Watermarking in the MPEG-4 context

D.Nicholson[1], P.Kudumakis[2], J.-F. Delaigle[3]

[1] THOMSON-CSF COMMUNICATION, 66 rue du Fossé Blanc, 92230 Gennevilliers, France
didier.nicholson@tcc.thomson-csf.com

[2] Central Research Labs, Dawley Road, Hayes, MiddleSex, UB3144, United Kingdom
pkudumakis@crl.co.uk

[3] Laboratoire de Télécommunication et Télédétection, Université Catholique de Louvain, Bâtiment Stévin - Place du Levant, 2, B-1348 Louvain-la-Neuve, Belgium
delaigle@tele.ucl.ac.be

Abstract. This paper presents the constraints involved by MPEG-4 to copyright protection systems based upon watermarking technology. It proposes also an assessment methodology in order to evaluate such systems in terms of robustness to compression and quality.

1 Introduction

MIRADOR European ACTS project was set up to evaluate and upgrade existing watermarking techniques developed within the MPEG-2 framework, to the new issues arising within the MPEG-4 standard. The project's objectives are to:

- integrate MPEG-2 watermarking technologies into MPEG-4 for both video and audio
- assess how these technologies behave in this new environment
- optimise the techniques to the new MPEG-4 constraints
- actively participate in the MPEG-4 IPMP (Intellectual Property Management and Protection) Ad Hoc Working Group to push forward the use of watermarking in MPEG-4

It must be stressed that the project is intended not only to be innovative with the watermarking algorithms, but to work closely with the standards body to ensure that watermarking is integrated and recognised as a key enabling technology for content protection of MPEG-4 objects. Consequently, the project has as an important objective to analyse and actively participate to the MPEG-4 ad hoc working groups so that the technology is accepted and integrated at the level of the MPEG-4 system and that associated hooks for coupling watermarking and monitoring are specified.

The major applications for watermarking in MPEG-4 are :

- monitoring of multimedia object usage (usually to monitor copyright liability)
- fingerprinting (to create an audit trail showing transfer of media objects)

- copy control (to facilitate authorised access to, and copying of, media objects)

The major application segments involving IPR are: interactive television and multimedia to set top boxes, infotainment and streaming media over Internet. This wide spectrum of applications segments, together with the new technologies for encoding objects in MPEG-4, make the issue of watermarking far more complex than with MPEG-2.

As watermarking techniques become more widely known, quite complete lists of basic requirements have been made. Such lists can be found in ACCOPI, TALISMAN [2] for classical coding schemes. However, since MIRADOR wants to deal with new MPEG4 coding schemes in the new interactive multimedia environment, new requirements clearly appear for watermarking techniques. Watermarking should not alter the quality of a work but should resist to a number of manipulations: cuts, scaling, new compression schemes, etc. This document will overview the way how the major issues identified in MIRADOR D1 document, publicly available on MIRADOR web site [1], should be evaluated with existing watermarking technologies, regarding the new MPEG-4 constraints. The evaluation process specification describes the system used for determining the visual quality of the protection mechanism developed into the project.

The items covered are :

- Watermarking robustness evaluation process. The process for the verification of the degree of resistance offered by the watermark present into an object towards its re – encoding.

- Watermarking audibility/visibility evaluation process. The process of verification on a subjective basis the audibility/visibility of a watermarked decoded object into its different representations. Audibility/Visibility constraints will be specified in relation with the concerned applications. Quality of embedding is less constraining for non-professional applications.

2 Rights holder requirements and technical constraints

This Chapter briefly overviews the MPEG-4 underlying concepts to establish the pre-requisites for the IPR requirements analysis.

2.1 The Requirements of Creators and Rights Holders for Associating IPR Identifiers within Digital Content

The control mechanisms required to manage the activities of licensing, monitoring and tracking, and the enforcement of legitimate usage within the distribution chain will rely on the implementation of four key infrastructure tools.

2.2 Persistent Identification

Perhaps the most important of these tools is persistent identification, which should be interpreted as the ability to manage the association of identifiers with digital content [4].

This will achieve the critical link between the one or more component creations that may exist within a piece of digital content and the environment which stores the related descriptive data, current rights holders, license conditions and enforcement mechanisms. The association of the identifier with each creation must be both persistent and resistant. Digital content can and will be modified, whether legitimately or not, and so the persistence of association between identifiers and their creations is a critical requirement. As the imprinting of identifiers into digital content provides the key to associate creations with the control mechanisms required managing intellectual property rights they must also be resistant to attack and removal.

Among the best candidates, we can list

ISAN, ISBN/ISSN/BICI/SICI, ISRC, ISWC-T, ISWC-L, ISMN

2.3 Global Resolution for Identifiers

We see from the above definition of persistent identification that its function is to provide the link between component creations within digital objects and the metadata associated within them. A structure such as the International DOI [4] Foundation provides the necessary level of trusted neutrality to establish routing services for all types of digital content and their associated metadata.

2.4 Information Management Standards

The resolution of identifiers with the storage of associated metadata will present a range of information, in both a numerical and textual format, which will describe amongst other things the information about the creation, its rights holders and licensing terms and conditions. The organisation of this information in a standardised form is a critical requirement if the community it is designed to serve is to benefit from the common interpretation of the information and therefore derive maximum benefit from this level of integration. The Common Information System (CIS) [5] is a clear example where a community of interest has established such an initiative **Trusted Certification Authority services**

The role of the certification authority is probably the least well defined within an architecture for conducting electronic commerce as its activities cannot easily be associated with equivalent services in the world of physical commerce. The conceptual diagram of the IMPRIMATUR Business Model [6], to which we refer in MIRADOR, implies the certification authority provides a validation service to support transactions between the creation provider, the media distributor and the purchaser. On this basis it operates as an independent trusted third party to manage identification certificates which can uniquely identify the purchaser and the media distributor by using a system of public and private keys. To participate in this trading environment, the Media Distributor and the Purchaser must first register themselves with the Certification Authority where they will be assigned a unique name and identification number. These parameters can be watermarked or imprinted into the purchased Creation to enable the Purchaser to prove it was purchased according to predetermined rules. Consequently, any content that violates these rules can, in principle, be more easily detected because the imprinted keys will either be missing or show some evidence that tampering has taken place.

