COMPUTATIONAL MODELING OF *KANSEI* PROCESSES FOR HUMAN-CENTERED INFORMATION TECHNOLOGY

Toshikazu KATO

Dept. of Industrial and Systems Engineering, Chuo University
kato@indsys.chuo-u.ac.jp

Abstract

This paper introduces the basic concept of computational modeling of perception processes for multimedia data. Such processes are modeled as hierarchical inter- and intra- relationships amongst information in physical, physiological, psychological and cognitive layers in perception. Based on our framework, this paper gives the algorithms for content-based retrieval for multimedia database systems.

1. Introduction

Value and effect of information is an important aspect in human communication as well as semantics and amount of information. *Kansei* is a subjective criterion for evaluating information based on a personal knowledge, taste, feeling, emotion, intention and ideas, which can be summarized as subjectivity. In human-to-human communication, we have to cover the difference of personal backgrounds building up each person's *kansei* model in our mind through mutual interaction to understand the messages with his background.

We need *kansei* mediation mechanisms to perform human-friendly information services with multimedia human interface to understand his request with his background. Modeling each user's *kansei* process for multimedia information enables to share and re-use subjective information amongst human beings with different personal backgrounds. Noiseless transmission and/or recognition of the multimedia message itself are not sufficient for communication.

2. *Kansei* Model for Perception Process – A Framework of Artificial *Kansei* –

As a working hypothesis, we have been developing a *kansei* model on multimedia perception process (Fig.1.)

(1) Physical level interaction: A visual cue may often remind us similar images or related pictures. This process is a kind of similarity and associative retrieval of pictorial data by physical level interaction with pictorial database.

(2) Physiological level interaction: Early stage of mammal vision mechanism extracts graphical features such as intensity levels, edge, contrast, correlation, spatial frequency, and so on. Visual perception may depend on such parameterized graphical features.

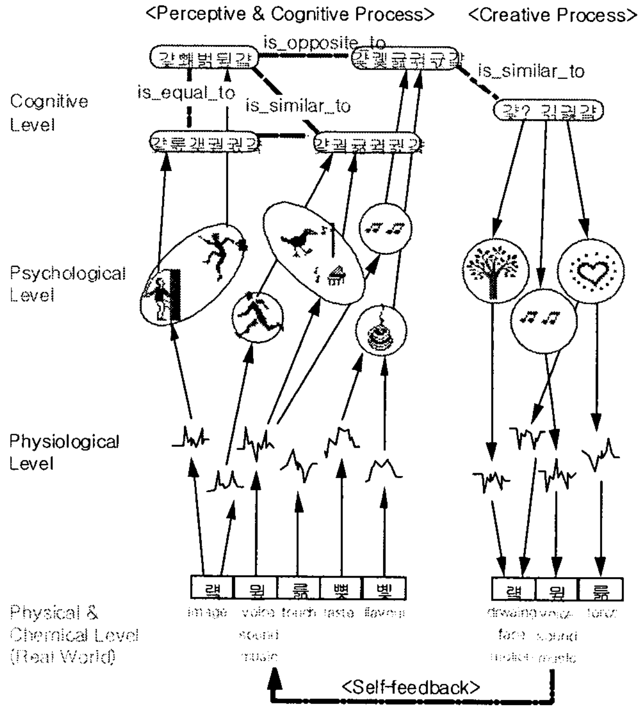

Fig.1. Hierarchical Structure of Kansei Process

(3) Psychological level interaction: We have to notice that the criteria for similarity belong to a subjective human factor. Although human beings have anatomically the common organs, each person may show different interpretation in classification and similarity measure. It means each person has his own weighting factors on graphical features. The computer should evaluate similarity according to each person's subjective criterion.

(4) Cognitive level interaction: We often have differ-

ence impressions, even when viewing the same painting. Each person may also differently give a unique interpretation even viewing the same picture. It seems each person has his own correlation between concepts and graphical features and/or subjective features.

3. Physical Level Interaction

3.1 Physical Perception Model

The general composition of a picture is one of the major parts in visual stimuli to human beings. We can assume that a person may remember an outline of a picture and can draw it as a rough sketch.

Boundaries of objects roughly approximate the general composition of the original picture. We can automatically extract such boundaries by our differential. In our algorithm, the boundaries are adaptively selected not only from global criteria as well as from local one. (When a person is paying attention to an object, he is involuntarily trying to trace the clearly perceived boundary points in the local scope.) We will refer such a map of boundary points as an abstract image of the original picture.

