An Image Retrieval Method Based on a Genetic Algorithm Controlled by User's Mind

Syuko KATO

This paper describes an image retrieval method based on an interactive Genetic Algorithm controlled by the minds of individual users. The method dynamically enhances the ability to reflect the mind of the user in the retrieval process and improves the retrieval efficiency. By using this method, without explicit representation of the desired image, the image is nevertheless induced through feature selection based on the diverse, fragmentary, and ambiguous subjective intent of the user. This points the way to a powerful system that is capable of guessing the mind of the user and is not dependent on overt representation of images.

Keywords
Image retrieval, Genetic Algorithm, Interaction, Subjectivity, Feature selection

1 Introduction

We live in a world flooded with an ever rising tide of information, making information retrieval technologies more important than ever before in helping users find their way in a user-friendly and efficient manner. Since users see a variety of implications in an image when searching for images, a method for making nonsuperficial inferences regarding each user's intentions is required, particularly when a system handles abundant and diversified images.

Statistical and liner algebraic methods have been often used in conventional research on image retrieval methods, which seek to assess user intentions. One method involves entering language keys, as in the case of picture retrieval[1] based on the entry of "impressional words" and on the entry of image keys, as in the case of similar image retrieval[2][3] for trademark designs and butterfly patterns. However, those methods entail calculations of the relations between language keys and image features and a priori weighing of image features based on training images in order to reflect the user's mind. Such approaches fare poorly when user intentions are unclear or change during image retrieval.

In recent years, techniques intended to facilitate interactions between the user and the retrieval system have received increasing attention. Some reports[4][5][6][7] indicate that it is effective to directly integrate the user's subjective valuation directly into the retrieval system; this method is not limited to just image retrieval. In particular, genetic algorithms (GA), an evolutionary interactive method, can be used to automatically provide very effective solutions by assigning a fitness value to images, based on user estimations, without requiring the definition of complex estimation functions. This technique has been explored for image retrieval applications such as photomontages of wanted crime suspects[8], nature images retrieval[9], and line drawn portrait retrieval[10]. It is regarded as an effective method for users who cannot articulate their intentions in the form of keys before image retrieval, but can estimate each presented images.

Nevertheless, the conventional image retrieval methods have at least the following two problems in terms of the expression and handling of image features from large-scale, diversified image databases.

First, they limit the range of image features according to application area. In most cases, they use only physical and graphical image features[11] such as color, geometry, and position that can be directly derived from the image; or even when using feature expressions of the semantic layer[11], such expressions are primarily limited to so-called *Kansei* words, or adjectives. However, an individual user's impressions of a desired image clearly cannot be depicted by limited sets of feature expressions. For example, the user may connote a sense of relaxation when seeing a picture of a lying dog, or the picture of a falling apple may bring to mind a typhoon or "flash," due to associations unrelated to color or *Kansei* words. We must not neglect the role of feature expressions of the semantic layer that indicate meanings suggested and derived from an image indirectly. Despite technological challenges to the automatic acquisition of feature expressions of the semantic layer, the required retrieval mechanism must be able to handle a variety of features involving the physical, graphical, and semantic layers in order to develop an image retrieval method capable of deep inferences into complex human minds.

Second, conventional methods fail to address image features as fragmentary expressions typical of complex mental states. In most cases, the relationship between an image and an image feature is uniquely determined with uniform feature expressions; all explicit image features are counted and distances calculated in the feature space reflecting the subjective amounts (quantified subjective weights) to infer the desired image. However, large-scale and diversified image databases are more likely to have non-uniform feature expressions for each image, including those involving the semantic layer. The calculation of distance in a large feature space reduces retrieval efficiency, with no offsetting advantages. In addition, this approach ignores the implications of images that are not successfully described, and the range of inferences based only on explicit feature expressions is necessarily limited. During image retrieval, the user may not have in mind a clear image; may provide poor expressions of desired images; experience different sensory associations that those registered in the system; or have in mind more than one solution, with their images being ambiguous, fragmented, diversified, and multimodal – that is, complex. The image features expressed by the system are only fragments reflecting limited aspects of the user's mental state. Although we believe it is possible for the user to consciously control the image retrieval process, we believe that the complexities of the mental states cannot be regarded as static quantities, and are not subject to precise expression. After all, we constantly have a "change of mind". To approach the human mind, we need an approach that can dynamically infer the user's intentions based on information derived from an interactive process between the user and the system, while leaving ambiguities in place. In this sense, we call such an image retrieval method an "image retrieval method controlled by user's mind."

Keeping in mind these aspects, we propose an image retrieval method, which tolerates the imperfection of feature expressions, based on an interactive GA[12][13][14][15] that can address the image features of the physical, graphical, and semantic layers. In our proposals, we show that the results of feature selection respond flexibly according to the user's implicit expectations[12] when the user estimates the images presented by the system. In the body of research concerned with exploring various levels of feature expressions, there is a layered model[16] that is primarily based on the perspective of the semantic layer. Our approach differs from this approach in that we do not conduct separate matching operations for each level of feature expressions, but implement matching at an integrated level. Even our approach has problems providing the matching images required for user estimations, and image searches can thus be quite time-consuming, although performance will greatly depend on user estimations based on perusal of many images. Furthermore, our approach relays on the emergent retrieval capability of GA to guess user's mind, only reflecting user estimations to image fitness.

In this paper, we propose a technique that improves on the method of the reference[12]. To control the image retrieval process based on the user's mind, we increase the interaction between the user and the system by allowing the user to change his/her estimation of the presented image and the number of presentable images at any time, even during an image search. The user's intentions are faithfully tracked by applying the information obtained from such interaction to a crossover technique. Our method also boosts retrieval efficiency by using information obtained during the retrieval history and interactions involving the acquisition and presentation of matching images.

In the following sections, we will provide an outline of the image retrieval method touched on in Section 2, describe its effectiveness in reflecting user intentions. Section 3 describes a proposed method for improving retrieval efficiency, while Section 4 discusses computer simulations and results. In this paper, we will refer to the method we have developed thus far as the "basic method," while referring to the improved basic method as the "proposed method." When no distinction needs to be made, we will refer to both simply as "this method."

