# Visualization of Crowd-Powered Impression Evaluation Results

Erika GOMI*, Yuri SAITO*, Takayuki ITOH*
*(*)Graduate School of Humanities and Sciences, Ochanomizu University*
*Tokyo, Japan*
{*erika53, yuri, itot* } *@itolab.is.ocha.ac.jp*

***Abstract*—There have been many collective knowledge on the Web, such as evaluation of restaurants, hotels, and manufactured products. Even though each of the participants on such Web sites usually just evaluate the small number of contents, these kinds of crowd-powered contents evaluation services bring us fruitful information. Visualization is a useful tool to carefully observe the evaluation results and discover complex trends of the evaluation. This paper presents our study on visualization of the crowd-powered contents evaluation. Firstly we developed a contents evaluation technique applying an interactive genetic algorithm, which presents contents estimated to be highly or poorly evaluated. Then we had a case study with various appearances of female face images to collect the evaluations. Finally, we visualized the result by applying an image browser CAT. This paper discusses how the visualization result depicts the trends of the evaluation on appearance of women.**

***Keywords*-Interactive genetic algorithm, crowd-powered contents evaluation, image browser.**

## I. Introduction

Collective knowledge based on the evaluation by ordinary consumers is recent important information on the Web. For example, we can look at the numeric evaluation of restaurants or hotels on various Web sites. Also, it is helpful information if catalogs of products such as automobiles show scores of the products evaluated by the consumers. Generally we cannot suppose that the customers input evaluations for all the products or other contents. On the other hand, we can construct fruitful crowd-powered knowledge even if each of the customers input evaluations for only small number of products or other contents.

This paper presents a crowed-powered contents evaluation and collective knowledge visualization technique based on the above supposition. Here, it is often important to collect highly or poorly evaluated answers in many impression analysis problems which construct collective knowledge. The presented contents evaluation technique iteratively shows contents to users and requests them to evaluate the contents, where it preferentially shows the contents which the technique predicts the user will highly or poorly evaluate. The technique applies interactive genetic algorithm (iGA) to explore the search space to specify the contents which the user will highly or poorly evaluate. Our study efficiently collects the interesting knowledge with small number of users' input by optimizing the contents search to find users' preferable or dislike contents.

The collective knowledge visualization technique applies an image browsing technique [1]. The technique displays a set of highly or poorly evaluated contents as a collection of images so that users can overview the set of the contents and quickly understand the trends of the evaluation results.

The paper shows our case study on impression analysis of appearance of women. Recent application software for makeup and hair style synthesis realizes us to easily simulate appearance of women. However, it is not always easy to self-evaluate what kinds of makeup or hair styles look fine for themselves. We therefore supposed that evaluation and visualization of appearance of women should contribute to encourage the decision of their own appearance change. This paper defines "appearance of women" as combinations of the types of eye, brow, nose, and frame of the faces, choice of makeup, and types of hair styles. We created over 1500 faces by the image synthesis of the appearance of women, and asked participants to input evaluations of the faces suggested by our evaluation technique. This paper introduces the visualization of the collective evaluation result.

## II. Related work

Crowd-powered analysis is a recent hot research topic to construct collective knowledge. Especially, this concept is useful to obtain subjective evaluation for various contents, such as Web design [2], 2D images [3], and 3D shapes [4]. Koyama et al. presented a generalized technique [5] to explore numeric parameter spaces with inputs of large number of participants. This paper also aims a general technique for optimal contents exploration; however, we do not suppose numeric parameters are given to the contents.

Interactive genetic algorithm has been used by various applications such as image retrieval and music recommendation [6] [7] [8]. The technique presented in this paper is different from these studies since it explores highly and poorly evaluated contents simultaneously, and specializes to the crowd-powered evaluation systems.