3.2 Sketch Retrieval on Pictorial Index

How can we organize sketch retrieval mechanism for pictorial database systems? Our goal is as follows. A user has only to draw an outline sketch of a picture to retrieve the original one. We call this sort of content-based operation "query by visual example (QVE)". The system automatically evaluates similarity by image matching between the sketch and the abstract images in the database. The collection of abstract images is referred to as pattern-type pictorial index.

Let us show an example of sketch retrieval of full color painting in Fig.2. While user's sketch is quite partial and deformed, the best three candidates have quite similar composition; trees appear in the right half of each painting. The average recall ratio was 94.4% from 205 paintings, if the sketch was not too rough to recognize the general composition.

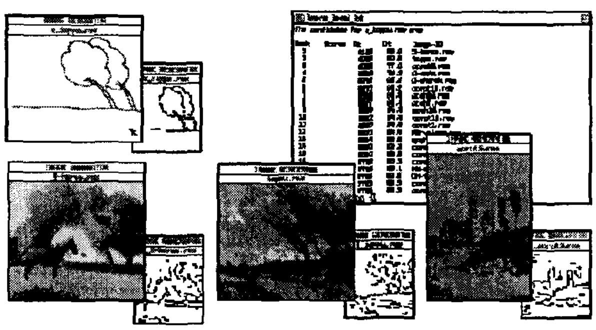

Fig.2. Sketch retrieval of paintings
by showing a rough sketch

4. Physiological Level Interaction

4.1 Physiological Perception Model

We will adopt graphical feature (GF) representation to describe the visual stimuli at physiological level of human vision mechanism. At the early stage of human vision mechanism, many sorts of low-level graphical features are extracted through the neural networks. We can summarize such graphical features into spatial distribution of the intensity level, spatial frequency, local correlation measure and local contrast measure. We can simulate the visual perception process by a GF space. Systems refer to these GF vectors as the pictorial index of physiological level representation.

4.2 Sketch Retrieval on GF Space

We may expect that similar pictures give the similar GF values and are mapped into the neighboring points in the GF space. For example, a GF value of a fine copy and that of a rough sketch are mapped into neighbors in the GF space. The system refers to these GF values as the pictorial index of pictures. Therefore, the system can retrieve original pictures by comparing their GF values.

Fig.3 shows the outline of the whole QVE mechanisms. A visual example as well as every picture in the database is automatically mapped into the GF space by analyzing the low level graphical features when it is registered to the system. We can expect that the closest point in GF space is the original picture.

