

多視点画像群を用いた DNN による 3 次元形状の印象推定 ～視線方向の影響に関する検討～

○阪下 啓祐[†] 飛谷 謙介[†] 田口 皓一[‡] 谷 伊織[†] 橋本 翔[†] 片平 建史[†] 橋本 学[‡] 長田 典子[†]

[†]: 関西学院大学理工学部人間システム工学科/感性価値創造研究センター

[‡]: 中京大学工学部機械システム工学科

sakashita.keisuke@kwansei.ac.jp

概要: 本研究の目的は, 高精度な 3 次元形状の印象推定モデルを構築することである. まず, 先行研究で構築された多視点画像群を入力とした DNN モデルを用いて, 印象推定に関わる視点重要度を定量化した. その結果, 人間の印象に寄与すると考えられる視点は, モデルにおいても重要度が高い傾向にあることが確認された. 次に, 視点重要度を学習可能な DNN の構造を提案し, 従来モデルとの精度比較を行った. その結果, 推定された印象評価値と人間の印象評価値との誤差が従来手法よりも減少したことから, 本手法の妥当性を確認した.

<キーワード> 3 次元形状, DNN, 印象推定, 感性

1. はじめに

近年, ModelNet40[1]を代表とする 3 次元モデルデータベースの整備や 3D プリンターの普及に伴い, 個人によるものづくり(パーソナルファブ리케이션)が注目を集めている[2]. しかし, ものづくり経験の少ない多くの個人は, 所望のデザインを実現するための専門的な知識や技術を有しておらず, このことが個人によるものづくりを困難にしている. 一方で, これらの個人であっても, 感性的な印象(滑らかな, 派手な, 硬い等)を用いたデザイン表現であれば比較的容易に行えると考えられる. そこで, 感性的な印象に基づいた形状の提案や検索等は, 個人によるものづくりの支援として有効だと考えられる. これらの支援を実現するには, 形状とそこから喚起される印象の関係性を明らかにするモデルが必要である.

これまでに田口らは, 3 次元形状に対して仮想的な複数視点から撮影した多視点画像群を用いた Deep Neural Network (DNN) により, 形状とそこから喚起される印象の関係性をマッピングした[3]. このような, 多視点画像群を入力とした DNN では, 視点毎の影響度が推定結果に大きく影響すると考えられる. しかし, 先行研究[3]では上述したような視点毎の影響度に関しては報告がなされていない.

そこで本研究では, まず従来の印象推定モデル[3]を用い, 印象推定時における各視点の影響度を定量化する. 次に, 本研究の目的である印象推定モデルの高精度化を達成するため, 定量化から得た知見をもとに, 新たな印象推定モデルを提案する.

2. 印象推定における各視点の影響

本章では, 印象推定に寄与する各視点の相対的な影響度を視点重要度として定義し, その算出方法と結果を示す. 加えて, モデル構築に用いるデータセットについても示す.

2.1. 視点重要度の算出方法

視点重要度の算出には, 先行研究[3]で用いられた印象推定モデル(図 1)を利用する. このモデルは, 表 1 で示す構造を持つ DNN である.

重要度の算出処理としては, 特に View-Pooling 層(図 1 中の②)に着目する. この層では, 平坦化した各視点の Convolutional Neural Network (CNN, 図 1 中の①)の出力を入力とし, それらを行方向に結合した後, 1 列ずつ視点間で比較し, 最大値を持つ視点の値を抽出することで, 次元圧縮を行っている. すなわち, View-pooling 層は印象推定に有効な視点の選択を行っていると言える.

そこで, 各視点が最大値をとる頻度を視点重要度とし, 視点毎の影響度の定量化をする. 本研究では, 学習済みモデルから各サンプルの 20 視点分の CNN の出力を取得し, 各視点が最大値を持つ頻度を算出した.

