basis for digital television, interactive graphics and multimedia applications. These objects can be natural or synthetic e.g. textures, 3D objects, videos or sounds. In this paper we suggest an integrity approach to protect the content of MPEG-4 data. The essential part of this approach is to embed a robust watermark into each visual, audio and 3D object. The content fragile watermark verifying the integrity of a scene is the sum of all information retrieved from the robust watermarks extracted from the objects of the scene. The information of the fragile watermark will be distributed redundantly to all robust watermarks of the scene. Another essential part of our approach is to embed a part of the scene description or object descriptors as a watermark message into the video or audio streams. The amount of embedded information depends on the payload of the watermarking algorithms. We also analyze the possibility of embedding equivalent information into 3D models, depending on the application.

Keywords: MPEG-4, digital watermarking, authentication, integrity, scene protection

1. INTRODUCTION

The MPEG-4 standard steadily becomes more and more important in multimedia processing and applications. With its extensions MPEG-4 became an international standard that provides technological elements for digital television, interactive graphics and multimedia applications. It enables the production of content based on audio and video objects for authors. These objects can be e.g. textures, 3D objects, videos or sounds. As an improvement to former compression standards MPEG-4 provides two different types in the video and audio part. The first type has a natural aspect, in the audio part music and speech elements and in the video part MPEG-2 defined video streams. The second type of objects is of synthetic nature. Furthermore, MPEG-4 provides elements to manage and protect content owner rights. End users are able to interact with the content within the limits set by the author.

In this article we discuss security aspects and possible risks of MPEG-4 application scenarios. With respect to the new features provided by this standard data authentication, protecting the integrity of the content is a main security aspect. Based on the scenario special situations can be relevant. For audio objects it could be important, that the stereo channels must be in the right order. Furthermore, it must be guaranteed that no part of the object is missing. With respect to video objects the spatial and temporal position respectively the visibility can be essential. If a video object becomes visible earlier or its position is changed then the basic message of the scene could be changed. Different scenarios with possible consequences will be discussed.

One of the new features introduced by MPEG-4 is the Intellectual Property Management and Protection (IPMP) framework. The interface controls the access to MPEG-4 objects. Because MPEG-4 does not describe which mechanisms should be used for content protection digital watermarks can also be a solution to solve that kind of problem. Together with cryptographic methods the security aspect of the used techniques can be increased.

We present a first proposal for a concept to solve data authentication problems. Based on the discussion of the application scenarios we suggest a theoretical solution to protect the content of MPEG-4 scenes and objects. The first part of our concept embeds the information of the scene description into the objects of the scene itself. For that reason we will analyze important information, which should be protected. The essential part of our approach is to protect the integrity of objects and scene. We will embed a robust watermark into each visual and audio as well as 3D object with existing robust and content-fragile approaches. Furthermore, the integrity of the whole scene will be protected by a message loop distributed to all objects. If an object was removed the loop has a leak and the manipulation can be detected. To be able to embed such a mass of information we need watermarking algorithms with a high payload.

This paper is organized as following: Chapter 2 gives a short overview about existing robust and fragile watermarking approaches in the context of MPEG-4. In chapter 3 we will discuss new watermark relevant features and challenges introduced by MPEG-4. Furthermore, some typical applications introduced by the new standard will be described. Possible security problems will be presented and discussed. Chapter 4 presents possible solutions for the risks described in chapter 3. The concept consists of three important elements, which will be envisaged. The paper will be finished with a conclusion.

2. RELATED WORK

Over the last few years, several projects have been concentrating on digital watermarking for MPEG-4. One of the first projects was MIRADOR [1], coordinated by Thomson CSF. It started in March 1998 for a length of 18 months. MIRADOR was initiated to evaluate and upgrade existing watermarking techniques developed within the MPEG-2 framework, to the new issues arising within the MPEG-4 standard. An extensive framework for copyright protection was developed, compliant with the MPEG-4 Intellectual Property Rights (IPR) group specifications, including still pictures, video and audio protection. The main goals of MIRADOR were the delivery of a MPEG-4 watermarking system to MPEG-4 productions and to come up with the first watermarked MPEG-4 production acting as a shop-window. Furthermore, the project had the goal to integrate watermarking technologies into the MPEG-4 codes. The limitation of this project was that it has been concentrating only on copyright watermarking but not on integrity protection. Beside some other institutes worldwide currently the Fraunhofer institutes IPSI and IIS are also active in developing new watermarking technologies for MPEG-4. In the context of the project TRANSMARK [2] a new bit-stream-watermarking system for MPEG-4 audio and video objects will be developed.

