データベース・システム 77－4
（1990．5．17）

マルチメディア情報環境

--「電子美術館」の経験から --

加藤　俊一　　栗田　多喜夫
（電子技術総合研究所）
水鳥　哲也
（明電舎）

概　要： 本稿では、システムがマルチメディア情報を意味的に解釈して利用者と対話できるような、マルチメディア情報環境の構成法を紹介する。マルチメディア対話の実際例として、利用者の描くスケッチをキーとして内容検索する QVE (Query by Visual Example) および、主観的な言葉を手掛かりに最も相応しい画像を検索する QBD (Query by Subjective Description) での処理方式を示す。

このような情報環境を実現するためには、画像処理的な視点から画像をモデル化すると同時に、認知的な視点から利用者をモデル化することが重要である。本稿では、これらを統合するための意味的データモデルの枠組みについても考察する。

A Cognitive Approach to Visual Interaction

- Electronic Art Gallery -

Toshikazu KATO and Takio KURITA
Electrotechnical Laboratory
1-1-4, Umezono, Tsukuba Science City
305, Japan

Tetsuya MIZUTORI
Meidansha Corp.
2-1-17, Oosaki, Shinagawa
Tokyo 141, Japan

Abstract

This paper describes the ideas of visual interaction and their implementations on an image database system. The visual interaction includes a query by visual example (QVE) and a query by subjective descriptions (QBD). The former provides a sketch retrieval function and the latter provides a sense retrieval function. This paper shows the detail algorithms of these functions developed on our experimental database system ART MUSEUM. These functions use a pictorial index created by image analysis and a personal index automatically learned as the user model. They formed visual interaction in a truly user-friendly manner. This paper also summarizes the requirements to a multimedia data model from a cognitive aspect.

1. Introduction

A visual interface plays an important role in multimedia information systems. We need a visual interface to enable flexible man-machine communication in the user-friendly manner. Especially we request the visual interface for image database systems to communicate the visual information to and from the system [Iyen 88, Gros 89].

Modern engineering workstations provide visual interfaces using icons, menus and a mouse. Window systems also provide us with several drawing tools. In spite of visual presentations, the computer system receives only alphanumeric data as the result of an operator action, such as typing commands or selecting options from a window system device. Even though they provide a user-friendly interface, these systems do not interpret the visual information itself such as the meaning or similarity of the images. Therefore, the window systems do not provide enough facilities for visual interaction with a database.

What are needed in visual interaction? We can summarize the essential needs in visual interaction as follows [Kato 89a, 89b].

(a) We want to process the visual information itself within the database.

(b) We need to communicate this information to and from the database in a user-friendly manner.

We can answer these needs taking a cognitive approach. The cognitive approach integrates both image model and user model on a semantic data model to interpret and operate the visual information semantically. The image model describes the physical constraints of image data, while the user model reflects the the visual perception process of the user. The semantic data model gives the framework for describing the relationships.

This paper describes the ideas of visual interaction from our cognitive aspect in Chapter 2. We will show our approach by typical user's query requests. We will show our approach by typical user's query requests. A query by visual example (QVE) provides a sketch retrieval function and a query by subjective descriptions (QBD) provides a sense retrieval function. This chapter also introduces our experimental database system ART MUSEUM[1]. The algorithm for QVE are described in Chapter 3. The system automatically creates the pictorial index for sketch retrieval. This index describes the general composition of paintings. Chapter 4 describes the detail algorithm for QBD. We have designed this algorithm to refer the user model in sense retrieval. This model describes the correlation between the color feature of paintings and the subjective impression of each user. This chapter also includes the learning algorithm to build the user model. Chapter 5 discusses the requirements for the multimedia data model from our experience.

2. Intelligent Visual Database System

2.1 Visual Interface

"A picture is worth a thousand words" Visual information is a good man-machine communication media. We will regard visual information as not only image data itself but also linguistic data related to some image data. For example, a photograph and a hand-written sketch belong to the former category. Subjective remarks on an artistic painting belong to the latter category. We can communicate the visual information to and from the system in the visual interaction process.

