

Automatic Indexing and Content-Based Retrieval of Captioned Images

Rohini K. Srihari
State University of New York, Buffalo

The interaction of textual and photographic information in an integrated text/image database environment is being explored at the Center of Excellence for Document Analysis and Recognition (CEDAR), SUNY, Buffalo. Specifically, our research group has developed an automatic indexing system for captioned pictures of people; the indexing information and other textual information is subsequently used in a content-based image retrieval system. Our approach presents an alternative to traditional face identification systems; it goes beyond a superficial combination of existing text-based and image-based approaches to information retrieval. By understanding the caption accompanying a picture, we can extract information that is useful both for retrieving the picture and for identifying the faces shown.

In designing a pictorial database system, two major issues are (1) the amount and type of processing required when inserting new pictures into the database and (2) efficient retrieval schemes for query processing. Searching captions for keywords and names will not necessarily yield the correct information, since objects mentioned in the caption are not always in the picture and vice versa. Performing a visual search for objects of interest (that is, faces) at query time is computationally expensive, not to mention time-consuming. Thus, selective processing of the text and picture at data entry time is clearly required.

Our research has focused on developing a computational model for understanding pictures based on accompanying, descriptive text. "Understanding" a picture can be informally defined as the process of identifying relevant people and objects. Several current vision systems employ the idea of top-down control in picture understanding by providing the general context of the picture (for example, airport scene or typical suburban street scene). We carry the notion of top-down control one step further, exploiting not only general context but also picture-specific context.

THE PICTON APPROACH

To demonstrate the viability of this approach, we have implemented Piction,¹ a system that identifies human faces in newspaper photographs based on the information contained in the associated caption. Most newspaper photographs have factual, descriptive captions, which are necessary qualities for this task.

A new approach to face recognition integrates textual and photographic techniques. It extracts information from photo captions to identify human faces and provide content-based retrieval.

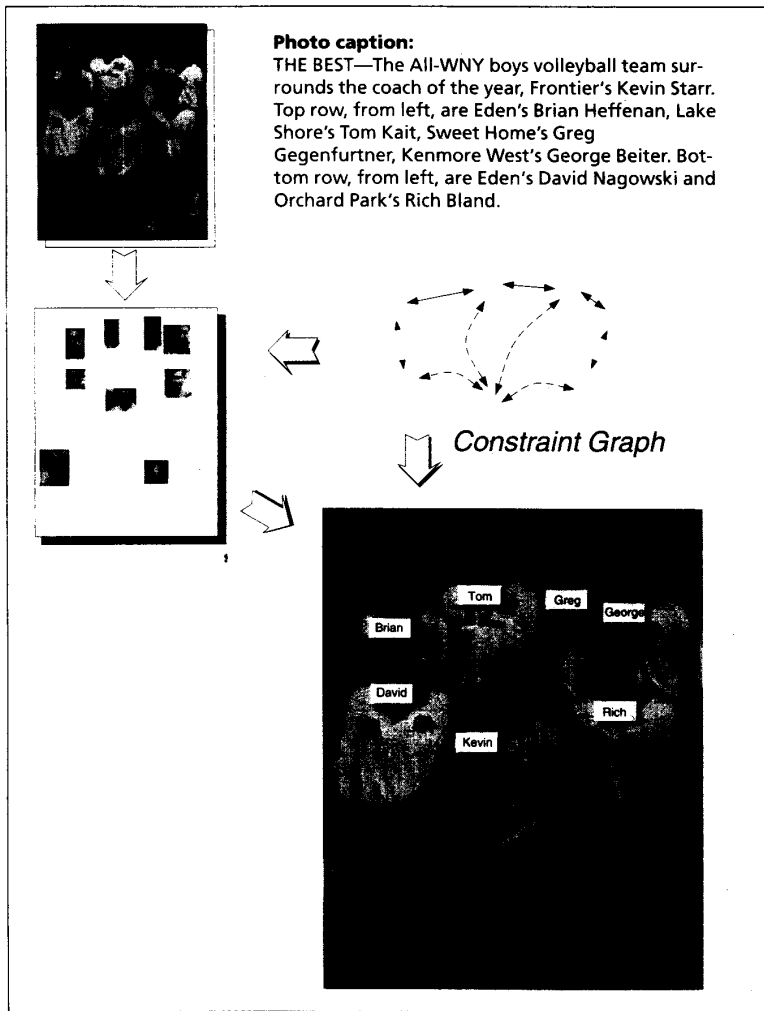


Figure 1. Working example from Piction using a captioned group photograph from the *Buffalo News*. Two false examples generated by a face locator were eliminated in the final output (lower right).

