Toward Confident Level of Image Quantization in Image Database Management System based on Color Features

Irena Valova, Boris Rachev

Abstract *- In this paper, we describe our research in content based image retrieval systems, based on color features. We describe our experiments, based on global and local color properties of an image and examine the effect of the degree of image quantization on similarity. We drew the following conclusions from the analysis of the results of our experiments: 12X12 quantized images may be used to distinguish similar images from dissimilar images for achieving the purpose of image retrieval and the amount of details in an image is not correlated with retrieval performance at different quantization levels, provided the quantization level is greater than 12.*

Keywords*: Image databases, content based image retrieval, image features, image quantization*

INTRODUCTION

There is a rapid increase in the size of digital image collections together with the fast growth of the Internet in the recent years. Digital images have found their way into many application areas, including Geographical Information System, Office Automation, Medical Imaging, Computer Aided Design, Computer Aided Manufacturing, and Robotics.

Content-based image retrieval represents a promising and cutting-edge technology to address these needs. The fundamental idea of this approach is to generate automatically image descriptions directly from the image content by analyzing the content of the images. Such techniques are being developed by many research groups and commercial companies around the world.

A typical Content-based Image Retrieval (CBIR) system deals not only with various sources of information in different formats (for example, text, image, video) but also user's requirements. Basically it analyzes both the contents of the source of information as well as the user queries, and then matches these to retrieve those items that are relevant. The major functions of such a system are described in [4]. Color is the first and most straightforward visual feature for indexing and retrieval of images [7, 8, 9.]. It is also the most commonly used feature in the field.

DEFINITION OF THE PROBLEM

In our previous work in the field of color content based image retrieval systems [1, 2, 3, 5] we proposed two types of image descriptors. First one is based on global color features of the images and represents the dominant colors in the images. It is used for hierarchical classification of the images in image database.

Another color descriptor, called color descriptor matrix, we propose to describe local color features or color distribution in the images. The original images were NxN quantized and were represented as NxN blocks (or sub images). We examined the creation and retrieval time and database size depending on the size of N in [2] and made the conclusion that N=16 is the most adequate for our purposes. In order to create this index structure the whole image is divided into 256 equal parts. This matrix stores the coefficient of the dominant color from the selected color code book in the corresponding part of the image. The advantages and disadvantages of these two structures are presented in [1, 3]. Also in [3, 5] is described the main algorithm for image database organization and retrieval.

The main problem is ***to discover how low the resolution of the quantized image may be used keeping a balance between loss of image information about color distribution and the loss of differentiation between different images*****.**

SOLUTION OF THE PROBLEM, EXPERIMENTS AND RESULTS

We propose an experimental solution of the problem by using statistically probability methods for processing the information for similarity between predefined and appropriate grouping images and query image. We formulate and verify the hypotheses that there is a minimal quantization level to distinguish similar images from dissimilar images.

In this experiment we use the algorithm of similarity measures to calculate *the similarities* between a query image and 19 sample images by using *29 different quantization levels*. Our prototype system supports following query types: Browsing; Query by Global Color Distribution, Query by User Sketch and Query by Image Example [3, 5]. Specifically, the method of "Query-by-Image-Example" is used here.

Figure 1 The query image (first left one) and the selected sample images, used in the experiment

These 19 sample images were randomly selected from an image collection, which were expected to include some "very similar", "similar", or "dissimilar" images to the query image. The query image and the selected sample images are illustrated in Figure1. We use our prototype system and create color content based image database, compound of these 20 images for 29 different quantization levels and examine the similarities of these 19 images with the query image. The results of this experiment about the similarities with 29 different levels of quantization (from 32X32 to 4X4) were listed in Table 1. Graphical representation of these results is shown in figure 2.

