A Qualitative Examination of Content-Based Image Retrieval Behavior using Systematically Modified Test Images

James C. French, Worthy N. Martin and James V. S. Watson

Department of Computer Science
University of Virginia
Charlottesville, VA 22904, USA

ABSTRACT

We describe the outcome of an effort to understand the behavior of content-based image retrieval (CBIR) technology by examining the behavior of a CBIR system in response to carefully constructed input query images. This work is preliminary and only considers a single CBIR system, SIMPLIcity[4]. We chose this particular system for expediency. It is our intention to develop a methodology suitable for examining any CBIR system.

1. INTRODUCTION

Content-based image retrieval (CBIR) has received much attention in recent years. The issues are well known and are now covered in general texts, e.g., [2,3]. Its utility and promise are generally accepted as self-evident. There has been considerable research into specific technologies, but the full promise has yet to be realized. More importantly, we have not yet clearly explicated the issues surrounding this lack of success. In this paper we begin to consider an approach for qualitatively assessing the behavior of CBIR technology by appeal to carefully controlled queries over a testbed. This blackbox approach seeks to discover biases in CBIR technologies towards various features of the images being retrieved. We hope to understand CBIR issues more clearly by analyzing these biases.

2. TEST IMAGES

We are trying to qualitatively access how various image features are used by the underlying CBIR technology. Our approach to understanding the behavior of content-based image retrieval (CBIR) technology is by examining the behavior of a CBIR system in response to carefully constructed input query images. Figure 1 shows a query image (Original) together with nine transformations of the image. We submit each test query image to a CBIR system and compare the output (ranked list of images) to better understand the behavior of the system. The rationale behind the choice of these images is given next.

In the sequel we restrict our attention to CBIR technologies based on feature extraction and analysis. These approaches generally implement a similiarity measure that is a function of one or more of: color, shape, and texture. Shape only makes sense when some form of implicit or explicit segmentation is employed. We are specifically excluding the explicit incorporation of spatial information from similarity computations, for example, by means of templates. The next sections describe query formulations designed to probe the color and shape/texture responses of a CBIR system.

2.1 Color Response

The goal here was to investigate the impact of color on the retrieval algorithm. The idea is to alter the color of the image in specific ways to gauge the impact of these changes on the performance of the algorithm. Four images were constructed from the original query image. The phrase given in parentheses is the name used for each transformed image.

The grayscale (Gray).
A color negative with each channel "inverted" (Negative).
The RGB values rotated to GBR (Rotate1).
The RGB values rotated to BRG (Rotate2).

2.2 Shape/Texture Response

The goal here was to investigate the impact of shape or texture information on the retrieval algorithm. We created four test images, again each is a transformation of Original.

Pixels in each 4 X 4 patch randomly permuted (Local Shuffle).
All pixels randomly permuted (Global Shuffle).
Pixels sorted and stored row major order (Sort1).
Pixels sorted and stored column major order (Sort2).

These test queries have an identical global color histogram to the original query, however, they manifest decidely different textures.

2.3 Abstract Mimic Response

This image was motivated specifically by the SIMPLIcity system. Here we attempted to mimic the SIMPLIcity clustering algorithm by presenting a test image that was very similar in composition to the clustered image of the original query image as derived by SIMPLIcity.

Hand-crafted image - one foreground item (Two Color Block).

3. RESULTS

Each of the test images described in Section 2 was presented to the SIMPLIcity CBIR system. The image identifiers of the top 31 responses to each query are summarized in Table 1. The image id for Original is 6248. Thumbnails of the actual responses

0-7803-7523-8/02/$17.00 ©2002 IEEE

are given in French et al.[1]. Space limitations precluded their inclusion here. The specific outcomes and their implications are discussed below.

Each column of Table 1 (except the leftmost) is the rank-ordered response list returned by SIMPLIcity when the image named at the top of the column was given as the query image. Thus, the Original column constitutes the standard against which the other response lists are compared. Note that the Original column is not quite a full "ground truth" because there was no prior "intent" specified with regard to the choice of Original. It is true that the rank one response to Original is indeed Original, however, the reader is left to imagine whether it constitutes a query for "horse" pictures, "green and red" pictures, "nature" pictures or "mother and child" pictures. Unlike Original, the test images are not actually elements of the image repository.

