DEVELOPMENT OF IMAGE RETRIEVAL SYSTEM USING MULTIPLE KEY IMAGES

Kazuki YASUKAWA [*a], and Toshikazu KATO [a]

[a] *Chuo University, Japan*

ABSTRACT

Many image search systems, designed to search on only one sample image, share the following 3 problems: (A) Appropriate search results can be expected only when users have images at hand which clearly possess characteristics they want. (B) The same search results are presented when searched by the same images, even if image characteristics intended by users vary. (C) Characteristics which are difficult to express, such as the general color tone and texture, cannot be defined as search keys. To solve these problems, we developed an algorithm to extract common image characteristics from multiple sample images and to estimate likely searches. We developed 3 prototype image search methods: using multiple sample images as a search key, explicitly highlighting details wanted by users, and estimating characteristics wanted by users from among multiple sample images presented. We evaluated these systems on a database of 1500 landscape images and obtained good results.

Keywords: *Image Retrieval System, Multiple Images, Graphical Feature*

1. INTRODUCTION

Currently, amid increasing popularity of the Internet, we can find a lot of images on the Internet including images used for commercial purposes such as advertisements and images created for personal hobbies, etc. To search for specific images from among the millions available on the Internet, several systems have been developed [1]. One widely used method is based on keywords. This method searches for images by a search keyword which matches keywords pre-assigned by other people.

The impressions and feelings evoked by an image vary depending on who is looking at it. Therefore, it is difficult to search for specific images by using keywords pre-assigned by

* **Corresponding author**: 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551,Japan

* **E-mail :**{k_yasukawa，kato}@indsys.chuo-u.ac.jp

other people. It is also difficult to describe landscape and other images accurately by using simple keywords.

Another major method of searching for images is the similar-images search method, which uses an image itself as the key. It compares image characteristics (by quantification of colors, shapes, etc.) between the images. This method, unlike the keyword-based search, can take into account the impressions and atmospheres evoked by images without using words, enabling users to find images that are difficult to find by keyword-based searches.

However, existing similar-images search systems work on general similarities to only a single sample image, and have the following problems:

(A) Users cannot search for images when they do not have an image with the desired characteristics.

(B) The same search results are returned when the same image is used as a key, while the image characteristics each user wants varies.

(C) Users cannot define as search keys any characteristics that are difficult to define, such as the general color tone and texture.

For these reasons, we propose methods for searching images by using multiple sample images as a search key, by explicitly presenting features that users want, and by estimating characteristics that users want from among multiple sample images.

2. PROBLEMS OF SIMILAR-IMAGES SEARCH SYSTEMS

As the existing similar-images search systems offer image results with similar overall characteristics to the sample image, they have the following problems:

(A) Response to a search key: It is difficult for users to find images when they do not have a sample image at hand with the desired colors and textures.

(B) Different focus by users: Each person may have a different focus while looking at the same image. For example, when performing a similar-images search against the image shown in Fig. 1, some people may focus on the shape of the sunflower while others may focus on the color of the sky. When users judge the similarity of images subjectively, they usually evaluate images by focusing on specific characteristics (such as the color or texture) of the images, instead of equally evaluating each characteristic of the image [2]. However, existing-similar images search systems offer the same results when given the same image as a sample image, whereas the characteristics and areas that each user wants may differ.

(C) Entering search keys in systems: When users stress characteristics or areas which are difficult to specify on an image, such as the general color tone, texture, or regularity (Fig. 1), or when they do not have a clear focus ("I cannot tell which part of the image I like but I love its general atmosphere"), it is difficult for them to enter search keys in the systems.

Figure 1: Problems of similar image search systems

3. METHOD OF SEARCHING IMAGES BY USING MULTIPLE SAMPLE IMAGES AS A SEARCH KEY

To solve the problem identified in Chapter 2(A), we propose a method for searching images by using multiple sample images as a search key, which takes into consideration human visual performance and uses a search key which integrates individual characteristics of multiple images.

