INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Exploring CBIR concepts in the CTRnet Project

N. P. Kozievitch *S. Codio* *J. A.Francois*
E. Fox *R.S. Torres*

Technical Report - IC-10-32 - Relatório Técnico

November - 2010 - Novembro

Exploring CBIR concepts in the CTRnet Project

Nadia P. Kozievitch[1], Sherley Codio[2], Jennifer A.Francois[2],
Edward Fox[2], Ricardo da S. Torres[1]

[1] Institute of Computing, University of Campinas (UNICAMP),
13084-851, Campinas-SP, Brazil

{nadiapk, rtorres}@ic.unicamp.br

[2] Department of Computer Science,
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

{shachou8, jenniff, fox}@vt.edu

Abstract. The project crises, Tragedy and Recovery Network (CTRnet) is an effort to build an integrated distributed digital library for providing a rich suite of CTR-related services. This report describes an independent study conducted at Virginia Polytechnic Institute and State University, consisting of collecting and archiving information related to the Haiti earthquake, later used to explore content-based image retrieval (CBIR) concepts. The objective was to collect and categorize relevant pictures related to the earthquake, followed by the exploration of practical CBIR concepts, such as descriptors, feature vectors, and experiment design.

Palavras-Chave: CTRnet, Content-Based Image Retrieval.

1. Introduction

Increasingly, individuals, groups, and communities are using information and communication technologies in innovative ways to learn from crises and tragedies, and to recover more quickly and effectively. However, the context and technologies involved in communication (e.g., Internet, WWW, online communities, mobile devices) make it exceedingly difficult to integrate content, community, and services. The project crises, Tragedy and Recovery Network (CTRnet) is an effort to address the gap between the availability and use of information and communication technologies in the context of the CTR domain.

CTRnet is an extension of prior work relating to the shootings of April 16, 2007 at Virginia Polytechnic Institute and State University [1]. The repository sources include Web 2.0, news, images, books, and scholarly articles. The objective is to develop better approaches toward making technology useful for archiving the events and conducting analysis on rescue, relief, and recovery, from a digital library (DL) perspective. CTRnet has several modules, including the Facebook application, user visualization, crawling, metadata search, and Content-Based Image Retrieval (CBIR).

Particularly, the CBIR module is important for retrieving multimedia information. CBIR is an approach that uses image visual properties, such as color, texture, and shape of objects for indexing and retrieving images. One advantage of this approach is that the image retrieval process is free from textual annotations or metadata, which are usually subjective and arduous to create.

The task of determining similarities among images relies on the use of image descriptors. The image descriptor is involved in extracting image features and comparing the images based on extracted feature vectors. It is very important to know which descriptors are more adequate for a system, so the image collection can be better explored.

This report describes an undergraduate research undertaken by the second and third authors of this report, with grad students, and faculty support, related to an independent study and term project along in CS4624 (Multimedia, Hypertext and Information Access - Spring 2010) at Virginia Polytechnic Institute and State University, using CBIR as a means to categorize/search two possible collections (human and non-human images) in CTRnet. Images related to the earthquake in Haiti were used as the primary dataset. Eva [2], a tool for evaluating image descriptors, was used for exploring the practical use of CBIR concepts (such as descriptors, feature vectors, experiments, image distances, etc.).

This document has the following organization: section 2 presents the background and related literature; section 3 shows the conducted study and challenges of building an image collection for CTRnet; and section 4 presents conclusions and discusses future work.

2. Background and Related Literature

This report is related to, and builds upon research in three areas, including crises and disaster management systems, CBIR, and particularly, the tool called Eva, used for exploring CBIR concepts.

2.1 Crises and Tragedy Management Systems

Professionals and communities use information and technology to better prepare for and to recover more quickly and effectively from disasters. From school shootings to earthquakes, communities, groups, and societies have been collaborating in innovative ways to learn from crises and tragedies. Sahana[6], Ushaidi[7], and CrisesCamp[1] are examples which bring together domain experts and community to improve technology and practice for humanitarian crises management and disaster relief. These initiatives explore a variety of technologies (such as web based tools and geographic information systems

1 http://crisescamp.eventbrite.com/

(GIS)) for the intensive exchange of textual data, images, and video.

The CTRnet effort is an extension of prior work relating to the shootings of April 16, 2007, at Virginia Polytechnic Institute and State University, USA. The advantages of CTRnet (if compared to other initiatives) is that, coming from a DL community, digital library services can be exploited, along with additional efforts, like Facebook interfacing, ontology buildings, and content-based image retrieval (CBIR).

2.2 Content-Based Image Retrieval (CBIR)

In general, two different approaches have been applied to allow searching on image collections: one based on image textual metadata and another based on image content [12].

The first retrieval approach is based on attaching textual metadata to each image and uses traditional database query techniques to retrieve them by keywords. One limitation of this approach relies on the annotation of the database images, which is a time-consuming task. Furthermore, the annotation is not usually made in a systematic way, decreasing the performance of the keyword-based image search.

The term CBIR was first introduced by T. Kato [2] to describe automatic retrieval of images, processed by algorithms, extracting feature vectors that represent image properties such as color, texture, and shape. In this approach, it is possible to retrieve images similar to one chosen by the user (query-by-example). One of the main advantages of this approach is the possibility of an automatic retrieval process, contrasting to the effort needed to annotate images.

