THE COLOUR IN THE UPCOMING MPEG-7 STANDARD

Charilaos Christopoulos¹, Daniel Berg¹ and Athanassios Skodras²

¹MediaLab, Ericsson Radio Systems AB, SE-164 80 Stockholm, SWEDEN
E-mail: {charilaos.christopoulos, daniel.berg}@era.ericsson.se

²Electronics Laboratory, University of Patras, GR-261 10 Patras, GREECE
E-mail: skodras@cti.gr

ABSTRACT
The colour spaces supported by the different image and video coding standards are presented in the present communication. Extensive reference to the colour spaces and the relevant colour descriptors supported by the MPEG-7 standard is given, and experiments on colour-based image retrieval efficiency are illustrated.

Keywords: colour spaces, image retrieval, content-based search, feature histogram, MPEG-7

1 INTRODUCTION
The representation of colour is a very important issue in colour image processing and in image compression in particular. A lot of effort has been devoted to this during the development of the still and moving picture standards, such as the JPEG, MPEG and H.26x. Colour representation plays a crucial role in lossy and lossless compression efficiency, in image retrieval efficacy and in the required computational complexity. The monochrome, RGB and YC₁C₂ colour spaces are those supported by all coding standards. The conversion from the RGB to the YC₁C₂ colour space is a linear non-reversible (i.e. lossy) transformation (Appendix I). In order for the JPEG2000 standard to achieve lossless compression of colour images, a new reversible colour transformation has been defined (Appendix II). In the absence of quantisation errors this decorrelating transformation introduces no errors (reversible / lossless). Three goals are achieved by this transformation, namely colour decorrelation for efficient compression, reasonable colour space with respect to the human visual system for quantisation and ability of having lossless compression, i.e. exact reconstruction with finite integer precision [1,2].

While MPEG-1 and MPEG-2 concentrated almost entirely on compression, MPEG-4 moved to a higher level of abstraction in coding objects and using content-specific techniques for coding content. MPEG-7 has moved to an even higher level of abstraction, i.e. to cognitive coding. In other words, MPEG-1, -2 and -4 made content available, while MPEG-7 allows the localization of the required content. Colour is the main visual feature, along with texture, shape and motion, towards content localization [3,4].

MPEG-7 supports three more colour spaces, in addition to monochrome, RGB and YC₁C₂. These are the HSV, the HMMD and the linear transformation matrix with reference to the RGB (see Appendices III to V respectively). Both the HSV and the HMMD represent non-linear transformations. They provide better results in search and retrieval applications, due to the fact that they are closely related to the human perception of colour. With the inclusion of an arbitrary linear transformation with respect to RGB, a large number of additional colour spaces are supported. An overview of the colour spaces supported by each of the coding standards is depicted in Table I.

| Table I: Colour spaces supported by the coding standards |
|-----------------|-----------------|-----------------|-----------------|
|                 | JPEG | JPEG2000 | MPEG | MPEG |
| Monochrome      | ✓    | ✓        | ✓    | ✓    |
| RGB             | ✓    | ✓        | ✓    | ✓    |
| YC₁C₂           | ✓    | ✓        | ✓    | ✓    |
| YUV             | ✓    | ✓        | ✓    | ✓    |
| HSV             | ✓    | ✓        | ✓    | ✓    |
| HMMD            | ✓    | ✓        | ✓    | ✓    |
| Linear Matrix   | ✓    | ✓        | ✓    | ✓    |

In the present communication the importance of colour in content-based image retrieval is examined. In section 2 an outline of the different colour descriptors supported by the MPEG-7 is given. The performance of these descriptors in content-based image localization in a non-annotated image database is demonstrated in section 3. Finally, conclusions are drawn in section 4.

2 THE COLOUR DESCRIPTORS IN THE MPEG-7
In the upcoming ISO/IEC standard of MPEG-7, multimedia is tagged with information regarding its content and origin. MPEG-7 is formally named “Multimedia Content Description Interface” and is
scheduled to become an international standard in July 2001. It standardises Descriptors for defining syntax and semantics of each feature representation and Description Schemes for specifying relationships between components (both Descriptors and Descriptor Schemes). MPEG-7 automates the process of extracting features from the multimedia. In the MPEG-7 eXperimental Model (XM), tools for feature extraction and multimedia search using various algorithms are included. Each method can extract feature information from media files in a database and match it with features of a reference image or movie clip.

The data is structured using the MPEG-7 Description Definition Language (DDL), which is an XML based mark-up language. The DDL information is written to a MPEG-7 bitstream file. When a search is performed, features are extracted from the reference image / movie / audio clip and compared to the features listed in the bitstream file [4,5]. Below, the Descriptors based on colour information are outlined.