Based on this broad definition the function of the certification authority can be introduced wherever there may be a requirement to enforce variable ‘rules’ or ‘conditions’ to control transactions between different roles**Protection of IPR**

The ability of rights holders to track and monitor the usage of their intellectual property is an essential requirement in both a physical and virtual trading environment. Electronic distribution, however, presents a different set of problems to physical distribution models which will require different solutions. The predicted high volume of transactions to be conducted by consumers combined with frequency of use, ease of digital reproduction, and at low cost, present new challenges to the task of intellectual property rights protection. As a summary, for an efficient protection scheme, the following functionalities should be guaranteed:

- Automated monitoring and tracking of creations
- Prevention of illegal copying
- Tracking object manipulation and modification history (i.e. persistent identification)
- Support transactions between Users, Media Distributors and Rights Holders

2.7 Licensing

The IMPRIMATUR Business Model proposes that the action of issuing a licence constitutes a transaction between the rights holder and one of the other roles that wish to acquire rights in a creation depending upon the type of usage which is required. The most likely roles to seek a licence from the rights holder are the creator, creation provider and the media distributor. Purchasers will also require a licence and these will form part of the terms and conditions of use encompassed by the acquisition from the media distributor.

To effectively control the activity of licensing and to be capable of determining the terms and conditions specified by a licence in a timely manner, licence metadata must be stored in an electronic environment and be accessible to:

- The parties who are signatory to the licence
- The party responsible for enforcing the terms and conditions of the licence

The Rights Holder is the most suitable candidate for managing the responsibility for the control and maintenance of a licence repository for the storage of licence metadata. It is the party closest to the source of the metadata and therefore is able to provide the most authoritative information. The licence Repository will require an interface with the IPR database to establish a link with the identity of the current rights holder. The Rights Holder is also the main source of data supplied to the IPR Database and so it is logical that they also take responsibility for the management of the licence metadata within the repository.

Licence Repositories can provide an archive for storing information about licences, which have been issued to specify the licensing conditions for any types of creation. It may be convenient to manage such a repository in a centralised way on behalf of a number of organisations that have a collective interest in sharing this information. Alternatively, a number of repositories may be managed independently by different rights holders, perhaps covering different geographical areas, but conforming to standards to achieve interoperability between them. Rights societies already maintain their own repositories but these are largely maintained for their exclusive use.. Increasingly, societies are discussing the need to share certain information about their licensing activities and the need to introduce greater interoperability between their licensing systems. So the need for a global repository of licence information will increase.

The purpose of a licence repository is to provide essential knowledge about licences to enable:

1. Identification (a mechanism by which licences for all types of rights can be uniquely identified)
2. Enforcement (by the Certification Authority, according to the parameters which are specified by the licensee in a trust management system)
3. Monitoring and Tracking (to verify the quantity of actual sales against the amount specified in a licence according to its period of validity).

To achieve these objectives it is essential that information about each licence is stored in the licence repository and it is **accessible**. Once a licence can be identified uniquely, routing devices can be implemented (such as DOI) to provide the navigational infrastructure to pinpoint its exact location.

2.8 Enforcement

The licence number can be embedded within a digital object as part of the process of certification in order to provide the enquiry reference identifier for the Monitoring and Tracking Authority to consult the licence repository. It will use the Licence Repository to carry out its primary activity of accurately collating usage information for the Rights Holder. In addition, however, the certification authority can verify that the type of use falls within the terms and conditions of the licence. The licence number provides the key which links to all the other pieces of metadata which are required in order to issue a licence. This will include the identification of the creation, its associated creators, the current rights holder, and other creations and interested parties to which it may be related (i.e. the sound recording of a musical work).

3 New concepts introduced by MPEG-4

The MPEG group initiated the MPEG-4 standardisation process in 1994 (ISO/IEC JTC1 SC29/WG11) with the mandate to standardise algorithms and tools for coding and flexible representation of audio-visual data to meet the challenges of future multimedia applications [7].

MPEG-4 addresses the need for :

- Universal accessibility and robustness in error prone environments
- Interactive functionality, with dynamic objects rather than just static ones
- Coding of natural and synthetic audio and visual material
- Compression efficiency
- Integration of real time and non-real time (stored) information in a single presentation

These goals were to be reached by defining two basic elements:

- A set of coding tools for audio-visual objects capable of providing support to different functionalities such as object-based interactivity and scalability, and error robustness, in addition to efficient compression.
- A syntactic description of coded audio-visual objects, providing a formal method for describing the coded representation of these objects and the methods used to code them.

The coding tools have been defined in such a way that users will have the opportunity to assemble the standard MPEG-4 tools to satisfy specific user requirements, some configurations of which are expected to be standardised. The syntactic description will be used to convey to a decoder the choice of tools made by the encoder.

Audio-visual scenes are composed of several Audio Visual Objects (AVOs), organised in a hierarchical fashion. At the leaves of the hierarchy, we find primitive AVOs, such as :

- A 2-dimensional fixed background,
- The picture of a talking person (without the background)
- The voice associated with that person;

MPEG standardises a number of such primitive AVOs, capable of representing both natural and synthetic content types, which can be either 2- or 3-dimensional. In addition to the AVOs mentioned MPEG-4 defines the coded representation of objects such as:

- Text and graphics;
- Talking heads and associated text to be used at the receiver's end to synthesise the speech and animate the head;
- Animated human bodies.

In their coded form, these objects are represented as efficiently as possible. This means that the bits used for coding these objects are no more than necessary to support the desired functionalities. Examples of such functionalities are error robustness, allowing extraction and editing of an object, or having an object available in a scaleable form. It is important to note that in their coded form, objects (aural or visual) can be represented independently of their surroundings or background.

3.1 Scene description

In addition to providing support for coding individual objects, MPEG-4 also provides facilities to compose a set of such objects into a scene. The necessary composition information forms the scene description, which is coded and transmitted together with the Audio Visual Objects (AVOs).

