Data mining from photographs using the KeyGraph and genetic algorithms

Data mining from photographs using the KeyGraph and genetic algorithms

Hiroyuki MORITA and Takanobu NAKAHARA

Abstract

Data mining is utilized in a variety of commercial and other real-world applications. The subject data are not only numerical, but also character, text, and image data. Of these, image data are the most difficult to mine directly for effective information or knowledge. After digitizing image data, we can apply various mining methods to them.

In this paper, we propose a method to mine useful information from photographs. Using photographs of various foods, our objective is to identify the factors (such as lighting, shape, or color) that give people the impress that the food in the image must be delicious. After we identify some useful explanatory variables using the KeyGraph, we extract these variables from photographs by using genetic algorithms and analyze them with regression analysis. The results confirmed that an image of food can be made more appetizing by increasing the size of the food domain (that is, the part of the image occupied by the food itself) relative to the whole image and by using large numbers of red pixels while avoiding the use of blue pixels.

1 Introduction

As information technology progresses, ever-larger quantities of data can be analyzed. As a result, the range of potential for data mining broadens, and the kinds of data that can be applied to mining methods expand from the purely numerical to text, sound, and image. Among these, however, detailed images, such as photographs, use such large quantities of data that it is still difficult to mine useful information from them.

In day-to-day life, people are exposed to large volumes of image data in forms such as photographs, computer graphics, and movies. Images such as those on banners or billboards and in television commercials are effective at influencing the thoughts and emotions that advertisers want to target. Although the possibility of mining image data is promising, it is not easy to mine valuable knowledge from them, not only because of the enormous quantity of data involved but also because we cannot use the values directly as explanatory variables, even if all of the data is digital. Instead we have to process large amounts of data and extract promising explanatory variables according to each objective.

In this paper, we propose a method to mine useful information from photographs. Using photographs of various foods, the objectives of the mining in this paper are to identify the factors that make people think food in an image must be delicious, and to clarify the influences of these factors. Each photograph shows a different dish. Some dishes are fried, others are boiled, and so on. Naturally, whether or not a dish is delicious is a personal, subjective impression. Likewise, a dish in a photograph may look delicious to one person but not to another. Further, just how appetizing a dish looks will vary even among people who agree it looks good. Finally, two different photographs of the same dish will produce different impressions in the same person, depending on the angle, brightness, and other factors. It follows that if we can know which photographic factors make food look appetizing, we can make images of food more appealing to more people. For example, in a restaurant or supermarket, menus that have photographs that maximize the appeal of the food could be expected to increase sales.