The creation of CBIR systems involves research on databases and image processing, handling problems that vary from storage issues to friendly user interfaces. A typical CBIR solution requires the construction of image descriptors, which are characterized by:

- An extraction algorithm to encode image features into feature vectors;

- A similarity measure to compare two images based on the distance between the corresponding feature vectors; if the distance value is large, the images are less similar whereas if the distance is small, the images are more similar.

The database images are ranked according to their similarity and the most similar images are forwarded to the user.

Several CBIR systems have been proposed recently [12], such as QBIC, Chabot, and Photobook. Even though a few of them became commercial products, many CBIR systems were proposed as research prototype [14], being developed in universities and research laboratories. Additional efforts include the use of CBIR systems in education [15].

2.3 Eva Tool

Eva [2] comprises a tool for evaluating image descriptors for CBIR. Eva integrates the image database, descriptor, experiment, and visualization management, and provides functionalities to facilitate the comparison of image descriptors in the context of content-based image retrieval.

Eva has been implemented using PHP, Python, and C, and it has a Web interface that makes it accessible by any browser, as shown in Figure 1. The query image is highlighted at the top of the page and the retrieved image list is shown below it.

The PostgreSQL relational database was used to store information about image descriptors, image collections, experiment configurations, and results from experiments. Examples of experiments' results data include time required for extracting feature vectors, time for distance computations, effectiveness values from user effectiveness evaluations, and distances between images.

Other attempts in the community research to standardize the performance comparison in CBIR were already presented [3, 4, 5]. The advantage of the Eva evaluation tool is that additional descriptors and image databases can be managed, along with an web interface for experiment evaluation [2].

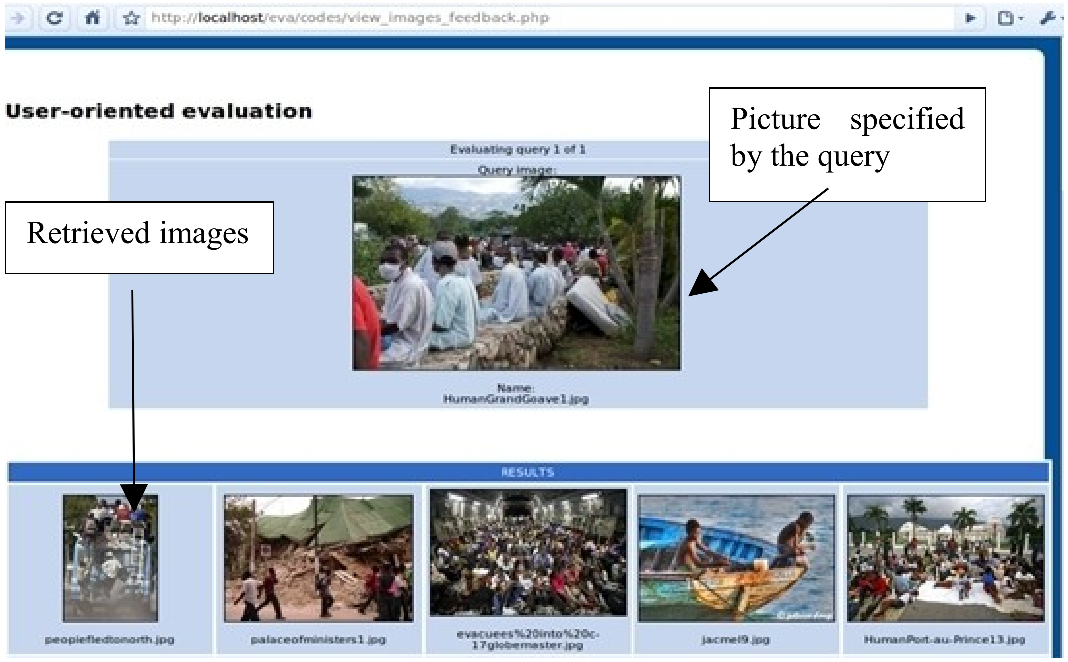


Fig.1. The web interface of Eva Tool.

3. Conducted Study

The conducted study with students from CS4624 (Spring 2010) had the following tasks: (1) each student selected approximately sixty images (before and after the disaster), and categorized them as human or not human (buildings, roads, etc.); (2) the students tested Eva with the final set of images; (3) human detection algorithms were tested; (4) and finally, the students presented a report about the experiments. These tasks are detailed below.

3.1 The Dataset

Building a collection in terms of documenting a disaster is not easy. Several questions arise, such as (i) how many images are enough to digitally preserve an event, so the impacts can be comprised in a collection; (ii) what stages of the healing and recovery process should be presented; (iii) what should be considered important (government buildings, schools, roads, etc.); (iv) what should the applications provide to the user; etc. In addition, the images are compelling, since they register the survivor's own emotional reflections of loss.

We concluded that pictures taken before the incident were as important as pictures taken after the incident to show the recovery process of the affected nation after witnessing such disaster. The set of pictures contains both images before and after the earthquake that destroyed a lot of its cities on January 12, 2010.