In order to facilitate the development of authoring, manipulation and interaction tools, scene descriptions are coded independently from streams related to primitive AV objects. Special care is devoted to the identification of the parameters belonging to the scene description. This is done by differentiating the parameters that are used to improve the coding efficiency of an object (e.g., motion vectors in video coding algorithms), and the ones that are used as modifiers of an object (e.g., the position of the object in the scene).

Since MPEG-4 should allow the modification of this latter set of parameters without having to decode the primitive AVOs themselves, these parameters are placed in the scene description and not in primitive AV objects.

4 New technical constraints introduced by MPEG4 coding

The robustness of watermarking within Visual objects is extremely sensitive to all manipulations that may be applied to the objects. This was already true for JPEG and MPEG-2 but becomes even more relevant and more complex with MPEG-4 capabilities.

4.1 Visual objects

4.1.1 New coding tools

MPEG4 is specifying new coding tools for natural and synthetic visual objects. The robustness of watermarking technologies to the loss of information introduced by lossy compression has to be guaranteed. Watermarking technologies are efficient protection tools if and only if they resist to a compression ratio such that the quality of the compressed image and thus its commercial value is low. So, tests have to be conducted with the new MPEG4 coding tools, even if they are close to the ones specified in MPEG2. The same tests have to be performed for wavelet compression, EZW specified for still images. The tests already achieved by MIRADOR are presented in section 7

4.1.2 Objects with arbitrary shapes and sizes

After MPEG4 encoding, two different entities carry information: the shape of the object and its texture. The watermarking may be embedded into both entities. In the case of the shape watermarking, the edge BABs or their transparency could be slightly modified. However, the possibility of objects merging (for instance when transcoding takes places) and of shape modification by the user (for instance the deformation and cropping) should be borne in mind.

Although watermarking the texture enables the use of more classical methods, the problem of synchronisation when attempting to retrieve the watermark is crucial, since a reference is always necessary. Difficulties start when the shape of the object has been modified, for instance when compressed with loss. Besides, practical difficulties also occur because of the arbitrary shape, since most algorithms are designed to work on rectangular images. In addition to this, it is obvious that a minimal size is required for the object to carry watermark information.

4.1.3 Geometric transforms

Geometric transforms concerning sprites objects are defined by the MPEG-4 standardisation:

- Translation (which may involve cropping of the object)
- Rotation
- Rescaling
- Generic affine transform

Moreover, taking into account the interactivity possibilities offered to the user, all kind of objects may be affected by geometric transforms. Thus, the watermarking techniques have to deal with this constraint. On the one hand, the watermark has to resist the deterioration introduced by the transformations. On the other hand, the watermark has to be associated with a so-called “re-synchronisation mechanism”, which allow to look for the watermark in the good location. Finally, the watermark must be sufficiently redundant so as not to suffer from the loss of information introduced by cropping. It is important to mention that cropping is often associated geometric transformations.

4.1.4 Multiscale objects

Concerning scalability, the main issue is the degree of protection that must be provided by the watermark. The available bandwidth at each level, as well as the picture quality (directly connected with its commercial interest), must be taken into account.

Temporal scalability is not a problem with today's methods, since the message is repeated in each image; but it may be quite disturbing when the small size of the objects has to be compensated for by disseminating the message on several pictures. An interesting approach may consist of using temporal wavelet transforms in order to spread the message over several components (DC components corresponding to the static parts of the image as well as AC components connected with the moving parts of the sequence).

Regarding the spatial scalability, the main question is: what actually has to be protected, taking into account the fact that it will be very difficult to write a message in the lowest layer without visibility problems ? It could be interesting to have a hierarchical message with a hierarchy on the message quality or its content. For instance, the message could be split into sub-messages corresponding to the Authors' Society reference, to the author, to the work, etc., and each message hidden in a different layer. This would enable a partial but pertinent decoding of the message at the lowest level of scalability. The watermarking method should also be compliant with the wavelet.

4.1.5 Transcoding problems

The problems involved do not seem to be overwhelming as long as we stay in the original MPEG-4 bitstream (accepting some minor modifications). The transcoding of a MPEG-4 bitstream into another MPEG-4 or MPEG-2 bitstream seems to be much more critical. It can be regarded as a multiple watermarking problem, with different messages coming from the merged objects. When reading the watermarks, synchronisation might therefor be much more difficult. Due to the difference in size of the objects, the message carried by a small one may be hidden by another message.

5 Watermark usage in MPEG4 environment

5.1 IPMP

Versions of MPEG prior to MPEG-4 have not included the mechanisms to allow Rights Holders to adequately protect their works. With this in mind, as a part of the MPEG-4 standardisation process, a separate committee was set up to discuss how best to provide these facilities. This became known as the IPMP Group (Intellectual Property Management and Protection Group).

Initially, in the MPEG-4 IPMP, it was thought that it may be possible to include facilities such as encryption and watermarking within the MPEG-4 Standard. However, because of the need to finalise the standardisation process quickly, the large range of potential MPEG-4 applications combined with their widely differing IPMP requirements, and finally the legal implications of recommending particular techniques which could later be proven inadequate, it was decided not to take this major step. Nevertheless, in order to meet the needs of the creative industries and to encourage them to use MPEG-4, it was considered necessary to provide a mechanism whereby IP could be protected if required in any given application. MPEG-4 IPMP standardises a generic interface to (possibly private) IPMP tools. This interface is referred to as the IPMP interface.

The main issue is that MPEG-4 Version 1 was offering no IP protection and could have presented a “back-door” route to persons attempting to “attack” a protected work. This situation was resulting from the general MPEG versioning philosophy, which maintains that backward compatibility with Version 1 is essential. This implies that Version 2 works should be playable on Version 1 players, albeit without any “advanced features” introduced by Version 2.

