Multiresolution Organization and Browsing of Images Using Multimedia Knowledge Networks

Ana B. Benitez
Dept. of Electrical Engineering, Columbia University
1312 Mudd, #F6; 500 W. 120th St., MC 4712
New York, NY 10027, USA
Phone: 1-212-854-7473 / Fax: 1-121-932-9421
ana @ ee.columbia.edu

Shih-Fu Chang
Dept. of Electrical Engineering, Columbia University
1312 Mudd; 500 W. 120th St., MC 4712
New York, NY 10027, USA
Phone: 1-212-854-6894 / Fax: 1-121-932-9421
sfchang @ ee.columbia.edu

ABSTRACT
This paper presents novel methods for organizing and browsing annotated images using multiresolution networks that represent knowledge about the images (e.g., objects, events and interactions). At the highest resolution, images are organized by perceptual knowledge (e.g., image clusters and visual relations), semantic knowledge (e.g., word senses and semantic relations), and statistical interrelations discovered from the collection. This process drives on the integrated processing of both images and annotations and the use of the electronic dictionary WordNet. Knowledge networks at lower resolutions are constructed by clustering similar concepts together. Users can then browse the annotated images by navigating the resulting knowledge network pyramid. Ideas from fish-eye views and spring modeling are exploited for displaying concepts using text and image example, and for drawing networks, respectively. Although the network pyramid is hierarchical, the navigation is not restricted to the hierarchy. Experiments are being conducted with users to evaluate the effectiveness, efficiency, and subjective satisfaction of the users in performing common browsing tasks such as image search. In these experiments, the proposed techniques are being compared to the sequential navigation of concepts in the initial knowledge network.

Categories and Subject Descriptors

General Terms
Algorithms, Management, Performance, Experimentation, and Human Factors.

Keywords
Image organization and browsing, knowledge discovery and summarization, multiresolution knowledge networks, multimedia knowledge, semantics, and perception.

1. INTRODUCTION
In recent years, there has been a major increase in available multimedia and in technologies to access multimedia. Users need and want tools for effectively and efficiently organizing and browsing multimedia, preferably, at the semantic level (e.g., people and objects in multimedia). Through browsing, users can gain a quick insight into the content of a collection and perform a variety of exploration tasks, with or without a particular goal in mind (e.g., finding a specific image or answering some questions). However, current multimedia browsing approaches often are based on feature descriptors at the perceptual level (e.g., color and texture) failing to meet the user needs. As an example, the most popular user operation in the web image search engine WedSEEk [25] is subject hierarchy browsing, over visual feature-based searches. This paper focuses on the organization and browsing of collections of images with annotations. Related approaches lack flexibility: they are often constraint to a unique space, hierarchy or network structure.

In this paper, we present innovative approaches towards multiresolution organization and browsing of images using visual and text features. The main contribution of this work is the organization of annotated images as a pyramid of concept networks based on knowledge extracted from both the images and the annotations (see Figure 1). Knowledge is usually defined as facts about the world and represented as concepts (e.g., dog and color pattern) and relations among the concepts (e.g., specialization and similar color). In addition, we propose to browse the image collection by navigating the knowledge network pyramid by using fish-eye views, spring network modeling and non-hierarchical navigation. Finally, this work evaluates the proposed techniques by measuring the effectiveness, efficiency and user satisfaction with which users carry out common browsing tasks such as searching for a specific image or concept. These methods are developed and used within the IMKA (Intelligent Multimedia Knowledge Application) system [3], which aims at extracting useful knowledge from multimedia and at implementing intelligent applications that use that knowledge.

Multimedia knowledge networks are constructed from annotated images [2] by clustering the images based on visual and text descriptors (perceptual knowledge), and disambiguating the senses of the words in the annotations using the electronic dictionary WordNet [19] and the image clusters (semantic knowledge). Visual, statistical and semantic relationships among concepts (e.g., image clusters and words senses) are discovered using visual descriptor distances, statistical inference, and WordNet, respectively. The concepts in the initial knowledge network are iteratively clustered together based on an information content measure of their distance. This process results in knowledge networks at different resolutions of detail forming a pyramid, as shown in Figure 1. Users can browse the annotated images by navigating the knowledge network pyramid. The visualization of the knowledge networks uses fish-eye views [6] for displaying concepts using text and image examples and spring

