Impression Estimation Model for Clothing Patterns Using Neural Style Features

Natsuki Sunda[1], Kensuke Tobitani[1], Iori Tani[1], Yusuke Tani[1], Noriko Nagata[1(✉)], and Nobufumi Morita[2]

[1] Kwansei Gakuin University, 2-1, Gakuen, Sanda, Hyogo, Japan
nagata@kwansei.ac.jp
[2] Couture Digital Ltd, Sankyo Sakaisujihonmachi Building F9 1-8-2, Bakurou-machi, Chuo-ku, Osaka, Osaka, Japan

Abstract. In product design, the impressions evoked by surface properties of objects attract attention. These impressions are called affective texture and demand is growing for technologies that quantify, index, and model it. Even in the fashion industry, the diversification of user needs necessitates the customization and personalization of products. Consequently, there has been a focus on custom-made services. However, enormous amounts of time and human costs are needed to find clothing patterns to suit one's own preferences and ideas from among the countless patterns available. This research focused on affective texture related to visual impressions, and we proposed a method for automatically estimating the affective texture evoked by clothing patterns. To this end, we conducted the following steps: (1) quantified the visual impressions for patterns; (2) extracted style features as physical characteristics; and (3) modeled the relationships between visual impressions and physical characteristics. Afterward, based on the obtained models, we estimated the impressions for unlabeled patterns. Then, we verified their validity through relative evaluation and absolute evaluation, and we confirmed that our models estimated the impressions corresponded to the impressions that people actually felt. In addition, we implemented a system to enable users to intuitively search for patterns.

Keywords: Fashion · Texture · CNN · Lasso regression

1 Introduction

In product design, not only designs, such as colors and shapes, but also the impressions (e.g., "luxurious" or "antique") evoked by surface properties of objects attract attention. These impressions are called affective texture, which is an important factor in preference for or the quality judgment of an object. Therefore, demand is growing for technologies that quantify, index, and model affective texture.

Even in the fashion industry, the diversification of user needs necessitates the customization and personalization of products. Consequently, there has been a

C. Stephanidis and M. Antona (Eds.): HCII 2020, CCIS 1226, pp. 689–697, 2020.
https://doi.org/10.1007/978-3-030-50732-9_88

focus on services that allow customers to order custom-made clothing on the Internet. However, enormous amounts of time and human costs are needed to design unique clothing. For example, it is difficult to find clothing patterns to suit one's own preferences and ideas from among the countless patterns available.

This research focuses on affective texture related to visual impressions, and we propose a method for automatically estimating the affective texture evoked by clothing patterns. To this end, we model the relationships between visual impressions and physical characteristics. In addition, based on the obtained models, we implement a system to enable users to intuitively search for patterns.

2 Previous Research

Research on texture analysis has been conducted for a long time, and various texture features have been proposed [1,2]. In recent years, Gatys et al. [3] has been proposed an algorithm for transferring an image style using VGG-19 [4], which is a convolutional neural network used for object recognition. In their research, it is suggested that content features are necessary for object recognition and that style features represent the detailed appearance in images, such as multi-scale color information and the pattern information, and are not as important for object recognition.

Thereby, Takemoto et al. thought that style features were strongly related to affective texture based on unique texture properties (uniformity, density, coarseness, roughness, regularity, linearity, directionality, direction, frequency, and phase [5]). Thus, they proposed a texture synthesis method with the desired affective texture using style features and showed high-precision results [6].

In this research, we think that clothing patterns are also part of texture. Therefore, based on Takemoto et al., we use style features as physical characteristics representing patterns and clarify the relationships with the visual impressions they evoke.

3 Proposed Method

In this research, we propose a method for automatically estimating the affective texture evoked by clothing patterns. Figure 1 shows an overview of our approach. First, we conducted a subjective evaluation experiment and a factor analysis, and quantified the visual impressions of a small number of patterns. Next, we extracted style features as physical characteristics using a VGG-19 network. Afterward, we conducted a lasso regression and constructed impression estimation models. Finally, using our models, we estimated the impressions for a large number of unlabeled patterns.

4 Quantification of Visual Impressions

4.1 Subjective Evaluation Experiment

We conducted a subjective evaluation experiment in order to quantify the visual impressions evoked by patterns.