In order to cope with these requirements and not to offer a “back-door” route to persons attempting to “attack” a protected work, it has been decided to include an interface to non-MPEG-4 Standard IP protection systems. Such systems will almost certainly be different for different applications but, as they are not part of the Standard, this does not present a problem. In fact this is seen as an advantage because the IPMP system can be optimised for individual applications.

The Dublin meeting saw the introduction of the IPMP “hooks” (control points) architecture. This represented a significant step in the development of IPMP infrastructure within MPEG. The group consensus after the May 1998 New York IPMP *ad hoc* meeting, and the subsequent evolution of the New York proposal at Dublin led the Convenor and others to suggest that IPMP should be considered for inclusion in MPEG-4 version 1.

MPEG-4 provides the facility to establish an IPI Data Set (Intellectual Property Identification Data Set). This provides no protection as such, but does allow a complete set of IP information to be included within the MPEG-4 bit stream. For further information please refer to WG11/N1918. The IPI Data Set can be used by IPMP systems as input to the management and protection process. For example, this can be used to generate audit trails that track content use. The IPI Data Set includes:

Type of Content	e.g. audio visual, book, musical work etc.
Type of Content Identifier	e.g. ISAN, ISBN, ISRC etc.
Content Identifier Code	i.e. a unique number identifying the content.
Supplementary Data Items	i.e. other data as required (not defined).

The aims of the IP Protection & Management, previously planned to be included in the Version2, were reviewed with a view to coping with the following issues:

- Persistent protection of IPI Data Sets.
- Management of intellectual property, conditional access permissions, transactions, user authentication and addressibility.
- Audit trails and modification history.
- Integrity and authenticity of intellectual property information, modification history information and payload.
- Real time issues and synchronisation.
- Interfaces between MPEG-4 and external systems (e.g. CORBA, DCOM, COM etc.).
- External security systems, watermarking and cryptography.

In order to provide appropriate solutions for the wide range of applications the MPEG-4 IPMP Group have proposed a modular IPMP System. A clear point of separation is defined between non-normative IPMP systems and the normative part of MPEG-4. This point of separation is the IPMP interface, on one side, being part of the Standard and on this other side, specific to an application and not part of the Standard. It should be emphasised that the interface is common to all applications and is part of the MPEG-4 Standard. This approach allows the design of application specific IPMP-S's (IPMP Systems).

While MPEG-4 does not standardise IPMP systems, it standardises the MPEG-4 IPMP interface. This interface was designed to be a simple extension of basic MPEG-4 systems constructs. It consists of *IPMP-Descriptors* (IPMP-Ds) and *IPMP-Elementary Streams* (IPMP-ES). IPMP Elementary Streams are like any other MPEG-4 elementary stream and IPMP Descriptors are extensions to MPEG-4 object descriptors. The syntax of these constructs is described in great detail in ISO/IEC 14496-1.

IPMP-Ds and IPMP-ESs provide a communication mechanism between IPMP systems and the MPEG-4 terminal. Certain applications may require multiple IPMP systems. When MPEG-4 objects require management and protection, they have IPMP-Ds associated with them. These IPMP-Ds indicate which IPMP systems are to be used and provide information to these systems about how to manage and protect the content.

Communication between the IPMP-S and the MPEG-4 unit is by means of IPMP-D's (IPMP-Descriptors) which may arrive via the MPEG-4 bit stream or through a side channel connected to the IPMP-S (such as a smart card). In the case of communications using the interface the IPMP-D's will be contained in either an IPMP ES (Elementary Stream) or in other ES's. The IPMP-S will communicate with the other ES handlers whenever objects in those streams have IPMP-D's associated with them. The IPMP-S is responsible for managing access to objects in protected streams.

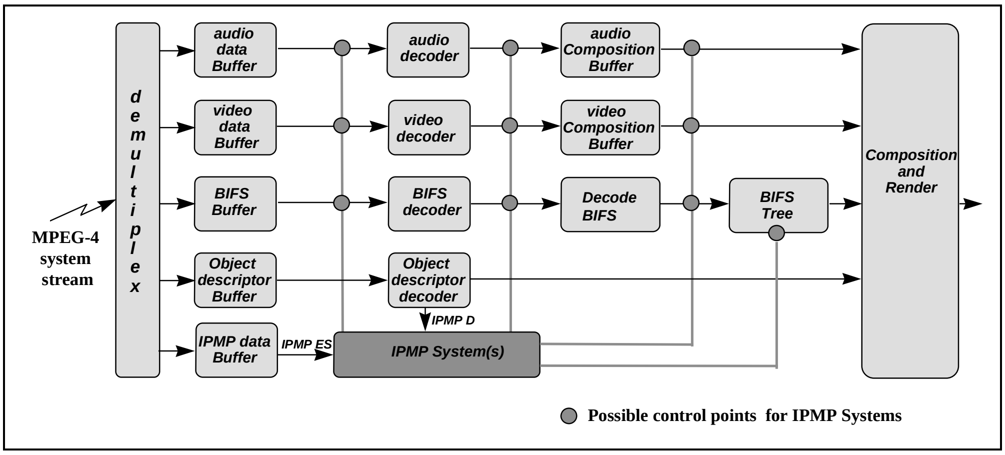


Figure 1: MPEG-4 IPMP framework

Watermarking is an emerging technology that aims at protecting IPR of diverse kinds of contents, such as sound, Still images, video, text or even 3D objects. The principle of watermarking is to embed invisibly (hide) IPR data directly into the Content. In other words, the Content is imperceptibly modified in order to carry IPR data. The embedding process is generally parameterised by a secret key. At the other end, the retrieval process extracts IPR data from watermarked contents with the use of the secret key. Without this key, it must be impossible to retrieve and thus remove IPR data. In certain cases, the above mentioned IPR data can be only a few bits and the retrieval process only checks the presence of a watermark. Watermarks can be applied before (Classical watermarks) or after MPEG4 encoding (bit-stream watermarks).

