A TRAINABLE RETRIEVAL SYSTEM FOR CARTOON CHARACTER IMAGES

Miki Haseyama and Atsushi Matsumura

School of Engineering, Hokkaido University
N-13 W-8 Kita-ku Sapporo 060-8628, JAPAN
{mikich, matsu}@media.eng.hokudai.ac.jp

ABSTRACT

This paper proposes a novel method to retrieve cartoon character images in a database or network. In this method, partial features of an image, defined as Regions and Aspects, are used as keys to identify cartoon character images. The similarities between a query cartoon character image and the images in the database are computed by using these features. Based on the similarities, the cartoon images same or similar to the query image are identified and retrieved from the database. Moreover, our method adopts a training scheme to reflect the user's subjectivity. The training emphasizes the significant Regions or Aspects by assigning more weight based on the user's preferences and actions, such as selecting a desired image or an area of an image. These processes make the retrieval more effective and accurate. Experiment results verify the effectiveness and retrieval accuracy of the method.

1. INTRODUCTION

The popularity of the World Wide Web has created needs for search engines, which find contents with retrieval methods. One of the greatest needs is "image retrieval", i.e., retrieving the desired images from databases or network, in which cartoon image retrieval is a big share.

Many image retrieval methods have been proposed [1]-[4]. These methods use various keys, such as words in "the key word retrieval method"[1], sample images in "the query by images method (QBI)"[2], same patterns in "the pattern matching method"[3], and color histograms in "the color based retrieval method"[4]. However, these keys may not work well for retrieving the cartoon character images due to the following reasons: A large number of data which do not include the same character images are identified, if the name or title of a cartoon image is used as a key word; and since the same cartoon character images sometimes have different colors or drastically changed shapes.

We propose a novel retrieval method specialized for cartoon character images. In our method, partial features of an image, which are defined as Regions and Aspects, are used as retrieval keys. The Regions are the areas which segment the image based on its boundaries; Each Region has nine Aspects, which represent different properties of the Region from various viewpoints. With the Regions and Aspects, the similarity of two images, an input image and an image in the database, is computed. Moreover, we adopt a training scheme to reflect the user's subjectivity during the process of the similarity computation. The training emphasizes the significant features by assigning more weight so that it improves the retrieval accuracy. Experiment results show the effectiveness and accurate performance of the method.

This research is partly supported by Grant-in-Aid for Scientific Research (C) 14550343, Japan Society for the Promotion of Science (JSPS).

2. COMPUTATION OF SIMILARITIES

Our method consists of two parts: computation of similarities and training. In this section, the first part is presented. Its flow is as follows:

i. Given an input image I^α.
ii. Segment I^α into N Regions, R_i^α $(i = 1, 2, \cdots, N)$.
iii. For each Region R_i^α, nine Aspects, $\mu_l^\alpha(i)$, $\mu_c^\alpha(i)$, $\mu_p^\alpha(i)$, $\mu_v^\alpha(i)$, $\mu_g^\alpha(i)$, $\mu_o^\alpha(i)$, $\mu_{a_{1,1}}^\alpha(i)$, $\mu_{a_{2,1}}^\alpha(i)$ and $\mu_{a_{1,2}}^\alpha(i)$, are computed.
iv. The similarities between I^α and the images in the database, $I^{\beta_1}, I^{\beta_2}, \cdots,$ and I^{β_L}, are computed by using the above Regions and Aspects as feature keys.
v. The images $I^{\beta_1}, I^{\beta_2}, \cdots,$ and I^{β_L} are sorted in descending order from the best matching and their thumbnails are displayed accordingly.

The details of (ii), (iii) and (iv) are described as follows.

2.1. Segmentation of an Image into Regions

Cartoon character images consist of several areas enclosed with boundaries in many cases as shown in Fig.1. We define these areas as *Regions*. For instance, the input image I^α in Fig.1 is composed of five regions R_i^α $(i = 1, \cdots, 5)$ with the center of gravity G^α. For each region R_i^α, its center of gravity is denoted as g_i^α, and θ_i^α is the angle between the horizontal line l and line $\overline{G^\alpha g_i^\alpha}$, which are counterclockwise rotated from l.