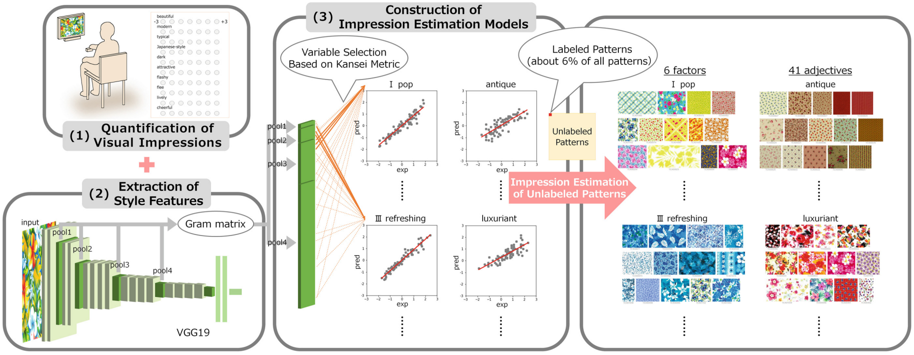

Fig. 1. Overview of our proposed method

First, it was necessary to gather and select comprehensive and representative words for evaluating visual impressions. Based on Tobitani et al. [7], we conducted a free writing experiment using 10 pattern images. Then, we verified whether the obtained words were suitable for visual impressions. As a result, we selected 28 adjectives with high fitness. We also added 12 adjectives that previous research used for clothing materials and the feel of materials [8–10] and "cute" from the viewpoint of actual use. Finally, we used 41 adjectives as the evaluation words in this experiment.

Next, we specifically focused on flower patterns and collected 1,158 such images. Then, in order to select comprehensive and representative stimuli, we conducted clustering based on style features extracted from the images. Based on the divided clusters, we selected 75 images as the stimuli. The extraction of style features is described in Sect. 5.1.

A total of 40 graduate and undergraduate students (18 men; 22 women; average age: 21.9 ± 2.8 years) participated in this experiment. The participants observed stimuli shown on an LCD monitor and evaluated their impressions on a seven-point Likert scale. In order to reduce the burden on them, we had them use 10 or 11 out of the 41 evaluation words. To take the order effect into consideration, the order of the evaluation words was randomized for each stimulus, and the order of the stimuli was randomized for each participant.

Through this experiment, we obtained data for 10 participants per stimulus and per evaluation word. Then, we scored each scale from -3 to 3 in one-point increments. For each evaluation word, we calculated the average score, and defined it as the evaluation score for the impression of each stimulus.

4.2 Factor Analysis

In order to extract and quantify latent factors that contributed to the impressions evoked by the patterns, we conducted a factor analysis of the evaluation scores using the maximum likelihood method and promax rotation. As a result, six factors were extracted, and the cumulative contribution ratio was 81.1%

(Table 1). We interpreted each factor as follows: Factor 1 was "pop"; Factor 2 was "elaborate"; Factor 3 was "refreshing"; Factor 4 was "novel"; Factor 5 was "tidy"; and Factor 6 was "stylish".