The MPEG-4 standard brings out new challenges for digital watermarking. The first new challenge is that the methods must work on the new object-based structure. To watermark a whole scene every object must be watermarked. This requirement increases the complexity of the embedding and retrieval process. Furthermore, MPEG-4 differentiates between natural and synthetic objects. Watermarking approaches, which work on natural objects, are mostly not applicable for synthetic objects. Several methods for watermarking MPEG-4 objects can be found in the literature.

As an example for watermarking of natural video the following two approaches can be used. Wu et al. [3] proposed a multi-resolution object watermarking approach based on the 2D and 3D shape adaptive wavelet transforms. The multi-resolution watermarking method is robust against image/video compression and is computational saving. In [4], Kim et al. used the shape adaptive discrete cosine transforms (SA-DCT), which is applied in MPEG-4, for object-based video watermarking because the object in MPEG-4 video could be in an irregular shape. In the aspects of robustness and quality the SA-DCT method is superior to other padding methods using common DCT. However, all the methods mentioned above concentrate on robust watermarking and did not consider the integrity protection of the object and the whole scene. In our concept for authentication of MPEG-4 data, we combine several robust watermarks together to compose a content-fragile watermark to provide integrity authentication.

Beside methods for watermarking natural video objects there exist some algorithms to embed messages into synthetic video objects. One of the new features introduced by MPEG-4 is the basic animated texture profile. Some watermarking algorithms have been proposed to integrate the watermarking technique with 2D meshes. In [5], Liao proposed a method for 2D meshes watermarking, using a temporal-domain wavelet transform to extract the important feature locations for embedding. The approach is robust to affine transform and random noise attacks but it needs original 2D meshes as help for watermark extraction. Based on the polygonal models, several kinds of watermarking methods are proposed. Ohbuchi [6,7] proposed Triangle Similarity Quadruple (TSQ) and Tetrahedral Volume Ratio embedding (TVR) algorithms, using some methods to organize geometry into embedding primitives and modify these primitives to embed information. Kanai [8] proposed a private watermarking scheme in frequency domain. The approach uses Lazy wavelet to decompose polygonal model into multi-resolution representation and to change the geometry by perturbing the wavelet coefficient vectors for embedding information.

Another new feature is the animation of synthetic faces. To reduce the mass of the transferred data a texture of a face will be mapped on a mesh. Furthermore, the facial animation parameters (FAP) will be transferred. Hartung et al. [9] proposed a method to embed a digital watermark into the MPEG-4 FAPs. It derives from an idea of spread spectrum communication method for still images and videos with a low computational complexity. For the human head a set of 66 FAPs is defined in the MPEG-4 Standard. Small changes of the FAPs are not visible and can be used for the watermarking method. Via correlation the watermark can either be retrieved from the watermark parameters or from a rendered video sequence with estimation of the FAP parameter from two successive frames. The payload is very limited (2.8 bits/sec in a video with a frame rate of 25 frames/sec) and not robust against all attacks.

3. NEW CHALLENGES AND APPLICATIONS

The MPEG-4 standard offers many new features and possibilities for different kinds of applications. The applications can be divided into two classes: one class is for stand-alone applications, like DVD players or home cinema systems; the other class is for online applications, e.g. interactive broadcast applications or multicast streams for cinemas. Beside the new applications MPEG-4 offers new challenges in the case of security and digital watermarking. One of the new features introduced by MPEG-4 is the direct communication between client (in most cases a player or set-top-box) and server. Therefore it is possible to create interactive applications together with features provided by older standards like MPEG-2. With respect to security aspects it is important to protect the communication channels between client and server. Otherwise a conflict regarding the anonymity or confidentiality can occur.