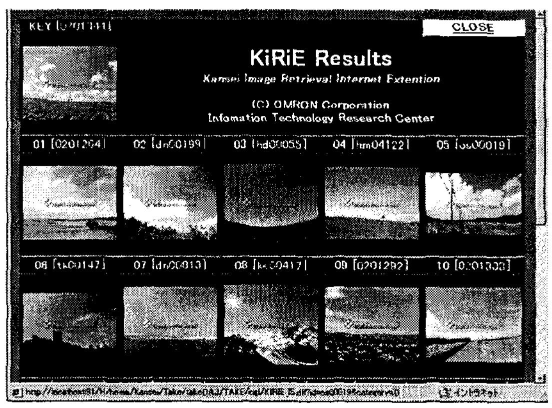

Fig.4. Query by visual example; a color photo

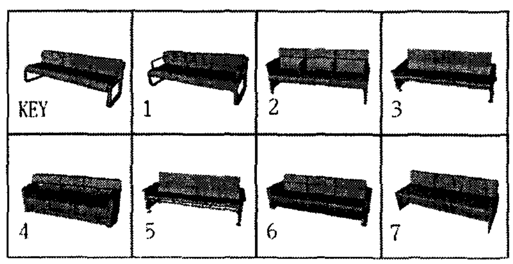

Fig.5. Query by visual example; a 3D-object model

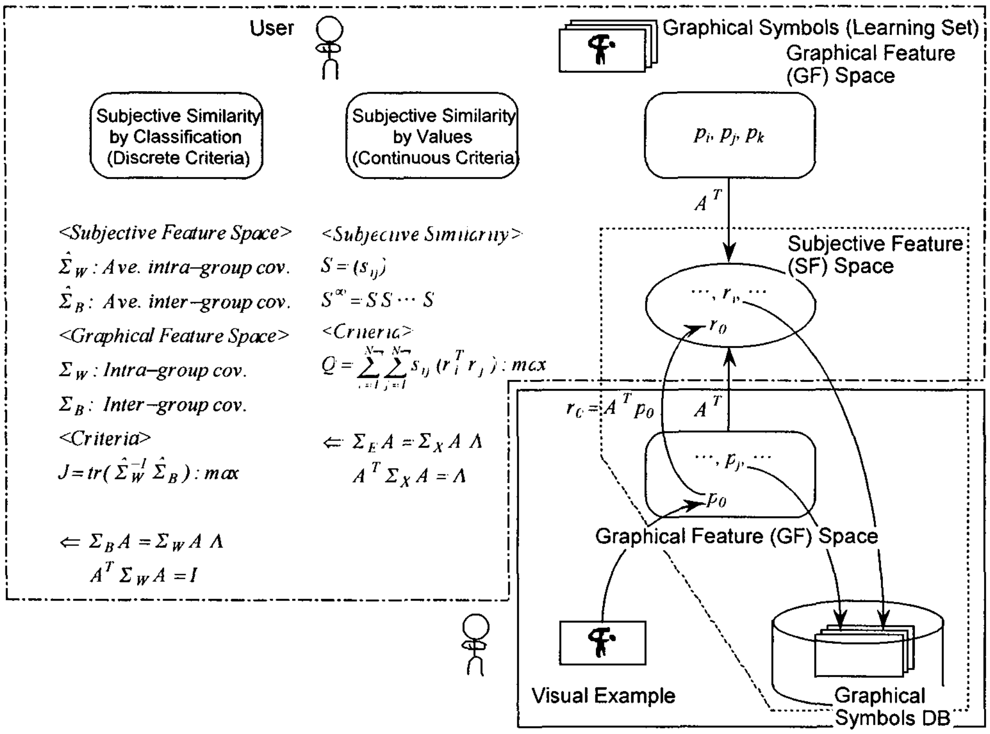

Fig.3. Overview of Query by Visual Example (QVE) mechanism on GF snd SF spaces

In Fig.4, a color photo is shown as a visual example, and similar stamps are retrieved as candidates.

We can extend this idea to three-dimensional objects. In Fig.5, an object (3D model) is picked up as a visual example, and similar objects are retrieved as candidates.

5. Psychological Level Interaction

5.1 Psychological Perception Model

We also have to consider a psychological aspect of similarity measure. Such similarity differs with each user even for the same figures. The system should learn the subjective similarity measure as a personal view from each user.

We need a subjective feature (SF) space which reflects the subjective similarity measure of a specific user. We can construct such an SF space by the statistical analysis. The discriminant analysis is one of the methods to analyze and simulate the classification (Fig.3.) We will refer to the SF space of as the personal index. Once the system has learned the linear mapping A, it can automatically construct the personal index only from the GF vectors. This algorithm reduces the personnel expenses for indexing.

5.2 Sketch Retrieval on SF Space

We may expect that subjectively similar pictures may give similar SF values and are mapped into the neighbors in the SF space for the user. Showing an example, which a user wants to see, the system retrieves similar pictures by comparing their SF vector, and shows suitable candidates (Fig.3.)

Fig.6 shows an example of similarity retrieval. The upper QVE window in this figure shows the ten candidates for similarity retrieval on the SF space, while the lower one shows the ten candidates on the GF space. The second to the eighth candidates on the SF space have matched with the classification by this user.

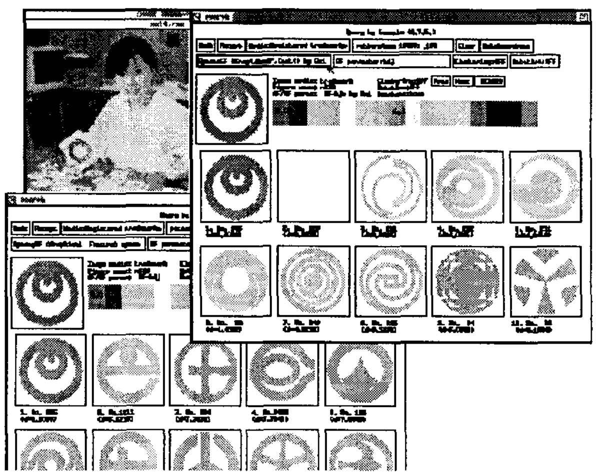

Fig.6. Example of similarity retrieval based on a subjective criterion

For a Patent Office application, modeling judging criteria of patent officers, who are experts in evaluating design and similarity among them, corresponds to create a knowledge base on graphical symbols. We can use the database system as a multimedia expert system referring the SF space as expert's knowledge.

