TECHNICAL REPORT

ISO/IEC TR 15938-8

Information technology — Multimedia content description interface —

Part 8:
Extraction and use of MPEG-7 descriptions

AMENDMENT 5: Extraction and matching of image signature tools

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

Partie 8: Extraction et utilisation des descriptions MPEG-7

AMENDEMENT 5: Extraction et correspondance des outils de signature d'image

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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

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

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

Amendment 5 to ISO/IEC TR 15938-8:2002 was prepared by Technical Committee ISO/IEC 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 use of MPEG-7 descriptions

AMENDMENT 5: Extraction and matching of image signature tools

Add after 4.8:

4.9 Visual Signatures

4.9.1 Image Signature

The visual content descriptors in ISO/IEC 15938-3 clauses 6 to 9 are very useful when trying to find images with *similar* content. However, such descriptors are intended to be general and are found to be unsuitable for the task of finding duplicate images. 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.

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

4.9.1.1 Feature Extraction

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

4.9.1.2 Similarity Matching

4.9.1.2.1 Global Signatures

To perform matching between two global signatures B_1 and B_2 both of length N, the Hamming distance should be 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 for independent images 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 1 ppm.

4.9.1.2.2   Local Signatures

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
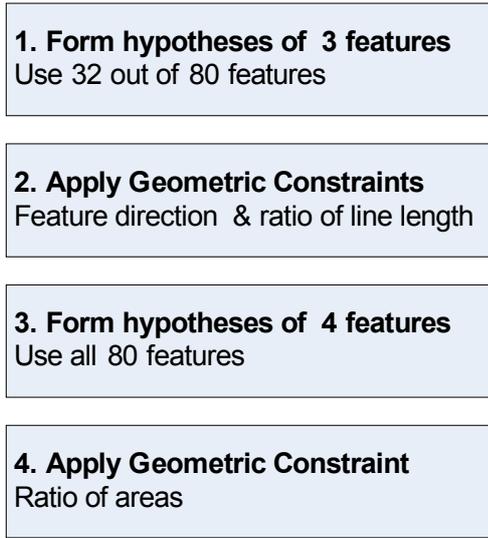
4. Apply Geometric Constraint
Ratio of areas

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

Matching images can be performed by comparing the local signatures which form a part of the image signature. For efficiency four stages can be used, as shown in Figure AMD5.1, to match local signatures from a pair of images. Hypotheses can be formed in stages one and three. A series of geometric test are performed in stages two and four, these tests should 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.

The first stage involves forming hypotheses for potential matching feature points by comparing the first 32 feature points from a first image with the first 32 feature points from a second image. To perform matching between feature points the local signatures (B_1 and B_2) for the feature points can be compared using the Hamming distance:

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

A list of feature point 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 feature points from the list. A hypothesis is made up of a set of three pairs of feature points $((a_1, b_1), (a_2, b_2), (a_3, b_3))$, where a_1, a_2, a_3 are feature points from a first image and b_1, b_2, b_3 are feature points from a second image. A set of three pairs of feature points is declared a match if the cumulative Hamming distance between the local signatures, of corresponding feature points, is below a constant threshold T_B. The set of hypotheses is ordered by their cumulative distance, with the lowest distance (i.e. highest likelihood 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 can be applied to reduce false matches, i) feature direction and ii) the ratio of line lengths.

A direction, which is quantized into 16 levels as defined in ISO/IEC 15938-3:2002/Amd.3:2009, 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

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

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. The angle distances may also be found between the other pairs of feature points, for example

$$\Omega_2 = \left| \left(\theta_{a2} - \theta_{a3} \right) - \left(\theta_{b2} - \theta_{b3} \right) \right| \text{ and } \Lambda_2 = \left| \left(\theta_{a2} - \theta_{a3} \right) + \left(\theta_{b2} - \theta_{b3} \right) \right|,$$

$$\Omega_3 = \left| \left(\theta_{a3} - \theta_{a1} \right) - \left(\theta_{b3} - \theta_{b1} \right) \right| \text{ and } \Lambda_3 = \left| \left(\theta_{a3} - \theta_{a1} \right) + \left(\theta_{b3} - \theta_{b1} \right) \right|.$$

Using the locations of features selected in each hypothesis a ratio of line lengths between the features is used as a constraint. By $\left| a_1 a_2 \right|$ we denote the Euclidean 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

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

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

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

where L_n^A denotes the nth line length ratio from image A and

L_n^B denotes the nth line length ratio from image B.

Two thresholds can be applied to the distance measures, one T_G to constrain the magnitude of the individual measures and the other T_H 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 can be 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 Hamming distance of the local signatures for the fourth feature.

In stage four geometric analysis can be 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

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

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

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

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

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

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

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 signature matching

Threshold	Suggested Value
T_A	15
T_B	38
T_C	0.2
T_D	0.55
q	1
T_G	0.06
T_H	0.1

4.9.1.2.3   Signature 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 AMD5.2. 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
,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0
,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,
1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0
```