2.2. Aspects of a Region

Although we can calculate the similarities with the regions, additional computable indexes, or the properties of a region, the relationship among the regions, etc., are necessary in order to obtain the similarities more objectively and accurately. We define nine kinds of parameters called *Aspects*, which represent features and properties of the regions from various viewpoints. They are classified into three different groups, and detailed below.
(i) The Aspects $\mu_l^\alpha(i)$, $\mu_c^\alpha(i)$ and $\mu_p^{\alpha_i}$, which indicate the properties of the target Region R_i^α, are defined as follows:

- $\mu_l^\alpha(i)$ is computed from the number of the opened lines in R_i^α as

$$\mu_l^\alpha(i) = \frac{1}{1 + (\text{the number of the opened lines in } R_i^\alpha)}. \quad (1)$$

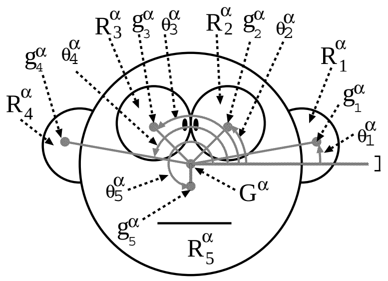

Fig. 1. This image consists of five areas which are labeled as R_1^α, ..., R_4^α, and R_5^α.

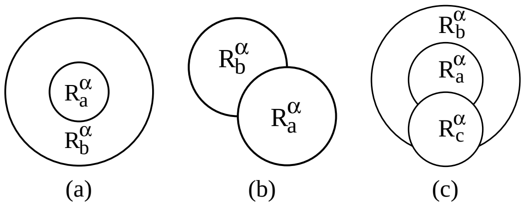

Fig. 2. Definition of three types of occlusions for $\mu_o^\alpha(i)$.

- $\mu_c^\alpha(i)$ indicates the complexity of the boundary of R_i^α as

$$\mu_c^\alpha(i) = \frac{4\pi(\text{the size of } R_i^\alpha)}{(\text{the length of the boundary of } R_i^\alpha)^2}. \tag{2}$$

- $\mu_p^\alpha(i)$ indicates the partial analysis of the boundary of R_i^α. It is described by the positive/negative maxima/minima values of the curvature from the boundary of R_i^α based on the definitions of Process Grammar[5], which is a method to describe the shape of boundaries.

(ii) The Aspects $\mu_v^\alpha(i)$ and $\mu_g^\alpha(i)$, which indicate the relationship between the target Region R_i^α and the input image I^α, are defined as follows:

- $\mu_v^\alpha(i)$ indicates the ratio of sizes between R_i^α and I^α,

$$\mu_v^\alpha(i) = \frac{\text{the size of } R_i^\alpha}{\text{the size of } I^\alpha}. \tag{3}$$

- $\mu_g^\alpha(i)$ indicates the relative distances between g_i^α and G^α,

$$\mu_g^\alpha(i) = \frac{|g_i^\alpha - G^\alpha|}{\text{Heywood Radius}^{\,1} \text{ of } I^\alpha}. \tag{4}$$

(iii) The Aspects $\mu_{a_{1,1}}^\alpha(i)$, $\mu_{a_{2,1}}^\alpha(i)$, $\mu_{a_{1,2}}^\alpha(i)$ and $\mu_o^\alpha(i)$, which indicates the relationship among the target Region R_i^α and the other Regions, are defined as follows:

- $\mu_{a_{1,1}}^\alpha(i)$, $\mu_{a_{2,1}}^\alpha(i)$ and $\mu_{a_{1,2}}^\alpha(i)$ indicate the relative position of R_i^α and are computed using the following equations:

$$\mu_{a_{m,n}}^\alpha(i) = \begin{cases} \angle g_{i-m}^\alpha g_i^\alpha g_{i+n}^\alpha/\pi & \angle g_{i-m}^\alpha g_i^\alpha g_{i+n}^\alpha < \pi \\ \angle g_{i-m}^\alpha g_i^\alpha g_{i+n}^\alpha/\pi - 1 & \angle g_{i-m}^\alpha g_i^\alpha g_{i+n}^\alpha \geq \pi \end{cases}$$

$$m = 1, 2, \quad n = 1, 2.$$

- $\mu_o^\alpha(i)$ indicates the degree of occlusions for R_i^α. In this method, "R_a^α occludes R_b^α" is defined for the following three conditions as shown in Fig.2: (a) R_a^α is enclosed with R_b^α; (b) R_a^α occludes part of R_b^α; and (c) represents "R_c^α occludes R_a^α" since (b) and (c) exist simultaneously, but (b) is given higher priority. By using these definitions, $\mu_o^\alpha(i)$ is computed by the order for the Regions, where the occluding Regions are higher priority than the occluded Region. If these regions are indicated as $R_{o_1}^\alpha < R_{o_2}^\alpha < \cdots < R_{o_i}^\alpha < \cdots < R_{o_N}^\alpha$, where $R_{o_1}^\alpha, R_{o_2}^\alpha, \cdots,$ and $R_{o_N}^\alpha$ correspond to $R_1^\alpha, R_2^\alpha, \cdots,$ and R_N^α without overlapping, $\mu_o^\alpha(o_i)$ is defined as

$$\mu_o^\alpha(o_i) = \frac{i-1}{N-1}. \tag{5}$$

2.3. Computing similarity

By using these Regions and Aspects, the similarity between the input image I^γ and an image I^β ($\beta \in \{\beta_1, \beta_2, \cdots, \beta_L\}$) in the database is computed in two steps: (1) computes the similarity between two Regions, and (2) computes the similarity between the two images by using the results from the previous step.

2.3.1. Similarity between two Regions

Suppose each Region, $R_1^\gamma, R_2^\gamma, \cdots, R_i^\gamma, \cdots,$ or R_N^γ, in I^γ is mapped to one of the Regions, $R_1^\beta, R_2^\beta, \cdots, R_j^\beta, \cdots,$ and R_M^β, in I^β without overlapping. The similarity $s_{i,j}^{\gamma\beta}$ between the two Regions, R_i^γ and R_j^β, is computed as

$$s_{i,j}^{\gamma\beta} = \sum_{k=o,c,v,g,a,l,p} w_k^\beta(j) f_k(i,j). \tag{6}$$

For $k = o, c, v, g, a, l$,

$$f_k(i,j) = \frac{2}{\sqrt{2\pi}\sigma_k^\beta(j)} \int_0^{\left|\mu_k^\gamma(i)\right|} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_k^\beta(j)}{\sigma_k^\beta(j)}\right)^2\right) dx,$$

where $\sigma_k^\beta(j)$ ($k = o, c, v, g, a, l, p$) are used to normalize $f_k(i,j)$, which are standard deviations of $\mu_k^\beta(j)$ ($k = o, c, v, g, a, l, p$) of the images in the database. And $w_k^\beta(j)$ ($k = o, c, v, g, a, l, p$) are the weights of the Aspects, which indicate the importance of the Aspects, and are controlled by training that will be described in the next section. The function $f_p(i,j)$ is defined as

$$f_p(i,j) = \frac{2}{\sqrt{2\pi}\sigma_p^\beta(j)} \int_0^{\left|\mu_i^\gamma(p)-\mu_j^\beta(p)\right|} \exp\left(-\frac{x^2}{2\sigma_p^\beta(j)^2}\right) dx,$$

$$\left|\mu_i^\gamma(p) - \mu_j^\beta(p)\right| = 1 - \frac{1}{1 + d\left(\mu_p^\gamma(i), \mu_p^\beta(j)\right)}, \tag{7}$$

where $d\left(\mu_p^\gamma(i), \mu_p^\beta(j)\right)$ is the minimum number of deformation required in order that the shape of R_i^γ transforms into R_j^β, which is used in Process Grammar.