A list of 111 picture samples were collected online, representing different areas affected by the earthquake. The list initially was divided into two categories, human and non-human pictures. The human category contained injured people pictures and humans living in Haiti before the incident. The non-human category contained all other pictures related to the earthquake such as collapsed buildings, road ruptures, and the buildings before the earthquake.

Fig.2. National Palace before and after.

Fig.3. An injured person being rescued.

Fig. 4. U.N. Secretary in front of the U.N Headquarters.

Title	Creator	Suject	Keywords	Description	Publisher	Contril	Date	Ressource	Forma	Res	Source
HumanLeogar	CareIP from	Leogane, Haiti-	Human, Haiti, Le	People gather	Flickr		1/15/2010	social site	jpg		http://www.flickr.com/pho
HumanLeogar	CareIP from	Leogane, Haiti-	Human, Haiti, Le	People gather	Flickr		1/15/2010	social site	jpg		http://www.flickr.com/pho
HumanLeogar	CareIP from	Leogane, Haiti-	Human, Haiti, Le	People gather	Flickr		1/15/2010	social site	jpg		http://www.flickr.com/pho
HumanLeogar	CareIP from	Leogane, Haiti-	Human, Haiti, Le	People gather	Flickr		1/15/2010	social site	jpg		http://www.flickr.com/pho

Fig.5. Excel sheet showing the Dublin Core fields.

The final picture samples had 71 human images, 39 buildings, 5 having humans and buildings, and 6 without both humans or buildings. Fig. 2 is an example of a non-human category image, presenting the National Palace before and after the earthquake. Fig. 3 is an example of a human being injured. Fig. 4 is an example of an image which combines human and non-human elements. The pictures were collected from different news websites and social network sites. One of the websites that was used a lot when looking for good samples was MSNBC[2]. It had a collection categorized by days and weeks. Some of the pictures in that collection had a short description that helps viewers understand the meaning of them (examples are Fig. 2, Fig. 3, and Fig. 4). Those descriptions facilitated the task of retrieving metadata from images. In some cases, an appropriate annotation was associated, since the students who took part in the study were born in Haiti, and were familiar with the areas devastated by the earthquake.

Dublin Core fields were used for storing basic metadata, such as Title, Creator, Subject, Keywords, Description, Publisher, Contributor, Date, Resource type, Format, Resource identifier, Source, Language, Relation, Coverage, and Rights Management (shown in Fig. 5).

3.2 Experiments with Eva Tool

The second task considered the execution of experiments with the Eva tool. For this, the following applications/technologies were used:

2 http://www.msnbc.msn.com/

- Eva: Evaluation Tool for Comparing Descriptors in Content-Based Image Retrieval (see Section 2.3), installed in a linux server;
- pgAdminIII: designed from writing simple SQL queries to developing complex databases. The graphical interface supports all PostgreSQL features and makes administration easy;
- Browser: for accessing the experiments and visualizing the results.

For computing the similarity among images, four algorithms were selected, BIC, EOAC, LAS, and SASI. The algorithms were written in C, and were submitted to the application using the web interface. The selection of the descriptors relies on their good performance on experiments with other datasets [2]:

1. The Border/Interior Pixel Classification (BIC) algorithm classifies image pixels as *border* or *interior* pixels. Two color histograms are computed considering pixels of both categories. The two histograms are concatenated and stored into a feature vector [8].

2. The Edge Orientation AutoCorrelogram (EOAC) is a shape-based scheme that considers the shape similarity of images. This requires achieving some information related to the approximate shape of the objects in images [9].

3. The Local Activity Spectrum (LAS) descriptor captures textures, i.e., spatial activity in four different directions separately: horizontal, vertical, diagonal, and anti-diagonal. The four activity measures are computed for a pixel (i, j) by considering the values of neighbors in the four directions [8].

4. The Statistical Analysis of Structural Information (SASI) descriptor is based on second order statistics of clique autocorrelation coefficients, which are the autocorrelation coefficients over a set of directional moving windows [10].

Basic constructs of the CBIR systems (such as feature vectors, image distance, experiment details, ranking, precision-recall values, etc.) can be easily introduced when students dealt with Eva, and could be investigated through the web interface, postgreSQL database, and configuration files. The comparison of the different algorithms (using color, shape, and texture) also could be visually explored using a browser, as exemplified in Fig. 6 and Fig. 7. In this example, both images use Cathedral_P_A_P.jpg as query image, and they present different ranking results if we consider color (BIC descriptor in Fig. 6) and texture (SASI descriptor in Fig. 7).