SPATIAL:
 surrounded-by(Kevin Starr, (Brian Heffernan, Tom Kait...))
 strict-left-of(Brian Heffernan, Tom Kait)
 below(David Nagowski, Brian Heffernan)

in-row(class: human; number: 4; names: (Brian Heffernan, Tom Kait...))
 in-row(class: human; number: 2; names: (David Nagowski, Rich Bland))

CHARACTERISTIC:
 gender(name: Brian Heffernan, male)
 gender(name: Tom Kait, male)

CONTEXTUAL:
 in-picture(name: Kevin Starr; class: human)
 in-picture(name: Brian Heffernan; class: human)

Figure 2. A subset of the constraints generated from the photo caption in Figure 1.

Traditional methods of face recognition employ model-matching techniques and thus require face models. Most face-recognition systems² use "mug shots" as input. These posed pictures, with their standardized location and homogeneous scale, facilitate detection of facial features. Our system does not require face models. It recognizes faces that have been automatically segmented out of an image, which is a much more difficult problem than model matching.

Current information-retrieval technology that is relevant to content-based retrieval of captioned photographs falls into two classes: image-based techniques and text-based techniques. Commonly used image-based techniques include color indexing and similarity-based retrieval. Text-based techniques have been investigated primarily in the context of document retrieval systems. Taken independently, existing techniques for text and image retrieval have several limitations. Image-based methods compute general similarity but are not designed for object identification. Text-based methods, while powerful in matching context, do not have access to image contents. By combining Piction's output with text similarity, we have obtained improved responses to focus-of-attention queries.

We have divided the problem of caption-aided face identification into two major subareas. The first area, discussed in the next section, deals with the processing of language input. The second major area is the design of an architecture (including a control paradigm) that exploits this information efficiently and that incorporates existing image-understanding technology.

VISUAL SEMANTICS

We have defined a new theory, called *visual semantics*,³ which describes a systematic method for extracting and representing useful information from text pertaining to an accompanying picture. This information is represented as a set of constraints.

Visual information in collateral text tells who or what is present in the accompanying scene and provides valuable information to locate and identify these people or objects. When combined with a

