Internal Use Only
非公開

TR－H－019

Context-Driven Retrieval and Saliency of Facial Features

Takashi KATO　Masaomi ODA

1993. 8. 9

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

Context-Driven Retrieval and Saliency of Facial Features[1]

Takashi Kato and Masaomi Oda

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan
Email: tkato or oda@hip.atr.co.jp

Abstract — An experimental evaluation of a facial image retrieval system showed that our previously proposed "context-driven retrieval mechanism" facilitated retrieval (or externalization) of ambiguous as well as better defined facial images. Some features, such as face shape, eyebrow tilt, and eye shape, were found to be more salient than others. The context-driven retrieval mechanism, while maintaining the relative importance of facial features, facilitated retrieval by reducing the variance in the less salient features.

I. INTRODUCTION

Recognizing a known face as such constitutes the first step in face-to-face communication, whether it takes place in the office or on the street. Recalling the facial image of a particular person may be needed when that person is the subject of discussion. At times it may be necessary to communicate to others a facial image that does not exist in the real world but only in one's mind. "My favorite face" may be one such example.

How can we facilitate retrieval of mental images which are more or less ambiguous? How can we facilitate the process of defining and subsequent externalization of internally-held mental images? We previously proposed a "context-driven retrieval mechanism" [1] for database systems to support retrieval (or externalization) of ambiguous, mental images, as well as of concrete, target images.

We subsequently evaluated a facial image retrieval system that used an initial version of the context-driven retrieval mechanism. The main purpose of this article is to discuss the effectiveness of this context-driven retrieval mechanism in light of the saliency of facial features shown in the experimental results.

1 We are grateful to Yoh'ichi Tohkura and Shigeru Akamatsu for their support and encouragement throughout the project reported here.

II. FACIAL IMAGE RETRIEVAL SYSTEM

The facial image retrieval system described below was designed to support efficient retrieval of image data, whether the target image is concrete and clear in the user's mind prior to retrieval or is yet to be clarified and better defined during retrieval. In the latter case, the system's objective is to help retrieve as close image data as possible to the mental image in the user's mind. While most standard database systems assume that the user has a clear image of the target prior to a retrieval attempt and/or that the target is definable by verbal labels (i.e., key words), our facial image retrieval system does not make such assumptions.

Our experimental system was developed for initial investigation of the previously proposed context-driven retrieval mechanism. Line-drawn schematic faces were used as image data to simplify database construction and manipulation of the experimental variables.

The system contains nearly 60,000 line-drawn faces, each of which is drawn by choosing particular values for 10 facial features (see Figure 1 for examples of line-drawn schematic faces). The facial features used were: eyebrow position (BP), eyebrow tilt (BT), eye position (EP), eye shape (ES), face shape (FS), mouth position (MP), mouth length (ML), nose position (NP), nose length (NL), and ear position (RP). There are three possible values for each feature, thus producing 59,049 (i.e., 3^{10}) possible faces.

The experimental system presents 10 line-drawn faces at a time on a computer display, from which the user selects the face or faces that she or he thinks satisfy the retrieval goal. The user is allowed to select as many faces as he or she wishes, so long as they meet his or her selection criteria. The system can keep up to 10 of