In contrast to MPEG-2 and older standards MPEG-4 is not a frame-based but an object-based framework, which provides a container for each object. The objects can be audio and video data, 3D models or other multimedia data. Furthermore, the standard differentiates between two types of objects - natural and synthetic. Natural means natural speech, music, images or video, which was recorded by a microphone or camera. Computers and other devices create the synthetic objects. Examples of synthetic objects can be speech generators or 3D models. The complete scene depends on the scene graph, which contains detailed information e.g. the time when the objects will be displayed and the movement inside the scene. This information is very important to create the output of a MPEG-4 stream at the receiver's side. If the scene graph is damaged or lost while the scene description was transferred, it is not possible to create the output correctly.

Based on MIRADOR [1] different applications, which use the new features of MPEG-4, will be described in the following sections and possible security risks and their consequences will be discussed.

3.1. Multimedia conferencing

In multimedia conferences new technologies provided by MPEG-4 are applied. They can be used e.g. for business, medical or private applications. Users who want to participate in such a conference are able to use both natural and synthetic audiovisual objects. Natural objects could be in this case a video of the participant himself captured by a camera or his voice. Synthetic objects are e.g. slide shows, desks and fonts. With respect to economical aspects the advantage of multimedia conferences is that the meetings can be held in a virtual environment. Costs and time for business trips can be saved. With respect to technical aspects it is very important that in contrast to normal video conferencing systems (e.g. video telephones) presentations, movies and the videos of the participants can be combined. Furthermore, high-end video conferencing systems and multimedia PCs can communicate in the same environment. An additional advantage is the scalability of objects. If the bandwidth of one of the receivers is not high enough to transmit the video of the participant a synthetic head that is synchronized with his voice can represent his face [10].

A multimedia conference is based on the technology of shared communication space. Every user, who would like to participate in the conference, connects to it in the same way as to a normal web site (e.g. through a URL). The conference server sends MPEG-4 data to the receiver, which represents the shared space. After the receiving process is finished the participant can send his streaming data to the shared space, e.g. audiovisual data representing him, presentations or other related data. Figure 1 shows the framework of a multimedia conference. Participants A, B and C connect to the server and receive the data representing the shared space. After sending the streams to the server the data will be displayed in the virtual environment, in this case a virtual conference room.

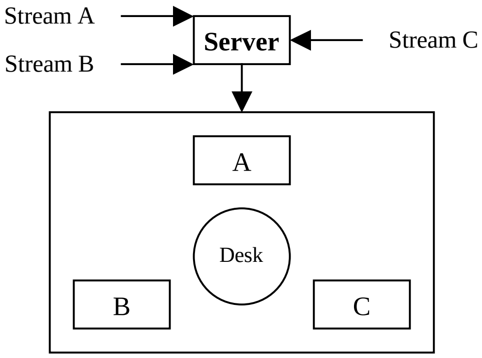

Figure 1: Framework of a multimedia conference

Especially for business and medical applications the following risks can be identified:

- It must be ensured that the address, where the user connects to, is authentic. The connection, which will be used for sending MPEG-4 data by the receiver or by the server, has to be protected. Furthermore, the server should be identified as authentic, e.g. by certificates or a trusted third party. Otherwise it could be possible that important data can be sent to wrong positions and to wrong persons.
- Another problem of streaming data is that it is difficult to authenticate the identification of the data sent by the receiver. A third party could have access to the stream and could send a modified stream to the server. The modified data could influence the tenor of the conference. So every stream should be protected by copyright and integrity information. Furthermore, information about the recording time should be delivered to prevent replay attacks.

The discussed problems can have effect to business contracts or to the corporate image. In medical applications private information about patients could be sent to the false persons or in the worst case a communication problem could have effect to the health of the patient. As can be seen security problems in multimedia conferences can have main effects to different groups of people.

3.2. Interactive broadcast

Beside conventional networks the Internet can be considered as a new instance of digital broadcast networks. In the case of interactive broadcast applications the MPEG-4 receiver can be the traditional home set-top-box (STB) or a conventional multimedia terminal connected to the broadcast network at its input end [10]. In order to receive the broadcast content and to interact with the program the user needs to connect to the broadcast server. The client terminal acquires the necessary scene description information for the tuned channel. The necessary audio and video streams, constituents of the program on the tuned channel, are also acquired and afterwards the user is able to watch and possibly interact with the program. Many applications can be supported, such as interactive home shopping, advance electronic services and interactive entertainment (e.g. sports programs or quiz shows).