modeling [14] for drawing networks. Although, the knowledge network pyramid is hierarchical, the navigation of this structure is not strictly hierarchical. When a user expands a concept into a higher resolution, the system displays not only the sub-concepts but also the nearest concepts in the higher resolution. At any given point, a network with at most eight concepts is displayed to the user because that is the about the size of the human short-term memory [20]. Ongoing experiments with users aim at evaluating the proposed techniques in terms of effectiveness, efficiency and satisfaction of users in performing common browsing tasks. Examples of user tasks are locating or searching for a specific image or concept in the collection. The proposed browsing techniques are being compared to a sequential navigation of the knowledge in the highest resolution. Effectiveness and efficiency are measured in terms of the accuracy achieved, and the time spent, in performing the designated tasks. The users are questioned afterwards about their preferences of browsing system and the difficulties encountered during the experiments.

The paper is organized as follows. Section 2 reviews relevant prior work on image organization and browsing. Some work for evaluating these techniques is also discussed. Section 3 summarizes the knowledge discovery and summarization for organizing the images. The way the images and the knowledge are browsed is detailed in section 4. Section 5 presents the experiment setup for evaluating the proposed techniques. Finally, section 6 concludes with a summary and some future work.

2. PRIOR WORK

Relevant prior work on image organization includes dimensionality reduction of visual feature descriptors [6][22], image clustering based on visual features and/or annotations [1][4], and thesaurus-based approaches [26][27]. Several works map images onto one- and two-dimensional spaces, [6] and [22], respectively, based on feature descriptors extracted from the images. The interest of these approaches is limited because they only display visual relationships and do not provide a structure to organize the images. Image clustering approaches group images hierarchically using feature descriptors extracted from the images [4], or by modeling the distribution of visual descriptors and words [1], among others. However, these structures are restricted to hierarchies. Finally, several works propose to organize images based on existing thesauri [26][27]. Relevant concepts in thesauri are found for the images based on image annotations [26] or user feedback [27]. However, only one network is used to organize the images, which may become of considerable size and complexity. There is some work in concept network reduction, EZWordNet [18] and VISAR [5], but the reduction operators are manually defined and lacking generality.

Prior work on image browsing is limited to showing the images in a low dimensional space [6][22], in the clusters at a given level of the hierarchy [4], or in association with a concept in a thesaurus, usually visualized as a text hierarchy [27]. More sophisticated techniques make use of fish-eye views (e.g., change size and orientation images based on current viewing point) to represent similarity relationships among the images [6]. Another uses simple and predefined graphical visualizations of few connected concepts [26]. The most similar work to the techniques proposed in this paper is the Information Navigator [10]. The Information Navigator displays text documents and keywords graphically as networks based on occurrence statistics. It also supports fish-eye views and overview diagrams to facilitate the navigation of the networks. The proposed techniques differ from the Information Navigator in two ways: (1) in integrating knowledge from external resources such as WordNet, and (2) in summarizing knowledge at multiple resolution levels.

Most work on information browsing and visualization do not report any evaluation. An exception is the thesaurus approach [27] that used a retrieval user task to evaluate the performance of the proposed browsing and retrieval system. The evaluation of these techniques is difficult because there is of the wide range of possible navigation tasks (e.g., locate, distinguish and categorize; see [21] for more examples) in addition to the complications of interacting with users. More recent work has proposed a taxonomy of user tasks [21] and several information theory metrics [28] for evaluating visual displays of information. We follow a strategy similar to [27] by monitoring the time and the correctness with which users execute common navigation tasks. We also question users after the experiments about their browsing preferences.

3. IMAGE ORGANIZATION

This paper proposes to organize annotated images using multimedia knowledge discovered and summarized from the image collection. This process consists of five steps as shown in Figure 2. First, images and annotations are segmented and feature descriptors extracted from them (basic image and text processing). Then, images are clustered based on the feature descriptors, and similarity and statistical relationships discovered between the clusters (perceptual knowledge). The next step is to disambiguate the senses of the words in the annotations using WordNet and the image clusters, and to find semantic relationships among the detected senses (semantic knowledge). Other interrelations among concepts are discovered by learning statistical dependencies among concepts using Bayesian networks (knowledge interrelation). Finally, knowledge is summarized into a pyramid of concept networks by clustering similar concepts together at different resolutions (knowledge summarization). Each step is briefly described in this section; see [2] for more details.