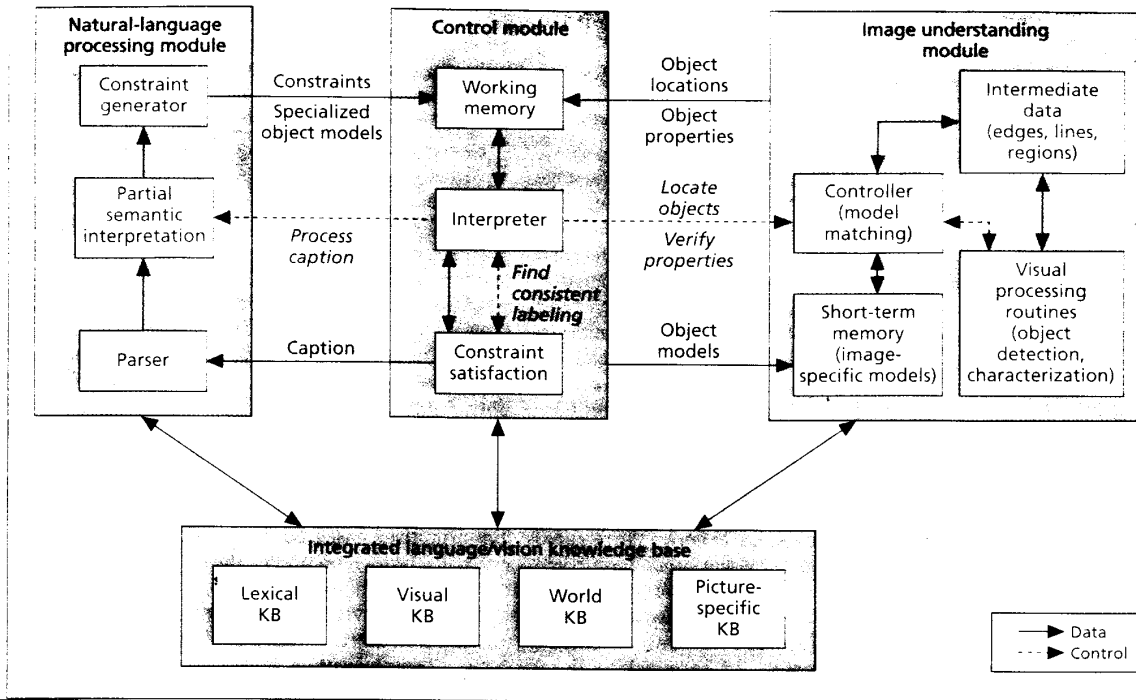


Figure 3. Computational model for collateral-based vision. The natural-language processing, image understanding, and control modules operate on a common, integrated language/vision knowledge base.

priori knowledge about the appearance of objects and the composition of typical scenes, visual information conveys the semantics of the associated scene. The resulting semantics provides the basis for top-down scene understanding.

Visual semantics leads to a set of constraints on the accompanying scene. The set is divided into three types of constraints: spatial, characteristic, and contextual. *Spatial constraints* are geometric or topological constraints, such as left-of, above, and inside. They can be binary or *n*-ary and can describe either relative, interobject relationships or absolute relationships relative to the image. Spatial constraints are used to identify and disambiguate objects of the same class (for example, people). They also express information about the location of objects in the picture. The information conveyed may be procedural. For example, if there is a chair in the corner, it results in the following high-level procedure construct: `loc(chair, region(corner(entire)))`. *Characteristic constraints*, which describe object properties, are unary in nature. Examples include gender and hair color. *Contextual constraints* describe the picture's setting and the objects that are expected—for example, the people present (explicitly mentioned in the caption), whether it is an indoor or outdoor scene, and general scene context (apartment, airport, and so on).

Figure 1 is an example of a digitized newspaper photograph and accompanying caption that Piction can handle. The task is to correctly identify each of the seven people mentioned in the caption. As Figure 1 shows, infor-

mation from the caption (in the form of a constraint graph) is used at an early stage, namely, when attempting to locate faces. Figure 2 illustrates the constraints derived from the caption. The number of faces and the fact that they are aligned in rows is used by a program that locates faces in the image. The row constraint is used on a second pass by the face locator, since the first pass failed to produce the required minimum number of candidates. By relaxing parameters in the appropriate areas, the additional face candidates were found. Identification of the faces revolved around the correct interpretation of the word "surround." We have interpreted it as an *n*-ary constraint involving minimum distances from a given candidate to the others. The simultaneous satisfaction of this and the remaining constraints resulted in the final, correct labeling.

ARCHITECTURE FOR COLLATERAL TEXT-BASED VISION

The architecture for collateral text-based image interpretation (Figure 3) consists of four main components: a natural-language processing module, an image understanding module, a control module, and an integrated language/vision knowledge base. All four components employ a common, intermediate knowledge representation scheme suitable for both natural-language and vision processing. Piction has been implemented in Loom,⁴ a high-level programming language and environment for constructing knowledge-based systems. Loom is based on the KL/1 family of semantic networks.

NLP module: Deriving constraints from text

The input to the natural-language processing (NLP) module is the original newspaper caption; the output is a set of constraints on the picture enabling the system to identify people. The NLP module has three phases: (1) syntactic parsing, (2) semantic processing, and (3) constraint generation.

The parser is implemented in two stages. The first stage eliminates directive phrases, such as "left" and "center," that do not affect the parse but are used later in constraint generation. The preprocessor also attempts to detect and classify proper-noun sequences by employing statistical techniques including part-of-speech tagging.⁵ The second stage employs a modified LR parser,⁶ which outputs a disambiguated parse tree. In this domain, typical problems in natural-language understanding, such as prepositional phrase attachment, are minimal.

The output of the semantic processing stage is a set of constraints on the picture. The constraint generator uses the various knowledge sources to derive the required semantics. Since visual information in language appears in a variety of syntactic and semantic constructs, it must be extracted systematically. For this reason, we have taken a goal-driven approach to visual semantics. The face labeling process requires that the NLP module correctly determine four types of information: object classes, who or what is in the photo, spatial constraints, and significant visual characteristics. A fifth type of information, determining contextual information, is optional.

