# Color features performance comparison for image retrieval

Daniele Borghesani, Costantino Grana, Rita Cucchiara

Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Modena e Reggio Emilia, Via Vignolese 905/b, 41100 Modena, Italy
{daniele.borghesani, costantino.grana, rita.cucchiara}@unimore.it

**Abstract.** This paper proposes a comparison of color features for image retrieval. In particular the UCID image database has been employed to compare the retrieval capabilities of different color descriptors. The set of descriptors comprises global and spatially related features, and the tests show that HSV based global features provide the best performance at varying brightness and contrast settings.

**Keywords:** color features, HSV, image retrieval, feature comparison

## 1 Introduction

The increasing availability of multimedia digital libraries (publicly shared or personal) and low cost devices to produce them, raised the need of appropriate tools for the search within this enormous amount of data. Classical search methodologies in desktop and web contexts are based on textual keywords. In order to reuse the majority of all preexisting searching techniques with multimedia data, the most immediate solution is tagging, but it is generally boring from the user prospective and unfeasible if the amount of data to annotate is too high. The search and retrieval based on content (CBIR) is the most difficult but at the same time the most effective and elegant way to solve the problem. A lot of literature background has been produced so far, focusing on specific components (like the learning algorithms, the features to use, the way to select the most effective features, etc…), sometimes specialized on some real world context (news, sports, etc…). A lot of functioning systems have also been proposed. Columbia University proposed its semantic video search engine (CuVid [1]), including 374 visual concept detectors and using different combinations of input modalities (keyword, image, near-duplicate and semantic concepts). IBM multimedia search and video system (Marvel [2]) uses multimodal machine learning techniques for bridging the semantic gap, recognizing entities such as scenes, objects, events and people. University of Amsterdam also proposed a semantic video search system [3] featuring a fascinating user interface (RotorBrowser) and 500 automatically detected semantic concepts. We also proposed a general framework called PEANO (Pictorially Enriched ANnotation with Ontologies) which allows to automatically annotate video clips by comparing their similarity to a domain specific set of prototypes [4].

Considering the way scientific community tried to solve the problem, we can highlight two fundamental functionalities we would like in a CBIR system:

- The ability to search and retrieve specific visual objects: given an image, we want to retrieve in our digital library all images containing the object depicted within the query, that is "I want to find my friend John in all my image library"
- The ability to search and retrieve images by appearance similarity: given a sample image or a keyword (textual representation of a pictorial prototype), we want to retrieve the most similar images to the query, that is "I like this seaside landscape, I want to find all similar images I've got in my image library".

The global appearance similarity task, especially if it is fast to compute, has also an important side effect, that is the possibility to prune from the digital library all images that will not likely matter for other more specific retrieval techniques. This is a major advantage even because usually the local features exploited for the objects recognition or other more sophisticated global features, and the learning algorithms employed, are quite weighty to compute.

The most straightforward representation for global features is the histogram. It allows a scale independent representation, suitable both for color and gradient information, and it has a robust and easy metric to evaluate similarity, that is histogram intersection. Besides, one of the most discriminative characteristic for global features is undoubtedly the color. It brings information about the nature of what we see, it allows inferences about environment depending on brightness conditions, and the way humans perceive the chromatic appearance helps recognition process of an object and changes understanding of the environment self.

In this paper, we aim at analyzing in detail the discriminative capabilities of several well known color descriptors. We adopt a standard image database called UCIDv.2, freely available and complete with ground truth. To test their performance, we modify the brightness characteristic of images in order to test the behavior of these features in extreme conditions.

## 2   Color features for image retrieval

The standard definition of HSV histogram is proposed as MPEG-7 scalable color description [5]: a 256-bins histogram providing 16 different values for H, and 4 different values for S and V. Normally, H is defined in the range [0..360], while S and V are defined in [0..1]. HSV36 has been presented in [6]. This procedure aims to improve the representation power of HSV color space. Firstly a non uniform quantization of hue has been introduced, dividing the entire spectrum into 7 classes: red, orange, yellow, green, cyan, blue and purple. Then another quantization has been proposed for S-V plane, in order to distinguish the black area (V <= 0.2), and the grey area (S <= 0.2), providing a fixed general subdivision of the chromatic  area of the plane into 4 subparts.