2.1 Colour Descriptors

Eight colour descriptors are supported by the MPEG-7, namely colour space, dominant colours, colour histogram, colour quantisation, group of frames / group of pictures colour histogram, colour-structure histogram, colour layout and Haar transformed binary histogram. Those that have been used in our experiments are briefly presented below:

2.1.1 Colour Space

This descriptor defines the colour space to be used in a certain application, usually in combination with other descriptors such as dominant colour or colour histogram. The colour spaces have been already presented in section 1.

2.1.2 Colour Quantisation

This Descriptor defines the quantisation of a colour space. Quantisation is the reduction of the number of unique colours in an image. The linear, non-linear and lookup table quantisation types are supported. In the linear case, the normalised colour value range is divided into equal intervals. Each quantised colour is represented by a colour value index that can be decoded to the correct colour value in three components according to the quantisation type used. When a lookup table is used it can be done instantly, otherwise it has to be calculated from predefined formulas and the current index. The Descriptor is used in the Colour Histogram, the Colour Structure Histogram and the Colour-Structure Histogram Descriptors below.

2.1.3 Colour Histogram

The colour histogram descriptor is a compound descriptor that expresses the colour features by means of a histogram. The extraction is performed by sorting the quantised colour pixel values into colour bins that represent the colour histogram. The matching is done by calculating the distance as the difference between the normalised histograms. The best match is the one with the lowest weighted difference between the corresponding colour histogram bins.

2.1.4 Dominant Colour(s)

This colour descriptor is best suitable for representing local (object or image region) features where a small number of colours is enough to characterize the colour information in the region of interest. Whole images are also applicable, for example, flag images or colour trademark images. Colour quantization is used to extract a small number of representing colours in each region/image. The percentage of each quantized colour in the region is calculated correspondingly. A confidence measure on the entire descriptor is also defined, and is used in similarity retrieval.

2.1.5 Colour Layout

This descriptor is designed to capture spatial distribution of colour for either whole image or any part of the image. It needs only 64 bits in default setting to describe spatial distribution and its similarity calculation process is simple enough to achieve very high-speed retrieval. This descriptor allows not only natural pictures but also sketches as queries. Any other colour descriptors do not support the sketch queries, which is very useful for user friendly interface. The descriptor consists of DCT coefficients of an 8x8 icon, which is defined as a set of dominant colours on 8x8-grid layout. The number of coefficients included in the descriptor provides the resolution scalability. The descriptor is suitable especially for video browsing and retrieving applications, like a home video server.

2.1.6 Colour-Structure Histogram

The colour-structure histogram descriptor is a colour feature descriptor and intended for still image retrieval, i.e. the main functionality is image-to-image matching. The colour-structure histogram descriptor embeds local colour structure information into the histogram; that is, the extraction method takes into account the colours in a local neighborhood of pixels, instead of considering each pixel separately. The colour-structure histogram descriptor provides additional functionality and improved similarity-based retrieval performance compared to the regular colour histogram descriptor for natural images.

3 COMPARATIVE RESULTS

Tests were performed using a database with 2045 JPEG images depicting various scenes. The reference image, shown in Fig. 1, was one of the database images (Img01155.jpg). Colour Histogram and Colour Layout are the only colour feature extraction and search algorithms already fully implemented in the XM. Furthermore, the Colour Quantization Descriptor is implemented and used by the Colour Histogram Descriptor. For comparison reasons we conducted also experiments using the Edge Descriptor, which belongs in the class of texture descriptors and the Region Shape Descriptor, which belongs in the class of shape
descriptors. It was seen that the image retrieval efficiency was similar to that achieved by the colour layout descriptor only.

![Query image](Img01155.jpg)

**Figure 1.** Query image (Img01155.jpg)

### 3.1 Colour Histogram

As described above this algorithm checks only the amount of different colours in the image or frame. It does not take into account where colours are found. The search result (four best matches) based on the Colour Histogram is shown in Fig. 2. Apart from the reference image itself, none of the search result images with high ranking depicted the same scene. They were sunsets over other scenes.

![Search result based on Colour Histogram](Img00138.jpg, Img00135.jpg, Img00136.jpg)

**Figure 2.** Search result based on Colour Histogram

### 3.2 Colour Layout

The Colour Layout Descriptor takes into account where in the image or frame the different colours are found. Therefore, it appears that the colour distribution as represented by the colour layout descriptor is more important than the Colour characteristics as they are represented by the Colour Histogram Descriptor. This becomes clear when comparing the search result above (Fig. 2) with the following four best matches from the Colour Layout search (Fig. 3).

