Image Retrieval Based on Text and Visual Content Using Neural Networks

EVIC 2009

Diego A. Castro - Leticia M. Seijas

{dcastro, lseijas}@dc.uba.ar

Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

Introduction In the last few years there has been a dramatic increase in the visual information available and retrieving images from big databases has become a challenging task. Typically, images are described by their textual content (TBIR) or by their visual features (CBIR), but recently the hybrid approach was introduced. It combines both characteristics to improve the benefits of using text and visual content separately. Considering that CBIR is still far from being as well-matured as TBIR, in this work ***we concentrate on CBIR and propose a new SOM model (ParBSOM), which can be used for indexing images efficiently. In addition, we study how these techniques can be applied to the hybrid approach and provide computational results to assess their performance***.

Image Descriptor

Color is one of the most intuitive features of an image.

In CBIR, ***color histogram*** [1] is one of the most widely used features in the area and it is constructed by counting the number of pixels of each color.

It can work with different color spaces such as RGB or HSV and with different distance metrics.

In many works, ***HSV*** with ***L1 distance*** has shown improved results [2,3].

In order to eliminate irrelevant images from the results list, we propose a ***scoring function*** that allows us to define a threshold (between 0 and 1) and filter those images below it.

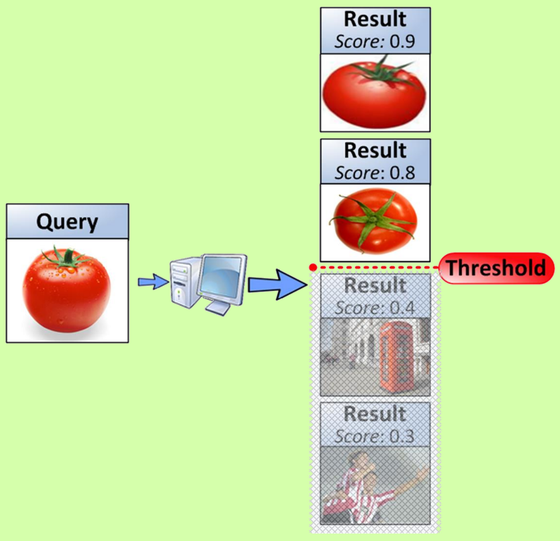

Figure 1: Threshold to eliminate irrelevant images during retrieval

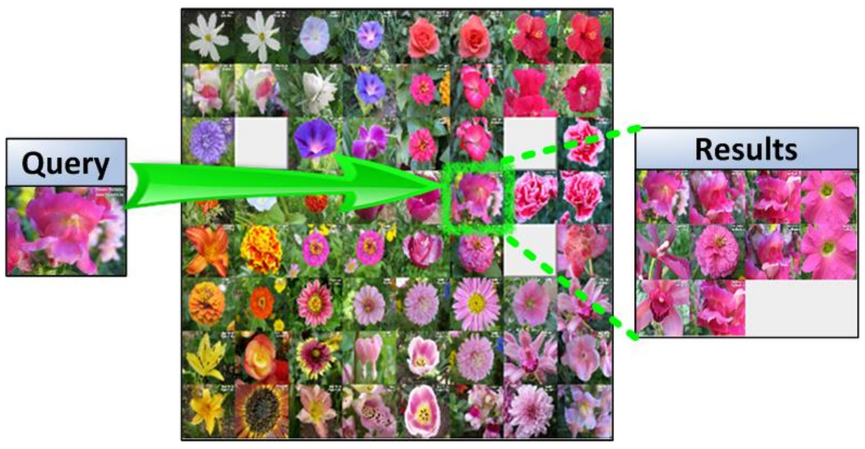

Figure 2: Retrieving images from a trained SOM

Image Index

One of the main problems faced in CBIR is that image descriptors are usually high-dimensional.

Current techniques such as R-Trees [4] or KD-Trees [4] are not scalable for dimensions higher than 20.

SOM [5] **can act as an image classifier**, mapping images to neurons in the network and generating maps where similar images are close in the network (this characteristic is used during retrieval).

SOM allows us to work with high-dimensional descriptors.

