Integration of Pattern Information and Natural Language Information for Image Analysis and Retrieval

Yasuhiko Watanabe[†] **Makoto Nagao**[‡]

[†]Dept.of Electronics and Informatics, Ryukoku University, Seta, Shiga, Japan
[‡]Dept.of Electronics and Communication, Kyoto University, Yoshida, Kyoto, Japan
watanabe@rins.ryukoku.ac.jp, nagao@kuee.kyoto-u.ac.jp

Abstract

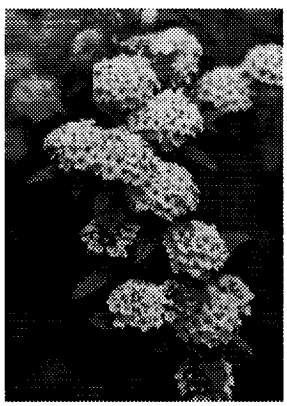
Cooperative use of pattern information and natural language information is quite effective for sophisticated and flexible information processing. Therefore, it is important to investigate the integration of these kinds of information (especially in multimedia). For this purpose, we propose a method for image analysis by using the natural language information extracted from the explanation text of image data. First, we describe the method of extraction of color information from the explanation text. Then, we describe how this color information is used for the extraction of objects from the image data. We report experimental results to show the effectiveness of our method. Also, we report an experimental multimedia database system for pictorial book of flora which we developed by using the results of the experiment.

Introduction

Pattern information and natural language information used together can complement and reinforce each other to enable more effective communication than can either medium alone (Feiner & McKeown 91) (Nakamura, Furukawa, & Nagao 93) (Watanabe & Nagao 96). One of the good examples is a pictorial book of flora (**PBF**). In the PBF, readable explanations combine texts and pictures, as shown in Figure 1. Taking advantage of this combination, we propose a method for analyzing the PBF pictures by using natural language information. Furthermore, we propose a retrieval method for the PBF by using the results of this image analysis.

It is difficult to analyze pictures of the PBF. It is because there are several problems as follows:

- Pictures of the PBF contain many kinds of objects. In addition to flowers and fruits, there are leaves, stems, skies, soils, and sometimes humans in the PBF pictures.
- The position, size, and direction of flowers and fruits are varied quite widely in each picture.



コデマリ
spiraea cantoniensis
落葉低木、高さ１．５ｍに達する。枝は細く、円柱形、無毛、幼時に暗紅褐色を呈し、しばしば先が枝垂れる。葉はひし状狭卵形またはひし状楕円形、鋭頭で基部はくさび形、長さ２～５ｃｍ、幅６～２０ｍｍ、中部以上に不整鋸歯がつき、両面無毛、裏面はしばしば粉白色で、脈が突出する。葉柄は長さ４～７ｍｍで無毛。花序は今年枝に頂生し、散房状、無毛。花は４～５月に咲き、白色で、径７～１０ｍｍ、単弁。萼裂片は３角形で鋭頭、花後に反曲しない。花托は内面に白短毛が密生する。袋果は無毛。中国(中南部)原産。日本では広く庭園に栽培される。名は球形の花序の様子を小型の手毬に見たてたもの。

Figure 1: An example article of a PBF (in Japanese)

- Each flower and fruit has its unique shape, color, and texture which are commonly different from those of the others.

Because of these problems, it is difficult to build the general and precise model for analyzing the PBF pictures in advance. In order to solve these problems and to realize an efficient and reliable image analysis, we propose a method for image analysis using natural language information. Our method works as follows. First, we analyze the PBF explanation texts for extracting the color information on each flower and fruit. Then, we analyze the PBF pictures by using the results of the natural language processing, and finally obtain the color feature of each flower and fruit.

