SICE Annual Conference 2012
August 20-23, 2012, Akita University, Akita, Japan

Features for image retrieval: the impression degree of a human image by overexposure occurring in the facial area

Tatsuki Murakami[1], Yoichi Kageyama[1], Makoto Nishida[1] and Yoichi Shirasawa[2]

[1]Akita University, Akita, Japan
(Tel : +81-18-889-2473; E-mail: kageyama@ie.akita-u.ac.jp)
[2]Alpha Systems, Inc., Kanagawa, Japan

Abstract: To improve the retrieval accuracy of content-based image retrieval systems, it is important to reduce the "semantic gap" between the visual features and the richness of human semantics. Overexposure in the facial area of an image, which makes parts of the face look unnaturally white, directly affects the impression of the image. If the impression of a human image can be automatically calculated on the basis of its content, an automatic indexing system focusing on human semantics can be developed. This can help to improve the efficiency of image retrieval and accuracy in searching images on the basis of facial expressions. In this study, we investigated the overexposure impression degree (OID), a feature that expresses the impression degree of a human image by overexposure. We also developed an algorithm for judging the OID.

The algorithm includes three steps. First, it detects areas of the face, eyes, lips, and skin of a human in each image in a preprocessing phase. Next, it uses the color difference in the CIELAB color space and extracts overexposed areas from the face on the basis of color information on overexposure. Finally, the algorithm classifies an image into three types of OIDs on the basis of the extent of overexposure in prone areas of the face.

The results of experiment conducted with 11 evaluators for analyzing images suggest that the proposed method can search and retrieve images that match the search queries with an F-measure of 94.7%.

Keywords: overexposure, image retrieval, human image, face

1. INTRODUCTION

Nowadays, people have access to a vast number of pictures because of the widespread use of digital cameras, and they share these pictures in numerous ways. Among all the categories of subjects of pictures, a "human" is an important category, and numerous opportunities are available for taking human subjects' pictures. Accordingly, an efficient image retrieval system that can search and retrieve human images from numerous images needs to be developed.

Examples of opportunities and occasions where pictures of humans can be acquired include parties, weddings, etc. Moreover, users who search human images often target one person from an event as the subject for image retrieval. The events are often held indoors, and a high contrast between lit and shaded areas is likely to occur in the images taken in such places. Thus, overexposure, which makes some parts of the image seem unnaturally white, can occur in such images. If overexposed images are included in the databases of image retrieval systems, the images retrieved may not match the requirement of the users' feeling; this requirement is, for example, to get the image whose human photograph "natural."

In a conventional image retrieval system, images are indexed in advance to aid search and retrieval; this retrieval is achieved by comparing the indexes and the keywords used for searching. However, manual indexing is impractical, because the workload increases with increasing number of pictures stored on the system. Hence, we need a content-based image retrieval (CBIR) system [1]. In particular, an automatic indexing system that focuses on the characteristics of feeling is needed for automatically calculating the impression of a human image by CIBR. The automatic calculation contributes to the improvement in the efficiency of human image retrieval, thus facilitating search of images that suit the requirement of the users' feeling.

In this study, we investigated the overexposure impression degree (OID), a feature that expresses the impression degree of a human image by overexposure. Moreover, we developed an algorithm for calculating the OID and determined the retrieval performance from the results of this method.

2. IMAGE DATA USED

Image data of full-faced subjects were acquired using a digital camera (Canon EOS Kiss Digital X equipped with Speedlite 580EXII flash). The conditions for capturing the image data used in this study are listed below. Table 1 presents details of the image data used.

(1) The distance between the camera and the subject was fixed at about 1.8 m.
(2) Exposure settings were fixed manually (f-stop: 5.6, shutter speed: 1/125 s, ISO speed: 1600).
(3) Bounce lighting was used: flash was directed onto the ceiling so that reflected light fell on the subject.

Table 1. Details of the image data used.

Dataset	A	B	C
Subjects (people)	5	5	9
Flash strength (kinds)	12 (12 kinds of strong flash in order of flash strength)		
Indoor brightness (kinds)	1 (700lx)	3 (400lx, 700lx, 1000lx)	
Flash angle (kinds)	1 (60°)		2 (60°, 75°)
Total (number of pictures)	60	180	648

Images from Dataset A were used to study the relationship between the brightness and the impression of the human images. Images from Dataset B were used to set the parameters of the proposed method. Images from Dataset C were used for estimating the retrieval performance of the proposed method. The data used in the study was acquired in accordance with ethical regulations regarding research on humans at Akita University, Japan.