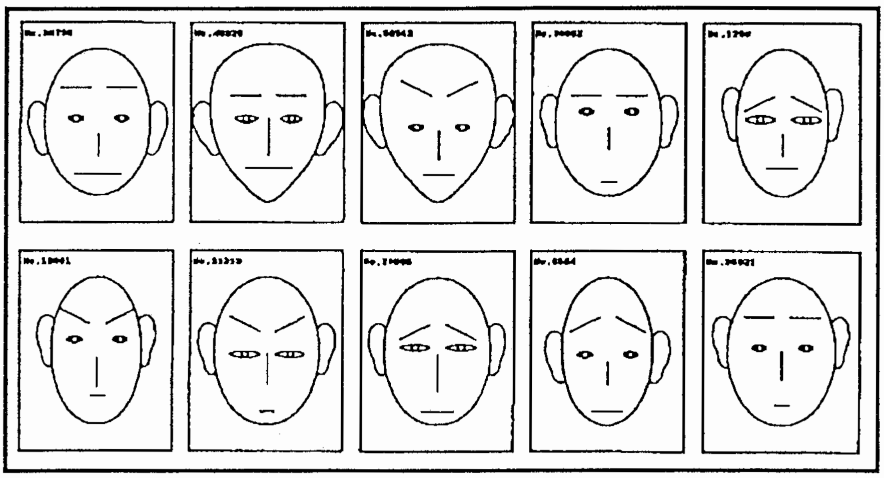

Fig. 1: Examples of Line-Drawn Faces

the most recently selected faces in a background buffer and the user can at anytime call them up and re-evaluate them as to whether to retain or discard any of them.

With the context-driven (CD) retrieval mechanism, the system presents on a selection screen (1) the seven faces that are computed to be most similar to the user-selected faces in the buffer, (2) a randomly chosen user-selected face, and (3) two faces randomly chosen from the database. By selecting the seven candidates based on the computed similarity to the user-selected faces, the system attempts to help narrow the search space. One of the user-selected faces is randomly chosen and presented on the selection screen to provide the user with a continuing opportunity to re-evaluate the previously selected faces. The two randomly chosen faces are presented so that the user, if he or she so wishes, can change the direction of the search-space pruning.

III. EXPERIMENTAL METHOD

A. Subjects

Forty students were recruited from Doshisha University and paid to participate in the experiment. There were 13 male and 27 female subjects.

B. Apparatus and Material

The facial image retrieval system described above was used. It contained nearly 60,000 line-drawn faces. Each face was a different combination of the three possible values of the 10 facial features.

C. Design and Procedure

There were two different retrieval tasks. In the favorite-face retrieval task, one group of subjects (6 male and 14 female subjects) was asked to retrieve 10 of their favorite faces. In the similar-face retrieval task, another group of subjects (7 male and 13 female subjects) was asked to retrieve the 10 faces most similar to a particular target face. In each case, the subjects completed the retrieval task once with the assistance of the context-driven retrieval mechanism and once without it. Half of the subjects attempted context-driven (CD) retrieval first and the other half random (RN) retrieval first with the context-driven mechanism turned off. In all cases, the subjects were asked to continue the retrieval task until they were satisfied with the 10 faces collected in the buffer. The retrieval tasks (i.e., favorite- vs. similar-face retrieval) were thus manipulated between-subjects and the retrieval conditions (i.e., with or without the context-driven retrieval mechanism) within-subjects.

The subjects were randomly assigned to one of four experimental conditions: 2 retrieval tasks x 2 orders of the retrieval conditions. After receiving retrieval task instructions and an explanation of the facial image retrieval system, the subjects were given an opportunity to familiarize themselves with the system and mouse operation. The practice session lasted for five minutes. Following the practice period, the subjects in the similar-face retrieval task were asked to spend five more minutes to select their favorite face. A printed copy of that face was then given to them as their retrieval task target. At the end of each retrieval session, the subjects were asked to identify the most satisfactory face from among the 10 selected faces and to orally explain why they judged it to be the best among the favorite faces or the faces similar to the target.

IV. EXPERIMENTAL RESULTS

Several measures were taken during the experimental evaluation of the facial image retrieval system.

A. Total Retrieval Time and Trials

The mean time the subjects took to complete the retrieval of favorite faces was 376 seconds for context-driven (CD) retrieval and 452 seconds for random (RN) retrieval. To complete the retrieval of similar faces, the subjects took on average 353 seconds for CD retrieval and 553 seconds for RN retrieval.

The mean number of the faces examined by the subjects in favorite-face retrieval was 190 for CD retrieval and 479 for RN retrieval. In similar-face retrieval, the numbers were 186 and 792 for CD and RN retrieval, respectively.

B. Variance of Parameter Values

For each of the 10 facial features, we calculated the variance of the parameter values among the 10 faces in the final set produced by each subject in each retrieval task.

