

**INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11
MPEG2009/m16748
June/July 2009, London, UK**

Source

Status Proposal

Title **Evaluation of Core Experiment for MPEG-7 Video Signature Tool**

Author **Marzia Corvaglia, Fabrizio Guerrini, Riccardo Leonardi, Pierangelo Migliorati,
Eliana Rossi (DEA-SCL, University of Brescia, Italy)**

Contacts: firstname.lastname@ing.unibs.it

Table of Contents

Abstract	1
1 Introduction	1
2 Experiment conditions as proposed in VCE-7 Core Experiment on Video Signatures	2
3 The alternative approach	2
4 Experimental results	2
5 Recall and Precision	5
6 Proposals for next CEs	6
Conclusions	6
References	6

Abstract

The present document evaluates and discusses the experiment conditions proposed in the VCE-7 CE on Video Signature, discussed in the previous MPEG meetings [MPEG-VS-09-01] [MPEG-VS-09-02].

In particular we present some simulation results that show the weaknesses of the evaluation procedure, proposed in the VCE-7 CE on Video Signature, and that highlight some aspects that could impair the performance evaluation of the here considered technology in a realistic environment. Therefore, we propose a new Core Experiment which will consider this issues.

1 Introduction

The Video Signature is the natural extension of the Image Signature Descriptor, which has been recently standardized, to the video signals.

The purpose of the Image Signature is to find identical or modified images of a given query image. The purpose of the Video Signature is, given a query video (also called ‘*copy*’), to find the video (also called ‘*original clip*’) where the query has been taken from, even if:

- the query has been *attacked*, that is it has been edited, modified, re-encoded, etc.;
- the query has been immersed in a dummy video;
- both.

The Video Signature can be extracted from both query and original clips. The basic idea consists of detecting the original clip comparing the Video Signatures of the query with the Video Signature of the original clips. This problem is better known in the literature as Content Based Copy Detection (CBCD).

2 Experiment conditions as proposed in VCE-7 Core Experiment on Video Signatures

The details of the VCE-7 CE proposed in MPEG for Video Signature are reported in [MPEG-VS-09-01] [MPEG-VS-09-02]. In the next Sections we will refer to these documents, in terms of methods of evaluation, test conditions and terminology.

3 The alternative approach

The performance of a method of search and retrieval of a multimedia document is usually tested with the exhaustive approach, briefly explained in the next paragraphs.

Given a database of N multimedia items (e.g., images, videos or audio clips) I_i with $i \in [1, N]$, given a set of features F_i that characterises each item I_i , given a query item Q whose similar items or modified copies need to be found in the database of multimedia objects, given the feature F_Q extracted from the query, chosen an appropriate distance measure D between the features, the exhaustive approach consists in calculating all distances D_i ($i \in [1, N]$), between F_Q and F_i and to select the most similar item (or the modified copy) I_j whose feature F_j has the minimum distance with F_Q : $D_j = \min(F_i, F_Q)$ with $i \in [1, N]$.

To evaluate the system performance two common parameters are usually considered:

- *Precision*, which is an indicator of the retrieved relevant items in the retrieved items;
- *Recall*, which is an indicator of the retrieved relevant items in the overall number of relevant items in the database.

In some cases these parameters are summarized in a single parameter F which is given by the weighted harmonic mean of *Precision* and *Recall*.

We have adopted this method also for the case of Content Based Copy Detection.

4 Experimental results

The exhaustive approach has been tested using the method proposed by University of Brescia for Video Signature tool [MPEG-UNIBS-09-2].

The experiments have been performed using Color Layout (CL) Descriptor on DIRECT 2 seconds with Light modifications, in three different conditions:

1. Exhaustive approach considering the minimum distance;
2. MPEG evaluation considering three thresholds, extracted by setting a false positive rate $\leq 5\text{ppm}$, $\leq 50\text{ppm}$ and $\leq 500\text{ppm}$ in the Independence test [MPEG-VS-09-01] [MPEG-VS-09-02];
3. MPEG evaluation considering a threshold extracted with an alternative method.

In Figure 1 the simulation results are plotted and compared with the adopted technology (NEC-Mitsubishi, blue stars in Figure 1).

