Face Image Annotation in Impressive Words by Integrating Latent Semantic Spaces and Rules

Hideaki Ito, Yuji Kawai, and Hiroyasu Koshimizu

School of Information Science and Technology, Chukyo University
101 Tokodachi, Kaizu-cho, Toyota, Aichi, 470-0393 Japan
{itoh@sist,h10806m@st,hiroyasu@sist}.chukyo-u.ac.jp

Abstract. This paper describes a mechanism to annotate face images in impressive words which express their visual impressions. An annotation mechanism is developed by integrating latent semantic indexing, decision trees, and association rules. Moreover, visual and symbolic features of faces are integrated, which are corresponding to lengths and/or widths of face parts and impressive words, respectively. Relationships among these features are represented in a latent semantic space, their direct relationships in decision trees, and co-occurrence relationships among symbolic features in association rules, respectively. Efficiency of annotation results is improved by integrating these mechanisms, since their features are utilized effectively.

Keywords: face image annotation, impression, latent semantic indexing, latent semantic space, decision tree, association rule.

1 Introduction

In recent, several types of face image processing systems are developed, such as face recognition systems, multimodal interfaces, etc. In especial, to develop a face image database is required according to the progress of such processing systems. Then, face images are retrieved not only in terms of visual features of face images, but also in terms of words which represent visual impressions of face images. Therefore, it is necessary that suitable words are assigned to face images, i.e., face images are annotated in impressive words.

A face image annotation system is being developed, named FIARS (Face Image Annotation and Retrieval System)[5]. Its annotation mechanism is realized by integrating three mechanisms based on latent semantic spaces constructed by latent semantic indexing[6], association rules and decision trees[8]. So far, each mechanism is developed independently, and these mechanisms were not yet integrated[5]. The latent semantic space consists of descriptions of face images and words. Co-occurrence relationships among words are specified in association rules. These rules are useful for inferring additional words. Decision trees specify requirements on visual features to assign a specific word. By integrating these mechanisms systematically, annotation results are able to be improved since these mechanisms work complementarily.

J.D. Velásquez et al. (Eds.): KES 2009, Part II, LNAI 5712, pp. 591–598, 2009.

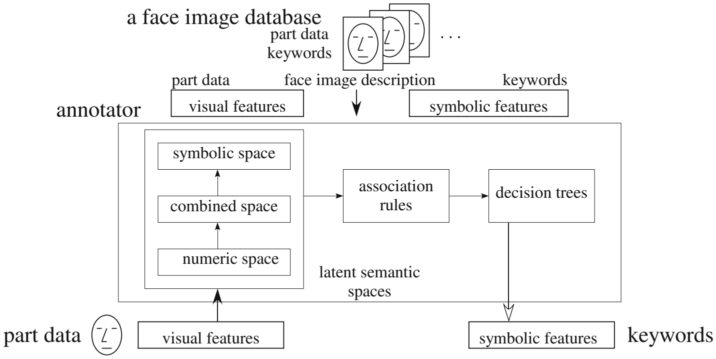


Fig. 1. An overview of an annotation mechanism and components of FIARS

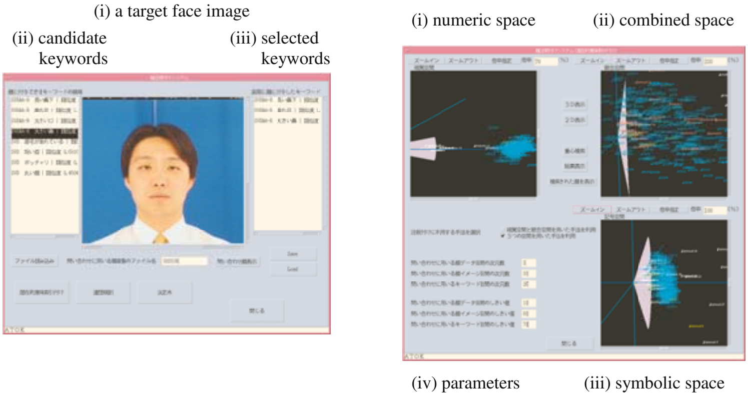
In recent, many mechanisms for retrieving and for annotating face images are developed[1]. In order to retrieve face images, to annotate them in keywords is required, like usual natural images[2]. Person identification is to identify the name of a person. Emotions indicated in faces are analyzed based on the facial action coding system[7,3]. [3] utilizes latent semantic analysis. Moreover, a face retrieval mechanism using mental words is developed[4].