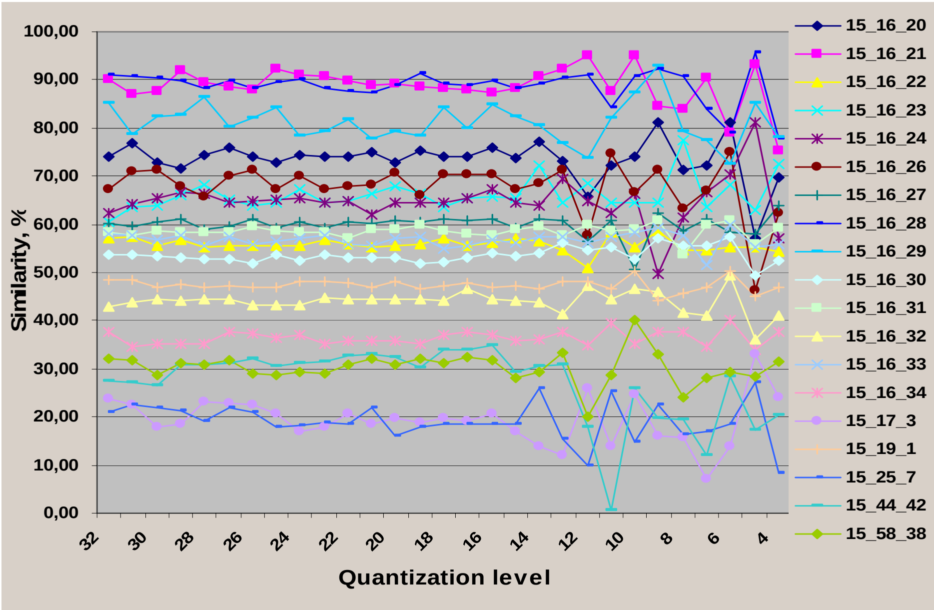

Figure 2 Graphical representation of the similarity change depending on the image quantization level

To examine the size of the quantization level of images, the experimental data was analyzed by using a statistical method, *two-tailed statistical test*. The null hypothesis is that ***the similarity measure maintains the rank order of images independent of the quantization level used***.

Give a 95% confidence interval for 29 different levels of resolutions; *the critical value* of the statistic that separates *the rejection region and the acceptance region is* 2.0484 as the degrees of freedom is 28 (i.e. *D=29-1)* [6].

By using the *t*-test [6], the *t*-scores of the data were listed in table 2 and figure 3. Since the confidence interval can be decided when the absolute value *t* should be smaller than 2.0484. Therefore, by observing the results of the *t*-test for 19 image samples, overall the sample data which had ***larger than 12 quantization level*** are correct theoretically.

CONCLUSIONS AND FUTURE WORK

In conclusion, after the analyses and the summary of these results of the experiment we assert that *at quantization level 12, there is 95% confidence (or better) that dissimilar images are in fact still dissimilar*. That is, the quantized images may be used to distinguish similar images from dissimilar images for achieving the purpose of image retrieval.

In addition, *the amount of detail in an image is not correlated with retrieval performance at different quantization levels, provided the quantization level is greater than 12*. Therefore, the use of the size of 16X16 in the prototype of the Image Database Management System by Color Features is reasonable.

The next stage in the development of our prototype system is to analyze the effect of the selected colors and their number in the proposed color code book on the similarity of images in process of organization and retrieval.