A 2 (CD/RN retrieval conditions) x 2 (favorite-/similar-face retrieval tasks) two-way analysis of variance (ANOVA), applied to the variance data collapsed across the 10 facial features, showed that both the main effects and interaction were significant: $F(1, 38)=769.26$, $p<0.0001$, for CD/RN retrieval conditions, $F(1, 38)=14.46$, $p<0.0005$, for favorite-/similar-face retrieval tasks, and $F(1, 38)=7.91$, $p<0.01$, for the interaction between these two factors.

A 2 (CD/RN retrieval conditions) x 10 (facial features) two-way ANOVA, applied to the variance data in the favorite-face retrieval task (see Figure 2), showed that both the main effects of CD/RN retrieval conditions ($F(1, 361)=388.37$, $p<0.0001$) and facial features ($F(9, 361)=11.75$, $p<0.0001$) were significant. There was a significant interaction effect between CD/RN retrieval conditions and facial features: $F(9, 361)=3.73$, $p<0.001$.

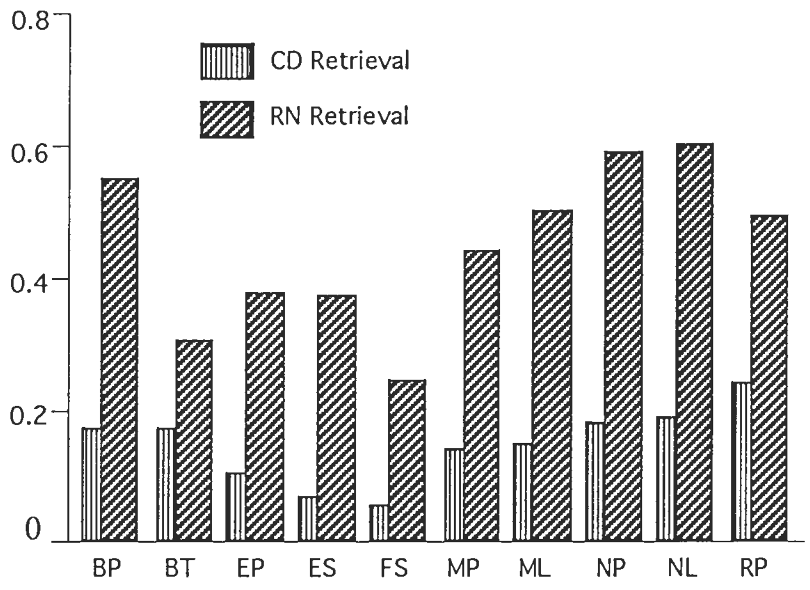

Fig. 2: Variance in favorite face retrieval as a function of retrieval conditions

Scheffe's tests (α=0.05) revealed that the variance of a given facial feature was significantly lower in the CD than the RN retrieval condition for all facial features except eyebrow-tilt.

A Tukey pairwise comparison (α=0.05) of the 10 facial features in the CD retrieval condition showed that the variance of the face-shape feature was significantly smaller than that of the ear-position and nose-length features and that the variance of the eye-shape and eye-position features was significantly smaller than that of the ear-position feature only.

A Tukey pairwise comparison (α=0.05) in the RN retrieval condition showed that the variance of the face-shape feature was significantly smaller than that of the nose-length, nose-position, eyebrow-position, mouth-length, ear-position and mouth-position features and that the variance of the eyebrow-tilt feature was significantly smaller than that of these features except the mouth-position feature. We also found the variance of the eye-shape feature was significantly smaller than that of the nose-length, nose-position, and eyebrow-position features while the variance of the eye-position feature was significantly smaller than that of the nose-length and nose-position features.

