CONTENTS

1

A METADATABASE SYSTEM FOR SEMANTIC IMAGE SEARCH BY A MATHEMATICAL MODEL OF MEANING

Yasushi Kiyoki, Takashi Kitagawa*
and Takanari Hayama*

Faculty of Environmental Information, Keio University
Fujisawa, Kanagawa 252, Japan
E-mail: kiyoki@sfc.keio.ac.jp

**Institute of Information Sciences and Electronics, University of Tsukuba*
Tsukuba, Ibaraki 305, Japan
E-mail: takashi@is.tsukuba.ac.jp, taki@is.tsukuba.ac.jp

ABSTRACT

In the design of multimedia database systems, one of the most important issues is to extract images dynamically according to the user's impression and the image's contents. In this chapter, we present a metadatabase system which realizes the semantic associative search for images by giving context words representing the user's impression and the image's contents.

This metadatabase system provides several functions for performing the semantic associative search for images by using the metadata representing the features of images. These functions are realized by using our proposed mathematical model of meaning. The mathematical model of meaning is applied to compute specific meanings of context words which are used for retrieving images unambiguously and dynamically. The main feature of this model is that the semantic associative search is performed in the orthogonal semantic space. This space is created for dynamically computing semantic equivalence or similarity between the metadata items of the images and context words.

1 INTRODUCTION

The design and implementation of metadatabase systems is one of the key issues in the field of multimedia database research. In the design of the metadata for images, the important issues are how to define and represent the metadata items of images and how to extract images dynamically according to the user's impression and the image's contents.

There are many approaches to image retrieval. Two major approaches are direct retrieval using partial pattern matching and indirect retrieval using abstract information of images. We use the latter approach for extracting images.

In this chapter, we present a semantic associative search method for images and its implementation in the metadatabase system for extracting appropriate images according to the user's impression and the image's contents. In this method, the images are selected by using a *mathematical model of meaning*[3, 6] which realizes the semantic associative search. This model has been originally applied to support semantic interoperability for multidatabase environments[6]. The realization of semantic interoperability is one of the important issues in the research field of multidatabase systems[13, 14, 17].

The mathematical model of meaning is applied as a fundamental framework for representing the metadata and extracting images. This model realizes the semantic associative search for extracting information by giving its context words. The main feature of this model is that the semantic associative search is performed unambiguously and dynamically in the orthogonal semantic space. This space is created for computing semantic equivalence or similarity between the user's impression and the image's metadata items which represent the characteristics of each image.

We point out that context recognition is essentially needed for multimedia information retrieval. The meaning of information is determined by the relation between contents and the context. The machinery for realizing dynamic context recognition is essentially required for multimedia information acquisition. We have proposed new methodology for realizing the machinery[3, 4, 6].

The advantages and original points of the proposed methodology are as follows:

1. The semantic associative image search based on semantic computation for words is realized by a mathematical approach. This image search method surpasses the search methods which use pattern matching for associative

search. Users can use their own words for representing impression and image's contents for image retrieval, and do not need to know how the images of retrieval candidates are characterized in databases.

2. Dynamic context recognition is created using a mathematical foundation. The context recognition can be used for obtaining multimedia information by giving the user's impression and the contents of the information as a context. A semantic space is created as a space for representing various contexts which correspond to its subspaces. A context is recognized by the computation for selecting a subspace.

The MMM (Mathematical Model of Meaning) consists of:

1) A set of m words is given, and each word is characterized by n features. That is, an m by n matrix is given as the data matrix.

2) The correlation matrix with respect to the n features is constructed. Then, the eigenvalue decomposition of the correlation matrix is computed and the eigenvectors are normalized. The orthogonal semantic space is created as the span of the eigenvectors which correspond to nonzero eigenvalues.

3) Images and context words are characterized by using the specific features(words) and representing them as vectors.

4) The images and context words are mapped into the orthogonal semantic space by computing the Fourier expansion for the vectors.

5) A set of all the projections from the orthogonal semantic space to the invariant subspaces (eigen spaces) is defined. Each subspace represents a phase of meaning, and it corresponds to a context or situation.

6) A subspace of the orthogonal semantic space is selected according to the user's impression or the image's contents, which are given as a context represented by a sequence of words.

7) The most correlated image to the given context is extracted in the selected subspace.

Several information retrieval methods, which use the orthogonal space created by mathematical procedures like SVD (Singular Value Decomposition), have been proposed. Our model is essentially different from those methods using the SVD (e.g. the Latent Semantic Indexing (LSI) method [1]). The essential difference is that our model provides the important function for semantic projections which realizes the dynamic recognition of the context. That is, in our model, the context-dependent interpretation is dynamically performed for computing the distance between words and images by selecting a subspace from the entire orthogonal semantic space. In our model, the number of phases of the contexts is almost infinite (currently 2^{800}, approximately). Other methods do not provide the context dependent interpretation for computing equivalence and similarity in the orthogonal space, that is, the phase of meaning is fixed and static.

In [16], a three-layered data model has been introduced to extract image features and semantics for content-based image retrieval. In this model, a high level semantic retrieval for medial image information is realized. Similarly to our approach, the features of images are obtained from the images themselves, and image contents are extracted with conceptual terms. The image retrieval is realized with the association between the conceptual terms and the image features. Our approach is different from this approach in modeling image contents. In our approach, we realize semantic image search by obtaining impression and object terms of image data by the observation of the entire image. As the medical application is assumed in the approach in [16], spatial and temporal characteristics of objects in images and their spatial relationships are important in image abstraction, and the shape and spatial relationships are described and used in the retrieval.

In this chapter, we present three methods for representing the metadata items for images and basic functions which extract the appropriate images from the orthogonal semantic space.

2 METADATABASE SYSTEM

2.1 The overview of the metadatabase system

The metadatabase system selects appropriate images for requests of database users by using metadata items and basic functions. This system consists of the following subsystems(Figure 1):

(1) Metadata Acquisition Subsystem: This subsystem supports the facilities for acquiring metadata from the database storing the source images.

Figure 1 The Metadatabase System Architecture

(2) Metadatabase Management Subsystem: This subsystem supports the facilities for keeping metadata consistent in the orthogonal semantic space. This subsystem also provides the facilities for interpreting users's queries.

(3) Image Selection Subsystem: This subsystem supports the facilities for selecting appropriate images by using the MMM. Three methods are provided for representing the metadata items for images. This subsystem maps the metadata items for images into the orthogonal space. By receiving metadata items for context representation, it selects the most correlated image to the context.