Table 1. Factor analysis results

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
cheerful	**0.971**	-0.171	-0.083	0.120	0.194	0.034
bright	**0.916**	-0.202	0.075	0.090	0.291	-0.001
gloomy	**-0.906**	0.145	-0.134	0.084	-0.155	-0.149
dark	**-0.881**	0.240	0.109	-0.272	-0.414	-0.024
mature	**-0.864**	0.177	0.070	0.011	0.280	0.442
vivid	**0.852**	0.256	-0.006	-0.138	0.116	0.121
plain	**-0.845**	-0.182	-0.104	0.050	-0.005	-0.083
stand-out	**0.844**	0.158	-0.023	-0.142	-0.296	0.003
lively	**0.830**	0.276	-0.046	0.005	0.109	-0.065
flashy	**0.775**	0.326	-0.148	-0.065	-0.079	0.060
calm	**-0.736**	-0.131	0.151	-0.054	0.340	-0.009
glittering	**0.727**	0.273	0.004	-0.050	-0.114	-0.109
colorful	**0.691**	0.275	-0.282	0.041	0.274	-0.118
Japanese-style	**-0.634**	0.260	-0.003	0.005	0.292	-0.374
modest	**-0.593**	-0.459	-0.153	0.074	0.209	-0.038
antique	**-0.564**	0.046	-0.320	-0.165	0.125	-0.214
luxuriant	**0.556**	0.427	-0.133	-0.222	0.458	0.261
rustic	**-0.489**	0.142	-0.268	0.034	-0.392	-0.290
contemporary	**0.433**	-0.078	0.335	0.274	-0.019	0.222
eccentric	**0.392**	0.286	-0.138	0.332	-0.341	0.052
complex	-0.142	**0.987**	0.084	0.204	0.186	-0.135
simple	0.056	**-0.955**	-0.035	-0.232	-0.202	0.068
multilayered	0.100	**0.869**	0.232	-0.110	-0.009	-0.126
rattling	0.325	**0.748**	0.148	-0.166	-0.204	-0.269
gorgeous	0.218	**0.613**	-0.241	-0.171	0.023	0.464
commonplace	-0.122	**-0.598**	0.009	-0.365	0.107	-0.058
mysterious	-0.495	**0.530**	-0.130	0.466	-0.119	0.118
cool-looking	-0.008	0.042	**0.973**	-0.091	0.179	-0.218
cool	-0.323	0.217	**0.965**	-0.026	-0.189	0.070
breezy	0.184	-0.035	**0.696**	-0.025	0.466	0.001
typical	0.029	-0.099	0.083	**-0.915**	-0.014	0.188
free	0.501	0.054	0.013	**0.696**	0.281	-0.206
unique	0.054	0.590	-0.076	**0.602**	0.101	-0.247
futuristic	0.222	0.254	0.115	**0.587**	-0.105	0.071
cute	0.271	-0.152	-0.013	0.186	**0.851**	-0.171
elegant	-0.563	0.293	-0.080	-0.154	**0.695**	0.290
beautiful	0.159	0.261	0.405	-0.166	**0.638**	0.237
sophisticated	-0.176	-0.232	0.323	-0.067	0.102	**0.610**
Western-style	0.360	-0.209	-0.152	-0.417	0.081	**0.537**
attractive	-0.074	0.216	0.362	0.221	-0.131	**0.413**
modern	0.213	0.192	0.180	0.186	-0.036	**0.347**

Finally, we defined the evaluation scores obtained from the subjective evaluation experiment and the factor scores obtained from the factor analysis as the quantified impressions (impression values) for the patterns.

5 Modeling the Relationships Between Visual Impressions and Physical Characteristics

5.1 Extraction of Style Features

We used the style features proposed by Gatys et al. [3] as physical characteristics representing patterns. The style features were Gram matrices of feature maps output from hidden layers of a pretrained VGG-19 network. Since they represent the detailed appearance in images, such as multi-scale color information and pattern information, they can be considered to be strongly related to visual impressions for clothing patterns. We extracted style features from pooling layers 1, 2, 3, and 4 of VGG-19 using the images labeled in Sect. 4. Their dimensions were 64×64, 128×128, 256×256, and 512×512, respectively.

5.2 Lasso Regression

We considered a regression problem to model the relationships between visual impressions and physical characteristics. In this research, over-fitting was expected since the style features were high-dimensional features with respect to the number of labeled patterns. Therefore, we adopted a lasso regression based on Takemoto et al. [6]. In a lasso regression, significant parts are selected from independent variables, and regression models are constructed, while over-fitting is prevented.

We conducted a lasso regression by setting the impression values as the dependent variables and the style features extracted from each pooling layer as the independent variables. Then, we used the penalty parameters obtained when mean square errors were minimized during K-fold cross validation (K = 7). For each factor and adjective, we constructed a regression model for each pooling layer and adopted the model with the greatest coefficient of determination among the four models as the impression estimation model. The impression estimation models were as follows: "bright," "breezy," and "cool-looking" were pooling layer 1; all six factors were pooling layer 2; "rustic" and "cool" were pooling layer 3; and the other 36 adjectives were pooling layer 4. The average coefficient of determination for all of the impression estimation models was 0.84 for six factors and 0.70 for 41 adjectives, and we confirmed that our models had high precision.

6 Validity of the Proposed Method

6.1 Impression Estimation for Unlabeled Patterns

Based on the models constructed in Sect. 5.2, we estimated the impressions for the remaining 1,083 unlabeled images. First, we extracted style features from the pooling layer adopted as each impression estimation model used them. After that, we input style features into the models and had the models calculate impression values for each pattern. In addition to the right side of Fig. 1, Fig. 2 shows the top 20 patterns with high calculated impression values for several of the factors and adjectives.