Natural Language Information for Image Analysis

Information on flowers and fruits in explanation text

In the PBF explanation texts, following kinds of information on flowers and fruits are described:

- size
- shape
- number
- color
- spatial relationship
- flowering and harvesting season

Among these kinds of information, the size, shape, number, color, and spatial relationship are important to recognize flowers and fruits in a picture. However, in this study, we use the color information for the image analysis. In other words, we do not use the size, shape, number, and spatial relationship. The reasons are as follows:

- It is difficult to estimate the actual size of flowers and fruits in a picture.
- Each flower and fruit has its unique shape which is generally complicated and different from the shape of others. For this reason, it is difficult to build a general model which would be able to represent all possible shapes of flowers and fruits.
- The occlusion of flowers and fruits might occur in the PBF pictures. As a natural result, it is unreliable to analyze the PBF pictures by using the shape, number, and spatial relationship against the occlusion. On the contrary, it is reliable to analyze the pictures by using the color information against the occlusion and the diversity of position, size, and direction.
- It is easy to convert color names (natural language information) into color vectors in a color space (pattern information). It is because many color names are aligned to color vectors in a color space.

In this way, in order to analyze the pictures of the PBF, we extracted the color information on flowers and fruits from the PBF explanation texts.

Extraction of color information from explanation text

Expressions of color information in explanation text The color information on flowers and fruits is described in the sentences whose topic words are as follows:

- "*hana* (flower)" and its subconcept words
- "*kaben* (petal)" and "*kahi* (perianth)" (these are generally called "*hanabira* (floral leaf)")
- "*kajitsu* (fruit)" and its subconcept words

Table 1: Verbs which govern a case or an adverb which represents the color

Verb		Preposition	Adverb
obiru	(tinged with)	*wo*	–
naru	(grow)	*to, ni*	yes
jyukusuru	(ripen)	*ni*	yes
teisuru	(turn)	*wo*	–
kawaru	(alter)	*ni*	yes

In Japanese text, a topic word is generally accompanied by a postposition "*ha*". Therefore, a topic word can be found by using the postposition "*ha*".

In this connection, the colors of flowers and fruits are described in the expressions such as:

- predicative noun
  (S–1) *hana* (flower) *ha hakusyoku* (white color)
  Observing the explanation text, we found 115 predicative nouns which represent the color features of flowers and fruits.
- predicative adjective
  (S–2) *kaben* (petal) *ha shiroi* (white)
  Observing the explanation text, we found 5 predicative adjectives which represent the color features of flowers and fruits: "*shiroi* (white)", "*kuroi* (black)", "*akai* (red)", "*aoi* (blue)", and "*kiiroi* (yellow)".
- verbalized noun (in Japanese called "Sahen" noun)
  (S–3) *kajitsu* (fruit) *ha sekijyuku suru* (red–ripen)
  Observing the explanation text, we found 5 verbalized nouns which represent the color features of flowers and fruits: "*sekijyuku suru* (red–ripen)", "*koujyuku suru* (deep–red–ripen)", "*oujyuku suru* (yellow–ripen)", "*kokujyuku suru* (black–ripen)", and "*kokuhen suru* (black–alter)".
- verb which governs a case or an adverb which represents the color
  (S–4) *kajitsu* (fruit) *ha sekisyoku* (red color) *wo obiru* (tinged with)
  (S–5) *kajitsu* (fruit) *ha kuroku* (blackly) *jyuku suru* (ripen)
  Observing the explanation text, we found 5 verbs, shown in Table 1, which govern cases or adverbs which represent the color feature. Also, Table 1 shows with each verb:
  - which postposition follows the word which represents the color feature
  - whether the verb governs an adverb which represents the color feature

  For example, in (S–4), "*obiru* (tinged with)" governs 2 cases. Among them, "*sekisyoku* (red color) *wo*", which includes the postposition "*wo*", represents the color feature. In (S–5), the adverb "*kuroku* (blackly)" represents the color feature.

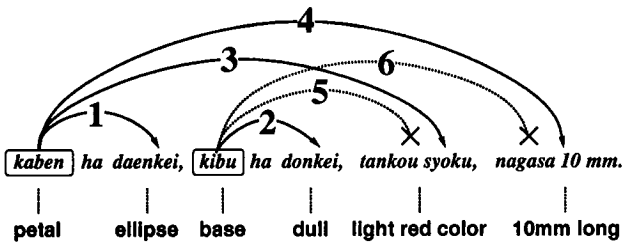


Figure 2: An example of the dependency structure

Dependency Analysis between Topic Words and Predicates In the PBF explanation texts, a topic word often depends on a predicate which is located after a posterior topic word. Figure 2 shows the dependency structure of an example sentence (S–6).

