ISO/IEC JTC 1 N9590

2009-06-09

Replaces:

ISO/IEC JTC 1
Information Technology

Document Type: text for DAM ballot

Document Title: Text for DAM Ballot of ISO/IEC TR 15938-8:2003/Amd.5:2009(E), Multimedia content description interface -- Part 8: Extraction and use of MPEG-7 descriptions, AMENDMENT 5: Extraction and matching of image signature tools

Document Source: SC 29 Secretariat

Project Number:

Document Status: This document is circulated to JTC 1 National Bodies for a 3 month DAM ballot. National Bodies are asked to vote and submit their comments via the on-line balloting system by the due date indicated.

Action ID: VOTE

Due Date: 2009-09-10

No. of Pages: 11

Secretariat, ISO/IEC JTC 1, American National Standards Institute, 25 West 43rd Street, New York, NY 10036; Telephone: 1 212 642 4932; Facsimile: 1 212 840 2298; Email: lrajchel@ansi.org

ISO/IEC TC JTC 1/SC 29

Date: 2009-04-24

ISO/IEC TR 15938-8:2003/Amd.5:2009(E)

ISO/IEC TC JTC 1/SC 29/WG 11

Secretariat: ANSI

Information technology — Multimedia content description interface — Part 8: Extraction and matching of image signature tools

Technologies de l'information — Interface de description du contenu multimédia

Document type: Technical Report
Document subtype: Amendment
Document stage: (60) Publication
Document language: E

STD Version 2.1c2

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

Amendment 5 to ISO/IEC TR 15938-8:2003 was prepared by Technical Committee ISO/IEC/TC JTC 1, *Information Technology*, Subcommittee SC 29, Coding of audio, picture, multimedia and hypermedia information.

Information technology — Multimedia content description interface — Part 8: Extraction and matching of image signature tools

Add after 4.8:

4.9 Visual Signatures

4.9.1 Image Signature

Most visual descriptors are very useful when trying to find images with *similar* content however, such descriptions are intended to be general: finding *identical* content is not of special importance. The image signature descriptor is robust (unchanging) across a wide range of common editing operations, but is sufficiently different for every item of "original" content to identify it uniquely and reliably – just like human fingerprints. Therefore this descriptor is able to distinguish between two images which have similar content and two images which are the same.

There are three components within the image signature, the first two are global signatures of the complete image and the third is composed of a set of local image signatures.

4.9.1.1 Feature Extraction

11.3.5 and 11.3.6 of ISO/IEC 15938-3:2002/Amd.3:2009 describes the extraction of the image signature.

4.9.1.2 Similarity Matching

4.9.1.2.1 Global Descriptors

To perform matching between two signatures B_1 and B_2 both of length N, the Hamming distance is taken:

$$H(B_1, B_2) = \sum B_1 \otimes B_2 \ ,$$

where $\otimes$ denotes the exclusive OR (XOR) operator.

This can be normalised to the range 0-1 using the normalised Hamming distance:

$$\hat{H}(B_1, B_2) = \frac{1}{N} \sum B_1 \otimes B_2 \ .$$

Evaluation on test data has shown that a Hamming distance $H(B_1, B_2)$ of 147 corresponds to a false alarm rate of less than 0.05 parts per million (ppm). A distance of 169 corresponds to a false alarm rate of 1ppm.

4.9.1.2.2 Feature Descriptors

1. Form hypotheses of 3 features
Use 32 out of 80 features

2. Apply Geometric Constraints
Feature direction & ratio of line length

3. Form hypotheses of 4 features
Use all 80 features

4. Apply Geometric Constraint
Ratio of areas

Figure AMD5.1 — Feature matching is carried out in a four stage process which combines hypothesis forming (using the feature signature) and geometric contstraints.

Matching images can be performed by comparing the feature signatures which form a part of the image signature. For efficiency four stages are used, as shown in Figure AMD5.1, to match feature signatures from a pair of images. Hypotheses are formed in stages one and three. A series of geometric test are performed in stages two and four, these tests must be passed in order for a hypothesis to progress to the next stage. The stages become increasingly computationally complex so each stage aims to minimise the number of hypotheses that are accepted for subsequent processing.

A first stage involves forming hypotheses for potential matching features by comparing the first 32 out of upto 80 feature signatures from one image with the first 32 out of up to 80 feature signatures from the second image. To perform matching between two feature signatures B_1 and B_2, the Hamming distance is taken:

$$H(B_1, B_2) = \sum B_1 \otimes B_2 \ .$$

A match-list of feature pairs with a Hamming distance less than a predefined threshold T_A is created. A set of hypotheses (candidate pairs of potentially matching features) is generated by taking all combinations of three pairs of matching features from the match-list. A hypothesis is made up of a set of three pairs of feature points ($a_1, a_2, a_3, b_1, b_2, b_3$). A set of three pairs of features is declared a match if the cumulative distance is below a constant threshold T_B. The set of hypotheses is ordered by their cumulative distance, with the lowest distance (i.e. highest probability of being a match) first. In order to minimise complexity the number of hypotheses may be limited to the hypotheses corresponding to the lowest distance.