2 Image Retrieval Method Based on a Genetic Algorithm

This technique executes image retrieval while selecting features, using an interactive GA technique to control the image retrieval process based on the user's mind. Each user's intention is interactively taken into the system as "environment," while GA controls the selection results through numerous generations. These results are internally represented as an "individual" indicate clues of the user's intention. This makes possible retrieval of images that match the user's intentions. Concretely, the user answers only questions involving the images presented (YES, NO, UNDECIDED) during the retrieval process. This method is a browsing type image retrieval technique by which the presented images become the retrieval results. The goal of this retrieval process is not to search for a best image, but to find more than one appropriate image. If the user wishes to enter search conditions at the start of search, he or she may enter an arbitrary feature or a combination of features of the images (logical product) defined by the system as an initial key. Unless the user thinks of a search condition, he/she is not required to enter any.

This method emphasizes the following two points, as described in Section 1. First, it is capable of handling all the features of the physical, graphical, and semantic layers; second, the feature selection results are matched with images based not on distance, but a set relationship. This method of determining correlation based on set relationship allows imperfections in expression and does not restrict implicit meanings. The following is a brief explanation of the design method[12].

(1) Coding

As each emerging chromosome describes the result of feature selection in GA, the gene, the smallest operational unit in GA, is defined as a feature, namely, a pair of a feature axis and a feature value. When $n(n \geq 1)$ genes have been collected, a chromosome is formed and becomes the feature selection result. The length of a chromosome n is determined when the user enters the initial key. One generation consists of as many as m ($m \geq 1$) chromosomes. The population size m may change.

In this method, since a feature is defined as a pair of a feature axis and a feature value each labeled with an integer, the features of all the physical, graphical, and semantic layers can be treated as genes. This facilitates handling a variety of features by labeling new features with integers and adding them to the already defined features, although in practice this approach is limited by computing resources. Moreover, the contents of the feature definition can be highly flexible: representing the combined intentions of the system designer, index designer, and the characteristics of each feature, a single feature axis allows one or more feature values. A feature value defined for a feature may be defined as the feature axis of another feature. Although we have determined that the chromosome length is a constant from a physical viewpoint, all the defined features need not occur on a single chromosome. Since a chromosome may have more than one of the same gene in it, chromosome length is variable from a logical viewpoint.

(2) Initial population

An initial population is composed from the initial key entered by the user at the start of the search, as a retrieval condition can be described by the initial population in GA. If the user does not enter an initial key, up to *m* chromosomes are randomly generated and used as the initial population. If the user does enter an initial key, up to *m* chromosomes made from chromosomes reflecting an arbitrary number of entered features (pairs of the feature axis and the feature value) and randomly generated chromosomes are used as the initial population.

(3) Genotype and phenotype

The relationship between the genotype and phenotype in GA is analogous to that between feature expression and image in image retrieval. Although an image is an object/entity of which a feature can be uniquely described as a collection of pixels, this method is based on the assumption that an image can be endowed infinite meanings by the complex user's mind, and that the feature expression and image are related by set theory. The genotype chromosome is regarded as a logical product of features, which becomes the feature selection result providing clues of user intention. Subsequently, all images of features, including such feature selection results, are regarded as phenotype images and provided as matching images corresponding to user intention.

(4) Fitness

Fitness is assigned to each chromosome. GA forms populations, placing priority on the chromosomes of the highest fitness, and proceeds to evolve generations. To select features that provide clues for inferring user's subjective intentions, fitness must reflect user's minds. Thus, in assigning fitness, an interactive method is adopted to seek out the user's direct estimation of system during retrieval. At this time, it is important for the system to present output in the form of an image (phenotype). Images having the features including all the features represented in chromosomes that match the given chromosomes are shown to the user, and the user estimates each image to determine image fitness. Note that fitness must be defined for the case in which more than one image matches a single chromosome. For convenience, the average of the fitness values assigned to matched images is adopted as the fitness of the corresponding chromosome. Regardless of user estimations, the fitness value is determined to provide a higher value for an image more likely to match. This is done for the sake of convenience. Table 1 lists six types of fitness corresponding to each condition. Introducing the concept of image fitness degree allows us to automatically provide intermediate fitness. Since image fitness is determined to be 25% when the user estimation is NO, the same effects are obtained as the bias fitness[17], which does not provide explicit differences based on estimations.

(5) Genetic operation

Chromosomes to which fitness are assigned in the population (characterized as *m*) are subject to genetic operations such as selection, crossover, and mutation. For selection, the number of the chromosomes of elite selection (characterized as *e*) are first taken in descending order of fitness. For chromosomes

Table 1 Type of fitness

Type of fitness No.	Type of fitness	User estimation	Situation	Value (%)
1	Image fitness	Present	Image for which the user estimation is YES	100
2	Image fitness	Present	Image for which the user estimation is NO	25
3	Image fitness	Present	Image for which the user estimation is UNDECIDED	75
4	Image fitness	Absent	Image for which the user estimation is thinned out	50
5	Fitness	Absent	Chromosome with no matching image	50
6	Fitness	Absent	Contradictory chromosome	0

other than elite chromosomes (m - e), the roulette method (selected based on probability proportional to fitness; overlap permitted) is applied. Chromosomes other than elite chromosomes are subjected to crossover. The pairs are removed in the order adopted during selection, and crossover operations are performed according to the crossover rate. Adopting the general idea of uniform crossover allows equal handling of all the genes in each chromosome. In the basic method, mutations occur only with chromosomes consisting of pairs of contradictory features (contradictory chromosomes) that cannot arise. A mask is created having the length of such a chromosome, and the determination of mutation according to the mutation rate is made based on such contradictory chromosomes. The mutation is then performed in the locus of the gene selected for mutation by randomly generating the feature axis and the feature value. The mutation operation is repeated until the features in the pair that appear in the chromosome do not contradict.

In the genetic operations adopted in this image retrieval method, selection is performed to select features that match user intentions. Crossovers are performed to obtain new pairs of features, while mutations are performed to retain features in pairs that do not contradict.

3 Proposal of Improved Basic Method

3.1 Outline

To achieve an image retrieval method controlled by the user's mind, we propose a technique that improves the degree of "subjectivity reflection" and the retrieval efficiency of the basic method described in Section 2. In the proposed method, the information obtained through interactions with the user is extensively deployed by a more interactive method. The outline of the proposed algorithm and its mechanism are shown in Fig.1. This image retrieval system consists of four major units: the image database unit, user interface unit, initial input processing unit, and image retrieval unit.