There are mainly three scenarios in which the watermark can be used:

- *Proprietary watermark search.* Typical uses of watermarks consist in monitoring contents on the Internet or on broadcast networks. The idea is to look for a proprietary watermark in order to detect illegal usage of a particular Content. This kind of application does not absolutely require any interaction with the MPEG4 playback, since it can be achieved after decompression and must be possible after decompression or re-encoding in any other formats. The major goal of this watermarking based monitoring is to deal with the leaks of encryption based security systems, which are no more efficient once the protected contents have been decrypted, copied, possibly re-encoded and eventually redistributed.

- *Copy Control.* The copy Control problem into MPEG4 player is primarily covered by the descrambling actions (IPMP). The idea is to use the watermark retrieval as an additional tool (additional to the scrambling) for copy Control. This can be done verifying whether the watermark read out from the object and the IPR data are coherent. This implies the existence of an IPR data flow providing copyright ownership of the objects being decoded. So the copy Control mechanism of the player should take advantage of two tools:

    1. Mutual identification of client and server: if the identification fails the copy Control acts on the descrambler to avoid the copy.

2. Coherence check between the IPR data conveyed with the stream and the watermark: if the check fails the copy Control mechanism acts on the player to avoid the copy or the rendering.

- *Fingerprinting.* Fingerprinting consists in associating the identification of the copying consumer device or the identity of the consumer with the copied Content, via a watermark insertion. Embedding fingerprints in consumer devices is a difficult challenge that absolutely needs interaction with MPEG4 playback platform.

  The watermarked object must replace the decoded object in the rendering chain into the player.

  To insert the watermark conveying the identities of the consumer are needed the following information:

  - identity of the user
  - key

6 Proposed Evaluation protocol

6.1 MPEG-4 selected applications for watermarking

MIRADOR evaluation methodology is the description of the way how watermarking experiments have been conducted in the context of MPEG-4. The evaluation process has two different parts, the robustness evaluation and the invisibility evaluation named quality evaluation. In order to establish an assessment protocol, it is necessary to overview what kind of evaluation has been made previously in other projects and domains. TALISMAN ACTS project has set up the first methodology to evaluate watermarking algorithms, and in order to analyse the robustness of the watermarks, it is totally possible to use the robustness evaluation process of TALISMAN. Concerning the quality evaluation, it is important, because the usage environment is different of the TALISMAN environment, to analyse what are the quality assessment used to conduct experiments regarding the quality of compression algorithms.

The next table of this document regroups the different possible applications for MPEG-4 and their suitability for the use of watermarking. Two criteria have been considered at the same level, the interest of using watermarking in those applications and the requirements for watermarking processes when it is of interest. When there was a risk that these requirements were unreachable, the application was rejected.

Application	Accept/ Reject	Reason
Real Time Communications	R	Watermarking not key to application
Surveillance	R	Watermarking not key to application.
Mobile Multimedia	R	1. Watermarking probably not key to the application. 2. Too wide to deduce requirements
Infotainment	A	Watermarking of great interest.
Content based Storage and Retrieval	R	Public watermarking of possible interest but beyond MIRADOR scope
Streaming Video on the Internet/Intranet	A(still)/ A(video)	Watermarking essential
Broadcast	A	Watermarking essential
Digital Television Set-Top Box	A	As for Broadcasting
Studio and Television Post-Production	R	Watermarking of great interest but difficult and closed domain to penetrate
Digital Audio Distribution and Sales	A	

Table 1: MIRADOR selected applications

7 Visual watermarking assessment

7.1 Robustness to compression measurement

The goal of the MIRADOR project is to test the behaviour of the watermarking techniques [3]within the MPEG-4 framework and MIRADOR targeted applications. The

MPEG-4 encoder / decoder used in the project is the MoMuSys software. In this software, only MPEG-4 natural video objects tools are available. Therefore the working domain of the MIRADOR project is limited to the natural image content tools (still image and video). Watermarking / monitoring efficiency with regard to SNHC objects or graphics objects will be not tested in the context of this framework. It means that, if application targeted by the MIRADOR project, requires SNHC tools and natural tools, only the behaviour of the watermarking in the context of the natural content tools will be tested. However, by testing the behaviour of tools such as EZW (wavelet compression), synthetic objects watermarking is partially treated. By partially treated, we mean that we don't watermark « animation parameters », e.g. motion vectors in animated 2D meshes, facial animation parameters, ... but, by the study of still image watermarking, we deal partially with the problem of texture mapping watermarking in synthetic applications. Through the MIRADOR project, first efficient evaluations will be carried out.

From MIRADOR targeted applications, useful objects types profiles and levels have been listed. From these object types, profiles and levels have been chosen and for every application, a set of parameters / tools has been proposed. These sets of parameters will allow defining a large number of MoMuSys encoder parameters to test our watermarking algorithms in the context of MPEG-4 applications. Therefore, at the end of the MIRADOR project, it will be possible to say if tested watermarking algorithms will be efficient and it will be possible to give bounds of watermarking efficiency in the MPEG-4 context. After this first experimental stage, we will develop new solutions so that the watermarking efficiency regards MPEG-4 encoder is improved

Targeted Applications	Profile chosen	Level chosen	Tested bit-rates (Mbits/s)	Objects type to evaluate
Infotainments	Core	level 2	2	Simple, Core
Infotainments	Core	level 1	0.384	Simple, Core
Streaming Video on Internet / Intranet	Main	level 2	1, 2	Simple, Core, Main, Scaleable texture
Streaming Video on Internet / Intranet	Simpl e scaleable	level 2	0.256	Simple, Simple Scaleable
Broadcasting	Main	level 3	2, 4, 8	Simple, Core, Main, Scaleable texture
Digital Television Set Top Box	Main	level 3	2, 4, 8	Simple, Core, Main, Scaleable texture

For each video object, a watermark is embedded in each frame. It is a 64 bits binary message. Robustness results will show the percentage of wrongly monitored embedded bits as well as the percentage of incorrectly protected frames. A frame is considered as not protected if at least one bit among the 64 bits message is wrong.

Error correcting code or time integration tools have not been used in these experiments. Even if these tools would enable to improve watermarking robustness, they could mask watermarking fine behaviour. When robustness results are given with error correction or post-processing, they must also be given as bare results, without any processing. This allow to analyse precisely the behaviour of watermarking algorithms, in order to be able to realise improvements.