A similar ANOVA of the variance data in the similar-face retrieval task (see Figure 3) showed that the main effects of CD and RN retrieval conditions and facial

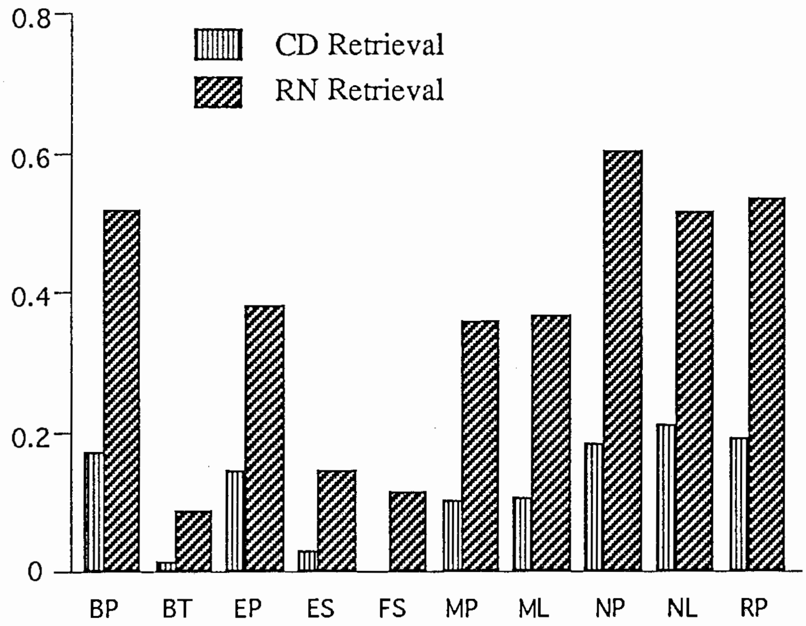

Fig. 3: Variance in similar face retrieval as a function of retrieval conditions

features were significant: $F(1, 361)=281.18$, $p<0.0001$, for CD/RN retrieval conditions, and $F(9, 361)=32.28$, $p<0.0001$, for facial features. The interaction between CD/RN retrieval conditions and facial features was also found to be significant: $F(9, 361)=6.00$, $p<0.0001$.

As with the case of the favorite-face retrieval task, Scheffe's tests ($\alpha=0.05$) indicated that in the similar-face retrieval task the reduction in the variance by using the CD retrieval mechanism was significant for all facial features except eyebrow-tilt.

A Tukey pairwise comparison ($\alpha=0.05$) of the facial features in the CD retrieval condition showed that the face-shape, eyebrow-tilt, and eye-shape features had significantly smaller variance than did the remaining features, except the mouth-position and mouth-length features.

For the variance data in the RN condition, A Tukey pairwise comparison ($\alpha=0.05$) indicated that the eyebrow-tilt, face-shape, and eye-shape features had significantly smaller variance than the remaining seven features, and that the mouth-position, mouth-length, and eye-position features had significantly less variance than the nose-position feature.

C. Hit Rates of Parameter Values

The goal in the similar-face retrieval task was to collect 10 faces that were most similar to a particular target face. We calculated for each of the 10 facial features the hit rates between the target face and the 10 faces in the final set (see Figure 4).

A 2 (CD/RN retrieval conditions) x 10 (facial features) two-way ANOVA showed that the main effect of facial features was significant ($F(9, 361)=18.07$, $p<0.0001$), but neither the main effect of CD/RN retrieval conditions ($F(1, 361)=1.46$, $p>0.2$), nor the interaction effect was significant ($F<1$).

A Tukey pairwise comparison ($\alpha=0.05$) indicated that the eyebrow-tilt feature had a significantly higher hit rate than did the other features, except the face-shape feature. The face-shape feature had a significantly higher hit rate than the remaining features, except the eye-shape and mouth-position features. The eye-shape feature showed a significantly higher hit rate than did the eyebrow-position, ear-position, nose-position, and nose-length features. The hit rate of the mouth-position feature was higher than that of the nose-length feature.