(S–6) *kaben* (petal) *ha daenkei* (ellipse) , *kibu* (base) *ha donkei* (dull), *tankou syoku* (light red color), *nagasa 10mm* (10mm long)

In Figure 2, the topic word "*kaben* (petal)" depends on "*tankou syoku* (light red color)" and "*nagasa 10mm* (10mm long)" located after the posterior topic word "*kibu* (base)". Because these kinds of dependency relation are often found in the explanation texts of the PBF, the ambiguity of dependency relation, like relation 3–6 in Figure 2, constitutes a serious problem for extracting the color information.

To solve this problem, we have proposed a method for analyzing the dependency relation using the following types of information about technical words used in the PBF texts (Watanabe & Nagao 95):

1. is–a information,
2. part–of information, and
3. attribute information.

In (Watanabe & Nagao 95), we examined our method for dependency analysis, and the correct recognition score was 73 %. In addition, these types of information are automatically obtained from diagrams and the explanation texts of the PBF.

Extraction of color information from explanation text We extracted the color information on flowers and fruits from the PBF explanation texts in the following way:

Step 1. Japanese morphological analysis

We used JUMAN (Matsumoto et al. 96) as a Japanese morphological analyzer.

Step 2. Dependency analysis between the topic words and predicates

Step 3. Extraction of color information from the results of the dependency analysis

The color information extracted from the PBF explanation texts is the color name of each flower and fruit.

Table 2: Results of natural language analysis

(a) Test sample

Color information	Number
described	26
not described	4
Total	30

(b) Results of color information extraction

Expression of color information	Result	
	Success	Failure
predicative noun	15	1
predicativee adjective	0	0
verbalized noun	3	0
case or adverb	6	1
Total	24	2

kajitsu (fruit) *ha kyukei mataha rankei* (sphere or oval) *de* (be), *kei 5–7cm* (5–7cm in diameter) *ni nari* (be), *7–8 gatsu ni jyukushi te* (ripen in July – August) *ouhaku syoku* (yellow white color), *ichibu* (part of them) *ha noukou syoku* (deep red color) *to nari* (alter), *hyoumen niha* (in the surface) *nanmou* (hair) *wo missei suru*(be a mass of).

Figure 3: An example of incorrect analysis

To evaluate our method, we used the explanation text of 30 pictures in our experiment. The results are shown in Table 2. One cause of incorrect analysis was the anaphoric expressions. Figure 3 gives an example of errors. This sentence came from the explanation text of "*momo* (peach)". From this sentence, the system could extract only "*ouhaku syoku* (yellow white color)" as the color information of "*momo* (peach)" fruits, and failed to extract "*noukou syoku* (deep red color)". It is because the system didn't estimate the referent of "*ichibu* (part of them)". In this experiment, we ignored the determination of referents of pronouns in Japanese sentences. But, in order to extract the color information on flowers and fruits correctly, we must realize the method for estimating the referents of pronoun in Japanese text. In any case, only "*ouhaku syoku* (yellow white color)" was utilized for analyzing the picture of "peach" fruits.

Image Analysis Using Natural Language Information

As mentioned previously, it is almost impossible to build the exact model which can predicate the locations of flowers and fruits in a picture of the PBF. But in order to realize the efficient and reliable analysis, it

shiro, haku	(white)	=	(1.00, 1.00, 1.00)
kuro, koku	(black)	=	(0.00, 0.00, 0.00)
aka, seki	(red)	=	(1.00, 0.00, 0.00)
midori, ryoku	(green)	=	(0.00, 1.00, 0.00)
ao, sei	(blue)	=	(0.00, 0.00, 1.00)
ki, ou	(yellow)	=	(1.00, 1.00, 0.00)
murasaki, shi	(magenta)	=	(1.00, 0.00, 1.00)
mizu, sui	(cyan)	=	(0.00, 1.00, 1.00)

Figure 4: Correspondence between the color names and the RGB color vectors

is desirable to focus attention on some regions where some flowers/fruits are assumed to be located. For this purpose, the process of our system roughly specifies some regions of flowers and fruits by using the natural language information. Then, the subsequence processes analyze these regions in detail in order to extract the regions of flowers and fruits with high degree of accuracy.

Rough Extraction Using Natural Language Information

The color information extracted from the explanation texts is the color name of each flower and fruit. Using this color information, we extract the regions of flowers and fruits roughly as follows: First, the color name is converted into a color vector in the RGB color space. Then, the regions of flowers and fruits are roughly extracted by means of this color vector.

