Content Based Image Retrieval Using a Metric in a Perceptual Colour Space

G.Tascini and A. Montesanto

Dip. Di Elettronica, Intelligenza Artificiale e Telecomunicazioniv,
Università Politecnica delle Marche
g.tascini@univpm.it

Abstract. The aim of the present work is building an evaluation method for the similarity between colour hues. The method is defined by studying the attribution process, by human subjects, of colour hue couple to similarity classes (from 'very similar' to 'little similar'). From the study of these categorical judgements it is derived that the relation between the hue and the colour similarity is 'not-isometric' and greatly depends on the colour category. This result allows to extract representative functions for the three colour of the subtractive system: Red, Yellow, Blue. Besides we used a new method for segmenting the colour, based on the similarity with the main colours. Our method defines a quaternary tree structure, named 'Similarity Quad-Tree'; it is capable of extracting, from the whole image, the belonging degree to the Red, Yellow and Blue colours and their similarity with the reference colour. The check on the method applicability has given good results both: in the user satisfaction and in the computation. The approach may be viewed as a simple and fast indexing method.

Keywords: Colour Perception, No isometric Similarity Metrics, Human Subjects, Content based image retrieval.

1 Introduction

Various fields, like art, medicine, entertainment, education, and multimedia in general, require fast and effective recovery methods of images. Among these are Content Based Image Retrieval, in which images are described not by keywords but by content. A main approach is using low level characteristics, like colour, for segmenting, indexing and recovering. This work presents a method for annotating and recovering images that uses a new evaluation method for the similarity between colour hues that corresponds to human colour perception. In addition a fast and effective method for image indexing, is presented. In literature many methods are presented to this aim. A simple and fast method is based on a set of key words that describes the pictorial content [1]. The drawbacks of this approach are various: the method is hard for the big data bases; the quality of key words is subjective; the search by similarity is impossible. A more general approach to multimedia recover, different from those based on visual or acoustic data. The main difference depends on extraction of features. A popular approach is the Query By Example (QBE), where

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 621 – 628, 2005.
© Springer-Verlag Berlin Heidelberg 2005

the query is a key object, in particular an image, in the data base or depicted at query time. The content-based methods allow recovering images by the visual language characteristics, like similarity, approximation and metric relations, research key as figures, structures, shapes, lines and colours. As consequence they are many modes of indexing, storing, searching and recovering visual data. More refined are the methods in which the images may be analysed in the query phase; the corresponding software are called: Content Based Image Retrieval (CBIR) Systems. As the Query by Colour**,** two types of approaches are important: 1) retrieval of images with global colour distribution similar to the query image one, interesting for the pictorial data bases; 2) recovering of an object in a scene, using its chromatic features. [2] We will briefly describe some of most popular CBIR. QBIC, that means Query By Image Content [4], uses various perceptual characteristics and a partition-based approach to the colour. Introduces the Munsell transformation and defines a colour similarity metric [3]. The system is limited in the search of spatial characteristics. Virage [3] that supports the query about colour, structure and spatial relations operated on the following four primitives: Global Colour, Local Colour, Structure e Texture. Photobook [5] is an interactive set of tools developed at M.I.T. Media Laboratory on the Perceptual Computing. The system interacts with user by Motif interface. The matching is performed on the feature vectors extracted by considering invariance, scaling and rotation. VisualSEEk [6,7] and WebSEEk [8] are academic information systems developed at the Columbia University. VisualSEEk is a hybrid image recovery system that integrates the feature extraction using the colour representation, the structure and the spatial distribution. The recovering process is enhanced with algorithms based on binary trees. WebSEEk instead is a catalogue-based engine for the World Wide Web; it accepts queries on visual properties, like colour, layout correspondence and structure. ImageRover [9] is an image recovery tool developed at the Boston University. This system combines visual and textual queries for the computation of the image decompositions, associations and textual index. The visual features are stored in a vector, using colour and histograms texture-orientation; the textual one are captured by using the *Latent Semantic Indexing* on the association of the words contained in the HTML document 10]. The user refines the initial query using the *relevance feedback.* The Munsell colour space is a three-dimensional polar space, the dimensions being Hue, Value and Chroma. Value represents perceived luminance represented by a numerical coordinate with a lower boundary of zero and an upper boundary of ten. Chroma represents the strength of the colour, the lower boundary of zero indicating an entirely achromatic colour such as black, white or gray. The upper value of Chroma varies depending upon the Value and Hue coordinates. The Hue dimension is polar and consists of ten sections that are represented textually each with ten subsections represented numerically, as shown in Fig. 1. Our work considers the only dimension Hue, while maintains constant the other 2 variable Saturation and Intensity. Differently from the Munsell colour space, it considers the space of this single dimension 'not-uniform' and 'not-linear', that is 'not-idometric'. A main difference with the Munsell space is the following: our work do not evaluates the belonging of a colour to a 'nominal' category, that may be invalidated also by conceptual structures related to social history of examined population. We evaluate how mutch it is similar to a target-colour a variation of it, performed in the only tolnality dimension. Then the not-linearity is related to the colour similarity and not their categorization. The results are related to the subject judgements on the similarity evaluation between two colours, and not on the hue. If the Munsell space is perceptually uniform and linear, then a variation Δh of hue would be proportional to the related similarity variation Δs : the results have shown that this direct proportionality, between hue-variation and similarity-variation, of two colours do not exists.