In the favorite-face retrieval task, there was no externally-given target face. However, at the conclusion of the retrieval task, the subjects were asked to specify the face they judged the best among the 10 favorite faces stored in the final set. We

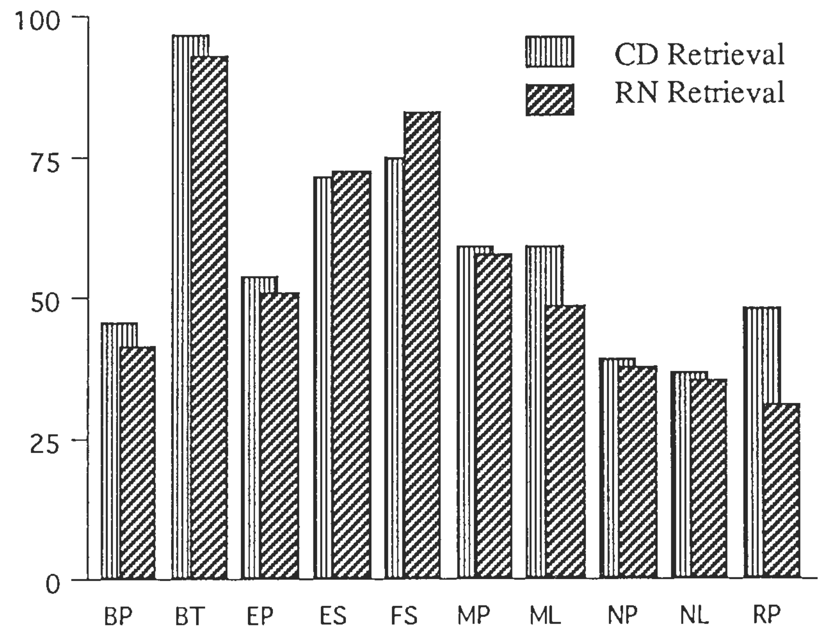

Fig. 4: Hits in similar face retrieval as a function of retrieval conditions

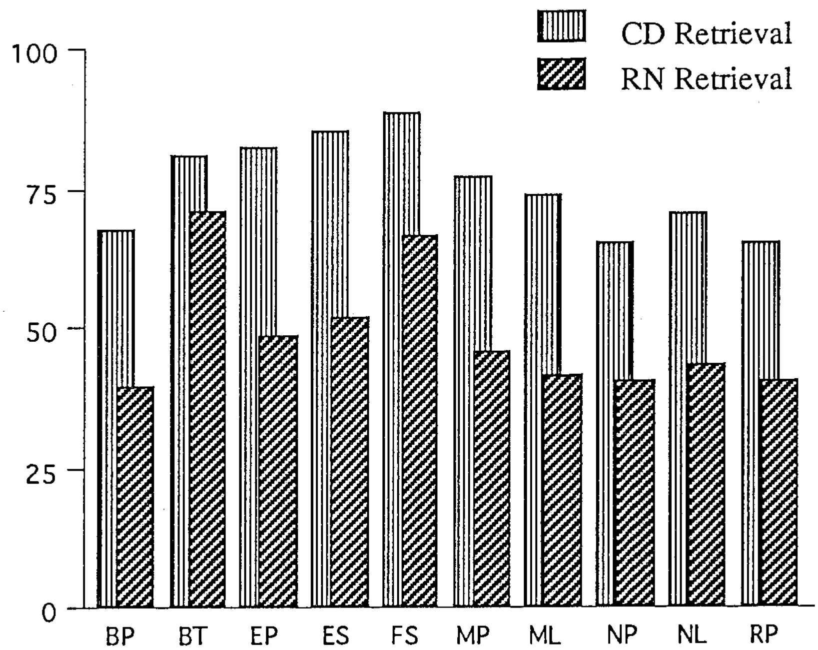

Fig. 5: Hits in familiar face retrieval as a function of retrieval conditions

calculated for each of the 10 facial features the frequency of the feature value shared between the best face and the remaining faces (see Figure 5).

A 2 (CD/RN retrieval conditions) x 10 (facial features) two-way ANOVA showed that the main effects of CD/RN retrieval conditions and facial features were significant: $F(1, 361)=152.88$, $p<0.0001$, for CD/RN retrieval conditions, and $F(9, 361)=7.36$, $p<0.0001$, for facial features. The interaction between these two factors was not significant: $F(9, 361)=1.08$, $p>0.3$.

