Impression Estimation for Deformed Portraits with a Landmark-based Ranking Network

Mari MIYATA Kiyoharu AIZAWA
The University of Tokyo

Background

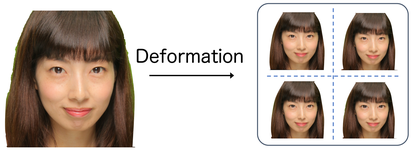
☐ Portraits play an important role in communication in a social networking service.

Original Filtering Digital makeup Deformation of Facial features

☐ Problem: difficult to manipulate portraits as intended without sufficient skills or experience

■ Our Goal:

A portrait manipulation method based on impression words



Which "deformation" makes this portrait more { CLEAR, SWEET, ELEGANT, MODERN, DYNAMIC } ?

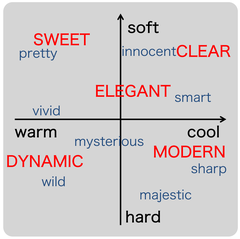
Data Collection

■ 492 Portraits: Japanese Females

☐ Restrict gender, nationality, age + facial expressions, head position

Example of portraits

■ 5 Impression Words

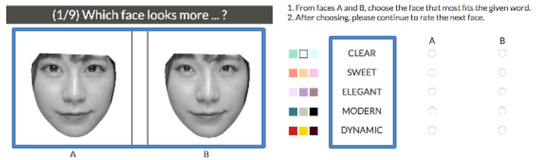


☐ Selected from Kobayashi's keywords[12]

☐ Conditions
- In different categories
- No overlap in meanings

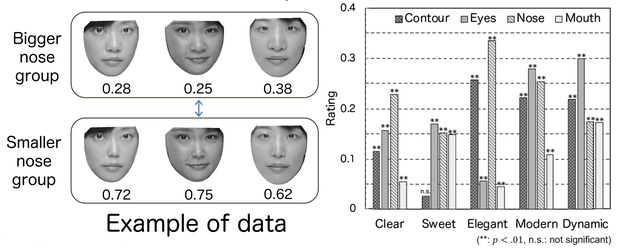
■ Impressions Associated with Deformed Portraits

☐ Crowdsourcing Pairwise Comparison
- Collect the evaluation data of portraits



Paired deformed portraits Impression words
- same subject
- scaling same facial feature

☐ Data Validation: paired t-test

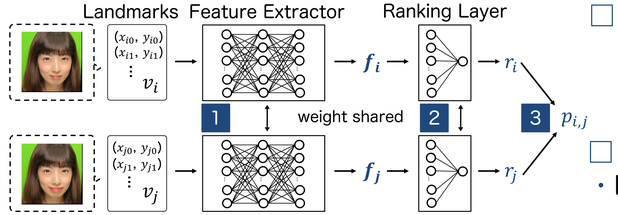


Example of data

- The deformation of facial features causes an impression difference in a portrait.

Impression Estimation for Deformed Portraits

Key Idea: Estimate how well fits a portrait to an impression word based on the relative relationship between two portraits



Landmark-based Ranking Network

1. Extract feature f_i, f_j from Landmarks.
2. Calculate the ranks r_i, r_j.
3. Calculate the relationship $p_{i,j}$ between ranks r_i, r_j.

☐ **Loss function**: solve as classification

$$L_{i,j} = -t_{i,j}\log(p_{i,j}) - (1-t_{i,j})\log(1-p_{i,j})$$

*$t_{i,j}$: target relative attribute label

☐ **Training data**
- Landmarks: 1870 pairs of deformed portraits
- Label: $t_{i,j} = \begin{cases} 1 & : \text{above 60\% of participants selected } I_i \\ 0 & : \text{above 60\% of participants selected } I_j \\ 0.5 & : \text{none of the above} \end{cases}$

☐ **During inference**
- Use a branch of the network to calculate Rank

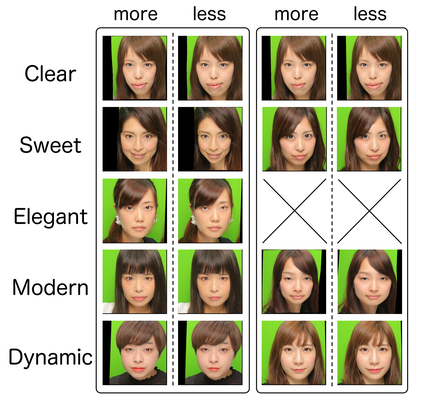
*Rank: value between -1.3 and +1.3 in our method.

■ Experiments

Percentage of correctly ordered pairs

	Method	Input	Clear	Sweet	Elegant	Modern	Dynamic	Mean
Rank SVM based	Relative	portrait	51.7	55.6	53.7	51.1	52.1	52.7
		landmark	53.0	55.8	53.7	51.9	51.9	53.3
	Fine-grained	portrait	54.5	55.6	53.9	52.0	56.7	54.5
		landmark	53.3	55.7	53.6	51.9	52.0	53.3
NN based	Deep Relative	portrait	71.4	67.3	72.8	79.1	73.2	72.8
		landmark	-	-	-	-	-	-
	Ours	portrait	-	-	-	-	-	-
		landmark	73.3	67.7	78.0	83.0	76.0	75.6

- Using landmarks as input works well.
- The neural-network-based methods perform better.
- Our method can rank the portraits according to the subject's facial features.



Comparison with Deep Relative[17]

Conclusion

■ Data collection and Impression estimation using the ranking method

■ Future work: Implementation of Generative network based on impressions