CAT [1] is a hierarchical image browser which we apply for visualization of the collective knowledge. Given a set of hierarchically organized images, CAT places them onto a display space by recursively applying a rectangle packing algorithm. It can represent the set of images and their hierarchical structure simultaneously; on the other hand, images would be displayed very small if large number of images are given. CAT features a zooming interface to solve the problem; it displays representative images of the clusters while zooming out, or individual images in the clusters while zooming in. Thanks to the interface, CAT can control the sizes and number of images to be displayed along users' interaction.

## III. Crowd-powered Contents Evaluation Technique

This section presents our contents evaluation technique. Given a large number of contents, the technique shows smaller number of contents to users and requests to input their evaluations. We can construct the crowd-powered knowledge by collecting the evaluations of many users by using this technique.

Here, we suppose that high or poor evaluation is especially important information in many cases of impression analysis. It should be efficient if we can collect high or poor evaluation of contents from small number of users, or small number of answers for each user. We developed a contents evaluation technique which preferentially shows the contents which the technique predicts the user will highly or poorly evaluate, applying interactive genetic algorithm (iGA). iGA is useful for this study, because it adopts an objective function based on users' input with their preferences or impressions, while it inherits the genetic operations of original genetic algorithm. Here, ordinary iGA just explores the highly adapted solutions, while our technique requires an algorithm which simultaneously explores highly and poorly adapted solutions. Therefore, our implementation of iGA applies the island model [9] to divide the individuals into two islands and separately explore highly or poorly scored contents respectively.

Following is the processing flow of the presented crowd-powered contents evaluation technique (see Figure 1).

Step 1: Initialize Population
: Randomly select the constant number (12 in case of our implementation) of contents as initial individuals.

Step 2: Display
: Show the contents to users.

Step 3: Evaluation
: Request the users to input the subjective evaluation for the contents. Our implementation provides three button widgets corresponding to "Good", "Soso", and "Bad", and requests the users to press one of them.

Step 4: Selection and Immigration
: Collect individuals evaluated as "Good" to the island of "Good". Similarly, collect individuals evaluated as "Bad" to the island of "Bad". Dismiss other individuals evaluated as "Soso".

Step 5: Crossover
: Generate new generation of the individuals in the two islands respectively.

Step 6: Mutation
: Randomly apply the mutation for the diversity of individuals.

Step 7: Termination
: Stop the iteration if it satisfies pre-defined conditions. (Our current implementation just terminates if the sequential number of the current generation exceeds the pre-defined number.)

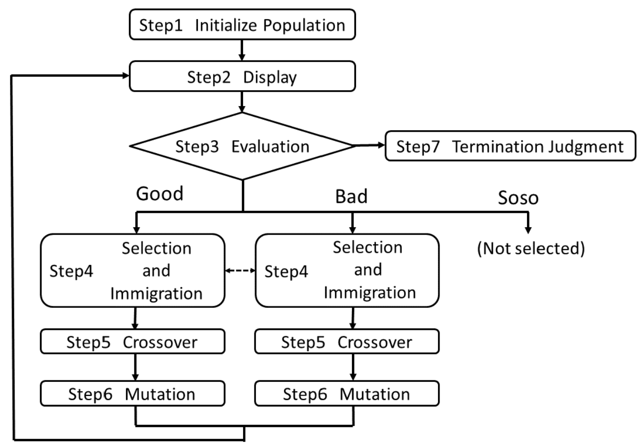


Figure 1. Processing flow of the presented crowd-powered contents evaluation technique.

## IV. Case Study: Appearance of Women

This section introduces our implementation and case study on evaluation of appearance of women.

### A. Face image preparation

As the preparation of the case study, we prepared the set of images of women's faces by the following process. We firstly took face pictures of 18 twenties women, and generated intermediate images by applying a morphing technique. As a result, we generated 16 types of intermediate face images as the combination of the following features.

- Length of the face: "long" or "short".
- Form around the chin: "thin" or "round".
- Impression of eyes: "bright" or "thin".
- Impression of nose: "thin" or "round".