Working with big networks can reduce the performance of the classical SOM, so different models such as BSOM [6] and ParSOM [7] have been developed.

BSOM is an alternative that modifies the training algorithm, reducing the time required to train the net.

ParSOM consists of dividing the network into many sections which are maintained by different processing nodes (training and retrieval can be performed in parallel) .

We propose a new model known as *ParBSOM* that combines both characteristics leading to a considerable improvement in training and retrieval times.

Hybrid Approach

In order to overcome TBIR and CBIR problems, recently the hybrid approach has been introduced.

CBIR and TBIR produce their own results and then both lists are merged (late fusion).

One of the merging strategies is known as **refinement** [8], which reorders TBIR results using the results of CBIR.

This strategy gives more importance to textual results as nowadays TBIR is a much more advanced area than CBIR.

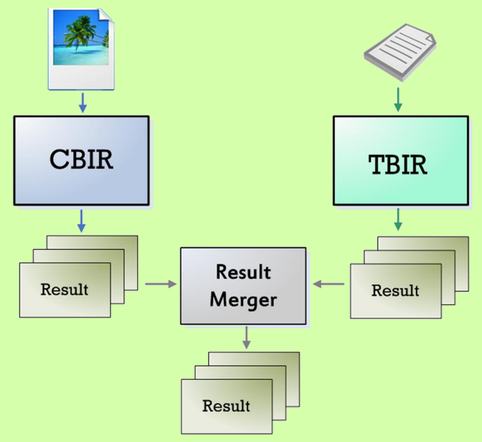

Figure 4: Typical organization of a hybrid system

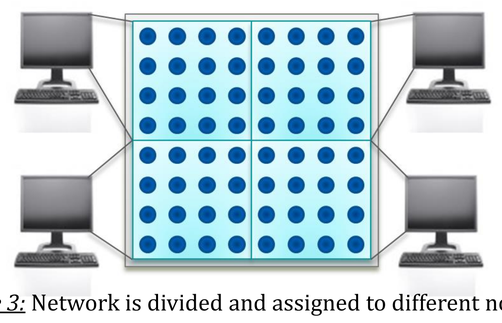

Figure 3: Network is divided and assigned to different nodes

Experimental Results

First, we compared **training times** for different SOM models: the traditional SOM, BSOM, ParSOM, and our proposed model ParBSOM.

Data sets of different size and dimension and two processing nodes –for parallel versions- were used in the experiments.

Data [size x dimension]	Map Units	Improvement BSOM vs. SOM	Improvement ParSOM vs. SOM	Improvement ParBSOM vs. BSOM	Improvement ParBSOM vs. ParSOM
[5000 x 250]	500	56%	33%	**34%**	**57%**
	1,000	52%	33%	**35%**	**54%**
[5000 x 500]	500	55%	31%	**34%**	**58%**
	1,000	50%	31%	**40%**	**57%**
[10,000 x 250]	500	60%	32%	**36%**	**63%**
	1,000	58%	32%	**37%**	**61%**
[10,000 x 500]	500	59%	31%	**38%**	**63%**
	1,000	56%	32%	**40%**	**61%**

Table 1: Training times for different models (10 epochs of training)

- BSOM vs. SOM: improvement above 50%
- ParSOM vs. SOM: improvement close to 30%
- **ParBSOM vs. BSOM**: improvement by about **40%**
- **ParBSOM vs. ParSOM**: improvement close to **60%**

Then, we compared the **quality** of the generated maps for ParSOM, ParBSOM, and the Brute Force algorithm (a linear search through the database).

We used image databases which are used in many works of the area [9,10,11].

Image Databases	Quality Loss ParBSOM vs. Brute Force	Quality Loss ParSOM vs. Brute Force	Quality Loss ParSOM vs. ParBSOM
ZuBuD	0.46%	1.12%	0.66%
UCID	8.1%	10.89%	3.04%
UK Bench	9.07%	9.94%	0.97%

Table 2: Quality loss in terms of F-Measure

- **ParSOM vs. ParBSOM**: similar quality
- **Brute Force vs. ParBSOM**: less than **10%** of quality loss in all DBs

We also calculated **retrieval times** for ParBSOM and the Brute Force method.