A second stage applies geometric constraints to each hypothesis generated from stage one. Two geometric constraints should be applied to reduce false matches, i) feature direction and ii) the ratio of line lengths.

A direction is associated with each feature, the direction for feature a_n is denoted θ_{an}. The difference between angles corresponding to features from the same image is taken to ensure that the angle distance measure is unaffected by a rotation. The distance measure addresses the issue of left-right and top-bottom flip modification as is done here using these two measures of angle distance

$$\Omega_1 = \left|(\theta_{a1} - \theta_{a2}) - (\theta_{b1} - \theta_{b2})\right| \text{ and } \Lambda_1 = \left|(\theta_{a1} - \theta_{a2}) + (\theta_{b1} - \theta_{b2})\right|$$

Thresholds are applied to the two angle distance measures Ω_1, Λ_1 to test whether they are in allowable intervals. Allowable values for the angle distances are given by

$$\Omega_1 \le q \text{ or } 16 - q \le \Omega_1 \le 16 + q \text{ or } \Lambda_1 \le q \text{ or } 16 - q \le \Lambda_1 \le 16 + q,$$

where q is a constant that can be used to vary the sensitivity to changes of angle.

Using the locations of features selected in each hypothesis a ratio of line lengths between the features is used as a constraint. By $|a_1 a_2|$ we denote the length of the line connecting the location of feature a_1 to the location of feature a_2. Three feature points a_1, a_2 and a_3 are used to find the ratio of lines lengths. Three line length ratios are computed

$$L_1 = \frac{|a_1 a_2|}{|a_1 a_3|},\ L_2 = \frac{|a_2 a_3|}{|a_3 a_1|} \text{ and } L_3 = \frac{|a_3 a_1|}{|a_1 a_2|}.$$

Measures of difference between the line length ratios (G_1, G_2, G_3) for two images A and B are defined as:

$$G_1 = \frac{|L_1^A - L_1^B|}{L_1^A + L_1^B},\ G_2 = \frac{|L_2^A - L_2^B|}{L_2^A + L_2^B} \text{ and } G_3 = \frac{|L_3^A - L_3^B|}{L_3^A + L_3^B}.$$

Two thresholds should be applied to the distance measures, one to constrain the magnitude of the individual measures and the other to constrain the sum of pairs of distances ($G_1 + G_2, G_2 + G_3, G_1 + G_3$).

Any hypothesis that fails one or more of the tests described above is removed from the set of valid hypotheses. If after testing all hypotheses the set is empty the images are declared non-matching. If the hypothesis has passed all tests then it progresses to the next stage to generate four feature hypotheses.

The third stage of matching compares all N features from the first image with all N features from the second image. Now, there is already a hypothesis for three feature pairs so the aim is to find a set of hypotheses with four pairs, hypotheses that add one more feature to the current three feature pair hypothesis. Potential features are found based on the signature distance as in stage one. The set of hypotheses are ordered by the distance of the fourth feature.

In stage four geometric analysis is carried out on the sets of four feature hypotheses to further reduce false acceptance rates. The ratio of lines used in stage two is invariant to similarity transforms. With four features the invariance can be relaxed to the more general affine transformation by using the ratio of areas as described below. This allows tighter thresholds to be set which accept true matches whilst rejecting false matches.

From a set of four feature locations (a_1, a_2, a_3, a_4), four triangles may be constructed ($a_1 a_2 a_3, a_2 a_3 a_4, a_3 a_4 a_1, a_4 a_1 a_2$), the areas of these triangles can be found from

$$A_1 = \sqrt{\left(s_1(s_1 - |a_1 a_2|)(s_1 - |a_2 a_3|)(s_1 - |a_3 a_1|)\right)},\ s_1 = (|a_1 a_2| + |a_2 a_3| + |a_3 a_1|)/2,$$

$$A_2 = \sqrt{\left(s_2(s_2 - |a_2 a_3|)(s_2 - |a_3 a_4|)(s_2 - |a_4 a_2|)\right)},\ s_2 = (|a_2 a_3| + |a_3 a_4| + |a_4 a_2|)/2,$$

$$A_3 = \sqrt{\left(s_3(s_3 - |a_3 a_4|)(s_3 - |a_4 a_1|)(s_3 - |a_1 a_4|)\right)},\ s_3 = (|a_3 a_4| + |a_4 a_1| + |a_1 a_3|)/2,$$

$$A_4 = \sqrt{(s_4(s_4 - |a_4 a_1|)(s_4 - |a_1 a_2|)(s_4 - |a_2 a_4|))},\ s_4 = (|a_4 a_1| + |a_1 a_2| + |a_2 a_4|)/2 \quad .$$

Area ratios can then be calculated as

$$A_{1/2} = \frac{A_1}{A_2},$$

and the distances between an area ratio $A_{1/2}$ of image *A* and an area ratio $B_{1/2}$ for image *B* can then be found by

$$D\left(A_{1/2}, B_{1/2}\right) = \frac{|A_{1/2} - B_{1/2}|}{A_{1/2} + B_{1/2}}.$$

Two thresholds T_c and T_D should be applied to the set of four area ratio distances between the feature sets. The first threshold rejects hypothesis which have any of the area ratios distance above a first threshold T_c. The second rejects hypothesis for which the sum of distances is larger than a threshold T_D.

