Evaluating Content-Based Image Retrieval by Combining Color and Wavelet Features in a Region Based Scheme

Fernanda Ramos[1], Herman Martins Gomes[2], and Díbio Leandro Borges[3]

[1] Faculdade de Filosofia, Ciências e Letras de Palmas, Palmas – Pr, Brazil
ramosrs@gmail.com
[2] UFCG - Universidade Federal de Campina Grande,
Departamento de Sistemas e Computação, Av. Aprígio Veloso s/n,
Bodocongó, Campina Grande – Pb, Brazil
hmg@dsc.ufcg.edu.br
[3] BIOSOLO, Goiânia – Go, Brazil
dibio.borges@terra.com.br

Abstract. Content description and representation are still challenging issues for the design and management of content-based image retrieval systems. This work proposes to derive content descriptors of color images by wavelet coding and indexing of the HSV (Hue, Saturation, Value) channels. An efficient scheme for this problem has to trade between being translation and rotation invariant, fast and accurate at the same time. Based on a diverse and difficult database of 1020 color images, and a strong experimental protocol we propose a method that first divides an image into 9 rectangular regions (i.e. zoning), second it applies a wavelet transformation in each of the HSV channels. A subset of the approximation and of detail coefficients of each set is then selected. A similarity measure based on histogram intersection followed by vector distance computation for the 9 regions then evaluates and ranks the closest images of the database by content. In this paper we give the details of the this new approach and show promising results upon extensive experiments performed in our lab.

1 Introduction

Most conventional content-based retrieval systems use color or spatial-color features for characterizing image content [5], [10]. Some interesting works have used additional low-level features that can be computed automatically, i.e., without human assistance, and associated with the color-based features. A promise one is texture. However, even after texture has been widely studied in Psychophysics, as well as in Computer Vision, our understanding of it is still very limited when compared with our knowledge of other features, such as color and shape. A difficult task when using texture is how to represent it or even how to combine it with other features, such as the ones based on color. Most of methods to represent texture are based on co-occurrence statistics, directional filter masks, fractal dimension and Markov Random Fields. Some interesting works have tried to represent texture using visual properties, such as in [4], where texture is characterized by the following features: coarseness, contrast, busyness, complexity and texture strength. In a similar direction, Rao and Lohse [8] have done an experiment to describe texture. They suggest that three perceptual

M. Lazo and A. Sanfeliu (Eds.): CIARP 2005, LNCS 3773, pp. 679 – 690, 2005.

features can describe texture: repetitiveness, directionality and granularity (complexity). Another strategy to characterize texture has been the use of a wavelet transform. In this direction, Brambilla et al. [2] have used multiresolution wavelet transform in a modified CIELUV color space to compute image signatures for use in a content-based image retrieval application. Other approach using wavelet coding on color is [4], where the authors propose a joint coding using texture, color, and shape with statistical moments on the wavelet coefficients. The work we present in this paper is close to those since it also uses a wavelet decomposition to derive descriptors for the images. However, it proposes a different way to represent and select the features, and also to compute and rank the closest matches. We also present a strong experimental protocol showing how they performed so we derived the final features for the approach. There is a large literature on Content Based Retrieval and we point the reader to the survey in [1], and to the works in [2], [5], and [7] for a more detailed covering of the area.

The remaining of this paper is organized as follows. Next, we present in detail our approach that combines color and wavelet features in a region based scheme. An extensive experimental protocol that we used to derive the best features in the three levels of decomposition and the image channels HSV is presented in Section 3. Section 4 summarizes the main results and points to future works.

2 Proposed Method

Figure 1 shows an overview of the proposed method. In the First Stage a signature is computed for each image in the database. It begins by converting the input image from RGB to the HSV color space, such as we could deal directly with perceptually more meaningful information [11]. The second step is used to divide the image into 9 regions. This has shown to be an interesting strategy to associate spatial information to color-based features. The third step corresponds to a Mallat wavelet decomposition [9] using Haar basis function. Finally, the combination of the best features is used as a signature for the input image. To obtain the best features, we have evaluated for each color channel (H, S and V) the sub-images corresponding to the following wavelet coefficients: approximation, horizontal, and vertical, all of them computed on three levels of decomposition. Using the 10% largest coefficients in magnitude for each band, i.e. LL, HL, and LH, has shown to be a sufficient selected set for the best features.