A Tukey pairwise comparison ($\alpha=0.05$) indicated that the face-shape and eyebrow-tilt features had a significantly higher hit rate than did the mouth-length, nose-length, eyebrow-position, nose-position, and ear-position features. The hit rate of the face-shape feature was higher than that of the mouth-position feature as well. Also, the eye-shape feature had a significantly higher hit rate than did the nose- and ear-position features.

D. Verbal Reports

As mentioned in the Method section, at the conclusion of each retrieval task, the subjects were asked to identify the best face and give the reason for their choice.

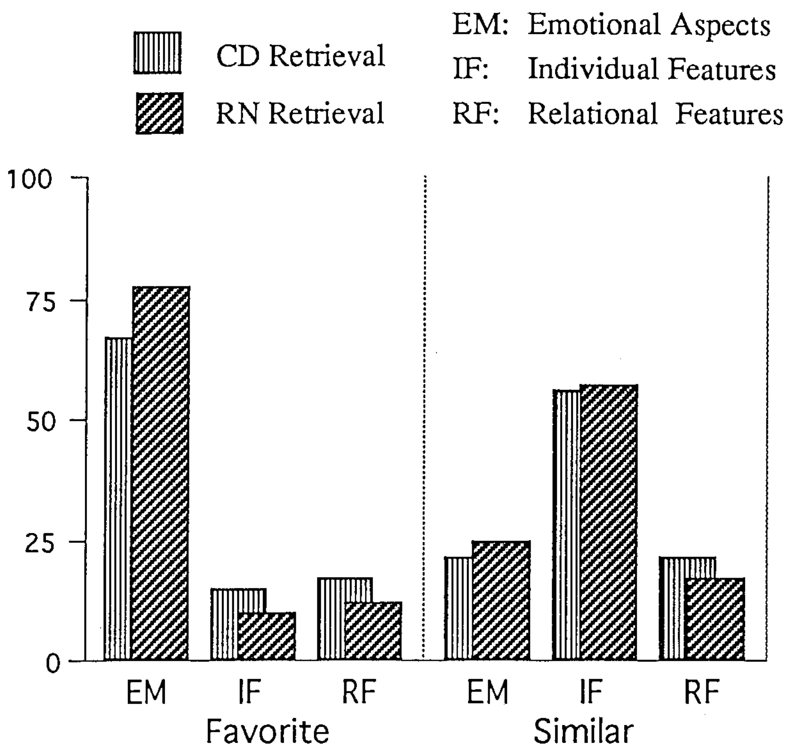

Fig. 6: Reports of selection reasons as a function of retrieval tasks

The reasons were classified into three categories depending on whether they referred to physical characteristics of individual features (e.g., big eyes), relations between features (e.g., space between eyebrows and eyes), or emotional aspects of the face (e.g., gentle). The frequency of occurrence in each category was then counted. If more than one category of reasons was given to the same face by the same subject, the frequency was divided by the number of categories. The percentage for each category as a function of retrieval tasks are given in Figure 6.

The pattern of the results was similar between the CD and RN retrieval conditions and indicated that the subjects tended to refer more to emotional aspects in the favorite-face retrieval task and more to individual features in the similar-face retrieval task.

E. Explicit and Implicit Features

The facial features so far discussed refer to the characteristics of individual features, which we might call "explicit features." There are also features of a rather implicit nature, such as spatial relations between individual features. We