The following possible risks can be identified for interactive broadcast applications:

- Pirate broadcast content could be plant by copying or redistributing received audio and video streams.
- The broadcast content can be manipulated. The content would be in this case audio and video objects in some of the delivered scenes so that the audio and video content becomes inconsistent.

As a consequence of the risks discussed above misleading information can arrive the broadcast receiver. Furthermore, it could be possible that illegal copies and unauthorized redistribution occurs. For that reason a fake evidence is needed (e.g. the concept discussed in chapter 4).

3.3. Content-based storage and retrieval

One of the new possibilities offered by MPEG-4 is the direct communication between clients (e.g. players or set-top boxes) and a server. Applications, which use content-based storage and retrieval, are asymmetric communication applications. Asymmetric communication means that the bandwidth of the upload process is lower than the bandwidth of the download process. In the upload process the client sends request information to the server. Afterwards the server sends results in the other direction. The request information contains metadata in the form of descriptions about the spatial and temporal content of the stored objects. With the metadata the content description and the object material itself an entity relationship model can be created, which is the base of a database on the server. In the download process the requested objects or parts of a stream with the content information will be extracted. Based on a content-hierarchical database the requested objects will be sent to the client. In contrast to former technologies the improvement is that not the whole stream must be sent to the player. If only a specific part or a specific object is requested then sending only the requested data can reduce the amount of transferred data.

The following possible risks can be identified:

- If the request information has changed the consequence could be that the server sends other streams to the client than requested. With the increasing relevance of the request information the protection of integrity for metadata should grow in the same way.
- The access point of the database can be changed. Therefore a wrong data record could be used to get the stream. A possible consequence is that another stream than requested will be sent to the client.
- To avoid spoofing attacks the client should be able to authenticate the server. If the request information will be sent to the wrong server the client gets no response or wrong data. Moreover the client cannot be sure that the received data is authentic. Therefore integrity protection mechanisms should be introduced.

3.4. Audio objects in digital cinemas and surveillance

Different applications exist, where the content of the audio objects is essential. If the position of the audio channels has changed or a part of an audio record was removed the main statement of the document can be changed. To make clear the risk we will describe two scenarios where the content of the audio objects respectively the position of the audio channels is important: once for the impression of a movie and once for the hearing of evidence at a court.

The first scenario is not a high-security application but demonstrates the possible consequences of content manipulation for a movie, played in a digital cinema at home or in a public cinema. For such a movie different audio layers in a MPEG-4 scene are available to create a sound impression as real as possible. Examples for such applications are multi-channel systems like THX. For a good impression of a movie it is necessary that the synchronization between audio and video channel is correct. If the position of the audio channels has changed it could happen that e.g. a train goes from the left side to the right at the screen but from the right side to the left in the audio part of the movie. The consequence will be that the audience is confused and the movie becomes a commercial flop. In the second scenario we are using MPEG-4 for surveillance cameras. In several situations the correct position of the stereo channels is important. Recordings of surveillance cameras can be used for hearings of evidence at courts. If there is e.g. a shot while a demonstration the position of the shooter can be identified by the audio channels of the surveillance recording. For that reason the correct position of the audio channels is important, otherwise an innocent person can be convicted. After storage and format conversion it must be ensured that the content can be authenticated. Furthermore, it is important that the content of an audio stream is complete. Otherwise e.g. the sentence “I am not guilty” could turn into “I am guilty” with juridical consequences.

With respect to these two example scenarios it is important to protect the content of the audio objects. We have to protect the content of the different channels and the position of the audio channels. Moreover it must be ensured that the video and audio stream belong together. In the scenarios described above the content protection could avoid commercial and juridical problems.

4. APPROACH TO PROTECT THE INTEGRITY OF MPEG-4 DATA

In this chapter we will discuss a concept to protect the integrity of a scene. For that reason we will introduce three different parts. Figure 2 demonstrates the points where the parts attach.

Part one protects the scene description by embedding important information into the objects (see the dashed rectangle around the scene graph). The second part is able to detect if the content of the scene has changed by introducing a loop between the objects of the scene (see the rectangle around the scene). Part three protects the objects themselves by using existing watermarking approaches for copyright and integrity protection. The different parts work together to protect the integrity of the complete scene by using robust as well as content-fragile watermarking schemes. The complete concept will be described in the following sections.