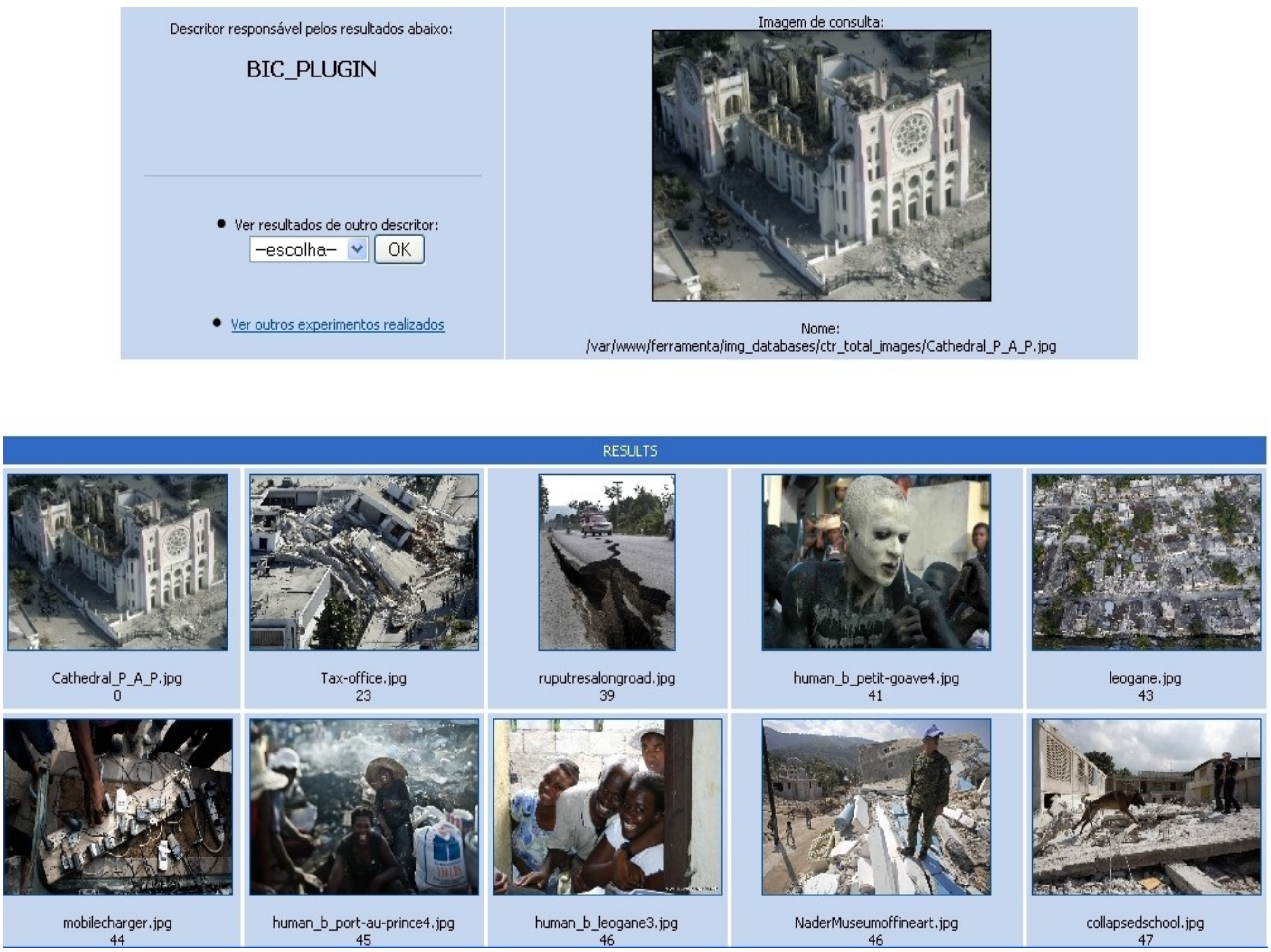


Fig.6. Image ranking created by the BIC descriptor.

Students were encouraged to query images and interact with the retrieved results. Sixteen pictures were chosen as query images, one by one used by searching, leading to a retrieved set, as shown in Fig.8. The green color indicates that the selected image was relevant.