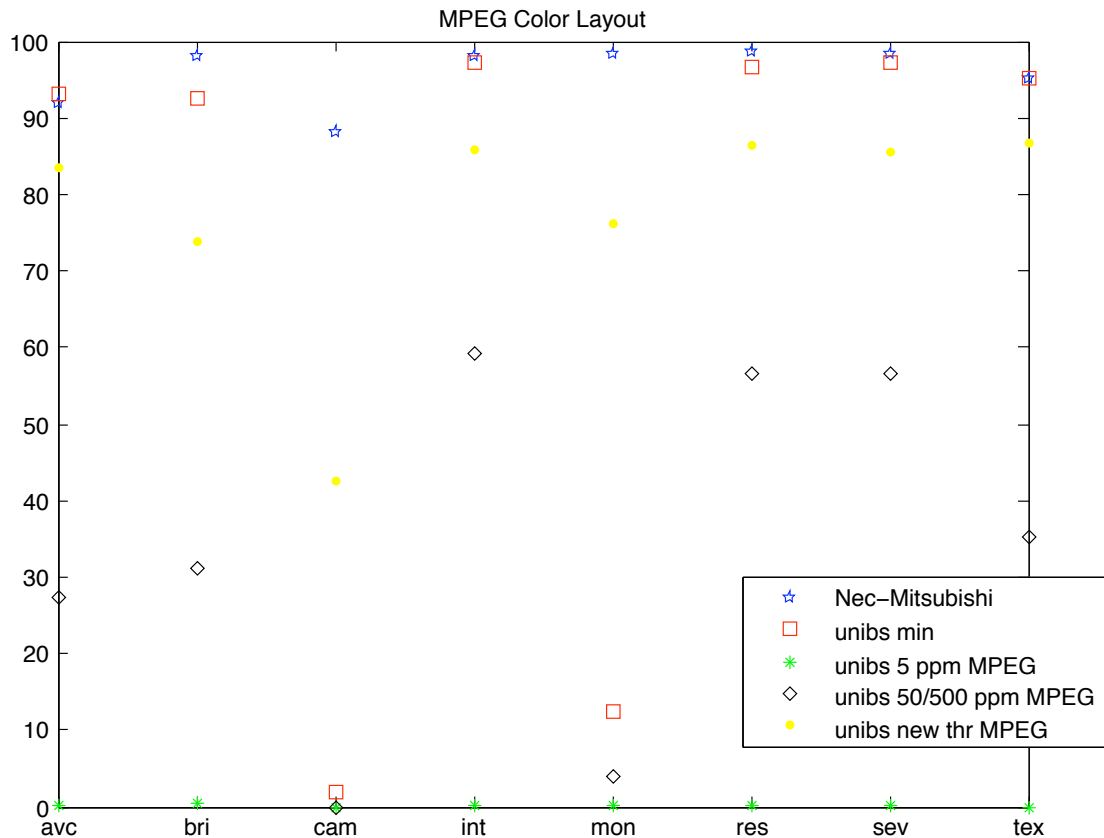


Figure 1 Experimental results (y-axes: percentage of the success ration – x-axes: type of modification).

Analyzing the results we can make some important considerations.

- First, the exhaustive method (red rectangles in Figure 1) provides comparable performance with the adopted technology (blue stars), except for ‘Capturing on Camera’ and ‘Color to Monochrome conversion’. On the other hand, the MPEG method (black squares and green asterisk) provides low performance and results not comparable with the exhaustive method. This experiment shows that the evaluation method can seriously compromise the performance of the considered system for copy detection, and hence could invalidate the actual performance of the proposed technology.
- Second, the choice of the threshold is a crucial aspect that should be dealt carefully with.

For example, the threshold obtained by the Independence test with false positive rate ≤ 5 ppm provides performance closed to null, while with false positive rate $\leq 50/500$ ppm the performance significantly increases.

In this sense the Independence test forces the threshold to a value that is not the optimum one for every technology. As a counterexample, an alternative approach for automatic definition of the threshold, mainly based on the training of a subset of the database, has been considered. The results obtained with the MPEG proposed evaluation are show in Figure 1 (yellow dots) and prove that the performance can significantly change with a different method for threshold definition.