3. DEFINITION OF THE FEATURE

3.1 Expert knowledge about overexposure

An image may be described as overexposed when there is a loss of highlighted detail. Some of the features of "overexposure," as stated by experts of photography in the interview are given below [2]. In this time, we asked about something regarding overexposure: For example, conditions when the digital image is overexposed, something to cause which overexposed area of the human image gives people unnaturally, and etc. were.

(a) In the overexposed parts, pixels have a maximum value or a value close to the maximum and are uniformly distributed.
(b) The nose, cheeks, and forehead are areas most prone to overexposure.
(c) The subject (a person) of the image will not look good if the entire face or the abovementioned areas are overexposed.

In this study, on the basis of points (a) and (b), we defined an overexposed pixel as being almost white, almost equal to the color of the surroundings, and present in the facial area. Figs. 1(a) and (b) show an example of a human facial image and its overexposed areas, respectively, as visualized on the basis of expert knowledge.

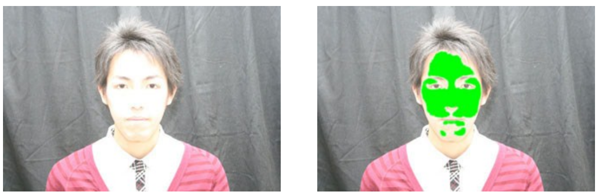
3.2 Relationship between brightness and the impression of a human image

In order to study the OID feature, five facial image sets (FISs) were created from Dataset A for a subjective evaluation using a passport-sized photo. An example of an FIS is shown in Fig. 2. Here, N is the brightness level. The numbers are given in ascending order of flash strength. The FIS image impressions were evaluated by 14 people in 2 experiments: a subjective evaluation and a questionnaire.

Subjective evaluation: This evaluation is carried out as follows. The FIS images are placed randomly, and an evaluator rearranges the images in order of improving impression. The rearranged order is classified into the following subjective evaluation ratings: (A) very good, (B) good, (C) neither good nor bad, (D) bad, and (E) very bad. The relationship between impression and overexposure in the facial image is shown in Fig. 3 as an average of the subjective evaluation values (ASV) of facial images that have the same N. Facial images with $N = 4$ or 5 have high ASVs of about 4.4–4.7 (the circle in Fig. 3). Facial images having lower ASVs can be brighter or darker than the abovementioned images. This result suggests that a facial image that has a good impression has near-optimal brightness.

Questionnaire: In the questionnaire, the subjective evaluation points ((A)–(E)) are broadly classified into three groups: GOOD ((A) and (B)), AVERAGE, ((C)), and BAD ((D) and (E)). An evaluator fills out a questionnaire on the standard subjective evaluation for each group. As a result, the most popular reason given for classifying an image into the BAD group is found to be "the entire face is white," and that for classifying an image into the AVERAGE group is "some parts of the face are white." These results suggest that the extent of exposure on a face can signify the impression of a human image.

On visual inspection of Dataset A, we found that the FIS images having $N = 5$ and 6 are overexposed on the nose but have good impression because the ASV is about 3.8. In this study, in addition to the abovementioned definition, the OID is also defined by taking into consideration expert knowledge (see Table 2). Examples of the OID are shown in Fig. 4. The OID



(a) Original image of human face. (b) Overexposed areas of image.

Fig. 1. Example of overexposed parts.

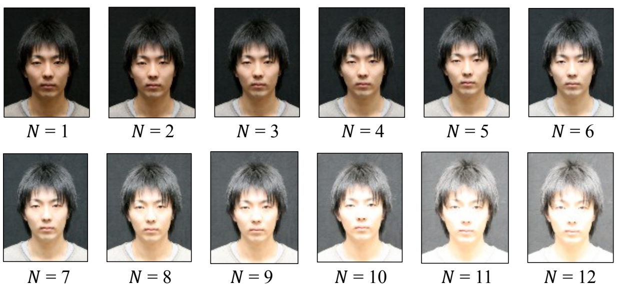


Fig. 2. Facial image set (FIS) for subject A_5.