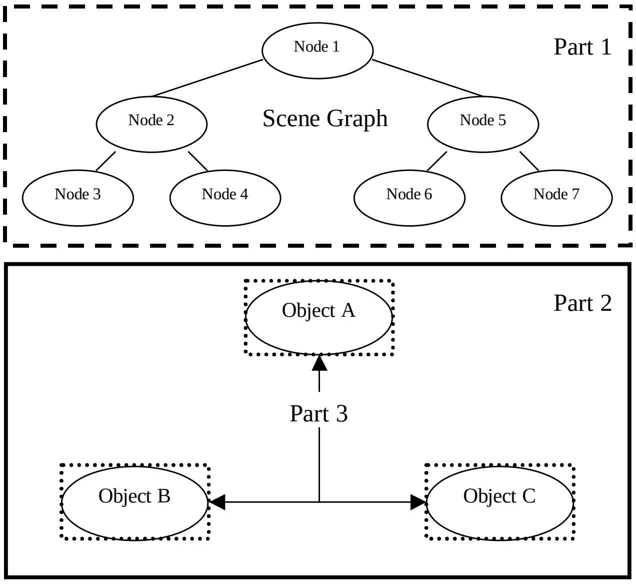

Figure 2: Framework of the content-fragile scheme

4.1. Redundancy between compression layer and scene

Based on [10] the compression layer will be defined as part of the MPEG-4 Systems architecture. It contains the audiovisual objects data as elementary streams, object descriptors, scene description information and up channel information. For the redundancy part of our concept we only use important information of scene description and object descriptors. Up channel information can be protected by other mechanisms like cryptographic methods and audiovisual objects are the elementary streams where the information of the compression layer should be embedded. To know, which information can be used for the first part of our concept, we have to analyze the framework of the scene description and the object descriptors.

The scene description defines the spatial and temporal positions of various objects. Furthermore, it stores their dynamic behavior and interactivity features. The commands of the Binary Format for Scenes (BIFS), which describe the structure of the scene, are based on the Virtual Reality Modeling Language (VRML). The directed non-cyclic scene description graph consists of nodes, fields and events. Each node consists of a list that defines the particular behavior of the object connected to it (e.g. a sphere node has a radius field). There exist several kinds of nodes [11]:

- Grouping node: structuring elements that construct the scene structure
- Children node: for representation of multimedia objects inside the scene
- Bind able node: can be activated at a special point of time
- Interpolator node: children node that represents interpolation data to perform key frame animation

- Sensor node: changes the environment for authoring interactive scenes

With an adoption of the DEFINE mechanism of VRML (DEF) the objects of a scene can be tagged as reference. The DEF names are integers, called node identifiers or IDs. To identify the different types of nodes special IDs will be defined for every node type.

The scene description refers to an object by pointers to object descriptors. Object descriptors (OD) represent the highest level of the object description framework. In the simplest case an OD contains the Object descriptor ID and a URL or an elementary stream descriptor (ES), e.g. an audio or video stream [12]. Furthermore, the OD can contain auxiliary information, e.g. descriptions about the intellectual property management and protection (IPMP).

To introduce the redundancy between scene description and the scene itself we have identified important information of the compression layer, which has to be protected. Because of a limited payload of existing watermark approaches we have chosen the node identifiers to describe the spatial constitution of the scene to be embedded as watermarking message. As temporal feature for the protection of integrity the positions of the interpolator nodes can be used. Interpolator nodes project a sequence of time events onto a sequence of interpolated positions, which describe the animation of the referenced object. As additional message the identifiers of the object descriptors and elementary stream descriptors can be embedded for redundancy.

The main reason for this part of the concept is that the scene description is important for all scenes. Therefore important information will be protected by embedding it as digital watermark in all audiovisual objects of the scene. Requirements for this part are the payload, robustness and real-time ability of the chosen watermarking algorithm. If the scene graph is manipulated or destroyed it should be possible to retrieve the watermarks of the objects and to rebuild several parts. Such an approach was previously proposed for videos of type MPEG-2. In [13] Dittmann et al. describe that it is not only important to protect the audio and video streams but also to introduce mutual content information embedded into the streams. Manipulation or the displacement of one stream can be detected with the mutual information. With respect to MPEG-4 scenes the user can compare the extracted watermarks with the information of the scene graph during the operating time. If there is a difference he has the evidence that the scene graph was manipulated and the content of the scene has been changed. If an attacker wants to change the content of the scene without detection of his manipulation, he needs to change the scene graph and the watermark in all objects. The secret key of the chosen watermarking algorithm should avoid this.