This paper is organized as follows. Section 2 shows an overview of the system. Section 3 presents an annotation mechanism. Experimental results are shown in Sec. 4. Finally, concluding remarks are described in Sec. 5.

2 An Overview of FIARS

Face images are described in impressive words which depict visual impressions inspired from faces and their face parts. In current, a sort of impressive words is restricted. They express sizes, lengths or shapes of face parts, e.g., a round face, a thin lip, etc. On the other hand, lengths and/or widths of face parts are measured, they are called part data. Annotation is to assign and to find impressive words as keywords based on part data.

Figure 1 shows an overview of FIARS. This system consists of three mechanisms for annotation, and a face image database. These annotation mechanisms are developed based on latent semantic spaces, association rules and decision trees. They are constructed from the face image database. The latent semantic spaces consist of three spaces which are a numeric space, a combined space and a symbolic space. The numeric space is constructed from part data only, the combined space from part data and keywords, and the symbolic space from keywords, respectively. Moreover, association rules specify co-occurrence relationships among keywords. When association rules are applied to a set of existing keywords, some additional keywords are obtained. Furthermore, decision trees are constructed from both part data and keywords. Decision trees represent conditions on part data, which specify whether a keyword is able to be assigned to a face image, or not. On the other hand, the previous system shown in [5] utilizes



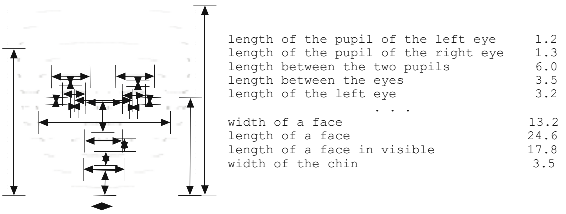
(a) An example of a window for annotating a face. (b) An example of windows for showing latent semantic spaces.

Fig. 2. Examples of screens for annotating a face image and for showing spaces

only a numeric space and a combined space. The symbolic space is constructed for making clear interrelationship among keywords, and for trying to improve efficiency of retrieved keywords. Moreover, latent semantic spaces, association rules and decision trees function independently each other in [5]. An entire annotation process has to be controlled by an individual user suitably. However, in this system, the system controls the procedure for applying each mechanism. They are working cooperatively, see Sec. 3.

This system provides some windows as an interface for assisting in annotation. Figure 2 (a) shows the main window of this system. (i) shows the face image to be annotated. (ii) and (iii) show candidate keywords inferred by the system, and selected keywords by a user, respectively. On the other hand, (i), (ii) and (iii) in Fig. 2(b) show a numeric space, a combined space and a symbolic space, respectively. (iv) shows some parameters for achieving keyword assignment.

