# CORTINA: Searching a 10 Million + Images Database

Elisa Drelie Gelasca
Pratim Ghosh
Emily Moxley

Joriz De Guzman
JieJun Xu
Zhiqiang Bi

Steffen Gauglitz
Amir M. Rahimi
B. S. Manjunath

Vision Research Lab., Electrical and Computer Engineering Department, mailto: drelie@ece.ucsb.edu
University of Santa Barbara, California 93106-9560

## ABSTRACT

We present an image search and retrieval system, Cortina, that indexes over 10 Million images using image content, text and annotations. This large collection of image data, gathered from the World Wide Web (WWW), poses significant challenges to automated image analysis, pattern recognition and database indexing. At the systems level, the components of Cortina include building image collections using a Web crawler, collecting category information and keywords, and processing images to compute content descriptors. Functionalities of Cortina include duplicate image detection, category and image content based search, face detection and relevance feedback. A MySql database is used for storing textual annotations and keywords, whereas the image features are stored in flat file structures. This combination appears to be effective and scalable for large collection of image/video data and is easily parallelizable.

## 1. INTRODUCTION

In the past decade, many general-purpose image retrieval systems have been developed. Examples include SIMPLIcity and ALIPR [10], Blobworld [5], VisualSEEK and WebSEEK [16] and the PicHunter [6]. The primary objectives of these systems include organizing the multimedia semantic content [2, 3]; image retrieval by similarity, duplicate detection [9] and enhancing performance with relevance feedback [7, 4]. Most of these systems have limited image content and diversity. Cortina is perhaps the first system (in published literature) to break the 1 Million image barrier and the current version scales this by an order of magnitude– to over 10 Million images– while adding new functionalities such as annotation and segmentation.

Cortina makes large scale, similarity and category based image retrieval on the web possible. Similarity search is performed in a combined feature space that includes color and texture. Powerful classifiers are being developed to automatically classify image content using these descriptors. Cortina provides a duplicate detection method that is fast and effective. The user can refine the search based on relevance feedback. The enhanced version of the system will be made available on the World Wide Web by July 2007. Compared to the previous version [14], the database size has grown by almost an order of magnitude (3 million to 10 million images). In addition, the new version of Cortina has an easy to use interface and offers new functionalities such as duplicate detection, face detection, category based search and image annotation and segmentation tools.


*VLDB '07,* September 23-28, 2007, Vienna, Austria.


## 2. SYSTEM OVERVIEW

This section gives an overview of the system shown in Fig. 1:

**Image Acquisition**: 11 million images and associated text are collected from the web using a web crawler and the DMOZ.org category information.

**Feature Extraction**: For each image, we compute five types of feature descriptors, including three MPEG-7 descriptors [12], the Homogenous Texture Descriptor (HTD), the Edge Histogram Descriptor (EHD), and the Dominant color Descriptor(DCD). A rotation and scale invariant descriptor is used for duplicate image detection (Compact Fourier Mellin Transform, CFMT) [8], and the SIFT descriptor is used for scene classification [3, 11].

**Clustering and Indexing**: For the 12 dimensional feature vector for duplicate detection (CFMT) using MySql spatial indexing took about 3 minutes (with no change to the original code). A sequential search in this 12-d space takes approximately 3 seconds for the 10 nearest neighbors on the average (see Table 1). A kd-tree implementation that we have built (for the 10 nearest neighbors) takes about 0.03 seconds. In addition, we are currently exploring different clustering methods for approximate nearest neighbor retrievals. Table 1 shows a results for different number of clus-

**Table 1: Comparison of speed and accuracy for duplicate detection in a 12-D space using sequential search and K-means clustering for the top 5 clusters and first 10 nearest neighbors. The search is over the entire 11 million image database.**

| # clusters | none | 32 | 64 |
|---|---|---|---|
| # points compared | 11033927 | 1085509 | 583381 |
| searching time (sec) | 3.014 | 2.841 | 1.826 |
| result accuracy | 1.00 | 0.95 | 0.80 |

