# THE VISUAL GOODNESS EVALUATION OF COLOR-BASED RETRIEVAL PROCESSES

*Petteri Kerminen*
Department of Information Technology, Pori, Tampere University of Technology,
P.O. Box 300, 28101 Pori, FINLAND
Tel: +358 2 6272737; fax: +358 2 6272727
e-mail: petteri.kerminen@pori.tut.fi

*Moncef Gabbouj*
Signal Processing Laboratory, Tampere University of Technology
P.O. Box 553, 33101 Tampere
Tel: +358 3 3653967; fax: +358 3 3653857
e-mail: moncef@cs.tut.fi

**ABSTRACT**

A content-based image retrieval system helps users to find suitable images or videos from large databases for their purposes. In this paper we present the results of analyses of visual outputs from our image retrieval system, based on color matching. These analyses are based on human evaluation of color content similarity. Computer-based simulation results have been analyzed visually with the opinion scores method, i.e. how good the results are according to the human visual system. Tests have shown that there is no superior retrieval method that is best in every case. The goodness results of each retrieval method will be presented in this paper. The usability and usefulness of each system will be presented as well as a possible utilization domain.

## 1. INTRODUCTION

Nowadays there are hundreds of thousands or maybe millions of digital images in image databases. Large image databases are difficult to browse with traditional text searches, if the user wants to search for images that have, for example, a similar color content. A content-based retrieval system solves this problem. In this case the system is a color matching based retrieval system. Video on demand (VoD) services also require an intelligent search system for end-users. The VoD systems' search methods differ slightly from those of the image databases. A color content-based retrieval system is one possibility also within the VoD business.

A reference book is a suitable option for a static image database, if the images are arranged using a method which is familiar to every user. An experienced user can use these systems, as well as textual searches (keywords have to be inserted in a database), efficiently. These systems are mostly indexed manually and therefore indexing is subjective. When the images' properties are picked up automatically, indexing is not dependent on an individual's subjective opinions. There are situations when a multi-language system has to be used. There the language independent search system's best properties can be utilized. A tool which is based on the images' properties can be implemented as a language independent system. These properties can be for example colors, shapes, textures, spatial locations of shapes etc. The properties can be saved as feature vectors without using a natural language. Queries can be made visually or with any written language, if a translation software module is available.

In the MuVi-project [1] content-based retrieval tools are under construction. They will cover most of the properties presented above. Research work on content-based image retrieval has been done for example in [2 – 6]. In this paper we present a summary of tens of simulations for each color system or color system combination which has been studied. The results of the analyses, which are presented in this paper, are based on visual output of MuVi's color content based retrieval algorithms. Other properties like textures and shapes are not used in this evaluation.

## 2. THE RETRIEVAL PROCESS

Colors and brightness are the most easily perceptible parts of images. Color-based retrieval mechanisms are based on similarity evaluations between histograms (feature vectors) from the search image and other images in a specified database. The comparison algorithm doesn't contain information on what the values on the histogram vector mean. Thus the histograms of all images have to be in the same color space and have the same number of bins. The database can contain several histograms from each image. This allows the use of searches with different precision levels or with different color systems.

The retrieval mechanisms that will pick up images with a similar color content are not as complex as mechanisms that utilize shapes or textures. However, there are some circumstances which are important for keeping the search process's speed and precision on a sufficient level. The first condition is that the histograms are calculated during the process when the images are inserted into the database. Only the histogram of the user-sketched search image or the user-defined search image is allowed be calculated on-line.

For small image databases search algorithms can be made with exhaustive search. This method calculates differences between the search image's histogram and all other histograms, and keeps a best match table, which is updated during the search loop. The recall value (how many percent of all correct matches are retrieved) is maximal, because all possible histograms are compared with the search image's histogram. Time consumption is a constant time slice for each image. The time used per search depends on the number of bins in the histograms, because the difference between the histogram vectors is calculated through bin by bin comparison. The time consumption depends also very much on the hard disk's reading capacity.