Then, we applied a makeup simulation service [1] and a hair style simulation service [2] to generate more variety of

[1] SHISEIDO Beauty check point makeup (URL: https://www.shiseido.co.jp/sw/check/makeup/)

[2] Hairstyle Simulator "ChouChou" (URL: https://itunes.apple.com/jp/app/chouchou-heasutairu-shimyureta/id573854005?mt=8)

face images. We applied the combination of the following features for the face image synthesis.

- Makeup type: “fresh”, “cute”, “cool”, or “elegant”.
- Length of hair: “long”, “medium”, or “short”.
- Bangs: “with” or “without”.
- Form of hair: “straight” or “waved”.
- Color of hair: “brown” or “black”.

We generated 1536 face images as a result. Figure 2 shows examples of the synthesized face images.

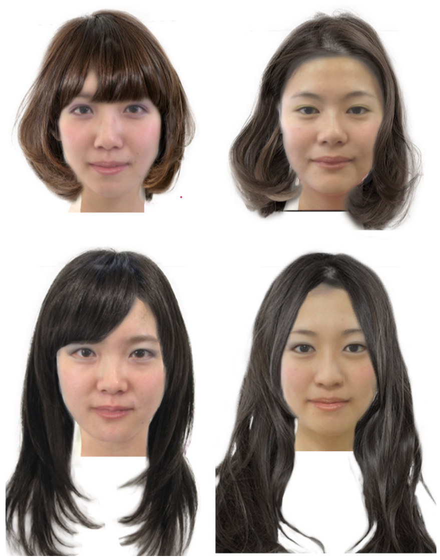

Figure 2. Example of synthesized face images.

### B. User interface for appearance evaluation

We developed a user interface featuring the iGA-based contents evaluation algorithm for the evaluation of appearance of women. Figure 3 shows a window capture of the user interface. We implemented this software with JDK (Java Development Kit) 1.6.0. The window features a display space to show two face images, and button widget to start the input, and input the evaluation “Good”, “Soso”, or “Bad”. Once we developed another user interface which displays just one face image and had a user experience. Many participants commented that it would be much easier if the window showed multiple face images and therefore they could comparatively evaluate the faces. We revised the user interface design reflecting such comments and had another user experience. Consequently we had better comments to the revised user interface from the participants.

Figure 3. User interface for the evaluation of appearance of women.

Our implementation displays face images pair-by-pair when a set of individual is generated by iGA. It supposes that a user presses a button as the evaluation of the left face first, then presses a button again as the evaluation of the right face. It switches the display to the next pair of faces at that time. The implementation proceeds iGA to generate the next set of individuals when the user finishes the input for all the faces corresponding to the current generation of individuals.

### C. User experience

We had a user experience with 30 twenties female participants using our implementation of the evaluation technique for appearance of women. The following is the setting of the iGA in our experience.

- Total number of face images: 1536
- Number of individuals in a generation: 12
- Crossover ratio: 1.0
- Mutation ratio:
  (if $n_{Soso} < 4$) : 0.05
  (if $n_{soso} \geq 4$) : $0.05(n_{Soso} - 2)$
  where $n_{Soso}$ is the number of images which a user evaluated as “Soso” in the previous generation.
- Termination condition: 20 generations

Average number of face images showed to participants was 169, which means that each of the participants evaluated just approximately 10% of the contents. As a result, one face image was evaluated as “Good” by 7 participants, and four face images were evaluated as “Good” by 6 participants. Three of these face images had the same hair style. On the other hand, five face images were evaluated as “Bad” by 6 participants; however, we could not find any commonality among these five face images.

Tables I and II shows the ratio of face images evaluated as “Good” or “Bad”. These results suggest that no strong correlation between the types of face parts and types of makeup is observed. On the other hand, we found a strong factor in the hair style that face images with bangs got totally higher evaluations rather than face images without bangs.

During the user experiences, 14 participants had at least one generation that they evaluated for all face images in the generation as “Good” or “Bad”. Also, we found that numbers of face images evaluated as “Soso” by 8 participants significantly decreased along the evolution. These results suggest that our technique successfully selected the face images which are estimated the users would evaluate as “Good” or “Bad”.