Table 1 Similarity change depending on the image quantization level

	Similarity, %																		
Quanti-zation	15_16_20	15_16_21	15_16_22	15_16_23	15_16_24	15_16_26	15_16_27	15_16_28	15_16_29	15_16_30	15_16_31	15_16_32	15_16_33	15_16_34	15_17_3	15_19_1	15_25_7	15_44_42	15_58_38
32	74,11	90,11	57,11	60,51	62,47	67,31	60,31	91,03	85,31	53,62	58,40	42,90	58,10	37,80	23,70	48,60	20,90	27,60	32,00
31	76,93	86,93	57,32	63,52	64,21	70,84	59,51	90,74	78,82	53,71	57,69	43,70	57,70	34,70	22,40	48,60	22,60	27,10	31,80
30	72,70	87,59	55,47	63,85	65,47	71,38	60,52	90,38	82,51	53,44	58,70	44,30	56,72	35,10	17,90	46,90	21,80	26,60	28,80
29	71,72	92,03	56,81	65,97	66,79	68,05	61,05	89,87	82,65	53,11	58,40	44,00	57,60	35,30	18,50	47,60	21,20	31,00	31,20
28	74,53	89,56	55,40	68,23	66,48	65,69	59,09	88,40	86,34	52,93	58,40	44,30	55,60	35,10	23,20	47,00	19,10	31,00	30,90
27	75,82	88,65	55,63	65,12	64,41	70,21	59,53	89,96	80,21	52,87	58,20	44,30	57,30	37,80	22,90	47,20	21,80	31,20	31,80
26	74,21	87,92	55,42	63,83	64,94	71,29	61,07	88,23	82,12	51,82	59,50	43,30	56,00	37,30	22,40	47,00	20,90	32,00	29,10
25	72,75	92,40	55,57	64,42	64,97	67,32	59,13	89,55	84,11	53,72	58,50	43,10	56,40	36,30	20,60	46,90	17,90	30,70	28,80
24	74,53	91,14	55,42	67,16	65,41	70,21	60,51	90,11	78,34	52,36	58,20	43,30	57,10	36,90	17,10	48,20	18,20	31,20	29,40
23	74,17	90,65	56,77	64,45	64,49	67,33	59,13	88,25	79,27	53,60	58,40	44,90	57,60	35,30	17,90	48,20	18,80	31,50	29,10
22	74,18	89,88	55,73	64,72	64,74	68,03	60,42	87,58	81,64	53,21	57,10	44,50	55,90	35,70	20,60	47,90	18,50	32,80	30,90
21	75,08	88,75	55,32	66,48	62,03	68,32	60,21	87,40	77,82	53,10	58,80	44,30	55,60	35,90	18,50	47,00	21,80	33,00	32,00
20	72,72	89,24	55,70	68,02	64,41	70,71	60,72	88,95	79,23	53,20	59,00	44,30	57,00	35,70	19,90	48,00	16,20	32,30	30,90
19	75,17	88,50	55,83	66,48	64,49	65,97	60,42	91,23	78,37	51,71	59,90	44,50	57,40	35,10	18,80	46,70	17,90	30,20	32,00
18	74,18	88,18	57,03	63,50	64,64	70,27	61,12	89,35	84,41	52,20	58,50	44,10	55,00	37,10	19,90	47,30	18,50	34,10	31,20