The Second Stage corresponds to the retrieval process in which after computing an intersection of the quantized histograms [11] for each channel and all three bands (i.e. (H,S,V) x (LL,HL,LH) = 9), and for each spatial region of the image, a distance measure computes and ranks the images based on a sum of the intersections.

2.1 Signature Extraction

For each channel (H, S and V), and, respectively, each of the 9 image regions, the wavelet transform is computed using three levels of decomposition for the tests. With the experiments we have noticed that performance was closely the same for the 3rd level of decomposition, so only the 3rd levels with less features were used as features. Of course this will depend upon the initial resolution of the images.

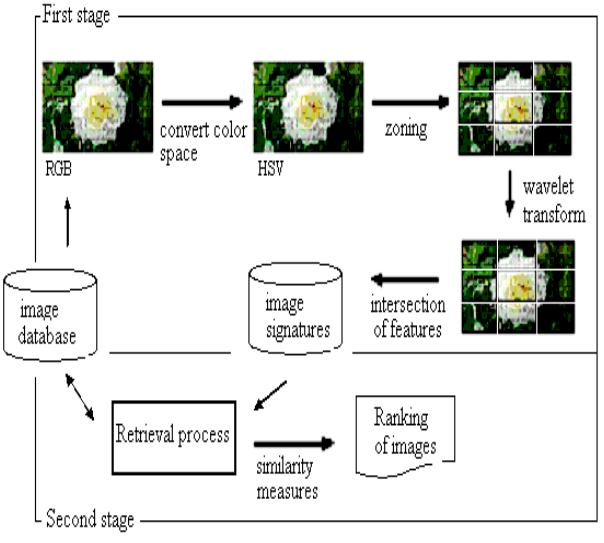


Fig. 1. Overview of the proposed method

Figure 2 shows the composition of the image signature computed. Each histogram used as signature is computed taking into account the coefficients quantized using the mean and standard deviation of each region at each decomposition level. Each histogram contains only 10% of the most significant coefficients of the image region being processed. The best results were obtained by using the combination of the approximation wavelet coefficients calculated on the H and V channel with the vertical coefficients calculated on the S channel, as it will be shown in the experiments.

	Wavelet coefficients	**Level of decomposition**
For each color channel (H, S and V)	Approximation (LL)	1st
		2nd
		3rd
	Vertical (HL)	1st
		2nd
		3rd
	Horizontal (LH)	1st
		2nd
		3rd

Fig. 2. Image signature structure. Each level of decomposition produces 3 histograms (LL, HL, LH) for each channel (i.e. H, S, and V).

2.2 Retrieval Process

Each image will then have as an index a feature vector of nine (9) histograms (HSV x LL,HL,LH) with 10% of the largest coefficients. All of them are referenced in 9 dif-

ferent spatial regions of the image. Retrieval then consists of measuring a distance for each pair of images (i.e. a query and one from the database) by computing the histogram intersection for each pair of regions and histogram features. A sum of these nine (9) intersection results for each channel (HSV) is computed as a distance and ranking measure in the end. In the experiments we have found a set of the three (3) best features that performed better in the end, and the final feature vector is selected as a combination of those. Table 1 shows the quantities and types of the classes used.

Table 1. Names and quantities of the 27 classes used for the experiments. Total number of images is 1020.

