HISA: A Query System Bridging The Semantic Gap For Large Image Databases

Gang Chen Xiaoyan Li Lidan Shou Jinxiang Dong Chun Chen
Department of Computer Science
Zhejiang University, HangZhou
P.R.China 310027
cg@cs.zju.edu.cn, kricel_lee@yahoo.com.cn, {should,djx,chenc}@cs.zju.edu.cn

ABSTRACT

We propose a novel system called HISA for organizing very large image databases. HISA implements the first known data structure to capture both the ontological knowledge and visual features for effective and efficient retrieval of images by either keywords, image examples, or both. HISA employs automatic image annotation technique, ontology analysis and statistical analysis of domain knowledge to pre-compute the data structure. Using these techniques, HISA is able to bridge the gap between the image semantics and the visual features, therefore providing more user-friendly and high-performance queries. We demonstrate the novel data structure employed by HISA, the query algorithms, and the pre-computation process.

1. INTRODUCTION

The explosive growth in the amount and complexity of image data has created an emergent need for efficient and accurate search and retrieval from a large image database or a collection of image databases. A variety of Content-Based Image Retrieval (CBIR) techniques have been proposed to address these issues. For example, a collection of research prototypes and commercial systems [8] have exploited the *visual features* of images, such as colors, textures, and shapes to represent and index image contents. However, it is widely noted that there is a "semantic gap" between the *visual features* and the *semantic meanings* of images, and it has been a major problem for most CBIR approaches. In the literature we see many efforts in the area of *Automatic Image Annotation* (AIA) [9, 2, 5, 1]. The AIA methods usually employ segmentation techniques to generate keyword-based annotations for the images being indexed to facilitate semantic searching. However, there are a few problems which the current AIA methods may not adequately solve: First, the segmentation techniques deployed on the images are often not robust enough to produce meaningful semantics. Second, clustering of images based on the keyword output may include noises, and is usually error-prone. Third, the AIA algorithms do not address the indexing problem.

In this demonstration, we present a novel system, namely HISA, which brings together the visual features of the images and their semantic meanings. Searching for images in this system can take both the image semantics and the visual features into consideration, therefore solving the problems described above. HISA employs a novel data structure which captures both the high-level ontological knowledge and the low-level visual features of image database. We present a two-phase query algorithm based on the HISA structure. We also use multi-semantic mining for individual images based on probabilistic analysis to improve the query performance. We have implemented a demonstration system incorporating the proposed techniques on a Pentium PC platform running Windows NT.

HISA is distinguished from other image retrieval systems in following ways:

1) Two-phase query The first phase of a query uses the high-level ontology structure, which is stored in a tree structure, to search for relevant nodes which contain generic image semantics. The relevant nodes can be located efficiently. The second phase searches for images using data which we call ASDs (Atomic Semantic Domain) referenced by the leaf nodes of ontology tree. The second-phase search is based on similarity comparison of the pre-computed dominant visual features of the indexed images. For large image datasets, this two-phase query technique achieves high retrieval accuracy without compromising the query speed. The reason for comparing visual features in the second phase is that we observe visual comparison is effective only when the semantics of the images being compared are well correlated. We note that HISA implements the first known data structure to capture both keyword semantics and visual features in a hierarchy and to answer a query.

2) Multi-semantic mining for individual images An image might be indexed by multiple leaf nodes based on the result of the probabilistic analysis of the keyword semantics. Compared to conventional image classification algorithms such as [1, 12, 3], HISA allows an image to be associated with multiple "classes" rather than a single one.

3) Post-annotation processing HISA is a system which uses the *output* generated from an AIA algorithm. The annotation algorithm used by HISA can be easily re-

VLDB '06, September 12-15, 2006, Seoul, Korea.

placed by another. Moreover, the structure can be easily extended to incorporate other annotation methods, such as manual annotation and personalized annotation techniques.

2. THE HISA-STRUCTURE

The HISA-structure combines generic ontological knowledge and more domain-specific semantics, to facilitate query for images efficiently. It consists of an *ontology tree*, a *keyword–node map* structure, and a set of *ASD* structures.

2.1 The ontology tree structure

The ontological knowledge is captured in a tree, where each node represents a category of the images that it indexes. Each *internal node* n of the tree contains the following information:

$\overrightarrow{K}$ A keyword vector which contains keywords implying the semantics of the category that the node represents.

$\overrightarrow{W}$ A weight vector where the ith element corresponds to the weight of the ith keyword in $\overrightarrow{K}$.

