Visual Information Retrieval from Large Distributed On-line Repositories

Shih-Fu Chang, John R. Smith, Mandis Beigi, & Ana Benitez
Department of Electrical Engineering
& Center for Telecommunications Research
Columbia University, New York, NY 10027

Digital images and video are becoming an integral part of human communications. The ease of capturing and creating digital images has caused most on-line information sources look more "visual". We use more and more visual content in expressing ideas, reporting, education, and entertainment. With the tremendous amount of visual information becoming on-line, how does one find visual information from distributed repositories efficiently, at least to the same extent as that of existing information retrieval systems. With the growing number of on-line users, how does one design a system with performance scalable to a large extent?

Visual information is rich in content. The same picture may invoke different responses from different users, at different time, and in different contexts. A picture may have different meanings at different levels, e.g., *description*, *analysis*, and *interpretation* described in [1]. Visual information may be represented in different forms — still images, video sequences, computer graphics, animations, stereoscopic images, and those for futuristic applications, such as multi-view video and three-dimensional video. Furthermore, visual information demands a large resource of network bandwidth and storage. All these factors have made the indexing and retrieval of visual information a great challenge.

Based on our experience in developing visual information retrieval system (VIRS) in the Web-based environment, we present our views of the present and future of VIRS, particularly those providing access to large distributed on-line repositories. We will discuss the role of content-based visual query, present a taxonomy for classifying existing VIRS, present a working Internet VIRS as a case study, describe a preliminary prototype of Internet meta VIRS, and finally make some bold conjectures about the dominant challenging in developing scalable VIRS for the future Internet environment.

Content-Based Visual Query

Recent research has produced much progress in visual information retrieval. Several systems, such as Virage, QBIC, Photobook, VisualSEEk, VideoQ [2,3,4,5,6] provide efficient tools for users to specify visual queries using image examples or visual sketches. Multiple visual features, such as color, texture, shape, motion, spatio-temporal composition, are used in combination with the textual features. The visual features used in these systems usually appear in the low level but can be automatically extracted without human intervention. One unique characteristics of these VIR systems is that the search is approximate, based on similarity ranking. Images returned on top of the result list have higher similarity with the query input and may not have any "exact" match to the attributes specified in the query. Figure 1 includes one image search example and one video search example based on the above mentioned functions.

FIGURE 1. Feature-based visual query. Query 1 uses the local color regions and spatial structures to find images of sunset or flower. Query 2 adds the motion feature to find relevant video clips, such as high jumps.

Other VIRS use the input from humans or supporting data to index the visual information in a direct, semantic way. Video icons are generated via manual annotations of objects (e.g., people, boat) or semantic events (e.g., sunsets) in [7]. Textual indexes are generated from the captions and transcripts of broadcast video [8,9,10] for news video retrieval.

A complementary function with visual search is summarization. Scene based techniques are used in efficient browsing interfaces and event detection and clustering [11,12]. Video analysis techniques are used to construct mosaic images for efficient browsing and indexing from continuous video sequences [13].

Work has also begun in a critical area which aims at automatic decoding of semantic meanings of visual content. Learning through iterative user interaction is used in [14] to map visual feature clusters to semantic classes. Models for visual classes (e.g., people and animals) are manually built using visual features and associated constraints and are used to classify unknown images [15].

Classification of VIR Systems

Many VIRS have been used in many different application domains, including libraries, museums, scientific data archives, photo stocks, and World-Wide-Web search engines. A high-level taxonomy will be useful in identifying the unique functions and critical issues. A taxonomy using the following criteria can be used to classify different VIR systems.

- Automation —
  Indexing features of images and video will determine the search functionalities users may use. Preparation of these indexes may be fully automatic or manual. Automatic methods usually are useful for low-level feature extraction only. High-level semantic indexes usually require human input in annotation or system training. Some systems provide interactive tools to assist users in selecting image objects and features and are thus qualified as semi-automatic systems.
- Multimedia Features —
  Indexing features may include those of images, video, text, audio, or any combinations. Most systems use individual features or multiple features independently. Integration of features have been shown in a few systems but has not been fully explored.
- Adaptability —
  Most systems use static indexing features, which are extracted in advance. Selection of features involves trade-offs of indexing cost and the search flexibility. However, due to the subjective nature of visual search, there exist needs of dynamic indexing features which adapt to changing user needs and application contexts.
- Abstraction —
  Images may be indexed at various levels, including feature (e.g., color, texture, and shape), object (e.g., moving foreground object), syntax (e.g., video shot), and semantics (e.g., image subject). Most automatic systems aim at low-level features, while the high-level indexes are usually done manually. Interaction among different levels is an exciting but unsolved area.
- Generality —
  The indexing schemes and database content may be customized to incorporate specific domain knowledge, such as those in medical and remote-sensing applications. Other systems may aim at unconstrained types of visual content such as those on the Internet.
- Content Collection —
  The population of content could be achieved by software robots, which roam freely over the World-Wide Web and automatically download visual content according to some heuristics. Or visual content may be manually prepared by domain operators, such as the on-line news archives and photo stocks. Or in the future, visual content may be submitted autonomously like the way people submit documents to the Usenet today.
- Categorization —
  When the database size grows, subject taxonomies will be very useful in providing hierarchical categories where users may freely navigate and browse through the entire database archive. Some systems simply provide browsing tools for users to interactively view images or video of interest, which can then be followed by a detailed query.

Other specialized factors may also be added to the above list of classification criteria. First, some systems perform feature extraction using the compressed images, rather than the original uncompressed pixel data [16]. This approach avoids expensive expansion of the coded data and manipulation in the decoded domain. Second, the idea of meta-search system for VIR [17] has been stimulated by similar work in the field of information retrieval. We will discuss this issue with more

details in a later section.