7.2 Robustness measurement results

7.2.1 Still images

Concerning still images, the relevant robustness tests consist in using DCT and wavelet compression schemes recognised by MPEG4. For this purpose, we used JPEG and EZW at different bit rates to compress the images. We then made the statistics of watermarking retrieval after decompression.

Robustness assessments toward JPEG compression has been carried out by using a 50 still images database. This database regroups diverse images having different characteristics, from low textured images to highly textured images or even synthetic images. The JPEG baseline algorithm has been used and watermarking robustness is expressed with regard to JPEG quality factor. This parameter varies from 1 (high compression and very poor image quality) to 100 (low compression and lossless image encoding). For our experiences, it varies from 10 to 90 with a step of 10. It enables to obtain a great variety of compression rates. After compression, each image has been decompressed and a monitoring stage has been conducted, that is to say a statistics of watermarking retrievals.

Quality factor	10	20	30	40	50	60	70	80	90
images correctly monitored	0%	0%	0%	0%	5.8%	51.2 %	59.6 %	76.9 %	86.5 %
bits incorrectly decoded	50	39.8	29.2	22.5	16	12	8.2	5.5	3.7

Table 2 : Watermarking robustness toward JPEG compression

The table presents for each quality factor, the mean percentage of wrong bits with regard to the quality factor, the percentage of images perfectly monitored (the percentage of wrong bits is null) and the mean PSNR of the compressed images luminance.

A deeper analysis of the results show that except for some synthetic images, the watermark can be considered as robust toward JPEG. As a matter of fact, the watermark becomes inefficient at low quality percentages. The corresponding decompressed images have lost a lot of commercial value.

As for the JPEG evaluation, robustness assessments toward EZW compression has been carried out by using a database of 50 still images. The EZW algorithm is coming from the MoMuSys ACTS project. The encoding parameters used for this evaluation are:

- filter type : default filter of the algorithm
- Number of wavelet decomposition levels : 5
- wavelet uniform
- encoding mode : Multi-quantization level
- zero tree scanning : band by band
- spatial scalability level : 1
- SNR spatial scalability : 1

The parameters that enable to modify compression ratio is the quantization step of the AC coefficients of the wavelet decomposition. Quantization steps used are : 2, 4, 8, 16, 32. It enables to obtain a great variety of compression rates.

The following table summarises the obtained results. It presents according the quantization step of the wavelet decomposition AC coefficients, the mean PSNR of the images decompressed, the mean percentage of wrong bits at the monitoring stage, obtained after EZW encoding.

AC quantisation step	32	16	8	4	2
average PSNR (dB)	31.27	34.65	39	43.8	48.54
average % of wrong bits	36.12	11.9	1	0	0
average bits rate (bit/pixel)	0.4	0.8	1.48	2.53	4

Table 3: Robustness toward EZW compression

The same conclusions can be drawn for EZW. When the efficiency of the watermrak decreases, the quality becomes much lower.

7.2.2 Moving Images

All along the project, experiments have been conducted with the following MPEG4 sequences : Akyio, Sean, News, Weather, Stephan, Flower and Table Tennis. Akyio and Sean are class A sequences (few motion), News is a class B sequence (intermediate motion), flower, tennis table and Stephan are class C sequences (lot of motion), weather is a class E sequences (Hybrid natural and synthetic). Each sequence consists in 300 images.

Sequences formats are QCIF (176 by 144 at 30 Hz), CIF (352 by 288 at 30 Hz) and ITU-R BT.601 (720 by 486 at 30 Hz or 720 by 576 at 25 Hz). QCIF and CIF sequences are progressive sequences and ITU-R BT.601 sequences are interlaced.

Several applications have been considered for the tests:

- Broadcasting ;
- Digital Television Set-Top Box ;
- Infotainment ;
- Streaming Video on the Internet/Intranet .

Each application requires different sets of MPEG-4 tools and parameters, which include :

- Rectangular video encoding with different images formats : QCIF, CIF, ITU-R BT.601;
- Arbitrary shape Video Object encoding ;
- Temporal and spatial scalability for video sequences encoding ;

- Progressive and interlaced encoding tools ;
- Still texture encoding with EZW (Embedded Zero tree Wavelet) ;

For each of them, experiments have been conducted. Analysis of the obtained results will allow to identify and highlight problems on the current algorithm , and then to set out new ways of research in order to solve those encountered problems. Consequently, only crude results (percentage of wrong monitored bits and wrong monitored images) will be given in this document Bare results are the results obtained without the use of data post-processing. In a real implementation, it is obvious that clever spatial and temporal redundancy should taken into account through the use of error correcting codes.

Sequence Name	targeted bit rate (Mbit/s)	Obtained bit rate (Mbit/s)	% of wrong bits	% of wrong images
Akyio	0,25	0,27	0,062	4
Akyio	0,5	0,52	0	0
Akyio	1	1,01	0	0
Akyio	2	2,00	0	0

Table 4: effects of the compression on CIF sequences : Akiyo

Sequence Name	targeted bit rate (Mbit/s)	Obtained bit rate (Mbit/s)	% of wrong bits	% of wrong images
Weather	0,5	0,55	19,3	100,0
Weather	1	1,05	9,4	99,67
Weather	2	2,02	4,2	94,67

Table 5: effects of the compression on CIF sequences : Weather

The tables above gather some of these results. The results are very good with these type of sequences. An improvement is needed when dealing with sequences having an important motion. As a consequence, we have worked in order to improve the algorithm. The new results are already promising, but they are too recent to be published. Moreover, this upgrading work is still on progress.

7.2.3 Arbitrary shapes objects

The algorithm has been modified in order to take into account arbitrary shape objects. Each of the objects is watermarked using a different value (identifier) and key, with regard to the binary shape. As the image sizes are quite small, few blocks are available to embed watermark . Moreover, as video objects are extracted and independently watermarked (one message associated with one video object), the number of blocks available is still more reduced . Each of the objects after watermarking is compressed separately. On the output of the decoder, each of the object is monitored separately.