2.2 Basic functions and metadata for images

The metadatabase system is used to extract image data items corresponding to context words which represent the user's impression and the image's contents. For example, the context words "powerful" and "strong" are given, the image with the impression corresponding to these context words is extracted. Each metadata item of images is mapped in the orthonormal semantic space. This space is referred to as "orthogonal metadata space" or "metadata space." The MMM is used to create the orthogonal metadata space. By this orthogonalization, we can define appropriate metric for computing relationships between metadata items for images and context representation. The MMM gives the machinery for extracting the associated information to the context.

Three types of metadata are used.

(1) *Metadata for space creation*: These metadata items are used for the creation of the orthogonal metadata space, which is used as a space for semantic image retrieval.

(2) *Metadata for images*: These metadata items are the candidates for semantic image retrieval.

(3) *Metadata for context representation*: These metadata items are used as context words for representing the user's imagination and the image's contents.

The basic functions and metadata structures are summarized as follows:

1. Creation of metadata space:

To provide the function of semantic associative search, basic information on m data items ("metadata for space creation") is given in the form of a matrix. Each metadata item is given independently one another. No relationship between the metadata items need to be described. The information of each data item is represented by its features. The m basic metadata items is given in the form of an m by n matrix M where each metadata item is characterized by n features. By using M, the orthogonal space is created as the metadata space $\mathcal{MDS}$. These metadata items are determined as mentioned in the following section.

2. Representation of metadata for images in n-dimensional vectors

   Each metadata item for images is represented in the n-dimensional vector whose elements correspond to n features. The metadata items for images become the candidates for the semantic associative search. Furthermore, each of the context words, which are used to represent the user's impression and the image's contents in semantic image retrieval, is also represented in the n-dimensional vector.

3. Mapping data items into the metadata space $\mathcal{MDS}$.

   Metadata items (metadata for space creation, metadata for images and metadata for context representation) which are represented in n-dimensional vectors are mapped into the orthogonal metadata space. Those data items are used as context words and target image data items which are extracted according to users' requests.

4. Semantic associative search

   When a sequence of context words which determine the user's impression and the image's contents is given, the images corresponding to the context are extracted from a set of retrieval candidate images in the metadata space.

2.3 A creation method of the metadata space

We introduce an implementation method for the creation of the $\mathcal{MDS}$.

The procedure for the creation of the $\mathcal{MDS}$ is as follows:

1. To create the data matrix M, we can use "General Basic English Dictionary[12]" in which 850 basic words are used to explain each English definition. Those

850 basic words are used as features, that is, they are used for characterizing metadata as the features corresponding to the columns in the matrix M. The 2,000 data items are used as "metadata for space creation." Those metadata items are used as the basic words in the English dictionary "Longman Dictionary of Contemporary English [9]." Each metadata item for space creation corresponds to a row of the matrix M. In the setting of a row of the matrix M, each column corresponding to the features which appear in the explanation of the data item is set to the value "1". If the feature is used as a negative meaning, the column corresponding to it is set to the value "-1". And, the other columns are set to the value "0". This process is performed for each of 2000 metadata items. And then, each column of the matrix is normalized by the 2-norm to create the matrix M.

2. By using this matrix M, the $\mathcal{MDS}$ is created as the orthogonal space. The creation method of the orthogonal space is described in Section 3.1. This space represents the semantic space for computing contexts and meanings of the metadata items.

To automatically create the data matrix M from the dictionary, several filters are used, which remove unnecessary elements (words), such as articles and pronouns, and transform conjugations and inflections of words to the infinitives. Those elements are removed from the features characterizing each data item. The unnecessary words are not used as features in the data matrix M.

(filter-1) This filter eliminates the unnecessary elements, such as articles and pronouns.

(filter-2) This filter transforms conjugations and inflections to the infinitives.

(filter-3) This filter transforms the large characters to the small ones.

(filter-4) This filter transforms clipped form words to the corresponding original words.

(filter-5) The rows of the matrix M are created for each data item by using the filtered features which characterize the data item.

Each metadata item (metadata item for images, metadata item for context representation) is mapped into the metadata space, by computing the Fourier expansion for the n-dimensional vector representing the metadata item itself. These metadata items are defined as metadata by using the n features. These metadata items are used as context words and metadata items for retrieval candidate images.

2.4 Creation methods of metadata for images

We present three methods for creating metadata for images.

In [15], Kashyap et al. have clearly identified and classified various metadata for digital media into three basic categories: *content-dependent metadata*, *content-descriptive metadata* and *content-independent metadata*. The metadata for images which is used in our semantic associative search is categorized in the content-descriptive metadata, because the metadata is associated with the original image without being extracted directly from the image contents themselves. Furthermore, in [15], the content-descriptive metadata is classified into two categories: *domain-dependent metadata* and *domain-independent metadata*.

In the following Method-1 and Method-2, the metadata for images is categorized into *domain-dependent* because the metadata is extracted from domain-specific concepts, which are used as a basis for the determination of the metadata itself. That is, the metadata type used in these methods is categorized as the *content-descriptive domain-dependent metadata*.

In Method-3, metadata for images is extracted from their color components which are used to characterize image features in an experimental psychology model of correlating colors and their impression words. This type of metadata is categorized as the *content-descriptive domain-independent metadata*.

Method-1

Each image is explained by using the n features which are used in the creation of the data matrix M. In this explanation, the impression or the content of the image is represented by using these features as the metadata for the image. As the result, each image is represented as n-dimensional vector in which the non-zero value is assigned to the corresponding elements of the vector to these features.

The image P is explained and defined by using some of the words which are used in the n features. Then, the image is represented as an n-dimensional vector.

$$P = (w_{i1}, w_{i2}, \dots, w_{in}).$$

Each metadata item is mapped into the metadata space by computing the Fourier expansion for the vector corresponding to the image data item itself.

Method-2

The image P is represented in t impression words $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_t$, where each impression word is defined as a t dimensional vector:

$$\mathbf{o}_i = (o_{i1}, o_{i2}, \ldots, o_{in}),$$

which is characterized by n specific features.

Namely, we define the image P as the collection of t impression words which represent the image.

$$P = \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_t\}.$$

Moreover, we define the operator union $\bigoplus$ of impression words $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_t$, to represent the metadata for the image P as a vector as follows:

$$\bigoplus_{i=1}^{t} \mathbf{o}_i \equiv (\mathrm{sign}(o_{\ell_1 1}) \max_{1 \le i \le t} |o_{i1}|, \mathrm{sign}(o_{\ell_2 2}) \max_{1 \le i \le t} |o_{i2}|,$$

$$\ldots, \mathrm{sign}(o_{\ell_n t}) \max_{1 \le i \le t} |o_{in}|)$$

where $\mathrm{sign}(a)$ represents the sign (plus or minus) of "a" and $\ell_k, k = 1, \ldots, n$, represents the index which gives the maximum, that is:

$$\max_{1 \le i \le t} |o_{ik}| = |o_{\ell_k k}|.$$

Method-3

In this method, the metadata for images is automatically and indirectly extracted from image data items themselves. Color is known as the dominant factor which affects the impression of images[7, 8, 18]. We use color to derive impressions of images.