The RGB histogram is a standard color histogram based on RGB components of the image. Each component is subsampled into 8 values, resulting in a 512-bin histogram.

Firstly introduced by [7], the color correlogram describes the spatial correlation of colors using an arbitrary spatial distance. It consists in a table indexed by color couple $(c_i, c_j)$ so that the $k$-th entry designates the probability $p$ of finding a pixel of color $c_j$ at a distance $k$ from a pixel of color $c_i$ in the image. So let $I$ be an $n$ x $n$ image, with $m$ quantized colors $c_1, \ldots, c_m$ and let a distance $d \in [n]$ be fixed a priori. The correlogram of $I$ is defined for $i, j \in [m], k \in [d]$ as

$$\gamma_{c_i, c_j}^{(k)}(I) \triangleq P\left\{\exists\, p_1 \in I_{c_i}, p_2 \in I_{c_j} : \left|p_1 - p_2\right| = k\right\} \tag{1}$$

In particular autocorrelograms are a subset of correlograms capturing the spatial correlation between identical colors only:

$$\alpha_c^{(k)}(I) \triangleq \gamma_{c,c}^{(k)}(I) \tag{2}$$

Correlograms may be computed on RGB or HSV color information [8].

Spatial chromatic histograms, were introduced in [9]. This feature proposes to merge spatial information with color information, in order to take in account spatial distribution of color blobs along the image: for each color bin add the center of mass and the average distance from it. The similarity between two queries is given by:

$$f_s(Q, I) = \sum_{i=1}^{c} \min(h_Q(i), h_I(i)) \cdot \left( \frac{\sqrt{2} - d(b_Q(i), b_I(i))}{\sqrt{2}} + \frac{\min(\sigma_Q(i), \sigma_I(i))}{\max(\sigma_Q(i), \sigma_I(i))} \right) \tag{3}$$

The similarity function $f_s$ acts also as a weighting function, increasing importance of large areas with equally colored pixels.

Color opponent process [10] is a well known color theory that exploits, instead of raw RGB channel information, color cones in an antagonistic matter. The motivation is that their respective wavelengths have some overlap, so it's more efficient to record differences between these cones than the original entire signal. The theory proposes three opponent channels: reddish green (red versus green), yellowish blue (blue versus yellow) and luminance (black versus white). An original 24 bin RGB pixel information can now be described as a compressed 8 bit color opponent pixel information, using 3 most significant bits for chrominance components and 2 most significant bits for luminance component. A 256-bins histogram may be computed to provide a color opponent histogram.

Edge detection is very used in content based retrieval: edge features locally extract more significant points in the image, that are perceptually more relevant from a human point of view, using only brightness information. A common and effective way to represent edge features is edge histogram [11], which globally describes the frequency and the directionality of edges and (as histogram itself) it's invariant to translations, rotations and scaling.

## 3   HSV histogram with achromatic selection

Regarding achromaticity issues in HSV color space, different solutions have been proposed. In [12] two regions are defined and separately treated, one for the chromatic and one for the achromatic colors. These areas are obtained with a complicated set of thresholds in the IHS color space. Similarly, in [13] a fuzzy technique has been proposed in order to distinguish among chromatic, achromatic, and black colors. Sural et al. [14] propose a histogram modification that takes into account the above mentioned regions. In particular, they identify the achromatic region by thresholding the saturation coordinate with a linear function of the intensity value and based on the outcome chose to represent the color with its hue or its value only. In [15] a detailed comparison of the MPEG-7 color descriptors can be found, proving that the Scalable Color Descriptor is not suitable for monochromatic images. Moreover in [6] an arbitrary subdivision of HSV color space has been proposed, in order to select an achromatic region and quantize the remaining chromatic one.