Several experimental image database systems provide visual interfaces. The QPE system provides the schema of graphic data in a tabular form [Chang 80]. The image browser navigates a user tracing its hypermedia-like indexes [Kasa 89]. In the icon-based system, an icon is the key of an image as well as the element of the visual query language [Chang 88]. Even for the alphanumeric database systems, we find several graphic languages to manage the database schema and to retrieve the data [Bryce 86, Czejdo 90]. While these systems use graphic devices to show schema, icons and guidelines, their queries are only substitutes for the query languages on alphanumeric data. These interfaces do not interpret the semantics of the visual information which we want to process.

We wish to process the semantics of visual information and to communicate

1 ART MUSEUM: Multimedia Database with Sense of Color and Construction upon the Matter of Art.

the visual information to and from the database in a user-friendly manner. We expect image database systems to have good visual interfaces to enable such visual interaction.

2.2 Visual Interaction Facilities

Then, how do we organize such a visual interface? We will show our approach by typical user's query requests.

A user often wishes to see some painting which he keeps in his mind. Then, he has only to draw its rough sketch and show the sketch as a pictorial key to the system. We call this kind of visual interaction a query by visual example (QVE). QVE should evaluate the similarity between the sketch and the images in the database. Therefore, we have to define a similarity measure on a robust image model [Kato 89d, Kurita 90].

Our second request to visual interaction is a query by subjective descriptions (QBD). A user often wishes to see several paintings which may leave him some artistic impressions. Then, he has only to describe such impressions by his own words to see several paintings which may leave him similar impressions. This kind of visual interaction is sense retrieval. Note that our impressions differ for each of us, even when viewing the same painting. Therefore, the system should analyze and learn the correlation between the subjective descriptions and the images with each user. Such correlation forms a user model. We have to develop a simple learning algorithm to adjust the model to each user.

Thus, we have to design a robust image model and a user model. These are the requirements and the technical problems associated with visual interface and visual interaction.

2.3 Cognitive Approach

Let us discuss the conventional approaches to image database systems with relation to the intelligent visual interaction. We may classify them into the image model approach in image processing and the semantic data model approach in database designing.

In the image model approach, the major topics have been physical data structures of spatial data for data compression and high-speed data access. Such database systems assume some image models only to design suited data and file storage structures in this aspect [Ahuja 81]. Pattern recognition facilities also assume some image models for pattern classification. They are only referred as the fixed templates, while image models describe the physical constraints of image data.

On the other hand in the semantic data model approach, we know several data models, such as ER model, to design the database schema. These data models describe the relationships among the entities and their attributes of images. An object-oriented model is an extended semantic data model suited to multimedia database systems [Masu 89]. Such data models are also useful to describe the knowledge about the real world [Gros 89]. Since semantic data models gives only a formal framework of knowledge representation, they do not concern the contents and the evaluation of image data.

Recently, the cognitive aspect was pointed out with man-machine system designing [Norm 86] and with database schema designing [Neu 89]. While they denoted the user model, their ideas cover only the operations for relationships on the semantic data model. The interpretation mechanism and the integration method were out of range.

Visual interaction requires interpreting the contents of image data to operate them semantically. Interpretation algorithms have to suite the visual perception process of each user. Such process belongs to a subjective human factor. Thus, our requirements on visual interaction suggests another approach to image database systems. In this context, we will call it cognitive approach. In the cognitive approach, we have to combine the image model and the user model on the semantic data model to support visual interaction.

2.4 ART MUSEUM System

We have been developing an electronic art gallery called ART MUSEUM [Kato 89a, 89b, 89c]. The ART MUSEUM is a collection of full color paintings of landscapes. The ART MUSEUM system provides both types of visual interaction. QVE is available for sketch retrieval using pictorial keys. QBD is also available for sense retrieval using some words from the user's personal taste. Figure 1 shows the overview of the ART MUSEUM system. This system has two types of special indexes for visual interaction. One is a pictorial index and the other is a personal index.

We have assumed following image models to design the pictorial indexes. The general composition and the coloring characterize each full color painting.

A contour image roughly approximates the general composition of the color painting. Applying a differential filter to the color painting, we get its contour image. A user has only to draw a sketch of the overall picture as a pictorial key in QVE. The system searches for a similar composition on the pictorial index, comparing the parameterized graphic feature of contour images. Then, the system finds the target picture and presents it in a full color representation.

An artist paints with many colors. We can parameterize the coloring by the distribution of the RGB intensity values. The system refers this pictorial index to create the following user model and personal index.