The basic idea of this method is to describe both images and impression words in the notion of *color* and compute correlations between images and words. Color used in this method is represented in the Munsell color system as it is more familiar to the human perception. Additionally, the color names defined by

ISCC(Inter-Society Color Council) and NIST(National Institute of Standards and Technology) are used to describe both images and impression words in the notion of color. By taking the correlations between images and impression words, we can obtain the suitable words which describe the impressions of images. The metadata for images is computed from the obtained impression words by the previously defined union operator $\bigoplus$. The detail of this method is described in Section 7.

3 CREATION OF A METADATA SPACE AND BASIC FUNCTIONS

In this section, we introduce a creation method of the metadata space $\mathcal{MDS}$ for systematically storing metadata and for implementing the semantic associative search for images.

3.1 Creation of a metadata space

The semantic associative search for images is created by using our mathematical model of meaning [3, 6]. For the metadata items for space creation, a data matrix M is created. When m data items for space creation are given, each data item is characterized by n features $(f_1, f_2, \cdots, f_n)$. For given $\mathbf{d}_i (i = 1, \cdots, m)$, the data matrix M is defined as the $m \times n$ matrix whose i-th row is $\mathbf{d}_i$. Then, each column of the matrix is normalized by the 2-norm in order to create the matrix M.

Figure 1 shows the matrix M. That is $M = (\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \cdots, \mathbf{d}_n)^T$.

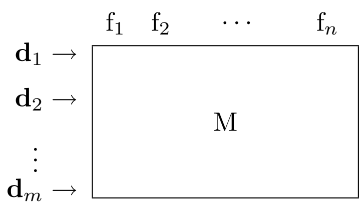

Figure 2 Representation of metadata items by matrix M

1. The correlation matrix M^TM of M is computed, where M^T represents the transpose of M.

2. The eigenvalue decomposition of M^TM is computed.

$$M^TM = Q \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_\nu & & \\ & & & 0 \cdot \cdot & \\ & & & & 0 \end{pmatrix} Q^T,$$

   $0 \le \nu \le n$.
   The orthogonal matrix Q is defined by

$$Q = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n)^T,$$

   where $\mathbf{q}_i$'s are the normalized eigenvectors of M^TM. We call the eigenvectors "semantic elements" hereafter. Here, all the eigenvalues are real and all the eigenvectors are mutually orthogonal because the matrix M^TM is symmetric.

3. Defining the metadata space $\mathcal{MDS}$.

$$\mathcal{MDS} := span(\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_\nu),$$

   which is a linear space generated by linear combinations of $\{\mathbf{q}_1, \cdots, \mathbf{q}_\nu\}$. We note that $\{\mathbf{q}_1, \cdots, \mathbf{q}_\nu\}$ is an orthonormal basis of $\mathcal{MDS}$.

3.2 The set of the semantic projections Π_ν

The projection P_{λ_i} is defined as follows:
$P_{\lambda_i} \overset{\mathrm{d}}{\iff}$ Projection to the eigenspace corresponding to the eigenvalue λ_i,
i.e. $P_{\lambda_i} : \mathcal{MDS} \to span(\mathbf{q}_i)$.
The set of the semantic projections Π_ν is defined as follows:

$\Pi_\nu :=$

$$\{\ 0\ ,\ P_{\lambda_1}, P_{\lambda_2}, \cdots, P_{\lambda_\nu},$$

$$P_{\lambda_1} + P_{\lambda_2}, P_{\lambda_1} + P_{\lambda_3}, \cdots, P_{\lambda_{\nu-1}} + P_{\lambda_\nu},$$

$$\vdots$$
$$P_{\lambda_1} + P_{\lambda_2} + \cdots + P_{\lambda_\nu}\}.$$

The number of the elements of Π_ν is 2^ν, and accordingly it implies that 2^ν different contexts can be expressed by this formulation.

3.3 Semantic operator

The correlations between each context word and each semantic element are computed by this process. The context word is used to represent the user's impression and the image's contents for images to be extracted. A sequence

$$s_\ell = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_\ell)$$

of ℓ context words and a positive real number $0 < \varepsilon_s < 1$ are given, the semantic operator S_p constitutes a semantic projection $P_{\varepsilon_s}(s_\ell)$, according to the context. That is,

$$S_p : T_\ell \longmapsto \Pi_\nu$$

where T_ℓ is the set of sequences of ℓ words and $T_\ell \ni s_\ell$, $\Pi_\nu \ni P_{\varepsilon_s}(s_\ell)$. Note that the set $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_\ell\}$ must be a subset of the words defined in the matrix M.

The constitution of the operator S_p consists of the following processes:

1. Fourier expansion of $\mathbf{u}_i (i = 1, 2, \cdots, \ell)$ is computed as the inner product of $\mathbf{u}_i$ and $\mathbf{q}_j$ u_{ij}, i.e.

   $$u_{ij} := (\mathbf{u}_i, \mathbf{q}_j)\ ,\ for\ j = 1, 2, \cdots, \nu.$$

   $\widehat{\mathbf{u}}_i \in \mathcal{I}$ is defined as

   $$\widehat{\mathbf{u}}_i := (u_{i1}, u_{i2}, \cdots, u_{i\nu}).$$

   This is the mapping of the context word $\mathbf{u}_i$ to the $\mathcal{MDS}$.

2. The semantic center $\mathbf{G}^+(s_\ell)$ of the sequence s_ℓ is computed as,

   $$\mathbf{G}^+(s_\ell) := \frac{\left(\sum_{i=1}^{\ell} u_{i1}, \cdots, \sum_{i=1}^{\ell} u_{i\nu}\right)}{\|\left(\sum_{i=1}^{\ell} u_{i1}, \cdots, \sum_{i=1}^{\ell} u_{i\nu}\right)\|_\infty},$$

   where $\|\cdot\|_\infty$ denotes infinity norm.

3. The semantic projection $P_{\varepsilon_s}(s_\ell)$ is determined as,

$$P_{\varepsilon_s}(s_\ell) := \sum_{i \in \Lambda\varepsilon_s} P_{\lambda_i} \quad \in \Pi_\nu,$$

where $\Lambda_{\varepsilon_s} := \{\ i \mid |\left(\mathbf{G}^+(s_\ell)\right)_i\ | \ > \varepsilon_s\}$.

3.4 Function for Semantic Image Search

We introduce a function to measure the similarity between images and context words. The function measures the quantity of association or correlation between context words and the candidate images. We also introduce a dynamic metric depending on the context between two images.

