Emotional Image and Musical Information Retrieval With Interactive Genetic Algorithm

SUNG-BAE CHO, MEMBER, IEEE

Invited Paper

Several techniques in artificial intelligence have shown a great potential to develop useful human–computer interfaces, but it is still quite far from realizing a system of matching the human performance, especially in terms of emotion, intuition and inspiration. To overcome this shortcoming, we present a promising technique called interactive genetic algorithm (IGA), which performs optimization with human evaluation, and with which the user can obtain what he has in mind through repeated interaction. To project the usefulness of the IGA to develop emotional human–computer interfaces, we have applied it to the problems of image and music information retrieval. Several experiments show that our approach allows us to design and search digital media not only explicitly expressed, but also abstract images such as "cheerful impression," and "gloomy impression." It is expected that the same approach can be applied to many other problems in musical information retrieval and manipulation based on intuition and inspiration.

Keywords*—Emotion, interactive genetic algorithm, musical information retrieval, subjective test.*

I. INTRODUCTION

As digital repositories of multimedia such as image and music are rapidly growing, effective ways of retrieving the information therein have been spotlighted in several fields. Unlike the retrieval of text information, music requires more effective management and retrieval because it is subject to users' subjective emotions. Even the same music might cause one user to feel gloomy and another user to feel happy.

However, most of the conventional methods lack the capability to utilize human intuition and emotion appropriately in the process of retrieval. Conventional music retrieval systems mainly focus on fast and effective extraction of musical patterns and transformation of the user's humming into systematic notes. It is difficult to retrieve appropriate information when the user cannot remember some parts of melody.

Manuscript received February 5, 2003; revised November 7, 2003. This work was supported by the Brain Science and Engineering Research Program sponsored by the Korean Ministry of Science and Technology.

S.-B. Cho is with the Department of Computer Science, Yonsei University, Seoul 120-749, Korea (e-mail: sbcho@cs.yonsei.ac.kr).

Digital Object Identifier 10.1109/JPROC.2004.825900

In order to solve this problem and supplement the lack of the user's expression capability, we have devised image and music retrieval systems based on human intuition and emotion by using interactive genetic algorithm (IGA) [1]. First proposed by J. Holland in 1975, the genetic algorithm (GA) as one of the computational implementations is an attractive class of computational models that mimic natural evolution to solve problems in a wide variety of domains.

The basis of GA is that a population of problem solutions is maintained in the form of chromosomes, which are strings of encoding potential solutions [2]. Strings can be binary or have many possible alternatives at each position. The strings are converted into problem solutions, which are then evaluated according to an objective scoring function. Often it is not possible to exhaustively test all aspects of a solution, and noise may be present on the objective function, so the assigned fitness is an estimate of the true fitness of a chromosome. It is important that this is a good estimate; otherwise, the selective pressure that favors truly high-scoring chromosomes can be lost in the noise caused by poor fitness estimates.

Following fitness evaluation, a new population of chromosomes is generated by applying a set of genetic operators to the original population. These are basically random copying and altering of individuals from the original population with the probability of copying of any individual from one generation to the next being proportional to its fitness.

In the meantime, most of the conventional applications of GA lack the capability to utilize human intuition and emotion appropriately in creative applications such as architecture, art, music, and design. There is no clear measure to give the evaluation of fitness other than the one in the human mind. IGA is a technique that performs optimization with the human evaluation. A human can obtain what he has in mind through repeated interaction with the method, when the fitness function cannot be explicitly defined. This allows us to develop effective human-oriented evolutionary systems, since this obtains from humans the fitness value for

Table 1
Summary of IGA Applications

Category	Examples
Artistic applications	Computer graphics, CG lighting, design, music
Engineering applications	Speech and image processing, hearing aid, virtual reality, database retrieval, data mining, control and robotics
Edutainment applications	Composition support, robot control, games

the problem at hand and produces better designs or images for the next generation.

There are some criticisms that IGA cannot provide conclusive results because the population size is small and the number of fitness calculations is limited because human beings perform the fitness evaluation. However, it is reported that the global optimum of IGA is not a point but rather an area from the conventional GA optimization perspective because every system output that a human user cannot distinguish is considered to be psychologically the same [3]. If we wanted to retrieve music of a gloomy impression, the optimal solution should not be just one but several different music notes that possess a common impression. IGA is used to converge within much shorter generations with smaller population, even though we still need some techniques to reduce the human fatigue problem that is common to all human–machine interaction systems.

The rest of this paper is organized as follows. Section II introduces the basic idea of the IGA and some possible applications briefly. In order to justify the usefulness of the IGA, emotion-based image and music retrieval systems are presented in Sections III and IV, respectively. Each section includes some experimental results with subjective tests to confirm the usefulness of the IGA for emotion-based image and music retrieval systems.

II. Backgrounds

A. IGA

GA applies some of natural evolution mechanisms like crossover, mutation, and survival of the fittest to optimization and machine learning. GA provides a very efficient search method working on population, and has been applied to many problems of optimization and classification [1].

The procedure of a simple GA is as follows.

```
t = 0;
InitializePopulation P(t);
Evaluate P(t);
while not done do
  t = t + 1;
  P' = SelectParents P(t);
  Recombine P'(t);
  Mutate P'(t);
  Evaluate P'(t);
  P = Survive P, P'(t);
end_while
```

Each chromosome is encoded by a bit string, and crossover operation swaps some part of the bit strings of parents. It emulates crossover of genes in the way that descendants inherit characteristics from both parents. Mutation operation inverts some bits in the bit string at a very low rate. In the real world, we can see that some mutants come out rarely. Each individual in the population gets higher fitness as it goes generation by generation.