OBJECT CLASSES. Proper-noun complexes (PNCs) are strings of proper nouns interspersed with common nouns, and they are ubiquitous in captions—for example, "Canadian Prime Minister Chretien." A primary task is to correctly detect and determine the object class for each of these PNCs. Examples of such classes include person, place, and organization. Mani et al.⁷ discuss a technique for classifying PNCs that looks up information in gazettes and uses syntactic and semantic context. But in Picition, we are primarily interested in determining whether PNCs correspond to people. In most cases, this determination can be made in the preprocessing stage.

As an example of how difficult this problem can become, consider this caption, "Winning Colors with her trainer Tom Smith prior to the start of the Kentucky Derby." Determining that Winning Colors is actually a racehorse and not a human involves understanding the meaning of the word "trainer" as well as contextual knowledge about the Kentucky Derby.

WHO/WHAT IS IN THE PICTURE. Once the PNCs corresponding to human names have been determined, predicting who is in the picture is based primarily on syntactic structure. Most of the concern is in rejecting those who do not appear in the picture; not all people mentioned in captions are in the pictures. For example, we would not expect to see President Clinton in a photo captioned "Tom Smith and his wife Mary

Jane prepare for the visit of President Clinton on Tuesday." In such cases, differences in event times are significant. Relative clauses introduced by words such as "before" and "after" are strong indicators of these situations. Phrases such as "the late" also provide valuable information.

SPATIAL CONSTRAINTS. Spatial constraints are the principal method for identifying people in captions. These constraints may be specified explicitly by terms such as "left," "right," and "top row." They may also be implicit—for example, when the order of mention in the caption reflects the order of appearance in the picture. Since caption writers are quite consistent, these situations can be identified.

Some terms, such as "standing" and "sitting," may not provide significant spatial constraints. However, in the domain of captions, "John Smith, standing" implies that John Smith's face appears above the others. Our system exploits this domain-specific interpretation of spatial terms. It can also process *n*-ary constraints such as "surround." Finally, its spatial reasoning component, which is part of the control module, quantitatively interprets spatial primitives based on the relationship between the bounding boxes representing the faces.

SIGNIFICANT VISUAL CHARACTERISTICS. It is necessary to determine any characteristic of an individual that can assist in face recognition. Determining gender is the most useful information in this respect. Captions for pictures containing both males and females do not usually provide spatial constraints to distinguish between them, since readers can do this easily. Our system needs to determine gender based on the caption and automatically classify facial images as male or female, thereby allowing face recognition. In text, it determines an individual's gender by either looking up the name in lists of male and female names or using kinship terms ("daughter of the late Sir John Smith") or pronouns ("Pinky Smith with *her* dog") appearing in the caption.

Other types of information useful in identifying people include distinguishing features such as hair color, beards, mustaches, and glasses. Such information may be gleaned through world knowledge (for well-known people) or be provided explicitly in the caption. Techniques for visual verification of gender and other such features are possible in cases where faces have been accurately located, are sufficiently large, and are of near-frontal orientation. Finally, captions sometimes identify individuals by their clothing ("wearing striped shirt") or the action they are involved in ("Tom Smith, holding the trophy"). Currently, our system cannot handle such situations, but they are being actively researched.

CONTEXTUAL INFORMATION. Captions often contain other information that could be used in object identification. For example, knowing whether the picture depicts predominantly an indoor or outdoor scene provides valuable low-level visual context. This information could be used in tuning edge detectors and segmentation algorithms, but we currently use it only for information retrieval. Similarly, if a complete labeling of the image was necessary, contextual knowledge would suggest the presence of other objects.

As an example of how difficult proper-noun classification can become, consider this caption, "Winning Colors with her trainer Tom Smith prior to the start of the Kentucky Derby."

LEXICAL RESOURCES. Piction employs several lexical resources in the natural-language processing phase. In particular, it effectively uses WordNet,⁸ a large-scale ontology of words, to determine the object class and other relevant visual characteristics. With WordNet, it can find meanings and synonyms for given words and access part-of and is-part-of hierarchies. We also employ the machine-readable version of *Longman's Dictionary of Contemporary English*. LDOCE provides syntactic information for parsing and "box codes," which are semantic categories for words that have been manually assigned. Finally, we use lists, obtained from public-domain sources, of proper nouns corresponding to male and female names, cities, countries, political leaders, and so on.