2.2. モデル構築に用いるデータセット

2.2.1. 印象評価実験による教師データ作成

3 次元形状(サンプル)の印象を定量化するため印象評価実験を行った. サンプルには, 物体認識タスクで広く利用されている公開データベース ModelNet40[1]から, 形状の多様性, 網羅性を担保す

るためにサンプル数が多い Car, Vase, Chair の 3 カテゴリーを採用した。そこから形状に欠陥のあるものを省き、Car 235 個, Vase 575 個, Chair 985 個を最終的なサンプルとして選定した。

続いて、印象評価を行う際の評価尺度(評価語)については、3次元造形物の主要な評価因子に関する検討を行った先行研究[4]で用いられた 18 形容詞対を選定した(表 2)。

実験において、評価者は液晶にランダムに提示されたサンプルを観察し、上述した評価語に関して各サンプルがどの程度当てはまるかを、7 段階 SD 法(-3:非常に当てはまらない, 2:かなり当てはまらない, -1:やや当てはまらない, 0:どちらでもない, +1:やや当てはまる, +2:かなり当てはまる, +3:非常に当てはまる)で評価した。評価者群は 20 代~60 代男女で、1 サンプルにつき 20 名~40 名の印象評価値が付与されている。この結果である印象評価値の分布(以降、評価値と記す)を、DNN の学習における教師データとした。

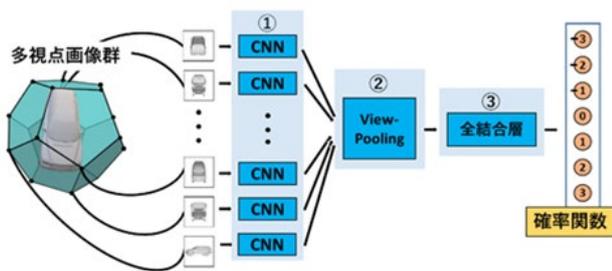


図 1 従来手法のモデルの概要

表 1 従来手法の DNN の構造

Layer		Input size	Output size	kernel	stride
CNN	Conv1	221×221×3	55×55×96	11×11	4
	Max-Pool1	55×55×96	27×27×96	3×3	2
	Conv2	27×27×96	27×27×256	5×5	1
	Max-pool2	27×27×256	13×13×256	3×3	2
	Conv3	13×13×256	13×13×384	3×3	1
	Conv4	13×13×384	13×13×384	3×3	1
	Conv5	13×13×384	13×13×256	3×3	1
	Max-Pool5	13×13×256	6×6×256	3×3	2
View-Pooling	reduce-Max	20×9216	9216	-	-
全結合層	FC1	9216	4096	-	-
	FC2	4096	4096	-	-
	FC3	4096	7	-	-

表 2 印象評価実験に用いた 18 形容詞対

なめらかな-粗い	弱い-強い
はげしい-おだやかな	はっきりした-ぼんやりした
柔らかい-固い	まとまった-ばらばらな
綺麗な-露骨な	重い-軽い
安定な-不安定な	動的な-静的な
活発な-不活発な	鋭い-鈍い
緩んだ-緊張した	好き-嫌い
規則的な-不規則的な	派手な-地味な
健康的な-不健康的な	陰気な-陽気な

2.2.2. 多視点画像群の作成

DNN への入力データとして、3D モデルから多視点画像群を作成した。

全サンプルの姿勢を統一した後、各サンプルを囲む正十二面体(図 1 中左部参照)の中心方向に向けて、正十二面体の各頂点(計 20 視点)からレンダリングを行った。これにより、位置と姿勢が統制された多視点画像群を作成した(図 2)。

2.3. 視点重要度の算出結果

2.1 節で述べた方法を用いて算出した視点重要度を図 3 に示す。横軸が視点番号、縦軸が視点重要度である。また、カテゴリによってサンプル数に差があるため、視点重要度をサンプル数で正規化した。

2.1 節で述べた学習済みモデルは、教師データに評価語「柔らかい-硬い」の評価値を用いて構築されている。また学習は Chair, Vase, Car のカテゴリ毎に行った。

なお、先行研究[4]において、評価語「柔らかい-硬い」は 3 次元造形物を表現する主要な評価因子であることが明らかにされている。このことから本研究でも、評価語「柔らかい-硬い」を代表として用いた。