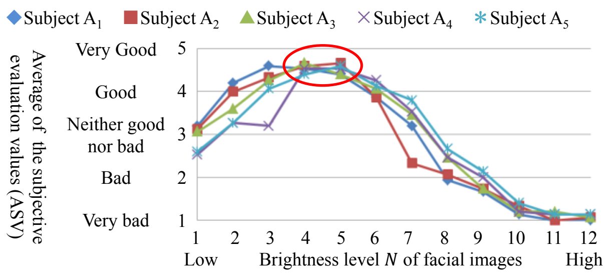


Fig. 3. Relationship between facial image brightness and ASV.

Table 2. Definition of OID.

OID (numerical data)	Overexposed parts	Impression of the image
Zero/Low (1)	None/Nose	Good
Average (2)	Nose, cheeks and forehead	Neither good nor bad
High (3)	Entire face	Bad

can express the impression degree of a human image by overexposure occurring in the facial area.

4. PROPOSED METHOD

4.1 Preprocessing method

In order to calculate the OID on the basis of the overexposure in prone areas, the proposed method detects areas of the face, eyes, lips, and skin of a person in each image in a preprocessing phase (see Fig. 5). First, it detects the facial area using an object detector of OpenCV [3, 4]. Next, it detects the lip and skin areas by applying a fuzzy interface technique on the basis of color information [5, 6]. Finally, it detects the eye area using the edge and color information [7].

4.2 Extraction of overexposed parts

On the basis of the definition of overexposed pixels (Section 3.1), white pixels and similar-colored pixels are extracted using the proposed method to obtain the overexposed parts in the facial area (see Fig. 6). Then, E_{ab} (the difference between color-1 (C_1) and color-2 (C_2) in the CIELAB color space) is employed, as defined below [8].

$$E_{ab}(C_1, C_2) = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \cdots\cdots (1)$$

E_{ab} of same amount in a color value produces a change of the same visual importance; therefore E_{ab} can be expressed abovementioned deference quantitatively. In the CIELAB color space, there are $L*$ for lightness, $a*$ and $b*$ for the color opponent. $L*$, $a*$, and $b*$ for color-x (C_x) are L_x, a_x, and b_x, respectively. The white pixels (see Fig. 6(b)) satisfy the following equation.

$$E_{ab}(C_{x,y}, W) \leq Th_{Eab} \cdots\cdots (2)$$

Here, $C_{x,y}$ is the color of pixel (x,y), and the values of $L*$, $a*$, and $b*$ for white color (W) are 100, 0.0, and 0.0, respectively. Pixels of similar colors (see Fig. 6(c)) satisfy equation (3).

$$\frac{1}{8} \sum_{i=x-1}^{n=x+1} \sum_{j=y-1}^{m=y+1} E_{ab}(C_{x,y}, C_{i,j}) \leq Th_{Eab} \cdots\cdots (3)$$

The overexposed parts are extracted using the logical product of white pixels and similar-colored pixels.

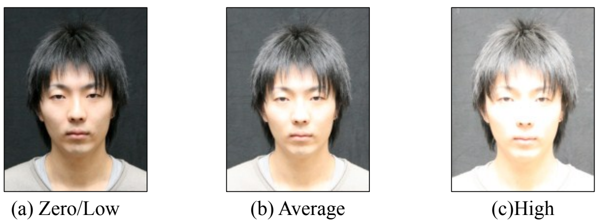


Fig. 4. Examples of OID.

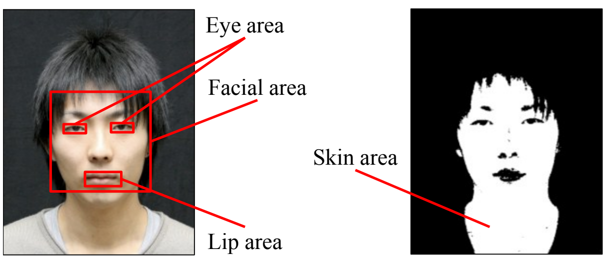


Fig. 5. Examples of detected areas of the face, lips, eyes, and skin.

4.3 Analysis of overexposure

On the basis of the definition of the OID (Section 3.2), the proposed method can be used to analyze the occurrence of overexposure in prone areas and classify human images into three types of OIDs.