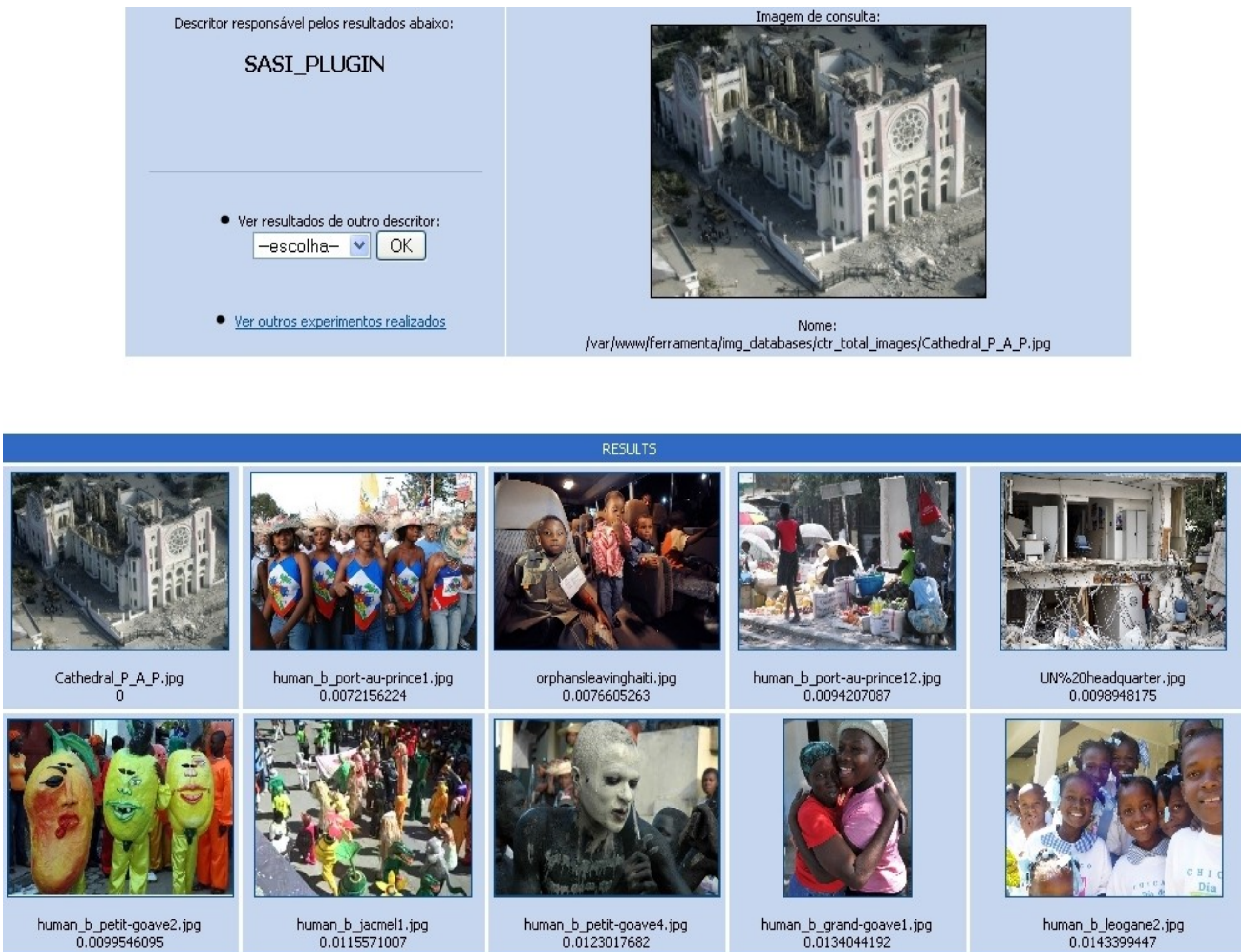


Fig.7. Image ranking created by the SASI descriptor.