Control module: Exploiting constraints in image interpretation

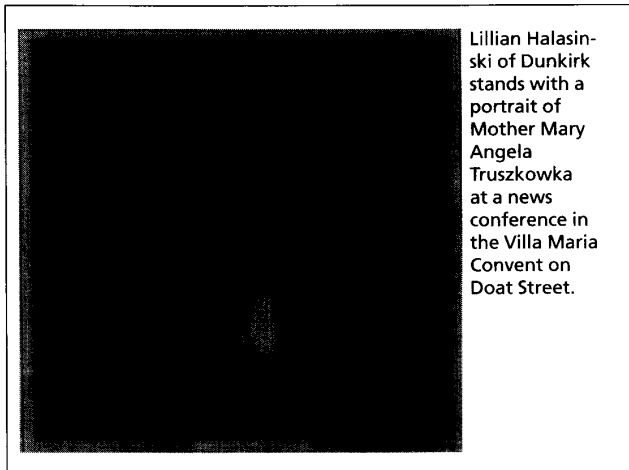
Traditional image-understanding systems employ mixed top-down/bottom-up control strategies (for example, blackboard systems). To detect low-level features such as edges and surfaces, they incorporate inexact graph matching, rule-based systems, and relaxation techniques. They also employ high-level relational model verification, using, for example, a model of a house or a typical neighborhood scene. Since exact image content is not known a priori, significant bottom-up computation is necessary before the appropriate high-level model can be invoked.

Strat and Fischler⁹ discuss the use of context in visual processing, but they focus on the exploitation of low-level collateral information (for example, lighting conditions). Our control strategy exploits (1) a confident hypothesis of the image contents and (2) all levels of contextual information to aid the visual processing. We have formulated control as a constraint satisfaction problem. This provides a single framework for incorporating spatial, characteristic, locational, and other general domain constraints without having to overspecify control information. However, the location of objects is still performed at the image understanding level, thereby facilitating integration of existing object detectors into the overall model. Some modifications to the traditional statement of constraint satisfaction problems were necessary; for example, to minimize computation time and effort, we modified the order in which constraints are verified.

For an example of how picture-specific constraints can enable a constrained search for objects, consider Figure 4. The word "portrait" implies a face that is enclosed by a boundary, typically a rectangle or an oval. Based on this, the NLP module generates as part of the set of constraints

contained-in-boundary(face-of(Mary Angela Truszkowka),portrait)

The object schema for portraits yields the information that it contains a frame/border and an interior and that the



Lillian Halasinski of Dunkirk stands with a portrait of Mother Mary Angela Truszkowka at a news conference in the Villa Maria Convent on Doat Street.

Figure 4. Constrained search. The word "portrait" in the caption for this Buffalo News photo implies a face enclosed by a boundary, and this information is included in the constraints generated by the natural-language processing model.

border is typically a rectangle or an oval. Since rectangles and ovals are easier to find than faces, the control module first calls for the detection of the frame. Having detected the frame, the face locator is called on to locate a face within the frame; this face is subsequently identified as Mary Angela Truszkowka. The search for the other face is conducted in the region outside the frame.

IU module: Face location and characterization

The image understanding (IU) module performs two basic functions: locating and segmenting objects and extracting visual properties. Currently, the only object class it handles is human faces.

The face-location module¹⁰ is an object locator that uses edge contours as basic features. It is supplied with the minimum number of faces to be found, based on caption information. The locator takes a holistic approach that uses a simple three-contour model to represent a face. These contours represent approximately the hairline and the left and right contours of the face.

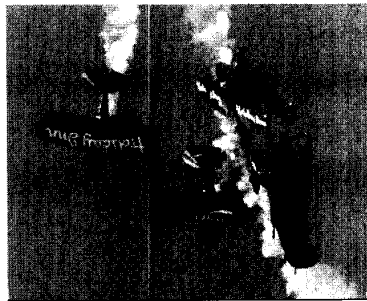
The face-location process begins with the application of a Marr-Hildreth edge operator. The zero crossings are pre-processed by thinning, spur removal, and other cleaning operations. The contours obtained are given weights corresponding to left, right, and top contours. Groups of three contours are formed, and a graph-matching technique generates possible face candidates. Since this technique does not rely on internal features such as eyes, it can detect faces in non-frontal orientations.