17	74,15	87,95	55,69	65,33	65,40	70,32	60,92	88,80	79,95	53,21	57,90	46,50	55,60	37,60	19,10	47,80	18,50	33,80	32,40
16	75,83	87,42	56,07	65,62	67,31	70,22	61,21	89,92	84,82	53,98	57,69	44,50	56,80	36,90	20,80	47,00	18,50	34,80	31,80
15	73,68	88,40	57,15	65,37	64,45	67,30	59,12	88,34	82,32	53,40	58,80	44,00	56,20	35,90	17,10	47,20	18,50	29,40	28,00
14	77,31	90,64	56,33	72,23	64,04	68,63	61,23	89,11	80,54	53,90	59,60	43,70	57,30	36,00	13,80	46,70	25,90	30,50	29,40
13	73,14	92,18	54,64	64,41	69,31	71,21	60,91	90,39	76,95	56,26	57,69	41,40	57,30	37,80	12,00	48,20	15,30	30,90	33,20
12	65,62	95,12	50,92	68,55	64,92	57,63	56,52	90,95	73,84	54,64	60,00	47,30	56,20	34,90	25,90	48,20	10,00	17,90	20,00
11	72,11	87,60	58,47	64,42	62,41	74,65	60,91	84,33	82,02	55,33	58,50	44,30	57,40	39,40	13,80	46,60	25,30	0,60	28,80
10	74,13	94,96	55,10	64,41	66,48	66,82	50,54	90,74	87,39	52,92	59,00	46,50	58,40	35,30	24,70	50,40	14,70	25,90	40,00
9	81,18	84,55	58,72	64,44	49,80	71,21	62,32	92,33	92,81	57,07	60,70	45,90	59,30	37,80	16,00	44,20	22,60	19,70	32,90
8	71,31	83,90	54,66	77,46	61,32	63,42	58,71	90,65	79,25	55,61	53,70	41,80	57,70	37,50	15,60	45,80	16,50	19,40	24,00
7	72,14	90,46	54,62	63,51	66,68	66,86	61,22	83,88	77,32	55,62	59,90	41,00	51,50	34,50	7,10	46,90	17,10	12,00	28,20
6	81,11	79,15	55,37	68,14	70,23	75,14	58,43	79,16	72,64	57,43	60,90	49,40	59,80	40,20	13,80	50,30	18,50	28,50	29,40
5	57,23	93,19	55,33	62,91	81,03	46,33	57,91	95,55	85,12	49,42	56,00	36,00	55,30	33,90	32,90	45,20	27,10	17,40	28,50
4	69,84	75,20	54,22	72,63	57,15	62,34	63,85	77,80	78,18	52,51	59,30	41,20	55,90	37,80	24,00	47,00	8,20	20,30	31,50
AVG	**73,50**	**88,70**	**55,79**	**66,06**	**64,84**	**67,76**	**59,88**	**88,72**	**81,18**	**53,65**	**58,53**	**43,91**	**56,75**	**36,43**	**19,34**	**47,40**	**19,06**	**27,01**	**30,28**
S	**4,30**	**4,14**	**1,43**	**3,39**	**4,82**	**5,42**	**2,28**	**3,58**	**4,17**	**1,66**	**1,39**	**2,34**	**1,52**	**1,50**	**4,95**	**1,26**	**4,08**	**7,75**	**3,29**