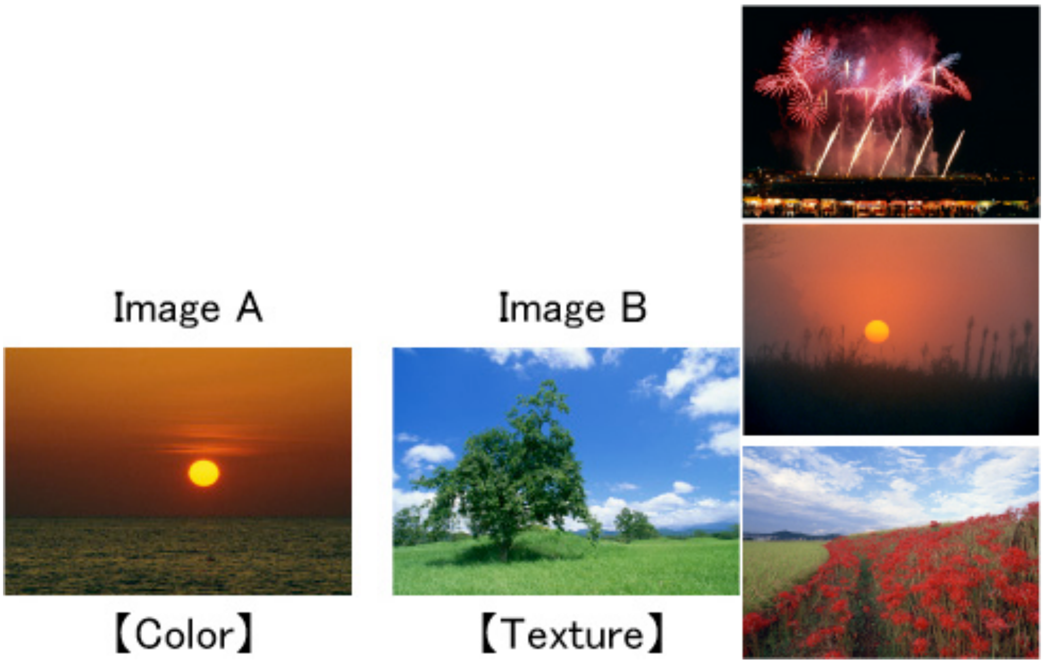
3.1. Human visual performance

The human visual system first processes an object by its properties, such as color, texture, and movement, and then integrates the characteristics [3]. However, as only 2 types of information can be integrated at a time, multiple types of information need to be integrated in multiple stages. During the integration process, the original information is highly likely to be lost [4]. Color and texture are said to be the biggest factors contributing to the human perception of images [5].

3.2. Searching images by using multiple sample images as a search key

The system of searching images by using multiple sample images as a search key compares the quantified characteristics retrieved from the sample images and those of images in the database, and then offers images which have large similarities between the two. We tested it on about 1500 landscape images.

Figure 2 (a) shows an example of a search using the color of Image A, which depicts the sun, and the texture of Image B, which depicts trees. Figure 2 (b) shows that images with high similarities in the color elements of red and black contained in Image A were offered, in preference to images with similar texture to that of Image B. We believe that this is because the similarity in the color characteristics was so strong that the similarity in the texture was de-emphasized.



(a)Search key (b)Search results

Figure 2: Example of a search using multiple sample images as a search key

3.3. Quantification of image characteristics

The similarity between different images is measured by comparing the quantified characteristics extracted from the images. In this section, we define quantification and similarity.

(a) Quantification of characteristics of color

The general color distribution of an image is measured by color histogram. As each of the R, G, and B values in a color image has a range of 0 to 255 in pixel, the number of color values is $256 \times 256 \times 256$ (16.8 million). The color histogram presents the color values in each pixel, but this presents too many color values to determine similarity. Instead, we used the number of pixels (level) per color after reducing the values to $8 \times 8 \times 8$, or 512 colors. The method of measuring the similarity between the 2 color histograms is called the histogram intersection. This method uses and accumulates the smaller of the values of each element of the 2 histograms, and determines the similarity by measuring the overlapping areas of the 2 histograms as shown in Fig. 3.

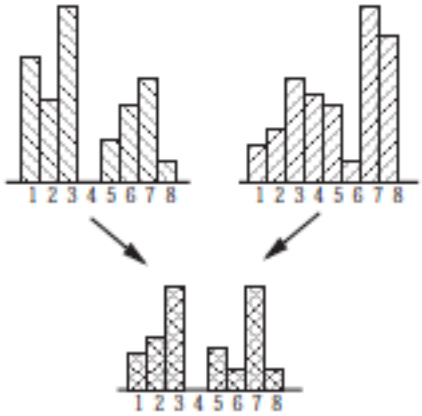


Figure 3: Histogram intersection

In the color histogram of 2 images p and q, the level of the color k in each image should be h_k^p, h_k^q (k = 1, …, N_c). The similarity level S_h measured by the histogram intersection can be defined as:

$$S_h = \sum_{k=1}^{N_c} \min(h_k^p, h_k^q) \qquad (\mathrm{Nc} = 512)$$

(b) Quantification of characteristics expressing texture

The process of visual perception is known to involve neural circuits which extract local and general characteristics such as brightness and colorfulness from images captured on the retina [6][7]. Humans perceive texture and shape by integrating these characteristics. Therefore, to simulate such human visual perception, it is necessary to extract from images not only the values of the general characteristics, but also the values of the local characteristics which describe the relationship between neighboring pixels.