WebSEEk — A Case Study of Internet VIR

The World-Wide Web includes a rich collection of visual information, which is also inter-linked with a vast variety of non-visual information. Although there have been many popular search engines for non-visual information, the arena of visual search engine is not yet fully explored.

Visual information on the Web is highly distributed, minimally indexed, and schema-less. To explore the information wealth on the Web, we have developed a semi-automatic image search and cataloging engine, WebSEEk [19]. The aim is to provide a visual search gateway which provides tools to collect, analyze, index, and search for the visual information on the Web. Figure 2 shows the front-end user interface of WebSEEk.

FIGURE 2. WebSEEk: A content-based image search and cataloging system on the Web.

Figure 3 depicts a high-level system diagram for WebSEEk. WebSEEk can be interfaced to distributed visual sources, serve as an aggregation point, and act as a server for general clients, who can query, retrieve, and manipulate the visual content indexed by the gateway. WebSEEk stores the metadata, the pointers to the source, but not the actual content.

FIGURE 3. A High-Level System Architecture of Internet VIR

According to the VIR system classification criteria mentioned above, WebSEEk is semi-automatic, uses static visual features and textual key terms, indexes unconstrained visual content, uses a customized taxonomy, and collects content by autonomous software agents.

The image and video collection process is achieved by autonomous Web agents. Visual content is detected by mapping the file name extensions to the object types according to the Multipurpose Internet Mail Extensions (MIME) labels (e.g., .gif, .jpg, .qt, .mpg, .avi). Based on the images and video collected during a 3-month period, it's interesting to note that about 85% of the visual content collected are colored still images. Only 1% of visual content is video. The current database has about 650,000 images and 10,000 video sequences.

WebSEEk adopts the following approaches to address the unique challenges of the Internet VIR.

- Use of Multimedia Metadata —
  Most visual content on-line does not exist alone. It usually is accompanied with a rich set of metadata. For implementation simplicity, we use the URL and html tags associated with the images or video to extract meaningful key terms for direct indexing and classification of the visual content.
- Efficient Feature Extraction and Indexing —
  The massive amount of images and videos in WebSEEk prevents use of any sophisticated feature set which requires intensive analysis and query computation. We use basic color features (i.e., binary color sets and color histograms) with novel indexing schemes to achieve a query response time within 2 seconds (not including the network delay). In addition to keyword-based search, users may also search for images using features of color distribution.
- Image Manipulation for Efficient Viewing —
  Visual query is an iterative, interactive process. Full-resolution images are not needed until the final stage when the user issues the explicit request. Reduced representations of images or video can be efficiently extracted on-the-fly or in advance. For video, automatic video shot segmentation and key frames selection are used.
- Us of Taxonomy and Subject Classification —
  A free browsing mechanism is not suitable for a large VIR system. Users tend to prefer guided navigation with a clearly defined hierarchical subject category. From our experiments, we found that subject navigation is the most popular access method in VIRS.

Unlike specialized systems, Internet VIR does not have well-defined schema or associated metadata. Therefore, a minimal set of common attributes available in most content sources are used. We use URLs and HTML tags to extract the key terms, which are then used to map the image or video to one or more subject classes in the taxonomy.

Figure 4 shows some descriptive key terms as well as non-descriptive key terms and their mappings to the taxonomy.

FIGURE 4. Examples of textual terms and their mappings to the subject classes.

The taxonomy is a key component of the Internet VIR. It contains more than 2000 classes and use a multi-level hierarchy. It is constructed in a semi-automatic way in which operators first design the basic hierarchy and then periodically inspect the important candidate classes suggested by the computer. Figure 4 also includes some example classes in the taxonomy.

Automatic subject classification using the simple key terms is not perfect. However, it does provide a satisfactory performance. Our subjective evaluation shows over 90% accuracy in most classes. This also conforms with the use statistics we have collected based on about 800,000 query and browse operations we have recorded in the initial deployment, which currently serves about 10,000 daily hits.

As shown in Figure 5, subject-based query is the most popular query method for images (accounting for 53.5% of image queries). Content-based image queries only accounts for 3.7% of the total queries. However, this may be partly affected by the limitation of the content-based search functions (color histograms only) implemented in the current system. Another important note is that we did not attempt to analyze the background of the user group, who may be highly specialized.

FIGURE 5. Use Statistics of an Internet VIR (WebSEEk)

Several related systems also provide visual content search on the Internet. [25] reported a system emphasizing the use of people portrait images. Commercial systems like those in [26] uses proprietary classification and annotation schemes, along with some basic feature-based search functions.

Meta Search Engines for Images

The proliferation of text search engines on the Web has motivated recent research in “integrated search” or so called “meta search engines”. The meta search engine serves as a common gateway linking users to multiple cooperative or competitive search engines. It accepts query requests from users, sometimes, along with user-specified query plans to select target search engines. The meta search engine may also keep track of the past performance of each search engine and use it in selecting target search engines for future queries.

A working meta search engine includes three basic components, as depicted in Figure 6 [17]. The “query interface component” accepts the query specification from the user and translate it to compatible query scripts to each target search engine. The “dispatching component” selects target search engines for each query. The “display interface component” merges the query results from each search engine. It may also perform required format conversion or image manipulation for producing a single list of displayable images for the client.

FIGURE 6. Basic components of meta search engines.

We have developed a prototype meta image search, MetaSEEk, for proof of concept [18]. The target search engines include VisualSEEk, WebSEEk, QBIC, and Virage[1]. The front-end interface allows for browsing of random images from different engines, selection and weighting of different visual features, and entering textual features (such as URL and keywords) used in the query. Figure 7 shows the user interface on the Web. It includes color percentage, color layout, and texture, which are supported by all three target search engines. Users may search for images based on examples, URLs, or text. Details of the implementation are discussed in [18]. We summarize our experiment

1. All subsequent discussions of each system is based on the implementation available on-line when the experiment was conducted.

experience and present some open issues in the following.