Function to measure the association

The function measures the association between context words and the candidate images. Suppose a sequence of associate context words is given to search an image, e.g. { dynamic, powerful }. We can regard the context words as the words forming the context s_ℓ. We can specify some subspace by using the semantic operator with weights c_j's which are given by

$$c_j(s_\ell) := \frac{\sum_{i=1}^{\ell} u_{ij}}{\| \left(\sum_{i=1}^{\ell} u_{i1}, \cdots, \sum_{i=1}^{\ell} u_{i\nu} \right) \|_\infty},$$

$$j \in \Lambda_{\varepsilon_s}.$$

Since the norm of the image, which can be calculated from the metadata of the image, reflects the correlation between the image and the semantic elements included in the selected subspace, we may use it as a measure for the association between the context words and the image data.

We introduce a function for computing the norm of the image. The function $\bar{\eta}_0\left(\mathbf{x}; s_\ell\right)$ is defined as follows:

$$\bar{\eta}_0\left(\mathbf{x}; s_\ell\right) = \frac{\sqrt{\sum_{j \in \Lambda_{\varepsilon_s} \cap \mathcal{S}} \left\{c_j(s_\ell) x_j\right\}^2}}{\|\mathbf{x}\|_2},$$

where the set $\mathcal{S}$ is defined by $\mathcal{S} = \{i | sign(c_i(s_\ell) = sign(x_i)\}$.

In this function, we eliminate the effect of the negative correlation. We note that the sum in the numerator of this function is sought over the selected semantic subspace from the context s_ℓ, while the norm in the denominator is sought over the whole metadata space $\mathcal{MDS}$.

Dynamic metric

We introduce a dynamic metric between the image data items, according to a context. Since each image data item can be represented as a vector via the union operator $\oplus$ defined in Section 2.4, we can utilize the metric, which we defined for two distinct words in [3, 4, 6], to compute the similarity between metadata items of images. The dynamic metric $\rho(\mathbf{x}, \mathbf{y}; s_\ell)$ for $\mathbf{x}, \mathbf{y} \in \mathcal{MDS}$ is introduced to compute the similarity between metadata items for images.

The metric $\rho(\mathbf{x}, \mathbf{y}; s_\ell)$ to compute the similarity between metadata items $\mathbf{x}$, $\mathbf{y}$ of two images in the given context s_ℓ is defined as follows:

$$\rho(\mathbf{x}, \mathbf{y}; s_\ell) = \sqrt{\sum_{j \in \Lambda_{\varepsilon_s}} \{c_j(s_\ell)(x_j - y_j)\}^2},$$

This metric, because of the presence of dynamic weights c_j's, can faithfully reflect the change of the context.

4 SEMANTIC ASSOCIATIVE SEARCH FOR METADATA FOR IMAGES

The proposed system realizes the semantic associative search for metadata items for images.

The basic function of the semantic associative search is provided for context-dependent interpretation. This function performs the selection of the semantic subspace from the metadata space. When a sequence s_ℓ of context words for determining a context is given to the system, the selection of the semantic subspace is performed. This selection corresponds to the recognition of the context, which is defined by the given context words. The selected semantic subspace corresponds to a given context. The metadata item for the most correlated image to the context in the selected semantic subspace is extracted from the specified image data item set $\mathcal{W}$. By using the function defined in Section 3.4, the semantic associative search is performed by the following procedure:

1. When a sequence s_ℓ of the context words for determining a context(the user's impression and the image's contents) are given, the Fourier expansion is computed for each context word, and the Fourier coefficients of these words with respect to each semantic element are obtained. This corresponds to seeking the correlation between each context word and each semantic element.
2. The values of the Fourier coefficients for each semantic element are summed up to find the correlation between the given context words and each semantic element.
3. If the sum obtained in the step 2 in terms of each semantic element is greater than a given threshold ε_s, the semantic element is employed to form the semantic subspace $P_{\varepsilon_s}(s_\ell)\mathcal{MDS}$. This corresponds to recognizing the context which is determined by the given context words.
4. By using the function $\bar{\eta}_0\,(\mathbf{x}; s_\ell)$, the metadata item for the image with the maximum norm is selected among the candidate metadata items for images in $\mathcal{W}$ in the selected semantic subspace. This corresponds to finding the image with the greatest association to the given context words from $\mathcal{W}$.

5 EXAMPLES OF CREATING METADATA

5.1 Method-1

To create the metadata as vectors for images by using the 850 basic words of "General Basic English Dictionary," the designer of metadata looks at an image and checks features that correspond to the image. If the feature corresponds to the image, the value 1.0 is put for that feature, if it does not correspond to the feature, the value 0.0 is put, and if it negates the image, the value -1.0 is put. Although the cost is very expensive, the simplest way is to check for 850 features one by one for each image. A vector is created for each image, and mapped into the $\mathcal{MDS}$.

5.2 Method-2

As the previous method, the same features from “General Basic English Dictionary” are used. This method creates metadata by giving impressions of the original image or referring to objects composing it. The impression words or object names which are extracted from the image are referred to from “General Basic English Dictionary,” and the explanatory words for each impression word or object name are checked as features, and the value for each feature in the vector corresponding to the impression word or object name is set. Then, the vector corresponding to the image is created from the vectors corresponding to the impression words or object names by the union operator defined in Section 2.4.

5.3 Method-3

In this method, metadata for an image is created by referring to color components. Digital images, which usually represented in the RGB color system, are transformed to color vectors in the Munsell color space by using the MTM(Mathematical Transformation to Munsell)[11]. The MTM is described in Section 7. Scaler values of color vectors for corresponding color are defined in the range of 0.0 and 1.0 according to given rules. One of the rules which we have used defines each value by referring occupancy of colors in images.

The difficulty of this method is to define the association between the colors and impression words. That is, how to create the descriptive metadata for context words. To solve this difficulty, we referred to the results from the field of the experimental psychology. Many word association tests have been done to make clear the relation between colors and psychological effects.