We can also extend this idea to three-dimensional objects. Once the system has learned each user model, such as of an industrial designer and also of a consumer, it can perform similarity retrieval based on each user model as shown in Fig. 7.

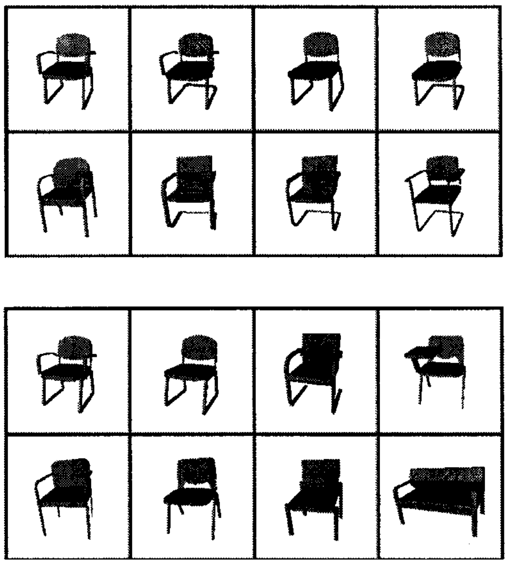

Fig.7. Similarity retrieval of 3D objects; base on different user models

Using this mechanism, we can also visualize the SF space of each user in 3D space form. W can easily understand the difference of the subjective criteria by viewing the space (Fig.8.)

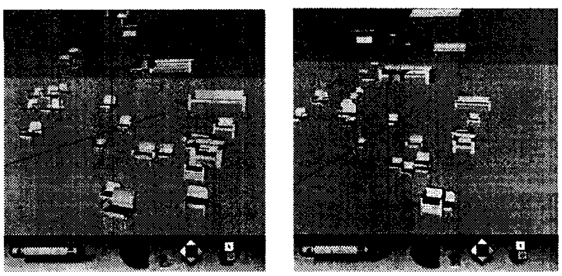

Fig.8. Visualization of SF space in 3D form

6. Cognitive Level Interaction

6.1 Cognition Model for Multimedia Data

The conventional pictorial database systems have the following problems.

The indexer has to assign suitable keywords to every picture when it is registered to the system. We need an automatic indexing method for effective database management.

Even when viewing the same picture, we often give the different descriptions. That means natural language description belongs to a subjective matter. Such subjectivity depends on personal character, education, skill and cultural backgrounds. We need a personal view mechanism for modeling subjective criterion of each person.

Keyword is keyword and picture is picture. Conventional database operations can be performed only on the textual data domain. We need a multimedia operation mechanism which can evaluate the values of multiple domains, such as textual domain and pictorial domain, directly, which is an essential requirement for multimedia database management systems.

Our approach aims to solve these problems at the same time in a single mechanism. Our basic idea is to unify the multiple domain data by finding correlation amongst them through statistical analysis and learning. In case of an electronic art gallery, it corresponds to learn how a person feels certain impressions when viewing a painting. The impressionism painting suggests that there is a reasonable correlation between the coloring and the descriptions.

We developed an algorithm for learning a personal view model for artistic impressions as shown in Fig.9. We may expect the strong correlation between the descriptions and the parameterized coloring feature. The system statistically analyzes such correlation, for example, by canonical correlation analysis method. We will regard the correlation as the personal view model for the specific user. We can construct a unified feature (UF) space on this model to compare the subjective descriptions and coloring features by the distance in the space.

Note that we do not have to assign the adjectives to every painting in the database. Once the system learned the linear mappings between two domains, it can automatically construct the personal index only from the GF values satisfying the specific user's criteria.

6.2 Sense Retrieval on UF Space

We may expect that a painting and its subjective description, given by a specific user, may be mapped into the neighbors in UF space. The user has only to show several words as "query by subjective descriptions (QBD)". The system evaluates the most suitable coloring pictures for the words by the distance in the UF space.

The UF space gives a criterion to evaluate the textual data and the pictorial data by their contents. The UF space enables multimedia operations on different domains. Note that the sense retrieval algorithm evaluates the visual impression in UF space. Therefore, we can retrieve paintings without assigning keywords to every painting.