For larger image databases it is reasonable to take into use a database management system (DBMS). It organizes images and histograms in the database so that the most similar ones are situated in neighbouring cells. The DBMS can pick *n* best matches directly form the database without calculation or exhaustive search, because the system knows how data is inserted into the database. Thus this version is not so critical regarding to the number of bins in the histograms, although the number of bins increases the database's dimensions and therefore also the database's build-up time and the search time a little. A disadvantage in this version occurs if the histograms are not inserted correctly to the database according to all components of the feature vector. In that case the recall value is not on the maximum level for all searches.

The retrieval process used during this evaluation has been described in [7,8]. The retrieval mechanisms used in MuViS have been presented in [1].

## 3. GOODNESS EVALUATIONS

In this evaluation the purpose was to test which color system and which number of bins gave visually the best retrieval results. The evaluation has been made using the opinion scores (OS) [9,10] to point at the visual goodness of the retrieval process's output. Only the color content has been used in this evaluation.

The color spaces used in this evaluation are RGB (red, green, blue), HSV/I (hue, saturation, value / intensity), XYZ (X and Z being chrominance values, Y luminance) and $L^*a^*b^*$ ($L^*$ being brightness, $a^*$ redness-greenness and $b^*$ yellowness-blueness) [11]. 27, 64 and 125 bin histograms have been used with the RGB system, 27, 39 and 166 bins have been used with the HSV system, 50 and 175 bins have been used with the $L^*a^*b^*$ system and 50 bins have been used with the XYZ system. These color systems with the bin numbers mentioned above and some combinations of those color spaces have been tested with tens of different query images. The visual results of the queries are not shown here, due to limitations of space and lack of colors.

The tests have been done with the MATLAB R11 (ver 5.3). The database contains 235 JPEG images of the size 320x240, on average. The I/O functions used have been MATLAB's own functions without any modifications. The mechanisms used have been presented in [8]. The results of the queries have been analyzed subjectively immediately after each query. A chosen set of 9 images with different color contents has been used as query images for all color spaces. The query results of these query images have been stored on a hard disk and analyzed more precisely off-line. The figures and tables in this paper are based on these off-line test evaluations.

The query results are analyzed with two subjective methods. First an opinion scores measurement 1 – 5 (from very poor to excellent) have been used. The method used is a variation from ITU-R's and ITU-T's testing methods presented in [9,10]. On this stage the purpose was to define how good the visual result of query is according its color contents. Spatial locations of colors are not used in evaluations. On the second stage the color spaces used are ordered in the rank order after each query image. The scale values depend form number of tested color systems / the number of bins combinations. The smaller the value is in rank order the better the system is.

### 3.1 Evaluation: A whole histogram case

In figure 1 and table 1 are the results of the simulations where all bins of the histograms have been used. The difference between the histograms has been calculated with (1), where *i* = bin, *n* =number of bins, *v* and *u* are histograms:

$$d = \sum_{i=1}^{n} \sqrt{\left(v(i) - u(i)\right)^2} \quad (1)$$

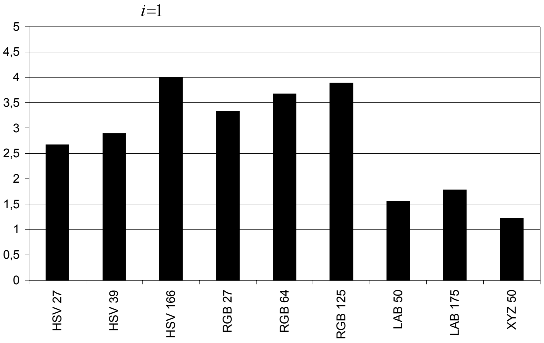


Figure 1: OS evaluation for tested color spaces.
1 = very poor, 5 = excellent. Whole histogram case.

These results show that the RGB and the HSV/I color spaces gave the best results. The best results appeared

| Color space / number of bins | hsv 27 | hsv 39 | hsv 166 | rgb 27 | rgb 64 | rgb 125 | lab 50 | lab 175 | xyz 50 |
|---|---|---|---|---|---|---|---|---|---|
| **Average ranking** | 5,67 | 5,56 | 2,11 | 3,44 | 2,89 | 2,11 | 7,78 | 6,78 | 8,67 |

Table 1: Rank order evaluation for tested color spaces. Whole histogram utilized.

when a large number of bins were used. The query images were natural images. When constructed images were used the best results were achieved with a smaller number of bins. This happened because constructed query images usually have fewer colors than natural images. Although the best results were achieved with a large number of bins, a smaller number of bins is recommended, because the retrieval is faster with smaller histograms.