3.1 Basic Image and Text Processing

During the first step, the images and the annotations are processed independently. The images are segmented into regions with homogeneous color and edge using a merge-and-split region segmentation method [30]. This technique has been proven to provide excellent segmentation results. After segmenting the images, descriptors are extracted from the images and the regions to represent visual features such as color. Color histogram and edge direction histogram are used globally for images [15][23]; and mean color, aspect ratio and pixel number locally for regions. These feature descriptors have been shown to be effective in image or video retrieval [15][23][30].

In an equivalent way, the words in the annotations are tagged with their part-of-speech information (e.g., noun and verb) using [17] and stemmed down to their base form (e.g., "burned" to "burn"). Then, stopwords, (i.e., frequent words with little information such as "be"), non-content words (i.e., not nouns, verbs, adjectives or adverbs), and infrequent words are discarded. The remaining words are represented as a vector using word-weighting schemes

[9] such as tf*idf - term frequency weighted by inverse document frequency- and log tf*entropy - logarithmic term frequency weighted by Shannon entropy of the terms over the documents.

3.2 Perceptual Knowledge Discovery

The second step consists on discovering perceptual knowledge from the image collection. Images are grouped into clusters based on their visual and text descriptors. Every image cluster is considered a perceptual concept. A diverse set of well-known clustering algorithms [11] is supported: the k-means, the Ward clustering, the K-Nearest Neighbors (KNN), the Self-Organizing Map (SOM) and the Linear Vector Quantization (LVQ) algorithms. Feature descriptors for images can be reduced using Latent Semantic Index (LSI) [7], concatenated, and normalized (bin mean and variance to zero and one, respectively).

Some clustering algorithms provide relationships among the clusters (e.g., neighboring in SOM and hierarchical in Ward). Additional relationships among image clusters are found based on centroid proximity and cluster statistics. For example, a cluster is said to have similar relationships with its k-nearest cluster neighbors based on their centroids' distances. Another example, if two clusters use the same feature descriptors and their conditional probabilities are one or very close to one, they are considered equivalent.

3.3 Semantic Knowledge Discovery

In this step, semantic knowledge is extracted by disambiguating the senses of the words in the annotations using WordNet and the image clusters (see section 3.2). Each detected word sense is a semantic concept in the knowledge. WordNet is an electronic dictionary that organizes English words into sets of synonyms (e.g., "rock, stone") and connects them with semantic relationships (e.g., generalization) [19].

The novel principle governing our approach is that the images in the same cluster are often related semantically although often very generally (e.g., images of animals and flowers share semantics such as "nature"). The proposed technique also follows the two widely accepted principles for Word-Sense Disambiguation (WSD) [29]: consistent sense for a word, and semantic relatedness of nearby words, in a document. During this process, the words annotating the images in each cluster are matched to the definitions of the possible senses of each word using word-weighting schemes (e.g., tf * idf). The definition of a sense (e.g., sense "rock, stone") is constructed by concatenating and weighting the words in the synonyms (e.g., "rock, stone"), the meaning (e.g., "lump of mineral matter") and the usage examples (e.g., "he threw a rock at me") of that sense together with the ones of directly or indirectly related senses (e.g., "lava", a specialization of "rock, stone") as provided by WordNet. Relationships and intermediated senses among detected senses are found in WordNet.

3.4 Knowledge Interrelation Discovery

During this step, statistical dependencies among concepts are learned. First, meta-classifiers are trained to predict the presence of concepts in images. Then, a Bayesian Network (BN), whose nodes are the meta-classifiers and whose initial topology is the one of initial knowledge, is trained to learn new statistical relationships among the concepts.

Individual classifiers are trained to predict the presence of a concept in images based on different combinations of visual and text descriptors. The classes for each classifier are labels such as {strong presence, weak presence, no presence} indicating the strength of the presence of a concept in an image. For image clusters, the class labels are the presence or absence of an image in the cluster; for semantic concepts, the quantized matching rank percentages for each detected sense. A diverse set of well-known classification algorithms [8] is used including Naïve Bayes (NB), Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN). Multiple classifiers for a concept (e.g., different descriptors) can be combined into a meta-classifier using boosting and stacking techniques.

Bayesian Networks (BNs) are directed graphical models that allow the efficient and compact representation of joint probability distributions for multiple random variables [8]. Interrelations among concepts are discovered by learning a Bayesian nework. The nodes of the Bayesian network are the classifiers; each node is thus representing a concept. The topology of the Bayesian Network is initialized with the topology of the initial knowledge network after removing cycles; our best guess based on prior knowledge. A new statistical relationship is added to the knowledge for each arc in the learned BN that does not have a corresponding relationship in the initial knowledge.