We have also assumed a user model to design the personal index. The subjective descriptions of each user correlate with the coloring of paintings. The system analyzes such correlation between the set of words and the parameterized coloring feature on the several test samples. The system uses the correlation as the personal index for the user. The user has only to show several words in QBD. The system evaluates the most suited coloring to the words on the personal index. Then, the system provides the paintings of suited coloring.

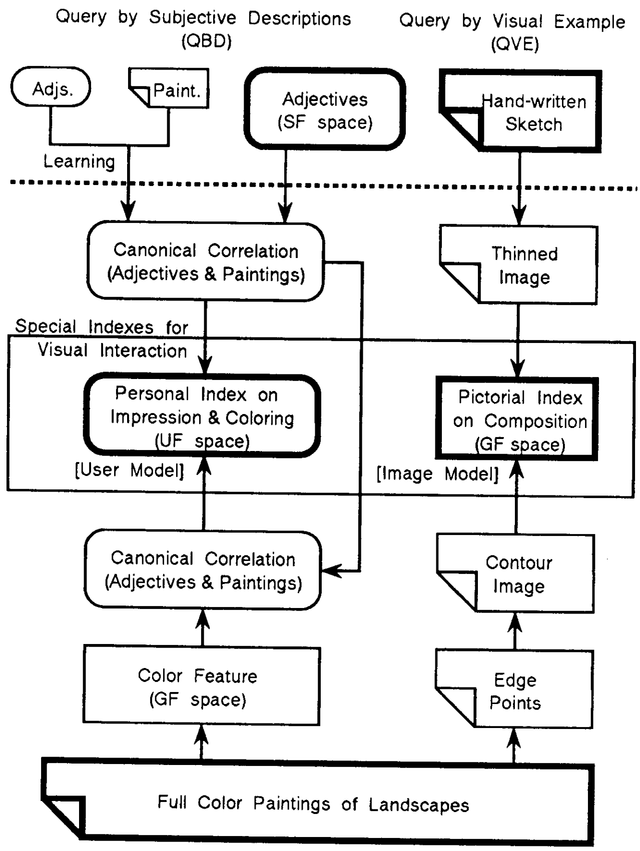

Figure 1. Overview of the ART MUSEUM System and Visual Interaction

3. Query by Visual Example

3.1 Pictorial Index

The general composition of painting is one of the major part in our impression. This chapter discusses an approach to the visual interaction, which is query by visual example (QVE). A user has only to draw a sketch of the overall paintings to retrieve several paintings of similar composition.

A contour image roughly approximates the composition of the original color painting. The ART MUSEUM system has a pictorial index of contour images for QVE. We have developed the following differential filter on RGB space to get a contour image.

[Construction of pictorial index]

(1) Apply an affine transformation to normalize the image size into 256 × 256 pixels. Then, apply a median filter of 3 × 3 pixels window to smooth the RGB intensity values.

(2) Apply the following nonlinear differential filter to evaluate the edge rate.

$\Delta S = \Delta I \,/\, I$.

Where for each pixel m_{xy} and for each of its 8-nearest neighbors $m_{x'y'}$.

$\Delta I = m_{xy} - m_{x'y'}$,

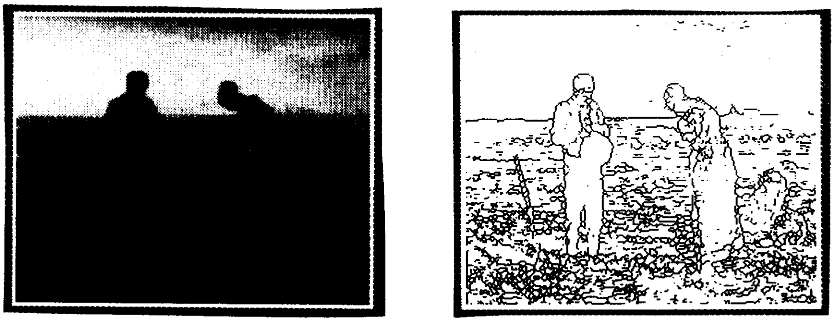

Figure 2. Color Image and its Contour Image

$I = m_{xy} + m_{x'y'}$.

(3) Choose higher edge rate points from the distribution of ΔS to get a binary contour image.

(4) Apply a thinning algorithm to the contour image.