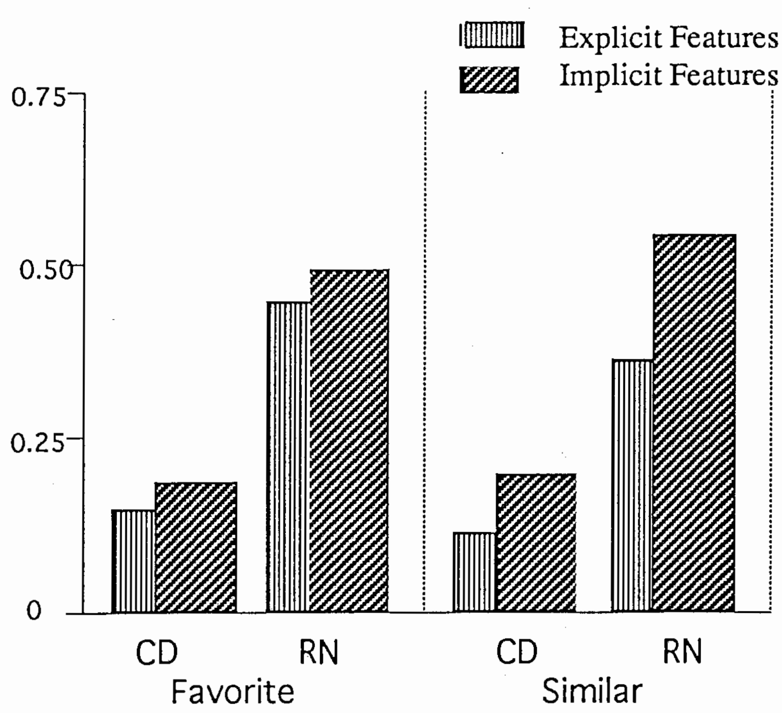

Fig. 7: Variance in explicit and implicit features as a function of retrival condition and tasks

defined six of such "implicit features"; the positional relations between eyebrow and eye, eyebrow and nose, eyebrow and mouth, eye and nose, eye and mouth, and nose and mouth. We then classified the parameter values of these implicit features into three categories so as to make their value scale compatible with that of the explicit features. Figure 7 shows the mean total variance as a function of retrieval conditions and tasks.

Since a 2 (explicit/implicit features) x 2 (CD/RN retrieval conditions) x 2 (favorite-/similar-face retrieval tasks) three-way ANOVA indicated that the three-way interaction was significant ($F(1, 38) = 7.36$, $p<0.01$), separate analyses were made for favorite- and similar-face retrieval.

A 2 (explicit/implicit features) x 2 (CD/RN retrieval conditions) two-way ANOVA, applied to the data in the favorite-face retrieval task, showed that both the main effects were significant: $F(1, 57)=7.65$, $p<0.01$, for explicit/implicit features, and $F(1, 57)=428.12$, $p<0.0001$, for CD/RN retrieval conditions. The interaction between these two factors was not significant: $F<1$.

A 2 (explicit/implicit features) x 2 (CD/RN retrieval conditions) two-way ANOVA, applied to the data in the similar-face retrieval task, showed that both the main effects were significant: $F(1, 57)=56.60$, $p<0.0001$, for explicit/implicit features, and $F(1, 57)=290.27$, $p<0.0001$, for CD/RN retrieval conditions. There was also a significant interaction effect between the explicit/implicit features and the CD/RN retrieval conditions: $F(1, 57)=8.12$, $p<0.01$.

V. DISCUSSION

We will first assess the effectiveness of the context-driven retrieval mechanism for the present retrieval tasks (favorite- and similar-face retrieval) and then discuss the relative importance of the facial features with respect to the differences in the retrieval conditions (i.e., CD vs. RN) and tasks (i.e., favorite- vs. similar-face retrieval).

A. Overall Assessment of the CD Retrieval Mechanism

Except for the mean time to retrieve favorite faces, context-driven (CD) retrieval showed substantial advantages over random (RN) retrieval. The pattern of the results suggests that in RN retrieval, the subjects encountered many faces that they were able to quickly discount as irrelevant to their goal. Caution should be exercised in interpreting these results beyond this general pattern, as the total retrieval time included the time the subjects spent examining the selected faces in the buffer screen, and the amount of viewing was at their discretion.

As expected, the CD retrieval mechanism helped reduce the total variance among the final set of 10 faces, whether the retrieval task was for favorite faces or similar faces. This reduction of the total variance seemed to result more from the reduced variance among the less salient features than that among the salient features, such as face- and eye-shape and eyebrow-tilt.