6 EXPERIMENTS

6.1 Metadata items in the experiment

In this experiment, we use 30 images and the corresponding metadata items representing the impressions of those images. Those metadata items are represented in the impression words shown in Table 1. Each impression word is explained by using several features from the 850 basic words of “General Basic

Table 1 Image data and their definitions(impression words)

image data name	impression
*chagalla*1	vivid,quiet,substance
*chagallb*1	grief,sombre,terrible
*chagallc*1	sober,dynamic,motion
*chagalld*1	shine,tender,calm
*corota*1	beautiful,calm,grand
*corotb*1	beautiful,delicate,calm
*corotc*1	grief,sombre,sober
*corotd*1	shine,beautiful,calm
*gogha*1	delight,shine,merry
*goghb*1	grief,sombre,terrible
*hokusaia*1	dynamic,motion,strong
*hokusaib*1	fight,motion,calm
*hokusaic*1	delicate,calm,quiet
*hokusaid*1	vivid,motion,speed
*loiranda*1	shine,grand,calm
*loirandb*1	delight,shine,calm
*loirandc*1	delight,grand,calm
*loirandd*1	quiet,substance,material
*nelsona*1	grand,dynamic,motion
*nelsonb*1	twilight,calm,quiet
*renoira*1	dim,tender,quiet
*renoirb*1	delight,dim,calm
*renoirc*1	loud,bustle,crowd
*renoird*1	fine,strong,quiet
*sarthoua*1	dynamic,motion,speed
*hiroa*1	twilight,grand,quiet
*hirob*1	cheer,dim,quiet
*hiroc*1	beautiful,quiet,calm
*hirod*1	fine,beautiful,shine
*hiroe*1	fine,beautiful,calm

English Dictionary." Method-2 is used for creating the metadata items for each image.

As the metadata items which are given as context words, we use several vocabulary entries shown in Table 2. These metadata items are used to express the user's impression. Each vocabulary entry is defined by using several explanation words listed in Table 2.

Table 2 Vocabulary entries used as context words

vocabulary entry	explanation (features)
brave	ready to go into danger, having self-control in danger or pain
clear	able to be seen through
cloud	(a mass of) mist high in the sky
dark	with little or no light
depress	make sad, unhappy
dynamic	of physical power, forces producing motion
excite	get (person, condition) worked up
glad	pleased, happy
grace	quality of being beautiful, sp. harmony of structure or motion
happy	full of pleasure, pleased
interesting	causing feeling of interest
light	that which makes things able to be seen
merry	laughing, happy, bright
might	great power, force
natural	of, forming a part of, in agreement with, nature or one's nature
pale	(sp. of face) having little colour
power	quality of being strong enough, able, to do something, sp., physical force
sad	unhappy
simple	of one substance, unmixed, without division into parts
stir	put or become in motion, make or become awake
strong	having, using, marked by, great, sp., physical, force

Table 3(a): The order of selected images in the context (power, strong).

context: power, strong

rank	image data	correlation*
1	hokusaia1	0.377136
2	renoird1	0.345938
3	nelsona1	0.306403
4	sarthoua1	0.303013
5	hokusaib1	0.262247
6	renoirc1	0.261547
7	loirandd1	0.221684
8	loirandc1	0.219976
9	hokusaid1	0.208567
10	corota1	0.206876
11	chagalla1	0.202815
12	loiranda1	0.202088
13	gogha1	0.201686
14	renoirb1	0.199490
15	loirandb1	0.195207
16	goghb1	0.193655
17	chagallb1	0.193655
18	corotc1	0.188142
19	hiroa1	0.176600
20	hiroc1	0.176152
21	chagalld1	0.175089
22	corotd1	0.172304
23	hirod1	0.167228
24	renoira1	0.156583
25	hirob1	0.156354
26	hiroe1	0.155545
27	nelsonb1	0.147163
28	chagallc1	0.139549
29	hokusaic1	0.136931
30	corotb1	0.130039

6.2 Experimental results

The experimental results are shown in Table 3(a), 3(b) and 3(c). In each experiment, 30 image data names are listed in the order of the selected images in each context. The function $\bar{\eta}_0\left(\mathbf{x}; s_\ell\right)$ is used for computing the norm of each image. The value of correlation in each image represents the norm of vector in the selected subspace. The norm of the vector is normalized as shown in Section 3.4. The values of correlation have only comparative significance for a given computation $\bar{\eta}_0\left(\mathbf{x}; s_\ell\right)$.

(1) Case-A(Table 3(a))

Table 3(b): The order of selected images in the context (light, strong).

context: light, strong

image data	correlation*
renoird1	0.313255
hokusaia1	0.272313
sarthoua1	0.191604
chagalla1	0.187024
gogha1	0.186192
nelsona1	0.184015
hokusaib1	0.183724
loirandd1	0.180893
loirandb1	0.177737
hokusaid1	0.175774
hiroa1	0.172693
loirandc1	0.168848
loiranda1	0.166349
renoirc1	0.163451
corota1	0.154491
chagallb1	0.150202
goghb1	0.150202
nelsonb1	0.149441
hiroc1	0.147595
renoirb1	0.146058
hirod1	0.145653
corotd1	0.144162
chagalld1	0.138508
hirob1	0.136352
renoira1	0.127080
hiroe1	0.123317
corotc1	0.117761
hokusaic1	0.114034
chagallc1	0.106145
corotb1	0.090652

Table 3(c): The order of selected images in the context (happy, glad).

context: happy, glad

image data	correlation*
gogha1	0.270147
hirob1	0.267835
hokusaic1	0.220439
loirandc1	0.213378
renoira1	0.209098
loirandb1	0.200794
corotb1	0.198720
hiroc1	0.197461
renoirb1	0.196746
renoirc1	0.194097
loirandd1	0.188176
chagalla1	0.187586
loiranda1	0.179469
chagalld1	0.179020
corota1	0.178073
nelsona1	0.170461
nelsonb1	0.168503
hokusaid1	0.167831
corotd1	0.165786
renoird1	0.165509
hiroa1	0.161292
sarthoua1	0.155030
hokusaia1	0.152708
hiroe1	0.152525
hokusaib1	0.146644
hirod1	0.142228
chagallc1	0.135303
goghb1	0.103767
chagallb1	0.103767
corotc1	0.088858

As shown in the experimental result, the image data named "hokusaia1" is in the first ranking (rank 1). This result reflects the recognition of the context "power, strong" to this image.

(2) Case-B(Table 3(b))

The image data named "renoird1" is in the first ranking (rank 1). This result reflects the recognition of the context "light, strong" to this image.

(2) Case-C(Table 3(c))

The image data named "gogha1" is in the first ranking (rank 1). This result reflects the recognition of the context "happy, glad" to this image.

7 AUTOMATIC CREATION OF IMAGE METADATA

In the semantic associative image search, each image is characterized in an n-dimensional vector as the metadata item in Methods 1 and 2. Currently, it is assumed that the designer of metadata checks each image data item, one by one, and defines n-dimensional vectors which represent image data items subjectively. As each image data item is defined subjectively, it causes individual variations. When several designers define metadata for each image data item, we need some sort of guideline.

It is advantageous to provide a method which calculates correlations between an image data item and impression words for Kurita and Kato et al have introduced a method which uses canonical correlation analysis[8]. In this method, 33 autocorrelation coefficients based on the color components, i.e. red, blue and green, of the images data item define its characteristics. They selected 30 adjectives as impression words, and the impression is expressed by selecting the corresponding adjectives. To create the space for calculation, learning is performed. It is done by showing sample images to testees and asking them to select most appropriate impression words. Then canonical correlation analysis is performed to make possible to calculate a space which allows to measure correlation between image vectors and impression word vectors. In this space, a corresponding impression word vector can be estimated from a image word vector.