(5) Calculate the number and the rate of edge directions p_i. (Currently, the edge directions are digitized into 36 patterns).

Figure 2 shows the original color images and their contour images derived from this algorithm. These contour images approximate the general composition of the original color paintings. Our differential filter satisfies Weber-Fechner's Law in visual psychology [Mac 86]. We will refer to the number and the rate of edge directions as the pictorial index of graphic feature (GF) vector on general composition.

3.2 Sketch Retrieval

Let us show the sketch retrieval algorithm in QVE. Figure 3 shows the outline of this algorithm. The user has only to draw a sketch of the overall pictures. The sketch is thinned and referred to as a pictorial key. The system searches for a similar composition on the pictorial index. Then, the system provides the target picture with a full color representation.

[Sketch retrieval]

(1) Apply an affine transformation to normalize the image size into 256 × 256 pixels. Then, apply thinning algorithm to the image, and calculate its GF vector p_0 of the number and the rate of edge directions.

(2) Choose the suited paintings whose GF vector p_i is near to the user's sketch.

In the current stage of our research, we have assumed that the user can write down the detailed sketch. The de-

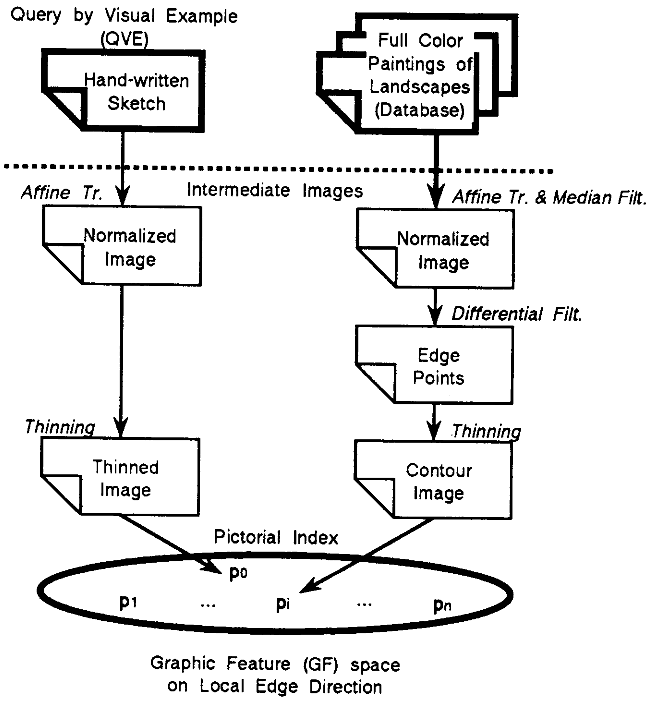

Figure 3. Sketch Retrieval in QVE

velopment of a powerful matching algorithm for rough sketches is the next stage of our research.

4. Query by Subjective Descriptions

4.1 Visual Impression and Color Feature

To retrieve some paintings which give us some impressions, key word indexes have been the popular approach. The indexer assigns several key words to each image as the index terms. The user retrieves some images by the preassigned key words. Even if the key word thesaurus is available, this approach has following two problems:.

(a) The indexer has to assign some key words to every painting in the database, which costs much personnel expense.

(b) Such descriptions may differ for each person according to his cultural backgrounds, even when looking at the same painting.

Therefore, we have to model how each user feels as the user model. Art critics views paintings from several aspects. They take notice on motif, general composition and coloring. [Chiji 83] reported that the dominant impression generated by paintings is coloring. From this aspect, we may expect there is a reasonable correlation between the coloring and the words in the reviews.

The ART MUSEUM system learns the correlation between such coloring and the subjective descriptions of the user's impressions. We may refer such correlation as the user model. In our current implementation, the artistic impressions are described by special 30 adjectives.

4.2 User Model and Personal Index

Let us show the algorithm for learning a user model. We cannot directly compare the words of subjective descriptions and the coloring of paintings, since they have different domains. Therefore, we have parameterized the words and the coloring into the subjective feature (SF) vector and the graphic feature (GF) vector at first. Here, the SF vector shows the weight of the adjectives for the user's impressions. The GF vector shows the distribution of the RGB intensity values. The SF vectors and the GF vectors make the SF space and the GF space, respectively. We adopted several test samples from the database to analyze the correlation between the SF space and the GF space.