The typographical conventions used in Table 1 are as follows. **Boldface** indicates a response image in common with the original query that has been placed at the same rank as in the original query. *Italic* indicates a response image in common with the original query that has been placed at a different rank. The other entries indicate response images not listed for the original query. At the bottom of each column, a total of each type is given. Also, the Jaccard measure of set (unordered) similarity is given.

Note that the first image in the thumbnail responses shown in [1] is the outcome of the SIMPLIcity clustering algorithm (from which Two Color Block was created).

3.1 Color Response

The intention here is to investigate color response without perturbing the shape/texture information. As can be seen from Table 1 the altered images produced 124 response images, 31 unique images per test image, and none of these images is contained in the response set of the original query. The obvious implication is that the SIMPLIcity algorithm is very strongly influenced by color.

Note that SIMPLIcity uses LUV encoding while we altered the images by manipulating RGB values. This may have had some effect on the clustering algorithm.

3.2 Shape/Texture Response

The intention here is to investigate the shape/texture response without perturbing the global color information. The local pixel shuffle test query has 28 of 31 images in common with the original query and overlaps the top 12 having the first 7 at identical ranks. The global shuffle test query had no images in common with the original query. Since both queries have identical color histograms, it is clear that texture is important to the SIMPLIcity algorithm. The local pixel shuffle results in a blurred image while the global pixel shuffle completely obscures the shape information in the image content. The test queries with sorted pixels do return some images in common with the original query. These images exhibit a strong horizontal (RM) or vertical (CM) separation of content, that is, the aggregate texture is a dominant factor. This can be seen in the thumbnails in [1].

3.3 Abstract Mimic Response

This test image is a two-color image roughly corresponding to the foreground and background of the original image. The brown region is approximately a bounding rectangle of the brown region of the original query. Table 1 shows that this test query has 15 images in common with the original query and both queries place the correct image at rank one.

4. CONCLUSIONS

The entries in Table 1 lead to a couple of observations. Local Shuffle is virtually identical to Original, but the response list has many order changes and more imterestingly, raises a new response image all the way up to rank 13. The Color Response images (Gray, Negative, Rotate1 and Rotate2) all leave the underlying "edge" information (at least perceptually) unchanged, yet absolutely no common responses are found. Again perceptually, Sort1 and Sort2 bear no resemblence to Original, yet a couple common images were returned. Finally, it is rather amazing that the rank one response for Two Color Block is Original (and three of the top five are in common). However we constructed Two Color Block to closely mimic Original in the form that SIMPLIcity has if for indexing purposes, so analytically that maybe should be expected. Then the question is, why were the other two of the top five completely new response images? Clearly, there are many non-intuitive factors in the similarity measure that this (and we would claim, all) CBIR system. In particular, the skew of the image collection that allows Two Color Block to get an exact hit, while Local Shuffle has many reorderings and a couple non-common response images.

5. REFERENCES

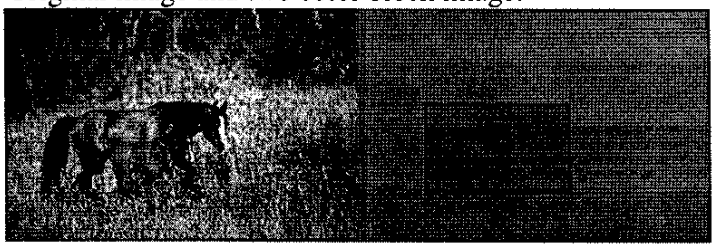
[1] French, J. C., W. N. Martin and J. V. S. Watson, "A Qualitative Examination of Content-Based Image Retrieval Behavior using Systematically Modified Test Images," Technical Report CS-2002-22, Department of Computer Science, University of Virginia, August 2002.

[2] Lew, M. S. (Ed.), Principles of Visual Information Retrieval. Springer, 2001.

[3] Santini, S. Exploratory Image Databases: Content-based Retrieval. Academic Press, 2001.

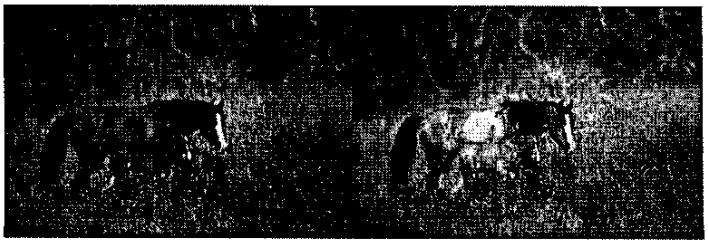
[4] Wang, J. Z. and Y. Du. "Scalable Integreated Region-based Image Retrieval using IRM and Statistical Clustering." *Proceedings of he First ACM/IEEE Joint Conference on Digital Libraries*, 2001, pp. 268-277.