Class	# samples	Class	# samples
1-Water	70	15-Flowers	43
2-Air	53	16-Football	44
3-Animals	55	17-Fruit	19
4-Wires	06	18-Girls	27
5-Trees	103	19-Trees2	55
6-Boxes	15	20-Windows	12
7-Cars	42	21-Foliage	30
8-Christmas	42	22-Mammals	16
9-Cement	14	23-Mosaic	25
10-Buildings	79	24-Mountains	31
11-Sunset	63	25-Bridges	07
12-Ducks	37	26-Sky	43
13-Flags	43	27-Snow	14
14-Texture	32		

3 Experimental Results

The experiments were carried out on 1020 images distributed in 27 classes. Table 1 gives the image classes and their quantities and Figure 3 shows sample images of each class. In the experiments each channel (H, S and V) was evaluated separately. Two sets were separated from each class, being 10% for training and 90% for testing, tests were performed individually for each image of the training set (against the testing set). The successful classification rank was finally computed considering the number of matches between query and result. The confusion graphics show the first classification in red (light gray), where for example a diagonal red (light gray) line would mean perfect matches for all classes. For each image region the wavelet transform is calculated considering three levels of decomposition. Figures 4, 5 and 6 gives the confusion graphics for channels H, S, and V respectively in the first level of decomposition for approximation, horizontal, and vertical coefficients. All the confusion graphics showed in this paper mark (red (light gray) curve in the figures) the respective class that was chosen as the first ranked in the experiment. The complete protocol evaluated the following:

1) In the first set of experiments, each color channel was considered in a separated way. The retrieval process was evaluated based on signatures created

from each of the three channels and considering 3 sub-images generated by the wavelet transform (approximation, horizontal and vertical) at each decomposition level. This totalizes 27 different signatures for each image.

2) The second set of experiments considers the combination of the most promising results of the first experiments, independent of the level of decomposition and trying to get the best results with the coarsest level of decomposition if possible. New signatures were created by combining signatures through the use of intersection of coefficient histograms.

3.1 First Set of Experiments

Independently, each of the 27 possibilities of features (i.e. HSV x 3 coefficients (approximation, horizontal, vertical) x 3 levels of decomposition) was tested in this stage. Ten percent (10%) of the images were separated for validation only, and the results were averaged using a cross-validation scheme. Figure 4 shows confusion graphics related to the experiment on the H channel using the approximation, horizontal and vertical details sub-images in the first decomposition level. Figures 5 and 6 show respectively the results for S and V channels. On the *x* axis the class of the query image is given, and on the *y* axis the class that had the best (highest) score for the classification result. During these experiments, we observed that the best retrieval results when using the channels H (Hue) and V (Value) were obtained from the signatures based on the approximation wavelet coefficients on the third decomposition level. Regarding the S channel best success rates were higher in the third level of decomposition using the vertical detail coefficients. Figures 7, 8 and 9 shows examples of images misclassified according to the labeled database used. Although a general comparison of the success rates obtained here is dependent on the database used, which we know it is not large enough for benchmarking purposes, there is a high correlation between some image classes and the main purpose here would be to design and test a small, however significant and efficient, set of features based on a combination of color and wavelet representations to be used in an image retrieval task. Since it was possible to evaluate out of the 27 sets of features which were the most efficient for retrieval we picked the 3 best ones and performed a second set of experiments in order to find a best combination of them for the final classification.

3.2 Second Set of Experiments

In the second set of experiments the signatures calculated on the channels H and V using the approximation coefficients (LL) and the signatures calculated on the channel S using the vertical coefficients (HL) were combined through the use of intersection of their histograms. No weight was given differently to each signature. In this case, the 10% corresponding to the most significant coefficients are chosen from the resulting histogram after the intersection process. The final results are shown in Figure 10. Only 3 out of 27 classes were not classified correctly as the first choice, which shows an improvement from any of the individual set of features in the first set of experiments.