Figure 4 shows the outline of the learning algorithm. A user describes his impressions by several adjectives with each painting in the test sample, which gives SF vectors. The system extracts the GF vector of each painting. We need a unified feature (UF) space where the correlation between SF space and GF space is maximum. We can construct such a UF space by the canonical correlation analysis. The canonical correlation analysis is one of the multivariate analysis methods to correlate the distinct domains [Cooly 71]. The algorithm to construct the UF space and the personal index is as follows. (See also Figure 4(a)).

[Construction of the UF space and personal index]

(1) Sample arbitrary paintings from the database as a learning set P. The user gives his impressions by $\boldsymbol{a}_k$ to each painting $k \in P$. (The SF vector $\boldsymbol{a}_k$ shows the weights of adjectives).

(2) Apply an affine transformation to normalize the image size into 256 × 256 pixels. Then, apply a median filter of 3 × 3 pixels window to smooth the RGB intensity values [Rosen 82]. Calculate the GF vector $\boldsymbol{p}_k$ of each painting $k \in P$. (The GF vector $\boldsymbol{p}_k$ is a histogram of RGB intensity values).

(3) Construct the linear mappings F and G to correlate $\boldsymbol{f}_k$ and $\boldsymbol{g}_k$, applying the canonical correlation analysis.

$$\boldsymbol{f}_k = F'\boldsymbol{a}_k ,$$

$$\boldsymbol{g}_k = G'\boldsymbol{p}_k .$$

(F' means the transposed matrix of F)

(4) Calculate the UF vectors with every painting in the database by $\boldsymbol{g}_i = G'\boldsymbol{p}_i$.

We will refer the UF space of g_i as the personal index. Note that we do not have to assign the adjectives a_n to every painting in the database. Once the system has learned the linear mappings F and G, it can automatically construct the personal index only from the GF vectors. This algorithm reduces the personnel expenses for indexing.

If we develop this image database system on the network, we can allocate the personal index on the local site for personal use. To the contrary, we had better place the image data and the pictorial index of coloring on the center site for common use.

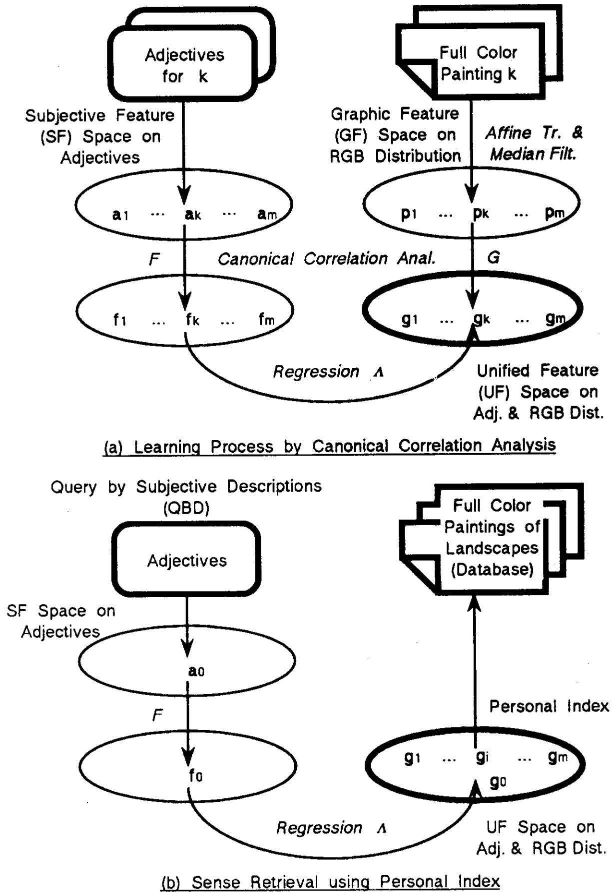

Figure 4. Sense Retrieval in QBD

4.3 Sense Retrieval

We may expect that the neighboring paintings on the personal index cause similar impressions with the user. (Of course, the test samples P must reflect the properties of the whole database). The user describes the target paintings he wants to appreciate by several adjectives in sense retrieval. The system infers some coloring which makes him such feeling on the personal index. The system then shows suited candidates in the database. The algorithm for sense retrieval on coloring is as follows. (See also Figure 4(b))

[Sense retrieval]

(1) Apply the linear mappings F and Λ to the vector a_0 of denoted adjectives in the user's query.

$$g_0 = \Lambda F' a_0.$$

Where Λ is the diagonal matrix of canonical correlation coefficients.