IGA is the same as GA except in the way of assigning the fitness value. In IGA the user gives fitness to each individual instead of fitness function. In this way IGA can interact with the user, and also can perceive the user's emotion or preference in the course of evolution. For this reason IGA can be used to solve problems that cannot be easily solved by GA, such as design and art [4]–[6].

More than 100 publications have been made with the application of the IGA to the problems that require creative solutions, and they can be classified into three categories: artistic, engineering, and edutainment applications [3]. For a good, quick overview of the classic examples, refer to Banzhaf [7, sec. C2.9]. Table 1 summarizes the possible applications of the IGA. Here is an example of design that shows the advantage of using IGA instead of GA. We can apply GA for design by initializing the population of individuals encoded from design characteristics, setting and evolving the fitness as how good the design is. However, there is no standard of the goodness of design, and it is almost impossible to organize the fitness function. IGA might be a solution in this case. IGA can reflect personal preference, because it perceives the fitness directly from humans instead of computing some function [8].

B. Music Retrieval Systems

To put things in perspective, music information retrieval is still a very immature field [9]. No comprehensive survey of user needs has ever been done. The results of the European Union's HARMONICA project are of some interest, but they focused on general needs of music libraries [10]. There are several publications on audio retrieval [11]; the research on music retrieval based on acoustic inputs is not popular [12]. Ghias *et al.* published one of the early papers on query by singing [13]. McNab *et al.*, in collaboration with New Zealand Digital Library, have published several papers on their experiments of query by singing [14].

Content-based music search techniques have been studied and some prototype systems have been developed. Users can input the key melody of the music via Web browser by uploading prerecorded humming [15] or typing the string to represent the melody contour [16], [17]. Combination of user modeling and music information retrieval is also proposed [18].

Some previous research on music retrieval systems is to generate queries for music retrieval and to extract regular patterns and themes from music. This has led some fast and effective retrieval methods of music through effective preprocessing [19], [32]. One of the successful systems is based on query by humming, which transforms the user's humming into some symbolized representation because most of the music information has been stored as symbols. MelDex is the representative system to retrieve the most similar music to what the user has sent out after recording [20].

Other research on music retrieval systems is based on monophonic music sequences. Doraisamy *et al.* investigated techniques for a full polyphonic music retrieval system [21]. Their strategy is to use all combinations of monophonic musical sequences from polyphonic music data. They investigated retrieval performance of monophonic queries made on a polyphonic music databases using an n-gram approach for full-music indexing [22]. Pickens *et al.* extended the "query by humming" music retrieval framework into the polyphonic realm [23]. It is the first to use polyphonic audio queries to retrieve from polyphonic symbolic collections. Song *et al.* have constructed a query system that can process both database queries and database records in raw audio format [24].

However, these kinds of systems have the shortcoming that the user should have appropriate information on the music that is required by the system. If the user does not provide the system with correct humming on some parts of music or some example, he could not get the music wanted. Themefinder, a music retrieval system based on themes, requires the query in regular expression about pitch names, musical scale, or tempo [25]. The new definition on musical symbols expedites the retrieval process of music, but it might hinder novice users from retrieving appropriate music.

III. Image Retrieval

To give a better understanding of the potential of the IGA in the field of emotional music retrieval, a preliminary study on image retrieval is illustrated in this section. The mainstream of previous approaches to image retrieval is the keyword-based method that indexes image database with keywords manually and queries are specified using these keywords. This method allows effective search with small image collections, but it costs much time and labor to construct an index when it applies to a large database, and the efficiency of search decreases when the viewpoint of index constructor is different from that of an actual user. The important problem is that it is inherently difficult to describe some visual aspects.

The content-based image retrieval is proposed to solve these problems. In this approach, image contents are modeled as a set of attributes extracted from the image, and queries are specified using these attributes. Generally, the procedure of feature extraction is automatic or semiautomatic. This method has the advantage of reducing time and labor to be needed to construct image database, but it is difficult to extract precise feature information. Features of image to be used for searching might be color, texture, shape, and relationship of objects within the image.

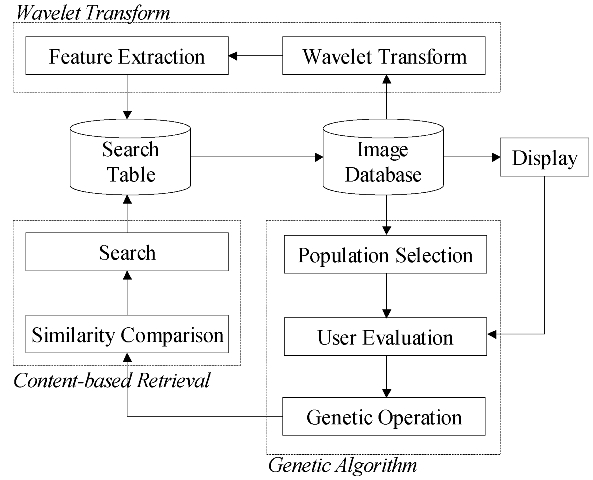

Fig. 1. Image retrieval system using IGA.