The image database unit stores/manages image data and extracts/manages information (image information) on the image data including defined features. It exchanges the required information with the initial input processing unit and the image retrieval unit. Large-scale image databases are built by procuring/controlling image data through networks and realized by expanding the image database unit.

The user interface unit is a site at which the system and the user interact directly. In the proposed method, the system accepts the initial key entered by the user and the historical information described in Section 3.3.1 as search conditions, as well as the user estimation of the images presented and the number of presentable images (browsing threshold) at any timing during retrieval, to achieve a flexible interactive operation. In addition to similar images or unregistered words, systems at some future date may be able to accept nonverbal information such as the viewpoints of the user and facial expressions. This would enhance user input flexibility and allow such a system to acquire even more mind information, which is the central factor for any control based on the user's mind. Future challenges will include further upgrades of the user interface and more sophisticated handling of

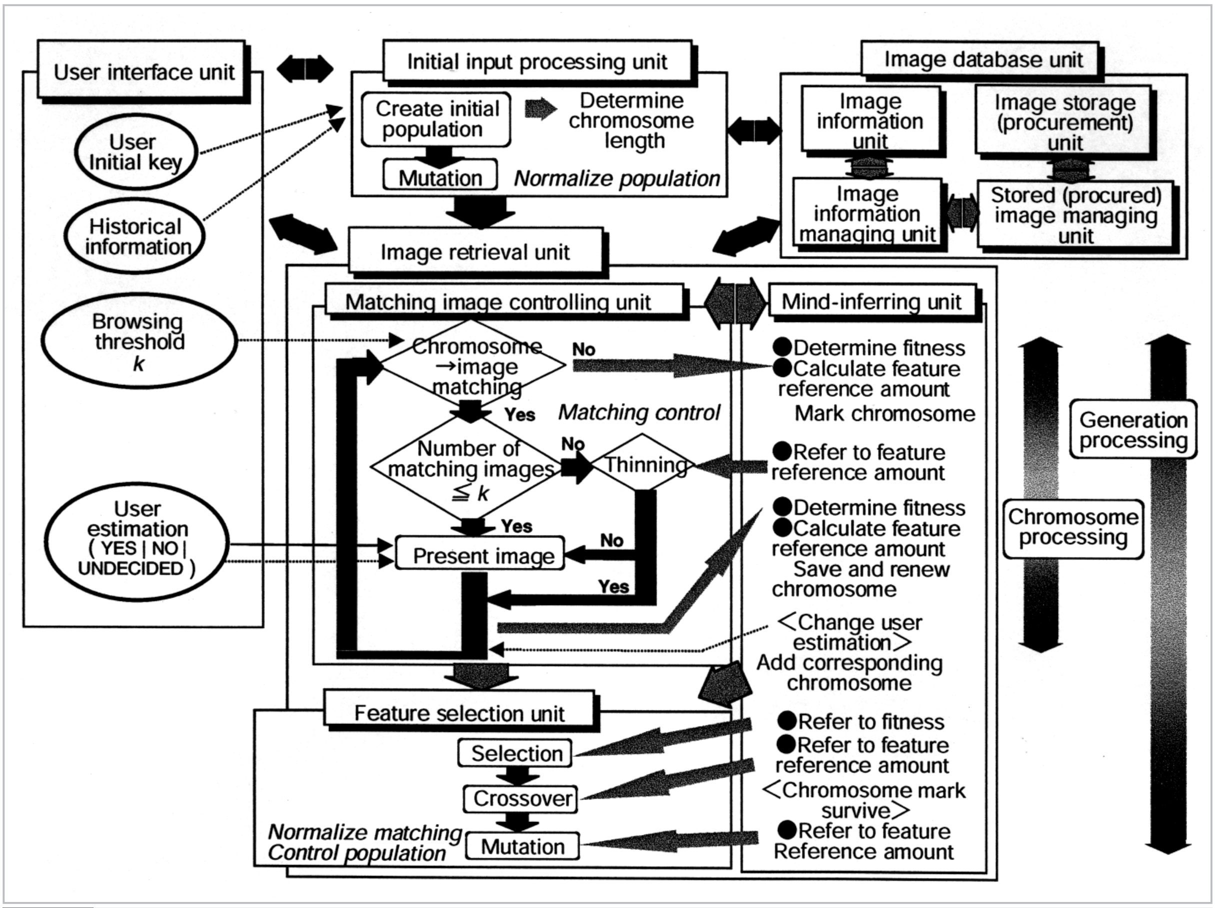

Fig.1 *Outline of the proposed method*

knowledge bases.

The initial input processing unit converts the search conditions entered through the user interface into a form used in GA. The proposed method generates an initial population based on the search conditions, determines chromosome length, and provides a normal population consisting of pairs of non-contradicting features for the image retrieval unit.

The image retrieval unit consists of a matching image control unit, feature selection unit, and mind-inferring unit, and executes the feature selection reflecting the user's mind by repeating generation processing and performing image retrieval. The proposed method defines the feature reference amount described in Section 3.2.1 to improve the way of reflecting the user's mind and applies this to thin out the matching images in the matching image control unit and in the crossover and mutation operations in the feature selection unit. The matching control unit repeats a series of chromosome processes consisting of chromosome-image correlating process, the matching image thinning process and image presenting process in proportion to the size of the population. The information provided by the user through interactive exchanges is processed as the control information reflecting the user's mind, and the management of the calculated fitness and feature reference amount are handled by the mind-inferring unit. The mind-inferring unit only manages and provides the fitness and feature reference amount; the proposed model does not require a specific model of mind. The feature selection unit performs selection, crossovers and mutations, referring to the fitness and feature reference amount managed by the mind-inferring unit after chromosome processing of one generation, in order to improve the subjectivity reflection and retrieval efficiency, and generate the next-generation population.

The following sections describe possible ways to improve improvement of subjectivity reflection (Section 3.2) and retrieval efficiency (Section 3.3.).