2 The Similarity Function

A test on humans has been performed to the aim of defining a colour-similarity function. To appraise the similarity with the image target they are presented, to the human subjects, 21 images that are similar but perturbed in the colour. Such images maintain spatial fixed disposition. Everything is repeated for the three colours used in the search: *yellow*, *red* and *blue*. They are the three fundamental colours of the subtractive model..

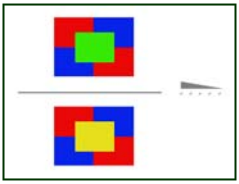

Fig. 1. The test image for the perception of the colours

2.1 Human Evaluation of the Similarity

The experiment on colour similarity evaluation uses the images shown in the Fig.1. They vary only the colour of the central rectangle. An image target and its variation are shown simultaneously to the subject. 63 couples of images are randomized for each proof (each set of 21 corresponds to a target image). The subject sees all the mixed images that are not subdivided for typology of basic colour. The subject have to assign a coefficient of "similarity to the target" of the images using scale from 1 to 5: 1= nothing, 2= very little, 3= little, 4= enough and 5= totally. Once submitted the test of similarity evaluation of the 12 subjects, we compute the assign frequencies of the 63 stimuli to the 5 categories. The average answer distributions, for each human subject, are not homogeneous and very asymmetric. The trend of the frequencies of the judgments of similarity for each colour furnishes an idea of how the different colours are represented at the cognitive level. We apply to the frequencies a unidimentional scaling to have a precise measure of the distance between the variation steps for each colour. The choice of the single dimension is because we consider only the variation of the hue, while the others two dimensions are fixed. The scaling is applied singularly to each scale of colour, so we have a scale of measurement for the Red, one for the yellow and one for the Blue. In this job the hypothesis is that the colours are different between them: 1- for the physical property of the bright energy, different for each one, 2- each colour has perceived psychologically in different manner and 3- the same stimuli are different entity. The colours could not be measured on the same scales, but they are subject to some transformations considering the verse that point out the affiliation to a particular category. The unidimentional scaling has been able to quantify the qualitative data of the first level: comparing the graphic of the two levels underlines their homogeneity. We have to notice that the graphic of the first level are not identical to those of the second level in as these last is more precise for what concern the metrics. Now we need a model for representing the relationship between hues (Hue with S=constant, I=constant) and related values of similarity. A not linear regression is used for finding a function (y) able to interpolate the points of the similarities related to the hues (x). This function is a polynomial. The values of the single parameters are different according to the derivation; each colour will weigh the value of similarity of the variation of tonality in different manner for the different values of the parameters. The resultant functions are the followings:

$$\text{Yellow: } y = -.00026 * x^3 + .046504 * x^2 - 2.6155* x + 43.9572; \tag{1.1}$$

$$\text{Blue: } y = .000061 * x^3 -.02622 * x^2 + 3.64571* x -163.38; \tag{2.2}$$

$$\text{Red: } y = .000001 * x^3 -.00096 * x^2+ .198553 * x -4.2372. \tag{2.3}$$

We underline the discrepancies between the form of the functions and the real disposition of the points in the space; to overcome this problem we have defined some broken functions to have a better representation of the data. So we have different range of tonality (h= hue) and they are divided for the three fundamental select colours: Blue (100< h< 204),(**2.1**); Yellow (26< h< 78),(**3.2**) and Red (-42< h< 22),that coincides to (0< h< 22) OR (198< h< 240),(**3.3**). Using these functions, given a value of hue with saturation of 240 and brightness of 120, we can see to which colour this hue is similar and how much is similar to the reference colour. For example: h= 0==> Red with sim.=- 3.199093,(**4.1**); 0<h<26==>Red with sim.= 0.016065* h2- 0.16173* h- 3.1260, (**4.2**) and 25<h<42==> Yellow with sim.=- 0.03198* h2+ 1.83212* h- 25.677, (**4.3**).

3 Similarity Quadtree

Once defined the relationship between Hue and similarity degree with reference colour, it is possible to represent an image using these concepts. To this purpose we have defined a new data-structure, named ‘Similarity Quad-tree’. In practice we depart from the original image and we recursively partition it quadrants: the procedure continues until a homogeneous quadrant is reached. We can stop the process also at a given level of subdivision: in this case the quadrant is not uniform: we can assign to the quadrant a colour equal to: a) prevailing colour, b) average colour. After this we have a value of Hue for each quadrant. At this time we calculate, for each quadrant, the value of similarity of the given Hue with the reference colour. The Similarity Quad-Tree (SQT) so obtained constitutes an image representation, related to the subjectivity with which the human perceives the similarity between colours; SQT have shown its usefulness in retrieval operations. In our colour based recovery we use two steps: 1) a decomposition of the image in a quad-tree structure, 2) matching between the obtained data-structure and a reference one. The whole procedure includes a series of operations; among these are size adjusting and pixel grouping; after we build a quaternary tree representing in each leaf the content in terms of primary colours Red Green Blue, named *RGB Quad Tree*: each leaf contains a string of three values, Red Green Blue. Then the *Similarity Quad- Tree* is built, departing from the RGB values, in which a leaf do not represent the colour but the values of the *similarity of the Hue* respect to the reference colour of the quadrant. The other nodes simply contain pointers to the children using the cardinal points notations: North West (NO), North East (NE), South East (SE) and South West (SO). The three colours of the SQT are defined in the RYB (Red, Yellow, Blue) space, while the *reference values* are extrapolated with the aid of our scaling computation on similarity judgments frequencies: a) *Red: ==> ref. =- 3,199093; b) Yellow: ==> ref. =- 3.544842;* c) *Blue: ==> ref. =- 4.768372.* The procedure for computing the similarity receives a leaf-node from an RGB_QuadTree in input extracts the three RGB values and from these computes the value of the hue (h). After the Hue range indicates the reference colour (3.1), (3.2), (3.3), as previous seen. Then the similarity to the main colour is computed, using the functions (2.1), (2.2) and (2.3) previous defined. The main colours are represented with the values 0, 1, 2 that stand respectively for *red*, *yellow* and *blue*. If the extracted colour from the RGB_QuadTree is perceptually not valid (for instance: little saturated or too bright), it is not inserted in the leaf. We have computed the hue of a given RGB colour, adopting a constant saturation value of 240 and a brightness of 120, corresponding to 1 and 0.5 in relative coordinates. The main procedure, that concerns the creation of the similarity tree, receives in input the roots of both trees: RGB_QuadTree and Similarity_QuadTree. When the RGB_QuadTree has reached a leaf-node, then the procedure insert the value of similarity in the Similarity_QuadTree. When it is not at a leaf-node the four children of the node are created. The SQT is suitable for indexing: here we associate with the database of imagines a file containing the data derived from the segmentation. This file is created before the retrieval, in such a way allowing fast access when these features are required. The structures of data-file will be composed by a sequence of information organized as in figure 5. An image may be selected for being used as query; with this the figure properties, of all database, will be compared to obtain a similarity matching. Our software gives the opportunity of sketching an image through a Java application. The user can also insert an image or a photo as query. The Matching process computes a score that identifies the similarity of a generic image with a selected one like key of search, assigning a value from 0 (completely different image), to 10000 (perfectly coincident image). This last value is divided in the following way: a) 4000 points derive from an appropriate comparison with the order of quantity of the principal red, yellow and blue colours; b) 6000 points result from the difference between the similarities of the two images for every principal colour. If the first two colours have identical quantity, then we assign to the variable score 2000 credits, while if also the seconds are equal then we assign 1250 credits, and so on. This to overcome the inconvenience of judging very similar two images only because they is a pixel of a certain colour similar to the average of all the colours of the other.

