

CNN 特徴量を用いた感性的質感認知に基づくテクスチャ生成

Texture Generation Based on Aesthetic Texture Perception Using CNN Feature

(キーワード：印象、スタイル、コンテンツ、Lasso 回帰)

(Keywords: Impression, Style, Content, Lasso Regression)

杉山幸音 (関西学院大学), 飛谷謙介 (長崎県立大), 長田典子 (関西学院大学)

1 はじめに

プロダクトデザインの分野において、感性的質感への関心が高まっている。感性的質感とは素材の表面性状から喚起される印象のことであり、ものの良さ悪しや好ましさを評価する上で重要であると考えられている。また、インターネットの普及とユーザーニーズの多様化によりプロダクトのカスタマイズ化やパーソナライズ化が進んでいる。その一例として、衣服のオーダーメイドサービスがある。衣服のオーダーメイドサービスでは、事前に用意された衣服の布地や柄、形状をユーザが選択し組み合わせることで個々の嗜好に沿った衣服をデザインできる。しかし、ユーザは事前に用意された選択肢の中から組み合わせを決めることしかできない。デザインの知見を持たないユーザにとってオリジナルなデザインをすることは難しい。そのため、選択肢を提供するだけでなくユーザが自らオリジナルなデザインを作れるように支援するシステムが必要である。これまでに著者らは、衣服の色柄の印象に基づいて多様なテクスチャを生成するシステムを構築することを目的に、CNN 特徴のうちスタイル特徴とコンテンツ特徴を用いたテクスチャ生成手法を提案してきた [1][2]。本研究では、コンテンツ特徴に基づいたテクスチャ生成方法を提案し、先行研究のスタイル特徴に基づいた方法との比較を行う。

2 先行研究

質感と関連が深いと考えられているテクスチャの解析に関する研究は古くから行われている。Gatys らは一般物体認識に用いられる畳み込みニューラルネットワークである VGG-19[3] から抽出される content feature (以下、コンテンツ特徴) 及び style feature (以下、スタイル特徴) に着目した画像変換アルゴリズムを提案した。この手法により、コンテンツ画像に描かれている物体の形状や構造にスタイル画像の画風を転移した画像を生成し、高精度な結果を示した。この研究により、コンテンツ特徴は形状情報を多く保持している特徴量であり、スタイル特徴は画像中の色やパターンの情報を多く保持している特徴量であることが示唆された [4]。特徴量抽出の外観を図 1 に示す。

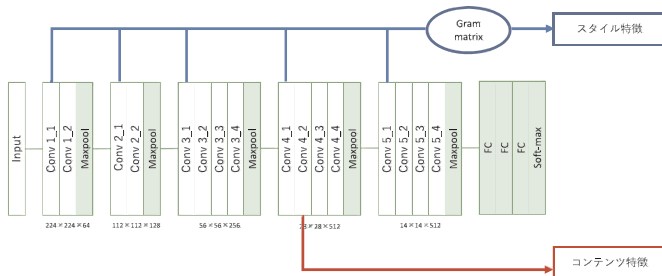


図 1 特徴量抽出の外観

また、竹本らはスタイル特徴がテクスチャと感性的質感に関連が強いと仮説を立てた。そこで、スタイル特徴を用い、感性

評価モデルを構築し、得られたモデルに基づき所望の感性的質感を有するテクスチャの生成方法を提案した [5]。

寸田らはスタイル特徴が衣服の柄と高い親和性を持つと仮説を立て、スタイル特徴を用いて感性的質感を推定する感性評価モデルの構築を行った。その後、構築した感性評価モデルを用いて、未知画像に対する感性的質感の推定値を求め、相対評価および絶対評価により検証実験による妥当性の検証を行い手法の有効性を示した [1]。そして、style と content をベースとしたテクスチャ生成システムを提案した。感性評価モデルに基づき所望の感性的質感を誇張したスタイル特徴を算出し、テクスチャの生成手法を提案した。その後、未知画像に対して所望の感性的質感を誇張するようにテクスチャを生成し、効果検証実験により元画像と比較して感性的質感が向上したことを確認し、提案手法が有効であることを示した [2]。