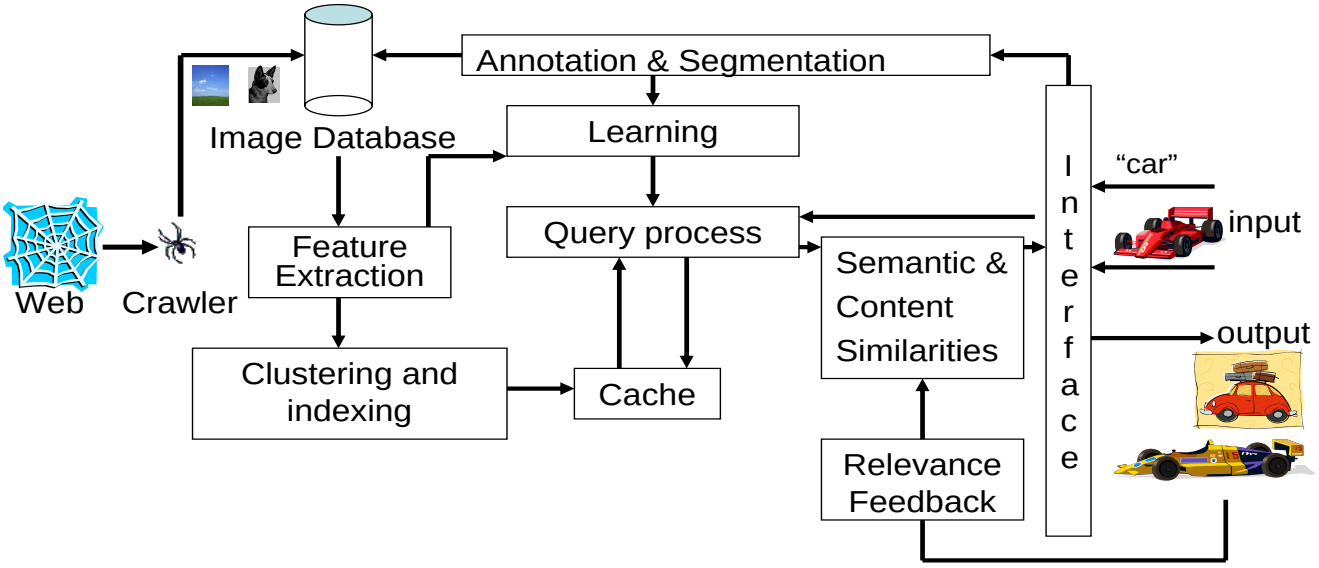


**Figure 1: A schematic view of Cortina.**

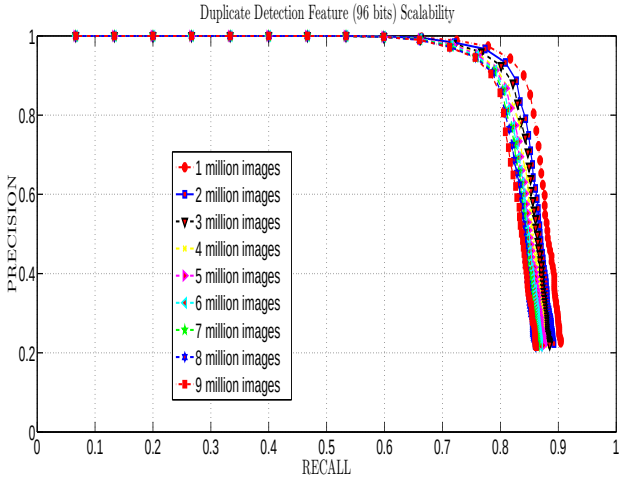


**Figure 2: Precision Recall values in Cortina dataset for near duplicate detection using AFMT. The results are averaged on 100 queries with 15 duplicates each.**

ters used for this approximate search and the corresponding accuracy. The search was done using a Intel(R) Xeon(R) with CPU 5140, 2.33GHz and 8G RAM. The accuracy is computed by assuming the sequential search (feature space without clustering) to be the ground truth.