[1]The method to compute the virtual radius from the size of an area, defined as V, by regarding the Region as the circle: $\text{Heywood Radius} = \sqrt{\frac{V}{\pi}}$[6].

2.3.2. *Similarity between two images*

The similarity $S^{\gamma\beta}$ between two images I^γ and I^β is computed as

$$S^{\gamma\beta} = \max \sum_i n_i^\gamma n_{r_i^\beta}^\beta F(i, r_i^\beta) s_{i,r_i^\beta}^{\gamma\beta}, \tag{8}$$

where n_i^γ and n_j^β are the weights of the Regions which indicate the importance of R_i^γ and R_j^β ; r_i^β indicates the label of the Region which is mapped to R_i^γ; and $F(i,j)$ is computed as the following if the number of the Regions in I^γ is larger than the number of the Regions in I^β,

$$F(i,j) = \max(f(i, r_i^\beta, a_{1,1}), f(i, r_i^\beta, a_{1,2}), f(i, r_i^\beta, a_{2,1})),$$
$$\mu_{r_i^\beta}^\beta(a_{m,n}) = \angle g_{r_{i-m}^\beta}^\beta g_{r_i^\beta}^\beta g_{r_{i+n}^\beta}^\beta,$$

otherwise

$$F(i,j) = \max(f(r_j^\gamma, j, a_{1,1}), f(r_j^\gamma, j, a_{1,2}), f(r_j^\gamma, j, a_{2,1})),$$
$$\mu_{r_j^\gamma}^\gamma(a_{m,n}) = \angle g_{r_{j-m}^\gamma}^\gamma g_{r_j^\gamma}^\gamma g_{r_{j+n}^\gamma}^\gamma,$$

wherer_j^γ indicates the label of the Region which is mapped to R_j^β.

3. TRAINING

Some particular features are particularly significant to increase the degree of similarity in many cases. For example, the features which indicate hairs or eyes may be more significant than others. By signifying these features, it can improve the retrieval accuracy. In order to incorporate this observation, we control the weights of Regions, n_i^γ, and the weights of Aspects, $w_i^\gamma(k)$. Further, since these features depend on the user's purpose and preference; the following training system is derived in order to reflect the user's subjectivity for computation of the similarity,

After the thumbnails of $I^{\beta_1}, I^{\beta_2}, \cdots, I^{\beta_L}$, which are selected as the similar images to the input image I^γ, have been shown, the user can perform one or more actions of the following, and then the weights are controlled accordingly:

Action1 Selecting the desired image from the thumbnails.

After the desired image, I^{β_S}, is selected, the weights of the Regions $n_{r_i^{\beta_S}}^{\beta_S}$ and the weights of the Aspects $w_{r_i^{\beta_S}}^{\beta_S}(k)$ in I^{β_S} are given by

$$n_{r_i^{\beta_S}}^{\beta_S\,new} = n_{r_i^{\beta_S}}^{\beta_S\,old} + n_i^\gamma \frac{u_i^\gamma}{L-1}, \tag{9}$$

where $n_{r_i^{\beta_S}}^{\beta_S\,new}$ indicates recomputed $n_{r_i^{\beta_S}}^{\beta_S}$, in the case that $n_{r_i^{\beta_S}}^{\beta_S\,old}$ indicates $n_{r_i^{\beta_S}}^{\beta_S}$; and u_i^γ is the total number of the Regions which satisfy $s_{i,r_i^{\beta_S}}^{\gamma\beta_S} > s_{i,r_i^{\beta_t}}^{\gamma\beta_t}$. Further,

$$m_{r_i^{\beta_S}}^{\beta_S\,new}(n) = m_{r_i^{\beta_S}}^{\beta_S\,old}(n) + \frac{u_i^\gamma(n)}{L-1}, \tag{10}$$

where n is one of the Aspects; $m_{r_i^{\beta_S}}^{\beta_S\,new}(n)$ indicates recomputed $m_{r_i^{\beta_S}}^{\beta_S}(n)$ in the case that $m_{r_i^{\beta_S}}^{\beta_S\,old}(n)$ indicates $m_{r_i^{\beta_S}}^{\beta_S}(n)$; and $u_i^\gamma(n)$ is the total number of the Aspects, which satisfy $|\mu_i^\gamma(n) - \mu_{r_i^{\beta_S}}^{\beta_S}(n)| > |\mu_i^\gamma(n) - \mu_{r_i^{\beta_t}}^{\beta_t}(n)|$.