The chief problems with the current technique are (1) sensitivity to scale, (2) accuracy of location, and (3) generation of false candidates due to incidental alignment of edges. To handle the first problem, two members of

The chief problems with the current technique are sensitivity to scale, accuracy of location, and generation of false candidates due to incidental alignment of edges.



President Clinton, right, talks with Colin Powell, left, during a ceremony at the White House marking the return of soldiers from Somalia on May 4.



Four aircraft performing daredevil stunts on US Armed Forces Day open house. President Bill Clinton took part in the celebrations and gave away awards to the best cadets from the US military and armed forces.



President Bill Clinton, right, talks with Colin Powell, left, during a ceremony at the White House marking the return of soldiers from Somalia on May 4.



President Bill Clinton and Vice President Al Gore walk back to the White House after they welcomed back US troops returning from Somalia at the White House May 5.



President Bill Clinton gives a speech to a group of eleventh graders at Lincoln High School on his visit there April 2.



President Bill Clinton, center, responds to questions put forth by interrogators.

Figure 5. Results of combining text and image information to satisfy queries: Top two "hits" in response to "Find pictures of military personnel with Clinton" (top row); "Find pictures of Clinton with military personnel" (middle); "Find pictures of Clinton" (bottom). Photos courtesy of Associated Press.

our team, Mahesh Venkatraman and Venu Govindaraju,¹¹ developed a multiresolution approach to face location. Due to wide variations in intensities and the presence of shadows, the edges are of different scales (that is, the gradient of the change varies over a wide range). To capture such wide variations, the zero-crossings of a wavelet transform are used to obtain multiscale edges. Face loca-

tion then proceeds as described above. Hypotheses at different scales are projected back to the original image, and the location decision is made. This approach has proven to be more effective than detecting faces at a single resolution.

The problem of inaccurate locations, that is, bounding boxes that are too large, too small, or not properly centered, requires the incorporation of face-specific features. Currently, this is only possible for near-frontal orientations. The third problem, the rejection of false candidates, is handled by segmenting the face region and searching for face-specific features through a relaxation approach.

Since gender discrimination is critical to our caption-based approach to face identification, we have focused extensively on this problem. The two-class problem of male-female discrimination has been implemented using a neural network type architecture. It should be noted that gender discrimination performs well only in cases where the faces have been accurately located and the orientation is frontal.

Integrated language/vision knowledge base

An integrated language/vision knowledge base is essential for extracting visual information from text. The model calls for four types of knowledge bases: a lexical KB that models word syntax, semantics, and interconnections; a visual KB that contains object schemas (declarative and procedural modeling of an object's shape designed to facilitate object detection) along with a hierarchy of these schemas; a world KB that contains facts about people, places, events, and general domain constraints; and a picture-specific KB that contains facts specific to previously processed pictures and captions. The latter is used in information retrieval but is not necessary for processing a picture and caption.

An integrated knowledge base is necessary to solve

examples where people are identified by phrases such as "Tom Smith, wearing striped shirt." In such a case, the NLP module must derive visual semantics to the effect that the torso region of Tom Smith is covered with a striped texture. In general, the NLP module must be able to access and modify object schemas used by the image understanding module to fully convey the visual semantics.

EVALUATION AND FUTURE WORK

In evaluating Piction, we've had to consider many types of complexity. Examples become more difficult to process based on complexities in both the image and natural-language processing.

The system was originally tested on a data set of 50 pictures and captions obtained from the *Buffalo News* and the *New York Times*. We used three success codes to evaluate results. SU (success) indicated that the system correctly and uniquely identified everyone in the caption. PS (partial success) indicated multiple possibilities for one or more people where the actual face was included. E (error) indicated incorrect identification of one or more people (that is, true face not included). The overall success rate (SU only) was 65 percent. Although the sample size was too small to be considered statistically valid, the results illustrate the viability of this approach to face recognition. The most common reason for a PS or E was the failure of the face locator to find one or more of the identified faces. In only one case was the error due to incorrect parsing (that is, predicting the wrong number of people in the picture). Other reasons for a PS or E included the failure of spatial heuristics and an inability to properly characterize faces (for example, male/female or young/old).