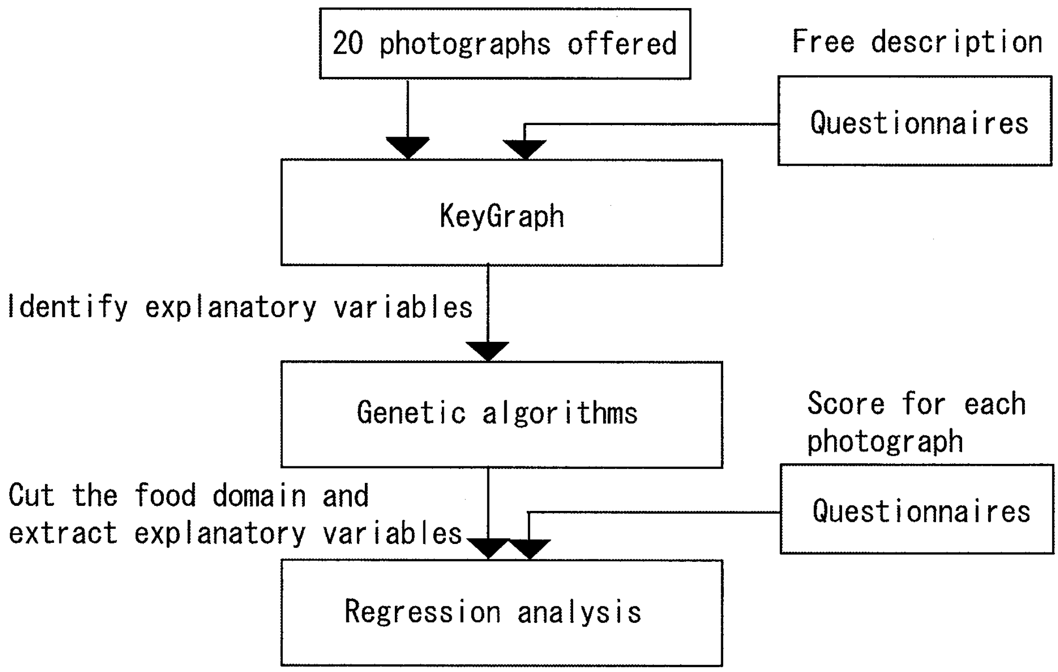

Figure 1: Flowchart for mining

To develop effective mining, we extract some promising factors from photographs, and analyze the influence of each explanatory variable (see Fig.1). In the first step, we obtain 65 data by having study subjects who are university male and female students fill out questionnaires, and use the results to identify some important factors using the KeyGraph method. In the questionnaires, after we show the subjects some photographs of various dishes, each subject records his or her opinion of whether or not the dish looks appetizing, then explains why freely. Then, the factors identified as playing a role in desirability are extracted from photographs using proposed genetic algorithms. Finally, the relationship between the objective values and explanatory variables are analyzed by the regression model. In this second step, another questionnaire is implemented. In this exercise, we show 20 different photographs for 65 subjects. Each subject rates the perceived deliciousness of the food in each photograph on a scale of 1 to 5 (5 is most delicious.)

2 Discrimination of explanatory variables using the KeyGraph

KeyGraph is an effective method for data mining, especially for text mining. KeyGraph is a graph-making method that uses the frequency in the use of a word and the co-occurrence among words, clarifies the important relationships among them, and extracts key factors from them [1]. Although KeyGraph was initially applied to text mining, it is now used for the other mining problems.

Figure 2 illustrates KeyGraph using data from the questionnaire for all subjects. In the KeyGraph figures, • denotes a vertex that occurs frequently, while ◦ denotes a vertex that occurs infrequently but that has a direct connection to •. ◎ and ⊙ denote vertexes whose Jaccard coefficients are large. A solid line connects each node in the case that two vertexes have different linkages to •. On the other hand, if one node is connected by a single linkage, a dotted line is used. The linkages between • and ◦ are connected by other dotted lines.

In Figure 2, we can see that there are three clusters that have strong linkages. A strong linkage means that there are some nodes with more than two degrees of the node. From these clusters, we can observe that the volume, size, and shape of the food are crucial.

Figures 3 and 4 are KeyGraphs for male and female subjects, respectively. The male subjects pointed out that the volume and the color of the food are important. On the other hand, although the female subjects frequently pointed out the color, they did not point out the volume. In addition, the feeling they get when they look at a dish seems important to females.

From these observations, we can guess that the size, color, and shape of a dish strongly influence its desirability. In psychology, it is pointed out that red and orange increase the appetite, whereas blue diminishes it [2].