For the content-based image retrieval, for example, several working systems have already been developed: the QBIC system of IBM [26]; the QVE of Hirata and Kato; Chabot of the University of California, Berkeley [27]; Photobook of the Massachusetts Institute of Technology, Cambridge; and Image Surfer of Interpix software. It can be applied to a digital library, medical management system, home shopping, and so on.

A. Method

The overall system is constructed as shown in Fig. 1. In the preprocessing step, at first, the wavelet transform is performed for every image in the database, and the average color and the indices and signs of the m magnitude wavelet coefficients are stored in a search table. The system displays 12 images, obtains the fitness values of the images from the human, and selects candidates based on the fitness.

Genetic operation, vertical or horizontal crossover, is applied to the selected candidates. To find the next 12 images, the stored image information is evaluated by each criterion. Twelve images of the higher magnitude value are provided as a result of the search.

For this problem, a chromosome is represented by an array that consists of the indices of wavelet coefficients. Fig. 2 shows the procedure that is used to construct a chromosome. The wavelet coefficients are obtained by decomposing an image using wavelet transform. The $r \times r$ matrix T obtained by the Haar wavelet transform has the average color of the image in entry $T[0,0]$ and the wavelet coefficients in the other entries of T. We can reconstruct the original image without loss using this information, but because we do not have to maintain all the information to search, we just extract the largest 50 coefficients in red, green, and blue (RGB)

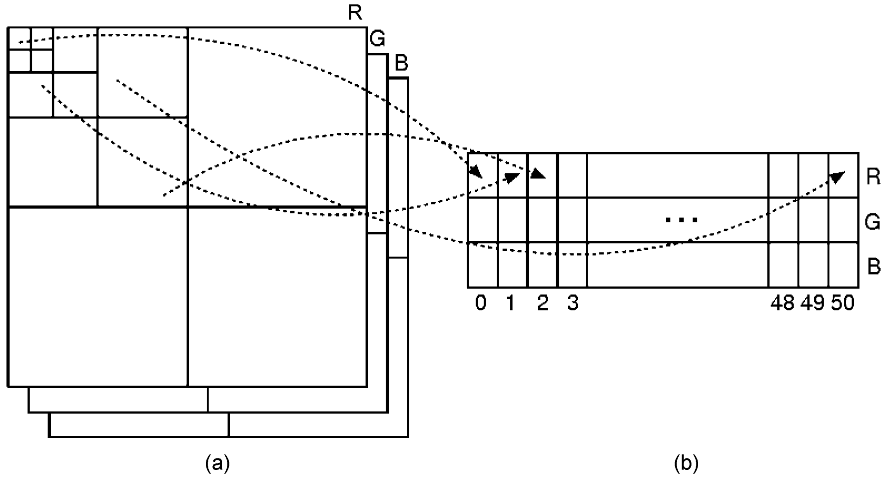

Fig. 2. Procedure used to construct a chromosome. (a) Wavelet transformed 256 × 256 images. (b) The chromosome constructed using the wavelet coefficients of (a).

Fig. 3. User interface of image retrieval system.

channels and use them for constructing a chromosome in a 3×50 array.

Jacobs' work shows that storing the 40∼50 largest magnitude coefficients in each color works best, and truncating the coefficients appears to improve the discrimination power of the metric [28]. Therefore, we have stored only the sign information of coefficient values into the chromosome. However, in addition to the wavelet coefficients, context should be considered to capture the essence of the mood of images, which remains for further works.

This system obtains the evaluation values from the human. It provides a solution that reflects user preference. The size of the population is 12. The strategy of selection is governed by the expected frequency of each individual. We use one-point crossover that proceeds in selecting a point and swapping a part of chromosome on the basis of it. In that case, two individuals having the highest fitness are selected and mated at random. In detail, horizontal and vertical crossovers are used. Those allow us to exchange the color and shape features respectively. Mutation is not adopted explicitly, but the search process explained below incorporates a sort of mutation.

In this sense, the image retrieval is to find 12 new images that have the highest fitness after applying crossover operation. At present, we consider an image of high fitness as a potential target image and an image in database as a candidate image, and compare the similarity between them to de-

Table 2
Within-Class and Interclass Similarities for Gloomy Images

	Within-class	Inter-class	Difference
Mean	1154.692	1114.939	39.752
Variance	27368.739	24196.524	3229.317
SD	165.434	155.552	56.827
t-test			5.508

termine a new image to be displayed. The similarity between potential target image and candidate image is calculated by the following:

$$\|Q, T\| = w_{0,0} \, |Q[0,0] - T[0,0]| + \sum_{i,j} w_{i,j} \, |Q[i,j] - T[i,j]| \, . \tag{1}$$

$Q[0,0]$ and $T[0,0]$ mean overall average intensities of single-color channels, and $Q[i,j]$ and $T[i,j]$ represent those color channels of wavelet decomposition of the query and target images. Here, $w_{0,0}$ and $w_{i,j}$ are determined empirically.

B. Experimental Results

The system is programmed in Microsoft Visual C++ on a Pentium PC. The size of the image database is 2000. In order to search more efficiently and quickly, a searching table is constructed by a batch job over the 256 × 256 JPEG images. It maintains signs and indices of wavelet coefficients. The crossover rate is 0.6. The initial population consists of 12 images selected randomly. As the user gives images fitness values based on how similar they are to what he wants, the system presents new images in the next generation using the GA. Fig. 3 shows the user interface of the system.