FIGURE 7. The query interface of MetaSEEk, which allows for browsing images from and issuing content-based queries to different target search engines. Image labeled in yellow are from QBIC, white from Virage, and blue from VisualSEEk.

- Feature Consistence —
  Although all target search engines support the same visual feature set, proprietary algorithms are used. For example, color histogram probably is the only consistent feature used among all target engines. Definitions of color layout and its computation are quite different. Mapping of each user-selected feature to those actually used in the target engines is based on the best effort. The relatively consistent features in most text search engines (e.g., TF*IDF) are not available in the visual domain.
- Query Results Merging —
  Unlike the text search engines, matching scores are not retuned from these VIRS. This is partly due to the lack of dominant features in the visual domain. Without the matching scores, the meta search engine may merge the result lists based on relative ranks in each individual list. Otherwise, the meta engine may use light-weight processes to compute simple visual features and rank the returned images on the fly.
- Special Functionalities —
  Each VIR system has specialized functionalities and individual limitations. Virage supports flexible weighting of each visual feature. QBIC and WebSEEk supports URL/ text search using images outside their own databases. But these functionalities are not shared among all target search engines. Usually, the special functionalities of each search engine cannot be easily ported to the common interface of the meta search engine.
- Performance Evaluation —
  Lack of effective evaluation metrics or benchmarks for VIR systems has been a critical but open issue. It's partly caused by the emerging status of this field and the strong dependence on application and user needs. Lacking a reliable performance metrics, it's difficult for the meta engine to monitor the performance of different target engines and make recommendations for subsequent queries. NO RESULTS and VISIT were used in [17] to assign query effectiveness. But in VIR, NO RESULTS metrics usually are not available since visual query is based on similarity ranking. There are always some images returned in a ranked order, unless a threshold is applied. However, determination of any threshold values will be tricky.

Towards a Scalable Solution for Internet VIRS

Although VIR is still in an emerging stage, we believe it's important to understand the important challenges we are facing along the path towards a scalable solution. A scalable Internet VIR solution needs to solve the following three most dominant problems: *heterogeneity*, *complexity*, and *bandwidth*.

- Heterogeneity
  Unlike text documents, visual materials do not share consistent formats, indexes, or metadata. This immediately became an important constraint in our development of meta search engines. Dozens of formats are used for images or video on the Web. Different algorithms are used for the same features in different VIR systems. There are no interoperable APIs among different VIR database systems. In the semantic level, customized taxonomies are used for cataloging images in each VIRS.

  Solutions are required for each level mentioned above. First, standardization of metadata and high-level content description will help solving the inconsistence problem in the semantic level. Such standardization will enhance the interoperability among VIRS and also encourage content users and creators to use metadata. Like the notable effort for the metadata for text documents, recent efforts such as that of Dublin Core [20] extend the standard set of core metadata elements to images. In addition, a standard taxonomy for graphic materials has been proposed for the use in the Library of Congress [21]. The audio-visual processing research community has also started to draft the standard for describing multimedia information content, i.e., MPEG-7 [22].

  In the feature level, the heterogeneity of algorithm and indexing scheme could be partly resolved by

extending a distributed database query protocol proposed in the field of information retrieval. For example, in the Z39.50 distributed database query protocol, the EXPLAIN facility is used for the client to learn the attributes and structures used in the server so that the client can to some extent self-configure and access the records in the server database [23]. The same type of facilities may be applied to VIR so that visual features and indexes can be "explained" to the clients. In addition, a common classification of features and query metrics can help the "explanation" process. Feature indexes may be classified to global vs. local, pixel-based vs. region-based, intra-object vs. inter-object attributes, spatial vs. temporal, and so on. Query metrics may be divided to count (i.e., histogram) vs. value, Euclidean vs. transform, etc. Such classification standard of features and associated metrics can also be used to develop an efficient translation framework for mappings features between different VIR systems. With this framework of classification and translation, meta search engines described earlier will also be able to merge the results from different VIR systems more meaningfully.

- Complexity

  Search of visual content is a complex process. Users may not know how to describe what he/she wants exactly. (It typically takes a historian or artist hours to find a picture from the library if he does not know the specific bibliographic information.) Cataloging images or video to a fixed subject hierarchy will partly solve this problem, but limitation of such traditional cataloging approach are also well known.

  On-line VIR systems could further solve this problem by supporting facilities for interactive browsing and dynamic querying. Users should be able to preview the search results, give relevance feedback, and refine queries efficiently. Particularly for Internet VIR, query preview is very important in providing summaries of the query results before transmission of full-resolution images or video. Summary may be provided in several levels. It could include the database statistics for each attribute or subject, like number of images with red regions or images classified to one particular subject class. With this type of cues, users are able to predict subsequent query outcomes to some extend and refine the detailed queries. Another type of summary is visual, like the mosaic image or key frame image of a video shot [13], topic transitional graph of a video [12], or an "average" image (e.g., eigenface) of a class of images in the database [4]. Detailed follow-up query can be made by pointing to the area of interest in the visual summary and adding more detailed query specifications.

  Diverse nature of Internet VIRS further increases the aforementioned issue of complexity. For example, in an image meta search system, how is one's subjective perception of query results fed back to individual VIR sources? How does one manipulate the query results from distributed VIR sources, e.g., use one image from one VIR to intersect the search result from other VIRs. How does one initiate a content-based meta search using a feature that is not globally supported in every VIRS?