Table AMD5.1 — Suggested threshold values for feature descriptor matching

Threshold	**Suggested Value**
T_A	15
T_B	38
T_C	0.2
T_D	0.55
q	1
G_1, G_2 and G_3	0.06
$G_1 + G_2$, $G_2 + G_3$ and $G_3 + G_1$	0.1

4.9.1.2.3 Descriptor Hashing

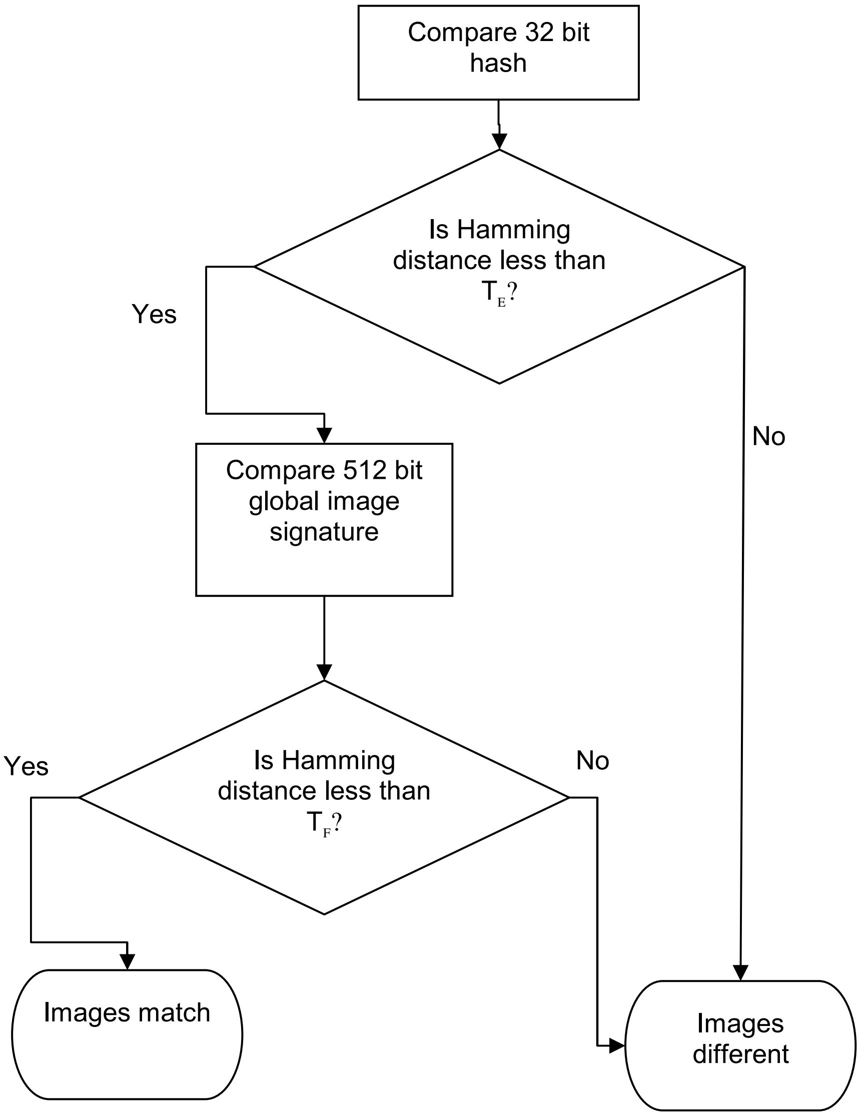

Figure AMD5.2 — Matching global image signatures using the 32 bit hash.

A 32 bit hash of the 512 bit global signatures can offer a significant improvement in search speeds, using the process shown in Figure AMD5.1. A 32 bit hash of the global image signatures can be chosen by numerous methods e.g. i) random selection, ii) by analysing robustness to specific modifications or iii) analysing the binary entropy of each bit on a large, independent dataset. By analysing and selecting the highest entropy bits 32 suitable bits are chosen, the bitmask is given below.

```
0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0
,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,
1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0
```


These 32 bits are taken from amongst the 512 bits of the global image signature during the matching stage, hence they do not in anyway affect the extraction or storage of the signatures.

Similarly an 8 bit hash can be formed for the 60 bit feature signature which again allows for significant improvements in search speeds. A suitable bitmask for the feature signature is given below.

```
0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0
,0,0,0,1,0,0,0,0,0,0,0,0,0
```

Suitable values for the thresholds T_E and T_F are 12 and 148 respectively, these correspond to a false alarm rate of less than 0.05ppm. For the 8 bit feature hash a suitable threshold is 3 with T_A the threshold for the complete signature.

4.9.1.3 Conditions of Usage

The image signature provides a unique, robust and compact representation of an image. It allows a user to rapidly detect duplicate, modified versions and near duplicates of an image.