**Querying**: to start a search the user has 4 options as shown in the Screen Shot in Fig. 3. *Keyword query*, to do a keyword or text search within the existing images, *upload* an image or insert the URL, *browse* images in the database randomly, or *cluster* to visualize images in its semantic clusters. We also adopted Viola Jones approach [17] to annotate up to 20 faces in each image and cluster them in the 'face' category. Our test results show that out of 400 randomly selected images we have 73% accuracy for the frontal faces and 81% accuracy for profile faces.

**Annotation and segmentation**: Manual annotations were collected through the Cortina web-site tools and used for learning the semantic categories. We trained a suite of classifiers, one for each scene (such as mountains, cityscape, etc..) and individual objects in the scene (cat, dog, etc..) using a large-scale concept ontology for multimedia [13]. We also integrated two segmentation tools, a Matlab based tool available online [1] and a web based labeling tool [15], to mark the regions of interests in each image. These segmentation results are stored in the database and can be displayed on demand.

**Visualization**: Two subset of results are displayed after a query image has been selected (see Fig.4). *Topic-related*: the images that are visualized in the first row are related to similar meta data associated to the image such as keywords or annotation. *Content-related*: the near duplicate images present in the database followed by the similar images according to the visual features, are displayed.

**Learning**: We tested different learning methods with global and local features for *topic* based retrieval. At the moment topic results are based on manual annotations but we plan to add the automatic annotation of images according trained categories in the demo. The topic results obtained are displayed in the first row of the Screen Shot of Fig.4.

**Relevance Feedback**: After the user enters a keyword-query or an image-query, one or more steps of relevance feedback follow. If the users choose a semantic approach to perform search, all images related to and clustered by that keyword will be ranked by their relevance to the query and returned as a result. Within the set of results, the user is able to mark an image as either relevant or irrelevant to the keyword provided. The idea is to model the conceptual/semantic relevance of an image and learn this over time to improve future semantic retrievals.