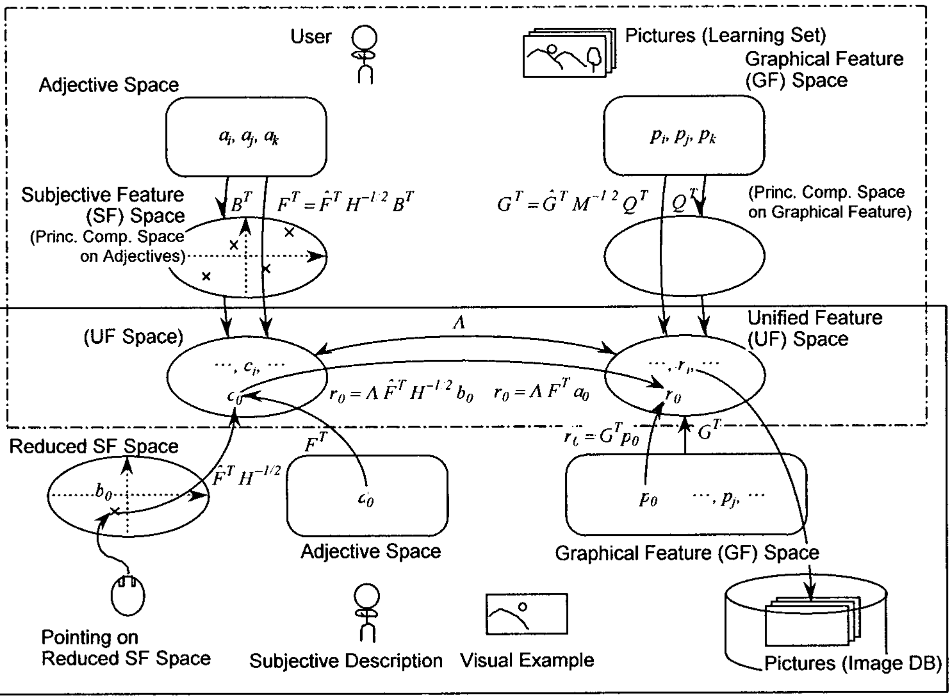

Fig.9. Overview of Query by Subjective Description (QBD) mechanism on UF space

Fig.10 is an example of sense retrieval showing the best eight candidates for the adjectives; "romantic, soft and warm". These paintings roughly satisfied the personal view of the subjects.

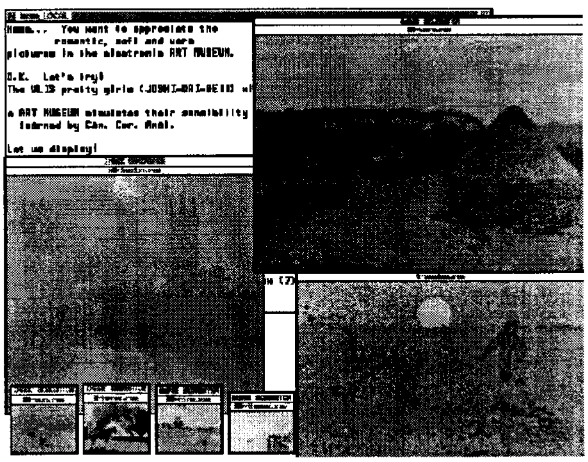

Fig.10. Example of sense retrieval
in Query by Subjective Description

We can also visualize the differences of the vocabulary of each user to describe their impressions. The principal component analysis (PCA) visualizes the spatial relationship of the impression words in the user's SF space. Fig.11 shows the results of PCA on SF spaces of female students and of male students. In the left window, the SF space of female students, "sober" and "placid" are opposite to "gorgeous" and "folksy / ethnic" on the first axis, and "charming", "elegant" are opposite to "hard," "wild" and "dynamic" on the second axis. In the right window, the SF space of male students, we see different relationships.

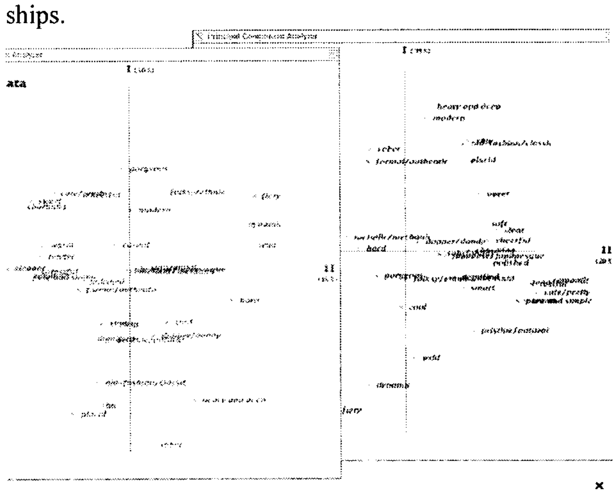

Fig.11. "Concept map" of two distinct users groups
(red: female student, blue: male student)

7. Toward Human-centered Information Technology (Human Media Technology)

Conventional information technologies have been taking computer-centered approach in designing computer systems and computer media. As a counter concept to this, human media technology takes a human-centered approach which can evaluate and produce wide variation of multimedia information including personal information, such as a taste, a feeling on something, on intention, and so on, as well as objective information in a natural manner.