The system is currently being tested on a larger database of several hundred pictures in digital form obtained from the Associated Press news wire service. We have been focusing on developing a more robust system at the cost of imposing more restrictions on the inputs. Restrictions have been imposed on the image with respect to minimum face size, near-frontal orientation, and general image quality. At the moment, images must be manually selected to meet these criteria. However, our system can use caption text to automatically reject most images where there is no clearly identifiable face (for example, nonface images and crowd scenes).

Using Piction for content-based retrieval

We have identified four distinct sources of information in computing the similarity between a query and a captioned image:

- text-based objective term similarity (exact match)
- text-based content term similarity (inexact match)
- image-based objective term similarity (exact match)
- image similarity (inexact match)

Text-based objective terms include keywords or other keys that have been assigned values manually. Examples include event type, location, and the general mood of the picture (happy, somber, serious, and so on). Chakravarthy¹² discusses methods of automatically assigning values to such keys. Although it is possible to derive values for some predefined keys, other robust methods of

measuring content-term similarity between a query and a captioned image should be considered. The availability of large-scale lexical resources such as machine-readable dictionaries and WordNet enable such methods. For example, for each content word w_q in the query, one could count the number of words in a caption with the same context by following pointers from w_q . Each pointer would represent a different type of relationship. The scores would be weighted by the distance (path length) from the original word. Other methods of capturing context include computing dictionary-definition overlap.

Any positive object/people identification made by Piction is represented in the database by the image coordinates. Similarly, any characteristic information that has been visually verified (for example, gender and hair color) is also noted. Image-based information useful in determining the presence of an individual can be quantified based on (1) whether the face was identified, (2) the size and orientation of the face, and (3) the method used to identify faces. The last measure of similarity concerns purely image-based techniques that have been discussed extensively in the image processing literature. Examples of such measures include texture similarity.

Based on the above measures of similarity, we can compute a combined similarity measure between a query and a captioned image as

$$\begin{aligned} \text{Sim}(\text{CapImage}, \text{Query}) = & \alpha \{ \text{text_based objective_term} \\ & \text{similarity} \} \\ & + \beta \{ \text{text_based content_based} \\ & \text{similarity} \} \\ & + \gamma \{ \text{image_based objective_term} \\ & \text{similarity} \} \\ & + \delta \{ \text{image similarity} \} \end{aligned}$$

We are in the process of empirically attempting the values for α , β , γ , and δ . Intuitively, we can see that higher emphasis should be placed on the exact-match components, especially the image-based exact-match component.

Dynamic satisfaction of emphasis in image retrieval

We have performed experiments where text and image information are dynamically combined to best satisfy a query. In such cases, users not only specify the context of the pictures they are seeking but also indicate whether the emphasis should be on image or text content.

The experiment involved three queries to a database of 140 digital images obtained from the Associated Press. Twenty of these images contained references to President Clinton in the caption. Figure 5 shows the top two "hits" on the three queries. In the first query, "Find pictures of military personnel with Clinton," the emphasis is on satisfying the context of "military personnel." The second query, "Find pictures of Clinton with military personnel," emphasizes Clinton. The final query, "Find pictures of Clinton," provides no context; presumably, the user was only seeking good pictures of Clinton.

Users not only specify the context of the pictures they are seeking but also indicate whether the emphasis should be on image or text content.

To measure how well the picture contents satisfy the query, we considered the following factors: (1) whether the required face was actually identified by Piction, (2) the size and orientation of the face, and (3) the centrality of the face in the image. The last factor was given a very low weight compared to the first two.

As the results show, the first query is weighted toward similarity of text context—the second hit does not even contain any people, let alone Clinton. The words “Armed Forces,” which are part of a larger title in the caption, caused a strong contextual match. We are attempting to refine our measures of context to overcome such problems. The second query results in pictures with Clinton for the most part; the picture with the airplanes is ranked very low. The last query produces the best pictures of Clinton without regard for context.

ALTHOUGH PICTON REPRESENTS only a preliminary foray into truly integrated text/image content-based retrieval, it illustrates the additional discriminatory capabilities obtained by combining the two sources of information. Much work remains, both in improving the language processing capabilities and in face location and characterization. We continue to work on the deeper research issues and on the development of a working model. ■

Acknowledgment


This work was supported in part by ARPA grant 93-F148900-000.