Action2 Selecting an important Region in I^γ.

After R_i^γ is selected as an important Region, the weight of R_i^γ, n_i^γ, becomes a constant value k_1.

Action3 Selecting a particularly different image from the thumbnails to decrease its significance.

After I^{β_s} is selected as a particularly different image, each weight of the Region in I^γ , n_i^γ $(i \in \{1, \ldots, N\})$, is recomputed as follows:

$$n_i^\gamma = k_2 \frac{t_i^\gamma}{L-1} + k_3, \tag{11}$$

where k_2 and k_3 are constant values; and t_i^γ is the total number of the Regions which satisfy $s_{ir_i^{\beta_s}}^{\gamma\beta_s} > s_{ir_i^{\beta_t}}^{\gamma\beta_t}$ $(t = 1, 2, \cdots, L)$.

Action1 is essential to obtain the desired image, and Action2 and Action3 are optional and can be used to assist retrieving the similar images.

4. EXPERIMENTAL RESULTS

We show the effectiveness of the proposed method in this section. The images used for the experiments are scanned binary images of cartoon characters: Fig.3 (a) – (g) are the input images and (A) – (F) are the images in the database. In the experiments, the desired images are defined as the same characters in the cartoon; the constants are given as $k_1 = 3.0$, $k_2 = 1.5$ and $k_3 = 0.5$.

4.1. The experiments of computing similarities

First, experiments to examine the effectiveness of the computation of similarities are performed. Both of the similarities and the orders of similarity are shown together in Table 1. The results of the same character image as each input image are indicated in bold. The larger the value, the more similar the two images and the higher the order of similarity. For example, for input image (a) in the first row, image (A) is the most similar and (D) is the least similar. The results indicate the similarity between the input image and the same character image is the highest in 6 out of 7 cases, i.e., the retrieval effectiveness is 86%. In the other experiments, which utilizes another 25 different characters, the retrieval effectiveness is 83.5%.

4.2. Training experiments

Next, we examine the effectiveness of the training. Assuming that image (c) in Fig. 3 is the input image, the following three trainings are performed separately, and the similarities and the orders are shown in Table 2:

i) Action1
Selecting image (B) as the desired image.

ii) Action2 + Action1
Selecting the Region which indicates hair in image (c) as an important Region, then selecting (B) as the desired image.

iii) Action3 + Action1
Selecting image (E) as the unnecessary image, then selecting (B) as the desired image.

These results show that training improves the retrieval effectiveness with 100% successful rate since all of the results show the

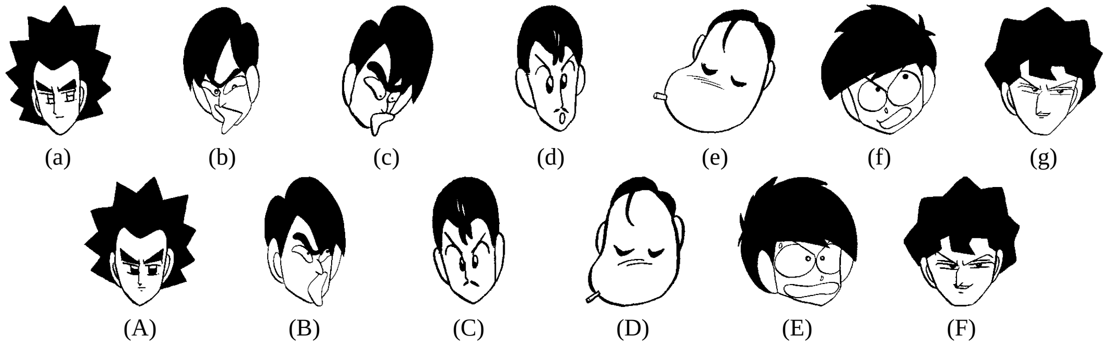

Fig. 3. (a)-(g) are the input images and (A)-(F) are the images in the database. The same character pairs are: {a, A}, {b, c, B}, {d, C}, {e, D}, {f, E}, and {g, F}.

same desired character image has the highest similarity to the input image.