Contrast is generally used to assess images and to measure the value of characteristics in local textures. Contrast quantifies the relationship between a reference point r and a neighboring pixel of displacement a, and can express the directions of linear changes of image data (Fig. 4). Contrast (CON) is defined as:

$$\mathrm{CON}^{(m)} = \sum_{r=1}^{N_r} \frac{f(\mathbf{r} - \mathbf{a}^{(m)})}{|f(\mathbf{r} + \mathbf{a}^{(m)})|} \quad (1) \qquad (\mathrm{m} = 1,\ldots,4)$$

where r represents the reference point (r = 1, …, N_r); $f(r)$ represents the brightness of the reference point; and m, the displacement of the pattern for measuring the contrast, is $a^{(m)}$ (m = 1, …, 4).

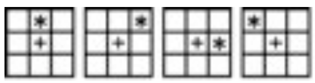


Figure 4: Measurement of contrast

The paper title should use Arial or any sans-serif substitute font, 16pt, aligned to left, with 1.2-interline space, big case (i.e. capital letters). Top space should be 54pt, and bottom one 20pt., If there is a subti

Eq. (1) takes into consideration Weber's law (the minimum difference in the amount of stimulus that humans can detect is proportionate to the level of the stimulus). The denominator represents the level of stimulus to the optic nerve, while the numerator represents the difference in the stimulus. Eq. (1) is the same as $\Delta S/S$, where S represents the level of stimulus.

However, it is also necessary to depict the curvatures as well as the directions of linear changes in order to express accurately the variety of information contained in image data. For this purpose, we propose a tri-contrast feature which mimics the mechanism of human eyes to extract characteristics, by improving the contrast in Eq. (1), and verified its effectiveness [2]. The value derived can express not only the direction of linear changes in images, but also the local characteristics such as curvature and its direction. Artificial objects such as buildings and bridges tend to show strong linear changes, while natural objects such as forests and trees tend to show strong curving changes. Thus, we proved the effectiveness of the tri-contrast feature in landscape images which include many natural and artificial objects [2].

The tri-contrast feature quantifies the relationship among 3 neighboring pixel points: a reference point r and 2 neighboring displacements (a_1 and a_2). Figure 5 shows 28 patterns in which contrast was measured. The tri-contrast feature C_m is defined as:

$$\mathrm{C}_m = \sum_{r=1}^{N_r} \frac{\{f(\mathbf{r}+\mathbf{a}_1{}^{(m)}) - f(\mathbf{r})\} + \{f(\mathbf{r}+\mathbf{a}_2{}^{(m)}) - f(\mathbf{r})\}}{|f(\mathbf{r}+\mathbf{a}_1{}^{(m)})| + |f(\mathbf{r}+\mathbf{a}_2{}^{(m)})| + 2|f(\mathbf{r})|} \qquad (2)$$

(m = 1,…,28)

where r represents the reference point (r = 1, …, N_r); $f(r)$ represents the brightness of the reference point; and m, the displacement of the pattern for measuring the contrast, is $a_1{}^{(m)}$, $a_2{}^{(m)}$ (m = 1, …, 28).

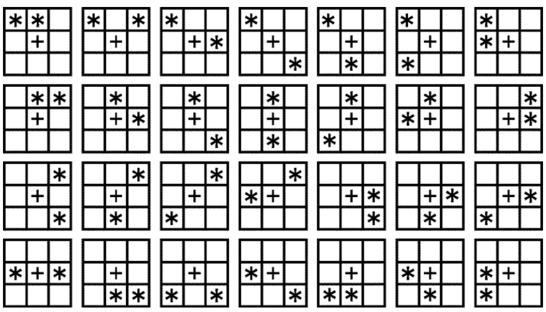


Figure 5: Measuring patterns by tri-contrast feature. +: reference point r; *: displacements a_1 and a_2.

However, the tri-contrast feature cannot offer information on the location of the textures and shapes while it can offer local textures and shapes. So we divided images into 4 × 4 areas and used the tri-contrast feature values of each of the areas as the values of texture. This method allowed us to retrieve the texture, shape, and location in each divided area.

The level of similarity, Sc, between these images can be defined as:

$$S_c = -\sum_{n=1}^{16}\sum_{m=1}^{28}\left|c_{m,n}^{p} - c_{m,n}^{q}\right|$$

The tri-contrast feature in area n in 2 images (p and q) divided into 16 (4 × 4) is represented by $c_{m,n}^{p}$, $c_{m,n}^{q}$ (n = 1, …, 16).

(c) Similarity level in characteristics values

The level of conformity, C, between each image in the database and the search key is defined as $C = S_h + S_c$, where S_h and S_c are the levels of similarity between a search key image and images in the database.