4 Results

We use two kind of image like keys of search in a small data-base. A subject gives a perceptive judgment on the similarity of the images to the keys, by concluding retrieval: this may be considered the optimal result that our program should reach. Then we perform the same two searches using our software. Fig. 2 show the images choices as keys of search. The subjects assign a score from 1 to 10 for each element of the database, where 10 stands for the best similarity and 1 for the smallest one. The total score associates a meaning degree of similarity to all database respect to the key image. Figure 3 shows the data-base images used in the example.

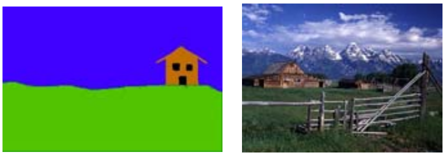

Fig. 2. Image key 1 and Image key 2

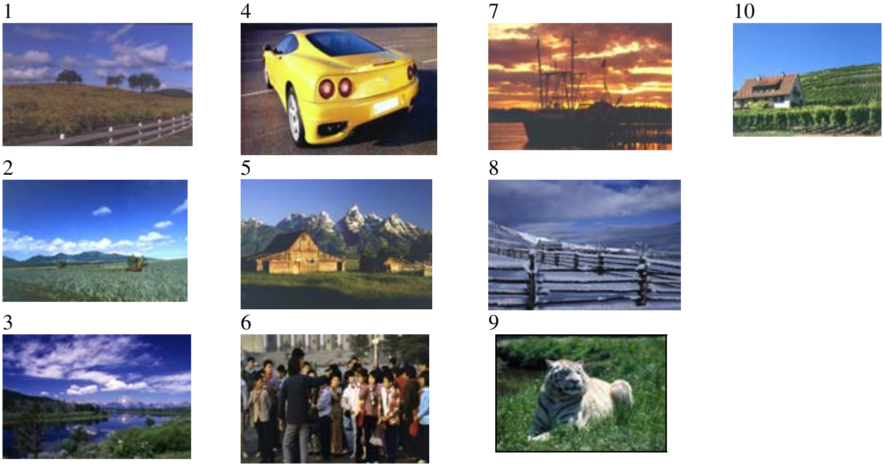

Fig. 3. The 10 images of the example

4.1 Comparison Between Human Decision and Software

In the tab 1 it is listed the numbers of the sorted figures, departing from the more similar to the key image 1 up to the less similar. The score that we find in the last column it is the associated value to the position of the orderly images from our CBIR system as regards the order of the same result from the Test. The used rule is that of assign 10 points to the images that they are in the same row, while two points for each line of distance are scaled. The last row corresponds simply to the percentage calculus of the score, therefore represents the empirical evaluation of the percentage of correctness in the measure of similarity of the software. This technique allows to have a range of percentages from 0 to 100 points because allows of assign also score negative, therefore the value gotten like a valid respect of the correctness of the created program could be considered. In tab.1 we compared the two methodologies using first as key the photo of the test 1. In this case we obtain a score of 88%, which points out that the program possesses a good analogy with the real perception human. Even in tab.1 we can see the test numbers 2, where the image key is a sketch. We used the same methods for the calculus of the score. The percentage of correctness in the measure of similarity of the software is 80%, so we have a good level of fidelity.