Further, the proposed method is applied to another 20 different characters, which are drawn by four different authors, and its retrieval effectiveness after the training is also 100%. Therefore, based on these experiments, it can be recognized that the proposed method including the training is effective and accurate for the cartoon character retrieval.

5. CONCLUSIONS

This paper has presented a retrieval method for cartoon character images. The proposed similarity was effective for retrieval of the same cartoon character images, and the training scheme greatly improves the retrieval accuracy. This method has several possible applications such as, search engines for cartoon images and indexing cartoon images of databases.

ACKNOWLEDGEMENTS

We would like to thank Mr. Masayuki Katayama, Mr. Koichi Aoyama at Kodansha, Mr. Takahashi at Takeshobou, Kodansha and Takeshobou, for allowing the use of the cartoon characters in our research.

Table 1. Similarities and orders between 2 images, the bold values are the same character pairs. (a) – (g) are the query input images while (A) – (F) are the images in the database.

	(A)	(B)	(C)	(D)	(E)	(F)
(a)	**0.5813 1st**	0.4596 4th	0.4653 3rd	0.4259 6th	0.4456 5th	0.5212 2nd
(b)	0.4557 5th	**0.5200 1st**	0.4546 6th	0.2858 3rd	0.4866 2nd	0.4760 4th
(c)	0.4966 2nd	**0.4893 3rd**	0.4797 4th	0.3951 6th	0.5134 1st	0.4125 5th
(d)	0.4903 2nd	0.4538 5th	**0.5008 1st**	0.4439 6th	0.4602 4th	0.4728 3rd
(e)	0.4878 5th	0.4350 4th	0.4592 2nd	**0.6062 1st**	0.4259 6th	0.4451 3rd
(f)	0.4578 2nd	0.4409 3rd	0.4168 5th	0.4175 4th	**0.5286 1st**	0.4099 6th
(g)	0.4804 2nd	0.4005 6th	0.4403 3rd	0.4104 5th	0.4233 4th	**0.5583 1st**

6. REFERENCES

[1] T. Hayashi and M. Hagiwara, "Image query by impression words-the IQI system," IEEE Transactions on Consumer Electronics, vol. 44, no. 2, pp. 347–352, May 1998.

[2] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele and Peter Yanker, "Query by Image and Video Content: The Qbic System," IEEE Computer, pp.23-32, Sep. 1995.

[3] Chih-Chih Liu Chen, "3D-List: a data structure for efficient video query processing," IEEE Transactions on Knowledge and Data Engineering, vol.14, No.1, pp.1541-1546. July 2002.

[4] Xia Wan, CC Jay Kuo, "Color Distribution Analysis and Quantization for Image Retrieval," Storage and Retrieval for Image and Video Databases IV, pp.8-16. 1995.

[5] M.Leyton, "A Process-Grammar for shape," Artificial Intelligence, vol.34, pp.213–247, 1988.

[6] H. Heywood, "Particle shape coefficients," J. Imperial College Chemical Society, 8, pp. 25–33.

Table 2. Similarities and orders of the training experiments. Image (c) is the query input image against images (A) – (F). (i) Action1; (ii) Action1 then Action2; and (iii) Action1 then Action3.

	(A)	(B)	(C)	(D)	(E)	(F)
(i)	0.4966 3rd	**0.6698 1st**	0.4797 4th	0.3951 6th	0.5134 2nd	0.4125 5th
(ii)	0.4966 3rd	**0.6626 1st**	0.4797 4th	0.3951 6th	0.5134 2nd	0.4125 5th
(iii)	0.4966 3rd	**0.6720 1st**	0.4797 4th	0.3951 6th	0.5134 2nd	0.4125 5th