In contrast to cryptographic methods like digital signatures this approach has two advantages. The first advantage is that the data that contains the information about the integrity of the scene description will not be delivered via separate channels but with the objects themselves. So the amount of information, which has to be delivered, does not increase. The other advantage is that in contrast to simple hash functions the concept is able to detect at which position the scene description has changed. With simple hash functions we are only able to identify if the integrity is granted or not.

4.2. Protection of the whole scene

The idea for the second part of our concept is based on the following question: How can we detect if an object was removed from the scene, added to the scene or an object of the scene was replaced by another object? If that happens the main message of the author could be adulterated.

The essential idea of our concept is to introduce a loop between the objects of the scene. If that loop was broken we have the evidence for a manipulation. The loop will be built by messages embedded as watermark into each object. Figure 3 demonstrates the loop of a scene. The embedded message contains three parts: the current ID, the ID of the following object in the loop and the number of all objects inside the scene, represented by the MaxID. The messages can be transferred into a binary vector and embedded as watermark. To increase the security a permutation based on a secret key can be applied. This approach has three advantages:

1. If an object was removed (e.g. object A), the loop is broken. This can be detected by the message extracted from object C in two cases: In the first case the ID of object A is missing and in the second case the total number of all objects, represented by the entry MaxID is incorrect.
2. If an object was added to the scene (e.g. an object D was inserted), the loop has also a leak. The total number of all objects in the scene will be incorrect and the added object contains no watermark. Furthermore, no message, extracted from the other objects, points to the inserted object.

3. If an object was replaced by another object it contains no detectable watermark because of the security key used for embedding and retrieval process. Furthermore, the loop has the same leak as in the first case; the ID of the replaced object, embedded in the object that is in the position before, is missing. Therefore we are able to detect the replaced object respectively the object that was added to the scene.

The requirements for the watermark approaches used to embed the messages into the objects are the maximal robustness, the real-time ability and the capacity to embed the amount of data, which is necessary.

In contrast to cryptographic methods (e.g. linked hashes) this part has the same advantages as the first part of our concept in contrast to digital signatures. The integrity information will be delivered with the stream itself but not with additional channels and the approach can detect the position where the content of the scene has changed. With simple cryptographic hash functions we only have the possibility to detect the manipulation of the content.

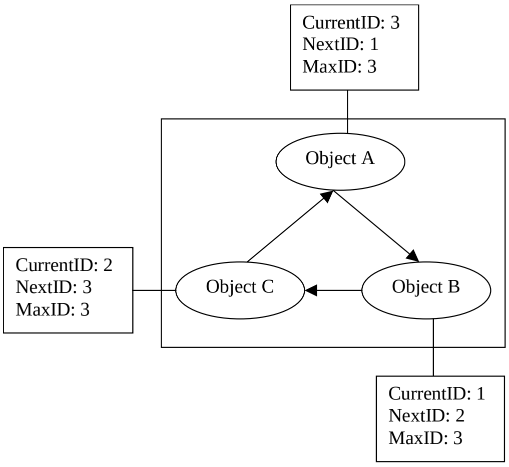

Figure 3: MPEG-4 scene with message loop

4.3. Protection of every object

Beside the protection of the whole scene it can be important to protect the copyright and integrity of every individual object. To solve this problem existing watermarking approaches can be applied.

The main requirements for copyright watermarks are robustness, transparency and capacity. As an example for existing robust MPEG-4 copyright watermarks the following two approaches can be used. Piva et al. [14] proposed the first approach. Every object is extracted from each frame of the video sequence as an individual image. The method applies the discrete wavelet transform (DWT) to the whole extracted object image and the watermark is embedded into the wavelet coefficients belonging to the three detail bands at level 0. After every extracted image is marked they are mixed together to obtain the frame containing the copyright information concerning the objects present. Because the watermark is inserted in raw video material it is robust against format conversion. Even if a video object is transferred from a sequence to another, the copyright data of the single object still can be detected correctly. The approach is also able to detect a watermark in very small regions of an image. For some applications (e.g. searching and indexing) bit-stream watermarking could be useful because in such cases full decompression and re-compression is not necessary and cannot satisfy the real-time requirement. Barni et al. proposed such a bit-stream watermarking scheme for MPEG-4 in [15]. The algorithm embeds a watermark in each video object of a MPEG-4 coded video bit-stream by imposing specific relationships between some predefined pairs of quantized DCT middle frequency coefficients in the luminance blocks of pseudo-randomly selected macro blocks. The quantized coefficients are recovered from the MPEG-4 bit-stream, modified to embed the watermark and then encoded again.