  To solve this problem, we believe more "intelligence" needs to be added to the content itself. Images will no longer be a passive array of pixel values, instead, it should become "active", including carefully defined "methods" to manipulate the signals. Images and video may be rendered with different viewing conditions without modifying the raw data. Clients should be able to self-configure to handle different coding algorithms used in the original sources. New indexing features may be computed dynamically based on the actual application needs. Recent work like MPEG-4 [22] for object-oriented multimedia representation, and FlashPix [24] for intelligent image transport and rendering have started to related issues.

  Finally, as the volume of visual content and user grows on the Internet, the fundamental paradigm of accessing and dissemination of visual content may change as well. What we have discussed above can be classified into the general paradigm using the "pull" technologies. Users request connection to the information provider, specify searching or browsing criteria, manipulate the search results, and finally retrieve the specific images or video of interest. This modality may work well in local specialized domain, but not in a large distributed environment. We believe for wide-area visual information access, a different paradigm incorporating "push" technologies could be used. Knowledge-bearing software agents could be trained to learn user preference and then help users to filter interesting visual content for further review by the users themselves. The aforementioned approach using summarization at various levels can also be used here to reduce the load on the user.

- Bandwidth

The slow response time on the Internet is, and will continue to be, one major limitation of Internet VIR. Technological advancements in computing, networking, and data compression have greatly reduced the scale of the bandwidth problem, but have not fully solved it. Downloading a 10 second-long MPEG-1 video clip takes 9 minutes over a 28.8Kbps modem, 2 minutes over a 128K bps ISDN line, and 10 seconds over a T-1 dedicated line.

Internet VIR should use techniques to reduce the bandwidth demand. Progressive retrieval can be used to search for images from rough to fine scales. Scalable, multi-resolution image/video storage and representation schemes can be used to support this type of search methods. MPEG-2 has already included scalable coding options in its high-profile standard. Open image standards like FlashPix has also included multi-tile, multi-resolution features [24]. In the emerging video standard such as MPEG-4, content-based scalability is being considered to support independent access to any object of interest in a video scene [22].

For existing compression standards such as JPEG, MPEG-1 and MPEG-2, the compressed-domain image/video manipulation techniques can be used to help solving the bandwidth problem also [16]. Given that most visual content will be stored or transmitted in a compressed form, the compressed-domain approach implements the same functions, such as key frame selection, feature extraction, and visual summarization, without expanding the coded visual content back to the large, uncompressed domain, thus reducing the need of large bandwidth and intensive computing.

We summarize the main issues of distributed Internet VIRS and our speculative views towards a scalable solution in Figure 8. It includes the following critical components: metadata standard, open query protocol, visual summary, query preview, a new paradigm combining push and pull technologies, progressive retrieval, scalable representation, and compressed-domain filtering. The large number of issues and solutions reflects the multi-disciplinary nature of this area.

FIGURE 8. A high-level system architecture for distributed Internet VIRS.

Conclusions

The wealth of visual content on the Internet is growing at a rapid rate. But the technology for managing and searching the visual information is lacking far behind. The bandwidth problem of the Internet may prevent the use of real-time interactive video services in the short term. But the wide deployment of visual content sources on the Internet is imminent in many applications such as digital museums, on-line news, and photo/video stocks.

Research of various search techniques using multiple media are intensively underway. Many interesting systems provide search methods of “Query by X”, where X can be image example, visual sketch, keyword, speech, etc. Web-based image search engines using these visual search tools as well as the traditional catalog-based navigation tools have started to emerge. However, extension of VIRS to a large-scale distributed Web environment is a challenging task. It includes several critical technical barriers, particularly those concerning heterogeneity, complexity and bandwidth issues on the Internet. Solutions to overcoming these barriers rely on contributions from many disciplines as well as knowledge of practical application needs and user preference.

References

[1] Sara Shatford, “Analyzing the Subject of a Picture: A Theoretical Approach,” Cataloging and Classification Quarterly, Vol. 6(3), Spring 1986, The Haworth Press Inc.

[2] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R.C. Jain and C. Shu, “Virage image search engine: an open framework for image management”, Symposium on Electronic Imaging: Science and Technology - Storage & Retrieval for Image and Video Databases IV, IS&T/SPIE, Feb. 1996.

[3] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker, “Query by Image and Video Content: The QBIC System,” IEEE Computer Magazine, Sep. 1995, Vol.28, No.9, pp. 23-32.

[4] A. Pentland, R.W. Picard, and S. Sclaroff, “Photobook: Tools for Content-Based Manipulation of Image Databases,” Proc. Storage and Retrieval for Image and Video Databases II, Vol. 2185, SPIE, Bellingham, Wash., 1994, pp. 34-47.

[5] J. R. Smith and S.-F. Chang, “VisualSEEk: A Fully Automated Content-Based Image Query System,” ACM Multimedia Conference, Boston, MA, Nov. 1996.

(demo http://www.ctr.columbia.edu/VisualSEEk)
(ftp://ftp.ctr.columbia.edu/CTR-Research/advent/public/papers/96/smith96f.ps)

[6] S.-F. Chang, W. Chen, H.J. Meng, H. Sundaram, and D. Zhong, "VideoQ- An Automatic Content-Based Video Search System Using Visual Cues," submitted to ACM Multimedia 1997, also CTR-TR #??.
(demo: http://www.ctr.columbia.edu/videoq)

[7] M. Davis, "Media Streams: An Iconic Language for Video Annotation",Telektronik 4.93: Cyberspace, 1993.

[8] A. G. Hauptmann and M. Smith, "Text, Speech and Vision for Video Segmentation: The Informedia Project," AAAI Fall Symposium, Computational Models for Integrating Languageand Vision, Boston, November 10-12, 1995.

[9] B. Shahraray and D. C. Gibbon, "Automatic Generation of Pictorial Transscript of Video Programs," SPIE Vol. 2417, pp.512-518, 1995.