Conversion from Color Names to RGB Color Vectors For the image analysis, we used the RGB color space which is represented by a unit cube. Fortunately, typical color names are aligned with color vectors in this RGB color space. Figure 4 shows the examples of the correspondence between the color names and the RGB color vectors. Taking advantage of this, the color name x was converted into a color vector $\boldsymbol{C}_x$ in the RGB color space as follows:

1. The color name x is segmented into color names $i(i = 1 \cdots n)$ which are aligned with the RGB color vectors $\boldsymbol{C}_i$. The segmentation follows the longest match principle.
2. Calculate the RGB color vector $\boldsymbol{C}_x$ with the color vectors $\boldsymbol{C}_i$.

$$\boldsymbol{C}_x = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{C}_i \tag{1}$$

For example, "*ouhaku syoku* (yellow white color)" was segmented into "*ou* (yellow)" and "*haku* (white)", then converted into its color vector $(1.00, 1.00, 0.50)$ by using the vectors of "*ou* (yellow)" and "*haku* (white)".

Extraction by using RGB Color Vector We roughly extract the regions of flowers and fruits by extracting the pixels the color of which is similar to the color of the flower/fruit described in the PBF explanation text. Namely, the system extracts a pixel whose color vector $\boldsymbol{C}$ satisfies the following inequality:

$$|\boldsymbol{C} - \boldsymbol{C}_x|^2 < d^2 \tag{2}$$

$$d^2 = \left(\frac{1}{4}\right)^2 \times 3 \tag{3}$$

where C_x is a color vector converted from a color name of a flower/fruit. In this way, Figure 5 (b) is obtained from Figure 5 (a).

Precise Extraction by Clustering and Heuristic Rules

As shown in Figure 5 (b), the extraction process using natural language information often extracts not only flowers and fruits but also leaves, stems, and so on. To solve this problem, the system chooses the regions representing flowers and fruits from the regions extracted by the natural language information in the following way. First, the system divides the regions extracted by using the natural language information into several clusters by a clustering process. Then, the system selects the regions of flowers and fruits from these clusters with the heuristic rules.

Clustering Process As mentioned above, the system divides regions into smaller regions by the clustering processes in the RGB color space. For clustering the regions, we used simple clustering and K–means clustering algorithm (Nagao 83). In this study, the simple clustering is used as a preliminary clustering process for giving a good initial state to the K–means clustering. It is because the result of the K–means clustering algorithm is affected by its initial state.

Heuristic Rules Observing the pictures in the PBF, we found the regions of flowers and fruits are:

- sufficiently large compared to the other parts,
- located in a remarkable position, and
- generally of gradating colors.

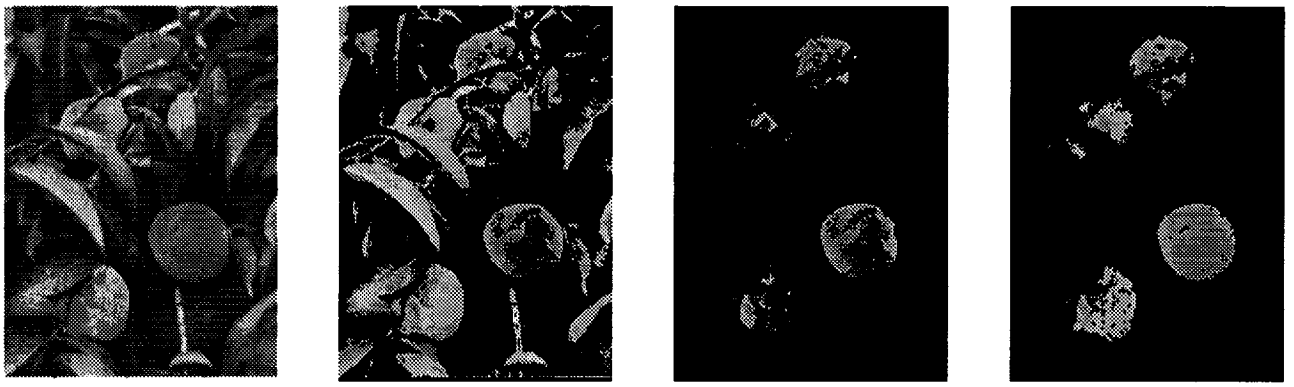
From these features, in order to extract the regions of flowers and fruits precisely, we introduced heuristic rules as follows:

Rule 1 Elimination of very small regions

We assumed that the regions of flowers and fruits are remarkably large compared to other regions. It is because the pictures in the PBF are taken plainly in order to recognize flowers and fruits in the picture easily. For this reason, this rule eliminates the clusters which contain less than 5% pixels of the entire pixels extracted by color information.