3.5 Hierarchical Knowledge Summarization

Finally, the knowledge network is summarized at different resolutions by clustering similar concepts together, either image clusters or work senses. The resulting structure is a knowledge network pyramid, as the one shown in Figure 1.

The distance among concepts in a knowledge network is calculated using a novel technique based on both concept statistics and network topology. It generalizes measure [13] from a concept tree to an arbitrary concept network. Assuming binary relations, the distance of a relationship r connecting concepts c and c' is calculated as follows:

$$\begin{aligned} dist(c, c', r) &= p(c)\ IC(c, r \mid c') + p(c')\ IC(c', r \mid c) \\ &= -p(c)\ \log(p(c, r \mid c')) - p(c')\ \log(p(c, r \mid c')) \end{aligned} \quad (1)$$

where IC(x) is the information content of x, p(c) is the probability of encountering an instance of concept c, and p(c,r|c') is the probability of encountering an instance of concept c through relation r given an instance of concept c'. The intuition behind the proposed concept distance is the following: the distance of a relationship between two concepts increases with their probabilities but decreases with the conditional probabilities for that relationship. Image occurrences are propagated on the knowledge network through the relationships because an instance of a dog is also an instance of an animal, which may not be explicit in the knowledge. A expert can assign a propagation weight for each relation and direction. This step is also necessary during the discovery of knowledge interrelation (see section 3.4). The distance between two concepts is calculated as the distance of the shortest distance path between the two concepts.

Finally, concepts are clustered hierarchically using a modified KNN clustering algorithm. At each step, the algorithm merges the clusters of the two data items with the largest number of shared neighbors. The KNN clustering algorithm [12] was selected because of the continuity of the clusters and because no distance function is required. The input to the clustering algorithm is the k nearest concepts for each concept. From the resulting binary cluster tree, a hierarchy in which each cluster has the same number of sub-clusters is obtained by removing intermediate clusters. The centroid of a cluster is set to the data item with maximum number of shared neighbors with the other data items in the cluster. A knowledge network is constructed for each level in the hierarchy by replacing each cluster for a super concept. The super concept inherits all the relationships and media examples of its sub-concepts. An example of a possible knowledge network pyramid obtained through this process is shown in Figure 1.

4. IMAGE BROWSING

Once some images have been organized into multiple resolutions, users can browse the collection by navigating the resulting knowledge network pyramid. The preferred number of concepts at each resolution of the pyramid is 1x8, 4x8, 42x8, …. In the knowledge network visualization, we propose to use fish-eye views [6] for displaying concepts using text and image examples, and spring modeling [14] for drawing networks. Although, the knowledge network pyramid is hierarchical, the navigation of this structure is not strictly hierarchical. When a user expands a concept into a higher resolution, the system displays not only the sub-concepts but also the nearest concepts in the higher resolution. At any given point, a network with at most eight concepts is displayed to the user because that is the about the size of the human short-term memory [20].

4.1 Concept Visualization

Concepts in knowledge networks are displayed using their corresponding media examples (see Figure 3). Two different views are created for each concept, a simple and an extended view, using fish-eye views [6]. The idea behind fish-eye views is to modify the attributes of the objects being displayed (e.g., size, orientation and shape) based on the current viewing point. For example, an object close to the viewing point is in focus as opposed to an object that is far away.

Perceptual concepts are image clusters based on visual and/or text descriptors (see section 3.2). The media examples of perceptual concepts are the images grouped by the cluster. The simplified view of a perceptual concept is set to the centroid image. In the extended view of a perceptual concept, the centroid image is surrounded by four (or another number) context images of half the size. The context images are the centroids images of clustering the image examples (excluding the centroid) into four clusters using a 2x2 SOM. The clustering uses the same feature descriptors used to construct the perceptual concept.

Semantic concepts are word senses in WordNet (see section 3.3). The media examples of semantic concepts include the synonyms of the sense, as provided by WordNet. Other media examples of a semantic concept are images on whose annotations the word sense was detected. The simplified view of a semantic concept is set to the synonyms of the sense. The extended view of a semantic concept consists, in addition, includes four context images surrounding the synonyms. The context images are obtained in the same way as the ones for the perceptual concepts. They are useful in clarifying the meaning of senses that share the same synonyms (e.g., "rock, stone" as a lump of mineral matter or as the material making up the Earth).