3.2 Improvement of subjectivity reflection

3.2.1 Definition of feature reference amount

In order to improve the accuracy of subjectivity reflection in the basic method and control image retrieval with the user's mind, the feature reference amount has been newly defined as a quantity that reflects the user's mind. The feature reference amount is a statistical quantity that reflects not just the user estimation of each image presented by the system upon the current chromosome in a form of fitness, but also strongly reflects it within the genes comprising the chromosome. The feature reference amount is a quantity added to each feature defined in the feature space with an initial value of zero, which emerges as genes across generations. When each feature emerges as a gene, the feature reference amount is renewed with reference to the fitness assigned to the chromosome upon its emergence. In the research for this paper, the following two calculation methods were studied to calculate the feature reference amount. One is a method for adopting the average of the fitness values obtained until the feature has emerged as a gene, as the feature reference amount (hereafter, Fa-value) of the feature; the other is a method for adopting the largest of the fitness values as the feature reference amount (hereafter, Fm-value) of the feature.

In reflecting the user's mind, combinations of features rather than a single feature often have greater significance. This is because the user changes the priority placed on each feature during feature combination, and because the features each play one role among supported feature, unsupported feature, or irrelevant feature. In this case, a combination of features is not limited to the combination of features explicitly described, and includes features that have not been explicitly described; it is thus hard to understand one's mind under such circumstances. Fitness is general information on the user's mind obtained through combinations of features dependent on the individuality of each image database, while the feature reference amount is statistical information depicting the user's mind, which depends on combinations of features and the emergence of each feature. In the proposed method, the fitness and the feature reference amount are utilized as simple parameters to control the image retrieval process based on the user's mind.

3.2.2 Dominant conditioned uniform crossover

A dominant conditioned uniform crossover using the feature reference amount is proposed to improve the subjectivity reflection in the feature section results arising through generations. In general, crossovers are performed upon pairs of chromosomes (Parent 1, Parent 2) determined to be subject to crossover according to the crossover rate. The uniform crossover is a crossover technique for generating a binary random mask the length of each gene and determining the location of the crossover according to the mask value (a crossover is performed when the mask value is "1"). This is the crossover technique employed in the basic method, referred to in this paper as a simple uniform crossover. In the dominant conditioned uniform crossover proposed here, an additional condition is considered to determine the location of the crossover. When the mask value is 1, the feature reference amounts assigned to features that have emerged in the gene pair located in the gene locus are compared, and a crossover is performed if the feature reference amount of Parent 1 is equal to or less than that of Parent 2. Provided that the gene of which feature reference amount is the larger is called superior gene and the other inferior gene, the above operation effectively classifies chromosomes into types including many superior genes and types including many inferior genes.

3.2.3 Ways of enhancing situation dependency

In this method, because an interactive

process has been installed in GA, the feature selection reflecting the user's mind proceeds in parallel with image retrieval. This method is appropriate for gradually accepting changes in the user's mind – that is, depending on circumstances. This is an advantage of adopting GA to realize an image retrieval method controlled by the user's mind. However, in the basic method, since no further estimation is permitted for an image that has already been estimated by the user, changes in mind cannot be quickly reflected in the search process. To improve the flexibility of that method, we propose a technique that enhances the situation dependence to improve subjectivity reflection.

The central part in Fig.2 is a window used for user estimations as implemented in the proposed method. When an image is presented by the system in the estimation-waiting window (right), the user estimates the image using a mouse. Images for which estimation have been completed are moved to the estimation-completed window (left) so that the latest estimation results (YES, UNDECIDED, NO) equal to the permitted scrolling a window may be identified. A change in the user's mind is accepted at any time, and the target image the estimation for which the user wants to change is moved from the estimation-completed windows with the mouse. At this time, the system reflects an appropriate feature selection result that has influenced the estimation of the image, i.e., chromosome, upon the determination of the population of the next-generation GA. As shown in Fig.1, a chromosome pair fulfilling the following condition is constantly held and renewed during the time of image retrieval for each image that matches the chromosomes. One such chromosome in that pair is the chromosome (Cfmax) having the largest fitness among those that have emerged so far, and another chromosome (Cfmin) that has the lowest fitness; these are what have emerged most recently and the numbers of their matching images are the lowest. If the user changes its estimation, the chromosome, either Cfmax or Cfmin is temporarily added as a potential chromosome for the next-generation population, according to the direction of the change, and the fitness of each chromosome is determined by the matching process between the chromosome and the image. They are then restored by selection to have predetermined population sizes. Four directions of change are permitted in this case, as shown below. When the user changes the estimation from YES to NO, or from UNDECIDED to NO, the Cfmax gene is added. For a change from NO to YES, or from UNDECIDED to YES, the Cfmin gene is added. The above rule is defined so that fitness changes the most corresponding to the change in user estimation.

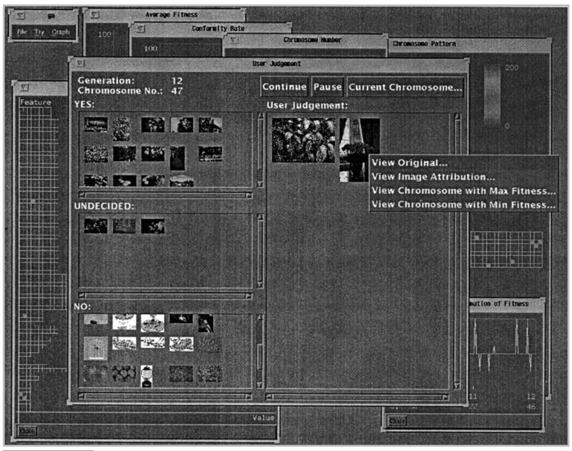

Fig.2 *Windows for user estimation*

The proposed method improves the degree of interaction with the user, and the user is permitted to change his or her past estimation at any time during image retrieval, enabling the user to interact with the system easily and naturally, for user-friendly image retrieval. In addition, changes in the user's mind can be quickly reflected in the feature selection to allow incorporation into the following trend of presented images. This refines and improves the subjectivity reflection in response to changes in the user's mind.

3.3 Improving retrieval efficiency

3.3.1 Use of historical information

As generations evolve, this method increases the number of chromosomes featuring high fitness - that is, the feature selection results that reflect the user's mind. We propose a technique of utilizing the feature selection results obtained for images in past searches as historical information to raise retrieval efficiency, saving high-fitness chromosomes as historical information along with appropriate examples of matching images and their estimation results. When the user makes a request to use this historical information, the historical information is added to the initial population unchanged. Note that one set of historical information is provided for each image retrieval process. For convenience, an appropriate population size of chromosomes is taken as a set selected from chromosomes in descending order of fitness, going back in time from the latest generation, with no overlap in order of appearance. From historical information of the several sets of past information saved in the system, saved images and their estimations, which are saved with historical information, are displayed in a suitable appropriate manner during image retrieval. Considering this information, the user determines whether to use this historical information and reflects this decision in the initial population. More than one set of historical information may be specified for use at this time.