Rule 2 Detection of the closest region to the center

The themes of PBF pictures are mostly flowers and fruits. Therefore, flowers and fruits are located in the



(a) Original image ("peach" fruit) Color information: yellow white color

(b) Image after extraction by color information

(c) Image after selecting by clustering and heuristic rule 1 and 2

(d) Image after extraction of the gradation

Figure 5: Image analysis by using natural language information

remarkable positions. One of the most remarkable positions in a picture is its center. For this reason, this rule chooses the cluster by measuring the average distance of its pixels from the center. Namely, the system computes the average distance d_j with cluster S_j,

$$d_j^2 = \frac{1}{N_j} \sum_{i=1}^{N_j} |\boldsymbol{X}_{ij} - \boldsymbol{X}_{center}|^2 \tag{4}$$

N_j	:	the number of pixels which belong to cluster S_j
$\boldsymbol{X}_{ij}$	:	the position of pixel i which belongs to cluster S_j
$\boldsymbol{X}_{center}$	:	the central position of the original image

and finds the cluster S_j the average distance of which is the smallest. Then, the system extracts the regions whose pixels belong to the selected cluster as the closest regions to the center of the picture. In this way, Figure 5 (c) is obtained from Figure 5 (b).

Rule 3 Extraction of gradations

The colors of flowers and fruits are generally nonhomogenous. It is caused by the native color variation, the growing environment, and so on. For example, in the Figure 5 (a), the central part of the peach right below is deep red, however, the other part of it is yellow white. Because of this color diversity, the system failed to extract the whole regions of this peach fruit, as shown in Figure 5 (b) and (c). In this way, because of the color variation, the system often failed to extract the whole regions of flowers and fruits.

But, in most cases, the surface colors of flowers and fruits are varied smoothly. That is, the surface of flowers and fruits is generally the gradation of color. Taking advantage of this, we extracted the gradations which adjoin the regions selected by rule 1 and 2. The procedure for extraction of the gradations consists of the following steps:

Step 1. Give the regions selected by the heuristic rule 2 as the regions of flowers and fruits to the system.

Step 2. Find the pixel i ($i = 1 \cdots N$) which adjoins the regions of flowers and fruits.

Step 3. Label these neighboring pixels of the pixel i:

- the neighboring pixels which belong to the regions of flowers and fruits.
- if the pixel i is extracted by the natural language information, the neighboring pixels which belong to the same cluster that the pixel i belongs to.
- if the pixel i is not extracted by the natural language information, the neighboring pixels which are not extracted by the natural language information, either.

Step 4. Select the pixel i the color of which is similar to the colors of the labeled neighboring pixels of it. Namely, the system extracts the pixel i whose color vector $\boldsymbol{C}_i$ satisfies the following inequality.

$$|\boldsymbol{C}_i - \boldsymbol{C}_{ij}|^2 < d^2 \tag{5}$$

$$d^2 = \left(\frac{1}{16}\right)^2 \times 3 \tag{6}$$

where $\boldsymbol{C}_{ij}$ is the color vectors of the labeled neighboring pixel $j (j = 1 \cdots k)$ of the pixel i. If the pixel which satisfies the inequality (5) is obtained, the system regards this pixel as the pixel which belongs to the regions of flowers and fruits, and goes to step 2. If no pixel is found, the process ends.

In this way, Figure 5 (d) is obtained from Figure 5 (c).