In addition to copyright authentication it is also necessary to protect the integrity for individual objects. For integrity authentication existing MPEG-2 based approaches could be integrated into MPEG-4. The method proposed by Celik et al. [16] is a block-based video authentication watermark, which enables the recovery of lost data. For that reason the watermark payload is divided into two parts: authentication and recovery packet. While the authentication packet contains the information about the integrity the recovery packet contains a highly compressed version of a distant block.

Together with the recovery information stored in a distant block the manipulated block can be reconstructed. This method can be used to authenticate the texture of video objects. For the authentication of still textures the approach proposed in [17] is applicable.

4.4. Usability for MPEG-4 applications

The proposed concept cannot solve all problems presented and described in chapter 3. However for partial problems the watermarking approach is a possible solution, especially for the protection of copyright and integrity. As an example for a partial solution the problem of authentication in multimedia conferences can be specified. The interaction between client and server (protection goals are authentication, integrity and confidentiality) cannot be protected with watermarking mechanisms but with secure Internet protocols and other security mechanisms. However watermarking methods can provide the content authentication. To protect the scene description and the content of the multimedia conferencing scene the approaches of section 4.1 and 4.2 are applicable. For content-based storage and retrieval watermarks can also partially be used. The integrity of the metadata can be protected by cryptographic methods (e.g. digital signature) and the access to the database by secure Internet protocols. The client could authenticate the server by additional information hidden as watermark in the received objects. The annotation watermark could contain the IP address of the server or another message that can identify the correct server. If that watermark cannot be found or the message is different from the expected message the receiver can assume that the delivered content is not authentic. Moreover the delivered and stored objects themselves should be protected by integrity and copyright watermarks as proposed in section 4.3.

There were two applications described in chapter 3 where our concept could solve all the specified problems. In the case of interactive broadcast the copied or redistributed audio and video streams can be protected by copyright and integrity watermarks. If a copy exists the copyright holder can be identified. The origin of an illegal copy can be identified if the set-top box, playing the streams, embeds a robust watermark while operating time. The watermark could contain e.g. the device ID. Furthermore, manipulated broadcast content can be detected by integrity watermarks. In chapter 3.4 two more scenarios were described where integrity watermarks could provide a solution. With respect to the digital cinema application the position of the audio objects representing the different audio channels can be protected by the concept of section 4.1. Furthermore, the content of the complete scene can be protected by the message loop introduced in section 4.2. Beside content authentication copyright identification of the material could be a further application. With respect to the surveillance scenario we have to work with robust watermarking methods. To detect the watermark after storage the watermark has to be robust against format conversions such like MPEG-4 to MPEG-2. Here the usage of robust watermarks, as introduced in section 4.3, is possible. The embedded message contains the position of the two stereo channels. Beside the aspect of protecting the direction of audio channels the content authentication of the video and audio stream is essential.

The analysis shows that the watermarking concept introduced in the last sections is applicable for all MPEG-4 applications discussed in chapter 3.

5. CONCLUSION

In this article we have discussed several possible security risks of MPEG-4. These security risks can be important for different new applications like multimedia conferencing and surveillance. To solve the problems of copyright protection and content authentication we have proposed a watermarking concept that protects the scene description, the content of the scene and the objects inside the scene themselves.

To realize the concepts, introduced in chapter 4, watermarking algorithms with the required robustness and payload have to be developed and evaluated. Furthermore, existing approaches need to be evaluated with respect to the usability for the concept. Given suitable solutions implementations based on the concept should follow. After the implementation process it is necessary to evaluate the limits of the proposed concept. It must be evaluated if the theoretical approach is really applicable in practice.