This procedure is repeated until the user obtains the image that is most similar to what he has in mind. If the results of the next generation are not satisfactory, the system allows the user to go back to the previous generation. Moreover, the user can increase or decrease the importance of the color to search image.

1) Usefulness of Wavelet Transform: At first, we test whether wavelet coefficients are appropriate for genetic representation or not. To do this, we classify the images into two classes: gloomy and the other. For each image, we calculate the similarity between images contained in the same class (within-class similarity) and the similarity between images contained in different classes (interclass similarity), respectively. The similarity is evaluated through the same method that is used to compare the target image with the candidate image in the system. The higher the value is, the closer the images are. High similarity would mean that the two images are in near points on the wavelet space. Most of the gloomy images produce higher similarity value in the same class. The averages of within-class and interclass similarities are 1154.692 and 1114.939, respectively. Table 2 shows that the mean of within-class similarities is larger

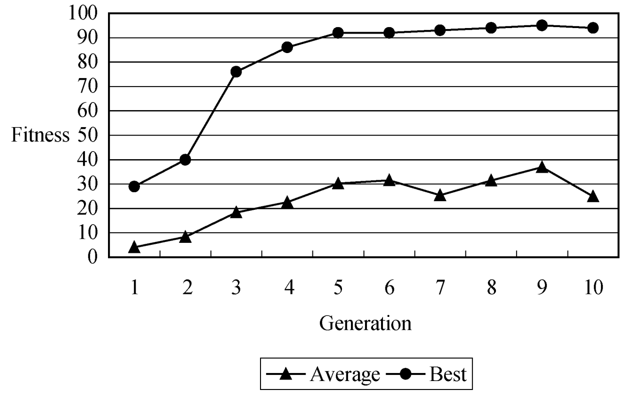

Fig. 4. Average fitness and best fitness in the case of searching the image of gloomy impression.

Fig. 5. Some typical images of converged solutions for gloomy impression.

than that of interclass similarities. In order to determine whether this difference is statistically significant or not, we have performed a paired t-test.

This is useful for testing whether the difference of mean values is significant or not. We hypothesize that the mean values of within-class and interclass similarities are not different, and the t-test value is calculated. The result of the t-test is 5.508 and our hypothesis is rejected at a 99.995% confidence interval. This means that the images in the same class are closer on the wavelet space and we can assert that wavelet coefficients represent the mood of images well.

2) Convergence Test: It is pretty hard to prove the convergence in the case of the IGA. Toward this goal we have attempted to show the change of fitness according to the generations and compare the results obtained. Fig. 4 shows the average and best fitnesses produced by the users for ten generations in the case of searching the image of gloomy impression. We can see that the fitness is effectively increased by adopting the user's evaluation, although the search space of chromosome encoded using wavelet coefficients is very large. Fig. 5 shows some images of converged solutions for gloomy impression.

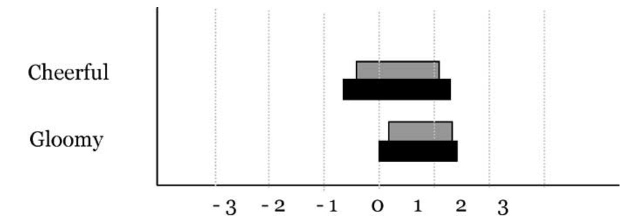

Fig. 6. Intervals of satisfaction of images searched according to cheerful and gloomy impressions. Gray and black bars represent confidence intervals of 95% and 99%, respectively.

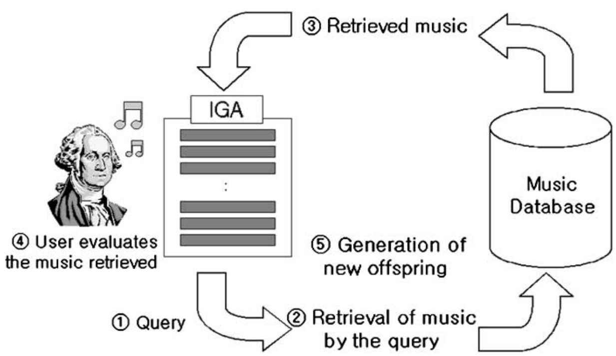

Fig. 7. Music retrieval system using IGA.

3) Subjective Test: A psychological test by Sheffé's method is conducted to see the user's satisfaction. Sheffé's method of paired comparison is a subjective test that requests subjects to give a score to the difference between a given pair of images [29]. It makes a psychological distance measure from the score. The significance of difference among evaluated images is tested by the analysis of variance. In order to test, we give three subjects two motifs that are gloomy and cheerful impressions, respectively, and they are requested to select eight images manually according to each motif from 500 images from the database. Next, we ask ten subjects to search images with the same motifs using the proposed system. Here, the number of images to be compared is nine, among which eight images are selected manually and one image searched using the system. Thus, the number of paired images used for this subjective test is ${}_9C_2 = 36$. We ask ten subjects to give a seven-step score to the difference between a pair of images while considering the given motif. The evaluation of subjects is statistically tested, and the result indicates that the users are satisfied with the images searched by the proposed interface. Fig. 6 shows the result of the statistical test. The x axis means a degree of satisfaction, and gray and black bars in this figure are confidence intervals of 95% and 99%, respectively.

IV. Music Retrieval

Fig. 7 shows the procedure of emotional music retrieval. In this figure, the system is composed of two main components. One is creating query in the form of chromosome using the IGA and the other is retrieving a music using the query from music database.