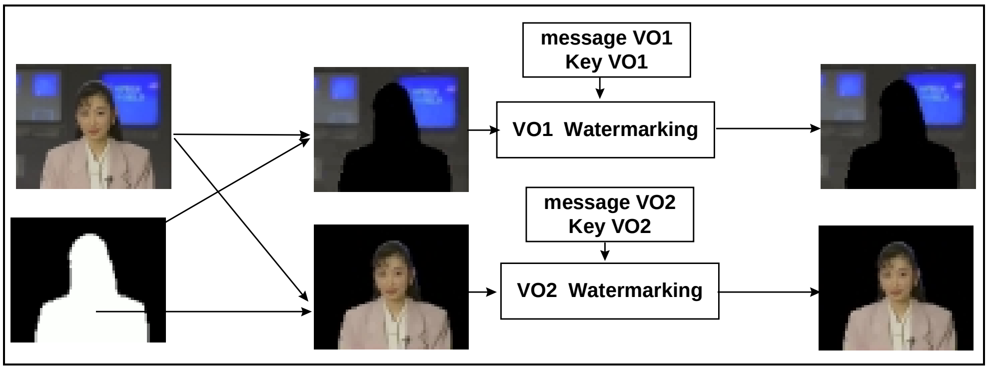


Figure 2: Watermarking of Arbitrary shape

Sequence Name	Bit rate (VO1) Mbits/s	% of wrong monitored images (VO1)	% of wrong bits (VO1)	Bit rate (VO2) Mbits/s	% of wrong monitored images (VO2)	% of wrong bits (VO2)
Akyio	0,19	0	0	0,74	12,7	0,218

7.3 Visual quality assessment protocol

Lots of work has been done on how to evaluate under standardised conditions the picture quality, necessary for the introduction of new television systems. For the moment, as far as the evaluation of codecs is concerned, only subjective assessments give confident results. Subjective measurements are the result of human observers, providing their opinion of the video quality.

Existing assessment methods have been developed and refined over years and have provided the reliable and sensitive test procedures, which have guided specification and design of actual television services. Formal subjective testing, as defined by ITU-R BT.500, has been used for many years with a stable set of standard methods. The current state of practice defines the best methods available and provides advice for the non-specialist on issues concerned with the choice of :

- basic methodology
- viewing conditions
- selection and screening of observers
- scaling method
- reference condition
- presentation timing
- test picture sequences
- analysis of voting
- presentation of results

In order to standardise subjective image quality evaluation, ITU Recommendation ITU-R BT.500 has been prepared and is regularly reviewed, to provide instructions on what seems the best available methods for the assessment of picture quality. In particular, ITU-R BT.500 defines common features of the tests.

A number of non-expert observers watch a series of test scenes for about 10 to 30 minutes in a controlled environment and are asked to score the quality of the scenes in one of a variety of manners. Advantages of subjective testing are valid results for both non-compressed (analog or digital uncompressed) and compressed television systems. A scalar mean is obtained (mean opinion score) and it works well over a wide range of still and motion picture applications.

The ITU-R BT.500 methods can be classified in three categories :

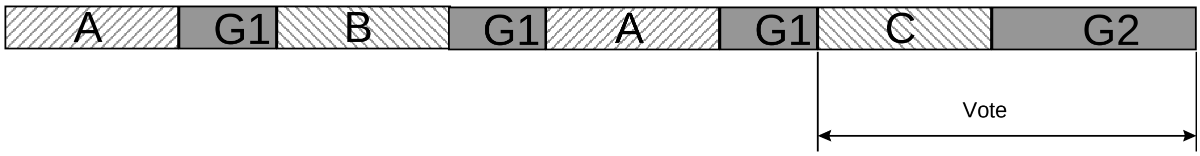
- Double stimulus methods (DSIS and DSCQS)
- Single stimulus methods (Adjectival categorical and Non-categorical methods, SSNCS, SSCQE)
- Stimulus comparison methods (Adjectival categorical judgement and Non-categorical methods)

These methodologies have been detailed in the MIRADOR D2 Deliverable, publicly available. As explained, in this document, in order to assess the quality of watermarked material and to discriminate small differences between the original and the modified pictures, the Double Stimulus Continuous Quality Scale (DSCQS) method is recommended.

7.3.1 Double Stimulus Continuous Quality Scale (DSCQS)

In the Double Stimulus method with a continuous quality scale, all the sequences are presented unimpaired (assessment reference) and impaired. The basic principle is to assess pairs of sequences.

One of the two sequences is a reference, the other has been subject to processing by the watermarking system being assessed. Observers are not informed of the position of these two images in the sequence and they allocate a score to each individual image using a continuous quality scale. This couple of sequences is repeated once. The sequencing of this method is summarised in the following figure :



A=10s	Test sequence A
G1=3s	Mid grey picture
B=10s	Test sequence B
G2=5-11s	Mid grey

Figure 3 : Presentation sequence for the DSCQS method

The scoring sheet includes pairs of scales. Each scale is divided into five segments, but only on the left-hand side, are the appropriate descriptions (excellent, good, fair, poor and bad). Observers express their opinion by marking the scales.

The Grading scales corresponds to the following notes:

5	excellent
4	good
3	fair
2	poor
1	bad

Table 6: Grading scale and quality note correspondence

The marks on the graphical scale are converted into numbers by taking the upper scale boundary as 5 and the lower as 0. They are read for A and B, separately. As a first step of processing, for each pair of votes the difference is taken: reference sequence minus processed sequence. Then for each combination of sequence, the average over observers of these difference scores is calculated as well as the standard deviation and the 95% confidence interval. Finally for each processed case, the average over observers and sequences of difference score is calculated, together with the appropriate standard deviation and 95% confidence interval.

At least 15 observers should be used. It has been found that means become stable when at least 15 observers are used. They should be non-expert, in the sense that they are not directly concerned with television picture quality as part of their normal work, and are not experienced assessors. Assessors should be introduced to the method of assessment, the types of impairment or quality factors, the grading scale, the sequence and timing. Training sequences demonstrating the range and the type of the impairments to be assessed should be used with illustrating pictures other than those used in the test.