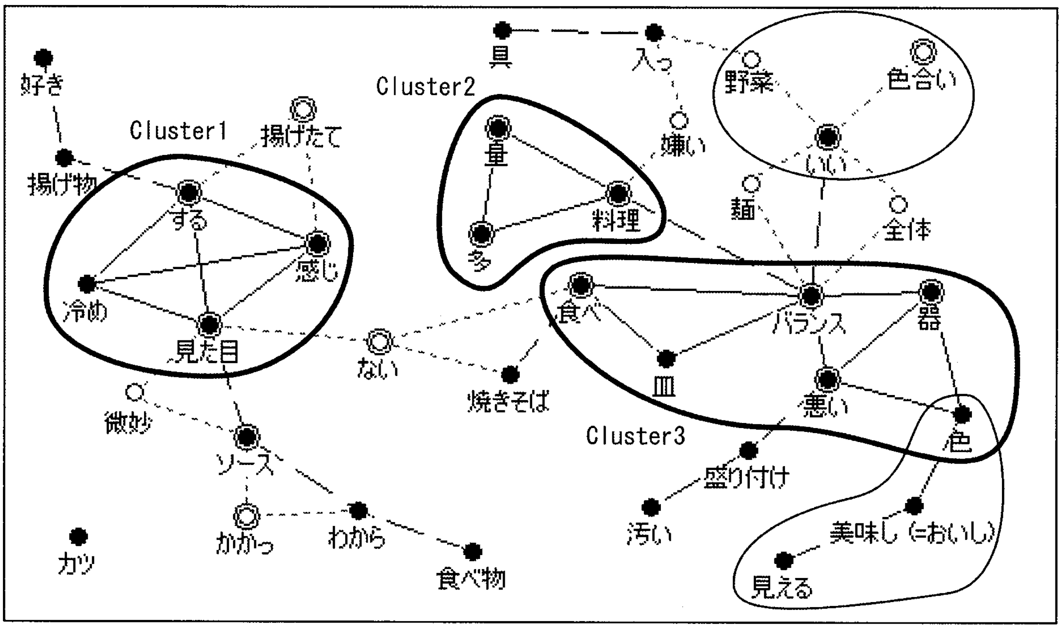

Figure 2: KeyGraph for all subjects

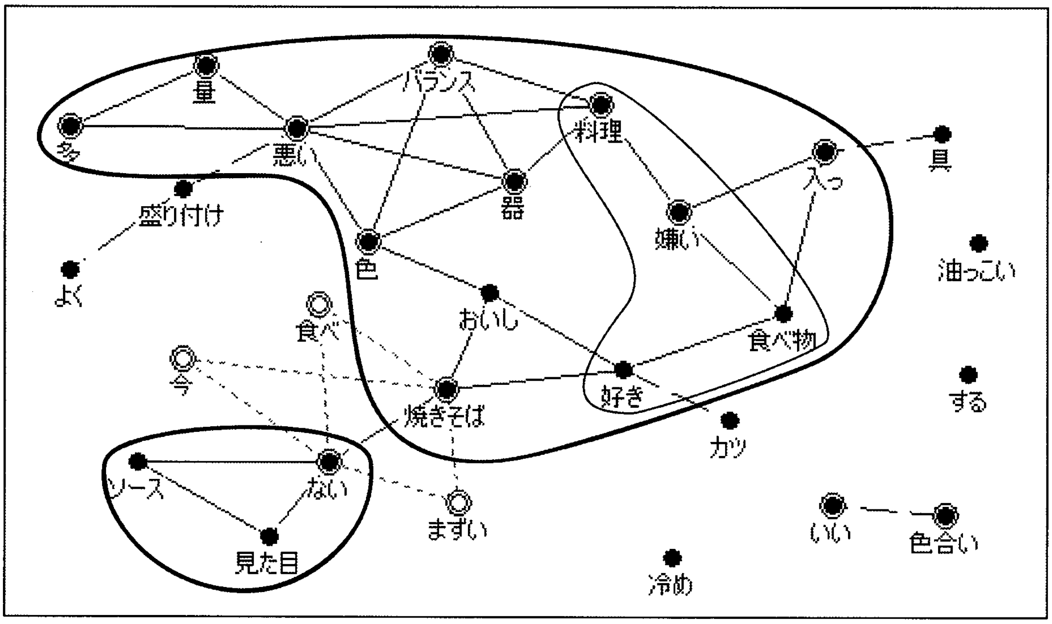

Figure 3: KeyGraph for male subjects

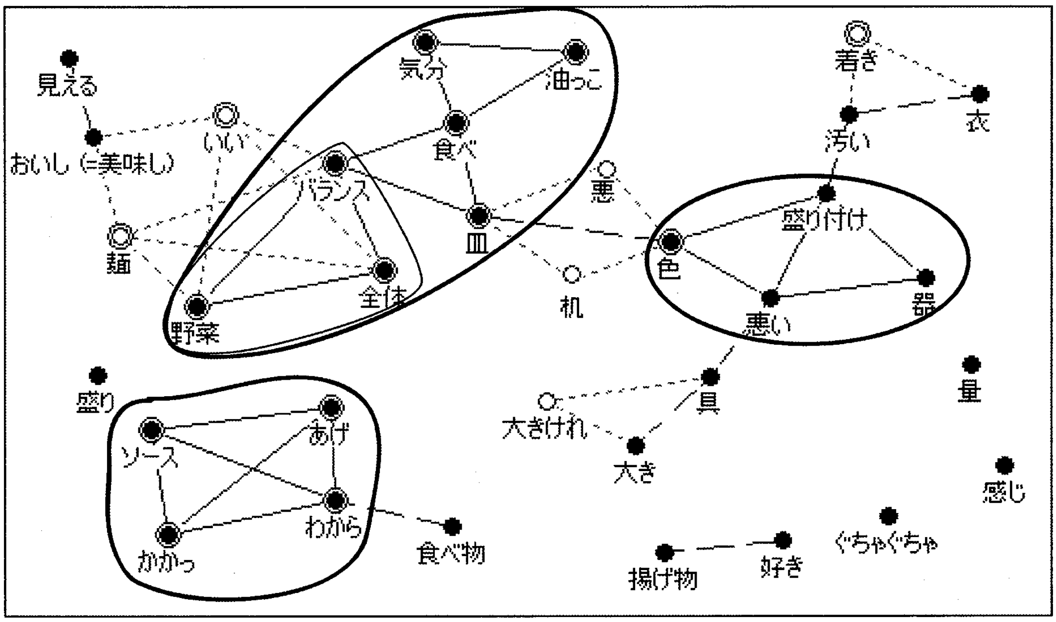

Figure 4: KeyGraph for female subjects

3 Extracting variables from photographs using GA

As discussed in a previous section, the size, color, and shape of a dish are important explanatory variables of a food's desirability. Of course, we cannot use these explanatory variables from photographs directly. So we have to extract them initially. It is possible to convert photographs to digitize photographs, even if they are analog photographs. Digital photographic data are in the form of pixels which are separated *m*(*columns*) $\times$ *n*(*rows*) vertexes. For example, in the case of $m = 320$ and $n = 240$, there are 76800 pixels. In addition, each pixel has color data. In black-and-white photographs, each pixel has data from 0 to u, where u denotes a maximum levels of black and white. In short, as u increases, so does the resolution of the photograph,but so also does the volume of data. In this paper, we use color photographs, so each pixel has three kinds of color data : one kind for each of the primary colors. So, given the same number of pixels and u, a color photograph has to hold three times as much as data as the black-and-white one. In our case, all photographs are color and have 590 $\times$ 393 = 231870 pixels. Each primary color (red, green, and blue) has a value from 0 to 255, R_{ij} , G_{ij} , and B_{ij} , where i and j denote the column and the row, respectively.
From these pixel data, we have to extract the following explanatory variables.

1. The size of the food domain
   The size of the food domain is an important factor. We cut the food domain from the entire image $x_{ij} \in \{0, 1\}$, where $x_{ij} = 1$ denotes that the pixel (i, j) is cut.

And we calculate the ratio of the domain to the entire image. This ratio is used as an explanatory variable that shows the size of the food, $FR=\Sigma_i \Sigma_j x_{ij} \ / \ 590 \times 393$.