本研究では、コンテンツ特徴を用いて感性的質感を推定するモデルの構築と、所望の感性的質感を制御したテクスチャの生成を行う。研究フローを図 2 に示す。はじめに、事前学習済みの VGG-19 の中間層からコンテンツ特徴とスタイル特徴の抽出を行う。次に、柄画像に付与された評価点と抽出したコンテンツ特徴の関係を Lasso 回帰を用いて定式化し感性評価モデルの構築を行う。最後に、構築したモデルに基づき、感性的質感を制御したコンテンツ特徴を算出しテクスチャ生成を行う。

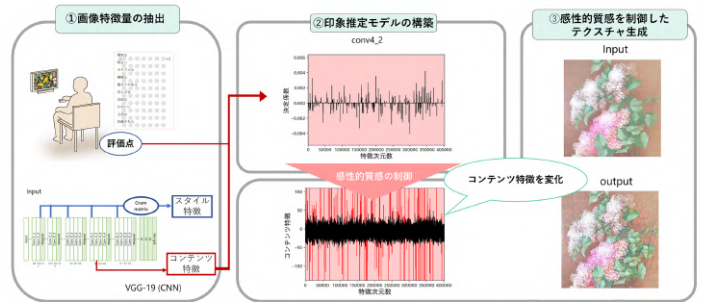


図 2 コンテンツベースのテクスチャ生成方法

3 感性的質感の定量化

画像特徴量の抽出および感性評価モデルの構築に、寸田らが提案したテクスチャ生成手法 [2] で用いられた柄画像 1158 枚と評価点のデータを用いる。寸田らは、はじめに自由記述実験により評価語の収集を行った。実験参加者 10 名に対し、PC 画面に表示される刺激について感じる印象を、自由に記述するよう求め、評価語の収集を行った。形容詞 98 語が得られ、意味が重複する形容詞を除き、計 60 語を収集した。次に適合度実験により評価語の選定を行った。実験参加者 10 名に対し、PC 画面にランダムに表示される刺激を観察し、収集した 60 語が柄から感じる印象を評価するのに適しているか、「非常に適している」、「適している」、「やや適している」、「どちらともいえない」、「やや適していない」、「適していない」、「非常に適していない」の 7 件法での評価を求めた。「非常に適していない」を 0 点、「非

常に適している」を6点とし、1点刻みの点数化を行い、平均3.6以上の28語を選定した。さらに、布や質感に関する先行研究[6][7]より形容詞12語と実利用の観点から「かわいい」を追加した計41語を評価語とした。

次に主観評価実験に用いる評価語と刺激の実験セットを構築した。評価語の中から「かわいい」を除いた計40語を用いて距離測定実験を行った。実験参加者10名に対し、ある評価語Aが他の評価語Bで置き換え可能であるか全780通りについて評価するよう求めた。その結果から、評価語間の距離行列を算出した。次に、得られた距離行列に対して階層クラスタ分析を行い、評価語40語を4クラスタに分割した。その後、各クラスタ中の評価語を2もしくは3個を含む計10語で構成される評価語セットA~Dを作成し、「かわいい」をC群に追加した。また、刺激セットは物理特徴に基づき階層クラスタ分析により構築した。物理特徴としてVGG19のプーリング層1, 2, 3, 4から抽出されるスタイル特徴を用いた。各特徴量に基づき全画像を20クラスタずつに分割し、計75枚を代表刺激セットとした。

そして予備実験として主観評価実験を行った。実験セットにつき1つにつき、実験参加者10名に対し、PC画面にランダムに提示される刺激について各評価語にどの程度当てはまるか、「非常に当てはまらない」、「かなり当てはまらない」、「やや当てはまらない」、「どちらでもない」、「やや当てはまる」、「かなり当てはまる」、「非常に当てはまる」からなる7件法で回答を求めた。「非常に当てはまらない」を-3点、「非常に当てはまる」を3点として1点刻みの点数化を行い、平均点を評価点として定義した。その後、得られた評価点を用いて因子分析を行い、5つの因子に分割した。因子分析結果から、各因子に含まれた評価語の数とその因子負荷量を加味し、「陽気な」、「明るい」、「カラフルな」、「複雑な」、「重なりのある」、「涼しげな」、「自由な」、「かわいい」、「上品な」、「洗練された」計10語を再選定した。その後追加実験を行った。柄画像1083枚をランダムに5群に分割し刺激セットを構築した。10語について実験参加者5名×5グループに対し、予備実験と同様の手続きで主観評価実験を行った[2]。