(2) Choose the neighboring paintings g_i on the personal index as candidates in sense retrieval.

The sense retrieval algorithm binds two attributes in different media. The one is the key words in linguistic media, and the other is the paintings in image media. Therefore, the sense retrieval algorithm seems to be a multimedia join operation between the linguistic media and the image media.

We have experimented the learning algorithm and the sense retrieval algorithm on our ART MUSEUM system. In this experiment, we used 50 paintings for the samples, while the whole database has 100 paintings. Figure 5 shows an example of sense retrieval. Figure 5 shows the first sixth candidates for the adjectives; refined, cool and bright.

In this experiment, we have adjusted the user model to five young girls. Their subjective estimation was almost good. Therefore, we may expect that the personal index is useful for the sense retrieval on image databases.

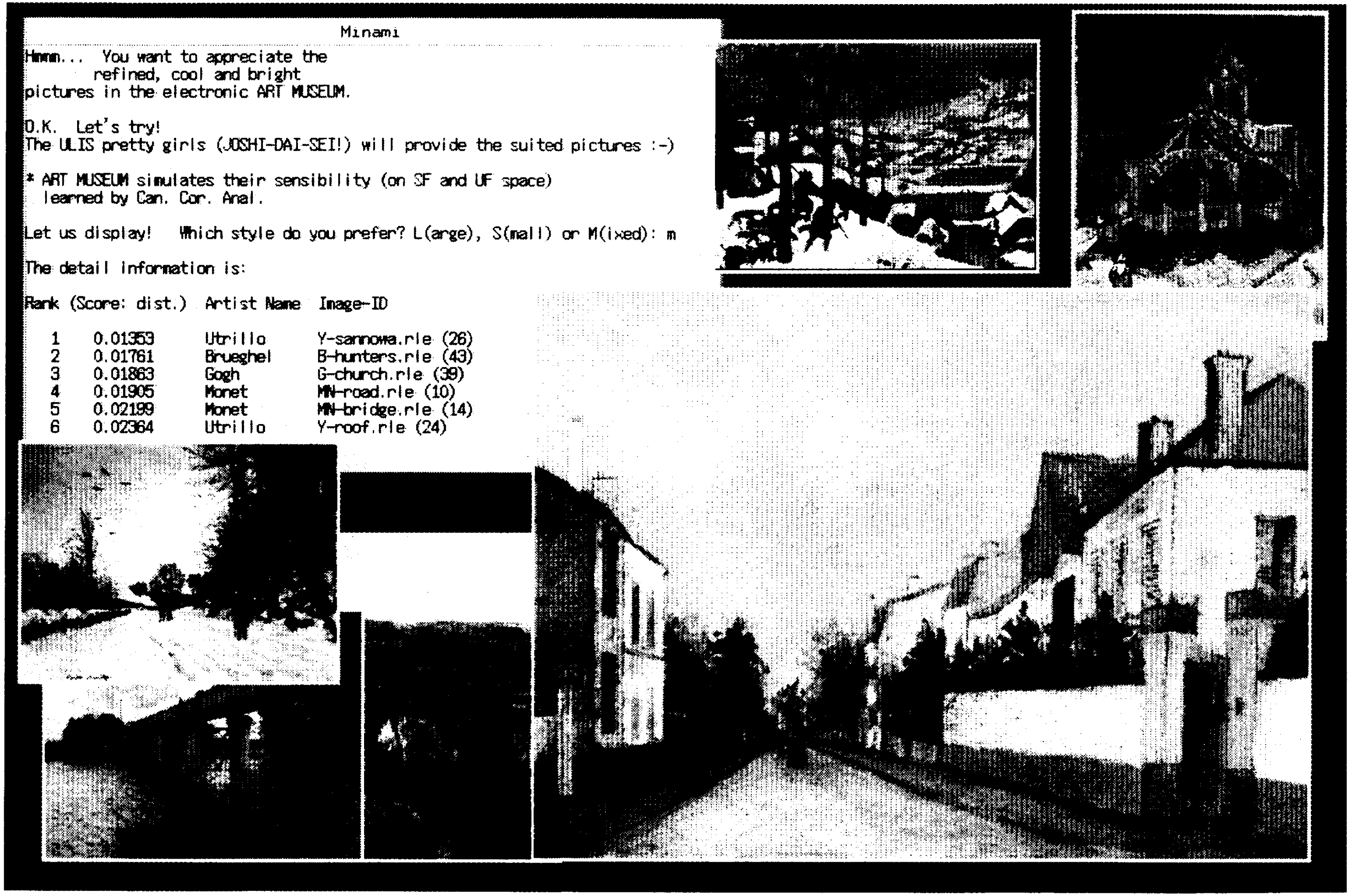

Figure 5. Sense Retriieval in QBD

Using the UF space, we can also retrieve several paintings which cause the similar impression on coloring by showing a painting. We can also infer suited key words for a painting, using the inverse mapping F^{-1} and Λ^{-1} as follows.

$$a_0 = F'^{-1} \Lambda^{-1} G' p_0 \quad .$$

5. Multimedia Data Model

5.1 Cognitive Aspect

Let us summarize the requirements for a multimedia data model from our cognitive aspect [Kato 89e].

(a) Spatial independence:. We want to provide the query by subjective descriptions in the user's view. User's view mechanism should manage the subjective indexes. Of course, the system has to manage also the pictorial indexes to process the spatial relationships and graphic features of images.

(b) Media independence:. We use many kinds of image data types suited to image compression and image processing. For example, a painting is stored as a matrix of RGB intensity values, a matrix of binary values, a series of run-length codes, and so on. We expect these data structures are transparent to the image processing algorithms. The abstracted data type is one of its approach.

(c) Application independence:. We developed many image processing algorithms. These algorithms depend on their data type of domain and range, which makes the control mechanism complicated. Therefore, we have to abstract these algorithms by their classifications according to their functions. This idea is method abstraction, which is useful to manage the relationships among the image entity, its data type and image processing algorithms. This mechanism provides the flexible wake-up of image processing algorithms in object-oriented environment.