2. The color felt by subjects
From the results of the KeyGraph and psychological observations, the colors in the photograph seem to be a very important factor. We regard a pixel's color as the primary color that has the largest value in that pixel. So we count the number of colors and determine which color has the largest value in each pixel. This is important for both the entire image and the food domain. Then, we identify the following six kinds of explanatory variables. NE_r, NE_g, and NE_b denote the numbers of pixels in which red, green, and blue are largest values, respectively, for the entire image. The same is true for NF_r, NF_g, and NF_b except that they refer to the food domain.

3. The subjects' impressions of shape as a factor in the desirability of a dish
The shape of the food domain and the entire image are important, too. Here we introduce six kinds of explanatory variables similar to those described above. SE_r, SE_g, and SE_b denote the average value for each color in the entire image. Similarly, SF_r, SF_g, and SF_b are the average value of each color in the food domain.

Thus we can identify some explanatory variables. To extract these variables from a photograph, we have to decide the scope of the food domain (x_{ij}). This is very difficult, because generally we cannot cut these domains by hand for all photographs. Additionally, the resolution is high in the photographs used in the study. So if a normal coding method is used, we cannot find the optimal solution within an acceptable computational time. In this paper, therefore, the domains must be cut by using a genetic algorithm.

3.1 A genetic algorithm to cut the food domain from color photographs

A genetic algorithm (GA) is a meta-heuristic method to solve combinatorial optimization problems. The GA is an approximate algorithm, so its solutions are not always the optimal ones. In general, however, GA solutions are acceptable, and a lot of successful applications of it have been reported in a variety of fields [4].