[10]Rakesh Mohan, "Text Based Search of TV News Stories," SPIE Photonics East, Intern. Conf. on Digital Image Storage & Archiving System, Boston, MA, Nov. 1996.

[11]H. Zhang, C.Y. Low, and S. Smoliar, "Video Parsing and Browsing Using Compressed Data," J. of Multimedia Tools and Applications, Vol. 1, No.1, Kluwer Academic Publishers, March 1995, pp.89-111.

[12]M.M. Yeung, B.-L. Yeo, W. Wolf, and Bede Liu, "Video Browsing using Clustering and Scene Transitions on Compressed Sequences," *IS&T/SPIE Symposium Proceedings,* Feb. 1995, San Jose, California. Vol. 2417, pp. 399-413.

[13]H.S. Sawhney, S. Ayer, and M. Gorkani, "Model-Based 2D and 3D Dominant Motion Estimation for Mosaicking and Video Representation," Proc. Fifth Int'l Conf. Computer Vision, 1995.

[14]T. P. Minka and R. W. Picard, "An Image Database Browser that Learns from User Interaction", MIT Media Laboratory and Modeling Group Technical Report, No. 365, 1996.

[15]D. A. Forsyth, J. Malik, M. M. Fleck, H. Greenspan, T. Leung, S. Belongie, C. Carson, C. Bregler, "Finding Pictures of Objects in Large Collections of Images," U. C. Berkeley, Dept. of EECS, CS Division, Technical Report, CSD-96-905, June 1996.

[16]S.-F. Chang, "Compressed-Domain Techniques for Image/Video Indexing and Manipulation", IEEE Intern. Conf. on Image Processing, ICIP 95, Special Session on Digital Image/Video Libraries and Video-on-demand, Oct. 1995, Washington DC.

[17]Daniel Drelinger and Adele E. Howe," Experiences with Selecting Search Engines Using Meta-Search," to appear in ACM Transactions of Information Systems, 1997.

[18]Mandis Beigi, Ana Benitez, and S.-F. Chang, "MetaSEEk: A Content-Based Meta Search Engine for Images," submitted to SPIE Conference on Storage and Retrieval for Image and Video Database, San Jose, Feb. 1997.
(demo and document: http://www.ctr.columbia.edu/metaseek)

[19] J. R. Smith and S.-F. Chang, "Searching for Images and Videos on the World-Wide Web," to appear in IEEE Multimedia Magazine, Summer, 1997.
(also Columbia U. CU/CTR Technical Report #459-96-25).
(demo http://www.ctr.columbia.edu/webseek)

[20]S. Weibel and E. Miller, Image Description on the Internet: A Summary of the CNI/OCLC Image Metadata on the Internet Workshop, September 24 - 25, 1996, Dublin, Ohio, D-Lib Magazine, January 1997.

[21]The Library of Congress, "Thesaurus for Graphic Materials I" http://lcweb.loc.gov/rr/print/tgm1.

[22]The MPEG Web Site, http://drogo.cselt.stet.it/mpeg/

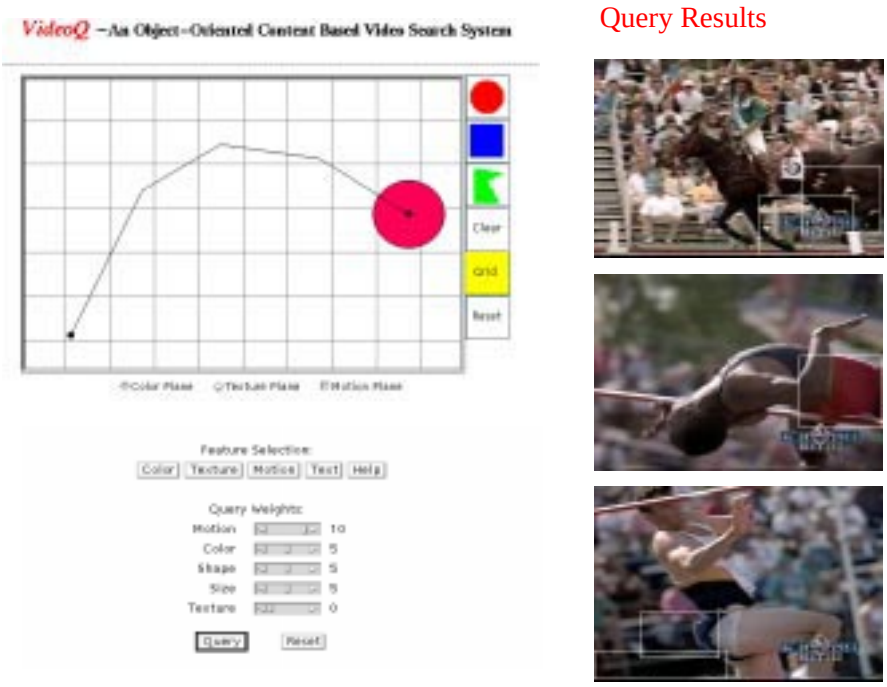
[23]Clifford A. Lynch, "The Z39.50 Information Retrieval Standard, Part I: A Strategic View of Its Past, Present and Future, " D-Lib Magazine, April 1997. (http://www.dlib.org/dlib)

[24]"FlashPix FAQ," http://www.kodak.com/daiHome/flashPix/flashPixFAQ.shtml.

[25]C. Frankel, M. Swain, and V. Athitsos, "WebSeer: An Image Search Engine for the World Wide Web", University of Chicago Department of Computer Science Technical Report TR-96-14, July 31, 1996.