In a subjective experiment, a large amount of data is collected. These data must be condensed by statistical techniques to present results, which summarise the performance of the system.

7.3.2 Quality assessment results

The conditions described in the ITU-R BT.500 have been respected. From the two types of environment, (laboratory or domestic environment), we chose the domestic environment which has the advantage of not requiring a special and standardised room for the assessments. Twenty-six observers respectively to the assessment conditions participated to the MIRADOR quality evaluation. They were between 25 and 45 years old, and were not involved in television as part as their normal work.

Assessor have been introduced to the method of assessment, the types of impairment or quality factors, the grading scale, the sequence and timing. Training sequences demonstrating the range and the type of the impairments to be assessed have been used with illustrating pictures other than those used in the test. TALISMAN ACTS project test sequences have been used for the explanation and training phase.

The training phase has the main objective of giving instructions to the assessors. During this phase a trial with two presentations has been done, to get the assessors used to the timing and the quality range shown during the test.

It is very convenient for the observers to have a sequence identifier; it avoids confusion between sequences A and B. The sequence identifier has been inserted (for example, A or B) in the centre of a mid-grey frame, before the corresponding sequence.

The following sequences have been used for the quality assessment, respectively to this presentation order :

Sequence number	Title	Duration	Source library tape
1	Mobile and calendar	30s	EBU
2	Flower garden	30s	EBU
3	Akyio	10s	MPEG-4
4	News	10s	MPEG-4
5	Weather	10s	MPEG-4
6	Bream	10s	MPEG-4
7	Sean	11s	MPEG-4
8	Silent	18s	MPEG-4
9	Foreman	16s	MPEG-4

Table 7: Video sequences used for MIRADOR quality assessment

The sequences to be assessed were between 10 s and 30s long.

For our experiments we are using a PC platform with professional video boards with components (YUV) and ITU-R BT.656 Digital video outputs. The sequences stored on a disk array and rendered by the system are not compressed.

The display monitor used has been adjusted as described in the methodology. The Monitor is connected through its YUV inputs to the PC video system. The height of the screen is 40 cm. In a domestic environment assessment type, the viewing distance corresponds to the preferred viewing distance PVD (according to ITU-R BT.500-8), which is 6H (6 times the height of the screen) in our case, i.e. 2.4m. In the case of the laboratory assessment type, in order to discriminate small differences it is also recommended to use 4H distances. Because watermarking has been designed to be invisible, the impairment that could appear should be very difficult to be seen. It is then appropriate to complete the normal evaluation in a domestic environment with the critical distance of 4H.

Assessments have been done using the two distances, through two sessions. In the first test session (Session 1) two of the three assessors have been installed at 4H and the third one at 6H. For the second test session (Session 2), we made the opposite, whom had been at 4H were placed at 6H and the third assessor be installed at 4H.

Visual quality assessment results

1 Results at the PVD (Preferred viewing distance)=6H=2.4m

Average value

	Calendar	Flower	Akiyo	News	Weather
original	3.577	3.452	3.837	3.875	3.808
watermarked	3.587	3.471	3.885	3.846	3.779
difference original- wk	-0.010	-0.019	-0.048	0.029	0.029

	Bream	Sean	Silent	Foreman
original	3.863	3.904	3.583	3.577
watermarked	3.844	3.865	3.525	3.587
difference original- wk	0.019	0.038	0.058	-0.010

Average value

	Calendar	Flower	Akiyo	News	Weather
original	0.657	0.542	0.460	0.445	0.375
watermarked	0.693	0.506	0.515	0.519	0.435

	Bream	Sean	Silent	Foreman
original	0.509	0.361	0.609	0.532
watermarked	0.504	0.394	0.618	0.528

95% Confidence interval

	Calendar	Flower	Akiyo	News	Weather

original	[3.324, 3.829]	[3.244, 3.660]	[3.660, 4.013]	[3.704, 4.046]	[3.663, 3.952]
watermarked	[3.320, 3.853]	[3.277, 3.666]	[3.687, 4.083]	[3.647, 4.046]	[3.612, 3.946]

	Bream	Sean	Silent	Foreman
original	[3.668, 4.059]	[3.765, 4.043]	[3.349, 3.817]	[3.373, 3.781]
watermarked	[3.651, 4.038]	[3.714, 4.017]	[3.287, 3.763]	[3.384, 3.789]

The obtained results are very closed together, having similar values for both average and standard deviation value. Even the quality notes distributions curves doesn't allow to separate original and watermarked sequences. We can conclude that at the preferred viewing distance, 6 times the height of the screen in our case (6H=2.4m), there are no perceived differences between original and watermarked sequences.

Conclusion

This document overviewed the constraints involved by MPEG-4 to watermarking algorithms. It presented the evaluation methodology followed in MIRADOR Europeen ACTS project in order to assess robustness to compression and quality of the watermarked material. The performance of an algorithm issued from previous work has been evaluated. The robustness results are quite promising and, concerning the quality assessment, the standardised evaluation process applied on the test sequence library in formal conditions did not allow to establish any difference between the original and the watermarked material. The invisibility of the potential watermarking artefacts has been demonstrated.

More information is available on http://www.tele.ucl.ac.be/MIRADOR/

References

1. http://www.tele.ucl.ac.be/MIRADOR/
2. http://www.tele.ucl.ac.be/TALISMAN/
3. V Darmstaedter, JF Delaigle, D Nicholson, B Macq “A block based watermarkingTechnique for MPEG2 signals: optimization and validation on real digital TV distribution links”, proceedings ECMAST 98, Springer, pp 190-206
4. http://www.doi.org
5. http://www.doi.org/workshop/minutes/CISoverview/
6. IMPRIMATUR: http://www.imprimatur.alcs.co.uk
7. MPEG4 overview document W2614, available at http://www.cselt.it/mpeg/

8. Recommendation ITU-R BT.500-8, Methodology for the subjective assessments of the quality of television pictures.