![Search result based on Colour Layout](Img00273.jpg, Img00135.jpg, Img01154.jpg)

**Figure 3.** Search result based on Colour Layout

### 4 CONCLUSIONS

In the present communication an overview of the colour spaces supported by the image and video coding standards has been presented. Special emphasis to the MPEG-7 has been given, since colour is one of its main features for searching of content in non-annotated image databases. It has been seen that the colour histogram feature alone is not adequate for the correct image retrieval, mainly due to the lack of all structure information about the image [3,6]. The use of the colour layout feature has led to impressive image retrieval efficiency. This has been verified further by extending our experiments for the case of image retrieval based on the Edge Histogram (i.e. texture feature) and Region Shape (i.e. shape feature). Results have shown that the retrieval efficiency was similar to that of Colour Layout only. Such capabilities can be used not only for image search applications, but also for video summary generation [7,8]. For example, colour histograms are widely used in key frame extraction and shot boundary detection.

**References**


[5] A. B. Benitez, S. Paek and S.-F. Chang, °Object-Based Multimedia Description Schemes and


Appendix I
RGB to YCbCr colour transformation (non-reversible):

\[
\begin{pmatrix}
Y \\
C_b \\
C_r
\end{pmatrix} = \begin{pmatrix}
0.299 & 0.587 & 0.114 \\
-0.16775 & -0.33126 & 0.5 \\
0.5 & -0.41699 & -0.08131
\end{pmatrix} \begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
\]

Appendix II
RGB to YUV and YUV to RGB colour transformations (reversible):

\[
\begin{pmatrix}
Y' \\
U' \\
V'
\end{pmatrix} = \begin{pmatrix}
0.299 & 0.587 & 0.114 \\
-0.16775 & -0.33126 & 0.5 \\
0.5 & -0.41699 & -0.08131
\end{pmatrix} \begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
\]

Appendix III
HSV expresses a non-linear transformation. It consists of the "value" v representing the lightness of the colour, the "saturation" s indicating how dominant a colour is, and the "hue" h representing the dominant spectral tone of the colour. Their values are derived from the normalised R,G,B values (ranging from 0 to 1) as follows:

\[
v = \max(R, G, B); \quad s = (v - \min(R, G, B)) / v; \quad h = \begin{cases} 0 & \text{if } (R = \max(R, G, B) \land G = \min(R, G, B)) \\ \frac{\min(R - G, B - G)}{2} & \text{else if } (R = \max(R, G, B) \land B = \min(R, G, B)) \\ \frac{\min(B - G, G - R)}{2} & \text{else if } (G = \max(R, G, B) \land R = \min(R, G, B)) \\ 0 & \text{else if } (B = \max(R, G, B) \land R = \min(R, G, B)) \end{cases}
\]

Appendix IV
HMMD has a double cone shape appearance consisting of blackness, whiteness, colourfulness and hue. There are five distinct attributes (components) in the HMMD colour space. However, three of them - Hue, Max or Min and Diff and Sum - are enough to define the colour space. The semantics of the six attributes are as follows:

\[
\text{Hue} : \text{Hue} \text{can be represented by an angle (0° to 360°) in the hue circle. When the angle increases, hue changes from red (0°), yellow (60°), green (120°), blue (240°) to red (360° = 0°).}
\]

\[
\text{Max} : \text{Max indicates how much black colour it has, giving the flavor of shade or blackness.}
\]

\[
\text{Min} : \text{Min indicates how much white colour it has, giving the flavor of tint or whiteness.}
\]

\[
\text{Diff} : \text{Diff indicates how much gray it contains and how close to the pure colour, giving the flavor of tone or colourfulness.}
\]

\[
\text{Sum} : \text{Sum simulates the brightness of the colour.}
\]

The transformation from RGB to HMMD is non-linear and reversible. If Max, Min, Diff and Sum is from 0 to 1 and Hue is from 0 to 360 then Max = max(R, G, B), Min = min(R, G, B), Diff = Max - Min, Sum = (Max + Min)/2. If Max = Min, Hue is undefined (for achromatic colour), otherwise

\[
\text{Hue} = \begin{cases} \frac{G - B}{\text{Max} - \text{Min}} \\ \frac{G - B}{\text{Max} - \text{Min}} \end{cases} \begin{cases} \text{60 if } (R = \text{Max} \land (G - B) > 0) \\ \text{60 + 360 if } (R = \text{Max} \land (G - B) < 0) \end{cases}
\]

Appendix V
RGB to linear matrix transformation:

\[
\begin{pmatrix}
C_1 \\
C_2 \\
C_3
\end{pmatrix} = \begin{pmatrix}
a_{00} & a_{01} & a_{02} \\
a_{10} & a_{11} & a_{12} \\
a_{20} & a_{21} & a_{22}
\end{pmatrix} \begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
\]