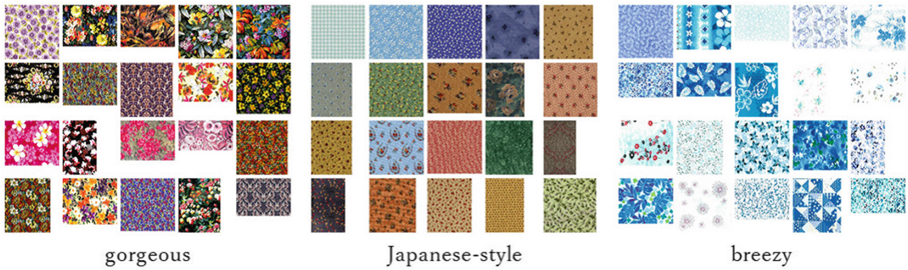

Fig. 2. Examples of the estimated results

6.2 Verification Experiment by Relative Evaluation

In order to confirm whether the results estimated in Sect. 6.1 corresponded to impressions that people actually felt, we first conducted a pairwise comparison experiment using Thurstone's paired comparison. Then, we psychologically scaled the actual impressions and obtained their ordinal structure. Finally, we compared it with that of the estimated impressions.

Among the evaluation words, after considering the number of adjectives included in each factor and their factor loadings, we selected the following 10 adjectives: "cheerful," "bright," and "colorful" in the "pop" factor; "complex" and "multilayered" in the "elaborate" factor; "cool-looking" in the "refreshing" factor; "free" in the "novel" factor; "cute" and "elegant" in the "tidy" factor; and "sophisticated" in the "stylish" factor.

Then, we selected the top two, the two around zero, and the bottom two of the calculated impression values for each evaluation word, and we used a total of 60 images as the stimuli in this experiment.

A total of 10 graduate and undergraduate students (5 men; 5 women; average age: 22.3 ± 1.1 years) participated in this experiment. The participants observed two stimuli shown on an LCD monitor and evaluated which one was applicable for each evaluation word. They used six images per evaluation word, and each participant conducted 150 trials. To take the order effect into consideration, the order of the stimulus pairs and evaluation words was randomized for each participant.

Figure 3 shows the psychological scale value of each pattern for each evaluation word. For "elegant," "bright," and "cool-looking," we confirmed that our models accurately estimated the ordinal structure, including for pattern groups whose calculated impression values were around zero. In addition, statistically significant differences existed between groups with high and low calculated impression values in seven out of 10 words.

6.3 Verification Experiment by Absolute Evaluation

We conducted a subjective evaluation experiment with the unlabeled images used in Sect. 6.1, and had people actually give them impression values. After

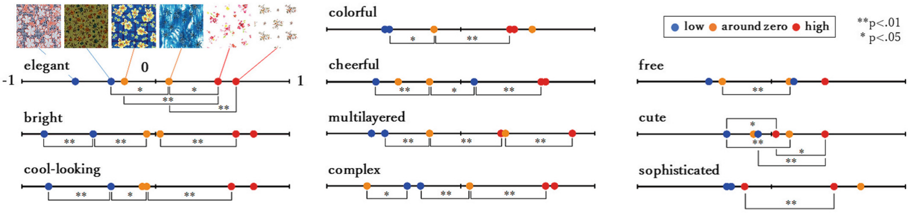

Fig. 3. Pairwise comparison results

that, we calculated the correlation coefficients between the impression values given in this experiment and the impression values calculated in Sect. 6.1, and then verified the validity of our models.

A total of 25 graduate and undergraduate students (15 men; 10 women; average age: 22.7 ± 3.3 years) participated in this experiment. The participants evaluated in the same procedure as in Sect. 4.1 with the evaluation words used in Sect. 6.2. However, in order to reduce the burden on them, we had them use one of the five stimuli sets into which the 1,083 images were divided.

Through this experiment, we obtained data for five participants to per stimulus and per evaluation word. Then, we obtained impression values in the same procedure as in Sect. 4.

Table 2 shows the correlation coefficients between the given impression values and the calculated impression values. There were extremely strong positive correlations for "cool-looking" and "bright," as well as positive correlations for "colorful," "cheerful," "complex," "cute," and "multilayered." "Free," "elegant," and "sophisticated" had weak positive correlations.