$\{L_i\}$ A number of links pointing to its child nodes. Each link denotes a hyponym relation to its child node.

l The level in the tree (the root node has level 0).

The *leaf nodes* of the tree contain the same fields excluding the children information. Additionally, each leaf node contains a pointer to an Atomic Semantic Domain (ASD), which is implemented as a VA-File containing the visual feature information.

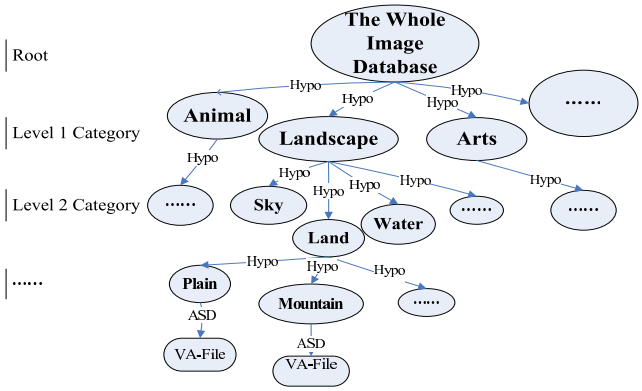

Figure 1: A simple example of the HISA-structure

Figure 1 shows a very simple example of the HISA structure with four levels. Nodes in the high-levels are marked by generic keywords. However, nodes in the lower levels of the ontology tree are much more domain-specific. Therefore, the respective keyword vectors are more likely to include specific keywords as hyponyms of their ancestors.

2.2 The keyword-node map

HISA maintains a map from each keyword to a set of nodes of the ontology. We also store the quantitative representations of relations among the keywords according to lexical knowledge(derived from WordNet [4]), and generate a dynamic keyword hierarchy. We use three binary relations: *synonym*, *hypernym* and *hyponym* in HISA. The synonymic keywords are stored together in this map structure. For each keyword K_i, we record the following information:

$\{Sign_i\}$ An encoded representation of the lexical knowledge of K_i, which can be used to determine if K_i has a binary relation with other keywords. That is, we can determine if K_i and K_j have one of the three relations by calculating a simple function $lex\_rel(Sign_i, Sign_j)$, which may return four possible values: NONE, SYNONYM, HYPONYM, and HYPERNYM.

$\{P_{hypo_i}\}$ The hyponym probability of K_i with respect to its hypernym, if any. In our keyword-node map, each keyword has at most one hypernym.

$\{N_i\}$ The ID set of nodes whose keyword vectors contain K_i.

If K_i has a binary relation with K_h, and K_h has the same kind of binary relation with K_j, K_i and K_j are said to have a *cascade relation* of length 2. Cascade relations can have any length l, where $l \geq 2$.

2.3 The ASD structure

We adopt the *VA-File* [11] as the storage structure for fast querying the candidate images in each ASD. We construct an adaptive VA-File for images in each ASD using the data distribution of their dominant features. We resort to the visual feature vectors of the candidate images and the annotation words for further pruning and ranking using a similarity measure which is based on the Euler distance.

3. THE QUERY ALGORITHM

Given a query keyword set, we firstly refer to the keyword-node map for query preprocessing. For example, if K_i has a hyponym or cascade hyponym relation with K_j, we keep K_i and remove K_j from the query keyword set. The query preprocessing produces a set of the most hyponymic keywords of the original query. Assume the query preprocessing output to be $\{K_1, \cdots, K_m\}$, we perform the procedure listed in Figure 2 to locate the overlapping leaf node set N_{leaf} that are implied by all the keywords in $\{K_1, \cdots, K_m\}$.

ALGORITHM **OverlappedLeaves**
Input: N_{leaf} includes all the leaves in HISA
BEGIN FOR1(each node n in N_{leaf}) Do{
  FOR2(each $K_i, i = 1, \cdots, m$) Do{
    IF(the keyword vector $\overrightarrow{K}$ of n contains K_i or a hyponym, or a cascade hyponym of K_i)
      Continue FOR2;
    ELSE
      Delete node n from N_{leaf}, continue FOR1;
  }
}
Output the remaining nodes in N_{leaf};
END