To solve the problem of achromatic images, we proposed an enhanced version of the HSV color space [16]. The Scalable Color Descriptor requires a quantization of the HSV color space, with 16 values in H and 4 values in each S and V (256 bins in total). Supposing every color channel in the range [0,1), the bin index may be obtained as:

$$\text{bin} = f\left(H, S, V\right) = \lfloor n_H H \rfloor n_S n_V + \lfloor n_S S \rfloor n_V + \lfloor n_V V \rfloor \qquad (4)$$

where $n_H$, $n_S$, $n_V$ are the quantization levels devoted to every color channel. Usually these are chosen to be powers of 2 for ease of representation.

Adopting, as is usual, a linear quantization of each coordinate leads to have, for example, 64 different bins for the darkest colors characterized by the lowest values of V. A possible error could be the erroneous assignment to different bins with dark color: for this reason a visually uniform background can be split on different bins. We propose to add $n_A$ bins to the HSV histogram that contains all the achromatic and dark colors. These $n_A$ bins correspond to gray levels, from black to white; for convenience, we choose to set $n_A = n_V = 4$, as the number of levels assigned to the V axis in the MPEG-7 standard. The dark and achromatic colors are selected by imposing a unique threshold $\lambda$ on the S and V coordinates respectively. The value of $\lambda$ has been empirically set.

Moving some of the colors from the original bins to the $n_A$ achromatic ones makes these original bins less used with respect to the others. In fact, it doesn't make sense anymore to uniformly subdivide the S and V channels, if part of it is then discarded. A better solution is to fully employ the chromatic bins to describe only the effective chromatic area (Fig. 1). To this aim we linearly quantize the remaining HSV space, by simply redefining the $f$ function of Eq. 4:

$$f'\left(H, S, V\right) = f\left(H, \frac{S - \lambda}{1 - \lambda}, \frac{V - \lambda}{1 - \lambda}\right) \qquad (5)$$

A different approach could be used to reduce the number of bins in the histogram, without affecting the chromatic area. The threshold $\lambda$ can be set to $1/n_V$, thus
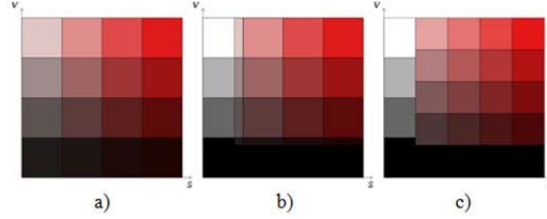
**Fig. 1.** a) original quantization of the SV plane with H = 0 (red); b) achromatic area detection (λ=0.2); d) Linear re-quantization of the chromatic area.

making the achromatic area exactly match the first set of bins for S and V. This forces these bins to 0, thus allowing their removal. This indeed induces a compression with respect to the color representation, but it is selectively applied to the least significant colors. For example with reference to the aforementioned 16,4,4 subdivision, this would lead to $16 \times 3 \times 3 = 144$ bins, plus 4 bins for the achromatic area.

HSV color space drawbacks are mostly significant in real world images and the consequent introduction of non-uniform brightness distribution. These phenomena is due to unprofessional cameras or indoor/outdoor situations, so the achromatic separation provided with aHSV can focus the descriptive power to the chromatic area.

## 4   Experimental results

The database used for our test was UCIDv.2 (UnCompressed Image Database), provided by [17]. It contains 1338 real world 512x384 uncompressed images, together with a ground truth: for each query, a variable number of models is indicated. To obtain a more significant corpus of data and perform our tests in different working conditions, we modified the original database producing 25 different versions. Brightness and/or contrast values have been varied in a systematic way in order to simulate different light conditions, producing a total amount of 1338x25 images.

To evaluate the performance of the system, we used the same metric indicated in [17], that is a modified version of match percentile metric:

$$MP_Q = \frac{100}{S_Q} \sum_{i=1}^{S_Q} \frac{N - R_i}{N - i}, \quad \text{with } R_i < R_{i+1} \tag{6}$$

Where $S_Q$ is the number of models for query $Q$, $N$ is the total number of images in the database, $i$ is the rank of the image in the ground truth and $R_i$ is the rank of the image computed by the system. The average match percentile may then used as a global measure of performance of a retrieval system.

The first two columns of the Table 1 report the percentage of decrease in brightness and contrast for the entire database. HSV based histograms perform better than other features in all cases, and with different light conditions proposed. It performs better that standard RGB histograms, and it's more stable at varying light conditions, due to the nature of HSV histograms.