One of those fields is image processing [3]. A recent case study by Agui and Nagao[3] extracts images of faces from entire black-and-white photographs. However, as mentioned above, in black-and-white photos there is only one value for each pixel, so extraction of specific image types is easier than in color photographs.

3.1.1 Coding and Initial solutions

In the normal method of coding an image, each pixel is coded either 0 or 1 to decide which food domain the pixel belongs to. This is a very simple and easy method, but in the photographs used in the present study we need 590 × 393 bits for it. It is impossible to code this way because doing so would consume too much memory. In our case, we know that every dish appears on a plate at the center of the photograph. So a new coded method is proposed, in which a starting pixel (sp_i) and a terminating pixel (tp_i) are decided for each row, where $sp_i = 0,..., 295$ and $tp_i = 295,..., 590$. For each row i, if we decide spi and tp_i, a cutting image is shaped. Then for each i, $x_{ij} = 1$ from $j = sp_i$ to $j = tp_i$, otherwise $x_{ij} = 0$.

Initial solutions are generated randomly in general. From previous experiments, we know that such initial solutions are very poor. So, for the rows near the top and bottom, the range of random values is set near the center, whereas for rows near the middle the range is set near the edge (see Fig.5).

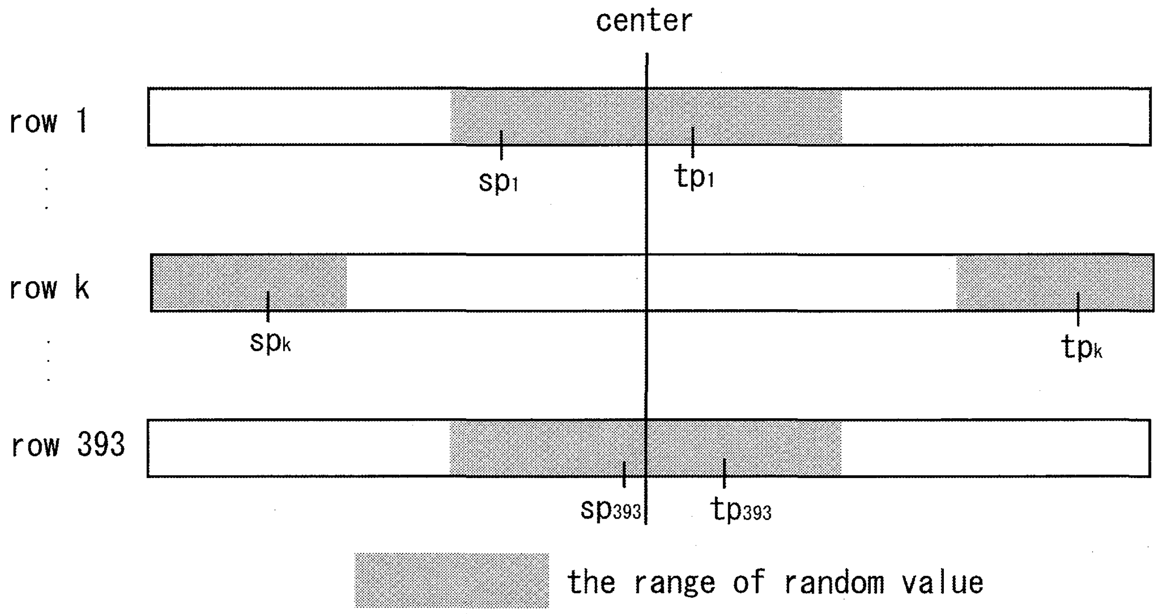

Figure 5: Coding and initial solutions

3.1.2 Crossover and mutation

In a crossover operator, a uniform crossover is used. In this case, two parent individuals (*$parent_A$* and *$parent_B$*) are selected randomly, and mask bits are generated randomly for each row. Next, if the mask bit is 0, *$offspring_A$* inherits a gene (in this case, a row) from *$parent_A$*, and if the mask bit is 1, a gene is inherited from *$parent_B$*. *$offspring_B$* inherits each gene contrary to *$offspring_A$* from *$parent_A$* and *$parent_B$*.