[26]Excaliber System, http://www.excalib.com/rev2/products/vrw/vrw.html

(a)



(b)

FIGURE 1. Feature-based visual query. Query 1 uses the local color regions and spatial structures to find images of sunset or flower. Query 2 adds the motion feature to find relevant video clips, such as high jumps.

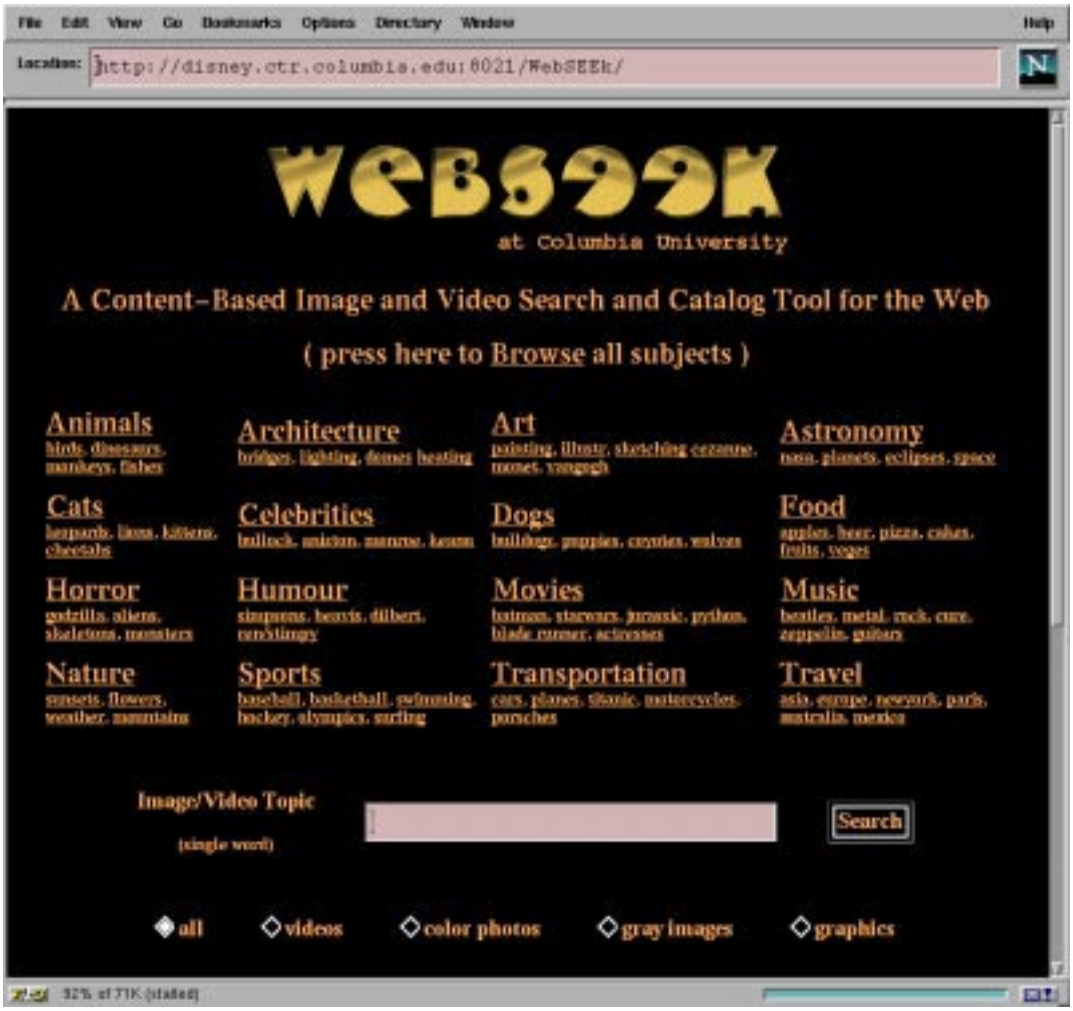


FIGURE 2. WebSEEk: A content-based image search and cataloging system on the Web.

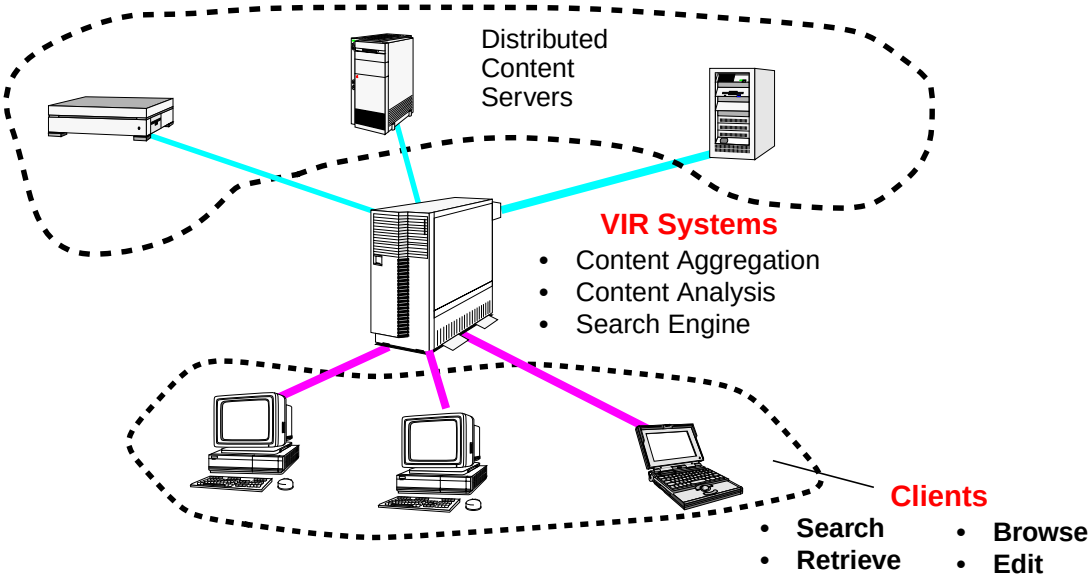


FIGURE 3. A High-Level System Architecture of Internet VIR

Non-descriptive terms		Descriptive terms		
term	count	term	count	mapping
image	86380	planet	1175	astronomy/planets
gif	28580	music	922	entertainment/music
icon	14798	aircraft	458	transportation/aircraft
pic	14035	travel	344	travel
img	14011	porsche	139	transportation/automobiles/porsches

FIGURE 4. Examples of textual terms and their mappings to the subject classes.