**Table. 1.** Comparison of different color features. HSV provides the best performance.

| Brightness | Contrast | HSV | RGB | HistColOpp | Gray | SpatialHSV | SpatialRGB | cHSV | cRGB | EH |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 92,72 | 91,05 | 90,29 | 76,16 | 91,05 | 89,44 | 88,21 | 87,26 | 76,28 |
| 0 | -20 | 92,66 | 90,11 | 89,93 | 76,35 | 90,53 | 88,72 | 88,71 | 87,17 | 75,83 |
| 0 | -40 | 92,33 | 89,55 | 89,4 | 76,34 | 89,35 | 86,15 | 88,43 | 86,63 | 75,18 |
| 0 | -60 | 92,24 | 88,04 | 88,98 | 76,45 | 85,74 | 83,4 | 87,67 | 82,78 | 74,58 |
| 0 | -80 | 91,69 | 81,61 | 88,09 | 76,5 | 77,17 | 83,37 | 86,21 | 77,4 | 73,64 |
| -20 | 0 | 92,71 | 90,34 | 89,72 | 76,26 | 90,34 | 88,48 | 88,2 | 86,97 | 76,01 |
| -20 | -20 | 92,72 | 90,47 | 89,76 | 76,35 | 90,36 | 88,09 | 89,08 | 86,65 | 75,76 |
| -20 | -40 | 92,42 | 89,3 | 89,45 | 76,34 | 90,18 | 87,44 | 88,62 | 85,05 | 75,19 |
| -20 | -60 | 91,99 | 87,92 | 89,14 | 76,45 | 87,58 | 82,9 | 88,29 | 84,14 | 74,62 |
| -20 | -80 | 89,87 | 83,26 | 88,04 | 76,5 | 78,77 | 74,13 | 85,98 | 79,24 | 73,65 |
| -40 | 0 | 91,84 | 89,54 | 89,14 | 76,51 | 89,83 | 87,15 | 87,67 | 86,27 | 75,59 |
| -40 | -20 | 92,81 | 89,89 | 89,71 | 76,38 | 89,62 | 87,16 | 88,42 | 85,59 | 76,03 |
| -40 | -40 | 92,84 | 89,65 | 89,07 | 76,34 | 90,07 | 86,97 | 88,87 | 86,45 | 75,18 |
| -40 | -60 | 91,77 | 87,62 | 89,16 | 76,45 | 89,1 | 84,16 | 88,12 | 83,33 | 74,62 |
| -40 | -80 | 90,49 | 81,33 | 87,42 | 76,5 | 81,79 | 81,33 | 85,99 | 80,18 | 73,74 |
| -60 | 0 | 91,31 | 88,36 | 88,4 | 76,5 | 88,25 | 85,53 | 87,18 | 86,09 | 75,75 |
| -60 | -20 | 92,14 | 88,33 | 88,66 | 76,73 | 88,89 | 85,05 | 87,02 | 85,78 | 74,99 |
| -60 | -40 | 92,68 | 88,7 | 89,26 | 76,33 | 89,37 | 84,22 | 88,38 | 84,91 | 75,4 |
| -60 | -60 | 92,67 | 88,12 | 88,78 | 76,45 | 88,98 | 83,44 | 88,68 | 82,74 | 74,58 |
| -60 | -80 | 92,28 | 82,36 | 88,42 | 76,5 | 85,29 | 84,19 | 87,42 | 78,23 | 73,64 |
| -80 | 0 | 90,34 | 86,42 | 86,93 | 76,62 | 87,34 | 84,78 | 86,46 | 84,61 | 74,78 |
| -80 | -20 | 91,03 | 87,26 | 87,36 | 76,95 | 87,4 | 84,1 | 87 | 84,02 | 75,03 |
| -80 | -40 | 91,92 | 87,22 | 88,15 | 76,71 | 87,19 | 82,5 | 86,86 | 83,37 | 75,18 |
| -80 | -60 | 92,7 | 86,5 | 89,11 | 76,41 | 86,79 | 80,89 | 88,67 | 83,92 | 74,57 |
| -80 | -80 | 91,39 | 84,27 | 87,56 | 76,5 | 84 | 76,74 | 87,85 | 79 | 73,76 |
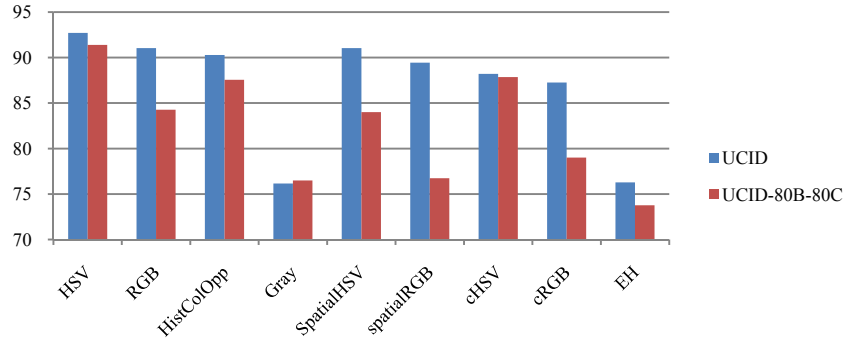