In this section, we show a simple and effective method for generating metadata items for image data. In the algorithm shown in this section, impression words of each image data item are derived via the color space. Impression words of image data are not derived directly from the image data themselves. We use colors as intermediates to derive impressions of images. First, we discuss how an image is transformed to its color representation. Then, we discuss how impression words are derived from the color representations of images. Finally, we present the method to derive impression words from image data items with its example.

7.1 Creation of Metadata

The procedure to extract impression words of an image is as follows:

1. **Color Transformation:** Transform color representation of an image I from the RGB color space to the Munsell color space by the MTM.
2. **Clustering:** Cluster each color component of the image into p clusters of colors. As the result, the image is represented in the form of image vector P.
3. **Derivation of Impression words:** Derive impression words from the image by calculating correlations between the image vector and impression word vectors.
4. **Creation of Metadata for Image:** The derived impression words are used to create metadata for the image I by Method-2 defined in section 5.2. The impression words for the image I are represented as the metadata item of the image by the operator union $\bigoplus$.

In the rest of the section, each of the steps is described in detail.

7.2 Impressions of images and colors

The impressions of images are derived from several factors. It is known that the color component of images is one of the dominant factors which affect the impressions of images strongly. That is, we can approximate the impressions of images from the color components. Concerning the color components of an image, we must consider the following points:

1. Colors appearing in the image,
2. Ratio of the area size of each color to the whole image.
3. Allocation of colors in the image,

In our current method, only the first two points are considered to define impressions of the images. It is assumed that the impression of an image roughly depends on colors appearing in the images, and it is fixed by complex combinations of various impressions derived from each color component. Allocation of each color appearing in an image is also important. In this method, we do not consider the structural information of an image. Therefore, the image I which consists of q color components is defined as follows:

$$I = \{c_1, c_2, ..., c_q\}.$$

Where c_i is the size of the area which the color components i occupies. It is defined as follows:

$$c_i = \frac{\text{Total area size of the color component } i}{\text{Total area size of the image } P}, \quad \text{where } i = 1, 2, ..., q.$$

To derive the impressions of images from color components, colors of images should be represented in psychological color solid[10]. Next, we discuss how the physical colors of the images are transformed to the psychological colors.

Psychological color solid

Normally, colors of image data items are expressed in the RGB color system. Each color in this color system is expressed as a mixture of three basic colors, Red, Green and Blue.

Although it is a convenient color system to handle colors, there is a gap between the color representation in the RGB color system and the psychological color representation. To deal with colors and their impressions, the psychological color solid, such as the well-known Munsell color system, is used. The Munsell color system identifies each color in three dimensions, (H, V, C) where H is the color, V is the lightness or darkness of the color, and C is the dullness or purity of the color[10]. This three-dimensional space has psychologically linear scales.

The transformation between the RGB color representation and the Munsell color representation can be performed mathematically by the MTM (Mathematical Transformation to Munsell)[11]. The MTM uses the Adams color space[1] to mediate the transformation between two color systems as $L^*a^*b^*$ of CIE. In our algorithm, color components of image data are transformed to the Munsell color data by the MTM.

Clustering of colors

Many colors are usually used in a single image, and they produce various impressions. To clarify these impressions, colors should be grouped into several meaningful clusters of colors.

To cluster the color components, it is necessary to calculate the difference between any two colors. In the Munsell color system, the color-difference between two Munsell colors (H_j, V_j, C_j) and (H_k, V_k, C_k) is calculated with the following Godlove color-difference formula[2]:

$$\Delta E = \sqrt{2C_jC_k\{1 - \cos(\frac{2\pi\Delta H}{100})\} + (\Delta C)^2 + (4\Delta V)^2},$$

where ΔH, ΔV and ΔC are expressed as follows:

$$\begin{aligned} \Delta H &= |H_j - H_k|, \\ \Delta V &= |V_j - V_k|, \\ \Delta C &= |C_j - C_k|. \end{aligned}$$

First, cluster centers are selected for each color component. By calculating the color-difference between these cluster centers and each color component, the cluster which the color component belongs to is fixed. Each color component belongs to the closest cluster, that is, the one with the smallest ΔE.

Then, the image I is defined as a p-dimensional vector, where p is a number of clusters. The p-dimensional vector is image vector P and defined as follows:

$$P = (v_1, v_2, ..., v_p),$$

[1]It is also known as ULCS(Uniform Lightness-Chromaticness Scale)

where v_i is a summation of c_j of color components in the cluster i. The impressions of the image I are derived from this vector P.

7.3 Color and impression

In the field of experimental psychology, many word association tests have been done on colors[18, 19, 20, 21]. For instance, to derive the association between colors and words, some experiments have been done by showing a single color to numbers of testees and asking them the impression words.

The results of such experiments show that there are associations between words and colors. For instance, the color 'strong orange' (e.g. 5YR7/14 in the Munsell color system) is likely to be selected as the associated color to the word 'warm'[18]. To describe the the impression words in the notion of colors, the results of these psychological experiments can be used effectively.

Additionally, those psychological experiments have shown that there are two types of associations. One type of association depends on individual mental background such as a culture. The other type of association is extracted by using universally common background. In image databases as the shared data resources, an image must be explained in the common impressions. By defining universally common impressions in terms of colors, the impressions can be acceptable.

As we use words to describe these impressions in our model, we call the words which describe impressions *impression words*. Assume that there are s impression words to describe the impressions of colors. The impression word W_j can be represented as a p-dimensional vector as follows:

$$W_j = (u_1, u_2, ..., u_p) \qquad \text{where } j = 1, 2, ..., s,$$

where p is the number of clusters as described in Section 7.2, u_i is a correlation coefficient between the cluster center (the color) of the cluster i and the impression words W_j, and s is the number of impression words for representing images.

7.4 Correlation between image vectors and impression word vectors

Impressions of the images are derived by finding impression words which correspond to the images. We have discussed how images and impression words are represented in the form of p-dimensional vectors. Correlation between images and impression words are calculated with these vectors.

Correlation is calculated by the inner product of each impression word vector W_j and the image vector P:

$$\mathcal{C}_j = (W_j, P) \quad \text{where } j = 1, 2, ..., s.$$

The impression word vectors which have a high correlation with the image vector P are selected, and the corresponding impression words to the selected vectors are extracted as the impressions of the images.