Retrieved images are displayed in descending order of C values.

3.4. Experiment to evaluate search system using multiple sample images

To verify the accuracy of our image search system, we used a database of about 1500 images of natural and urban landscapes with various color tones and textures as shown in Fig. 6. Ten subjects were asked to evaluate the resultsafter searching the images by using the system.

Figure 6: Example images in experiment

Each subject imagined 10 different images and selected 2 sample images representing its color and texture. They searched images by using a different search key image each time. The subjects were asked to evaluate the top 10 images offered as the search result as either "similar" or "not similar" in color, texture, and both. To verify the search accuracy of the system, we calculated the conformity level of the results of evaluation (Fig. 7).

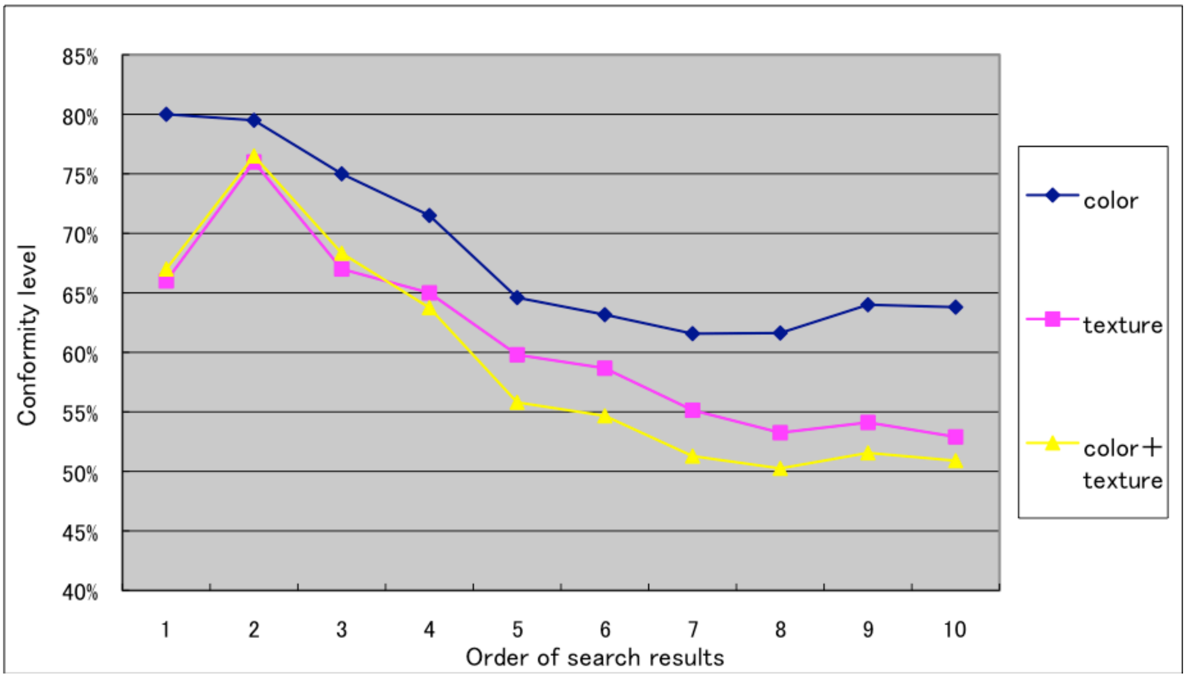


Figure 7: Mean conformity level of up to N results evaluated by subjects

The conformity level of the top 10 search results was 51% (Fig. 7). The conformity level of the top 4 results was 64%, and a mean of 2.6 images among the top 4 matched those imagined by the subjects. Thus, by using colors and textures selected by the subjects from different images as search keys, the system offered matching images even when the subjects did not have at hand one search key which included both color and texture.