Super concepts are groups of semantic and perceptual concepts, and other super concepts. The simplified view of a super concept is set to the extended view of the centroid of the concept cluster. The extended view of a super concept consists of the extended views of the centroid concept together with four context concepts. The context concepts are obtained in a similar way to the context images for the perceptual concepts, by clustering the elementary concepts based on the number of shared neighbors into four clusters using a 2x2 SOM.

4.2 Network Visualization

This section describes how (parts of) knowledge networks are drawn on a 2D display, i.e., how each node in the network is positioned on the display based on the relationships and distances to other nodes.

There are many methods for drawing graphs (i.e., a set of nodes connected by arcs). The spring modeling algorithm [14] was selected because it minimizes arc crossings when drawing the graph. This algorithm considers each pair of nodes in a graph to be connected by a virtual spring. The algorithm iteratively repositions the nodes of the graph so as to minimize the overall tension or energy of the system of springs. Consider a string of negligible mass, one end of which is attached to a wall. The energy stored in the spring whose free end is stretched a distance X is given as follows:

$$E = \tfrac{1}{2} K X^2 \tag{2}$$

where K is the spring force. The total energy of the system of springs connecting the n nodes of a graph is given by:

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \tfrac{1}{2} k_{ij} (| v_i - v_j | - l_{ij})^2 \tag{3}$$

where vi is the position of node i in the display, dij is the length of the shortest path between nodes i and j, and lij and kij are the length and the strength of the spring between nodes i and j, respectively. lij and kij can be calculated in terms of the distance among nodes i and j –dij-, the length of a shortest side of the display area -L0-, and the maximum distance between any two

nodes, using $l_{ij} = \dfrac{L_0 \, d_{ij}}{\max_{z<l}(d_{zl})}$ and $k_{ij} = \dfrac{K}{d_{ij}^2}$, respectively. (4)

The application of the spring modeling algorithm to draw (parts of) knowledge networks is straightforward. The concepts are the nodes of the graph; and the concept relationships, the arcs connecting the nodes. The distance of each arc is the distance of the corresponding relationship as calculated using equation (1). To make the display of knowledge networks less cluttered, relationships with the highest distances are sometimes omitted.

4.3 Network Pyramid Navigation

This section describes how users can navigate the knowledge networks pyramid for browsing image collections. These techniques follow the Visual Information-Seeking Mantra [24]: "Overview first, zoom and filter, then details-on-demand".

The preferred number of concepts at each resolution of the pyramid is 1x8, 4x8, 4^2x8, …, which is constructed as described in section 3 (the numbers four and eight can be adapted by applications). This way, a concept at one resolution has four sub-concepts in the higher resolution. When the user starts browsing the image collection, the entire knowledge network at the lowest resolution is displayed. This network of eight concepts provides an overview of the image and corresponding knowledge of the entire collection.

At this point, the user can expand any of the displayed concepts into the higher resolution. The system will respond to the zoom in command by displaying the sub-network containing the four sub-concepts and the four nearest concepts in the higher resolution. Although the knowledge network pyramid is hierarchical, users can navigate the pyramid in a not hierarchical fashion by then zooming into one of the nearest concepts instead of a sub-concept. The operation of zoom out to a lower resolution is also supported by the system.

Apart from zooming in and out operations, users can also pan through the knowledge network at a given resolution using the next and previous commands. For enabling these operations, the concepts at a given resolution are linearized or ordered in an array using the following strategy. The first concept is selected at random. Then, concepts with the largest number of shared neighbors with the already selected concepts are iteratively picked and added to the concept array. If a concept is being displayed (i.e., subnet with sub-concepts and nearest concepts), when the user selects the next concepts, the next concept in the array at that resolution is displayed (sub-net with sub-concepts and nearest concepts).

At any given point, networks of only up to eight nodes are shown to the user because human short-term memory has been shown to have a capacity of "the magical number seven plus or minus two" [20]. Additional browsing functionality available to users is to retrieve the images and other media examples of a concept, and vice versa.

5. EXPERIMENTS

Knowledge was extracted from a collection of 271 nature images and their annotations. The knowledge was then summarized into a knowledge network pyramid of dimensions as indicated in section 4.3. The knowledge network pyramid is being evaluated based on how well users can carry out some common browsing tasks in terms of correctness, speed and user preferences.

5.1 Experiment Setup

The test set was a collection of 271 nature images from the Berkeley's CalPhotos collection (http://elib.cs. berkeley.edu/photos/). The images in CalPhotos are categorized into plants (86), animals (82) landscapes (66) or people (37) and have brief annotations in the form of keywords (e.g., "people, culture, Zulu warrior, Africa").