In a mutation operator, an individual is selected from the population randomly, and a gene k (a row) is selected randomly in this individual. Then sp_k and tp_k are changed randomly in their respective ranges of random values.

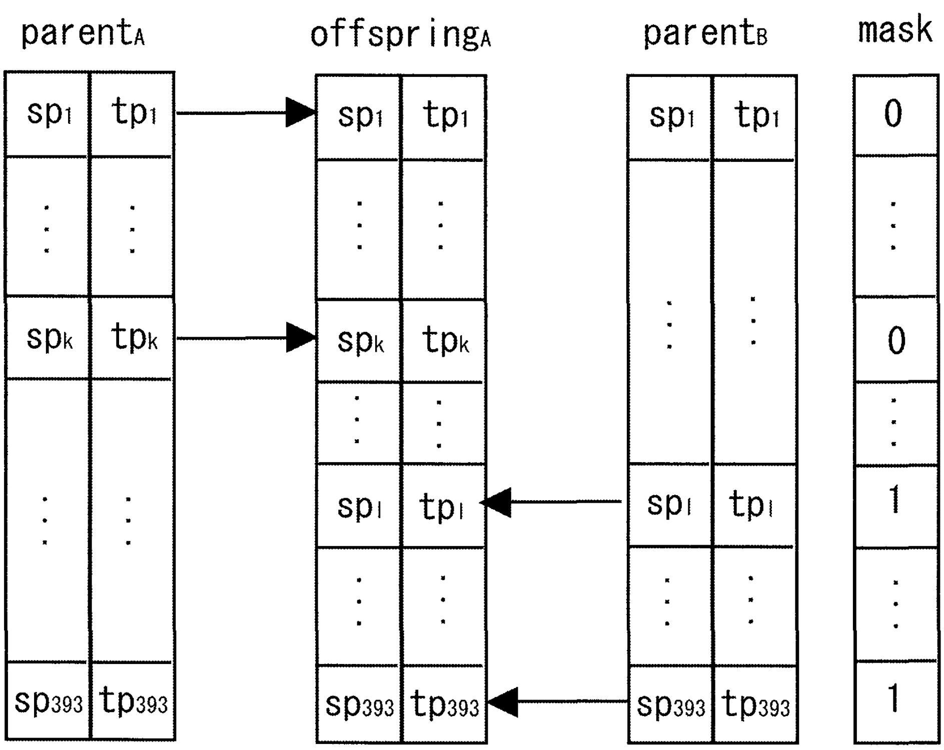

Figure 6: Uniform crossover in our paper

3.1.3 Additional operator

From preliminary experiments, we know that the performance is not good when only the above GA operators are used. Observation of the results points to two causes for this. One is that each gene is so independent that there are few linkages between neighborhood genes. Because of this, the frontier of the food domain is dispersed entirely. The other cause is that it is dificult to improve individuals entirely using only GA operators. These operators can change individuals, but it is a small part of the food domain. So it is necessary to change individuals more entirely.

Here, we add two additional operators. One is a method similar to that for mutation. As a mutation operator, an individual and a gene k in this individual are selected randomly. Then, sp_k and tp_k are changed by $(sp_{k-1} + tp_{k+1}) / 2$ and $(tp_{k-1} + tp_{k+1}) / 2$, respectively, and the fractions are rounded. This operator makes the shape of the food domain smooth. Another operator increases or decreases the size of the food domain entirely. After an individual is selected randomly, an increase or decrease is decided randomly. In the case of an increase, for every gene, sp_k is decreased by 1, and tp_k is increased by 1. In the case of a decrease, the converse is true. In this way, the food domain is increased or decreased entirely and easily.

3.1.4 The food domain

In experiments, the population size (|pop|) is set to 10, and 1000 generations are repeated. Each operator generates new |pop| solutions at every generation, and if a new solution is not identical with pop, it is included in pop as a candidate solution. At the end of the generation, if |pop| is more than 10, an elite solution and nine solutions selected randomly are kept for the next generation. The experiments in this study were were performed on a Pentium 4 PC running at 2.6GHz.Coding was done in the C language.