- Third, as hinted above, the exhaustive approach fails on some attacks ('Capturing on Camera' and 'Color to Monochrome conversion'). This behaviour could be seen as a clear flag to discriminate for which attack the considered technology works better.

An additional experiment has been performed. For every correct copy found in the database with the exhaustive approach, we have evaluated the temporal localization of the query in the original clip in four cases, namely: temporal localization within 1 seconds, 5 seconds, 10 seconds, 20 seconds. The results are plotted in Figure 2.

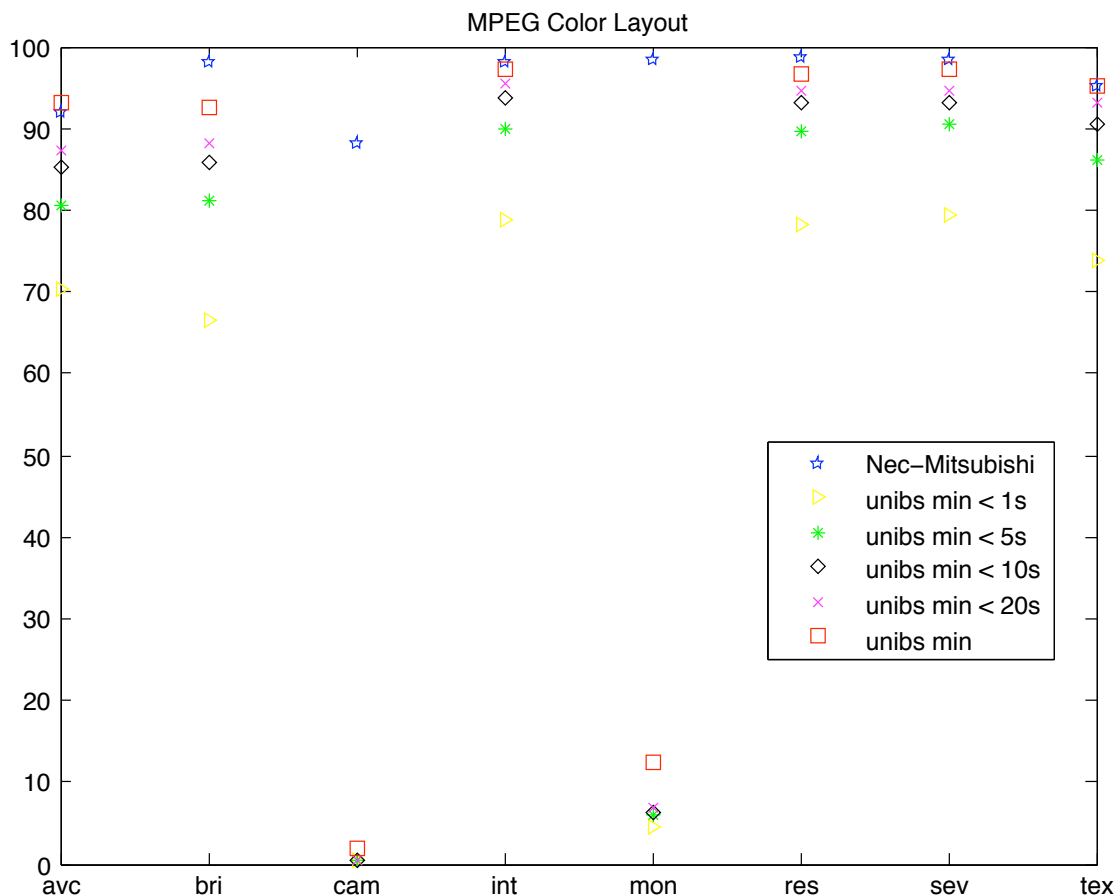


Figure 2 Experimental results for temporal localization (y-axes: percentage of the success ration – x-axes: type of modification)..

These experiments prove that the proposed CE conditions (1 second) are very restrictive. From our point of view, finding the query in the original clip is already appreciable and the temporal precision can be used as a further evaluation parameter of the considered technology.