For the proposed method, the method of constructing the initial population is expanded so that historical information in addition to the initial key (feature) entered by the user can be entered. In the basic method, the permitted initial input at the start of image retrieval was limited to the initial key or search condition, while the chromosome of the population size corresponding to the initial key was defined as the initial population. The concept of a sub-population is introduced at this point to allow historical information in the initial input. The population corresponding to the initial key is set aside as a sub-population without change, and the chromosome of one set of saved historical information comprises one sub-population reflecting historical information. As a result, in the event of a request for the use of s ($s \geq 0$) sets of historical information, up to s sub-populations of historical information are created. The sum of these sub-populations, the population size, becomes the initial population. In the generations that follow, no specific sub-population is recognized, but the generation processing is performed on a population size basis. Note that in dealing with historical information, since sub-populations may have different chromosome lengths, the longest chromosome length in the sub-populations is taken as the representative chromosome length for the population. Then sub-populations in which historical information in which chromosome lengths are shorter than the representative length compensate for the shortage of genes by randomly selecting features from those that have emerged in those chromosomes.

Introducing the concept of the sub-population allows natural handling of a search condition consisting of as many as s logical sums, the initial key entered by the user, through the general of s sub-populations for each logical unit. This definition of the initial population has a natural form, by which for many initial inputs of search conditions for logical product, the chromosome lengthens and the range of matching images narrows, while in the case of many initial inputs of search conditions for logical sums, the population size grows and the search range expands.

3.3.2 Control of matching images

There are two problems related to images corresponding to matching images, i.e., chromosomes, that affect retrieval efficiency. One is a problem related to the acquisition of matching images; the other is a problem related to the thinning of matching images. Each problem is discussed below, and a solution proposed. Both problems are solved by controlling matching images with feature reference amounts and their variance, which are statistical quantities reflecting the user's mind.

(1) Problem relevant to matching images and its solution

In this method, the chromosomes created through generations express the feature selection results that have been controlled by the user's mind. However, since most of the emerging chromosomes become feature selection results that do not match images in the database, images are not presented quickly to the user. Of course, this problem arises from the characteristics of each database, and whether (or not) the feature selection results happen to be insignificant in the current database; they may be significant in another database. Reducing the emergence of chromosomes that do not match images is reduced by considering the characteristics of each database improves retrieval efficiency.

We propose a method of controlling insignificant feature selection results. As shown in Fig.1, the chromosomes (feature selection results) that have not matched images are marked first. Then, if the marked chromosomes still remain in the next generation after selection and crossover, they are subjected to the following mutation using the feature reference amount:

- Replace the genes of small feature reference amounts and variance with those of large feature reference amounts and variance in the target chromosomes (If the feature is single, a number of other features equal to chromosome length are randomly created).

Since the above operation reduces the logical chromosome length, reflecting the user's mind, we can expect the rate of emergence of significant feature selection results matching images to improve. In addition, since they are replaced by genes of large variance, genes of high unpredictability are distributed to latter generations, contributing to the creation of various feature selection results that provide indications of the user's mind.

(2) A problem associated with thinning of matching images and a proposed solution

Since the basic method assumes a large-scale image database, not all matching images are shown to the user, but are randomly thinned to reduce the load on the user during estimation (the images thinned out are not presented to the user, but assigned the image fitness listed on Table 1). However, thinning out images randomly is inefficient, as the number of matching images changes with chromosomes, with the risk of missing a number of images wanted by the user. Retrieval efficiency improves when a thinning, which reflects more of the user's mind is performed, relying on the size of the database and tolerances of user estimation.

We propose a method for controlling image browsing in which the user enters a value *k*, a browsing threshold indicating how many images are permitted on screen per browsing session, as shown in Fig.1. The user may alter *k* at any time during image retrieval. Excess images are thinned out and the remaining images presented to the user on screen only when the number of matching images matching the chromosome during chromosome processing exceeds *k*. The order of presentation for images adheres to the following rule, based on the feature reference amount:

- Focusing on the features defined in other than the target chromosomes among the matching images, the system presents images in descending order of feature reference amount and ascending order of variance (randomly presented when exceeding *k*).

The above operation maintains user interaction and helps the user browse images that reflect his or her mind.

4 Computer Simulations

4.1 Outline of the simulations

Computer simulations have been performed to simulate improvements in subjectivity reflection and retrieval efficiency obtained by the proposed method. For qualitative estimation of the flexibility of the interface unit, some 2300 color images such as illustrations and photographs were used. In other quantitative experiments, for comparison against the basic method, 100 binary images (32×32 pixels) were used as stored images, as in the case of Reference[12]. The quantitative results shown in this section are those obtained without entering any initial key and provided by the user's implicit intention. For simplicity, no UNDECIDED estimation was made. The implicit user intention was regarded as a "horizontal line." The feature definition used in the experiment did not feature a clear feature of such a "horizontal line." However, as related features in the semantic layer, we find a feature of a "line" along the "geometry name" feature axis, while in the graphical layer we find a "black pixel" for each area of the 64-divided images along the "black pixel area" feature axis. It is possible to express black pixel areas consisting of horizontally arrayed black pixels. As images that match the user's expectation, manually drawn line images are defined as "other" features along the "geometry name" feature axis. The images defined by the "line" feature values include "vertical line" images that do not match user intentions. We have prepared tricks such as these in feature definition for the stored images used in the experiment in order to provide diversity in the feature expressions of images that meet user expectations and simulate the complex situations in which the same feature expression may have different meanings to the system and the user while the system and the user agree or cross at a point through different feature expressions. We assumed that the thin-out rate of matching images in the case of a large-scale image database to be 90% (the 10% given in the experiments referenced in References[12][13][14][15] are incorrect; 90% is correct). However, since the quantity of the stored images used in the quantitative experiment is merely 100, the matching image control method described in Section 3.3.2 was not implemented, and the thinning performed randomly.