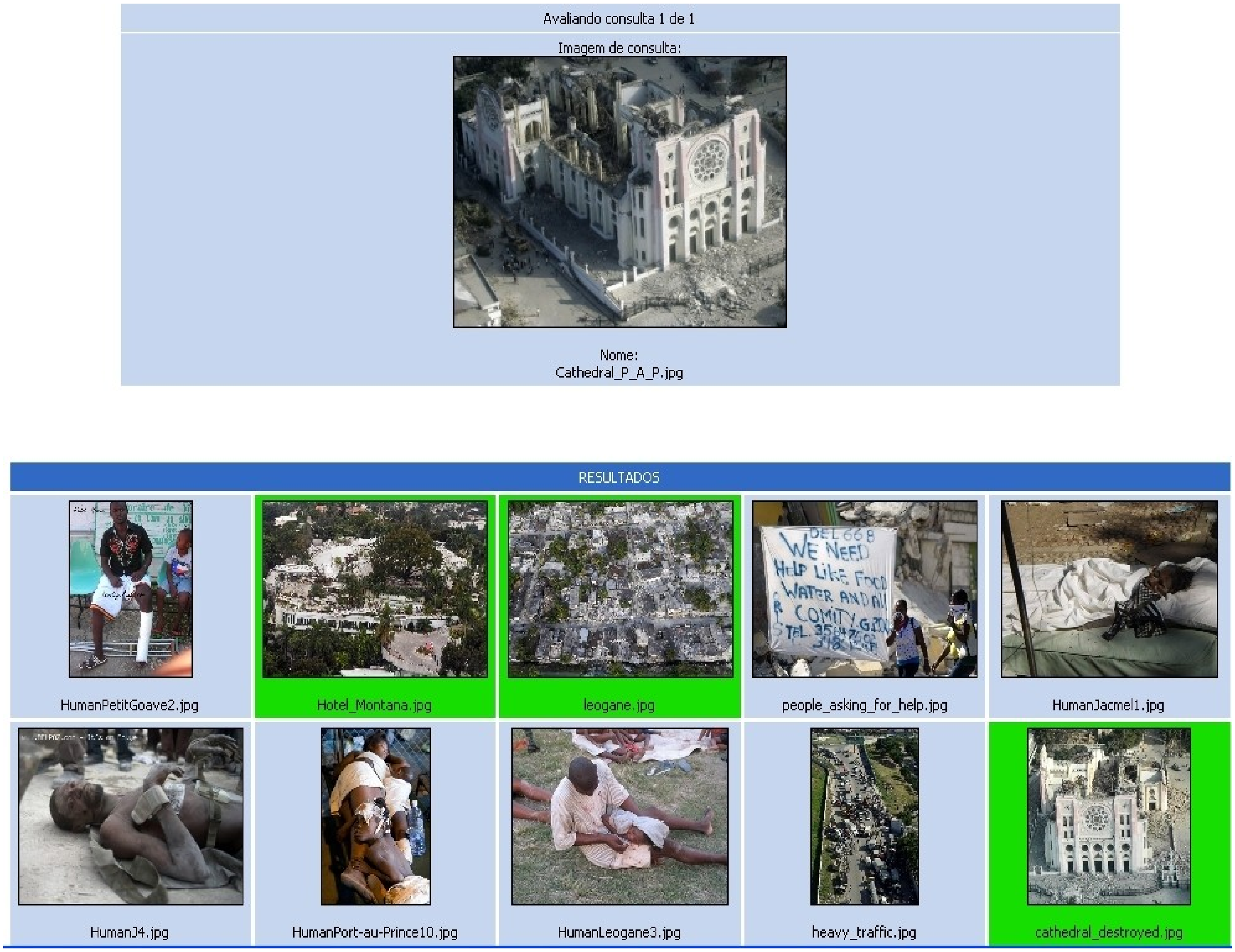
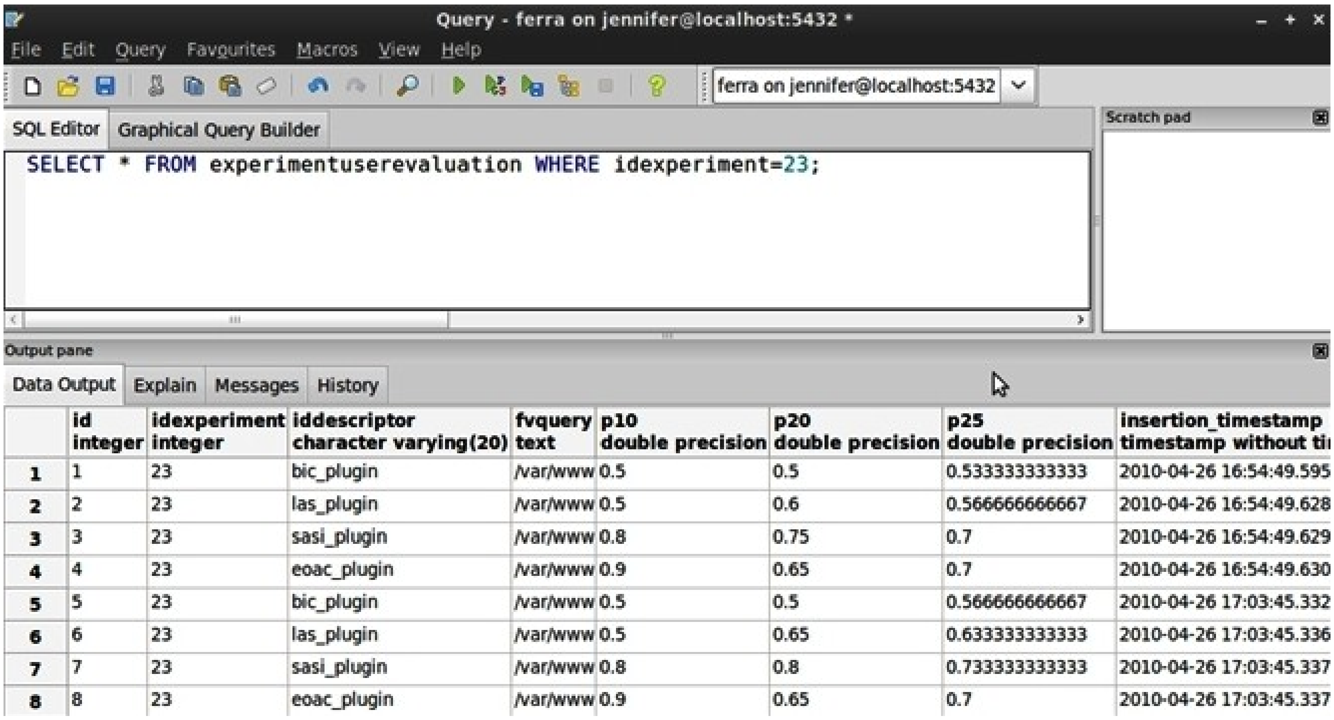


Fig.8. Query image and relevant results (colored in green).

Some data for the image comparison was stored in the postgreSQL database, as shown in Fig.9 and Fig. 10. In Fig. 9, the idexperiment column represents a particular experiment, the fvquery column represents the path where the image is located in the known directory. The p10, p20, and p25 columns are measures that stand for precision values and they indicate the percentage of images marked as similar among the top 10, 20, and 25 results, respectively. These measures are calculated for each descriptor, for each query image and for each user. Therefore, it is possible to calculate the descriptors'



	id integer	idexperiment integer	iddescriptor character varying(20)	fvquery text	p10 double precision	p20 double precision	p25 double precision	insertion_timestamp timestamp without ti
1	1	23	bic_plugin	/var/www	0.5	0.5	0.533333333333	2010-04-26 16:54:49.595
2	2	23	las_plugin	/var/www	0.5	0.6	0.566666666667	2010-04-26 16:54:49.628
3	3	23	sasi_plugin	/var/www	0.8	0.75	0.7	2010-04-26 16:54:49.629
4	4	23	eoac_plugin	/var/www	0.9	0.65	0.7	2010-04-26 16:54:49.630
5	5	23	bic_plugin	/var/www	0.5	0.5	0.566666666667	2010-04-26 17:03:45.332
6	6	23	las_plugin	/var/www	0.5	0.65	0.633333333333	2010-04-26 17:03:45.336
7	7	23	sasi_plugin	/var/www	0.8	0.8	0.733333333333	2010-04-26 17:03:45.337
8	8	23	eoac_plugin	/var/www	0.9	0.65	0.7	2010-04-26 17:03:45.337

Fig.9. Example of data stored for the p10, p20 and p30 precision.