Table 2. Correlation coefficients results

cheerful	bright	colorful	complex	multilayered	cool-looking	free	cute	elegant	sophisticated
0.574	0.791	0.597	0.563	0.476	0.818	0.341	0.528	0.336	0.298

6.4 Discussion

From the results in Sect. 6.2 and Sect. 6.3, we confirmed that our models had high generalizability. In particular, models such as the "bright" and "cool-looking" models had high precision. On the other hand, models such as the "free" and "sophisticated" models had lower estimation precision than the other adjectives.

We considered individual differences in evaluations caused by ambiguities of words. Therefore, for each evaluation obtained in Sect. 4.1 and Sect. 6.3, we calculated the standard error between the participants for each stimulus and calculated its average between stimuli (Table 3). "Bright," "colorful," and "cool-looking" had a smaller average standard error than the other adjectives did. This suggests that these simple impressions with high estimation precision are less affected by individual differences. On the other hand, “free,” “elegant,” and “sophisticated” had a larger average standard error was larger than the other adjectives did. This suggests that these complex impressions involving multiple factors with low estimation precision are greatly affected by individual differences. Based on these observations, for complex impressions, it is necessary to extend the models so that they not only use the average of the given scores but also consider the dispersion of individual evaluations.

Table 3. Standard error results

	cheerful	bright	colorful	complex	multilayered	cool-looking	free	cute	elegant	sophisticated
Sect. 4.1	0.362	0.356	0.352	0.451	0.451	0.410	0.378	0.425	0.498	0.491
Sect. 6.3	0.346	0.312	0.337	0.365	0.346	0.312	0.399	0.380	0.399	0.435

6.5 Implementation a Pattern Search System

Since we confirmed the validity of our proposed method, we implemented a pattern search system using the estimated results. This system enables users to intuitively search for patterns with the words used in Sect. 4 as queries. It was applied to the fashion-on-demand app COUTURE made by digital fashion ltd. and is already being used in practical application (Fig. 4).

Fig. 4. The app COUTURE

7 Conclusion

We have proposed a method for automatically estimating the affective texture evoked by clothing patterns and conducted the following steps: (1) quantified the visual impressions of a small number of patterns with a subjective evaluation experiment and a factor analysis; (2) extracted style features as physical characteristics using a VGG-19 network; and (3) modeled the relationships between visual impressions and physical characteristics with a lasso regression. Afterward, based on the obtained models, we estimated the impressions for a large number of unlabeled patterns. Then, we verified their validity through relative evaluation and absolute evaluation, and we confirmed that our models had high generalizability. In addition, we implemented a system to enable users to intuitively search for patterns.

In the future, we will extend our models to consider individual differences and correspond with general patterns and shapes.

References

1. Julesz, B.: Textons, the elements of texture perception, and their interactions. Nature **290**(5802), 91 (1981)
2. Portilla, J., Simoncelli, E.P.: A parametric texture model based on joint statistics of complex wavelet coefficients. Int. J. Comput. Vis. **40**(1), 49–70 (2000). https://doi.org/10.1023/A:1026553619983
3. Gatys, L.A., Ecker, A.S., Bethge, M: Image Style Transfer Using Convolutional Neural Networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423 (2016)
4. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint arXiv:1409.1556
5. Tuceryan, M.A., Jain, A.K.: Handbook of Pattern Recognition and Computer Vision. Texture Analysis, pp. 235–276. World Scientific (1993)
6. Takemoto, A., Tobitani, K., Tani, Y., Fujiwara, T., Yamazaki, Y., Nagata, N.: Texture synthesis with desired Visual impressions using deep correlation feature. In: IEEE International Conference on Consumer Electronics, pp. 739–740 (2019)
7. Tobitani, K., Matsumoto, T., Tani, Y., Fujii, H., Nagata, N.: Modeling of the relation between impression and physical characteristics on representation of skin surface quality. J. Inst. Image Inf. Telev. Eng. **71**(11), 259–268 (2017)
8. Doizaki, R., Iiba, S., Okatani, T., Sakamoto, M.: Possibility to use product image and review text based on the association between onomatopoeia and texture. Trans. Jpn. Soc. Artif. Intell. **30**(1), 57–60 (2015)
9. Mori, T., Uchida, Y., Komiyama, J.: Relationship between visual impressions and image information parameters of color textures. J. Jpn. Res. Assoc. Text. end-uses **51**(5), 433–440 (2010)
10. Mouri, C., Ueda, E.S., Terauchi, F., Aoki, H.: Relationship between Mimetic words and kansei and sensory characteristics. Bull. Jpn. Soc. Sci. Des. **58**, 209 (2011)