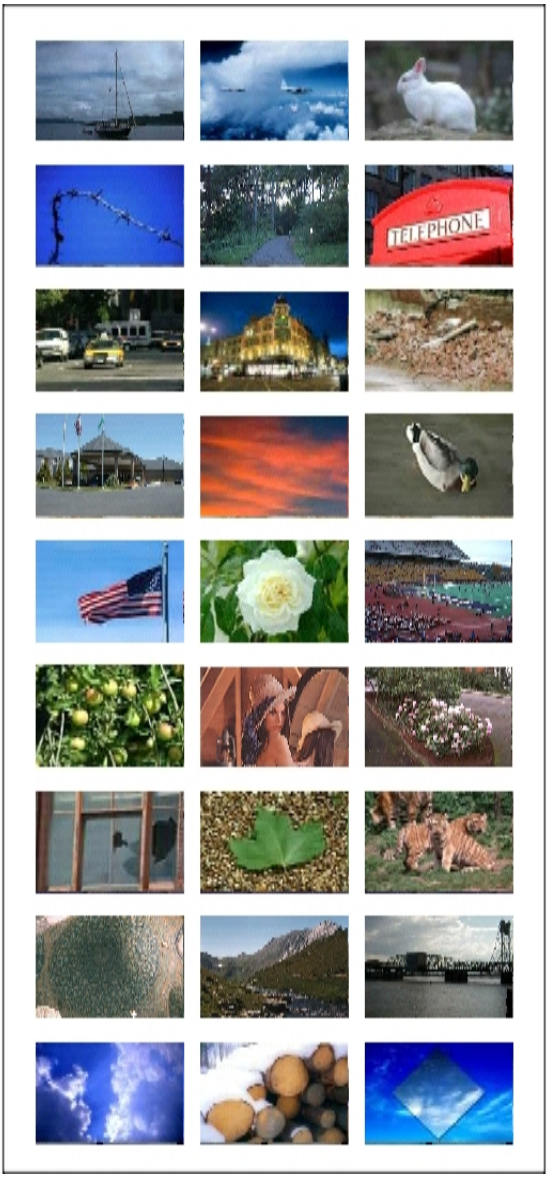
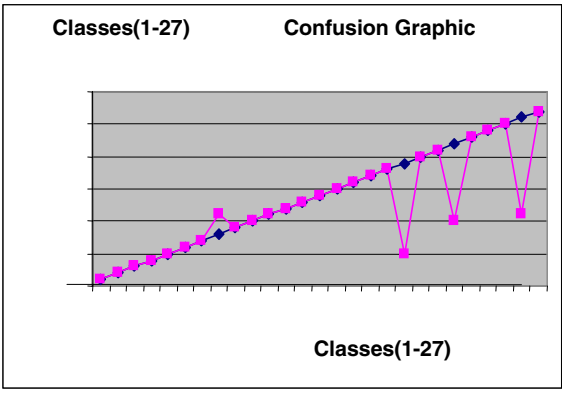
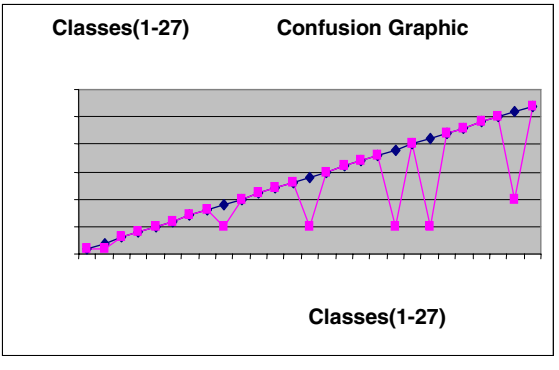


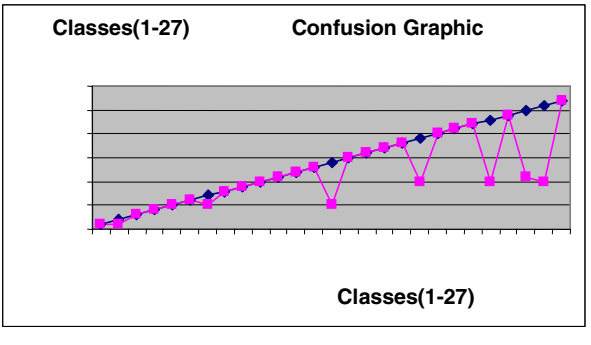
Fig. 3. Samples representative of the classes. From the top left to the right (1 to 27 as named in Table 1). The complete database has 1020 images.



Channel H 1st level (Approximation coef.)