In order to analyze this occurrence, the overexposed areas of the nose, cheeks, and forehead are determined using the areas of the face, eyes, and lips, as shown in Fig. 7. In this figure, for example, area S consists of four elements: S_x, and S_y (coordinates of the upper-left vertex of area S), and S_w, and S_h (width and height, respectively, of area S).

Next, the OID is determined using the overexposure rate (OR), obtained by dividing W by S, as shown in Fig. 8. Here, W is the number of overexposed pixels and S is the number of pixels corresponding to the skin area. For instance, OR_S (OR of the area S) is calculated using W_S (W of the area S) and S_S (S of the area S). The OID is calculated in the following steps.

(a) "Zero/Low" OID: Visual inspection of Dataset A

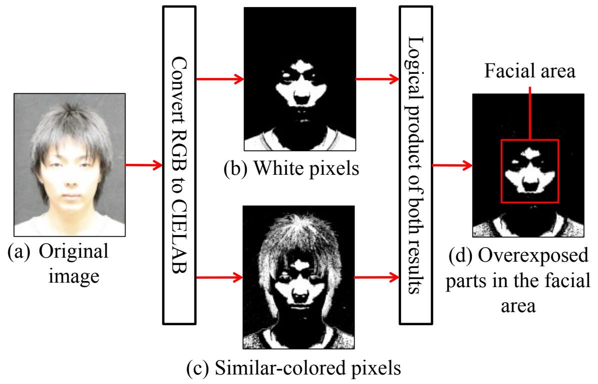


Fig. 6 Example of extraction of overexposed parts from facial area.

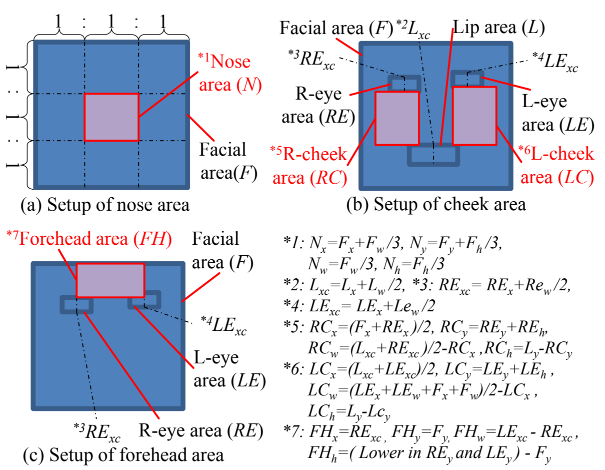


Fig. 7. Examples of overexposure-prone areas.

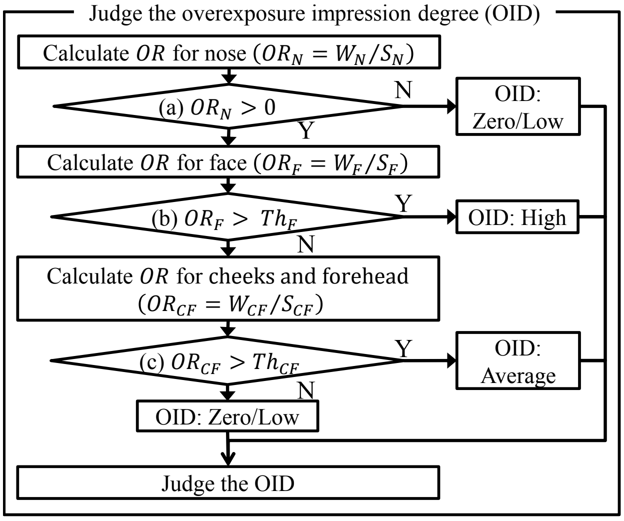


Fig. 8. Flowchart for judging OID.

revealed that the ridge of the nose was prone to overexposure. Hence, we conclude that there is no overexposure if OR_N (*OR* of the nose) is 0 and determine the OID to be "Zero/Low."

(b) "High" OID: If overexposure occurs in the entire face, we judge the OID to be "High," i.e., if OR_F (*OR* of a face) is greater than threshold Th_F.

(c) "Average" OID: If overexposure occurs in either of the cheeks or the forehead, we judge the OID to be "Average." Specifically, the following three *OR* values are calculated. OR_{RC} (*OR* of the right cheek), OR_{LC} (*OR* of the left cheek), and OR_{FH} (*OR* of the forehead). OR_{CF} is set equal to the highest *OR* calculated above, and the OID is judged to be "Average" if OR_{CF} is greater than Th_{CF}.