### 3.2 Evaluation: 3 dominant colors case

Figure 2 and table 2 show the results of simulations where only 3 dominant colors (bins) of histograms were used. In this test session the algorithm picks up the three highest values from each histogram and checks if all three dominants are the same, if some are the same or if all are different. According to this comparison procedure the images are arranged into order of similarity. HSV 39, $L^*a^*b^*$ 50 and XYZ 50 have been tested, but not with all test images, because the results in these color spaces seem to be of worse quality than in other tested color spaces. Thus the test results of these color spaces are not presented here. In this ”three dominant” test session the OSs are almost equal for all other color spaces except $L^*a^*b^*$ 175. The differences in the ranking table are bigger, but the OSs clarify the situation better.

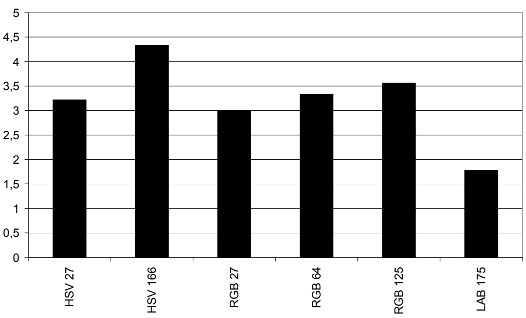


Figure 2: OS evaluation for tested color spaces. 1 = very poor, 5 = excellent. 3 dominant colors case.

| Color space / number of bins | hsv 27 | hsv 166 | rgb 27 | rgb 64 | rgb 125 | lab 175 |
|---|---|---|---|---|---|---|
| **Average ranking** | 3,56 | 1,89 | 3,33 | 4,22 | 2,44 | 5,56 |

Table 2: Rank order evaluation for tested color spaces.

### 3.3 Evaluation: 1 dominant color case

Figure 3 and table 3 show the results of simulations where only 1 dominant color (bin) of histograms are used. In this test session the algorithm picks up the highest value from each histogram and checks if the dominant is the same or a different bin. According to this comparison the procedure images are arranged into order of similarity. In this ”one dominant” test session the OSs are good or better for all the other color systems except $L^*a^*b^*$ 175.

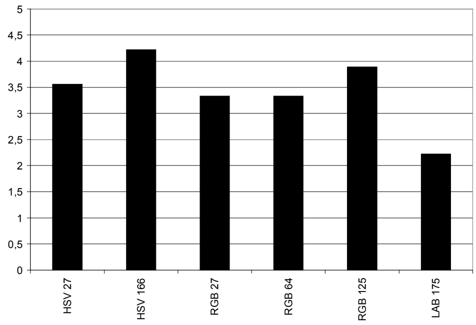


Figure 3: OS evaluation for tested color spaces. 1 = very poor, 5 = excellent. 1 dominant color case.

| Color space / number of bins | hsv 27 | hsv 166 | rgb 27 | rgb 64 | rgb 125 | lab 175 |
|---|---|---|---|---|---|---|
| **Average ranking** | 3,33 | 1,78 | 4,22 | 3,78 | 2,56 | 5,44 |

Table 3: Rank order evaluation for tested color spaces.

### 3.4 Evaluation: Dual-stage case

In figure 4 are the results of simulations where two different color spaces / number of bin combinations have been used. During the first search loop a smaller number of bins have been used (in most cases) and all the histograms have been checked. The most similar images have been kept in the ”best match table”. The number of images that are kept on this table is 10 times the desired number of retrieved images. In the second loop only the histograms of these images have been compared. That has been done to increase the speed of the system for cases when there is no DBMS available.

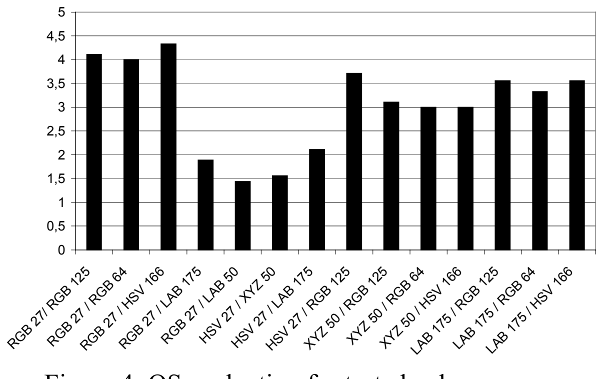


Figure 4: OS evaluation for tested color spaces. 1 = very poor, 5 = excellent. Dual-stage.