Original image and two color block image:



Gray and Negative:

Color rotations:



Local and global pixel shuffle:

Sorts:

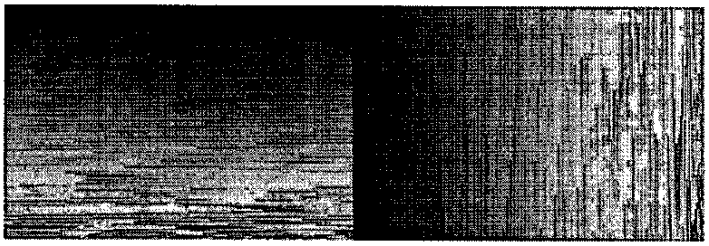


Figure 1: Original query plus nine test images.

Result Lists for CBIR Test Images										
Rank	Original	Two Color Block	Gray	Negative	Rotate 1 (GBR)	Rotate 2 (BRG)	Local Shuffle	Global Shuffle	Sort 1 (RM)	Sort 2 (CM)
1	6248	**6248**	9570	58447	43756	35068	**6248**	32352	*32427*	*32427*
2	6227	6246	35712	58317	29341	35262	**6227**	12991	34075	35111
3	5714	*6227*	51925	30001	43570	35138	**5714**	31904	6141	53313
4	25419	*5714*	35764	30468	31900	35225	**25419**	13264	19953	33350
5	46894	17446	24791	8254	23405	32523	**46894**	45417	*4921*	19953
6	1776	*6274*	35714	49332	42306	35187	**1776**	36653	33350	35103
7	12244	54434	55544	35795	23418	35762	**12244**	45111	12231	*46894*
8	6274	6238	22826	6792	13855	26468	*7315*	21379	19675	19675
9	7315	*58132*	7044	5872	57879	35191	*6274*	25437	52042	13231
10	6280	*7315*	20275	17543	31004	46361	*29288*	35227	35140	6141
11	25436	6241	40008	58529	49027	14616	*25436*	21394	52062	8673
12	29288	*46894*	27914	30896	30998	9027	*6280*	32304	4300	35140
13	7065	*54438*	35719	1324	23403	18392	55350	32230	52608	4300
14	4921	7066	53356	5878	22998	35590	*54438*	38498	27031	52608
15	58258	54408	38457	13340	12214	2202	*7065*	45141	26195	22860
16	37906	*6297*	10193	54719	31765	35699	*4921*	19530	*46894*	32378
17	54438	37922	1404	20232	54748	34496	*58258*	31848	5070	23168
18	5726	*4921*	24752	58832	17892	35545	*37906*	44635	20921	15117
19	50638	19378	31243	20202	56852	58409	**50638**	31843	13210	5410
20	6220	55345	53040	34408	37436	35718	*29234*	27312	26162	20252
21	13277	8704	20298	4259	30910	10632	*5726*	31849	55478	26195
22	29234	53631	50431	1403	24758	24895	*29278*	38745	33336	*4921*
23	6207	*29234*	49301	3523	27202	22508	*6220*	37616	5305	5480
24	4316	6260	24727	24312	35021	16727	*6207*	48532	10472	5070
25	29278	*5726*	28020	9681	35580	5564	*4316*	36659	24267	46793
26	58132	*13277*	55110	35056	35816	5272	**58132**	42420	*6248*	45915
27	6297	*1776*	17375	8218	24753	58413	*35264*	1726	12287	*12244*
28	16733	*50638*	21648	30932	14483	31096	*6297*	31233	4157	*26321*
29	35264	21771	14076	8222	32608	35721	*32427*	48520	12851	*54438*
30	32427	54401	39604	25540	35047	35124	55345	32368	14942	27925
31	35618	8708	20201	10804	20284	41765	53631	21307	31467	4199
Bold		1	0	0	0	0	9	0	0	0
Italic		14	0	0	0	0	19	0	4	5
Other		16	31	31	31	31	3	31	27	26
Jaccard		0.32	0.00	0.00	0.00	0.00	0.82	0.00	0.07	0.09

Table 1: Result lists for each query image in test set.