Latent semantic spaces, association rules and decision trees are constructed from a face image database. Figure 3 (a) shows 24 measured places as part data. (b) shows an example of part data. Face image I_d is represented in terms of a vector, which consists of two vectors $(\boldsymbol{v}_d; \boldsymbol{w}_d)$. They are called a part vector and a keyword vector, respectively. A part vector is $\boldsymbol{v}_d = (v_{d,1}, \ldots, v_{d,24})^T$. For constructing a combined latent semantic space, a normalized part vector is computed from a part vector, $\boldsymbol{v}'_d = (v'_{d,1}, \ldots, v'_{d,24})^T$, where $v'_{d,j}$ is a normalized value of $v_{d,j}$. This value is computed as $v'_{d,j} = (v_{d,j} - \mu_j)/\sigma_j + 1/2$, where μ_j and σ_j are the mean value and the standard derivation of face part j. On the other hand, a keyword vector is $\boldsymbol{w}_d = (w_{d,1}, \ldots, w_{d,43})^T$. Each element $w_{d,j}$ is 1 or 0. They represent whether keyword j is assigned to face image I_d, or not, respectively. Furthermore, when a decision tree is constructed, discretization is applied to part data. Part datum $v_{d,j}$ is normalized as $v''_{d,j} = (v_{d,j} - \mu_j)/\sigma_j$. This value is transformed into one symbolic value, a, b or c. They are interpreted as small/short, normal and large/long, respectively. Let α be a threshold. If $v''_{d,j} < -\alpha$ then the value is a, $-\alpha \leq v''_{d,j} \leq +\alpha$ then b, and $+\alpha < v''_{d,j}$ then c.



```
length of the pupil of the left eye      1.2
length of the pupil of the right eye     1.3
length between the two pupils            6.0
length between the eyes                  3.5
length of the left eye                   3.2
                    . . .
width of a face                         13.2
length of a face                        24.6
length of a face in visible             17.8
width of the chin                        3.5
```

(a) part data. (b) part data, lengths/widths of 24 places of a face.

Fig. 3. Part data and examples

If this threshold is 0.38, both a and c are assigned to about 35% of face images, respectively. The rest are assigned b.

3 An Annotation Mechanism

Figure 4 shows an overview of an annotation procedure. At first, a target face image to be annotated is given, which is specified in its part data.

Next, some keywords are obtained using three latent semantic spaces. To achieve this, the dimensions of spaces and the thresholds (see θ_1, θ_2 and θ_3 in Fig. 4) for query processing are specified. A part vector of a given face image is treated as a query vector(f_t). Some similar face images to a target are obtained using the numeric space. The vectors of the similar face images are found in the combined space. The centroid vector of them(f_c) is computed, which is used as a query vector for seeking keywords in the combined space. After query processing, some keywords are retrieved. The keyword vectors of obtained keywords are found in the symbolic space, the centroid vector of them(k_c) is computed, like the above described procedure. As this result, a retrieval result is obtained, which is a set of pairs of keyword k_i and its weight w_i, $K = \{< k_1, w_1 >, \dots, < k_n, w_n >\}$. The weight is similarity weight between the keywords and the given face image. This similarity weight is computed by a cosine similarity measure.

To expand K, association rules are applied to it. An association rule is represented in $A_i : t_{i,1}, \dots, t_{i,m} \rightarrow t_{i,m+1}, \cdots, t_{i,m+l}, (Sup_i, Con_i)$, where $t_{i,j}$, Sup_i and Con_i are keywords, support and confidence. Confidence is interpreted as a certainty factor. Rule A_i is applied to K, when all keywords appeared in the left-hand side of A_i are members of K. If so, words $t_{i,m+1}, \cdots, t_{i,m+l}$ are added to K. The weights of added keywords are $max\{w_{i,1}, \dots, w_{i,m}\} * Con_i$, where $w_{i,1}, \dots, w_{i,m}$ are weights of $t_{i,1}, \dots, t_{i,m}$.

Finally, decision trees are applied to individual members of K. Rule D_i obtained based on a decision tree is represented as $D_i : t_i \leftarrow p_{i,1}, \cdots, p_{i,m}, ER_i$, where t_i, $p_{i,j}$ and ER_i are words, a condition element related to a face part and an error ratio of the rule. Condition-part is constructed based on decision trees using a specified error ratio[8]. After this, an actual error ratio of each obtained rule, ER_i is computed. Then, $1 - ER_i$ seems its certain factor CF_i. For k_i in K, it is checked whether a given face image satisfies the condition-part of rule D_i