5. EXPERIMENTS

5.1 Visual judgments

The impressions of all images from Dataset B (180 images) and Dataset C (648 images) were judged visually by 11 evaluators in an experiment, and the judgment results were denoted as "VJ."

During the visual judgment, first, an evaluator classified the impressions of all images individually into the following three subjective evaluation points.

(a) "Overexposure does not occur" or "Overexposure occurs but the impression is good" (OID: Zero/Low).
(b) Overexposure exists and the impression is bad (OID: High).
(c) Neither (a) nor (b) (OID: Average).

Next, on the basis of the *VJ* obtained for the OIDs, the subjective evaluation points were treated as an evaluation value (*EV*): (a), (b), and (c) were "1," "3," and "2," respectively. Using 11 *EV*s for a human image i (i = 1–828), we calculated a_i (average of i-th *EV*), v_i (variance of i-th *EV*), and V_a (variance of all a_i).

Finally, we classified i human images into the three types of the OIDs using a_i, v_i, and V_a, and obtained the visual judgment results. These results were obtained by first comparing v_i and V_a. If v_i was larger than V_a, the OID was judged as being "Average" because of the assumption of variations in the evaluation. If v_i was less than or equal to V_a, we classified the images on the basis of the following conditions.

(1) If a_i is 1.5 or less, the OID is "Zero/Low."
(2) If a_i is 2.5 or more, the OID is "High."
(3) If neither of the above conditions is satisfied, the OID is "Average."

5.2 Setting parameters

In order to judge the OID using the proposed method, Th_{Eab}, Th_F, and Th_{CF} (see Section 4) should be set beforehand. In this study, we used the allowable color differences (ACD) of level 1 to 4. When an ACD level is lower, the color difference to determine whether 2 colors are same is lower. In this study, a threshold of ACD is defined as Th_{Eab}; Th_{Eab}s were 0.6, 1.2, 2.5, and 5.0 which were corresponded to the differences in levels 1 to 4 [7]. One value each of OR_F and OR_{CF} was calculated from each image of Dataset B. Further, one value each of Th_F and Th_{CF} was selected from these calculated OR_F's and OR_{CF}'s, respectively. We set the optimal parameters as follows.

Method for setting Th_F and Th_{CF}: In order to set Th_F, we classified the images from Dataset B into two categories—"High" and "Zero/Low/Average" of OID—by comparing OR_F with a candidate for Th_F, as shown in (b) in Fig. 8. Further, to set Th_{CF}, we classified the images into categories of "Average/High" and "Zero/Low" by comparing OR_{CF} with a candidate of Th_{CF}, as shown in (c) in Fig. 8. Next, by comparing these classification results and the visual judgment results for Dataset B, we calculated the classification performances (*CP*); CP_F (CP_{CF}) is the percentage of the number of the cases in which the classification result of Th_F (Th_{CF}) and the visual judgment result coincide. A higher value of *CP* indicates better performance. Thus, we set Th_F (Th_{CF}) as equal to a candidate of Th_F (Th_{CF}), when CP_F (CP_{CF}) becomes maximum.

5.3 Estimation of performance of proposed method

In this study, we determined the performance of conventional human image retrieval system under the assumption that the OID is used to search for images; we then determined whether the OID of the searched images was greater than or less than "High" or "Average" or "Zero/Low" OID. The following four queries were used to search for images.

(A) Search for images having good impression. That is to search for images having "Zero/Low" OID.
(B) Search for images not having bad impression. That is to search for images having "Zero/Low/Average" OID.

In order to estimate the retrieval performance of the proposed method, we calculated *Precision*, *Recall*, and *F-measure*, which are defined as follows.

$$Precision = \frac{R}{N} \quad (4)$$

$$Recall = \frac{R}{C} \cdots\cdots (5)$$

$$F\text{-}measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} = \frac{2R}{N+C} \cdots (6)$$

Here, N is the number of images selected in this retrieval; C, the number of images matched to a search query; and R, the number of images selected in this retrieval and matched to a search query. The values of N, C, and R for each search query are shown below. Here, the OID and the visual judgment result of the i-th human image (i = 181–828) are OID_i and VJ_i, respectively.