5.2 Framework of CHEMIN

In the rest of this chapter, we will show the framework of our multimedia data model CHEMIN[2] [Kato 89e].

The object-oriented models regard an atomic object as a frame composed of data item and related methods. In the visual interaction, we can define the semantics as explicit relationship among object hierarchy, e.g. is_a, and as implicit relationship by some methods, e.g. similarity operation. Therefore, we need a new knowledge representation form to abstract the relationships.

The CHEMIN model describes the micro-structure of the objects. This model classifies the information related to the image data into the following three categories. They are entities, media and methods. The entities conceptually describe the real world by class hierarchy of image objects. The media describe the class hierarchy of image data types. The methods classifies the algorithms for image operations and database operations. These relationships are currently described in first order predicates. This is one of the ways to abstract not only the data types but also the semantics and image operations. Therefore, this is an extensible framework for both database system and expert vision system.

6. Summary

We have described the ideas and their implementations of visual interaction with an image database system. The ART MUSEUM system provides sketch retrieval in query by visual example and sense retrieval in query by subjective description. These functions formed visual interaction in a truly user-friendly manner, based on the use of the image model and the user model.

This paper also summarized the requirements to a multimedia data model from a cognitive aspect. Multimedia database systems require method abstraction and media transparency as well as data abstraction.

Acknowledgments

We would like to thank the colleagues in Electrotechnical Laboratory (ETL) and the students from University for Library and Information Science (ULIS). This research is supported by the national research and development project of *Interoperable Database Systems*.

References

[Ahuja 81] Ahuja, N. and Schachter, B. J.: "Image Models", ACM Computing Surveys, Vol.13, No.4, pp.373-397, Dec. 1981.

[Bryce 86] Bryce, D. and Hull, R.: "SNAP: A Graphic-based Schema Manager", Proc. of Int. Conf. on

[2] CHEMIN: Composition of Hyperobjects on Entity, Media, Method and Instructions.

Data Engineering, pp.151-164, Feb. 1986.

[Chang 80] Chang, N. S. and Fu, K. S.: "Query-by-Pictorial Example", IEEE Trans. on Software Engineering, Vol.SE-6, No.6, pp.519-524, June 1980.