5 Recall and Precision

Given a query Q (copy) and a database of N items (original clips), we suppose to run the Video Signature Tool in order to find the original clip which includes the copy. Usually three possible outputs are available:

- the original clip is found by the system and it is correct (true positive = T_p)
- the original clip is found by the system, but it is not correct or it is not present in the database (false positive = F_p)
- the original clip, which is present in the database, is not found by the system (false negative = F_n)

Thus *Recall* and *Precision* are respectively defined as follow:

$$R = \frac{T_p}{T_p + F_n} \quad (1)$$

$$P = \frac{T_p}{T_p + F_p} \quad (2)$$

Using the evaluation method proposed by MPEG the common parameters R and P cannot be calculated with the traditional formulas (1) and (2).

In MPEG evaluation process the false positive value F_p is set in the Independence test ($\leq 5\text{ppm}$) and hence the *Precision* is approximately close to 100%.

The query set proposed by MPEG (Robustness set) consists in a set of queries whose original clips are always present in the database (SetA). This means that the false positive F_p , when the original clip is found by the system even if it is not present in the database, cannot be defined in the actual MPEG evaluation and hence the *Precision* (100%) is not reliable. A set of additional queries, obtained by “dummy” videos not available in the database, should be added to the query set in order to compute the correct *Precision*.

A further consideration can be done about MPEG methodology. If we suppose to have “dummy” queries in the query set, the MPEG evaluation is not applicable as explained following. The proposed algorithm for Video Signature Tool is required to output a binary decision: “1” if the query is *related* to the original clip or “0” if the query is *not related* to the original clip [MPEG-VS-09-01] [MPEG-VS-09-02]. This output is generated by comparing the video signature of the query with the video signature of the original clip which contains the query. Hence if we consider a “dummy” query, which is not included in any of the original clips, the video signature comparison cannot be performed. For this scenario only exhaustive approach can be used.

6 Proposals for next CEs

On the basis of the considerations reported in Section 4, some proposal for next CEs are reported in the following:

1. Test the adopted technology also with the exhaustive approach;
2. Modify the queryset by introducing “dummy” queries in order to:
 - a. validate the system by the computation of *Recall* and *Precision*, which also require the exhaustive approach, as argued in the previous Section;
 - b. make the experiments more realistic.
3. Relax the conditions for the threshold definition.
4. Relax the constraints for temporal localization of the query in the original clip.

Conclusions

In this document we have presented some experimental results and some considerations about the adopted evaluation method for MPEG-7 Video Signature Tool. Then, some inputs for new CEs are proposed.

References

[MPEG-VS-08] MPEG Video Sub-Group, “Updated Call for Proposals on Video Signature Tools”, ISO/IEC JTC1/SC29/WG11 MPEG 2008/N10155, October 2008, Busan, Korea.

[MPEG-VS-09-01] MPEG Video Sub-Group, “Description of Core Experiments in Video Signature Description”, ISO/IEC JTC1/SC29/WG11 MPEG 2009/N10345, February 2009, Lausanne, CH.

[MPEG-VS-09-02] MPEG Video Sub-Group, “Description of Core Experiments in Video Signature Description Development”, ISO/IEC JTC1/SC29/WG11 MPEG 2009/N10544, April 2009, Maui, US.

[MPEG-UNIBS-09-2] M. Corvaglia, F. Guerrini, R. Leonardi, P. Migliorati, E. Rossi, “A proposal for Video Signature Tool and Video Fingerprinting”, Document ISO/IEC JTC1/SC29/WG11 MPEG 2009/m16475, Maui, Awaii, April 2009.

[MPEG7-01] “Text of ISO/IEC 15938-3/FDIS Information technology – Multimedia content description interface – Part 3 Visual”, ISO/IEC/JTC1/SC29/WG11 Doc. N4358, July 2001, Sydney, Australia.

[MPEG7-02] “Text of ISO/IEC 15938-5/FDIS Information technology – Multimedia content description interface – Part 5 Multimedia Description Scheme”, ISO/IEC/JTC1/SC29/WG11 Doc. N4358, July 2001, Sydney, Australia.