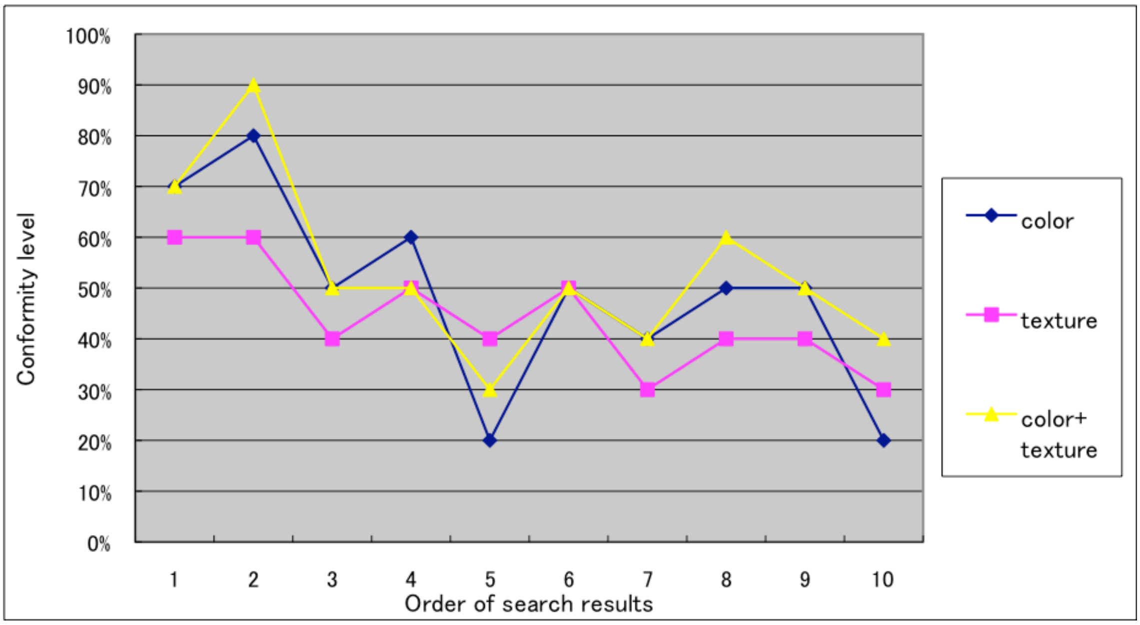


Figure 8: Conformity level of up to N results selected by color

The conformity levels of the texture and of color plus texture showed similar tendencies (Fig. 7), suggesting that the subjects determined that the general characteristics (color plus texture) of the images were similar if the textures were similar to those they imagined. However, the results by subject and by circumstance suggest that the subjects also determined that the general characteristics of the images were similar if the colors were similar to those they imagined (Fig. 8). Thus, characteristics emphasized varied depending on the person or the circumstance. Therefore, the accuracy of search results can be improved further by allowing searches to include information about which characteristics (and, further, which areas) users stress in their sample images.

4. SEARCHING IMAGES BY EXPLICITLY PRESENTING CHARACTERISTICS AND AREAS THAT USERS STRESS

To solve the problem described in Chapter 2 (B), we propose a function to explicitly present characteristics and areas that users stress and a function to adjust the level of focus on each.

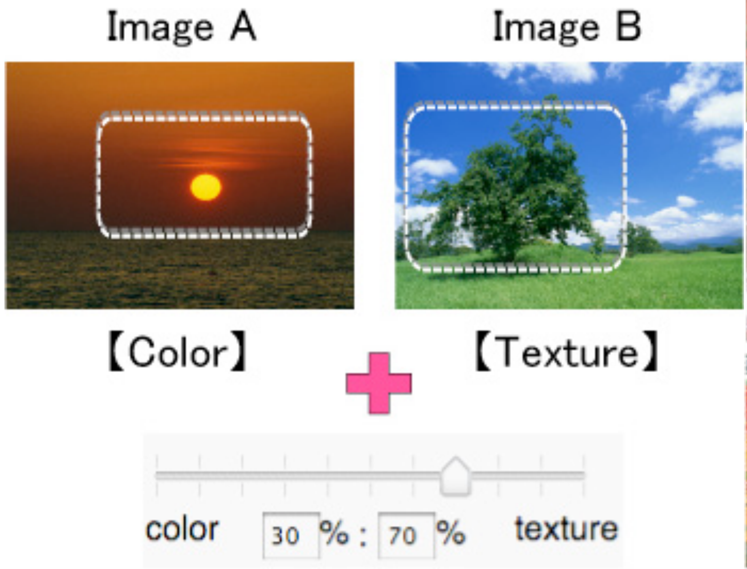
We developed a prototype system based on the assumption that we can develop a system designed to search images imagined by users by adding to the system described in the previous section a function to select characteristics and areas stressed by users, weighted by the level of focus on the image characteristics.

4.1. System for searching images by explicitly presenting characteristics and areas that users stress

We presented 2 images that resemble the imagined color and texture (Fig. 9).

(1) Selecting areas in images: Specific areas in images are highlighted. The sun in Image A and the tree in Image B were selected as examples (Fig. 9).
(2) Specifying search standard for selected areas: Specify which characteristics (color, texture, etc.) in the area selected in point (1) a user stresses. Image A was used for color and Image B for texture (Fig. 9).
(3) Weighting level of focus: Specify the weight for each of the characteristics presented in each of the sample images. The color in Image A and the texture in Image B were weighted at the ratio of 3:7 (Fig. 9).