視点番号	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
Car	
Vase	
Chair	

図 2 作成した多視点画像群 一例

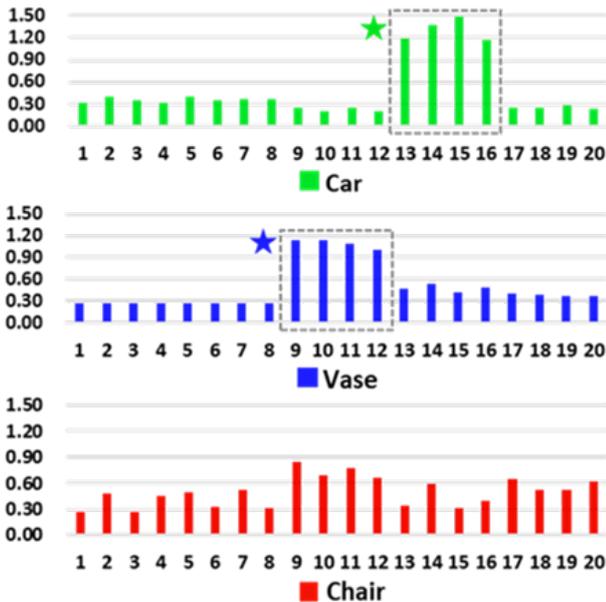


図3 評価語「柔らかい-硬い」におけるカテゴリ別の視点重要度



図4 Chair カテゴリのサンプル一例

3. 視点重要度を学習可能である新たな印象推定モデルの提案

3.1. 視点重要度に関する考察

図3より視点重要度はカテゴリに応じて異なる傾向をもつことが分かった。Car, Vase カテゴリでは重要度に顕著な偏り(図3の☆)が見られた。Car カテゴリではフロントやリア部分を写した視点の重要度が高く、ヒトが自動車のフロント部分から様々な印象を受けるという先行研究の報告[5]と整合的であった。Vase カテゴリにおいても、口や底などの特徴的な部位の視点における重要度が高くなった。さらに両カテゴリにおいて、同程度の重要度を示している視点群があった。これは形状の対称性により、当該視点群内でレンダリング画像が類似しているためと考えられる。

一方 Chair カテゴリでは、視点間で重要度に顕著な差は見られなかった。これは形状の自由度が高く(図4)、印象推定に必要な情報が複数の視点に分散したためと考えられる。このように、人間の印象に寄与する

と考えられる視点では重要度が高くなり、与える印象に差がない視点間では、その重要度は同程度になることを確認した。本結果は構築した印象推定モデルの妥当性を示唆している。

ここまでの議論において、従来手法の View-Pooling 層のように特定の視点のみに着目するアプローチでは、他の視点において有効な形状情報があっても損なわれる恐れがある。そこで次節では、視点重要度を学習したうえで、より多くの形状情報から印象推定を行う新たな DNN の構造を提案する。

3.2. 視点重要度に関する考察

3.1 節をもとに、視点重要度を学習可能にするため View-Pooling 層を拡張した(図5)。

まず、視点毎の重要度を学習する W 層を追加した(図5中の①)。W 層の活性化関数には式(1)の Rectified linear unit (ReLU 関数)を使用した。これは最終的に定まる視点重要度(W 層の重み)が負の値を取らないようにするためである。

次に、View-Pooling 層(図5中の②)において W 層の出力に対する平均をとることで、次元圧縮処理を最大値抽出から視点毎の重要度を用いた重み付き平均和に変更した。さらに全結合層にも変更を加えた。DNN の構造を表3に示す。これらの変更により、視点毎の重要度を学習したうえで全視点の情報を用いたモデルの構築が可能となった。

一方、損失関数や出力層の活性化関数は先行研究[3]と同様のものを使用した。教師データである評価値が7クラスあることから、DNN はクラス分類問題を採用した。出力層の活性化関数には式(2)の softmax 関数を使用した。

$$y = \max(x, 0) \quad (1)$$

$$p_i = \frac{\exp(y_i)}{\sum_{i=0}^n \exp(y_k)}, \quad i = 0 \dots n \quad (2)$$

$$E = - \sum_{i=1}^n t_k * \ln p_i, \quad i = 1 \dots n, \quad n = 7 \quad (3)$$

ここで、 y_i は各評価値クラスに対応する前層の各ユニットの出力とその結合重みを乗じた値である。この活性化関数を使用することで、評価値クラス*i*に対する確率関数 p_i を算出する。この確率関数 p_i と教師データの誤差 E を式(3)から算出し、 E が最小となるように学習アルゴリズム Adam[6]を用い、重みを更新する。式(3)はクロスエントロピー誤差関数であり、 t_k は教師データの*i*番目の値である。