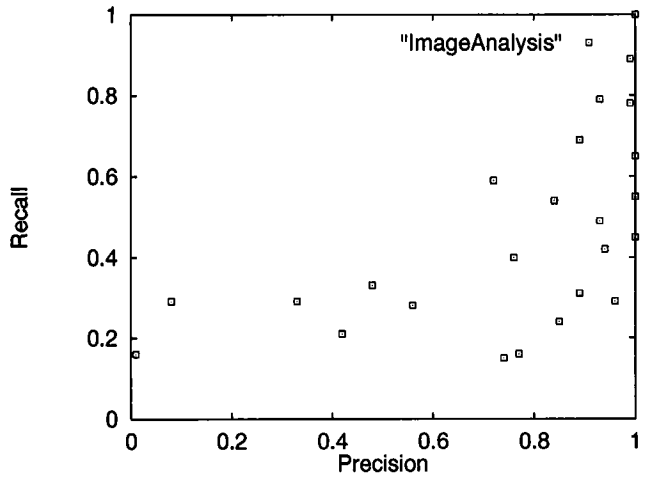


Figure 6: Results of Image Analysis

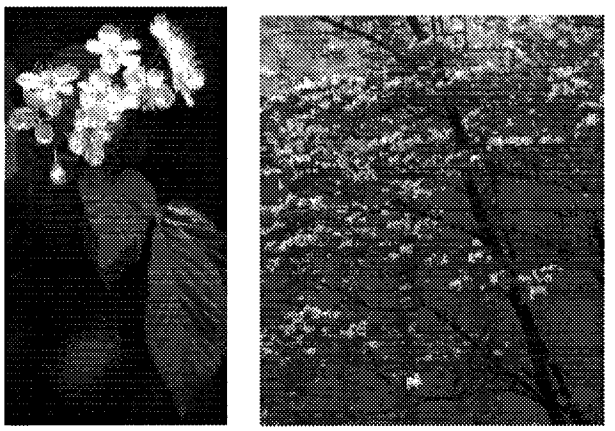
Table 3: Causes of failure of image analysis

Cause	Failure
a flower/fruit was as black as a shade, and the system failed to distinguish between them	3
a flower/fruit was as green as a leaf, and the system failed to distinguish between them	2
the color expressed in the picture was different from the color described in the explanation text.	1
Total	6

Experimental Results

Figure 6 shows the results of the image analysis using the natural language information. We used two measures for evaluating the results of the image analysis: precision and recall. Precision is the percentage of extracted pixels which are actually relevant. Recall is the percentage of pixels that the extraction process found successfully. In most cases, the precision scores were over 80%, and it was only in 6 cases when the precision scores were less than 60%. Table 3 shows the causes of these failures. To yield better precision scores, we must use not only the color information but also various other properties, such as, the shape, size, texture, spatial relationship with the other objects, and so on.

The recall scores were comparatively good in the case of narrow–angle pictures, such as Figure 7 (a), and tended to be worse in the case of wide–angle pictures, such as Figure 7 (b). These low recall scores were caused by the irregularity of lighting. For example, in Figure 7 (b), the top of the tree gets much sun and the bottom gets little sun. From this, the flowers blooming at the top of the tree are brighter than those at the bottom. Because of this difference in light, the clustering process failed to partition pixels into appropriate clusters. In consequence of this, we must investigate the extraction method which would not be affected by lighting.



(a) narrow–angle picture (b) wide–angle picture

Figure 7: Examples of narrow–angle and wide-angle pictures

Retrieval System for Pictorial Book of Flora

It is difficult to retrieve a relevant article from the PBF when a reader does not know the flower name. In such case, it is useful to consult the PBF by the feature of flowers and fruits, especially, by the color of them. For this purpose, we have developed an experimental retrieval system for the PBF. For constructing this system, we made use of the results of the image analysis using natural language information. Two important points for this system are as follows:

- user can consult the PBF without knowing the name of a flower/fruit which he wants to find.
- the system can deal with the fine distinction of colors which the user specifies.