The objective function is used as follows,

$$F(x_{ij}) = \left| \frac{\sum_i \sum_j RGB_{ij} \cdot C_{ij}^1}{\sum_i \sum_j C_{ij}^1} - \frac{\sum_i \sum_j RGB_{ij} \cdot C_{ij}^0}{\sum_i \sum_j C_{ij}^0} \right|$$

$$C_{ij}^k = \begin{cases} x_{ij} = k & 1 \\ otherwise & 0 \end{cases}$$

where, $RGB_{ij} = R_{ij} + G_{ij} + B_{ij}$. The objective value increases as the food domain is cut exactly. So the individual with the largest objective value is the most adaptive.

Table 1 illustrates the entire results of the GA. By using the GA, objective values are improved for all instances. The actual image cut from a photograph is shown in figure 7. Although the results are not perfect, we can see that the GA can cut the food domain approximately. Using x_{ij} , the other explanatory variables are calculated.

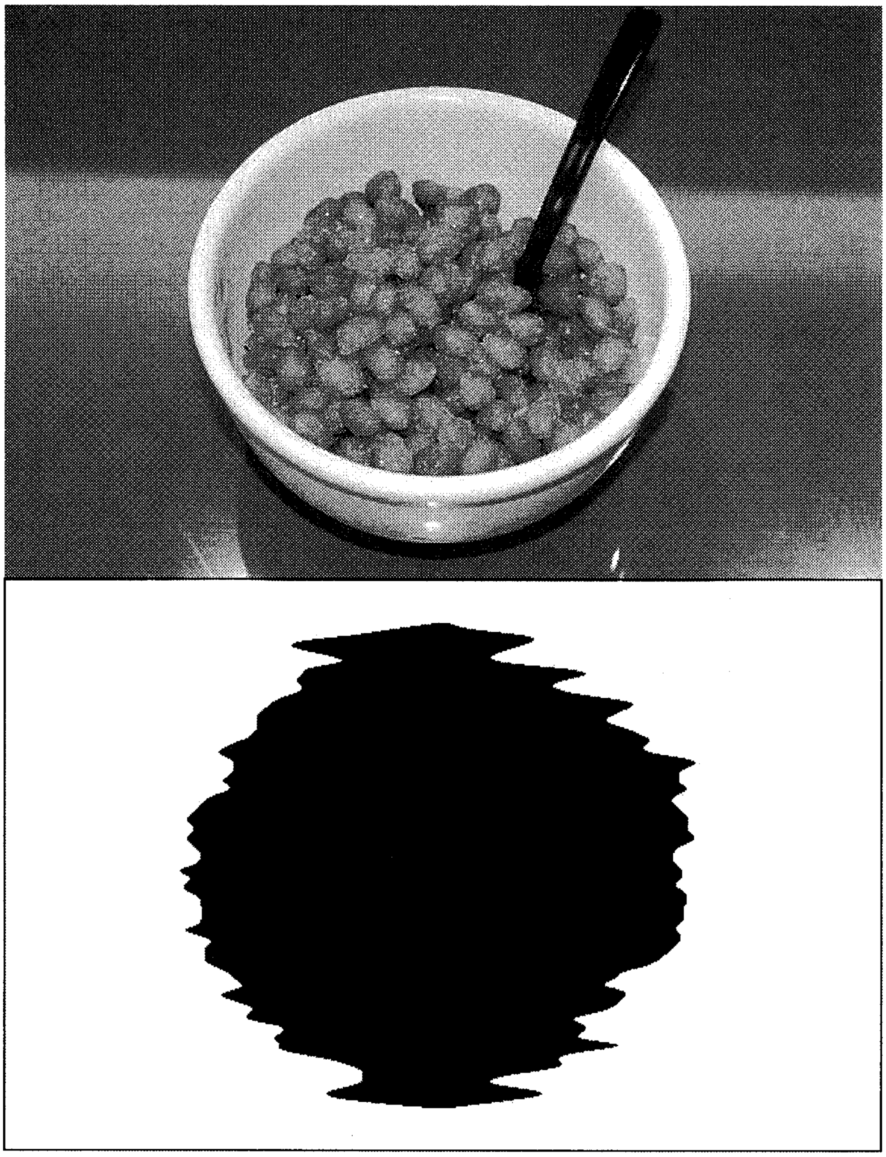

Figure 7: Cutting image from a picture

Table 1: Computational results

Picture	*Bint*[1]	*Bfinal*[2]	*FR*	*Picture*	*Bint*	*Bfinal*	*FR*
pic1	81	94	41.65	pic11	134	144	27.66
pic2	43	176	48.94	pic12	95	157	25.22
pic3	277	303	23.97	pic13	112	123	27.54
pic4	63	140	18.51	pic14	281	298	25.48
pic5	221	225	27.75	pic15	77	189	51.15
pic6	102	123	31.08	pic16	182	194	33.08
pic7	285	307	29.12	pic17	111	151	26.67
pic8	141	167	37.18	pic18	53	176	50.57
pic9	84	156	21.18	pic19	68	102	27.35
pic10	7	83	41.2	pic20	55	107	45.82

1 Best objective value in initial population
2 Best objective value in final population

4 Statistical analysis

In this section, statistical analysis is performed. For the dependent variables, we use the average value of the points that the subjects used, on the scale of 1 to 5 to rate the desirability of the dish in each photograph. Candidate explanatory variables are NE_r, NE_g, NE_b, NF_r, NF_g, NF_b, SE_r, SE_g, SE_b, SF_r, SF_g, SF_b, and FR.

Table 2: Model summary

Variables	R Square	Adjusted R Square	Std. Error of the Estimate
NF_r SF_b	0.697	0.662	0.2912

Table 3: The model fit

	Sum of Square	df	Mean Squares	*F*	*Sig.*
Regression	3.321	2	1.661	19.587	0.000
Residual	1.441	17	0.0848		
Total	4.762	19			

Table 4: Coefficients

	Unstandardized Coeffcients			
	B	Std. Error	t-value	p-value
(constant)	2.300	0.220	10.457	0.000
NF_r	0.00002	0.000	6.240	0.000
SF_b	-0.0058	0.002	-2.905	0.010

Tables 2 - 4 illustrate the results for the best model. FR correlates closely with NF_r, so we cannot use both variables simultaneously. In a comparison of the two models, the adjusted R-square in the case when NF_r is larger than that FR is used. In the better model, two effective variables, NF_r and SF_b, are identified.