REFERENCES

1. D. Nicholson, P. Kudumakis, J.-F. Delaigle, “Watermarking in the MPEG-4 context”, *Multimedia Applications, Services and Techniques - ECMAST'99*, 4th European Conference, Madrid, Spain, May 1999, Proceedings. Lecture Notes in Computer Science, Vol. 1629, Springer, 1999, ISBN 3-540-66082-8, pp. 472-492
2. http://www.iis.fraunhofer.de/amm/techinf/ipmp/transmark.html
3. X.Wu, W.Zhu, Z.Xiong, and Y.Zhang, “Object-based multiresolution watermarking of images and video”, *ISCAS’2000*, pp.212-215, Geneva, Switzerland, May 2000
4. G.Y.Kim, J.Lee, and C.S.Won, “Object-based video watermarking”, *ICCE’99*, pp.100-101, June 1999
5. Chun-Yen Liao, Shih-Hsuan Yang, Chin-Yun Hsieh, “Digital Watermarking for MPEG-4 2D Meshes”, *National Computer Symposium (NCS'01)*, Taipei, Taiwan, Dec. 2001
6. R. Ohbuchi, H. Masuda and M. Aono, “Watermarking 3D polygonal models”, *Proc. ACM Multimedia ’97*, 1997
7. R. Ohbuchi, H. Masuda and M. Aono, “Watermarking three dimensional polygonal models through geometry and topological modifications”, *IEEE Journal on Selected Areas in Communication*, **Vol. 16**, No. 4, pp. 551-560, 1998
8. S. Kanai, H. Date and T. Kishinami, “Digital Watermarking for 3D polygons using Multiresolution Wavelet Decomposition”, *Proc. Sixth IFIP WG 5.2 GEO-6*, pp. 296-307, Tokyo, Japan, December 1998
9. F. Hartung, P. Eisert, B. Girod, “Digital Watermarking of MPEG-4 Facial Animation Parameters”, *Computers & Graphics, Special Issue on Data Security in Image Communication*, v. 22, n. 4, Elsevier Science, p. 425 - 435, 1998
10. Olivier Avaro, Alexandros Eleftheriadis, Carsten Herpel, Ganesh Rajan, Liam Ward, “MPEG-4 Systems: Overview”, in: Atul Puri, Tsuhan Chen, *Multimedia Systems, Standards and Networks*, Marcel Dekker Inc., 2000, ISBN 0-8247-9303-X, pp. 331-365
11. Julien Signès, Yuval Fisher, Alexandros Eleftheriadis, “MPEG-4: Scene Representation and Interactivity”, in: Atul Puri, Tsuhan Chen, *Multimedia Systems, Standards and Networks*, Marcel Dekker Inc., 2000, ISBN 0-8247-9303-X, pp. 407-447
12. Carsten Herpel, Alexandros Eleftheriadis, Guido Franceschini, “MPEG-4 Systems: Elementary Stream Management and Delivery”, in: Atul Puri, Tsuhan Chen, *Multimedia Systems, Standards and Networks*, Marcel Dekker Inc., 2000, ISBN 0-8247-9303-X, pp. 367-405
13. Jana Dittman, Martin Steinebach, “Joint watermarking of audio-visual processing”, in: *2001 IEEE Fourth Workshop on Multimedia Signal Processing*, October 3 - 5, Cannes France, IEEE, Piscataway, NJ, USA, pp. 601 - 606, ISBN 0-7803-7025-2, 2001.
14. A. Piva, R. Caldelli, and A. D. Rosa, “A DWT-based object watermarking system for MPEG-4 video streams”, *Proceedings of ICIP’2000*, Vol. III, pp. 5–8, Vancouver, Canada, September 2000
15. M. Barni, F. Bartonlini, V. Cappellini, and N. Checcacci. “Object watermarking for MPEG-4 video streams copyright protection”, *Proceedings of SPIE Vol.3671*, S. Jose’, CA, January 2000
16. M. Celik, G. Sharma, A. M. Tekalp, E. Saber, “Video Authentication with Self Recovery”, *Proceedings of SPIE, Security and Watermarking of Multimedia Contents IV*, 21-24 January 2002, San Jose, pp. 531-541
17. J. Fridrich, “Security of fragile authentication watermarks with localization”, *Proceedings of SPIE, Security and Watermarking of Multimedia Contents IV*, 21-24 January 2002, San Jose, pp. 691-700