Table 2 Statistical results

Quanti-zation	t-score 15_16_20	15_16_21	15_16_22	15_16_23	15_16_24	15_16_26	15_16_27	15_16_28	15_16_29	15_16_30	15_16_31	15_16_32	15_16_33	15_16_34	15_17_3	15_19_1	15_25_7	15_44_42	15_58_38
32	0,1411	0,3406	0,9235	-1,6365	-0,4928	-0,0829	0,1883	0,6448	0,9888	-0,0191	-0,0935	-0,4298	0,8869	0,9103	0,8809	0,9512	0,4503	0,0756	0,5233
31	0,7965	-0,4266	1,0701	-0,7487	-0,1316	0,5688	-0,1626	0,5637	-0,5661	0,0350	-0,6042	-0,0883	0,6243	-1,1563	0,6182	0,9512	0,8668	0,0111	0,4626
30	-0,1866	-0,2674	-0,2212	-0,6513	0,1299	0,6684	0,2805	0,4631	0,3180	-0,1272	0,1223	0,1678	-0,0190	-0,8896	-0,2913	-0,3963	0,6708	-0,0534	-0,4479
29	-0,4143	0,8039	0,7141	-0,0260	0,4039	0,0537	0,5130	0,3205	0,3515	-0,3256	-0,0935	0,0397	0,5587	-0,7563	-0,1701	0,1585	0,5238	0,5143	0,2805
28	0,2387	0,2079	-0,2701	0,6406	0,3396	-0,3819	-0,3469	-0,0904	1,2356	-0,4337	-0,0935	0,1678	-0,7543	-0,8896	0,7799	-0,3171	0,0093	0,5143	0,1894
27	0,5385	-0,0116	-0,1095	-0,2767	-0,0901	0,4525	-0,1539	0,3457	-0,2331	-0,4698	-0,2374	0,1678	0,3617	0,9103	0,7192	-0,1585	0,6708	0,5402	0,4626
26	0,1644	-0,1878	-0,2561	-0,6572	0,0199	0,6518	0,5218	-0,1379	0,2246	-1,1008	0,6978	-0,2591	-0,4917	0,5770	0,6182	-0,3171	0,4503	0,6434	-0,3569
25	-0,1749	0,8931	-0,1514	-0,4832	0,0261	-0,0810	-0,3293	0,2310	0,7013	0,0410	-0,0216	-0,3445	-0,2291	-0,0897	0,2544	-0,3963	-0,2847	0,4756	-0,4479
24	0,2387	0,5891	-0,2561	0,3250	0,1175	0,4525	0,2761	0,3876	-0,6811	-0,7763	-0,2374	-0,2591	0,2304	0,3103	-0,4530	0,6342	-0,2112	0,5402	-0,2658
23	0,1551	0,4709	0,6862	-0,4744	-0,0735	-0,0792	-0,3293	-0,1323	-0,4583	-0,0311	-0,0935	0,4240	0,5587	-0,7563	-0,2913	0,6342	-0,0642	0,5789	-0,3569
22	0,1574	0,2851	-0,0397	-0,3947	-0,0216	0,0500	0,2366	-0,3196	0,1095	-0,2655	-1,0286	0,2532	-0,5573	-0,4896	0,2544	0,3963	-0,1377	0,7466	0,1894
21	0,3666	0,0125	-0,3259	0,1244	-0,5841	0,1036	0,1445	-0,3699	-0,8057	-0,3316	0,1942	0,1678	-0,7543	-0,3563	-0,1701	-0,3171	0,6708	0,7724	0,5233
20	-0,1819	0,1307	-0,0607	0,5786	-0,0901	0,5448	0,3682	0,0633	-0,4679	-0,2715	0,3381	0,1678	0,1648	-0,4896	0,1129	0,4756	-0,7012	0,6821	0,1894
19	0,3875	-0,0478	0,0301	0,1244	-0,0735	-0,3302	0,2366	0,7007	-0,6739	-1,1669	0,9855	0,2532	0,4274	-0,8896	-0,1094	-0,5549	-0,2847	0,4111	0,5233
18	0,1574	-0,1250	0,8677	-0,7546	-0,0424	0,4635	0,5437	0,1751	0,7732	-0,8724	-0,0216	0,0824	-1,1482	0,4437	0,1129	-0,0793	-0,1377	0,9143	0,2805