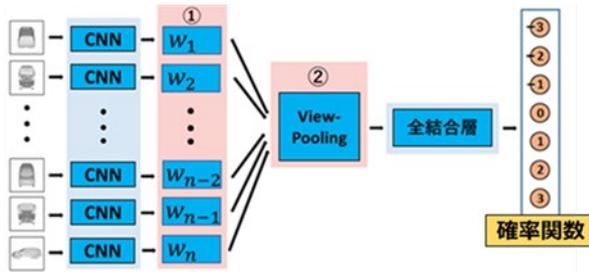


図 5 視点重要度を考慮したモデルの概要

表 3 視点重要度を考慮した DNN の構造

Layer		Input size	Output size	kernel	stride
CNN	Conv1	221×221×3	55×55×96	11×11	4
	Max-pool1	55×55×96	27×27×96	3×3	2
	Conv2	27×27×96	27×27×256	5×5	1
	Max-pool2	27×27×256	13×13×256	3×3	2
	Conv3	13×13×256	13×13×384	3×3	1
	Conv4	13×13×384	13×13×384	3×3	1
	Conv5	13×13×384	13×13×256	3×3	1
	Max-pool5	13×13×256	6×6×256	3×3	2
W層	-	6×6×256	6×6×256		
View-Pooling	reduce-Mean	20×9216	9216	-	-
全結合層	FC1	9216	4096	-	-
	FC2	4096	2304	-	-
	FC3	2304	1152	-	-
	FC4	1152	576	-	-
	FC5	576	288	-	-
	FC6	288	144	-	-
	FC7	144	72	-	-
	FC8	72	7	-	-

4. 精度検証

4.1. Jensen Shannon divergence (JSD) による精度比較

本節では、これまでに示した従来手法と提案手法、それぞれの手法による推定結果の比較を行う。比較する際の指標として、確率分布間の差異を表す JSD を用いた。これによって予測した検証データの評価値と人間による検証データの評価との分布差異を定量化した。

JSD の算出には式(4),(5),(6)を用いる。確率分布 p, q が等しいとき $D_{JS}(p, q)$ は 0 になる

$$D_{JS}(p, q) = \frac{DKL(p||m) + DKL(p||m)}{2} \quad (4)$$

$$m(x) = \frac{p(x) + q(x)}{2} \quad (5)$$

$$DKL(p||q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad (6)$$

9-fold 交差検証を行い、各検定における $D_{JS}(p, q)$ を平均した値で比較を行う。

結果を表 4 に示す。いずれのカテゴリにおいても、提案手法の $D_{JS}(p, q)$ が小さく、分布の再現度が高いことが示された。このことから提案手法の妥当性を確認した。

表 4 JSD を用いた各手法の推定精度

	Chair	Vase	Car
提案手法	0.081	0.064	0.054
従来手法	0.110	0.114	0.124

5. まとめと今後の方針

本研究では、高精度な 3 次元形状の印象推定モデルの構築を目的とし、まず先行研究のモデルを用いて印象推定に関わる各視点の影響を定量化した。その結果、人間の印象に寄与すると考えられる視点は、モデルにおいても、重要度が高くなることを確認した。

次に、視点重要度を学習可能な DNN の構造を提案し、従来手法との間で精度比較を行った。その結果、提案手法において、評価値の分布がよりよく再現されることが示された。このことから、視点重要度を学習可能な DNN の手法の妥当性を確認した。

今後は、先行研究[4]で挙げられている残り 2 つの 3 次元形状の主要な評価因子「安定-不安定」、「派手-地味」にも本提案手法を適用するとともに、さらなる推定精度向上を目指して DNN の構造を検討する。

謝辞 本研究は JST 研究成果展開事業 COI プログラム「感性とデジタル製造を直結し、生活者の創造性を拡張するファブ地球社会創造拠点」の支援によって行われた。

参考文献

- [1] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and, J. Xiao: 3d shapenets: A deep Representation for Volumetric Shapes, IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.1912-1920, 2015.
- [2] 田中浩也: パーソナルファブ리케이션時代におけるものづくりのオープンソース化の動向と Fab-Commons の提案, 情報処理, Vol.54, No2, pp.127-134, 2013.
- [3] 田口皓一, 橋本学, 飛谷謙介, 長田典子: 多視画像統合による 3 次元物体と感性的印象のマッピング, ビジョン技術の実利用ワークショップ講演論文集, pp.110-114, 2018.