	Browse			**Query**			
	total operations	subject navigations	visual browsing	subject-based queries	text-based queries	basic content-based queries (color distribution)	Advanced content-based queries (VisualSEEk)
#	828,802	354,512	189,612	175,359	120,832	30,256	1,336
%	—	42.8%	22.9%	21.2%	14.6%	3.7%	0.2%

FIGURE 5. Use Statistics of an Internet VIR (WebSEEk)

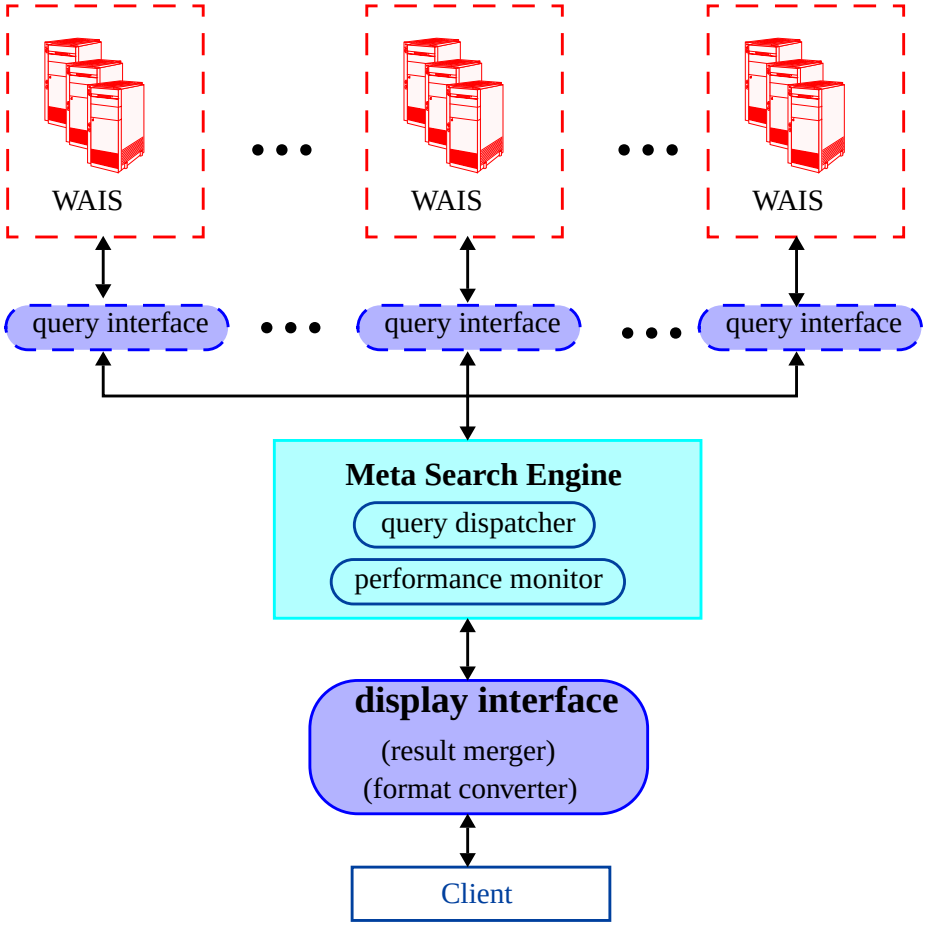


FIGURE 6. Basic components of meta search engines.