First of all, an initial population is generated randomly. The system retrieves the music that is the most similar with the chromosome and the user evaluates the corresponding music after listening. By iterating this process, searching the music database using the query evolved and evaluated by the user's subjective emotion is conducted.

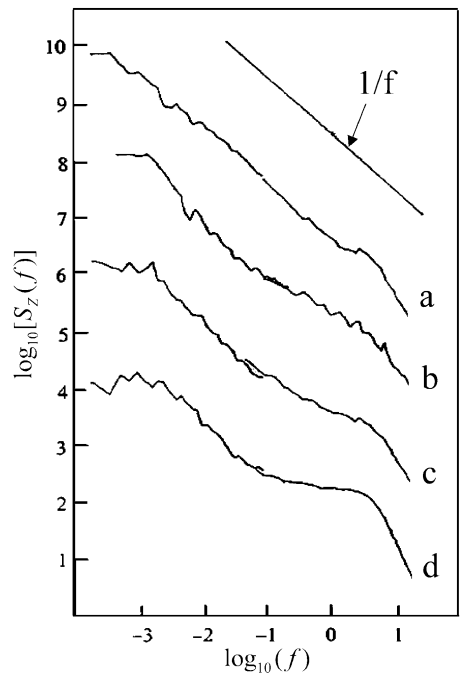

Fig. 8. Power spectrum analysis with respect to genre according to the change of pitches. (a) Classical channel. (b) Jazz channel. (c) Rock channel. (d) News channel.

A. Method

1) Representation of Music: An appropriate encoding scheme is needed to use the IGA in music retrieval. In this paper, the chromosome encodes the frequency of music interval in the music. In the 1975 experiment about analysis of music interval variation, high frequency (high variation of music interval) rarely appeared while low frequency is frequent [30]. The particular thing in this experiment is that the variation of music interval is inverse proportional to the frequency, as the music is more popular and better to listen to. This means that the characteristics of the music might be hopefully inferred from the analysis of music interval variation. Fig. 8 indicates that music of different genres such as classical, jazz, rock, and nonmusical (news channel) has different degrees of strength in frequency. The horizontal axis presents the variation of music interval and the vertical axis presents the power spectrum of frequency. This figure is about the analysis result of four different radio channels (classical, jazz, rock, and news). This shows that the variation of the music interval can represent the characteristics of music.

The appropriate representation of music is essential in music retrieval because this system evolves chromosomes by subjective emotion and uses them as queries. For this reason, statistics of frequency in variation of music intervals are used as chromosome representation. An analog music signal is analyzed for measuring frequencies of music interval variation and the power spectrum of them can be

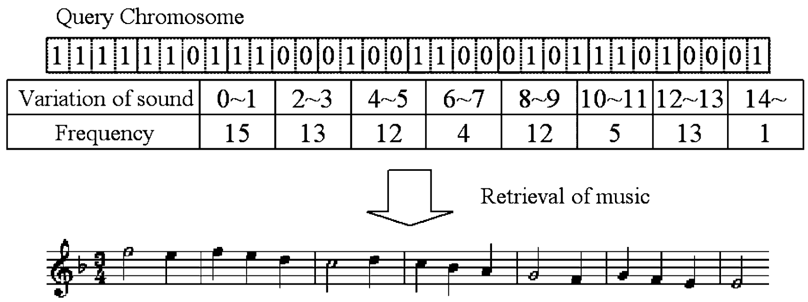

Fig. 9. Music encoding and retrieval.

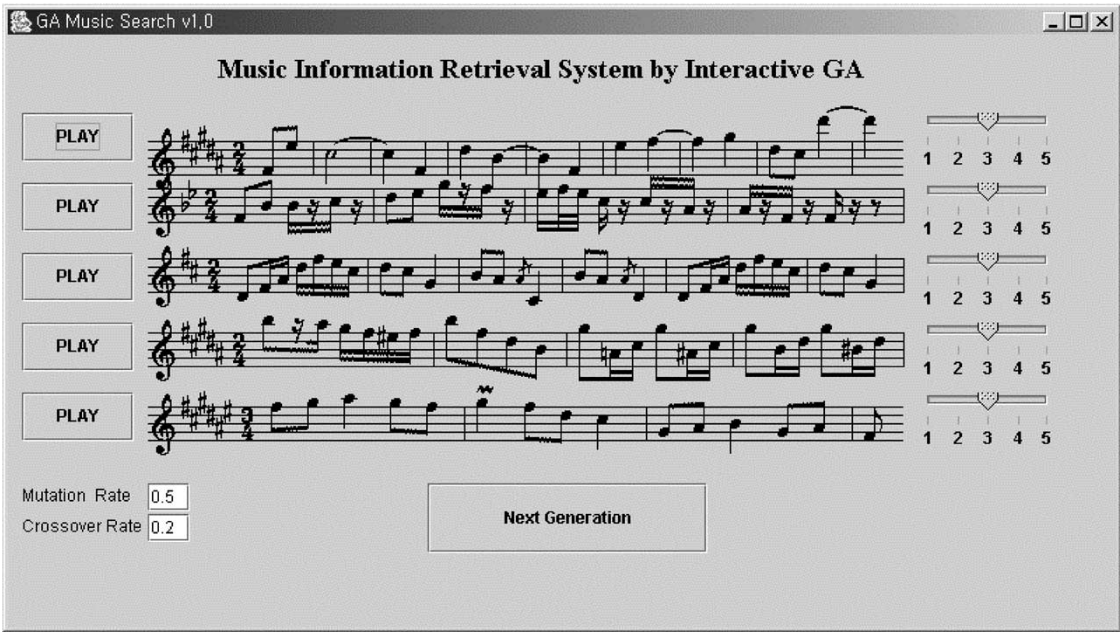

Fig. 10. User interface of music retrieval system.