We searched images in the database of 1500 landscape images. We searched images by placing more weight on the texture of the area of the tree in Image B than on the color of the area of the sun in Image A (Fig. 9a). Images containing the red included in Image A and many curves as included in Image B were determined by the system to have high similarity (Fig. 9b). As texture was emphasized, more images with similar texture were offered than in Fig. 2.



(a)Search key

(b)Search results

Figure 9: Example of search by explicitly presenting characteristics and areas.

4.2. Defining similarity determination criteria

The level of conformity, C, between each image in the database and the search key is defined as $C = W_b S_b + W_c S_c$, where S_b and S_c are …; and W_b and W_c represent the weighting (chapter 4.1(3)).

Retrieved images are displayed in descending order of C values.

4.3. Problems with method

The method for searching images by explicitly presenting characteristics and areas that users stress has problems in the entry of search keys:

1. when many types of image characteristics are available
2. when users cannot imagine which characteristics could affect the search results and how
3. when users focus on characteristics which are difficult to specify, such as the color tone and the texture of an entire image
4. when users cannot precisely specify characteristics they want, such as when the overall atmosphere of the image appeals to them.

5. SEARCHING IMAGES BY ESTIMATING CHARACTERISTICS THAT USERS STRESS FROM AMONG MULTIPLE SAMPLE IMAGES

To solve the problem described in Chapter 2 (C) , it is necessary to estimate the characteristics that users stress from among multiple sample images. We propose a method for such a search that asks subjects to select multiple sample images instead of characteristics, so that common characteristics can be extracted. We assumed that associated images can be retrieved by automatically weighting the search results.

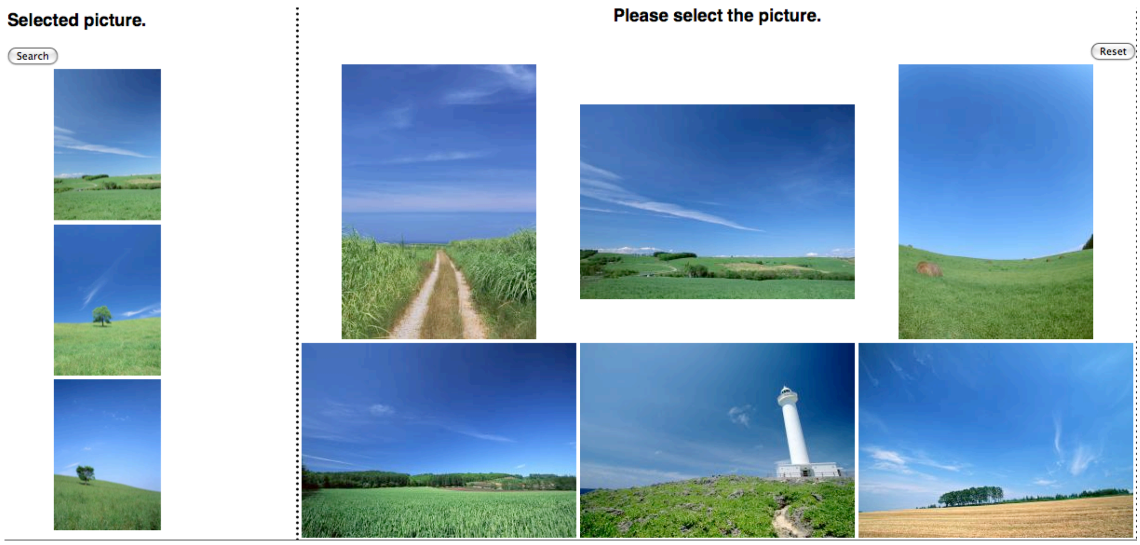
5.1. Estimating characteristics that users stress

We assumed that we can measure the extent of users' consistency in the selection criteria by using the dispersion in the distribution of the values of image characteristics in each image selected. We further assumed that as the distribution becomes more dispersed, the users' selection criteria are inconsistent or that the users do not focus on the values. Conversely, as the dispersion narrows, the users' selection criteria are consistent, or the users focus on the values.

5.2. System of Searching images by estimating characteristics users stress from among multiple sample images selected

We searched images in the database of 1500 landscape images. We selected 3 images of plains with similar color tones (Fig. 10a). Images with color elements of green and blue were determined to be the images with the highest similarity level (Fig. 10b).

We then selected 3 images of buildings with similar textures (Fig. 11a). Images with many straight lines were determined to have the highest similarity: images with buildings and other artificial objects (Fig. 11b).



(a)Search key (b)Search results

Figure 10: Example of searching images by estimating characteristics that users stress from among multiple sample images selected (plains)

(a)Search key (b)Search results

Figure 11: Example of searching images by estimating characteristics that users stress from among multiple sample images selected (buildings)

5.3. Evaluation of system of searching images by estimating characteristics users stress from among multiple sample images

We evaluated the accuracy of the system by asking 10 subjects to select from the database 3, 4, and 5 images each per theme of sky, plain, sea, buildings, and forest. We asked the subjects to evaluate the top 10 images returned as "similar" or "not similar." To verify the accuracy of the system, we calculated the level of conformity of the images. Fig. 12 shows the conformity level per theme.