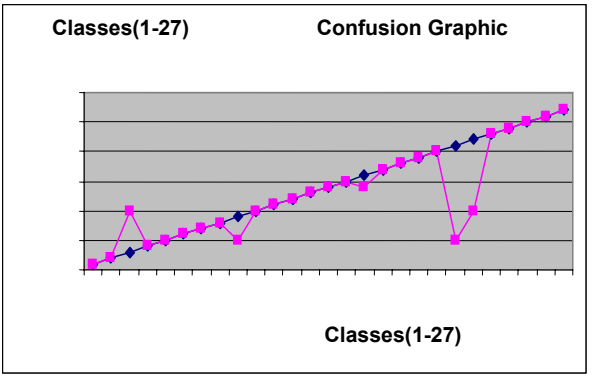


Channel H 1st level (Horizontal coef.)

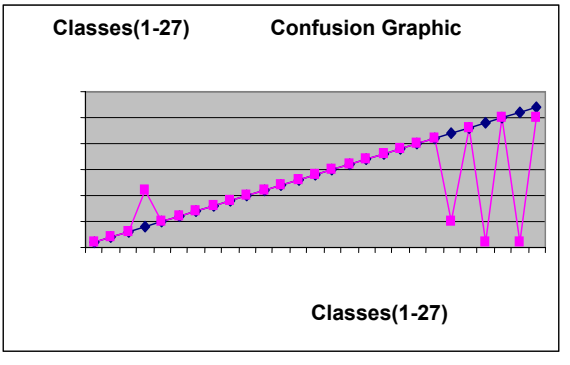


Channel H 1st level (Vertical coef..)

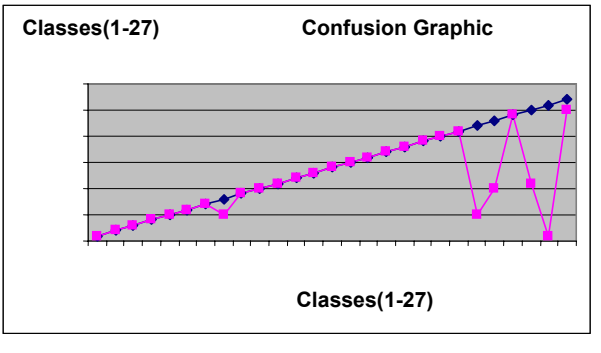
Fig. 4. Confusion graphics: experiment on channel H (Hue Value), all coefficients in the first level of decomposition. (*x* axis is the class of the query image, and *y* axis is the result classification obtained).



Channel S 1st level (Approximation coef.)

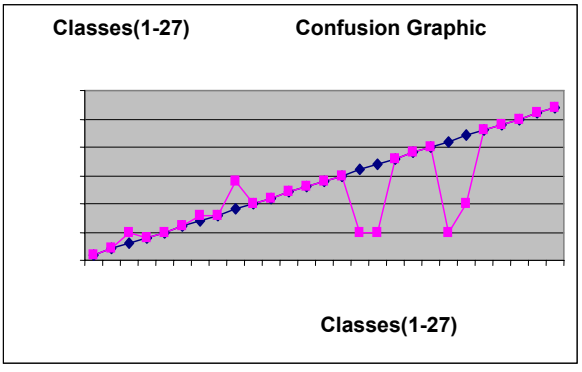


Channel S 1st level (Horizontal coef.)

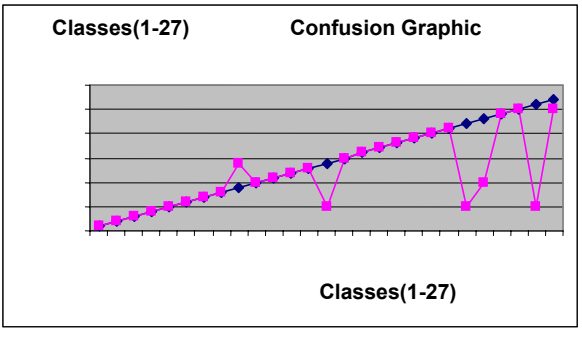


Channel S 1st level (Vertical coef.)