Figure 2: Computing the leaf node set implied by a keyword set

Let $P(n/K_i)$ denote the predictive probability of node n with keyword K_i. Assume that N_{leaf_i} denotes one leaf node

contained in set N_{leaf}, we compute the predictive probabilities $P(N_{leaf_i}/(K_1, \cdots, K_m))$ for keyword set $\{K_1, \cdots, K_m\}$ as

$$P(N_{leaf_i}/(K_1, \cdots, K_m)) = \sum_{i=1}^{m} P(N_{leaf_i}/K_i), \qquad (1)$$

where $P(N_{leaf_i}/K_i)$ is given by

- If K_i exists in the keyword vector of N_{leaf_i}: $P(N_{leaf_i}/K_i)$ is equal to its corresponding normalized weight W_i.
- Otherwise, there must exist one of its hyponyms or cascaded hyponyms, $K_{i_{hypo}}$, in the keyword set of N_{leaf_i}. Therefore, $P(N_{leaf_i}/K_i)$ is given by $W_{i_{hypo}}$ multiplied by the hyponym probability (or the cascaded hyponym probabilities as defined in subsection 2.2) between $K_{i_{hypo}}$ and K_i.

The possible results of the above algorithm are:

- If $\{N_{leaf}\}$ is NULL, empty;
- Otherwise, choose the top-k leaves from $\{N_{leaf}\}$ based on the ranked probability $P(N_{leaf_i}/(K_1, \cdots, K_m))$, where k is normally set to 1.

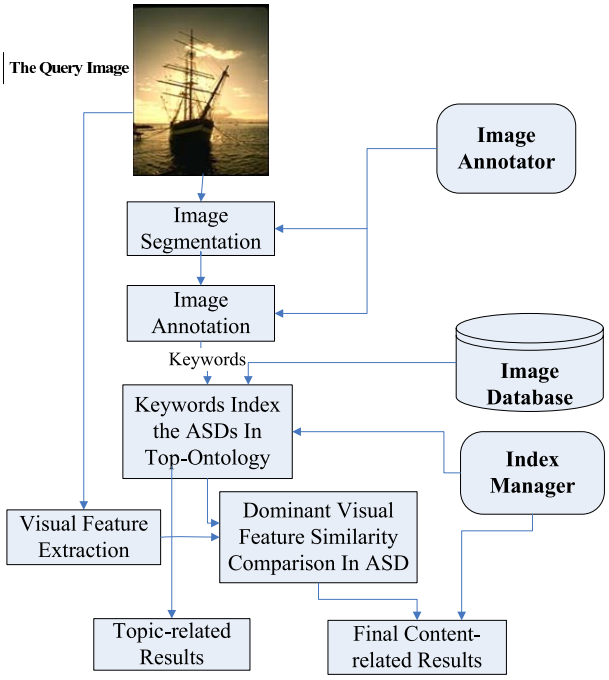

Figure 3: The Retrieval Process of Query-by-Image-Example in HISA

The query process of query-by-image-example is shown in Figure 3. HISA firstly uses the annotation words of the query image to prune the irrelevant semantic branches in the ontology tree, and then searches the ASD structures pointed to by the leaf nodes to fetch the candidate visually-similar images.

4. THE PRE-COMPUTATION OF HISA

The pre-computation process involves three steps: (1) image annotation, (2) ontology construction, and (3) domain-feature selection.

4.1 Image annotation

The salient objects and the main background contained in one image are perceptionally important for image semantic recognition. We start the pre-computation process by attaching these elements with proper annotations. Instead of obtaining elaborate annotation with enriched semantic information which can be directly used for keyword-match search, we make a trade-off by just annotating the salient objects and the main background for each image. The Annotator utilizes the image segmentation technique and the translation(words-to-regions) model to automatically generate annotation words of the images in the database. We adopt the Grabcut [6] algorithm, an efficient foreground/background segmentation technique, to extract the salient objects and the main background. We also use the co-occurrence translation model between keywords and blob-tokens to make the association between a keyword and a blob-token. In this way, our automatic annotation method is more reliable and robust.

4.2 Ontology construction

We construct the ontology using the keyword set, the common-sense knowledge, and the domain knowledge. The keyword set K generated from the annotation is used for the ontology analysis. The annotation words used by HISA are different from the conventional image annotation which we obtained from the image source (described as above). Figure 4 shows the difference between them. We apply a

Figure 4: Generic annotation words used by HISA vs. conventional keywords from the image source

Generative Hierarchical Clustering pattern (GHC) to construct a tree-like conceptual taxonomy in top-down fashion. As sketched in Fig 1, the root node is the whole image database, followed by the topics that may be of interest to users, and related sub-topics are list in a recursive way. For each node split we perform GHC algorithm in the following three main steps:

***Trial-Query Construction*:** Prepare a set of keyword-based trial-queries $Q = \{q_1, q_2, \ldots, q_m\}$ to induce related sub-categories.