The effectiveness of the CD retrieval mechanism depends not only on reducing the variance in the parameter values, but on the end result matching the retrieval goal. The hit-rate data for favorite-face retrieval showed that the CD retrieval mechanism effectively increased the hits (i.e., frequencies of sharing the identical parameter values) between the best face and the remaining faces in the final set. In the similar-face retrieval task, however, the CD retrieval mechanism did not significantly increase the hit rates of the feature parameter values between the target face and the 10 faces selected by the subjects, although the observed difference was in a direction favorable for the CD retrieval mechanism.

B. Feature Saliency and the CD Retrieval Mechanism

It seems reasonable to assume that subjects would select those faces with matching values for the more important features than for the less important ones. This implies that the difference in variance between facial features can be an index of the difference in their relative importance in retrieving facial images [2].

There is now converging evidence that the face-shape, eyebrow-tilt, and eye-shape features were important in the retrieval of line-drawn faces used in the present experimental system. It seems that the CD retrieval mechanism reduced the variance of the less prominent features more than that of the more prominent features. It also appears that the differences in variance among the facial features were less distinctive in favorite-face retrieval than in similar-face retrieval and that this reduction in the saliency was further augmented by the CD retrieval mechanism. Together with the hit-rate data for favorite-face retrieval showing the advantage of the CD retrieval mechanism, these results suggest that the CD retrieval mechanism was effective in increasing the overall quality of the 10 faces in the final set.

C. Explicit and Implicit Features

The variance in the implicit features was larger than that in the explicit features for both favorite- and similar-face retrieval. This suggests that the differences in the spatial relations between features were less obvious and/or less important to the subjects than those in the individual features, especially when the task was to collect those faces similar to the target. The interaction between the CD/RN retrieval conditions and the explicit/implicit features in the similar-face retrieval task suggests that the CD retrieval mechanism was more helpful in reducing the variance in the implicit features than that in the explicit features. Put another way, it seems that while the subjects somehow managed the variance in the explicit features, they were less successful in controlling the variance in the implicit features when the physical similarity was a primary criterion (i.e., similar-face retrieval). It might be that the subjects tended to (or even had to) pay more attention to the explicit similarities between the target and presented faces at the expense of the implicit similarities, but that the CD retrieval mechanism helped reduce the extent of such tradeoffs.

D. Improving the CD Retrieval Mechanism

Taken together, it seems safe to say that the CD retrieval mechanism was successful in supporting the retrieval of facial images. It helped reduce the variance in the spatial relations between the features (i.e., implicit features) as well as that in the individual features (i.e., explicit features), while improving or at least maintaining the same degree of correspondence between the mental image and the outcome of the retrieval (i.e., hits).

The present CD retrieval mechanism was intentionally kept simple to see if the basic idea of context-driven retrieval is workable. A closer and qualitative examination of the experimental results indicated that the CD retrieval mechanism might have come into play too soon and/or too strongly in directing the search space in the database. We therefore re-designed the CD retrieval mechanism to avoid the possible problem of "jumping the gun" by applying what we learned from the experimental evaluation reported here. A new algorithm was designed to exploit not only the global contextual information, such as those feature characteristics shared by the selected exemplars and the number of the selections made, but also the local contextual information, such as the user's actions on recent selection occasions (e.g., whether any selection was made from the latest screens). We are preparing an evaluation study of this new version of the CD retrieval mechanism in which we plan to employ a variety of retrieval tasks, such as "retrieve gentle or distasteful faces," to determine the general applicability of the CD retrieval mechanism and to further our understanding of feature saliency in face retrieval.

REFERENCES

[1] Oda, M.: Context Dependency Effect in the Formation of Image Concepts and its Application. IEEE SMC '91, 1673-1678, 1991.

[2] Kato, T & Oda, M.: Indirect Measurement of Feature Saliency in Face Processing. Paper presented at the International Conference on Face Processing, Cardiff, Wales, UK, 1993.