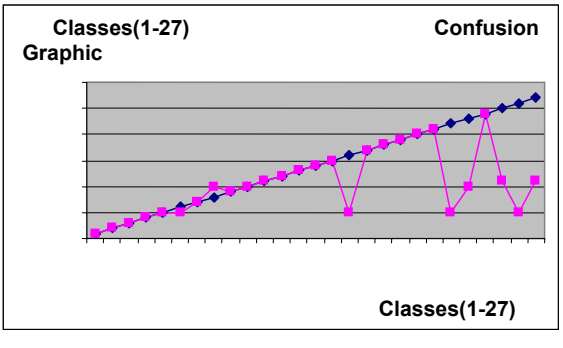
Fig. 5. Confusion graphics: experiment on channel S (Saturation Value), all coefficients in the first level of decomposition. (*x* axis is the class of the query image, and *y* axis is the result classification obtained).



Channel V 1st level (Approximation coef.)



Channel V 1st level (Horizontal coef.)



Channel V 1st level (Vertical coef.)

Fig. 6. Confusion graphics: experiment on channel V (intensity Value), all coefficients in the third level of decomposition. (*x* axis is the class of the query image, and *y* axis is the result classification obtained).

Fig. 7. Examples of misclassification (i.e similar images) occurred in Hue channel a) an image from Class 5 – Trees, and b) an image from Class 19 – Windows

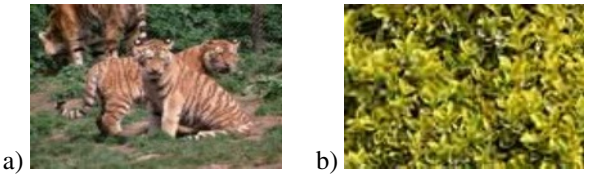


Fig. 8. Examples of misclassification (i.e. similar images) occurred in Value channel a) an image from Class 22 – Mammals, and b) an image from Class 5 – Trees

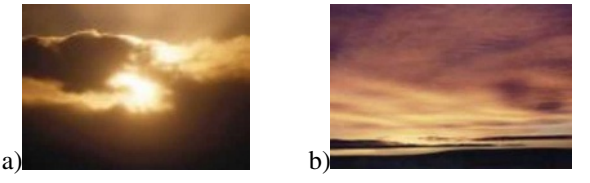


Fig.9. Examples of misclassification (i.e. similar images) occurred in Saturation channel. a) an image from Class 26 – Sky, and b) an image from Class 11 – Sunset.

The main purposes of this work were: 1) first, design a special set of features to be used efficiently as a reduced set of features in image retrieval tasks; 2) evaluate the conditions upon which they could work better and combine the best results in a particular set of features. The first was accomplished with the proposition of a combined set of features based on HSV channels and Mallat decomposition of each channel on approximation, horizontal, and vertical coefficients. A spatial grid, based on 9 regions, to improve localization, and a protocol of experiments using three levels of the decomposition in order to find the most reduced sets were proposed and evaluated. With the best results on those conditions, using 27 different individual sets, we picked the 3 best ones as a useful and efficient reduced set and evaluated the new combined feature. We have found that the new set improved the classification results in the database tested. Those results encourage us to explore further with this feature set, particularly in special purpose image retrieval tasks such as with medical databases, and with feedback relevance schemes where evidence combination together with small sets of features are good characteristics to hold.

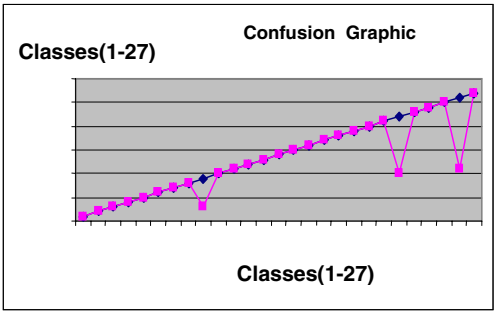


Fig. 10. Confusion graphic: combining the best results (i.e. selected feature set): H, V (approximation coefficients) + S (vertical coefficients), all in the third level of decomposition. (*x* axis is the class of the query image, and *y* axis is the result classification obtained).