From the results can be seen that there doesn't exist any superior color space / bin combinations. RGB and HSV with a large number of bins gave the best results in on average, but still get rank orders 4 – 6 for some specific queries. In the dual-stage simulations the best results were achieved with queries where RGB or HSV was used at the final stage. There is not a very large difference, if RGB 27, HSV 27, XYZ 50 or $L^*a^*b^*$ 175 has been used at the preliminary stage.

## 4. SUGGESTIONS FOR UTILIZATION

The lightness/brightness based retrieval algorithms, $L^*a^*b^*$ and XYZ, are good for the search cases where a query image has very large intensity values or very small ones (over/underexposured photographs). In these cases the colors are in the area of color spaces where the hue is not very useful. In the case of "successful" photographs the hue contains important information and the brightness can be coded with a smaller number of levels.

There are three circumstances, which affect the selection of the number of bins. First, if the query image is a natural image, a higher number of bins will give better results. In case of a user drawn query image a lower number of bins will give better results. Second, if there is no DBMS available and the database contains very many images a lower number of bins is recommended to be used, due to the time consumption. Third, if there is a DBMS available and the database contains very many images a large number of bins will give more accurate matches.

## 5. FUTURE WORK

There are two critical parameters, which the user has to choose before search: the number of bins and the color space used. In the future an automatic color space selection study has to be made. When the query system automatically chooses an ideal query method, according to the query image, the results will be optimal. However, there must be a possibility for the users to make their own decisions. In this study only overall color contents are estimated. Future work will contain position based color content measurements and retrieval based on that method.

## 6. CONCLUSIONS

There is no superior color space for all retrieval cases. The rank order evaluation's results are more difficult to analyze than opinion score results. There is a possibility that two systems give almost equal results, but one is ranked above the other. When a user interpreters this kind of results, one has to be very careful.

The systems which utilized color primaries or hue gave better results with natural images than the systems which utilized brightness. The selection of the best number of bins depends on the situation where the query is launched. A lower number of bins is suitable for user sketched queries and a larger number of bins for natural image based queries.

## References


[1] Alaya Cheikh F et. al. MUVIS: A System for Content-Based Indexing and Retrieval in Large Image Databases. Proc. of SPIE Storage and Retrieval for Image and Video Databases, San Jose, 1999,vol. 3656: 98 – 106.

[2] Sawhney H. S., Hafner J. L., Efficient Color Histogram Indexing for Quadratic Form Distance Functions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, vol. 17 no. 7: 729 – 736.

[3] Vellaikal A. et. al. Content-Based Retrieval of Color and Multispectral Images Using Joint Spatial-Spectral Indexing, SPIE vol. 2606: 232 – 243.

[4] Ng V. et. al. Adaptive Histogram Indexing, SPIE vol. 2606: 202 – 211.

[5] Smith J. R., Chang S. F. Querying by Color Regions Using the VisualSEEK Content-Based Visual Query System, in Intelligent Multimedia Information Retrieval, Maybury M. T. ed., The MIT Press, 1997, 23 – 41

[6] Flickner M. et. al. Query by Image and Video Content: The QBIC™ System, IEEE Computer 1995 vol. 28 no. 9: 23 – 32.

[7] Kerminen P, Gabbouj M. Image Retrieval Based on Color Matching. Proceedings of FINSIG'99, Oulun yliopistopaino, 1999: 89 – 93.

[8] Kerminen P, Gabbouj M. Prototyping Color-based Image Retrieval with MATLAB®, Proceedings of MATLAB DSP Conference 1999, Digicolor Oy, 1999, 91 – 95.

[9] ITU-R Recommendation BT.500-7, "Methodology for the subjective assessment of the quality of television pictures", International Telecommunication Union, 1995

[10] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia application", International Telecommunication Union, 1996

[11] Jain A. K. Fundamentals of Digital Image Processing, Prentice Hall, 1989: 66 - 71.