ID DescriptorFv2	**distance double precision**
bic_plugin Cathedral_P_A_P.jpg	0
bic_plugin Tax-office.jpg	23
bic_plugin HumanJ4.jpg	25
bic_plugincathedral_destroyed.jpg	32

Fig.10. Results from the distance table.

effectiveness for each query image independently. The last column shows when exactly the experiment was conducted.

The distances among images also were stored in the database. Fig.10 illustrates the distances among images using the BIC descriptor. The query image considered in this example is Cathedral_P_A_P.jpg. Note that the distance is zero when the query image is compared to itself. The second most similar image was Tax-office.jpg, whose distance to the query image is equal to 23.

3.3 Human Detection Algorithms

Human detection recognition (particularly, face detection) is one of the most difficult problems in the computer vision area. Human detection and recognition also receives a huge attention in the medical field, and in research communities including biometric, pattern recognition and computer vision [13]. Certain parameters have been taken into account for different algorithms, such as facial expressions variations, pose variations, details (such as glasses, hats, etc.), and resolution.

Two algorithms were tested: Kienzle et al. [11] and the OpenCV library (available at http://opencv.willowgarage.com/wiki/). Three images were considered for testing the algorithms: Fig. 2, Fig. 4, and Fig. 11. Fig. 2 is an example of non-human category pictures, and Figures 4 and 11 are examples of human category pictures. Both algorithms presented issues on automatically classifying human images, showing in a practical experiment how tricky it is to use image details for classification.

Note that Fig. 2 is a picture from the non-human category, therefore no faces or human bodies should be detected. Fig. 3 is a picture from the human category, where faces have specific details, such as hats, glasses, and different pose variations. Fig. 11 is another example from the human category, although the resolution is different if compared to Fig. 4.

Fig.11. Example of Human Category Image.

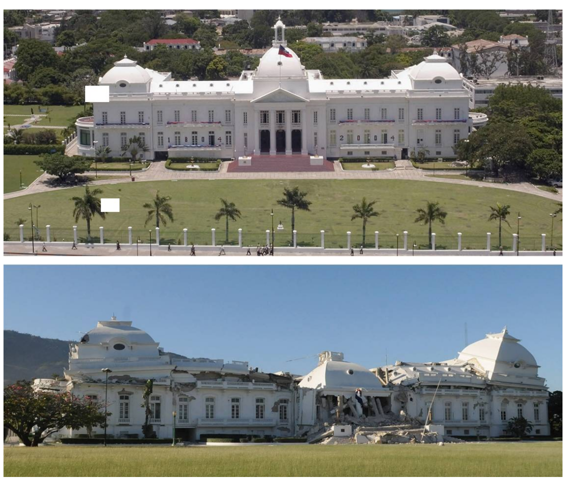


Fig.12. Face detection with Kienzle et al. [11].

Fig.13. Face detection with Kienzle et al. [11].

In Kienzle et al. [11], the algorithm returns the number of faces detected and their positions. Fig. 12 is the figure created after the algorithm execution: an example of a false positive. The image has two buildings, and presented two face detections, at the upper part (represented by two white squares). A possible reason for that relies on the fact that these regions should have high texture details, a very important parameter for face detection algorithms. In Fig. 13 only three faces were detected (represented by three white squares).

Note that besides having several people, the majority of them were wearing glasses and sunglasses, a detail that is not present in the training set of the algorithm.

The OpenCV human detection algorithm was also tested. The algorithm returns the image and the detected human bodies in green color. For image without human bodies (as those present in Fig. 2), the algorithm did not identify any face, as expected. A partial positive human detection is also present in Fig. 4 (the processed image is shown in Fig. 14). However the algorithm failed to detect people in Fig. 2.

Finally, a commercial software for frontal face detection[3] was also tested. This tool requires additional configuration of parameters: the detection threshold, face scale, horizontal and vertical shift, and the detection performance. For images without human bodies (as those present in Fig. 2), the algorithm did not identify any face, as expected. In Fig. 4, just one face was detected (the last person at the right). Possible issues for this algorithm include the combination of testing parameters: the detection can occur or not depending on how they are configured.

Fig.14. People detection with OpenCV Library (see green rectangles).

3 http://www.luxand.com/facesdk/

The human detection algorithms discussed in this report brought to our attention additional details (pose variations, resolutions, face expressions) and categories (partial detection, non-detection, false positives) which can vary depending on the algorithm, and that can be further explored for better image categorization.

3.4 Final Report

The students listed the following problems regarding the conducted experiments:

1. The management of descriptor output files provided Eva could be extended. Only the color descriptor generated text file output. The other three descriptors (one shape, two texture) generated binary files output. To solve this problem during the case study, a Perl script was used to convert the binary files. The generated feature vectors also had different range sizes, and their distances covered a different range of values.
2. Initial parameters (such as the number of pictures to represent a collection, or what should be the image categories, or how to classify images that belong to more than one category) also impacted the evaluation, since they could be responsible for additional time in the analysis.
3. Another problem encountered with the experiment was related to the non-human category pictures since this category included all kinds of scenes other than injured people. Some pictures had buildings and people included (Fig. 4.); others had neither buildings nor people. A solution would be to create sub-categories within the non-human category itself. We would suggest adding: humanitarian aid, roads, rescue, and evacuation team. But this would bring a secondary issue, since primary tags and descriptions would be necessary now for the image categorization, instead of only content-based retrieval.