Table I
RATIO OF FACE IMAGES EVALUATED AS “GOOD”.

| !! | Fresh | Cute | Cool | Elegant |
|---|---|---|---|---|
| Long face | 18.2% | 31.8% | 25.6% | 24.4% |
| Short face | 20.2% | 25.3% | 28.8% | 25.7% |
| Thin chin | 17.6% | 30.8% | 29.3% | 22.3% |
| Round chin | 22.0% | 23.4% | 25.6% | 29.0% |
| Bright eyes | 19.7% | 31.0% | 26.3% | 22.9% |
| Thin eyes | 19.3% | 23.6% | 29.2% | 27.9% |
| Round nose | 18.3% | 24.3% | 30.4% | 27.1% |
| Thin nose | 21.1% | 31.5% | 24.4% | 23.0% |

Table II
RATIO OF FACE IMAGES EVALUATED AS “BAD”.

| !! | Fresh | Cute | Cool | Elegant |
|---|---|---|---|---|
| Long face | 21.9% | 23.0% | 25.7% | 29.4% |
| Short face | 21.0% | 24.0% | 25.6% | 29.4% |
| Thin chin | 23.3% | 26.4% | 22.2% | 28.0% |
| Round chin | 19.8% | 20.8% | 28.8% | 30.6% |
| Bright eyes | 22.3% | 25.8% | 20.5% | 31.3% |
| Thin eyes | 20.8% | 21.5% | 30.1% | 27.7% |
| Round nose | 22.7% | 21.2% | 28.1% | 28.0% |
| Thin nose | 20.0% | 26.4% | 22.5% | 31.2% |

## V. VISUALIZATION OF COLLECTIVE KNOWLEDGE

This section introduces our case study on visualization of the contents evaluation results of the appearance of women. We conducted the evaluation of the appearance of women using the images synthesizing the features described in the previous section. This section divides the features into the following:

- invariant features (eyes, noses, and outlines of the faces), and
- variant features (makeups and hair styles).

Actually, we synthesized the face images by firstly composing the invariant features and then applying makeup and hair style simulations. We therefore concluded that users would be interested in what types of combinations of invariant and variant features got better impressions. We designed the data structure of the contents evaluation results for the visualization based on the above discussion.

### A. Selection of the feature combinations for visualization

We counted the number of participants which answered “Good” or “Bad” for each of the face images. We then calculated the total numbers and ratios of the answers for arbitrary combinations of invariant and variant features, and ranked the combinations based on the ratios. Finally we selected the combinations those ratios significantly differed from the ratios of other combinations which are similar but partially different. Tables III and IV shows the set of selected feature combinations those ratios of the answers of “Good” or “Bad” were especially high.

Table III
ESPECIALLY HIGHLY EVALUATED FEATURE COMBINATIONS.

| | Good |
|---|---|
| Long face | Cute make, Long hair |
| Short face | Cool make, Medium hair |
| Thin chin | Long hair, Brown hair |
| Round chin | Elegant make, Medium hair |
| Bright eyes | Cute make, Long hair |
| Thin eyes | Medium hair |
| Round nose | Medium hair, Black hair |
| Thin nose | Cute make, Long hair, Brown hair |

Table IV
ESPECIALLY POORLY EVALUATED FEATURE COMBINATIONS.

| | Bad |
|---|---|
| Long face | Elegant make |
| Short face | none |
| Thin chin | none |
| Round chin | Elegant make |
| Bright eyes | Elegant make |
| Thin eyes | Cool make, Medium hair |
| Round nose | none |
| Thin nose | Elegant make, Short hair |

### B. Hierarchical data construction

We generated a hierarchical dataset from the above feature combination selection result. We divided the set of feature combinations according to the invariant features, and then divided to two groups, “Good” and “Bad”. In addition, we listed the variant features in each of the groups, and collected the corresponding face images for each of the variant features. Finally, we selected the representative face image for each of the variant features, according to the ratios of “Good” or “Bad” answers. Figure 4 shows the constructed structure.