4 コンテンツ特徴を用いた印象推定モデルの構築

コンテンツ特徴はVGG19の中間層から抽出される特徴マップである。Gatysらの研究に基づき、Conv4.2からコンテンツ特徴の抽出を行った。抽出したコンテンツ特徴の次元数は $28 \times 28 \times 512$ である。説明変数が高次元であり過学習が起こると考えられるため、Lasso回帰を用いて定式化を行う。Lasso回帰は変数選択とL1正則化により、過学習を防ぎつつ回帰予測を行う手法である。Lasso回帰の罰則パラメーターはK分割交差検証により、平均二乗誤差が最小になった場合の値を使用する。スタージェスの公式により $K=11$ とする。目的変数を評価点、説明変数をコンテンツ特徴とし、Lasso回帰を用いて感性評価モデルの構築を行った。表1に評価語ごとにモデルの決定係数を示す。10個の評価語のうち9個の決定係数が0.5以上になった。決定係数の上位3個の評価語は「涼しげな」、「重なりのある」、「明るい」であり、下位3個の評価語は「洗練された」、「上品な」、「自由な」であった。決定係数が低い「洗練された」、「上品な」、「自由な」では評価点が3点以上-2点未満や、2点以上3点未満である極端な評価点を持つ柄画像が少ない(図3)。そのため、感性的質感と画像特徴量の関係を正確にモデル化できていないと考えられる。先行研究に基づき、決定係数が0.5以下である「洗練された」はテキストチャの生成に使用しない。

5 テクスチャ生成

テキストチャ生成するために、Lasso回帰で選択された変数が少ないと考えられるため、Lasso回帰で選択された変数と相関

表1 感性評価モデルの決定係数

陽気な	0.713
明るい	0.803
カラフルな	0.728
複雑な	0.800
重なりのある	0.832
涼しげな	0.887
自由な	0.616
かわいい	0.692
上品な	0.554
洗練された	0.371

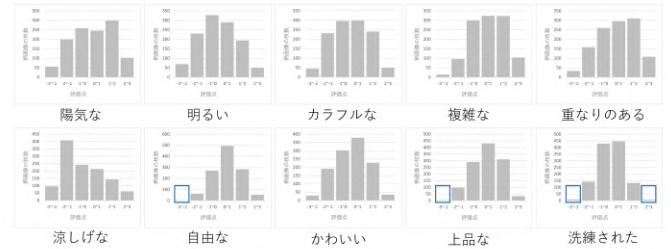


図3 評価点の分布(縦軸:柄画像の枚数,横軸:評価点)

が高い別の変数を併用して生成を行う。Lasso回帰で選択された変数と相関が0.8以上ある場合、選択された変数の決定係数の0.8倍の値を入力する前処理を行った。

テキストチャの生成は5段階で行われる。(i)入力画像からコンテンツ特徴とスタイル特徴を抽出する。(ii)コンテンツ特徴を誇張する。誇張は式1で行う。コンテンツ特徴を P 、モデルの構築により得られた回帰係数をコンテンツ特徴の形状に合うように変形したものを $\hat{\omega}_{reshape}$ 、重みを S とした。(iii)出力画像からコンテンツ特徴とスタイル特徴を抽出する。(iv)入力画像と出力画像の特徴量からコンテンツ特徴の誤差とスタイル特徴の誤差を求める。以降コンテンツ特徴の誤差をコンテンツ損失($L_{content}$)、スタイル特徴の誤差をスタイル損失(L_{style})とする。(v)コンテンツ損失に重み α 、スタイル損失に重み β を加えた合計を L_{total} とする。式(2) L_{total} を最小化するように出力画像を更新する。(iii)から(v)を n 回繰り返す。