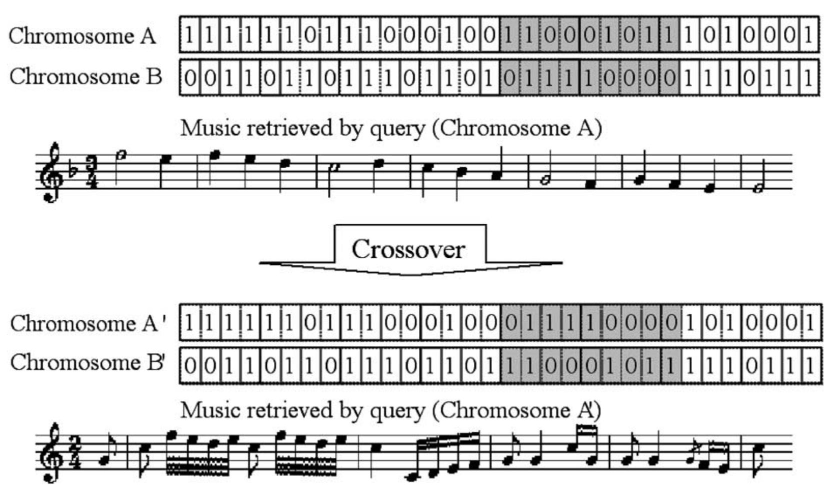

Fig. 11. Crossover operation and decoded music.

calculated. This system calculates the frequency of music interval variation from the information in a MIDI file.

At first, the chromosome of 32 b is generated by the GA. This chromosome contains frequency information of music interval variation per 4 b. The frequency information of the MIDI in the music database is also calculated. The distance between the query generated and the MIDI information in the music database is used and the music of minimum distance is retrieved. Fig. 9 shows the query form of a chromosome and retrieved music that is the closest to it.

2) Search and Evaluation: The user evaluates the real music that is the closest to the chromosome. Stored music has the frequency information of music interval variation through preprocessing. The potential query and the real music are compared by calculating the rate for each variation of music interval using the number of total notes, eliminating the bias

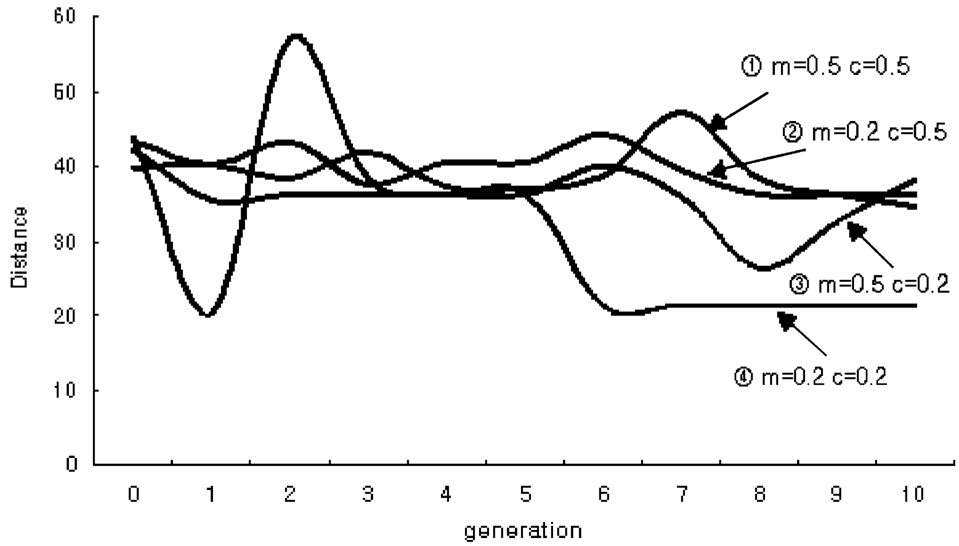

Fig. 12. Fitness changes on searching for music.

Variation of sound	0~1	2~3	4~5	6~7	8~9	10~11	12~13	14~
Frequency	7	5	6	2	0	0	0	0

Variation of Sound	0~1	2~3	4~5	6~7	8~9	10~11	12~13	14~	Distance
Gen = 0	0	8	14	9	9	0	6	9	43.68
Gen = 1	14	15	12	3	8	3	2	0	20.37
Gen = 2	0	1	6	8	5	6	6	13	56.99
Gen = 3	11	12	2	1	11	2	1	3	38.33
Gen = 4	13	8	14	6	14	6	1	14	36.39
Gen = 5	13	8	14	6	14	6	1	14	36.39
Gen = 6	6	5	13	1	4	13	7	2	40.30
Gen = 7	15	13	11	10	13	10	14	6	35.91
Gen = 8	15	12	12	0	0	12	0	1	26.53
Gen = 9	15	11	1	11	0	4	2	1	32.45
Gen =10	15	9	7	8	11	14	5	11	38.13

Fig. 13. Target music and query changes as evolution goes on.

by the size of total notes. $Q(f)$ is the frequency rate for each variation of music interval in query. $M(f)$ is the frequency rate for real music. Distance $D(Q, M)$ is measured by the following:

$$D(Q, M) = \sqrt{(Q(f1) - M(f1))^2 + \ldots + (Q(f7) - M(f7))^2}. \quad (2)$$

Simple Euclidean distance is used between frequency rates of real music stored after preprocessing and those of potential query. After selecting the closest music, the user evaluates by listening to the music while seeing the musical notes from the interface as shown in Fig. 10. This evaluation value is used as fitness for each chromosome.