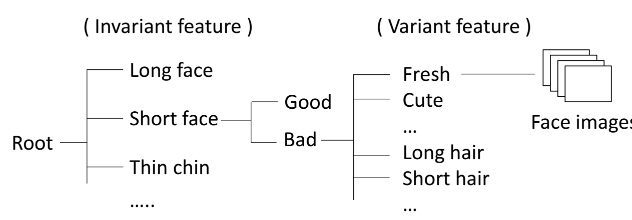


Figure 4. Hierarchical data structure.

### C. Visualization

We visualized the above mentioned hierarchical structure by applying an image browser CAT [1]. This visualization represents the hierarchy by nested rectangular borders. The most outer rectangles in the visualization results shown in Figures 5 to 8 depict invariant features. They enclose the painted two rectangles, where blue ones depict groups of “Good” face images while red ones depict groups of “Bad” face images. These painted rectangles enclose inner rectangular borders corresponding to variant features. Representative face images are displayed in the inner rectangular border.

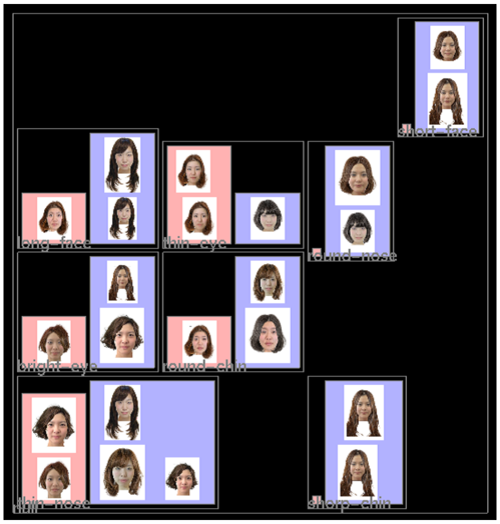

Figure 5. Example (1). CAT displays representative images while zooming out.

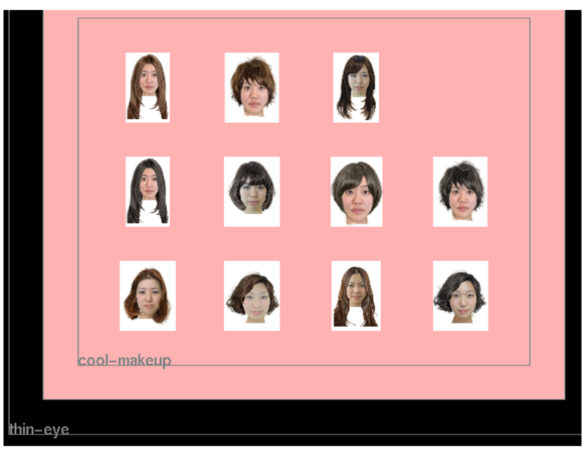


Figure 6. Example (2). CAT displays individual images corresponding to particular variant features while zooming in.

Figure 5 shows an example of the visualization while zooming out so that we can overview the contents evaluation result. We can look at the set of extremely highly or poorly evaluated representative face images at first. Zooming into the particular portion of the display, CAT switches to display all individual face images corresponding to particular variant features, as shown in Figure 6. We can interactively select the interested groups of face images and display all of them by this operation. CAT also features an operation for magnification of particular images so that we can carefully observe them, as shown in Figure 7.

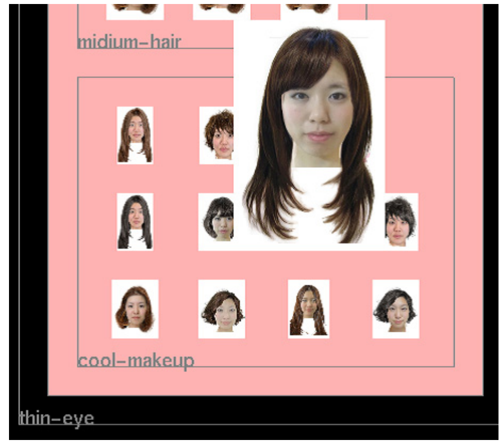


Figure 7. Example (3). CAT magnifies user-selected images.