With the experiments, the students were exposed to the practical use of concepts like descriptors, feature vectors, distance between images, evaluation of experiments and comparisons, user evaluation, image categorization for specific collections, metadata, and false positives. The conducted study helped in the integration of CBIR with CTRnet, highlighting the importance of appropriate metadata, the clustering of images considering different target categories, and the amount of data generated.

4. Conclusions

We present a report about activities related to use of Eva with CTRnet images during an independent study during the CS4624 at Virginia Tech. This experiment was a prototype to explore image categories and the understanding of CBIR concepts. Initial testings with a face detection algorithm presented future issues that we will have with bigger collections. With the experiments, the students were exposed to the practical use of CBIR concepts.

The data gathering should not stop there. Other algorithms for face/human detection also should be investigated, along with larger datasets. Another issue is concerning the performance of experiments with potential users. User evaluation also could be taken into account.

5. Acknowledgments

We thank NSF for supporting this project via grant #0916733, as well as CAPES, FAPESP, and CNPq. The CTRnet project gratefully acknowledges the many online sources that have and continue to document the crises, tragedies, and community recovery efforts around the globe. We store metadata giving details on those sources in our digital library. In the interest of brevity we omit that information with regard to the sample pictures included in this paper but urge those interested to go to original online sources.

References

[1] E. A. Fox et. al. *A Digital Library for Recovery, Research, and Learning from April 16, 2007*, Virginia Tech. Traumatology, v. 14, Issue 1, p. 64-84, March 2008.

[2] O. Penatti, and R. S. Torres. *Eva - An Evaluation Tool for Comparing Descriptors in Content-Based Image Retrieval Tasks,* MIR '10: Proceedings of the international conference on Multimedia information retrieval, p. 413-416, 2010.

[3] R. Datta, D. Joshi, J. Li, J.Z.Wang, James Z. *Image retrieval: Ideas, influences, and trends of the new age,* ACM Comput. Surv., v. 40, n. 2, p. 1-60, 2008.

[4] H. Muller, W. Muller, D. Squire, S. Marchand-Maillet, T. Pun. *Performance evaluation in content-based image retrieval: overview and proposals,* Pattern Recogn. Lett., v. 22, n. 5, p. 593-601, 2001.

[5] J. French, J. Watson, X. Jin, W. Martin. *Integrating Multiple Multi-Channel CBIR Systems*, International Workshop on Multimedia Information Systems (MIS 2003), pp.85-95, Ischia, available at http://www.cs.virginia.edu/~xj3a/research/CBIR/index.htm, Italy, May 26-28, 2003.

[6] M. Careem, C. Silva, R. Silva, L. Raschid, S. Weerawarana. *Demonstration of Sahana: free and open source disaster management,* Proceedings of the 8th annual international conference on Digital government research, p. 266-267, 2007.

[7] K. Banks, E. Hersman. *FrontlineSMS and ushahidi - a demo,* ICTD'09: Proceedings of the 3rd international conference on Information and communication technologies and development, p. 484-484, 2009.

[8] J. A.Santos, O.A.B Penatti, R.S. Torres. *Evaluating the Potential of Texture and Color Descriptors for Remote Sensing Image Retrieval and Classification*. Technical Report IC-07-05, University of Campinas, 2009.

[9] F. Mahmoudi, J. Shanbehzadeh, A. Eftekhari-Moghadam, H. Soltanian-Zadeh. *Image retrieval based on shape similarity by edge orientation autocorrelogram,* Pattern Recognition, v. 36, Issue 8, August 2003.

[10] A. Carkacroglu, Y. F. Yarmun- Vurul. *Sasi: a generic texture descriptor for image retrieval,* Pattern Recognition, v. 36(11), p. 2615-2633, 2003.

[11] W. Kienzle, G. Bakır, M. Franz, B. Schölkopf . *Face detection - efficient and rank deficient,* Advances in Neural Information Processing Systems 17, p. 673-680, 2005.

[12] R. S. Torres, A.X. Falcão. *Content-Based Image Retrieval: Theory and Applications,* Revista de Informática Teórica e Aplicada, v. 13, p. 161-185, 2006.

[13] N. Ismail, M. I. M. Sabri. *Review of Existing Algorithms for Face Detection*,

CIMMACS'09: Proceedings of the 8th WSEAS International Conference on Computational intelligence, man-machine systems and cybernetics, p. 30-39, 2009.

[14] N. P. Kozievitch, T. R. C. Falcão, and R. da S. Torres. *A .Net Implementation of a Content-Based Image Search Component.* In: V Sessão de Demos, 2008, Campinas. XXIII Simpósio Brasileiro de Banco de Dados, 2008.

[15] N. P. Kozievitch, R. Da S. Torres, F. Andrade, U. Murthy, E. Fox, E. Hallerman. *A Teaching Tool for Parasitology: Enhancing Learning with Annotation and Image Retrieval*. ECDL '2010: the 14th European conference on Research and Advanced Technology for Digital Libraries, Glasgow, UK, 2010.