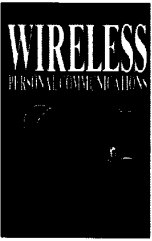
References

1. R.K. Srihari, “Use of Collateral Text in Understanding Photos,” *Artificial Intelligence Review*, special issue on integration of natural-language processing and vision, Vol. 8, No. 5-6, 1995, pp. 409-430.
2. R. Chellappa, C.L. Wilson, and S. Sirohey, “Human and Machine Recognition of Faces: A Survey,” *Proc. IEEE*, Vol. 83, No. 5, May 1995, pp. 705-740.
3. R.K. Srihari and D.T. Burhans, “Visual Semantics: Extracting Visual Information from Text Accompanying Pictures,” *Proc. AAAI 94*, American Association for Artificial Intelligence, Menlo Park, Calif., 1995, pp. 793-798.
4. *Loom Users Guide*, Version 1.4, Univ. of Southern Calif. Information Sciences Institute, Marina del Rey, Calif., 1991.
5. D. Cutting et al., “A Practical Part-of-Speech Tagger,” Tech. Report, Xerox Palo Alto Research Center, 1993.
6. M. Tomita, “An Efficient Augmented-Context-Free Parsing Algorithm,” *Computational Linguistics*, Vol. 13, No. 1-2, 1987, pp. 31-46.
7. I. Mani et al., “Identifying Unknown Proper Names in Newswire Text,” *Proc. Workshop on Acquisition of Lexical Knowledge from Text*, Assoc. for Computational Linguistics, Somerset, N.J., 1993, pp. 44-54.
8. R. Beckwith et al., “WordNet: A Lexical Database Organized on Psycholinguistic Principles,” in *Lexicons: Using On-Line Resources to Build a Lexicon*, Lawrence Erlbaum, Hillsdale, N.J., 1991, pp. 211-232.
9. T.M. Strat and M.A. Fischler, “Context-Based Vision: Recognizing Objects Using Information from Both 2D and 3D Imagery,” *IEEE Trans. PAMI*, Vol. 13, No. 10, Oct. 1991, pp. 1,050-1,065.
10. V. Govindaraju, S.N. Srihari, and D.B. Sher, “A Computational Model for Face Location Based on Cognitive Principles,” *Proc. AAAI 92*, American Association for Artificial Intelligence, Menlo Park, Calif., 1992, pp. 350-355.
11. M. Venkatraman and V. Govindaraju, “Zero-Crossings of a Nonorthogonal Wavelet Transform for Complex Object Location,” *Proc. Int’l Conf. on Image Processing*, IEEE Signal Processing Society, to be published in Oct. 1995.
12. A. Chakravarthy, “Representing Information Need with Semantic Relations,” *Proc. 15th Annual Conf. on Computational Linguistics (COLING 94)*, Assoc. for Computational Linguistics, Somerset, N.J., 1994.

Rohini K. Srihari is a research scientist at the Center of Excellence for Document Analysis and Recognition (CEDAR) and a research assistant professor of computer science at the State University of New York, Buffalo. Her research centers on using linguistic information to interpret visual data. She is the principal investigator on two projects: an NSF/ARPA project on language models for recognizing handwritten text and a DoD/ARPA project on the use of collateral text in understanding photos in documents. Srihari received a B.Math degree in computer science from the University of Waterloo, Canada, and a PhD in computer science from SUNY, Buffalo.

Readers can contact the author at CEDAR, UB Commons, 520 Lee Entrance—Suite 202, State University of New York, Buffalo, NY 14228-2567; rohini@cedar.buffalo.edu.





Wireless Personal Communications: The Future of Talk


by Ron Schneiderman

Provides you with a fascinating account of how the development of new wireless communications technologies, including cellular phones and mobile satellite services, will completely alter the way we communicate. The book presents a complete overview of the wireless market, including emerging technologies, market projections, competitive analysis, and overseas opportunities. In addition, it includes a directory of over 150 companies, trade associations, and regulatory agencies who are involved in wireless activities. From this book you will gain a better understanding of the dynamics of this fast-emerging new electronics market.

*208 pages. 1994. Hardcover. ISBN 0-7803-1010-1.
Catalog # RS00004 — \$24.95 Members / \$29.95 List*

Call toll-free:
+1-800-CS-BOOKS

Order by FAX:
+1-714-821-4881



**IEEE
COMPUTER
SOCIETY**