17	0,1504	-0,1805	-0,0676	-0,2148	0,1154	0,4728	0,4560	0,0214	-0,2954	-0,2655	-0,4532	1,1070	-0,7543	0,7770	-0,0488	0,3171	-0,1377	0,8756	0,6447
16	0,5409	-0,3084	0,1976	-0,1293	0,5118	0,4543	0,5832	0,3345	0,8714	0,1973	-0,6042	0,2532	0,0335	0,3103	0,2948	-0,3171	-0,1377	1,0047	0,4626
15	0,0412	-0,0720	0,9515	-0,2030	-0,0818	-0,0847	-0,3337	-0,1072	0,2725	-0,1513	0,1942	0,0397	-0,3604	-0,3563	-0,4530	-0,1585	-0,1377	0,3079	-0,6907
14	0,8848	0,4685	0,3791	1,8204	-0,1669	0,1608	0,5920	0,1081	-0,1540	0,1492	0,7697	-0,0883	0,3617	-0,2896	-1,1200	-0,5549	1,6754	0,4498	-0,2658
13	-0,0843	0,8401	-0,8005	-0,4862	0,9270	0,6371	0,4516	0,4659	-1,0141	1,5675	-0,6042	-1,0702	0,3617	0,9103	-1,4838	0,6342	-0,9218	0,5014	0,8875
12	-1,8320	1,5494	***-3,3971***	0,7349	0,0157	-1,8698	-1,4744	0,6224	-1,7592	0,5939	1,0574	1,4485	-0,3604	-1,0229	1,3256	0,6342	-2,2203	-1,1760	-3,1186
11	-0,3237	-0,2650	***1,8728***	-0,4832	-0,5052	1,2721	0,4516	-1,2281	0,2006	1,0086	-0,0216	0,1678	0,4274	1,9769	-1,1200	-0,6342	1,5284	-3,4082	-0,4479
10	0,1458	1,5108	***-0,4795***	-0,4862	0,3396	-0,1733	***-4,0978***	0,5637	1,4872	-0,4397	0,3381	1,1070	1,0839	-0,7563	1,0830	***2,3781***	-1,0688	-0,1437	2,9512
9	1,7843	-1,0009	***2,0473***	-0,4773	***-3,1227***	0,6371	***1,0701***	1,0082	***2,7857***	***2,0543***	1,5610	0,8509	1,6747	0,9103	-0,6753	***-2,5366***	0,8668	-0,9437	0,7964
8	-0,5096	-1,1577	***-0,7866***	***3,3629***	***-0,7315***	-0,8010	***-0,5136***	0,5385	***-0,4631***	***1,1769***	***-3,4744***	-0,8994	0,6243	0,7103	-0,7562	***-1,2683***	-0,6277	-0,9824	-1,9047
7	-0,3167	0,4251	***-0,8145***	***-0,7516***	***0,3811***	-0,1659	***0,5876***	-1,3539	***-0,9255***	***1,1829***	***0,9855***	-1,2410	***-3,4458***	-1,2896	***-2,4741***	***-0,3963***	-0,4807	-1,9373	-0,6300
6	1,7680	***-2,3038***	***-0,2910***	***0,6140***	***1,1179***	1,3625	***-0,6364***	***-2,6733***	***-2,0467***	***2,2706***	***1,7048***	***2,3450***	***2,0029***	***2,5102***	***-1,1200***	***2,2988***	-0,1377	0,1918	-0,2658
5	***-3,7819***	***1,0837***	***-0,3189***	***-0,9286***	***3,3597***	***-3,9557***	***-0,8646***	***1,9082***	***0,9433***	***-2,5431***	***-1,8199***	***-3,3755***	***-0,9512***	***-1,6896***	***2,7404***	***-1,7439***	1,9694	-1,2405	-0,5390
4	***-0,8513***	***-3,2568***	***-1,0937***	***1,9383***	***-1,5971***	***-1,0003***	***1,7414***	***-3,0535***	***-0,7194***	***-0,6861***	***0,5539***	***-1,1556***	***-0,5573***	***0,9103***	***0,9416***	***-0,3171***	***-2,6613***	-0,8663	0,3715