Gestalt psychology was developed in the early 20th century to help counter the associationist view that stimuli are perceived as parts and then built into complete images. *Gestalt* is the German word for form. Gestalt's developers proposed a theory of pattern perception that relies on the overall form and is not predictable by considering its components [31]. The similarity measure used in this paper does not compare components of two music but global characteristics of them.

3) Evolution of Search Query: Chromosomes evaluated by the user are evolved to the next generation by the GA. The roulette wheel method selects offsprings by the selection probability determined by evaluation value. Crossover and mutation generate a population of the next generation by changing the offsprings with the user-defined crossover and mutation rates. Two individuals selected randomly exchange their chromosomes using two-point crossover. Mutation changes as many bits of the chromosome as determined by the mutation rate probabilistically. Fig. 11 shows the queries crossovered and the corresponding music changed by the operation. Here, shaded bits in the chromosomes represent the part exchanged.

B. Experimental Results

For some preliminary experiments, we used 200 short music files in MIDI format. Subjects are asked to select a target music at random among the 200 files stored and to listen to it repeatedly. Based on the feeling about the music acquired after listening, the retrieval experiments are conducted by changing the crossover and mutation rates. The search process of music is investigated by recording the distance between the target music and the query generated by evolution in every generation from the initial population. Fig. 12 shows the changes of distance to the target music during evolution according to the changes of crossover and mutation rates, which depicts the closest individuals to the target at the population in each generation. Here, c is the crossover rate, and m is the mutation rate.

It can be seen that the retrieval by the query obtained by evolution produces music closer to the target than that generated initially at random, even though the detailed results depend on the crossover and mutation rates. Fig. 13 compares the queries being evolved and the target music. Initially the variation of sound with the highest frequency tends to exist at the latter part, but it moves to the front part like the target music as the evolution goes on.

V. Conclusion

This paper has presented an approach that implements image and music retrieval systems with human preference and emotion using the IGA. The systems give some room to enable the user's subjective retrieval of image and music by evolutionary generation of appropriate queries through interaction with users. They provide a quite different retrieval environment from the conventional systems that search indices based on classification and compare the patterns extracted. With this kind of interface, novice users who do not know the exact title and melody can hopefully find images and music that they want.

Several experiments show that our approach can allow one to search digital media not only by explicitly expressed image, but also by abstract image such as "cheerful impression," "gloomy impression," and so on. However, there are several problems that we have to solve. One of the biggest problems might be in human fatigue that hinders us from applying the IGA to real-world problems. We have to devise effective methodologies to reduce the human burden with better interface and acceleration methods of evolution. Also, we need to investigate further about better representation and feature to encode the music information, which can characterize the music genre sufficiently.