The results confirm the following. First, the size of the food domain is important, and it is better to zoom in on it, when we take photographs. This is consistent with the results suggested by KeyGraph. Secondly, in the food domain, it is effective to use a large proportion of red pixels. Finally, the use of large numbers of blue pixels should be avoided. These suggestions are consistent with previous psychological research. Although the results are not surprising, they are valuable insofar as they confirm the results of mining from photographs and questionnaires.

5 Results

In this paper, we propose a method for mining image data and clarify some of the factors that create the impression that food shown in images is delicious. In this method of mining, explanatory variables are identified by KeyGraph, the food domain is cut from the photograph using GA, the explanatory variables are extracted from the domain, and the influence of each of these explanatory variables is confirmed by the regression analysis. As a result, we clarify that increasing the sizes of the food domain relative to the whole image and increasing the proportion of red in the food domain improve the appeal of the food, while the use of blue in the food domain decreases the appeal. These could serve as guidelines for photographing food menus.

In future research, it will be helpful to obtain more data by using a wider variety of photographs. In the present study, explanatory variables are not assumed in advance. This may create the impression that only a few factors have much influence. From the beginning, if we can use photographs that assume some explanatory variables that KeyGraph can identify, other factors may emerge as effective factors.

KeyGraph and genetic algorithms are powerful tools for data mining. KeyGraph points out promising explanatory variables precisely, and the genetic algorithm in our paper cuts effective food domains from the whole images. This combination is very useful for this type of data mining. The present study demonstrates one case of the effectiveness of these tools for mining image data. The tools can be applied to mining problems involving other kinds of image data as well.

References

[1] Y. Osawa, et al., "KeyGraph", The Institute of Electronics, Information and Communication Engineers, Vol. J82- D1, No.2, pp.391 - 400 , 1999 (in Japanese).
[2] O. Katayama and M. Tajima, *Shokuhin to Iro*, Korin, 2003 (in Japanese).
[3] T. Agui and T. Nagao, *Introduction to Image Processing using programming language C*, Shoukoudou, 2004, pp.187 - 212. (in Japanese)
[4] T. Bäck et al. ed., *Handbook of Evolutionary Computation*, Institute of physics publishing and Oxford university press, 1997.