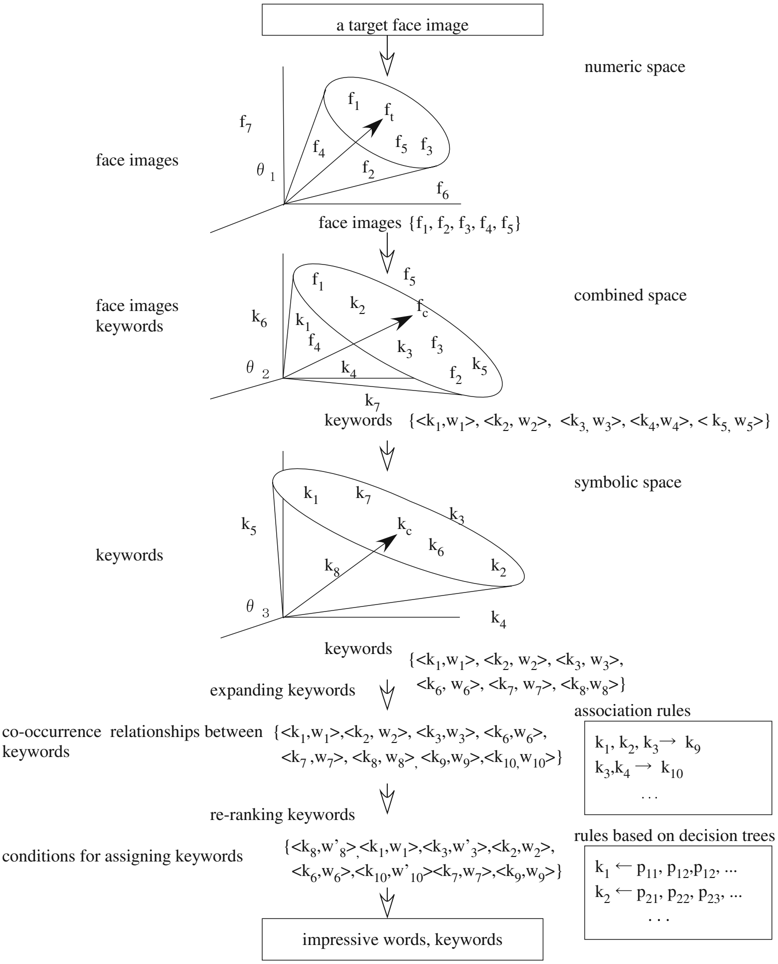

a target face image
numeric space
face images
f_7
f_1
f_t
f_4
f_5 f_3
θ_1
f_2
f_6
face images $\{f_1, f_2, f_3, f_4, f_5\}$
f_5
combined space
face images
keywords
f_1
k_2
f_c
k_6
k_1
f_4
k_3
f_3
k_5
θ_2
k_4
f_2
k_7
keywords $\{<k_1,w_1>, <k_2, w_2>, <k_3, w_3>, <k_4,w_4>, < k_5, w_5>\}$
symbolic space
k_1 k_7
k_3
k_5
k_c
k_6
keywords
k_8
k_2
θ_3
k_4
keywords $\{<k_1,w_1>, <k_2, w_2>, <k_3, w_3>,$
$<k_6, w_6>, <k_7, w_7>, <k_8,w_8>\}$
expanding keywords
association rules
co-occurrence relationships between keywords
$\{<k_1,w_1>,<k_2, w_2>, <k_3,w_3>, <k_6,w_6>,$
$<k_7 ,w_7>, <k_8, w_8>, <k_9,w_9>,<k_{10},w_{10}>\}$
$k_1, k_2, k_3 \rightarrow k_9$
$k_3,k_4 \rightarrow k_{10}$
...
re-ranking keywords
rules based on decision trees
conditions for assigning keywords
$\{<k_8,w'_8>,<k_1,w_1>,<k_3,w'_3>,<k_2,w_2>,$
$<k_6,w_6>,<k_{10},w'_{10}><k_7,w_7>,<k_9,w_9>\}$
$k_1 \leftarrow p_{11}, p_{12},p_{12}, \dots$
$k_2 \leftarrow p_{21}, p_{22}, p_{23}, \dots$
...
impressive words, keywords

Fig. 4. A conceptual overview of a procedure for assigning keywords

for k_i, or not. When its conditions are satisfied with the given face image, the weight of k_i is replaced to CF_i if its weight w_i is lower than CF_i.