7.5 Examples

Table 4 and Table 5 are definitions of the impression word vectors and the image vectors, respectively. Table 4 is created from the psychological experiments[18]. Each impression word corresponds to an emotional property of colors. Two impression words are selected for each emotional property of colors to describe the opposite meanings. We selected seven colors, red, yellow, green, blue, purple, white and black, as the primitive colors which we have set for describing impressions. Selected cluster centers, therefore, are 5R/14 (red), 5Y8/12 (yellow), 7.5G5/8 (green), 10B4/8 (blue), 10P4/8 (purple), N9 (white) and N1 (black). As the example, we use eight images painted by Marcestel[22]. The images are shown in Figure 3. Although the images have the same motif of bouquet, each image is different in color components. Table 5 shows the definition of image vectors. These image vectors are obtained by the method described in Section 7.2.

By computing correlations between the image vectors and the impression word vectors, the impression words shown in Table 6 are obtained for each image data item. The correlation coefficients in this table are the results of the inner product between impression word vectors and image vectors. Bouquet1, for instance, has a blue color as its basic background color. The obtained impression words are satisfactory words to describe its impressions.

Table 4 Impression words and their vectors

Impression Words	Red	Yellow	Green	Blue	Purple	White	Black
showy	1.892	1.752	0.775	-0.508	-0.879	0.188	-1.083
sober	-1.892	-1.752	-0.775	0.508	0.879	-0.188	1.083
natural	0.000	0.318	1.807	0.339	-1.319	1.032	0.083
artificial	-0.000	-0.318	-1.807	-0.339	1.319	-1.032	-0.083
warm	1.558	1.035	-0.258	-2.202	-1.026	-0.375	0.083
cold	-1.558	-1.035	0.258	2.202	1.026	0.375	-0.083
bright	1.224	1.752	1.549	-1.016	-1.319	1.314	-1.833
dark	-1.224	-1.752	-1.549	1.016	1.319	-1.314	1.833
intellectual	-1.558	-0.637	1.033	1.016	-0.147	1.032	0.583
emotional	1.558	0.637	-1.033	-1.016	0.147	-1.032	-0.583
soft	0.000	0.717	0.516	-1.186	-0.879	0.657	-1.333
hard	-0.000	-0.717	-0.516	1.186	0.879	-0.657	1.333
light	0.000	1.274	0.775	-0.678	-1.759	1.032	-1.583
grave	-0.000	-1.274	-0.775	0.678	1.759	-1.032	1.583
plain	-0.891	0.239	-0.516	-0.508	-1.319	1.595	-0.333
insistent	0.891	-0.239	0.516	0.508	1.319	-1.595	0.333
strong	1.670	1.433	1.549	1.863	0.733	-0.282	1.333
feeble	-1.670	-1.433	-1.549	-1.863	-0.733	0.282	-1.333
favorite	0.445	0.478	0.775	0.339	-0.879	1.314	0.917
disliked	-0.445	-0.478	-0.775	-0.339	0.879	-1.314	-0.917
static	-1.336	-1.035	-0.516	1.016	0.586	1.032	0.750
dynamic	1.336	1.035	0.516	-1.016	-0.586	-1.032	-0.750
cheerful	1.224	1.274	-0.258	-1.016	-1.319	0.375	-1.083
gloomy	-1.224	-1.274	0.258	1.016	1.319	-0.375	1.083
beautiful	0.557	0.478	1.033	0.169	-0.440	1.314	0.333
ugly	-0.557	-0.478	-1.033	-0.169	0.440	-1.314	-0.333
manly	-0.891	-0.239	1.033	1.355	-1.026	-0.469	2.000
womanly	0.891	0.239	-1.033	-1.355	1.026	0.469	-2.000
vivid	0.891	1.354	1.807	1.016	-0.293	0.845	1.000
vague	-0.891	-1.354	-1.807	-1.016	0.293	-0.845	-1.000
clear	0.111	0.955	1.033	0.508	-0.879	1.408	-0.083
muddy	-0.111	-0.955	-1.033	-0.508	0.879	-1.408	0.083
simple	0.223	0.955	1.033	0.508	-1.319	1.314	-0.250
complicated	-0.223	-0.955	-1.033	-0.508	1.319	-1.314	0.250
graceful	-0.223	0.159	0.258	0.678	0.000	1.408	0.750
vulgar	0.223	-0.159	-0.258	-0.678	-0.000	-1.408	-0.750
new	0.445	0.717	0.516	0.169	-0.879	0.751	-0.250
old	-0.445	-0.717	-0.516	-0.169	0.879	-0.751	0.250
excited	0.891	0.796	-0.516	-1.016	-1.026	0.000	-0.750
melancholy	-0.891	-0.796	0.516	1.016	1.026	-0.000	0.750

`bouquet1`

`bouquet2`

`bouquet3`

`bouquet4`

`bouquet5`

`bouquet6`

`bouquet7`

`bouquet8`

Figure 3 Image Data Items

Table 5 Image vectors

Image Data (Title of the Image)	Red	Yellow	Green	Blue	Purple	White	Black
Bouquet1 (The Birth of Love)	0.0363	0.1611	0.0087	0.4494	0.0798	0.2262	0.0385
Bouquet2 (Marbled Flowers)	0.2279	0.4124	0.0037	0.0004	0.0039	0.3446	0.0071
Bouquet3 (Bouquet of Remembrance)	0.0034	0.6351	0.2100	0.0008	0.0016	0.1489	0.0003
Bouquet4 (The Extraordinary Bouquet)	0.3888	0.1161	0.0038	0.0182	0.1073	0.3289	0.0370
Bouquet5 (The Resting Butterfly)	0.0011	0.0268	0.0065	0.0003	0.0023	0.5854	0.3777
Bouquet6 (Outburst of Happiness)	0.0605	0.3885	0.0060	0.0129	0.2416	0.2744	0.0163
Bouquet7 (The Closeness of Your Bouquet)	0.0840	0.2443	0.4436	0.0003	0.0483	0.1783	0.0013
Bouquet8 (The Radiant Bouquet)	0.3769	0.0233	0.0019	0.0005	0.00100	0.2521	0.3444

The metadata items for images are created with the obtained impression words by the union operator of Method-2.

8 CONCLUSION

In this chapter, we have introduced new methodology for retrieving image data according to the user's impression and the image's contents. We have presented functions and metadata for performing semantic associative search for images. The functions are realized on the basis of mathematical model of meaning.

For the creation of the metadata for images, we have introduced three methods (Method-1, Method-2 and Method-3). The metadata created by those methods is categorized in the type of *content-descriptive metadata* according to the metadata classification for digital media presented in [15]. Furthermore, the metadata created by the first two methods, Method-1 and Method-2, is categorized into the content-descriptive domain-dependent metadata, and the metadata by the third method is classified as the type of content-descriptive domain-independent metadata.