System Overview

Figure 8 shows the overview of the retrieval system for the PBF. The system consists of following modules:

query window shows color samples that a user has specified. He can specify the color of flowers and fruits only by pointing the color sample in the window. Then, the window converts user's request into a color vector.

output window shows a list of the flower names in the order of reliability scores. Also, this window

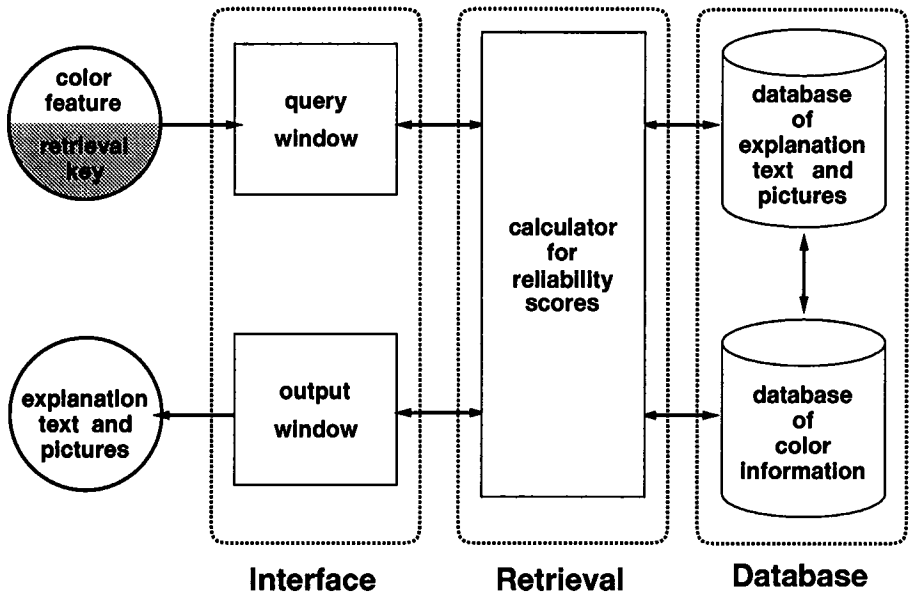


Figure 8: Overview of PBF retrieval system

shows the explanation text and pictures which the user had selected from the list.

database of explanation text and pictures stores the explanation text and the pictures of the PBF. These are used in the natural language analysis and image analysis for extracting the color information on flowers and fruits.

database of color information stores the color information extracted from each picture by the image analysis. Each color information is normalized by making equal the number of pixels extracted by the image analysis.

calculator for reliability scores calculates reliability score for each picture. Each picture is linked to the color information represented by the pixel $i(i = 1 \cdots N)$. The reliability score is the number of the pixel i which satisfies the following inequality.

$$|C - C_i|^2 < d^2 \tag{7}$$

$$d^2 = \left(\frac{1}{16}\right)^2 \times 3 \tag{8}$$

where C and C_i are the color vector of user's request and pixel i, respectively.

Example of Retrieval

In Figure 9, the system answers "*momo* (peach)" and "*anzu* (apricot)" when a user specifies deep red near orange. In addition, the system displays the explanation text and picture of "*momo* (peach)" according to the user's selection.

It should be kept in mind that a user can retrieve this result by specifying deep red near orange as a retrieval key. As we mentioned previously, the system could extract only "yellow white color" from the explanation text as the color of peach fruit. For this reason, if the user specifies "deep red color near orange" as a retrieval key, the system fails to retrieve the article for "*momo* (peach)" only using textual information. On the contrary, using our method, the system extracted the color information of peach fruit that is not only "yellow white" but also "deep red", "orange", and neutral tint among them. That is to say, this is a good example that cooperative use of pattern information and natural language information is quite effective for refined and flexible information processing.

Conclusion

At the moment, this retrieval system is provided only for 18 species of the rose family. We would like to improve the performance by increasing the number of species registered in the system. And, we would also like to study the image analysis using not only the color features, but also the shape, spatial relationship, and so on.

Acknowledgement

This research owes much to the thoughtful and helpful comments of Takako Kageyama, Yuri Hisamatsu, and Harumi Tange.

References

Feiner, S.K.; and McKeown, K.R. 1991. Automating the Generation of Coordinated Multimedia Explanations,