Table 1. Results of testing the two key images

Image key 1					**Image key 2**			
	Similarity Order	***Image n°***		**Score**	**Similarity Order**		***Image n°***	**Score**
		HUMA N	**SOFT WARE**			**HUMA N**	**SOFT WARE**	
	1	5	5	10	***1***	10	10	10
	2	3	3	10	***2***	2	2	10
	3	2	2	10	***3***	5	3	8
	4	1	10	8	***4***	3	5	8
	5	10	1	8	***5***	1	9	8
	6	8	8	10	***6***	8	1	6
	7	9	6	8	***7***	9	6	6
	8	6	9	8	***8***	7	8	8
	9	7	4	8	***9***	6	7	6
	10	4	7	8	***10***	4	4	10
Percentage of fidelity:				***88%***	***Percentage of fidelity:***			***80%***

5 Conclusions

The paper has presented a method of Content Based Image Retrieval, whose originality is related to two main aspects: 1) the definition of a *perceptual approach* that allows to build a new method for the similarity between colour hues evaluation, and that represents the abstraction of the content in a simple and efficient way; 2) the introduction of a new *methodology for indexing* the images, based on the Similarity Quad-Tree. So we can extract the properties only related to the colour, excluding the features like form and spatial relationships between objects of the image. The efficiency of this idea derives from the similarity function application. This function is derived from experimental evaluation of the perceptual metrics used by the human while judge the similarity between colours. The indexing methodology quality is related to a fast access to the representative features of an image that are stored in a vector: this necessarily involves high computation speed and cost minimization. Preliminary results give on 8000 image data-base, about 15 seconds of image-seek, for an ancient Pentium III at 750 MHz, where about 9 seconds are used for loading graphic interface, 2 second for feature extraction and less the 4 seconds for searching in the data-file and for matching. In a commercial PC with over 3 GHz of clock we go down under 2 second for all computation.

References

1. Y. Rui, T. S. Huang, S. F. Chang – "Image retrieval: Past, present, and future" - *Journal of Visual Communication and Image Representation* – pages 1÷23 - 1999.
2. J. R. Smith – "Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression" - *PhD thesis, Graduate School of Arts and Sciences, Columbia University* - 1997.
3. J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, C. F. Shu – "The Virage image search engine: An open framework for image management" - *In Proceedings of the Storage and Retrieval for Still Image and Video Databases IV* - pages 76÷87 - San Jose, CA, USA - February 1996.
4. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker – "Query by image content and video content: The QBIC System" - *In IEEE Computer* - pages 23÷32 - September 1995.
5. Pentland, R. Picard, S. Sclaroff – "Photobook: Content-based manipulation of image databases" - *In International Journal of Computer Vision* - pages 233÷254 - San Jose, CA, USA - June 1996.
6. J. R. Smith, S. F. Chang - *Querying by Color Regions using the VisualSEEk Content-Based Visual Query System* - Intelligent Multimedia Information Retrieval, AAAI/MIT Press - 1996.
7. J. R. Smith, S. F. Chang – "VisualSEEK: a fully automated content-based image query system" - *In Proceedings of the 4th ACM International Conference on Multimedia* - pages 87÷98 - Boston, Massachusetts, USA - November 1996.
8. J. R. Smith, S. F. Chang – "Visually searching the web for content" - *IEEE Multimedia Magazine* – pages 12÷20 - 1997.
9. S. Sclaroff, L. Taycher, M. La Cascia – "Imagerover: A content-based image browser for the world wide web" - *In Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries* - pages 2÷9 - San Juan, Porto Rico - June 1997.
10. M. La Cascia, S. Sethi, S. Sclaroff – "Combining textual and visual cues for content-based image retrieval on the world wide web" - *In Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries* - pages 24÷28 - Santa Barbara, CA, USA - June 1998.