Now, an example of the process to assign keywords is described. Let a face image be given as a target, shown in Fig. 2(a)(i). Also, some parameters are specified, see Sec. 4. At first, some keywords are captured by using three latent semantic spaces. Four keywords are obtained as follows; (1) 'dropping eyebrow', (2) 'short eyebrow', (3) 'chubby face' and (4) 'round face', in the descending order. Although weights of these keywords are computed, they are not shown here for simplicity. Next, by applying the association rules to the obtained keywords, other four keywords are derived. They are (5) 'long length between the nose and the upper lip', (6) 'large nose', (7) 'dropping eye' and (8) 'large mouse' as an ordered list. Many of these keywords seem suitable to the given face image. These two kinds of keywords are arranged by their weights, the ordered keywords are

(5), (6), (1), (7), (2), (8), (3) and (4). Continuously, decision rules are applied to each keyword for confirming individual keywords. Finally, the ordered list is obtained, which is {(5), (7), (8), (6), (1), (2), (3), (4)}. This list is presented to a user as candidate keywords. The keywords obtained using association rules are located higher in this case. Moreover, it is considered that five keywords among retrieved eight keywords are suitable, which are (5), (7), (6), (3) and (4). Figure 2 shows two windows used in this process.

4 Experimental Results

Five types of experiments are tried as follows:

(1) LSS. Three latent semantic spaces are used, only.
(2) LSS, association rules. After keywords are obtained using (1), association rules are applies to them.
(3) decision tree, association rule. After keywords are obtained by applying decision trees, association rules are applied to them.
(4) decision tree. Keywords are obtained by applying decision trees, only.
(5) top 9. After keywords are retrieved by (2), decision rules are applied. The nine keywords in a higher rank among the retrieved keywords are tested, i.e., keywords are obtained by the proposed method in Sec. 3. Moreover, one existing face image description defined in the face image database has 8.7 keywords on average. So, nine keywords are evaluated.

Some parameters are required for building latent semantic spaces, association rules and decision trees, and for achieving keyword retrieval, as shown in Table 1. The dimensions and the threshold for the numeric space are 3 and 10°, respectively. A cumulative contribution ratio is used for deciding dimensions of a combined space and a symbolic space. When the cumulative contribution ratio is over 0.8, the number of cumulated singular values is treated as the dimensions of these spaces. On the other hand, the minimum support and the minimum confidence are specified for association rules. Moreover, the threshold used in discretization and the error ratio are specified for decision trees.

In each experiment, 30 face images are given as targets. Retrieval keywords are tested in sense of precision and recall, which are defined as *precision = the number of retrieved correct keywords / the number of retrieved keywords* and

Table 1. Some parameters for setting FAIRS

latent semantic space	dimension	threshold(a degree)
numeric space	3	10
combined space	33	80
symbolic space	26	70
association rule	minimum support	minimum confidence
	0.1	0.4
decision tree	threshold for discretization	error ratio
	0.38	0.2

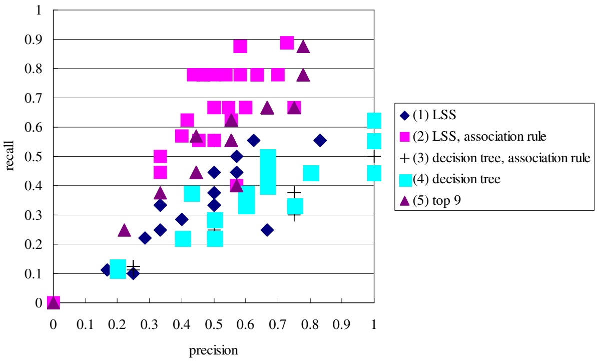


Fig. 5. Precision and recall of retrieval keywords

Table 2. Mean values of precisions and recalls of retrieval keywords

	precision	recall
(1) LSS	0.52	0.36
(2) LSS, association rule	0.53	0.68
(3) decision tree, association rule	0.63	0.35
(4) decision tree	0.69	0.31
(5) top 9	0.57	0.57