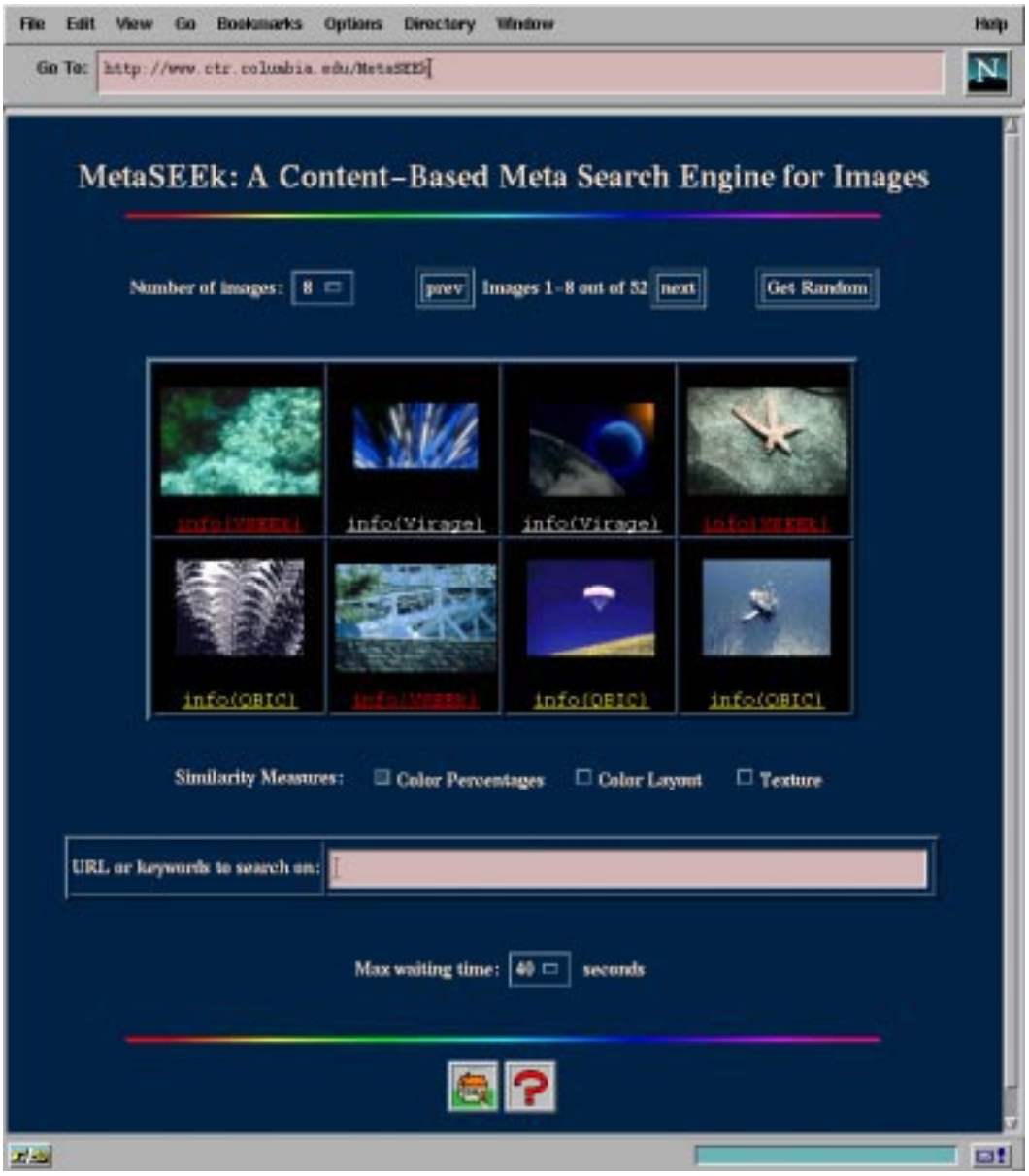


FIGURE 7. The query interface of MetaSEEk, which allows for browsing images from and issuing content-based queries to different target search engines. Image labeled in yellow are from QBIC, white from Virage, and red from VisualSEEk.

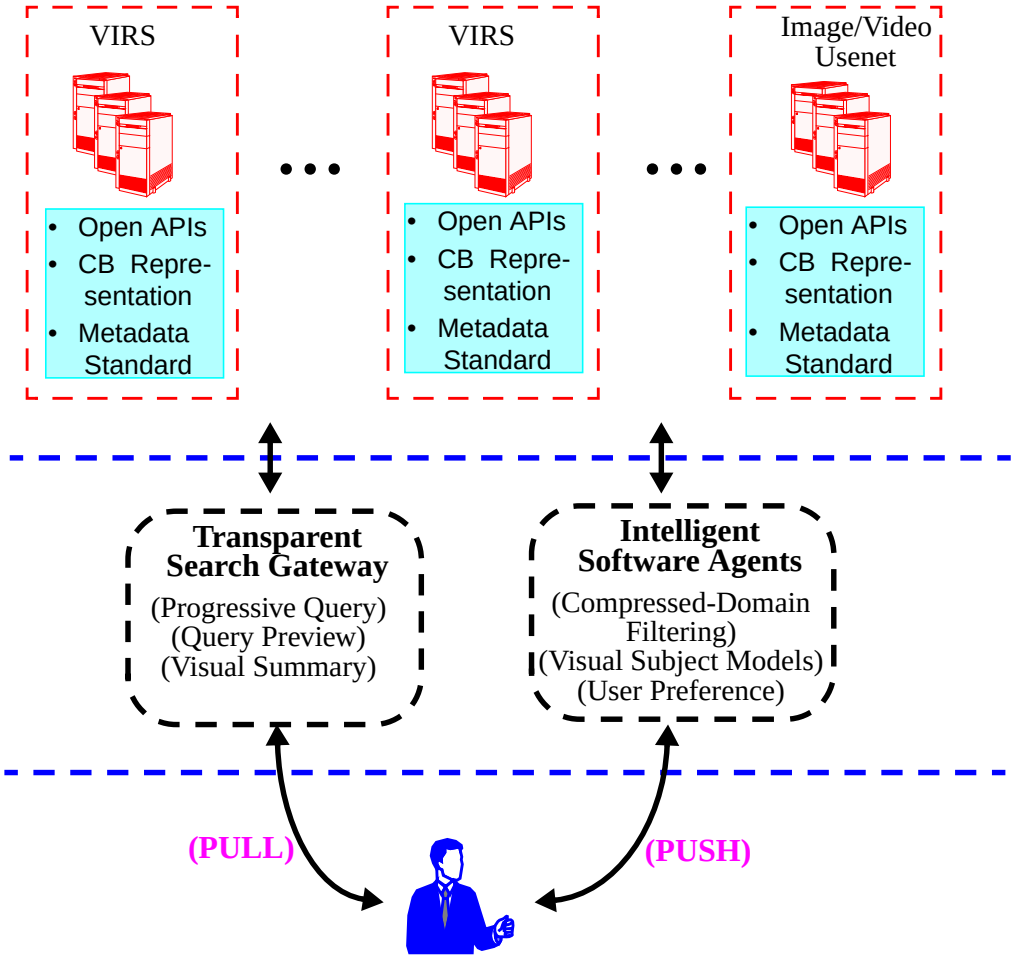


FIGURE 8. A high-level system architecture for distributed Internet VIRS. (CB: Content-Based, API: Application Programming Interface)