[Chang 86] Chang, S. K.: "Image Database Systems", in "Hand Book of Pattern Recognition and Image Processing", Young, T. Y. and FU, K. S. (ed.), Academic Press, Chap.16, pp.371-393, 1986.

[Chang 88] Chang, S. K., Yan, C. W., Dimitroff, D. C. and Arndt, T.: "An Intelligent Image Database System", IEEE Trans. on Software Engineering, Vol.SE-14, No.5, pp.681-688, May 1988.

[Chiji 83] Chijiiwa, H.: "Chromatics", Fukumura Printing Co., Chap.5, pp.128-163, 1983 (in Japanese).

[Cooly 71] Cooly, W. W. and Lohnes, P. R.: "Multivariate Data Analysis", John Wiley & Sons, Inc., 1971.

[Czejdo 90] Czejdo, B., Elmasri, R., Rusinkiewicz, M. and Embley, D. W.: "A Graphical Data Manipulation Language for an Extended Entity-Relationship Model", COMPUTER, Vol.23, No.3, pp.26-36, Mar. 1990

[Gros 89] Grosky, W. I. and Mehrotra, R. (ed.): "Image Database Management", IEEE Computer (special issue), Vol.22, No.12, pp.7-71, Dec. 1989.

[Iyen 88] Iyenger, S. S. and Kashyap, R. L.: "Image Databases", IEEE Trans. on Software Engineering (special selection), Vol.SE-14, No.5, pp.608-688, May 1988.

[Kasa 89] Kasahara, H. and Abe, S.: "Image Mnemonics — Indexing and Communication —", Proc. of Advanced Database System Symposium'89, pp.293-300, Dec. 1989.

[Kato 89a] Kato, T.: "Multimedia Interaction on Advanced Image Database System — Intelligent Personal Interface —", Proc. of Computer World'89, pp.174-181, Sep. 1989.

[Kato 89b] Kato, T., Kurita, T. and Shimogaki, H.: "Multimedia Interaction with Image Database Systems", Proc. of Advanced Database System Symposium'89, pp.271-278, Dec. 1989.

[Kato 89c] Kato, T., Kurita, T. and Sakakura, A.: "Electronic Art Museum: Full Color Image Database with Visual Interaction on Color and Sketch", Tech. Rep. of SIG-IE of IEICE Japan, IE88-118, pp.31-38, Mar. 1989 (in Japanese).

[Kato 89d] Kato, T., Shimogaki, H. and Fujimura, K.: "TRADEMARK: Multimedia Image Database System with Intelligent Human Interface", Trans. of IEICS Japan, Vol.J72-D-II, No.4, pp.535-544, Apr. 1989 (in Japanese).

[Kato 89e] Kato, T. and Mizutori, T.: "Multimedia Data Model for Advanced Image Information Systems", Proc. of Advanced Database System Symposium'89, pp.113-120, Dec. 1989.

[Kurita 90] Kurita, T., Shimogaki, H. and Kato, T.: "A Personal Interface for Similarity Retrieval on an Image Database System", Trans. of IPS Japan, Vol.31, No.2, pp.227-237, Feb. 1990 (in Japanese).

[Mac 86] MacLeod, D. I. A. and Thomas, J. P. (ed.): "Basic Sensory Processes I", in "Hand Book of Perception and Human Performance (I)", Boff, K. R., Kaufman, L. and Thomas, J. P. (ed.), Sec.2, 1986.

[Masu 89] Masunaga, Y.: "An Object-Oriented Approach to Multimedia Database Organization and Management", Proc. of the Int. Symp. on Database Systems for Advanced Application, pp.190-200, Apr. 1989.

[Rosen 82] Rosenfeld, A. and Kak, A. C.: "Digital Picture Processing (2e)", Academic Press, 1982.

[Neu 89] Neuhold, E. J. and Kracker, M.: "Cognitive Aspects of Accessing Multi-Media Information", Proc. of Computer World'89, pp.119-126, Sep. 1989.

[Norm 86] Norman, D. A.: "Cognitive Engineering", in Norman, D. A. and Draper, S. W. (ed.) "User Centered System Design", pp.31-61, Lawrence Erlbaum Associates, Pub., 1986.

(*) This paper is a modified version of "Visual Interaction with Electronic Art Gallery", by Kato, T. and Kurita, T., ETL TR-90-13, Mar. 1990.