4.2 Experimental Results and Discussion

4.2.1 Subjectivity reflection

The degree of subjectivity reflection can be determined from five requisite properties: individuality, situation-dependence, autonomy, diversity, and reproducibility.

Individuality and situation-dependence are central properties for reflecting one's mind. In the proposed method, through interactions between the user and the system, the user can enter search conditions representing his/her mind and estimate the presented images on a real-time basis at any time during image retrieval. The image that fits user intentions can be provided on each occasions by incorporating the user estimation into the GA fitness and feature reference amount. Individuality and situation-dependence are thereby reflected upon the image retrieval process. In the experiment, the user specified historical information, looking at past retrieved images, and entered logical sum search conditions in addition to logical product conditions as initial keys. It was confirmed that the interface grew more flexible and reflections of individuality and situation-dependence more accurate. Because changes in the user's mind were constantly taken into account in an interactive manner, the user changed estimations occasionally, with a breakthrough occurring when the feature selection froze, no longer providing new matching images.

Autonomy is a property of finding an image that matches the user intention without being restricted by initial keys entered by the user. The proposed method is provided with autonomy. GA is characterized by emergence, with an original mechanism that ensures a high probability of survival into the next generation of high fitness chromosomes. Although the user did not enter initial keys in the experiment, desired retrieval results were nevertheless obtained and the feature selection results providing clues of the user's mind during the retrieval process obtained with high fitness, confirming the quality of autonomy. We also confirmed that the feature selection results (series of black pixels) reflecting differences in user intentions according to either implicit expectation (anticipation of "horizontal line" or "vertical line") emerged with high fitness. Even when the user entered initial keys, the feature selection results, including those superficially contradicting the initial keys, emerged with high fitness, and images matching the user's mind were provided. Thus, the proposed method was able to reconcile differences in "sensitivity" between the user and the system.

Diversity is a latent ability indicating system adaptability, appearing when various users try to reflect their minds on the system. Considering that the present method reflects the user's mind upon the system in a form of the feature selection, the degree of diversity of feature combinations seen in the feature selection results should be related to the readiness to reflect one's mind. This heavily draws on the diversified searchability of GA, and its crossover method is an important factor in setting its abilities. We have recognized the diversity of the proposed method based on the frequency of fitness maximum feature patterns (the rate of emergence of chromosomes that have a fitness of 100% and differ in the combination of features). Fig.3 shows the cumulative frequency of fitness maximum feature patterns plotted against the consecutive number of chromosomes (chromosome consecutive number) that have emerged through all generations, wherein the following four types of crossover T0-T3 are compared.

T0: No selection or crossover. Chromosomes are randomly generated through all generations.
T1: No feature reference amount is set in each feature. Simple uniform crossover is performed. (Basic method)
T2: Fa is adopted as the feature reference amount, and a dominant conditioned uniform crossover is carried out. (Proposed method)
T3: Fm is adopted as the feature reference amount, and a dominant conditioned uniform crossover is carried out. (Proposed method)

Note that the imperfection in expression was taken into account in all types and that the relationship between chromosomes and images is defined as an inclusion relation. The experiment was performed at population size 50, chromosome length 10, elite selection rate 50%, crossover rate 100%, and mutation rate 100%. As generations evolve, the cumulative frequency increases in every type. In comparison, T3 shows slightly higher values than others at early generations, while T2 starts to increase significantly at a consecutive chromosome number around 50000 (1000 generations). From the viewpoint of diversity, T1 is inferior even to T0, in which genes are generated randomly. The results indicate that T2 type of the proposed dominant conditioned uniform crossover is significantly improved in terms of diversity, although it requires some time for startup.

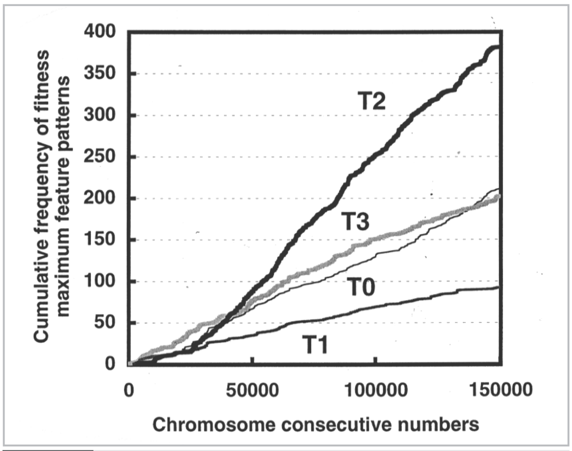

Fig.3 Diversity (Comparison between crossover modes)

Reproducibility refers to the ability to provide images matching the user's mind. In this method, we focus on the analogy between the relationship between the GA phenotype and genotype that have been expressed but not correlated one-to-one due to shortage of information and the implicit relationship between the image and the user's mind. We define the relationship between the genotype and phenotype as an ambiguous, inclusion relation between the feature selection results shown by the chromosomes and image features. The reproducibility of this user-control method is maintained by allocating the image fitness from the user's mind to phenotype images. Fig.4 demonstrates the changes in reproducibility (ratio of retrieved images to the total images that match the user intention). The experimental conditions were the same as those of Fig.3. Although the potential for obtaining 100% reproducibility within a predetermined generation range is highly contingent on variables, this value improves as generations evolve regardless of differences in types. In the early stages, T3 and T1 show higher reproducibility, while T2 shows lower reproducibility. However, as generations evolve, T2 and T0, obtain high reproducibility, piling up short step-ups, while T1 and T3 are likely to settle in local levels. This is also found from the results of the average multiplexing degrees of the maximum fitness patterns where T1>T3>T2>T0. T1 and T3 show sharp increases in early generations because they produce many feature selection results that overlap; frequent re-matching operations are carried out on images that have been thinned out and removed, thereby eluding user estimations. The reproducibility of T0 fluctuates significantly in each trial. This is because the randomness of T0 produces many feature selection results that do not match images. This is evident from its average fitness – around 47% - throughout all generations. The results also imply that T0 is a mode that does not learn the relations between the features defined by the stored images and user intentions, even through interactive exchanges.

The above results on subjectivity reflections demonstrate that the proposed method improves upon the basic method. The T3 crossover mode obtains simplified image retrieval results quickly. The T2 mode is appropriate when the user wishes diversified image retrieval results that reflect his/her mind, with no requirement for speed of processing. Following the proposed method, it is easy to change crossover modes from T3 to T2 halfway through the evolution of generations, and the degree of subjectivity reflection should increase.