During the knowledge extraction process, the images were scaled down to a maximum height and width of 100 pixels. Words that appeared less than 5 times were discarded for the extraction of the text feature descriptors, whose dimensionality was further reduced to 125 bins using LSI. Perceptual knowledge was constructed by clustering the images based on color histogram, log tf * entropy, and a concatenated color histogram/log tf * entropy feature vector, into 16 clusters each. Semantic knowledge was discovered by disambiguating the senses of words annotating the images. The initially extracted knowledge had a total of 116 concepts, 48 of which were perceptual concepts (image clusters). A knowledge network pyramid of 1x8, 4x8, 4^2x8, …, was constructed with a total depth of seven levels.

On-going experiments aim at evaluating the proposed techniques for image browsing based on how well some designated tasks are carried out by users. Examples of browsing tasks that are being considered in our experiments are locating a specific image or concept in the collection, and identifying or categorizing the subject of the image collection. The specific criteria used in these experiments are the effectiveness, efficiency and satisfaction of users in performing the tasks. The effectiveness is being measured as the correctness or accuracy of the finished tasks, for example, how often users find the target images. Efficiency is calculated as the ration between the task effectiveness and the time spent in completing the task. Finally, the data for user satisfaction is obtained by questioning the users about their preference of browsing system and the difficulties encountered in performing the tasks. These experiments are comparing the performance of the proposed techniques to the baseline browsing technique of sequential navigating the concepts in the highest resolution's knowledge network.

6. CONCLUSIONS

This paper proposes novel techniques for automatically organizing and browsing annotated images. Annotated images are organized in concepts network pyramids, which are constructed based on knowledge extracted from the images and the annotations. The initial knowledge discovered from the collection is clustered hierarchically resulting in the network pyramid. The discovered knowledge includes semantic knowledge (e.g., image clusters and relationships), semantic knowledge (e.g., word sense and relationships) and statistical interrelations among these. Users can browse the image collection by navigating the knowledge network pyramid. Experiments are being conducted to evaluate the effectiveness, efficiency, and subjective satisfaction of users in performing common browsing tasks such as image search. In these experiments, the proposed techniques are being compared to the sequential navigation of the concepts in the initially discovered knowledge network.

7. ACKNOWLEDGMENTS

This research is partly supported by a Kodak fellowship awarded to the first author of the paper.

8. REFERENCES

[1] Barnard, K., P. Duygulu, and D. Forsyth, N. de Freitas, D. Blei, and M.I. Jordan, "Matching Words and Pictures", submitted to JMLR.

[2] Benitez, A.B., and S.-F. Chang, “Automatic Multimedia Knowledge Discovery, Summarization and Evaluation”, submitted to IEEE Trans. on Multimedia; available at http://www.ee.columbia.edu/dvmm/publications.htm.

[3] Benitez, A.B., S.-F. Chang, and J.R. Smith, "IMKA: A Multimedia Organization System Combining Perceptual and Semantic Knowledge", ACM MM-2001, Ottawa, CA, 2001.

[4] Chen, J., C.A. Bouman, and J.C. Dalton, "Hierarchical browsing and search of large image databases", IEEE Trans. on Image Processing, Vol. 9, No. 3, 2000.

[5] Clitherow, P., D. Riecken, and M. Muller, "VISAR: A System for Inference and Navigation in Hypertext", ACM Conference on Hypertext, Pittsburgh, USA, Nov. 5-8, 1989.

[6] Craver, S., B.-L. Yeo, and M. Yeung, "Multi-Linearization Data Structure for Image Browsing", IS&T/SPIE-1999, Vol. 3656, pp. 155-166, Jan. 1999.

[7] Deerwester, S., S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, "Indexing by Latent Semantic Indexing", JASIS, Vol. 41, No. 6, pp. 391-407, 1990.

[8] Duda, R.O., P.E. Hart, D.G. Stork, "Pattern Classification", John Wiley & Sons, Second Edition, USA, 2001.

[9] Dumais, S.T., "Improving the retrieval of information from external sources", Behavior Research Methods, Instruments and Computers, Vol. 23, No. 2, pp. 229-236, 1991.

[10] Fowler, R.H., A.W. Bradley, W.A.L. Fowler, "Information Navigator: An Information System Using Associative Networks for Display and Retrieval", University of Texas - Pan American, Technical Report NAG9-551, #92-1.