Gray histograms (that carries information only about the brightness) provide poor performance. The use of spatial information (with RGB and HSV histograms) have also proved to be less effective than using color histograms only. Their dependency on the color distribution, even if very useful for the classification of particular images with a similar color distribution, is not always useful for global appearance similarity.

We tested color correlograms and edge histograms, in order to capture the aggregation level and the salient edges in the image respectively, but none of them reached the same performance of HSV histograms. Both of them are intrinsically

**Table. 2.** Comparison of different HSV color features. B and C indicates respectively the decrease of brightness and contrast in the modified images.

| B | C | HSV | AHSV | HSV36 |
|---|---|---|---|---|
| 0 | 0 | 92,72 | 93,25 | 90,49 |
| 0 | -20 | 92,66 | 92,93 | 89,97 |
| 0 | -40 | 92,33 | 93 | 88,63 |
| 0 | -60 | 92,24 | 91,69 | 84,38 |
| 0 | -80 | 91,69 | 88,18 | 73,67 |
| -20 | 0 | 92,71 | 93,1 | 89,73 |
| -20 | -20 | 92,72 | 93,06 | 89,79 |
| -20 | -40 | 92,42 | 93,01 | 89,55 |
| -20 | -60 | 91,99 | 92,27 | 86,56 |
| -20 | -80 | 89,87 | 88,61 | 76,61 |
| -40 | 0 | 91,84 | 91,97 | 88,78 |
| -40 | -20 | 92,81 | 93,14 | 89,02 |
| -40 | -40 | 92,84 | 93,17 | 89,38 |
| -40 | -60 | 91,77 | 92,83 | 88,46 |
| -40 | -80 | 90,49 | 89,8 | 79,86 |
| -60 | 0 | 91,31 | 90,93 | 87,68 |
| -60 | -20 | 92,14 | 92,28 | 88,32 |
| -60 | -40 | 92,68 | 93,04 | 88,7 |
| -60 | -60 | 92,67 | 92,76 | 88,05 |
| -60 | -80 | 92,28 | 91,52 | 83,66 |
| -80 | 0 | 90,34 | 90,09 | 86,56 |
| -80 | -20 | 91,03 | 90,68 | 86,58 |
| -80 | -40 | 91,92 | 91,48 | 86,47 |
| -80 | -60 | 92,7 | 92,51 | 85,87 |
| -80 | -80 | 91,39 | 92,14 | 81,93 |
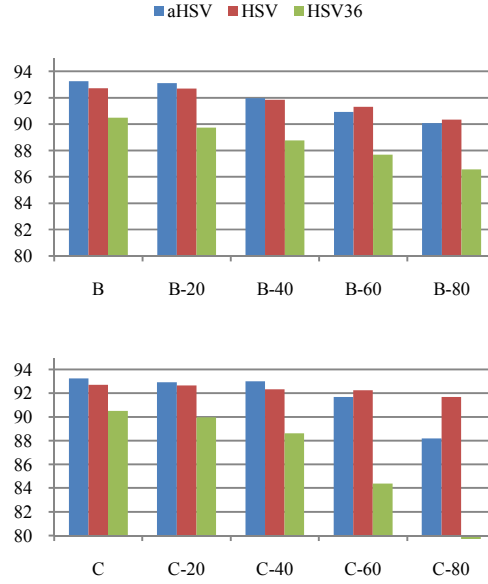