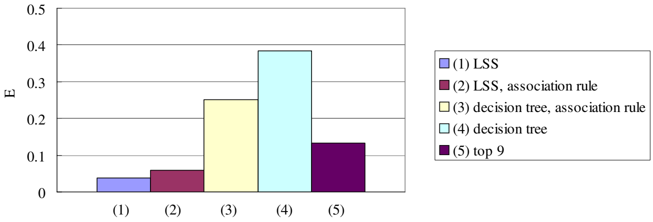


Fig. 6. Accuracy of retrieval keywords

recall = *the number of retrieved correct keywords / the number of keywords assigned to a target in advance*. Precisions and recalls in each experiment are shown in Fig. 5, and their mean values are summarized in Table 2. Recall in (2) is better than one in (1) by applying association rules. Precisions in (3) and (4) are better than ones in (1) and (2). The result using the decision trees is more precise than others. The result using (2) is obtained using the latent semantic spaces and association rules, however, decision trees are not applied to the retrieved keywords since their order is not evaluated in this case.

To evaluate ranked retrieval keywords, accuracy is computed. Accuracy is defined as $E =$ *(the number of correct keywords − the number of wrong keywords) / the number of retrieved keywords)*$(-1 \leq E \leq 1)$[9]. The result in each experiment is shown in Fig. 6. Recall in (2) is better than others, so, recalls in (3) and (4) are worse than one in (2), as shown in Table 2. However, accuracies in (3) and (4) are better than ones in (1) and (2). As shown in this table and Fig. 6, balance between precision and recall in (5) is better than others. Therefore, the proposed method is more effective than the method which utilizes only either latent semantic spaces or decision trees.

5 Concluding Remarks

This paper describes a mechanism for annotating face images in impressive words, which is developed by integrating latent semantic spaces, association rules and decision trees. Precision and recall of final retrieval keywords are more effective than ones using a single method. On the other hand, to develop some mechanisms is planed for improving capability of the system. It is required that some parameters are (semi-)automatically determined. Moreover, to develop a mechanism for retrieving appropriate face images using impressive words is necessary. Furthermore, the interface makes easy to understand assignment process.

Acknowledgement

The face images are used by permission of Softopia Japan, Research and Development Division, HOIP Laboratory.

References

1. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and Machine Recognition of Faces: A Survey. Proceedings of the IEEE 83(5) (1995)
2. Datta, R., Joshi, D., Li, A., Wang, J.Z.: Image Retrieval: Ideas, Influence, and Trends of the New Age. ACM Computing Survey 40(2) (2008)
3. Fasel, B., Monay, F., Gatia-Perez, D.: Latent Semantic Analysis of Facial Action Code for Automatic Facial Expression Recognition. In: Proc. MIR. ACM, New York (2004)
4. Fang, Y., Geman, D., Boujemaa, N.: An Interactive System for Mental Face Retrieval. In: Proc. MIR. ACM, New York (2005)
5. Ito, H., Kawai, Y., Koshimizu, H.: Face Image Annotation based on Latent Semantic Space and Rules. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part II. LNCS, vol. 5178, pp. 766–773. Springer, Heidelberg (2008)
6. Manning, C.P., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)
7. Pantic, M., Rothkrantz, L.J.M.: Facial Action Recognition for Facial Expression Analysis from Static Face Images. IEEE Tran. on SMC. Part B 34(3) (2004)
8. Tan, P.-N., Steinback, M., Kumar, V.: Introduction to Data Mining. Addison-Wesley, Reading (2005)
9. Wang, X.-J., Zhang, L., Ma, W.-Y.: Annotating Images by Mining Image Search Results. IEEE Tran. on PAMI 30(11), 1919–1932 (2008)
10. Zhang, Z., Zhang, R.: Multimedia Data Mining, A Systematic Introduction to Concepts and Theory. CRC Press, Boca Raton (2009)