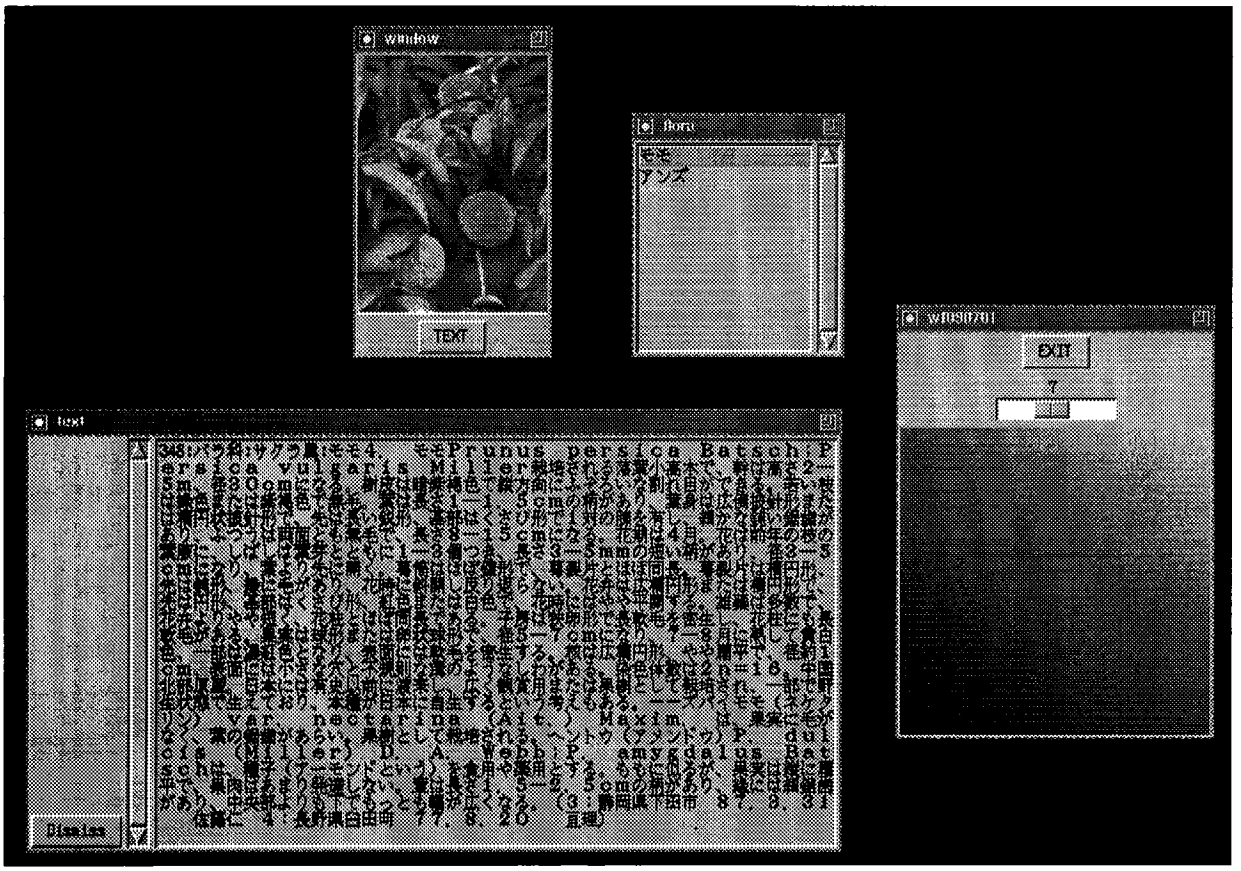


Figure 9: An example of PBF retrieval by color information

IEEE Computer, Vol.24 No.10.

Nakamura, Y.; Furukawa, R.; and Nagao, M. 1993. Diagram Understanding Utilizing Natural Language Text, 2nd International Conference on Document Analysis and Recognition.

Matsumoto, Y. et al. 1996. JUMAN Manual version 3 (in Japanese), Nagao Lab., Kyoto University. [1].

Nagao, M. 1983. Methods of Image Pattern Recognition (in Japanese), CORONA.

Watanabe, Y.; and Nagao, M. 1995. Automatic Acquisition of Hierarchical Information of Words Using Diagrams and Text of Pictorial Book of Flora (in Japanese), IEICE–WGNLC 95–10.

Watanabe, Y.; and Nagao, M. 1996. Diagram Understanding for Pictorial Book of Flora Using Integration of Pattern Information and Natural Language Information, ECAI-96 Workshop on Representations and Processes between Vision and Natural Language.

[1]The source file and the explanation (in Japanese) of Japanese morphological analyzer JUMAN can be obtained using anonymous FTP from
`ftp://pine.kuee.kyoto-u.ac.jp/pub/juman/juman3.11.tar.gz`