4 Conclusions

We have presented in this paper an evaluation of a combined set of features to be used in image retrieval tasks. As can be seen from the experiments the combined new feature is an efficient approach to content-based image retrieval that combines color and wavelet coding in a zoning scheme. A set of features is derived from the HSV channels by computing wavelet coefficients in each of them and selecting upon the most significant the ones to index the image. Different than others we did not use global features as moments (e.g. see [1], [2], [6] and [7] for other approaches in image retrieval using wavelet features), but tested to check the combined performance of HSV channels, coarse and detail coefficients in three levels of decomposition using the most significant ones. Extensive testing was done in order to end with only 3 small sets for final similarity measure and rank. One of the difficulties in content-based research is that the databases may have multiple and yet acceptable classifications. Instead of giving only a precision x recall curve we plotted where the misclassifications occurred, considering the voted highest result achieved with all the training set of the images. Of course we do not advocate our approach to be a final word for this, it is still a challenging problem. However we have shown it to be a direct, and generic method (i.e. deal with different types of image classes), and with competitive successful results, not evaluated before, to derive significant features for use in image retrieval. Also, performance measures on a reasonably sized database is given. Although tests were not performed in full using other databases such as Corel, and others with more than 20 thousand images, for the purpose of validating as a useful and reduced set of features the experiments given were meaningful. Future works will deal with relevance feedback for consistent uncertainty treatment [12], special purpose medical databases, and further tests with bigger available databases for benchmarking purposes in image retrieval.

Acknowledgments

This work was partially supported by CAPES/MEC, and CNPq/MCT.

References

[1] E. Albuz, E. Kocalar, ad A. Khokhar," Scalable color image indexing and retrieval using vector wavelets," *IEEE Trans. on Knowledge and Data Eng*., vol. 13, no. 5, pp. 851-861, Sep. 2001.

[2] Brambilla, C. Ventura, D.A., Gagliardi, I and Schettini R. "Multiresolution Wavelet Transform and Supervised Learning for Content-based Image Retrieval". *IEEE Multimedia Systems 99*, IEEE CS Press, Vol. I, pp. 183-188, 1999.

[3] S.-C. Chen, S. Sista, M.-L. Shyu, and R.L. Kashyap, "An Indexing and Searching Structure for Multimedia Database Systems," IS&T/SPIE Conference on Storage and Retrieval for Media Databases 2000, pp. 262-270, 2000.

[4] Du Buf J.M.H., Kardan M., Spann M. Texture feature performance for image segmentation. *Pattern Recognition*, Vol. 23, pp. 291-309, 1990.

[5] M. Flickner, H. Sawhney, W. Niblack, and J Ashley. Query by image and video content: the qbic system. *IEEE Computer*, 28(9), pp.23--32, Sep. 1995.

[6] Liang, K. and Kuo, C.C. WaveGuide: a joint wavelet-based image representation and description system. *IEEE Transactions on Image Processing*. Vol. 8, n. 11, pp. 1619-1629, 1999.

[7] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. IEEE Trans. Patt. Anal. Machine Intell., vol.18, pp. 837--842, Aug. 1996.

[8] Rao A.R., Lohse G.L. Identifying High level Features of texture Perception. *CVGIP: Graphic Models and Image Processing*, Vol. 55(3), pp. 218-233, 1993.

[9] Resnikoff, H. and Wells Jr., R. *Wavelet Analysis.* Springer-Verlag, 1998.

[10] Smeulders, A. W.M., Worring, M., Santini, S., Gupta, A., Jain, R. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 22, n. 12, pp. 1349-1380, 2000.

[11] Swain, M. and Ballard, D. Color Indexing. *Int. J. Computer Vision*. Vol. 7, pp. 11-32, 1991.

[12] Yavlinsky, A., Pickering, M., Heesch, D., Ruger, S.: A comparative study of evidence combination strategies. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Volume III, pp 1040—1043, 2004.