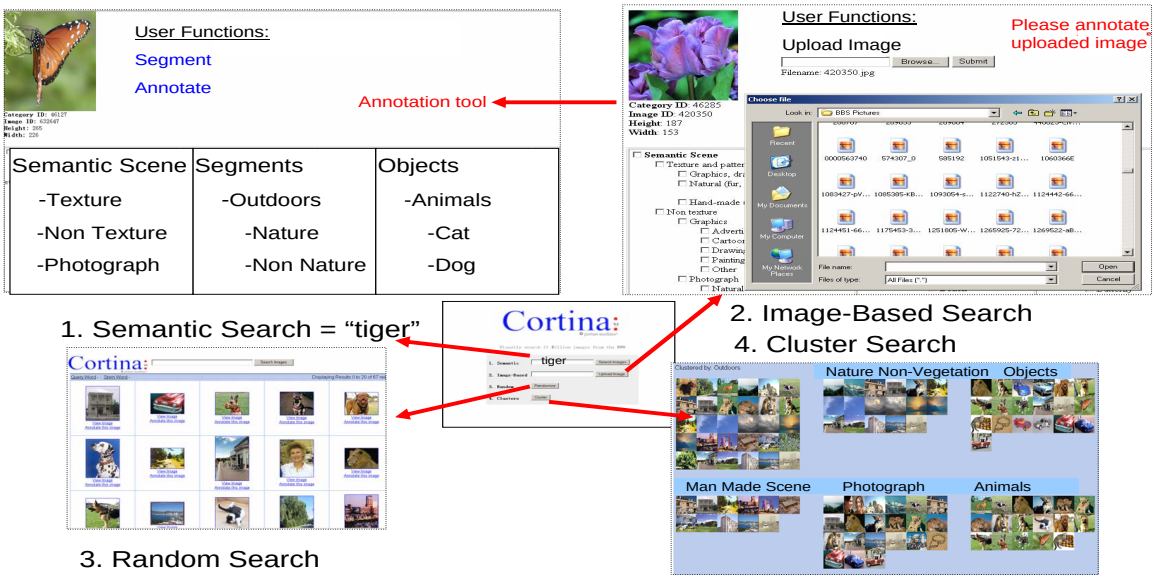


**Figure 3: A screen shot of the query with the 4 possibilities to start a search.**

## 3. CORTINA DATABASE

A web crawler stores images from websites traversed from the categorical structure of DMOZ. Textual information relevant to each image is also stored. Such textual metadata is extracted from the filename, ALT text and collateral keywords surrounding the image. Currently, we have approximately 460,000 categories and approximately 900,000 keywords in the database. Feature descriptors for each image are computed for newly acquired images and are stored in flat file structures. A batch process is run periodically to crawl the web for new images and compute their feature descriptors listed in Table 2. We have used a MySql database to store the textual information and centroids for the feature clusters.

**Duplicate detection** For duplicate detection the feature descriptor is 12 dimensional vector quantized and stored in single precision floating point (96 bits in total). The compact signatures are stored into a binary file which consumes only around 0.1 GB for more than 11 million images. The proposed signature (CFMT) involves Fourier-Mellin Transform, conventional PCA and Lloyd-Max non-uniform scalar quantization [8]. The high precision recall values are depicted in Fig. 2. **Similarity search** For the similarity search, retrieval in the visual feature spaces consist of K-Nearest Neighbor search. The $L_2$ norm is used to measure similarity in the HTD and EHD feature space and a quadratic distance measure [12] is used for DCD. The results of retrieval for each feature are combined for a joint search: the distances in the feature space of the three descriptors are summed in a linear way. To improve the retrieval results based on semantic associations between text and visual features, association rule mining is applied as in [14].

## 4. STRUCTURE OF THE DEMONSTRATION

To summarize, the main features of Cortina are:

- a system for large scale, web image categorization and retrieval is implemented;

**Table 2: List of image features used by‘ Cortina**

| Feature | dimensionality | #bytes | similarity |
|---|---|---|---|
| HTD | 62 | 62 | L2 |
| EHD | 80 | 80 | L1 |
| DCD | 32 | 32 | quadratic |
| CMFT | 12 | 12 | L2 |
| SIFT | 128 | varies | L2 |

- with over 11 Million of images, Cortina has one of the largest image collections that we are aware of for content based search and retrieval;
- several low level descriptors, both global and local, are implemented and tested with different classifiers;
- the system offers the possibility of collecting: manual annotation on a chosen ontology, segmentation and labels.
- it facilitates ongoing research on data mining, machine learning, pattern recognition and classification, and high dimensional indexing on very large image database. Visual descriptors from Cortina have been used by database researchers.

At the conference time, we plan to present an image retrieval system available on the WWW in the demo. Particulary in the demo we shall focus on the following aspects.

1) We shall demonstrate the relevance feedback procedure online. 2) We shall demonstrate how the images from the database can be easily downloaded according to a preselected category and segmented or labeled. 3) We shall show the effectiveness of results for content based query for both near duplicate detection and similarity search by using the different modalities of querying the database. 4) We shall give the possibility to the user to effectuate searches in the random and predefined clusters of images in Cortina.