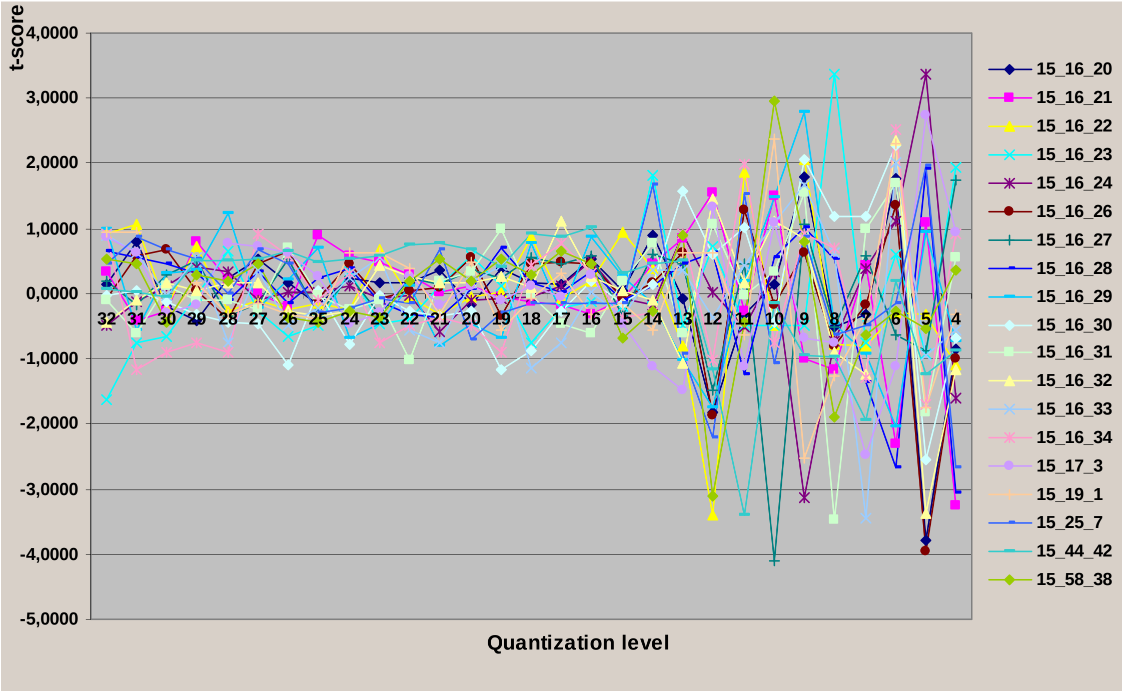

Figure 3 Statistical results – graphical representation

ACKNOWLEDGEMENTS: This work is supported by Bulgarian Ministry of Education and Science – Grant DBG 18/05 and in accordance with the Bulgarian-Greek Joint Research Programme

REFERENCES

1. Boris Rachev, Irena Valova, Silyan Arsov, An Approach for Image Organization and Retrieval in Realistic Image Databases, 7th EC-GIS Workshop, EG II-Managing the Mosaic, Potsdam, Germany, 13-15.06.2001
2. Valova, B. Rachev, Image Databases – an Approach to Image Segmentation&Color Reduction Analysis&Synthesis, International Conference CompSysTech'2003, Sofia, Bulgaria, 19-20 June 2003
3. Valova, B. Rachev, Retrieval By Color Features In Image Databases, Adbis'04, 22-25 September, 2004, Budapest, Hungary
4. Irena Valova, Boris Rachev, A Content Based Image Retrieval System Based on Color Features, November 2004, CODATA, Berlin Germany
5. Valova, Irena; Rachev, Boris (Bulgaria): "An Algorithm for Organization and Retrieval by Color Features in Image Databases", International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2004; July 21 - 25, 2004 in Orlando, Florida, USA, Volume IV;
6. Mitkov, At., D. Minkov Mathematical Methods for Engineering Research, Rousse 1985
7. Swain, M. J. and Ballard, D. H. (1991). Color indexing. International Journal of Computer Vision, 7(1):11–32.
8. Schettini, R., Ciocca, G., and Zuffi, S. (2000). Color in databases: Indexation and similarity. In Proc. of Int'l Conf. on Color in Graphics and Image Processing, pages 244–249.
9. Schettini, R., Ciocca, G., and Zuffi, S. (2001). Color Imaging Science: Exploiting Digital Media, Ed. R. Luo and L. MacDonald, chapter A Survey on Methods for Colour Image Indexing and Retrieval in Image Database. John Wiley

ABOUT THE AUTHORS

Irena Marinova Valova, University of Rousse, Phone:++359 82 888695, Irena@ecs.ru.acad.bg
Boris Rachev, TU – Varna, тел. 052 302431 (407), e-mail: RACHEV@ms.ieee.bg