***Relevance Feedback Refinement*:** We adopt a stochastic process, which is similar to [7], as the fuzzy clustering algorithm. An image may fall into more than one category according to a *relevance threshold*.

***Statistical Data Analysis*:** Extract the representative keywords for each sub-category. We use a keyword vector

$\overrightarrow{K}$ with corresponding normalized weight vector $\overrightarrow{W}$ to represent each node.

4.3 Domain-feature selection

The Domain-Feature Selector extracts the visual features for each image by incorporating color, and wavelet coefficients to form a high-dimensional feature vector. We adopt the Daubechies' wavelets [10] to extract the wavelet coefficients using the LUV color space. For each ASD, we construct a Matrix F_{n*m} where n is the total number of images indexed by this ASD, and m is the initial number of visual feature dimensions. We apply the principal component analysis (PCA) on the original data space (Matrix F_{n*m}) to reduce the dimensionality. The resulting principal components will give the dominant visual feature vector $\overrightarrow{V}$ and its corresponding normalized weight vector $\overrightarrow{W}$.

5. STRUCTURE OF THE DEMO

In this demonstration, we use a collection of 60000 images, which contains a variety of images with various contents and textures. We will illustrate the novel techniques in HISA as following:

- *The HISA structure*
  HISA captures a collection of concepts for large image database, and provides a dynamic snapshot for their interrelationships, which provides a more natural way for retrieval compared to conventional systems. Both the ontological knowledge and the visual features are captured in the HISA structure. In the prototype, we also show an extendable mechanism for adding new concepts to the ontology, removing or refining some old ones from it.

- *The Query Process*
  HISA provides high recall and high precision without compromising the speed in a large diversified image database. We shall illustrate the keyword-based queries and queries by image examples. The query performance is higher compared to conventional systems in terms of both precision and speed. We shall also demonstrate that multi-semantic mining can help to improve the query performance of the system.

- *The Pre-computation Results*
  We compare the annotation results generated in the pre-computation process with those based on a keyword-only system. We shall illustrate the effectiveness of PCA in improving the search performance inside the ASDs.

6. REFERENCES

[1] K. Barnard, D. Forsyth. Learning the semantics of words and pictures. Journal of Machine Learning Research, 3:1107-1135,2003.

[2] J. Jeon, V. Lavrenko, R. Manmatha. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models. 26th Annual Int.ACM SIGIR Conference, Toronto, Canada, 2003.

[3] J. Li, J. Z. Wang. Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. IEEE Trans.on Pattern Analysis and Machine Intelligence, vol.25, no.9, pp.947-963,2003.

[4] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to WordNet: an on-line lexical database. International Journal of Lexicography, 3, 235-244(1990).

[5] J. Y. Pan, H. J. Yang, P. Duygulu, and C. Faloutsos. Automatic Image Captioning. In Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME 2004), 2004.

[6] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. Proc.ACM Siggraph, 2004.

[7] M. L. Shyu, S. C. Chen, M. Chen, C. Zhang, C. M. Shu: MMM A Stochastic Mechanism for Image Database. Proceedings of the IEEE 5th International Symposium on Multimedia Software Engineering (MSE2003), pp. 188-195, December 10-12, 2003, Taichung, Taiwan, ROC.

[8] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain. Content-Based Image Retrieval at the End of the Early years. IEEE Trans.Pattern Analysis and Machine Intelligence, vol.22, no.8, pp.1349-1380, Dec.2000.

[9] M. Srikanth, J. Varner, M. Bowden, D. Moldovan. Exploiting Ontologies for Automatic Image Annotation. Proceedings of the 28th Annual International ACM SIGIR (SIGIR 2005), August 2005.

[10] J. Z. Wang, G. Wiederhold, O. Firschein, S. X. Wei. Content-based image indexing and searching using Daubechies' wavelets. Int.J.on Digital Libraries 1(4):311-328,1998.

[11] R. Weber, H. Schek, and S. Blott. A quantitative analysis and performance study for Similarity Search Methods in High Dimensional Spaces. In proceedings of the 24th International Conference on Very Large Data Bases (VLDB), 1998, pp. 194-205.

[12] R. F. Zhang and Z. F.(M.) Zhang. Image Database Classification based on Concept Vector Model. proceedings of the 2005 IEEE International Conference on Multimedia and Expo(ICME), Amsterdam, The Netherlands, July,2005.