Image Databases	Time Brute Force	Time ParBSOM	Improvement ParBSOM vs. Brute Force
ZuBuD	3.43 ms.	0.27 ms.	**92%**
UCID	4.58 ms.	0.32 ms.	**93%**
UK Bench	40.63 ms.	1.68 ms.	**96%**

Table 3: Time required to retrieve an image from the database

- **ParBSOM vs. Brute Force**: more than **90%** of improvement
- As database size increases, improvement is higher

Finally, we applied the studied methods (HSV color histograms and ParBSOM) to a hybrid system which uses the refinement strategy.

Metric	TBIR	TBIR + CBIR (Hybrid)	Improvement
MAP	14,94	16,59	**9,95%**
Precision	5,35	5,35	0%
Recall	49,27	49,27	0%
F-Measure	8,26	8,26	0%
Prec(10)	22,33	27,83	**19,76%**
Prec(20)	18,33	22,08	**16,98%**

Table 4: Different retrieval methods for ImageCLEFphoto 2007 [12]

- Precision, Recall, and F-Measure show no changes (they are not sensitive to image rankings and refinement alters TBIR rankings without modifying the results set)
- **MAP, Prec(10), and Prec(20)** results show an improvement between **10%** and **20%**

Conclusions

✓ Several techniques applied to the image retrieval area have been investigated.

✓ We have studied *color histograms*, comparing their performance in the *RGB* and *HSV* space.

✓ A *scoring function* for color histograms has been proposed in order to eliminate irrelevant images from the results list.

✓ We have investigated how SOM can be used as an index in CBIR.

✓ We have introduced a new SOM model (**ParBSOM**) that improves BSOM's training time by about 40% and also ParSOM's training time by about 60% and proposed to use it in CBIR.

✓ We have studied hybrid techniques and observed that the refinement strategy can actually improve textual results by using visual features.

References

[1] M. J. Swain and D. H. Ballard, "Color indexing," International Journal of Computer Vision, vol. 7, pp. 11–32, 1991.
[2] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in ICIP '95: Proceedings of the 1995 International Conference on Image Processing, vol. 3, p. 3528, IEEE Computer Society, 1995.
[3] O. Jonsgård, "Improvements on colour histogram-based CBIR," Master's thesis, Gjøvik University College, 2005.
[4] C. Böhm, S. Berchtold, and D. A. Keim, "Searching in high dimensional spaces: Index structures for improving the performance of multimedia databases," ACM Comput. Surv., vol. 33, pp. 322–373, 2001.
[5] T. Kohonen, "Self-organized formation of topologically correct feature maps," Biological Cybernetics, vol. 43, pp. 59–69, 1982.
[6] T. Kohonen, Self-Organizing Maps. Springer-Verlag, 3 ed., 2001.
[7] P. Tomsich, A. Rauber, and D. Merkl, "parSOM: Using parallelism to overcome memory latency in self-organizing neural networks," in High Performance Computing and Networking, pp. 61–5, Society Press, 2000.
[8] M. Grubinger, Analysis and Evaluation of Visual Information Systems Performance. PhD thesis, School of Computer Science and Mathematics, Faculty of Health, Engineering and Science, Victoria University, 2007.
[9] H. Shao, T. Svoboda, and L. van Gool, "ZuBuD — Zurich BuildingsDatabase for Image Based Recognition," tech. rep., Swiss Federal Institute of Technology, Switzerland, 2003.
[10] G. Schaefer and M. Stich, "UCID - An Uncompressed Colour Image Database," in Storage and Retrieval Methods and Applications for Multimedia 2004, vol. 5307 of Proceedings of SPIE, pp. 472–480, 2004.
[11] D. Nistér and H. Stewénius, "Scalable Recognition with a Vocabulary Tree," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 2161–2168, 2006.
[12] M. Grubinger, P. Clough, A. Hanbury, and H. Müller, "Overview of the ImageCLEFphoto 2007 Photographic Retrieval Task," Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Revised Selected Papers, pp. 433–444, 2008.