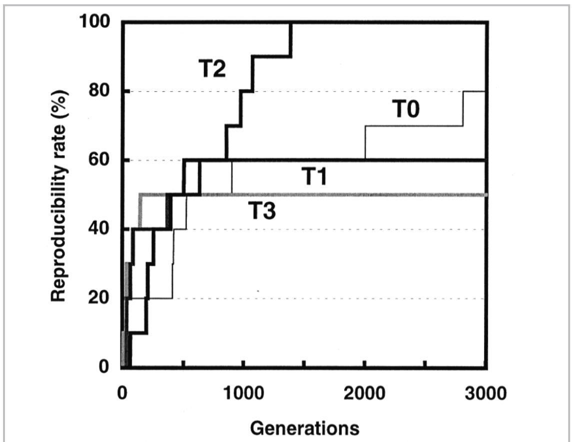

Fig.4 Reproducibility (comparison between crossover modes)

4.2.2 Retrieval efficiency

To improve retrieval efficiency, the efficiency of GA itself must be improved, as with pipeline processing[18], eliminating the concept of generations. However, in this paper, in an effort to improve retrieval efficiency by accurately reflecting the user's mind, we propose a method of raising efficiency in a semantic manner.

Our experiment demonstrates the usefulness of historical information. Fig.5 shows the cumulative frequency of fitness maximum patterns along with population size, comparing the results of use and non-use of historical information. The T2 crossover mode was adopted and the horizontal axis represented by generation. The sub-population size (SUB-SIZE) in the case of using the historical information was determined to be the same for both the initial keys and historical information. Compared to the same population size (SIZE), 100, the frequency of fitness maximum feature patterns is larger with historical information than without in an amount equal to the startup savings. In the experiment, where an equal 50 sub-population size was set for the initial keys, the use of historical information proved effective. Fig.6 shows the changes in reproducibility, while Fig.7 shows changes in conformity rate (the ratio of images matching the user's mind to the total retrieved images. Since they are images that have already appeared in the retrieval process, they do not include images that have eluded user estimations) in the early stages of generation. These figures demonstrate that the use of historical information helps improve reproducibility and conformity rates in the early stages of generation.

These studies demonstrate that the use of historical information improves retrieval efficiency by introducing sub-populations in the early search stages without losing the diversity associated with the degree of the user's subjectivity reflection. When the user relies on historical information, these results imply that it is unnecessary to search all stored images. The proposed method can be a handy technique for quickly finding likely images.

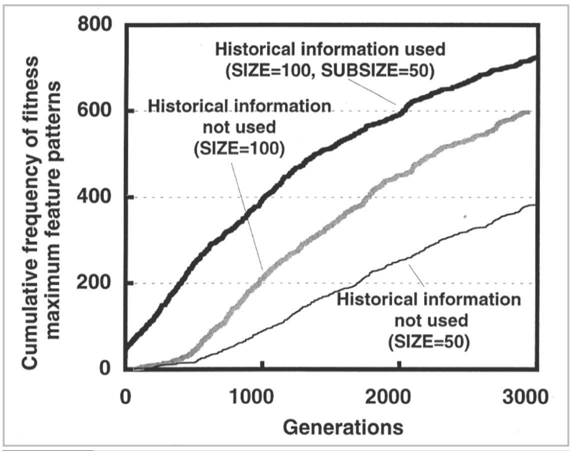

Fig.5 *Diversity (comparison between use and non-use of historical information and population sizes)*

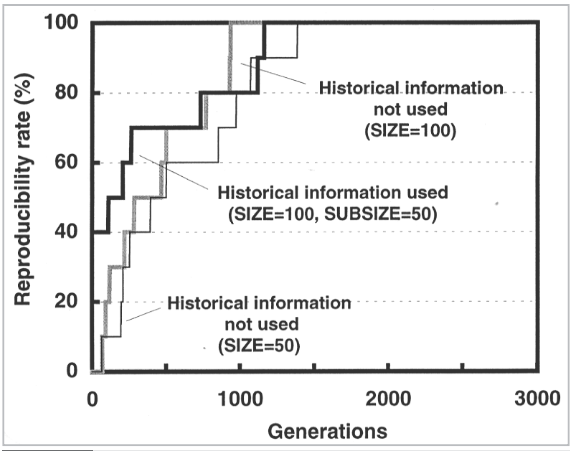

Fig.6 *Reproducibility (comparison between use and non-use of historical information and population sizes)*

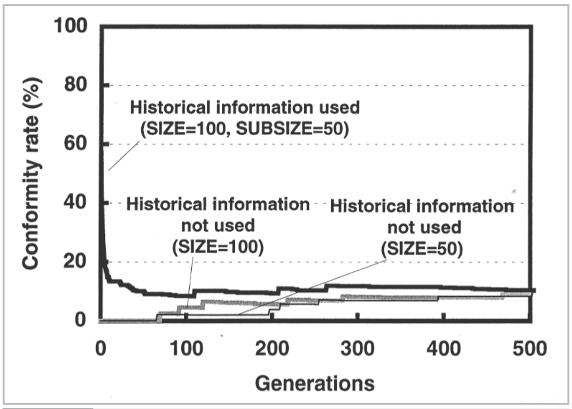

Fig.7 *Conformity rate (comparison between use and non-use of historical information and population sizes)*

5 Conclusions

This paper places our GA-based image retrieval as a method of controlling image retrieval by each user's mind. We have also proposed techniques for improving the degree of subjectivity reflection and retrieval efficiency. To enhance subjectivity reflections, we increased the level of interactivity and proposed a dominant conditioned uniform crossover that provides a variety of feature selection results matching the user's mind, based on a newly defined feature reference amount and a technique for enhancing situation-dependence capable of interactively reflecting changes in the user's mind in the system at any time. The effectiveness of these approaches has been confirmed by computer simulations. To improve retrieval efficiency, we also propose a technique that draws on historical information based on past data, confirming the effectiveness of this approach in the early search stages. In addition, we propose a matching image control method by which retrieval efficiency can be improved by acquiring matching images and controlling image presentation based on the feature reference amount that reflects the user's mind.