[11] Jain, A.K., M.N. Murty, and P.J. Flynn, "Data Clustering: A Review", ACM Computing Surveys, Vol. 31, No. 3, 1999.

[12] Jarvis, R.A., and E.A. Patrick, "Clustering Using a Similarity Measure Based on Shared Near Neighbors", IEEE Trans. on Computers, Vol. c-22, No. 11, Nov. 1973.

[13] Jiang, J.J., and D.W. Conrath, "Semantic Similarity based on Corpus Statistics and Lexical Taxonomy", ROCLING, 1997.

[14] Kamada, T., and S. Kawai, "An Algorithm for Drawing General Undirected Graphs", Information Processing Letter, Vol. 31, No. 1, pp. 7-15, April, 1989.

[15] Kumar, R., and S.-F. Chang, "Image Retrieval with Sketches and Coherent Images", ICME-2000, New York, Aug. 2000.

[16] Ma, W.Y., and B.S. Manjunath, "A Texture Thesaurus for Browsing Large Aerial Photographs", JASIS, Vol. 49, No. 7, pp. 633-648, May 1998.

[17] McKelvie, D., C. Brew and H. Thompson, "Using SGML as a basis for data-intensive NLP", Applied Natural Language Processing, Washington, USA, April 1997.

[18] Mihalcea, R., and D. Moldovan, "Automatic Generation of a Coarse Grained WordNet", NAACL, Pittsburgh, 2001.

[19] Miller, G.A., "WordNet: A Lexical Database for English", Comm. of the ACM, Vol. 38, No. 11, pp. 39-41, Nov. 1995.

[20] Miller, G.A., "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information", Physiological Review, Vol. 63, 1956.

[21] Morse, E., Lewis, M., and Olsen, K. A. (2000). Evaluating Visualizations: Using a Taxonomic Guide. International Journal of Human Computer Studies, 53: 637-662.

[22] Rubner, Y., L.J. Guibas, and C. Tomasi, "The Earth Mover's Distance, Multi-Dimensional Scaling, and Color-Based Image Retrieval", DARPA Image Understanding Workshop, 1997.

[23] Rui, Y., T. Huang, and S.-F. Chang, "Image Retrieval: Current Techniques, Open Issues, and Promising Directions", Journal of Visual Communication and Image Representation, 1999.

[24] Shneiderman, B., "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", IEEE Symposium on Visual Languages, Boulder, CO, USA, 1996. Sep. 3-6, 1996.

[25] Smith, J.R., and S.-F. Chang, “An Image and Video Search Engine for the World-Wide Web”, IS&T/SPIE-1997, San Jose, CA, Feb. 1997.

[26] Tansley, R., "The Multimedia Thesaurus: Adding A Semantic Layer to Multimedia Information", Ph.D. Thesis, Computer Science, University of Southampton, Southampton UK, 2000.

[27] Yang, J., L. Wenyin, H. Zhang, Y. Zhuang, "Thesaurus-Aided Approach for Image Browsing and Retrieval", ICME-2001, Tokyo, Aug. 2001.

[28] Yang-Pelaez, J.A., "Metrics for the Design of Visual Displays of Information", Doctor of Philosophy Thesis, Dept. of Mechanical Engineering, Massachusetts Institute of Technology, 1999.

[29] Yarowsky, D., "Unsupervised Word-sense Disambiguation Rivaling Supervised Methods", Association of Computational Linguistics, 1995.

[30] Zhong, D., and S.-F. Chang, "An Integrated Approach for Content-Based Video Object Segmentation and Retrieval", IEEE Trans. on CSVT, Vol. 9, No. 8, pp. 1259-1268, 1999.

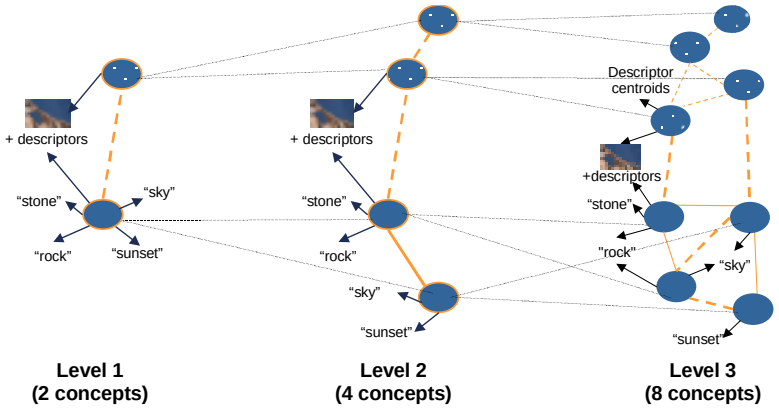

Figure 1. Multiresolution pyramid of three knowledge networks with two, four and eight concepts. A concept in a resolution has two sub-concepts in the higher resolution.