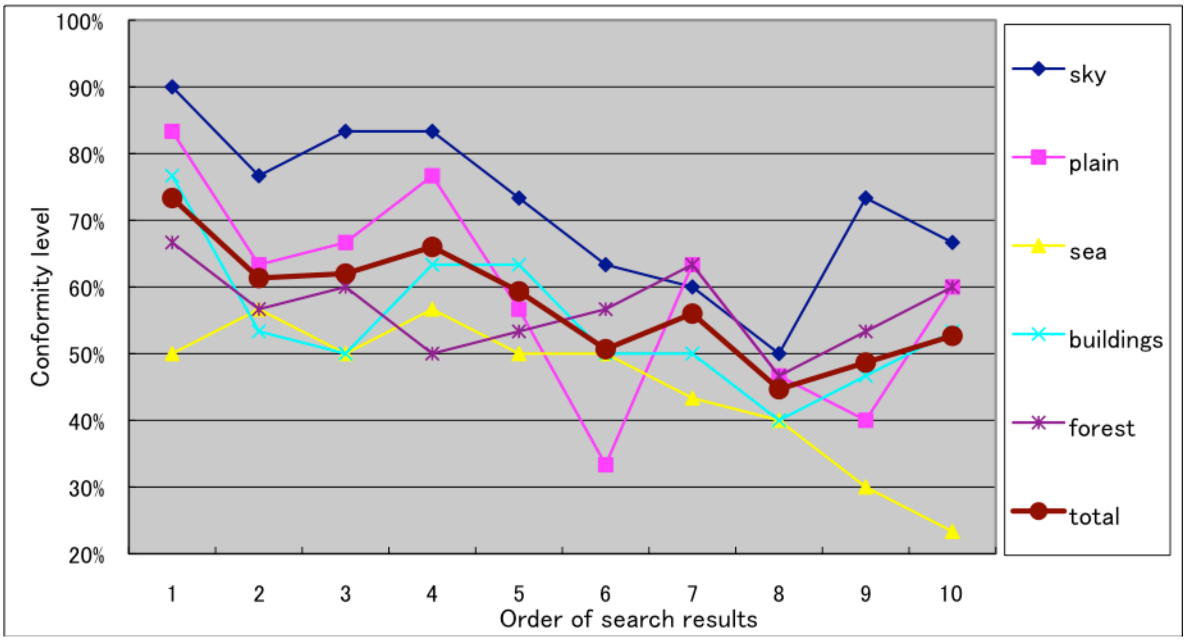


Figure 12: Conformity level of top N images per theme

The average conformity level of the top 10 images was 53% (Fig. 12). Images with sky had the highest conformity level, because these images had common characteristics in the blue expanse at the top and less similar textures.

The images with sea had the lowest conformity. Some have the sea at the bottom, while others have it at the top or over their entirety. We believe that this is because, unlike the images with sky, the characteristics of the images with sea are dispersed, reducing the conformity level.

Fig. 13 shows the relationship between the number of images selected by subjects as search keys and the conformity level. It shows no significant difference in the conformity level between different numbers of images used as search keys. This result suggests that the characteristics that users stress can be estimated from as few as 3 images. Good search results were offered from as few as 3 images when users could clearly imagine what kind of images they wanted (when using multiple similar images as search keys). When users cannot imagine them clearly, the search accuracy of the system can be improved by increasing the number of sample images.