### D. Discussion

We carefully observed the face images in the visualization result, and discovered many trends of the evaluation result. This section introduces several of the trends.

We found that many face images with bangs are shown in the blue rectangles, while many face images without bangs are shown in the red rectangles. It suggests that hair styles with bangs were relatively preferable for the participants of our case study. On the other hand, there are also a small number of face images without bangs in blue rectangles, or face images with bangs in red rectangles. It suggests that combination of invariant and variant features is also an important factor for the evaluation.

Next, we focused on all the hair styles. There are 24 types of hair styles as combination of length, form, color, and existence of bangs, in our case study. We found that the completely same types of hair styles were applied in multiple representative face images. It suggests that iGA evolved to show similar types of face images to many participants. Also, we found that the same face images are displayed in the multiple groups even though each group corresponds to different combinations of features. Figure 8 shows an example depicting the above fact.

## VI. Conclusion

This paper presented our case study on visualization of contents evaluation results. We developed a crowd-powered evaluation technique applying an interactive genetic algorithm, which presents contents estimated to be highly or poorly evaluated. We had a case study with various appearances of female face images to collect the evaluations, and visualized the result by applying an image browser CAT. This paper discussed what kinds of trends could be observed with the visualization result.

Following are our potential future issues. First, we would like to improve the aggregate calculation of the participants’ answers. In this case study we simply counted the number of participants answered as “Good” or “Bad” for each face


Figure 8. Example (4). Same image is selected as representative images of different groups.

image; however, we do not think the ranking brought from this process is sufficiently reliable. We would like to apply other techniques to rank the face images. Second, we would like to test with other types of visualization techniques or image browsers to more carefully observe the evaluation results. Finally, we would like to apply this framework to various applications, not limited to appearances of women. We may also need to customize the implementation of interactive genetic algorithm while applying our techniques to various applications.

## ACKNOWLEDGMENT

This work has been partially supported by Japan Society of the Promotion of Science under Grant-in-Aid fir Scientific Research.

## REFERENCES

[1] A. Gomi, R. Miyazaki, T. Itoh, J. Li, CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique, International Conference on Information Visualisation (IV08), 82-87, 2008.

[2] K. Reinecke, K. Z. Gajos, Quantifying Visual Preferences around the World, ACM SIGCHI Conference on Human Factors in Computing Systems, 2014.

[3] Y. Gingold, A. Shamir, D. Cohen-Or, Micro Perceptual Human Computation for Visual Tasks. ACM Transactions on Graphics, 31(5), 119:1-119:12, 2012.

[4] A. Secord, J. Lu, A. Finkelstein, M. Singh, A. Nealen, Perceptual Models of Viewpoint Preference, ACM Transactions on Graphics, 30(5), 109:1-109:12, 2011.

[5] Y. Koyama, D. Sakamoto, T. Igarashi, Crowd-powered Parameter Analysis for Visual Design Exploration, ACM Symposium on User Interface Software and Technology, 65-74, 2014.

[6] S. Cho, Towards Creative Evolutionary Systems with Interactive Genetic Algorithm, Applied Intelligence, 16(2), 129-138, 2002.

[7] Y. Saito, T. Itoh, MusiCube: A Visual Music Recommendation System featuring Interactive Evolutionary Computing, Visual Information Communication and Interaction Symposium, 2011.

[8] H. Takagi, S. Cho, T. Noda, Evaluation of an IGA-based Image Retrieval System Using Wavelet Coefficients, IEEE International Conference on Fuzzy Systems, 1775-1780, 1999.

[9] D. Whitley, S. Rana, R. B. Heckendorn, The Island Model Genetic Algorithm : On Separability, Population Size and Convergence, Journal of Computing and Information Technology, 7(1), 33-47, 1999.