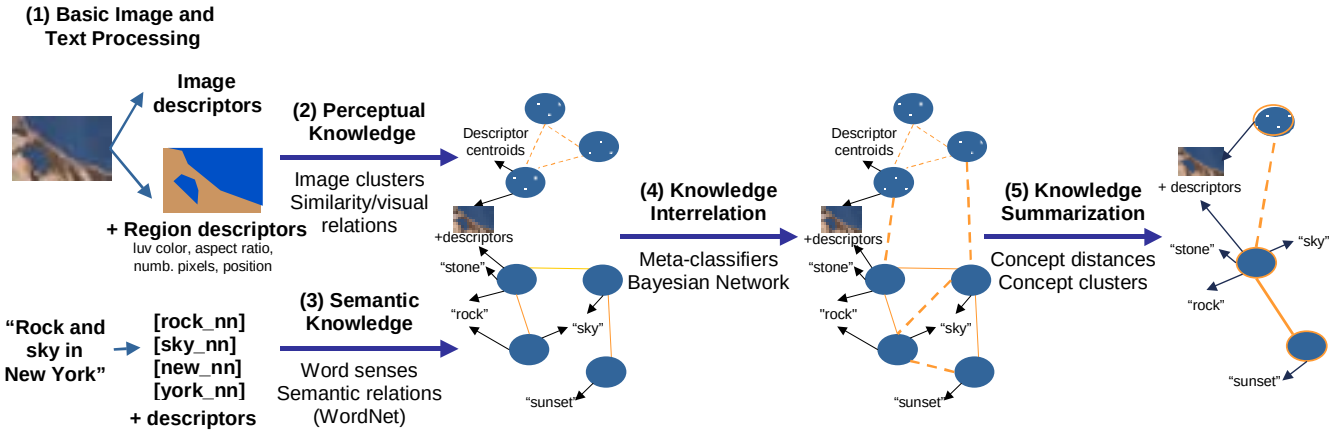

Figure 2. The organization of a collection of annotated images consists of five steps: basic image and text processing, perceptual knowledge discovery, semantic knowledge discovery, knowledge interrelation discovery and knowledge summarization. "_nn" indicates the preceding word is a noun. Blue circles represent concepts. The dotted circles are perceptual whereas the plain ones are semantic concepts. The dash lines between concepts represent perceptual relationships; whereas the plain lines are semantic relations. Arrow lines link concepts to their corresponding image and word examples.

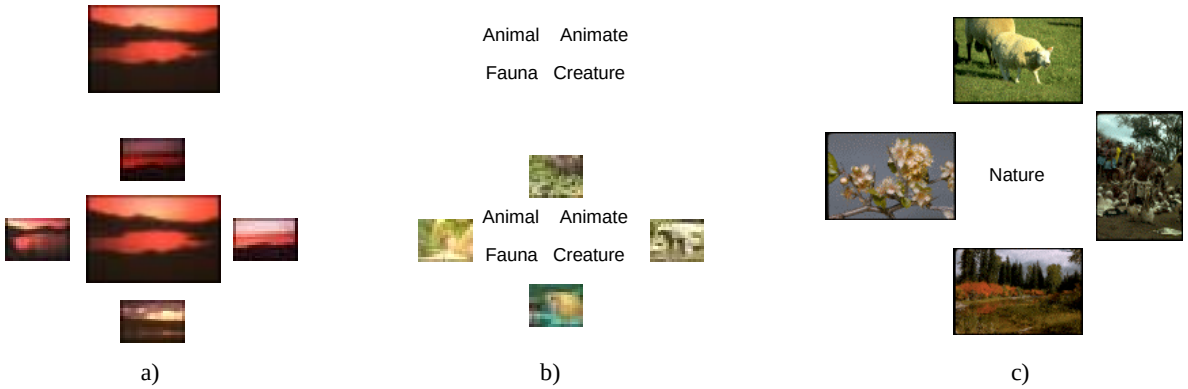

Figure 3. Examples of views for different kinds of concepts: a) simple (top) and extended (bottom) views of a perceptual concept; b) simple (top) and extended (bottom) views of a semantic concept; and c) simple view of a super concept.