The proposed method is quite useful for finding images that match user intentions without precisely expressing the user's mind. Since our image retrieval is controlled only by the user's mind, the information obtained through interactions between the user and the system is central to this method. As a result, the method is greatly dependent on user estimations, and the complexity of user estimations is a current problem. Future challenges include improving the function for inferring the user's mind in order to reduce complexity in user estimations, adaptation of the method to large-scale image databases, and further evaluations of its results and implications.

Acknowledgments

We are grateful to Mr. Katsuhiro Fujimoto and Mr. Yutaka Ohkoshi at M-KEN Co., Ltd., Dr. Shun-ichi Iisaku, former Intelligent Communications Division Director, and staff members at the former Visual Communication Section for their assistance and for numerous helpful discussions.

References

1 T. KURITA, T. KATO, I. FUKUDA, AND A. SAKAKURA, *"SENCE RETRIEVAL ON A IMAGE DATABASE OF FULL COLOR PAINTINGS"*, TRANS. OF IPSJ, D-II, Vol.33, No.11, pp.1373-1383, Nov., 1992(in Japanese).

2 H. NAGASHIMA AND Y. HIJIKATA, *"ON RETRIEVAL OF SIMILAR TRADEMARKS CONSIDERING SUBJECTIVITY"*, TRANS. OF IEICE, D-II, Vol.J74-D-II, No.3, pp.311-320, Mar., 1991(in Japanese).

3 K. TANABE, J. OHYA, AND K. ISHII, *"SIMILARITY RETRIEVAL METHOD USING MULTIDIMENSIONAL PSYCHOLOGICAL SPACE"*, TRANS. OF IEICE, D-II, Vol.J75-D-II, No.11, pp.1856-1865, Nov., 1992(in Japanese).

4 T. YAGYU, Y. HISAMORI, Y. YAGI, AND M. YACHIDA, *"COLOR COORDINATE SUPPORTING SYSTEM WITH NAVIGATING STATE OF USER'S MIND"*, TRANS. OF IEICE, A, Vol.J79-A, No.2, pp.261-270, Feb., 1996(in Japanese).

5 M. ODA, S. AKAMATSU, AND H. FUKAMACHI, *"ADAPTABILITY OF K-L EXPANSION TECHNIQUE FOR AMBIGUOUS FACE IMAGE RETRIEVAL"*, TRANS. OF IEICE, A, Vol.J79-A, No.2, pp.288-297, Feb., 1996(in Japanese).

6 H. TAKAGI AND K. AOKI, *"INTERACTIVE EVOLUTIONARY COMPUTATION: FROM CREATIVITY SUPPORT TO ENGINEERING APPLICATIONS"*, ET AL, WORKSHOPS:INTERACTIVE EVOLUTIONARY COMPUTATION, Mar., 1998(in Japanese).

7 G. Venturini, M. Slimane, F. Morin, and J. -P. Asselin de Beauville, *"On using interactive Genetic Algorithms for knowledge discovery in databases"*, Proc. of 7th Int. Conf. on Genetic Algorithms, pp.696-703, 1997.

8 C. Caldwell, and V. S. Johnston, *"Tracking a criminal suspect through "Face-Space" with a Genetic Algorithm"*, Proc. of 4th Int. Conf. on Genetic Algorithms, pp.416-421, 1991.

9 H. Shiraki and H. Saito, *"An interactive image retrieval system using Genetic Algorithms"*, Int. Conf. on Virtual Systems and Multimedia '96 in Gifu, 1996.

10 M. NAGAO, M. YAMAMOTO, K. SUZUKI, AND A. OHUCHI, *"EVALUATION OF THE IMAGE RETRIEVAL SYSTEM USING INTERACTIVE GENETIC ALGORITHM"*, JOURNAL OF JAPANESE SOCIETY FOR ARTIFICIAL INTELLIGENCE, Vol.13, No.5, pp.720-727, Sep., 1998(in Japanese).

11 S. KATO, *"INFORMATION SPACE FOR IMAGE RETRIEVAL"*, SPRING CONF. OF IEICE, D-499, pp.7-235, Mar., 1994(in Japanese).

12 S. KATO, *"FEATURE SELECTION IN FLEXIBLE IMAGE RETRIEVAL"*, TRANS. OF IEICE, D-II, Vol.J80-D-II, No.2, pp.598-606, Feb., 1997(in Japanese).

13 S. KATO AND S. IISAKU, *"AN IMAGE RETRIEVAL TECHNIQUE USING GENETIC ALGORITHMS AS A METHOD OF REFLECTING USER SUBJECTIVITY"*, PROC. OF THE 2ND SYMPOSIUM INTELLIGENT INFORMATION MEDIA, pp.91-98, Dec., 1996(in Japanese).

14 S. KATO AND S. IISAKU, *"A PROPOSAL OF HOW TO MAKE GOOD USE OF THE PERSONAL HISTORY INFORMATION REFLECTING USER SUBJECTIVITY IN AN IMAGE RETRIEVAL TECHNIQUE USING GAS"*, TECHNICAL REPORT OF IEICE, AI97-10, pp.71-78, May, 1997(in Japanese).

15 S. Kato and S. Iisaku, *"An image retrieval method based on a genetic algorithm"*, Proc. The 12th Int. Conf. on Information Networking, pp.333-336, 1998.

16 A. KITAMOTO AND M. TAKAGI, *"HIERACHICAL MODEL AS A FRAMEWORK FOR CONSTRUCTING SIMILARITY-BASED IMAGE RETRIEVAL SYSTEMS"*, TECHNICAL REPORT OF IEICE, IE97-27(PRMU97-58, MVE97-43), pp.25-32, Jul., 1997(in Japanese).

17 K. Nishino, M. Murakami, E. Mizutani, and N. Honda, *"Efficient fuzzy fitness assignment strategies in an interactive Genetic Algorithm for cartoon face search"*, Proc. of 6th Int. Fuzzy Systems Association World Congress, pp.173-176, 1995.

18 A. KITAMOTO AND M. TAKAGI, *"INTERACTIVE IMAGE BROWSING USING QUEUE-BASED GENETIC ALGORITHM"*, JOURNAL OF JAPANESE SOCIETY FOR ARTIFICIAL INTELLIGENCE, Vol.13, No.5, pp.728-738, Sep., 1998(in Japanese).

Syuko KATO

Researcher, Next Generation Internet Group, Information and Network Systems Division

Next Generation Media