(A) N: OID_i < "Average," C: VJ_i < "Average,"
R: OID_i < "Average," and VJ_i < "Average."

(B) N: OID_i < "High," C: VJ_i < "High,"
R: OID_i < "High" and VJ_i < "High."

("Low" = 1, "Average" = 2, and "High" = 3)

6. EXPERIMENTAL RESULTS

6.1 Setting parameters

We determined the relationship between the thresholds (Th_F and Th_{CF}) and the classification performances (CP_F and CP_{CF}), as shown in Fig. 9. As a result, we set Th_F (Th_{CF}) as equal to a candidate of Th_F (Th_{CF}) when CP_F (CP_{CF}) becomes maximum (see Table 3).

6.2 Determination of retrieval performance of proposed method

In order to estimate the retrieval performance using the OID, first, we acquired the OIDs of the images in Dataset C. In this study, Th_F and Th_{CF} were each set at

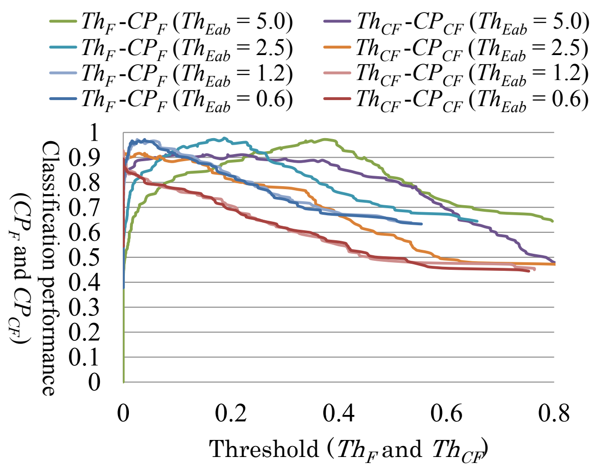


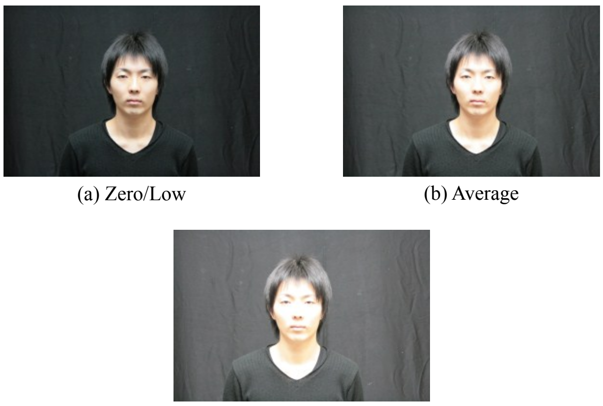
Fig. 9. Relationship between thresholds and classification performances.

Table 3. Thresholds used in this study.

Th_{Eab}	0.6	1.2	2.5	5.0
CP_F	0.972	0.972	0.978	0.972
Th_F	0.0389	0.0248	0.1729	0.3521
CP_{CF}	0.894	0.928	0.917	0.911
Th_{CF}	0.0002	0.0000	0.0399	0.2293

Th_{Eab} (Table 3). The results of the OID judgment are shown in Fig. 10. The concordance rates between the OIDs and the visual judgment results are presented in Table 4. The total concordance rates were over 80%, except for Th_{Eab} = 5.0, but the concordance rate of "Average" OID was low for Th_{Eab}. There was no case where the OID differed greatly from VJ, except in the case of the images for which VJ was "Zero/Low" and OID was judged as "High" (Th_{Eab} = 5.0) (see the colored cell in Table 5). This suggests that the proposed method cannot cause an egregious error. Moreover, the concordance rates of "High" and "Zero/Low" were high—greater than 80%—except when Th_{Eab} = 5.0. In particular, for Th_{Eab} = 0.6, these rates were highest (94.5% and 97.4%, respectively). As a result, it can be concluded that the proposed method can judge OIDs with high accuracy, especially when Th_{Eab} = 0.6.

Next, using the OIDs and VJs explained above, we calculated *Precision*, *Recall*, and *F-measure* as the retrieval performance indexes. These indexes were calculated for each search query ((A) and (B)) (Section 5.3). Table 6 presents the retrieval performance indexes



(a) Zero/Low (b) Average

(c) High

Fig. 10. Results of judging the OID.