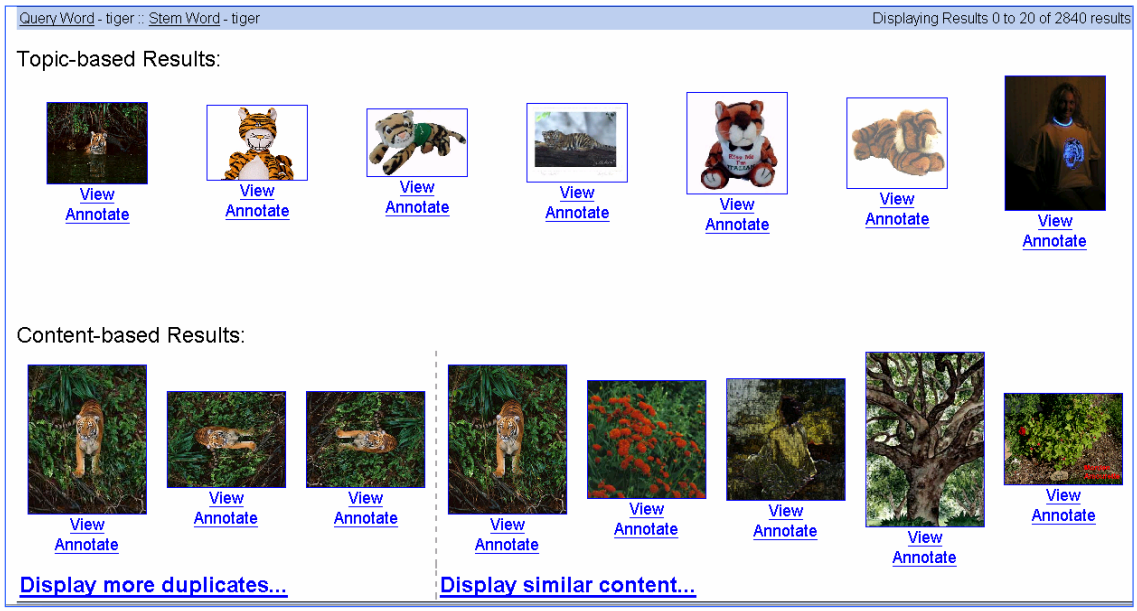


**Figure 4: A screen shot of the visualization with the 2 rows of first retrieved results according to topic and content.**

We expect to have all of the above functionalities fully integrated into Cortina by July 2007. Till that time a limited version of Cortina is available at *http://cortina.ece.ucsb.edu*.

## 5. ACKNOWLEDGMENTS

This project is supported by grants from NSF ITR #0331697.

## 6. REFERENCES

[1] http://vision.ece.ucsb.edu/download.html.

[2] K. Barnard and D. Forsyth. Learning the semantics of words and pictures. In *International Conference on Computer Vision*, volume 2, pages 408–415, 2001.

[3] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via pLSA. In *Proceedings of the European Conference on Computer Vision*, 2006.

[4] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[5] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*. Springer, 1999.

[6] I. J. Cox, M. L. Miller, T. P. Minka, T. Papathomas, and P. N. Yianilos. The bayesian image retrieval system, pichunter: Theory, implementation and psychophysical experiments. *IEEE Transactions on Image Processing*, 2000.

[7] L. Geng and H. J. Hamilton. Interestingness measures for data mining: A survey. *ACM Comput. Surv.*, 38(3):9, 2006.

[8] P. Ghosh, B. Manjunath, and K. Ramakrishnan. A compact image signature for rts-invariant image retrieval. In *IEE International Conference on Visual Information Engineering (VIE 2006)*, Sep 2006.

[9] Y. Ke, R. Sukthankar, and L. Huston. An efficient parts-based near-duplicate and sub-image retrieval system. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 869–876, New York, NY, USA, 2004. ACM Press.

[10] J. Li and J. Z. Wang. Real-time computerized annotation of pictures. In *Proceedings of the ACM Multimedia Conference, Santa Barbara, CA*, October 2006.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.

[12] B. S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG7: Multimedia Content Description Language*. 2002.

[13] M. Naphade, J. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis. Large-scale concept ontology for multimedia. 13(3):86 –91, July-Sept 2006.

[14] T. Quack, U. Monich, L. Thiele, and B. Manjunath. Cortina: A system for large-scale, content-based web image retrieval. In *ACM Multimedia 2004*, Oct 2004.

[15] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: a database and web-based tool for image annotation, MIT AI Lab memo AIM-2005-025, September 2005.

[16] J. R. Smith and S.-F. Chang. Visualseek: a fully automated content-based image query system. In *ACM Multimedia, Boston, MA*, November 1996.

[17] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001*, volume 1.