Table 6 Impression words derived automatically from the images

Images	Order	Impression Words	Correlation Coefficient
Bouquet1	1	strong	1.189
	2	cold	0.932
	3	vivid	0.929
	4	graceful	0.672
	5	clear	0.640
Bouquet2	1	bright	1.441
	2	showy	1.210
	3	vivid	1.065
	4	cheerful	0.920
	5	strong	0.893
Bouquet3	1	bright	1.634
	2	vivid	1.368
	3	showy	1.307
	4	strong	1.202
	5	light	1.122
Bouquet4	1	strong	0.891
	2	bright	0.890
	3	showy	0.860
	4	vivid	0.812
	5	beautiful	0.676
Bouquet5	1	favorite	1.132
	2	graceful	1.113
	3	vivid	0.921
	4	beautiful	0.914
	5	static	0.857
Bouquet6	1	strong	0.812
	2	vivid	0.781
	3	bright	0.762
	4	showy	0.614
	5	beautiful	0.487
Bouquet7	1	bright	1.386
	2	vivid	1.345
	3	strong	1.165
	4	natural	1.000
	5	showy	0.920
Bouquet8	1	strong	1.055
	2	vivid	0.928
	3	favorite	0.826
	4	beautiful	0.668
	5	warm	0.543

We have implemented the semantic associative search system to clarify its feasibility and effectiveness. Currently, we are designing a learning mechanism to adapt metadata for context representation and images according to the individual variation. The learning is a significant mechanism for semantic associative search, because the judgment of accuracy for the retrieval results might be dependent on individuals. In the learning, if an inappropriate retrieval result for a request is extracted, appropriate image data which must be the retrieval results is specified as suggestions. Then, the learning mechanism is applied to the semantic associative search system for adapting the metadata for context representation and images. The learning mechanism can be used to create individual semantic search environments by adjusting individual metadata for context representation to the semantic space.

We will use this system for realizing a multimedia metadatabase environment. As our future work, we will extend this system to support multimedia data retrieval for video and audio data. This system will be integrated in a distributed multimedia database environment[5, 6].

Acknowledgments

Special thanks to Amit Sheth and Wolfgang Klaus for their constructive and valuable comments and suggestions on this work. We would like to thank W. W. Chu for his valuable comments. We would also like to thank Marcestel who kindly gave us the opportunity for using his images in the experiments.

REFERENCES

[1] Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W. and Harshman, R. A., *"Indexing by Latent Semantic Analysis,"* Journal of the American Society for Information Science, Vol.41, No.6, pp.391–407, 1990.

[2] Godlove, I.H., *"Improved Color-Difference Formula with Applications to the Perceptibility and Acceptability of Fadings"*, Journal of the Optical Society of America, Vol.41, No.11, pp.760–772 Jul. 1951.

[3] Kitagawa, T. and Kiyoki, Y., *"A Mathematical Model of Meaning and its Application to Multidatabase Systems,"* Proc. 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130–135, April 1993.

[4] Kiyoki, Y., Kitagawa, T., and Hayama, T., *"A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning,"* ACM SIGMOD Record (Special issue on metadata for digital media), W. Klaus, A. Sheth, eds., Vol.23, No. 4, pp.34–41, Dec. 1994.

[5] Kiyoki, Y. and Hayama, T., *"The Design and Implementation of a Distributed System Architecture for Multimedia Databases,"* Proc. 47th Conference of International Federation for Information and Documentation, pp.374–379, Oct. 1994.

[6] Kiyoki, Y., Kitagawa, T. and Hitomi, Y., "*A Fundamental Framework for Realizing Semantic Interoperability in a Multidatabase Environment*," Journal of Integrated Computer-Aided Engineering, Vol.2, No.1(Special Issue on Multidatabase and Interoperable Systems), pp.3–20, John Wiley & Sons, Jan. 1995.

[7] Kurita, T., Kato, T., *"Learning A Cognitive Schema for Multimedia Indexing –Experiments on Image Database–,"* Technical Reports of IEICE, DE93–3, May 1993.

[8] Kurita, T., Kato, T., Fukuda, I., Sakakura, A., "*Sense Retrieval on a Image Database of Full Color Paintings*," Journal of Japan Information Processing Society, Vol.33, No.11, pp.1373–1383, Nov.1992.

[9] *"Longman Dictionary of Contemporary English,"* Longman, 1987.

[10] Newhall,S.M., Nickerson,D. and Judd, D.B., *"Final Report of the O.S.A Subcommittee on the Spacing of the Munsell Colors"*, Journal of the Optical Society of America, Vol.33, No.7, pp.485–418 Jul. 1943.

[11] Miyahara, M. and Yoshida, Y.,*"Mathematical Transformation of (R,G,B) Color Data to Munsell (H,V,C) Color Data"*, SPIE's Visual Communications and Image Processing '88, Vol.1001, No.118, pp.650–657, Nov. 1988.

[12] Ogden, C.K., *"The General Basic English Dictionary,"* Evans Brothers Limited, 1940.

[13] Sheth, A. and Larson, J.A., *"Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases,"* ACM Computing Surveys, Vol.22, No.3, pp.183–236, 1990.

[14] Sheth, A., *"Semantic Issues in Multidatabase Systems,"* ACM SIGMOD Record, Vol.20, No.4, pp.5–9, 1991.

[15] Kashyap, V., Shah, K., Sheth, A., *"Metadata for Building the Multimedia Patch Quilt,"* V.S. Subrahamanian, S. Jajodia, eds., Multimedia Database Systems, pp.297–319, 1996.

[16] Chu, W.W., Hsu, C.C., Ieong, I.T., Taira, R.K., *"Content-Based Image Retrieval Using Metadata and Relaxation Techniques,"* The Digital Media(This book), W.Klaus, A. Sheth, eds., McGraw-Hill, 1997.

[17] Shimizu, H., Kiyoki, Y., Sekijima, A. and Kamibayashi, N., *"A Decision Making Support System for Selecting Appropriate Online Databases,"*, Proc. 1st IEEE International Workshop on Interoperability in Multidatabase Systems, pp.322–329, Apr. 1991.

[18] Chijiiwa, H., *"Color Science"*, Fukumura Printing Co., 1983.

[19] Russel,J.A., Mehrabian,A., *"Evidence for a Three-Factor Theory of Emotions,"* Journal of Research in Personality, Vol.11, pp.273–294, 1977.

[20] Mehrabian,A., *"Measures of individual differences in temperament,"* Educational and Psychological Measurement, Vol.38, pp.1105–1117, 1978.

[21] Valdez,P., Mehrabian,A., *"Effects of Color on Emotions,"* Journal of Experimental Psychology: General, Vol.123, No.4, pp.394–409, 1994.

[22] Marcestel, *"Marcestel,"* Kyoto Shoin Co., 1993.