References

[1] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
[2] M. Srinivas and L. M. Patnaik, "Genetic algorithms: a survey," *IEEE Computer*, vol. 27, pp. 17–26, June 1994.
[3] H. Takagi, "Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation," *Proc. of the IEEE*, vol. 89, no. 9, pp. 1275–1296, 1998.
[4] J. A. Biles, "GenJam: a genetic algorithm for generating jazz solos," in *Proc. Int. Computer Music Conf.*, 1994, pp. 131–137.
[5] C. Caldwell and V. S. Johnston, "Tracking a criminal suspect through face-space with a genetic algorithm," in *Proc. 4th Int. Conf. Genetic Algorithm*, 1991, pp. 416–421.
[6] Y. Nakanishi, "Capturing preference into a function using interactions with a manual evolutionary design and aid system," *Genetic Programming Conf. 1996 Late-Breaking Papers*, pp. 133–138.
[7] W. Banzhaf, "Interactive evolution," in *Handbook of Evolutionary Computation*, T. Back, D. B. Fogel, and Z. Michalewicz, Eds. New York: Oxford Univ. Press, 1997.
[8] S.-B. Cho and J.-Y. Lee, "A human-oriented image retrieval system using interactive genetic algorithm," *IEEE Trans. Syst., Man, Cybern. A*, vol. 32, pp. 452–458, May 1998.
[9] J. Foote, "An overview of audio information retrieval," *Multimedia Syst.*, vol. 7, pp. 2–11, Jan. 1999.
[10] D. Byrd and T. Crawford, "Problems of music information retrieval in the real world," *Inform. Process. Manage.*, vol. 38, pp. 249–272, 2002.
[11] Accompanying Action on Music Information in Libraries: HARMONICA. HARMONICA, Amsterdam, The Netherlands. [Online]. Available: http://projects.fnb.nl/harmonica
[12] J.-S. R. Jang and H.-R. Lee, "Hierarchical filtering method for content-based music retrieval via acoustic input," in *Proc. 9th ACM Int. Conf. Multimedia*, 2001, pp. 401–410.
[13] A. J. Ghias, L. D. Chamberlain, and B. C. Smith, "Query by humming-musical information retrieval in an audio database," in *Proc. ACM Multimedia '95*, San Francisco, CA, 1995.
[14] R. J. McNab, L. A. Smith, and J. H. Witten, "Signal processing for melody transcription," in *Proc. 19th Australian Computer Science Conf.*, 1996, pp. 301–307.
[15] MelDex. University of Waikato, Hamilton, New Zealand. [Online]. Available: http://www.nzdl.org/
[16] Themefinder. Stanford Univ., Stanford, CA. [Online]. Available: http://www/ccarch.org/themefinder/
[17] Search by humming [Online]. Available: http://audio.ecs.soton.ac.uk/sbh/
[18] W. Chai and B. Vercoe. Using user models in music information retrieval systems. presented at Int. Symp. Music Information Retrieval 2000. [Online]. Available: http://ciir.cs.umass.edu/music2000/posters/chai.pdf
[19] P.-Y. Rolland, "Music information retrieval: a brief overview of current and forthcoming research," in *Workshop Engineering and Music*, G. Johannsen, O. Werner, and P. Zerweck, Eds. Kassel, Germany: Univ. Kassel, 2001.
[20] D. Bainbridge, R.-J. McNab, and L.-A. Smith, "Melody based tune retrieval over the World Wide Web," Dept. Comput. Sci., Univ. Waikato, Hamilton, New Zealand, Tech. Rep. 98-17, 1998.
[21] S. Doraisamy and S. Ruger, "An approach toward a polyphonic music retrieval system," in *Proc. Int. Symp. Music Information Retrieval*, 2001, pp. 187–193.
[22] ——, A comparative and fault-tolerance study of the use of N-grams with polyphonic music. presented at Int. Symp. Music Information Retrieval 2002. [Online]. Available: http://ismir2002.ircam.fr/proceedings/02-FP04-1.pdf
[23] J. Pickens *et al.*, "Polyphonic score retrieval using polyphonic audio queries: a harmonic modeling approach," in *Proc. Int. Symp. Music Information Retrieval*, 2002, pp. 140–149.

[24] J. Song, S.-Y. Bae, and K. Yoon, "Mid-level music melody representation of polyphonic audio for query-by-humming system," in *Proc. Int. Symp. Music Information Retrieval*, 2002, pp. 133–139.
[25] How Themefinder works, D. Huron and W. Hewlett *et al.*. (2001). [Online]. Available: http://www.themefinder.org
[26] W. Niblack *et al.*, "The QBIC project: querying images by content using color, texture and shape," in *Storage and Retrieval for Image and Video Databases*. Bellingham, WA: SPIE, 1993, pp. 173–187.
[27] V. E. Ogel and M. Stonebraker, "Chabot: retrieval from a relational database of images," *IEEE Computer*, vol. 28, pp. 40–48, Sept. 1995.
[28] C. E. Jacobs, A. Findkelstein, and D. H. Salesin, "Fast multiresolution image querying," in *Proc. SIGGRAPH 95*, 1995, pp. 277–286.
[29] H. A. David, *The Method of Paired Comparison*. London, U.K.: Griffin, 1969.
[30] R.-F. Voss and J. Clarke, "1/f noise in music and speech," *Nature*, vol. 258, pp. 317–318, 1975.
[31] K. V. Nesbitt and C. Friedrich, "Applying Gestalt principles to animated visualizations of network data," in *Proc. 6th Int. Conf. Information Visualization*, 2002, pp. 737–743.
[32] P.-Y. Rolland and J.-G. Ganascia, "Pattern detection and discovery: The case of music data mining," in *Lecture Notes in Computer Science, Pattern Detection and Discovery*, D. J. Hand, N.M. Adams, and R. J. Bolton, Eds. Heidelberg, Germany, 2002, vol. 2447.

Sung-Bae Cho (Member, IEEE) received the B.S. degree in computer science from Yonsei University, Seoul, Korea, in 1988 and the M.S. and Ph.D. degrees in computer science from Korea Advanced Institute of Science and Technology (KAIST), Taejeon, Korea, in 1990 and 1993, respectively.

From 1991 to 1993, he was a Member of the Research Staff at the Center for Artificial Intelligence Research at KAIST. From 1993 to 1995, he was an Invited Researcher of Human Information Processing Research Laboratories at ATR (Advanced Telecommunications Research) Institute, Kyoto, Japan. In 1998, he was a Visiting Scholar at University of New South Wales, Canberra, Australia. Since 1995, he has been a Professor in the Department of Computer Science, Yonsei University. His research interests include neural networks, pattern recognition, intelligent man–machine interfaces, evolutionary computation, and artificial life.

Dr. Cho is a Member of the Korea Information Science Society, INNS, the IEEE Computer Society, and the IEEE Systems, Man, and Cybernetics Society. He was awarded outstanding paper prizes from the IEEE Korea Section in 1989 and 1992, and another one from the Korea Information Science Society in 1990. In 1993, he also received the Richard E. Merwin prize from the IEEE Computer Society. In 1994, he was listed in *Who's Who in Pattern Recognition* from the International Association for Pattern Recognition and received the best paper awards at International Conference on Soft Computing in 1996 and 1998. In 1998, he received the best paper award at World Automation Congress. He was listed in Marquis' *Who's Who in Science and Engineering* in 2000 and in Marquis' *Who's Who in the World* in 2001.