コンテンツ特徴の制御は式1で行う。感性評価モデルの回帰係数に前処理を施したものを使用する。前処理後の回帰係数に重みパラメータをかけることで、変数が選択されなかった部分は0のままとなり、感性的質感と関連が強い部分の値のみ変化する。

$$P_{enhanced} = P \times (1 + \hat{\omega}_{reshape} \times S) \quad (1)$$

$$L_{total} = \alpha \times L_{content} + \beta \times L_{style} \quad (2)$$

6 結果と考察

入力画像のサイズを 224×224 、重みパラメータを $S=20000$ 、 $\alpha=10^6$ 、 $\beta=1$ 、 $n=300$ としてテキストチャの生成を行った結果を図4に示す。「複雑な」、「重なりのある」で先行研究のスタイルベース法で生成されたテキストチャと本研究のコンテンツベース法で生成したテキストチャを比較し、異なる傾向の質感の変化が観察された。図5に示すように、「複雑な」では画像中の明度が高い部分の明度が高くなり、明度が低い部分の明度が低くなるようなテキストチャが生成された。図6に示すように、「重なりのある」では明度が高い部分の明度が高くなり、彩度が高くなるようなテキストチャが生成された。また、コンテンツ特徴を制御

した場合にはスタイル特徴を制御した場合と比較し、明確な変化が起きていないことがわかる。Collomosse らの研究ではコンテンツとスタイルを分離し、画像検索のための視覚的類似性の尺度が提案された。スタイルのみを用いた検索において、従来の技術と比較し精度が向上した [8]。このことから、コンテンツ特徴と比較し、スタイル特徴は画像の質感に対してより関連がある特徴量であると考えられる。



図4 テクスチャの生成結果



図5 ”複雑な”におけるスタイルベース法、コンテンツベース法の比較



図6 ”重なりのある”におけるスタイルベース法、コンテンツベース法の比較

7 まとめ

本研究では、衣服の柄を対象とし、コンテンツ特徴を制御することで所望の感性的質感を有するテクスチャを生成する手法を提案した。はじめに、VGG19 からコンテンツ特徴を抽出し、Lasso 回帰により感性評価モデルを構築した。次に、感性評価モデルに基づきコンテンツ特徴を制御し、テクスチャの生成を行った。その結果、スタイル特徴を制御した場合と比較し変化が起きにくいことが分かった。

参考文献

参考文献

[1] Natsuki Sunda, Kensuke Tobitani, Iori Tani, Yusuke Tani, Noriko Nagata, and Nobufumi Morita.: Impression

estimation model for clothing patterns using neural style features. In International Conference on Human Computer Interaction, pages 689–697. Springer, 2020.

[2] 寸田菜月, 谷伊織, 飛谷謙介, 竹本敦, 谿雄祐, 長田典子, 森田修史: Neural style feature を用いた感性モデルに基づく質感表現. In 信学技報, volume 120 of MVE2020-18, pages 38–43, ONLINE, 9月2020.

[3] Karen Simonyan and Andrew Zisserman.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[4] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge.: Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[5] Atsushi Takemoto, Kensuke Tobitani, Yusuke Tani, Taishi Fujiwara, Yoichi Yamazaki, and Noriko Nagata.: Texture synthesis with desired visual impressions using deep correlation feature. 2019 IEEE International Conference on Consumer Electronics (ICCE), pages 739–740, 2019.

[6] 森俊夫, 内田裕子, 小見山二郎: 色彩テクスチャの視覚的印象と画像情報量との関係. 繊維製品消費科学, 51(5):433–440, 2010.

[7] 茂利千香子, 上田慎治エジウソン, 寺内文雄, 青木弘行: 擬態語と感性・感覚特性との関係について. 日本デザイン学会研究発表大会概要集, 58:209–209, 2011.

[8] John Collomosse, Tu Bui, Michael J. Wilber, Chen Fang, and Hailin Jin.: Sketching with style: Visual search with sketches and aesthetic context. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Oct 2017.