connected to the spatial arrangements of color blobs, so the efficiency is more or less closed to spatial histograms. Lastly we also tested color opponent histograms, exploiting color opponent theory, that has proved to be not as powerful as HSV based histograms but very stable varying light conditions indeed.

The second section of this analysis regards in particular the direct comparison between standard HSV histograms and aHSV histograms. For completeness, we also show comparison with another enhanced HSV (here called HSV36) previously described. Results are shown in Table 2. Standard HSV outperforms HSV36 in all cases, staying up to 10 percentage points over the rival with a more stable behavior decreasing brightness and contrast. Moreover aHSV performs quite and slightly better than HSV itself in most cases.

The gap between respective AMP can seem quite small, but it's an average estimation sparse on the entire database, that takes account not only the "binary" retrieval result (if a model was found or not) but also the rank of this model in the retrieval (in which position the system outputs that model). For this reason, a small gap can lead to significant differences from the visual point of view in the retrieval output. aHSV histogram maintains its supremacy, and only when images become too dark HSV histogram performs better than aHSV: in these cases, the gap between the respective performances are quite similar. In these cases a more optimized value for $n_A$ and $\lambda$ should be necessary: appearance (in terms of hue, saturation and brightness) turns out to be closer to gray colors, so the separation between chromatic and achromatic area leads to a less efficient representation of the achromatic area in the
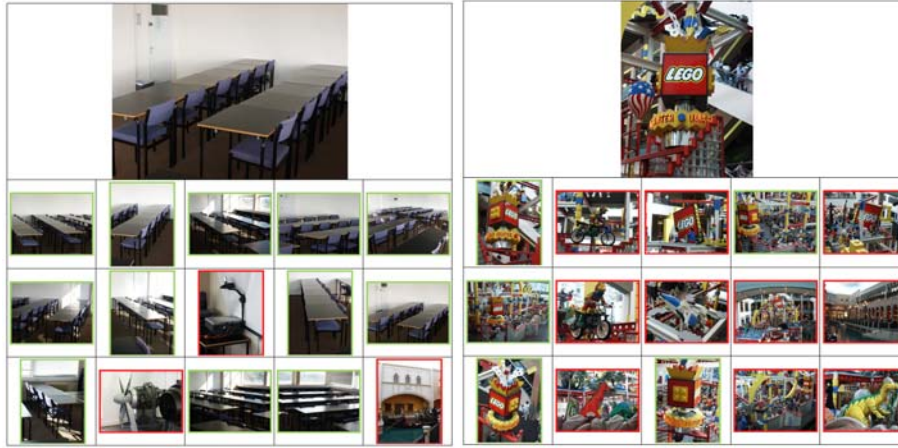
**Fig. 2.** The results page on the left has a MP=98.46, with 4 missing models of 16 in first 15 results. The one on the right has a MP=99.29, with 2 missing models of 7 in first 15 results.

image (now dominant). Even if in these cases image tends to grayscale, HSV still remains a more efficient retrieval feature than simple grayscale histograms. According to these tests, with real world images, in the best case scenario (generally with quite predictable and limited variations of brightness and quality), aHSV can potentially outperform standard HSV histogram. Instead in the worst case scenario aHSV turns out to have the same or just a little worse performance.

Some examples of retrieval are reported in figure **2**. The first image in the sequence represents the query submitted to the system, the following images are the result of the query. Green bordered images corresponds to effective model images of the ground truth provided, red bordered images are misclassifications.