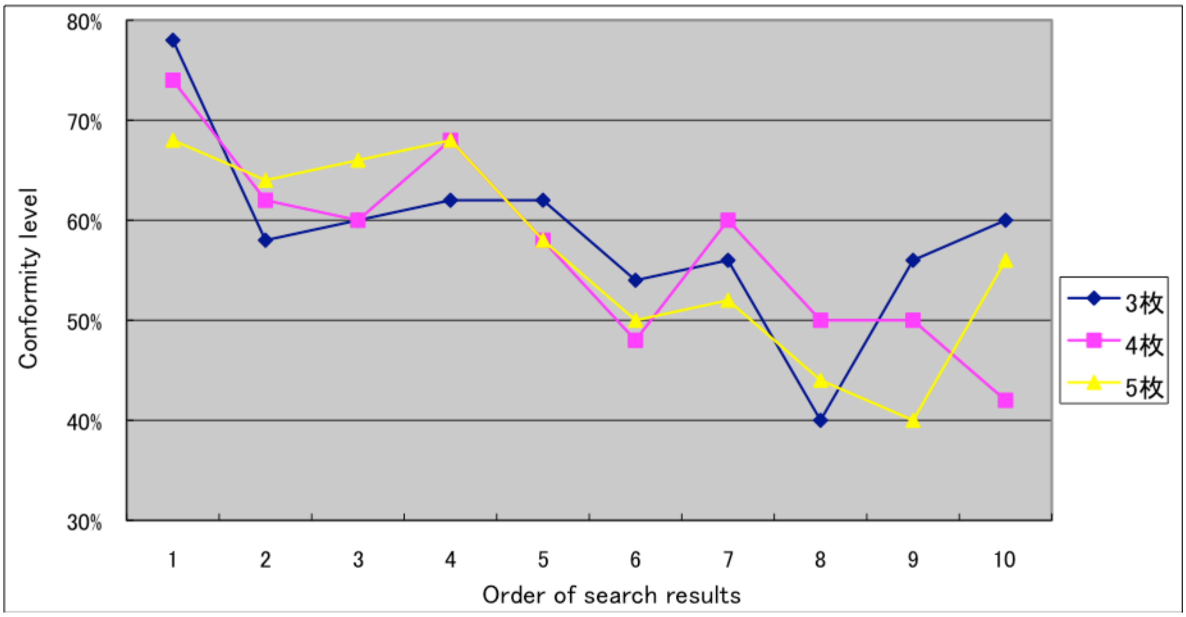


Figure 13: Conformity level of top N images used as search keys

Table 1 shows the relationship between the characteristics that users stress and themes. The images with sky, plain, and sea had a strong focus on color, while the images with buildings had a strong focus on texture. The images with forest had a strong focus on both color and texture. These results indicate that when determining similarity subjectively, people do not evaluate each of the characteristics in images equally, but focus on specific characteristics. Thus, estimating the characteristics that users stress by weighting the searches is effective.

Table 1: Relationship between themes and characteristics users focus on

	Level of focus on color	Level of focus on texture
Sky	0.542	0.458
Plain	0.539	0.461
Sea	0.515	0.485
Buildings	0.494	0.506
Forest	0.500	0.500
Total	0.518	0.482

6. SUMMARY AND FUTURE OUTLOOK

To solve the problems inherent in the similar-images search method, we propose 3 new methods. The method for searching images by using multiple sample images as a search key was able to offer images imagined by users when no single search key image was available. The method for searching images by explicitly presenting characteristics and areas that users stress was able to offer different search results dependent on users' foci on different characteristics and areas in the same images. The method for searching images by estimating characteristics that users stress from among multiple sample images enabled users to search images even when characteristics were difficult to specify or when users were unsure about which characteristics they should stress.

We plan to evaluate these methods further while adding other types of characteristics.

Acknowledgements

We extend our gratitude to all the people at Amana Holdings Inc. and Amana Images Inc. for their important advice and data; to Dr. Masahiro Tada at ATR Intelligent Robotics and Communication Laboratories for his advice on statistical analysis and learning algorithms; and to all the people at the Human Media Lab, Faculty of Science and Engineering, Chuo University, and at the Sensitivity Robotics Research Center for their discussions on research and their cooperation.

REFERENCES

[1] Masahiro Tada, Shunichi Kato, "Analysis and study of visual impressions using SVM and the application to automatic image classification", Technical Report of The Institute of Electronics, Information and communication Engineers, Vol.104, No. 573 (20050114), pp. 45-50, 2004

[2] Masahiro Tada, Shunichi Kato, "Analysis of characteristics of similar visual impressions and modeling of visual sensitivity", Thesis Report of The Institute of Electronics, Information and communication Engineers D-II, Vol.J87-D-II, No.10, pp.1983-1995 (2004)

[3] Treisman, Anne M., and Nancy G.Kanwisher, "Perceiving visually presented objects: recognition, awareness, and modularity,"Current Opinion in Neurobiology, 8 (1998), pp 218-226

[4] S.Morokami and M.Morita：「Two-attribute hypothesis in human visual feature integration, Proc. of the Second International Workshop on Man-Machine Symbiotic Systems」, Kyoto, 173-180, 2004.

[5] Hideaki Chichiiwa, "Viewing colors by mind ? Color psychology design", Fukumura Shuppan Inc.,

[6] L. Spillmann and J. S. Werner：「Visual Perception」, Academic Press, 1990

[7] Kenneth R. Boff, Lloyd Kaufman and James P. Thomas：「HANDBOOK OF PERCEPTION AND HUMAN PERFORMANCE Vol. 2」, Chap. 28, 1986

[8] Fujitaro Nanbu, Kozaburo Hachimura, "Image search system which automatically configure similar judgment criteria", Symposium on Human Science and Computer 2002