Table 4. Concordance rates between OIDs and VJs.

OID	Concordance rate			
	Th_{Eab} = 0.6	Th_{Eab} = 1.2	Th_{Eab} = 2.5	Th_{Eab} = 5.0
Zero/Low	94.6% (298/315)	83.2% (262/315)	80.3% (253/315)	72.4% (228/315)
Average	48.0% (47/98)	55.1% (54/98)	54.1% (53/98)	21.4% (21/98)
High	97.4% (229/235)	100% (235/235)	96.6% (227/235)	100% (235/235)
Total	88.6% (574/648)	85.0% (551/648)	82.3% (533/648)	74.7% (484/648)

Table 5. Number of mismatched cases.

VJ \ OID	Zero/Low	Average	High
Zero/Low		17/53/62/73	0/0/0/14
Average	29/7/7/5		22/37/38/72
High	0/0/0/0	6/0/8/0	

*In a cell, the number of mismatched cases starts from the left in order of Th_{Eab} = 0.6, 1.2, 2.5, and 5.0

Table 6. Retrieval performance indexes.

Th_{Eab} = 0.6		***Precision* (*R*/*N*)**	***Recall* (*R*/*C*)**	***F-measure***
Search query	**(A)**	93.9% (298/327)	94.6% (298/315)	92.8%
	(B)	98.5% (391/397)	94.7% (391/413)	96.5%
Average		96.2%	94.6%	94.7%
Th_{Eab} = 1.2		***Precision* (*R*/*N*)**	***Recall* (*R*/*C*)**	***F-measure***
Search query	**(A)**	92.5% (262/269)	83.2% (262/315)	89.7%
	(B)	100% (376/376)	91.0% (376/413)	95.3%
Average		96.2%	87.1%	92.5%
Th_{Eab} = 2.5		***Precision* (*R*/*N*)**	***Recall* (*R*/*C*)**	***F-measure***
Search query	**(A)**	91.2% (253/260)	80.3% (253/315)	88.0%
	(B)	97.9% (375/383)	90.8% (375/413)	94.2%
Average		94.6%	85.6%	91.1%
Th_{Eab} = 5.0		***Precision* (*R*/*N*)**	***Recall* (*R*/*C*)**	***F-measure***
Search query	**(A)**	87.5% (228/233)	72.4% (228/315)	83.2%
	(B)	100% (327/327)	79.2% (327/413)	88.4%
Average		93.8%	75.8%	85.8%

for all search queries; a colored cell denotes an index that is greater than 90%. Except for Th_{Eab} = 5.0, each index showed rates as high as greater than 90%. In particular, in the case of Th_{Eab} = 0.6, all indexes of each search query were greater than 90%; these are the largest values obtained for all Th_{Eab}; *Precision* was 96.2%; *Recall*, 94.6%, and *F-measure*, 94.7%. This suggests that human images can be searched using the OID with high accuracy, especially when Th_{Eab} = 0.6.

ACKNOWLEDGMENT

The authors thank Ms. C. Ishizawa and Mr. T. Takahashi, Akita University, Japan, for their help in conducting the experiments.

REFERENCES

[1] Y. Liu, D. Zhang, G. Lu, W. Ma: "A survey of content-based image retrieval with high-level semantics," Pattern Recognition, Vol.40, No.1, pp.262-282 (2007).
[2] DIGITABLE Web Site, http://www.digitable.info/.
[3] P. Viola and M. Jones: "Rapid object detection using a boosted cascade of simple features," Proc. IEEE CVPR (2001).
[4] Willow Garage Web Site, http://www.willowagrage.com/.
[5] Y. Shirasawa, M. Nishida, K. Nishi: "A Study on lip extraction due to fuzzy reasoning by using color information," T.IEE Japan, Vol.125-C, No.9, pp.1430-1437 (2005).
[6] Y. Shirasawa, K. Ootaki: Extraction method robust to differences in skin color temperature, Image lab, Vol.9, No.11, pp.36-41 (2008).
[7] Y. Toduka, Y. Kageyama, M. Nishida, Y. Shirasawa, K. Ootaki: Study to determine a particular person in digital images, ITE Annual winter conference 2008, 8-5 (2008).
[8] The Color Science Association of Japan: "Handbook of Color Science [Third Edition]," University of Tokyo Press (2011).