## 5   Conclusions

In this paper a single feature comparison between color features has been proposed in order to establish the respective discriminating value in a generic retrieval application. The analysis showed the advantages linked to the use of HSV color space in content retrieval application, and then the advantages linked to the use of the enhanced HSV color space. An improvement of performances in terms of quality of retrieval has been evidenced, without losing any advantages compared to the standard definition. Test were conducted using a general images database, with unprofessional photos taken in different environment, with different light conditions, with variable zooming, framing and orientation.

# References

1.  Shih-Fu Chang , Lyndon S. Kennedy, Eric Zavesky, Columbia University's semantic video search engine Proceedings of the 6th ACM international conference on Image and video retrieval, 2007, pp. 643-643, Amsterdam, The Netherlands
2.  Apostol Natsev, Jelena Tešić, Lexing Xie, Rong Yan, John R. Smith,  IBM multimedia search and retrieval system, Proceedings of the 6th ACM international conference on Image and video retrieval, 2007, p. 645, Amsterdam, The Netherlands
3.  Rooij, Cees G. M. Snoek, and Marcel Worring. Mediamill: Semantic video browsing using the rotorbrowser. In Proceedings of the ACM International Conference on Image and Video Retrieval. Amsterdam, The Netherlands, July 2007.
4.  Grana, C., Vezzani, R., Bulgarelli, D., Barbieri, F., Cucchiara, R., Bertini, M., Torniai, C., and Del Bimbo, A. 2006. PEANO: Pictorial Enriched ANnotation of VideO. Accepted for publication in Proceedings of the 14th ACM international Conference on Multimedia (Santa Barbara, CA, United States, October 23-27, 2006).
5.  Information technology - Multimedia content description interface - Part 3: Visual, ISO/IEC Std. 15 938-3:2003, 2003.
6.  Zhang Lei, Lin Fuzong, Zhang Bo, 1999, A CBIR method based on color-spatial feature, In TENCON 99. Proceedings of the IEEE Region 10 Conference, 1999, Vol.1, pp 166-169
7.  J. Huang, et. al., "Image indexing using color correlogram", Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, pp. 762-768, 1997.
8.  Ojala T, Rautiainen M, Matinmikko E & Aittola M (2001) Semantic image retrieval with HSV correlograms. Proc. 12th Scandinavian Conference on Image Analysis, Bergen, Norway, 621 - 627.
9.  Cinque, L.  Levialdi, S.  Olsen, K.A.  Pellicano, A., Color-based image retrieval using spatial-chromatic histograms, Multimedia Computing and Systems, 1999. IEEE International Conference on, V0ol.2, pp 969-973, 1999
10. Hurvich, Leo and Jameson, Dorthea. An opponent-process theory of color vision. Psychological Review, Vol. 64, pp. 384--390, 1957.
11. Efficient Use of MPEG-7 Edge Histogram Descriptor. Chee Sun Won, Dong Kwon Park, and Soo-Jun Park, ETRI Journal, vol.24, no.1, Feb. 2002, pp.23-30
12. D. C. Tseng, C.H. Chang. 1992. Color segmentation using perceptual attributes. In Proceedings of 11th IAPR International onference on Pattern Recognition, Vol.III. Conference C: Image, Speech and Signal Analysis, pp. 228-231.
13. M. Seaborn, L. Hepplewhite, J. Stonham. 1999. Fuzzy colour category map for content based image retrieval. In Proceedings of the British Machine Vision Conference, BMVC99.
14. S. Sural, Gang Qian, S. Pramanik. 2002. Segmentation and histogram generation using the HSV color space for image retrieval. In Proceedings of the International Conference on Image Processing. 2002, Vol.2, pp II-589- II-592.
15. H. Eidenberger. 2004. Statistical analysis of MPEG-7 image descriptions, ACM Multimedia Systems Journal, Springer, Vol. 10, No. 2, pp. 84-97.
16. C. Grana, R. Vezzani, R. Cucchiara, "Enhancing HSV Histograms with Achromatic Points Detection for Video Retrieval" in Proceedings of ACM International Conference on Image and Video Retrieval, CIVR 2007, pp. 302-308, 2007
17. G. Schaefer and M. Stich (2004) "UCID - An Uncompressed Colour Image Database", Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia 2004, pp. 472-480, San Jose, USA.