Human media technology focuses on the interaction manner between human beings and information space or amongst human beings through information space. It enables one to directly receive multimedia messages through his five senses, and directly send those on a feeling on something, an intention, and so on.

Human media technology takes a human-centered approach to the next generation information technologies for the advanced information infrastructure in the twenty first century.

The essential technologies for human media can be categorized into three. They are knowledge media technology, virtual media technology and kansei media technology. Let us summarize our essential requirement to human media technology from these aspects.

(a) Knowledge Media: Knowledge media technology supports to describe knowledge of the real world, common sense, context of a task and also each person's knowledge background in relation to multimedia data; ontology base and ontology management. It supports seamless communications amongst users on cyber space to access, extend, share and re-use the knowledge on application domains.

(b) Virtual Media: Virtual media technology supports smooth human-computer interaction mechanisms with multimedia representation in a natural extension form of our sense and behavior. We can act to, or look at the real world through a cyber space with super-reality and actuality. Thus, it enables us to share the real space in one-to-one through many-to-many communication amongst all the participants over a long distance in a multimedia manner.

(c) *Kansei* Media: *"Kansei"* media technology models the each user's cognitive process for multimedia information. Typical subjective information of a specific user is a personal knowledge of something, a taste, a feeling on something, an emotion, an intention, and ideas, and so on. It can create a multimedia message in the most familiar representation to each of the specific persons. It also enables to share and re-use subjective information amongst human beings with different personal backgrounds.

References

(1) Toshikazu Kato: "Multimedia Interaction with Image Database Systems", Proc. of Advanced Database System Symposium'89 (ADSS'89), Dec. 1989.

(2) Toshikazu Kato: "Visual Interaction with Electronic Art Gallery", Proc. of International Conference on Database and Expert Systems Applications (DEXA'90), Aug. 1990.

(3) Toshikazu Kato: "Cognitive View Mechanism for Multimedia Database System", IEEE Computer Society, International Workshop on Interoperability in Multidatabase Systems (IMS'91), Apr. 1991.

(4) Toshikazu Kato: "Database Architecture for Content Based Image Retrieval", SPIE / IS&T, Proc. of International Conference on Image Storage and Retrieval Systems, Feb. 1992.

(5) Toshikazu Kato: "Query by Visual Example -- Content Based Image Retrieval --", Proc. of International Conference on Extending Database Technology'92 (EDBT'92), Mar. 1992.

(6) Toshikazu Kato: "Cognitive View Mechanism for Content-based Multimedia Information Retrieval," in R. Cooper (ed) "Interfaces to Database Systems," pp.244-262, 1992.

(7) Tatsuya Shibata, Toshikazu Kato: "Modeling of Subjective Interpretation for Street landscape Image", Proc. of International Conference on Database and Expert Systems Applications, pp.501-510, Aug. 1998.

(8) Y. Kobayashi, T. Kato: "A High Fidelity Contrast Improving Model Based on Human Vision Mechanisms," Proc. of IEEE ICMCS'99, pp.578-584, June 1999.

(9) Tatsuya Shibata and Toshikazu Kato: "Kansei Image Retrieval System for Street Landscape --- Discrimination and Graphical parameters based on Correlation of Two Images ---", Proc. of IEEE International Conference on Systems, Man and Cybernetics, Vol.VI, pp.247-252, Oct. 1999.

(10) Motofumi Suzuki, Toshikazu Kato and Hideo Tsukune: "3D Object Retrieval based on Subjective Measures", Proc. of International Conference on Database and Expert Systems Applications, pp.850-856, Aug. 1998.

(11) Human Media Project (ed): "Selected Papers on Human Media Database Systems and Kansei Agent Systems", 2000.

(12) M. Tada, T. Kato, I. Shinohara: "Similarity Image Retrieval System Using Hierarchical Classification", Proc. of 13^{th} Database and Expert Systems Applications, DEXA 2002, pp.779-788, Sep. 2002.