- [4] 片平建史, 武藤和仁, 李奈栄, 飛谷謙介, 白岩史, 中島加恵, 長田典子, 岸野文郎, 山本倫也, 河崎圭吾, 荷方邦夫, 浅野隆: 3次元造形物の感性評価における主要因子, 日本感性工学会論文誌, Vol.15, No.4, pp.563-570, 2016.
- [5] 黄峻, 川澄未来子: タイ人向けの自動車フロントグリルデザイン提案のための感性の分析, 日本感性工学会論文誌, Vol.18, No.1, pp.128-137, 2019.
- [6] K. Diederik, B. Jimmy: Adam: A Method for Stochastic Optimization, International Conference on Learning Representations, arXiv preprint arXiv:1412.6980, 2014.

阪下啓祐: 現在関西学院大学理工学部人間システム工学科在学中. 感性情報学および質感に関する研究に従事.

飛谷謙介: 2002年早稲田大学理工学部応用物理学科卒業. 2004年岐阜県立情報科学芸術大学院大学(IAMAS) 修士課程修了. JST 地域結集型共同研究事業特別研究員を経て, 2010年岐阜大学大学院工学研究科博士後期課程修了. 同年岐阜大学産官学融合本部研究員. 2014年より関西学院大学理工学部/感性価値創造研究センター特任講師. 博士(工学). 主に感性工学, コンピュータビジョンに関する研究に従事. 電気学会, 精密工学会, 日本顔学会, ACM など各会員.

田口皓一: 2019年中京大学大学院工学研究科機械システム工学専攻修了. 3次元物体認識, 感性情報処理に興味を持つ. 精密工学会, 情報処理学会, IEEE 会員. 2017年度 IEEE Nagoya Section student Paper Award 受賞.

片平建史: 2004年大阪大学文学部卒業. 2011年同大学院人間科学研究科博士課程修了. 博士(人間科学). 2012年関西学院大学理工学研究科博士研究員, 2014年同理工学部/感性価値創造研究センター特任助教, 2015年同特任講師. 専門は感性心理学, 感情心理学, 社会心理学. 日本心理学会, 日本人間工学会など会員.

橋本翔: 2007年大阪大学人間科学部卒業. 2015年同大学院人間科学研究科博士課程修了. 博士(人間科学). 2015年関西学院大学理工学研究科博士研究員. 2017年同理工学部/感性価値創造研究センター特任助教. 専門は心理統計学, 多変量解析法の開発など. 日本行動計量学会, 日本計算機統計学会など各会員.

谷伊織: 2014年神戸大学理学研究科地球惑星科学専攻博士課程後期課程修了. 早稲田大学客員次席研究員, 総合研究大学院大学学術情報基盤センター助教を経て, 2018年より関西学院大学理工学部感性価値創造研究センター研究特別任期制助教. 博士(理学). 人工知能による感性情報処理に関する研究に従事. 自然計算手法を用いた感性情報処理に興味を持つ. 計測自動制御学会, 共創学会正会員.

橋本学: 1987年大阪大学大学院修了. 同年三菱電機(株)入社. 生産技術研究所, 先端技術総合研究所に勤務. 2008

年より中京大学教授. 2017年より工学部長. 3次元物体認識, ロボットビジョン, ヒューマンセンシングの研究などに従事. 2012/2017年度画像センシングシンポジウム優秀学術賞, 2015年精密工学会小田原賞など受賞.

長田典子: 1983年京都大学理学部数学系卒業. 同年三菱電機(株)入社. 1996年大阪大学大学院基礎工学研究科博士後期課程修了. 2003年より関西学院大学理工学部情報科学科助教授, 2007年教授. 2009年米国パデュー大学客員研究員. 2013年感性価値創造研究センター長. 2015年革新的イノベーション創出プログラム「感性とデジタル製造を直結し, 生活者の創造性を拡張するファブ地球社会創造拠点」サテライトリーダー. 博士(工学). 専門は感性工学, メディア工学等.