

## 研究速報

## 多視点画像群を用いた DNN による 3 次元物体の印象推定

飛谷 謙介<sup>†a)</sup> 田口 皓一<sup>††</sup>  
 橋本 学<sup>††</sup> (正員) 阪下 啓祐<sup>†</sup>  
 谷 伊織<sup>†</sup> 橋本 翔<sup>†</sup>  
 片平 建史<sup>†</sup> 長田 典子<sup>†</sup> (正員)

Impression Estimation of 3D Object by DNN Using Multi-View Images

Kensuke TOBITANI<sup>†a)</sup>, Koichi TAGUCHI<sup>††</sup>, *Nonmembers*,  
 Manabu HASHIMOTO<sup>††</sup>, *Member*, Keisuke SAKASHITA<sup>†</sup>, Iori TANI<sup>†</sup>,  
 Sho HASHIMOTO<sup>†</sup>, Kenji KATAHIRA<sup>†</sup>, *Nonmembers*,  
 and Noriko NAGATA<sup>†</sup>, *Member*

<sup>†</sup> 関西学院大学理工学部/感性価値創造インスティテュート, 三田市  
 School of Science and Technology, Research Institute for Kansei Value Creation,  
 Kansai Gakuin University, 2-1 Gakuen, Sanda-shi, 669-1337 Japan

<sup>††</sup> 中京大学大学院工学研究科, 名古屋市

Graduate School of Engineering, Chukyo University, 101-2 Yagoto Honmachi,  
 Showa-ku, Nagoya-shi, 466-8666 Japan

a) E-mail: tobitani@kansai.ac.jp

DOI:10.14923/transinfj.2020JDL8003

あらまし 本研究では、3次元物体の直感的なデザイン支援を最終的な目標とし、その最初の試みとして、3次元物体の印象推定手法を提案する。推定には多視点画像群を入力とするDNNを使用した。検証実験において推定値が実際の印象と正の相関関係を示し、本研究の有効性を確認した。

キーワード 3次元物体, 印象推定, DNN, 多視点画像

## 1. まえがき

近年、3Dプリンタ技術の急速な発展、及び低価格化による普及により、3次元物体を出力するための環境が整いつつある。このような環境変化は、幅広いユーザに対して個人的なものづくり活動の機会を提供する可能性を有する。しかしながら、3次元物体のデザインには専門的な知識や経験、技術が必要であり、デザインに対する潜在的なニーズをもっている、それを物体形状に反映させることは一般のユーザには難しいと予想される。そのため、幅広いユーザに対して個人的なものづくり活動の機会を提供するためには、ユーザが容易に3次元物体のものづくりに取り組めるような支援技術、例えば印象を表現する語(例:「もっとやわらかい感じに」)による形状の提案等が不可欠だと考えられる。これらの支援を実現するためには、物体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必要がある。

そこで本研究では、3次元物体の直感的なデザイン

支援を最終的な目標とし、その最初の試みとして、3次元物体の印象推定手法を提案する。その際、質感や色といった要素は考慮せず、形状に対する印象のみを対象とする。推定には、3次元物体を多視点からレンダリングした画像群を入力データとするDeep Neural Network(以下DNN)を使用する。更に、定量化した印象が付与された3次元モデルデータセットを用い、提案手法の精度検証のための実験を行うことで本研究の有効性を明らかにする。

## 2. 先行研究

対象物に対する印象を活用した研究事例として、印象を表現する語(以下評価語)をクエリとした検索技術が挙げられる[1]~[4]。これらの技術はいずれも対象の物理的な特徴量と喚起される印象との関係性を用いて実現している。このような印象推定に関する研究は主に2次元画像を対象にしてきたが、3次元物体に応用した事例も報告されつつある[5],[6]。しかしながら、これらの研究では幾何学的で単純な3次元形状を用いており、多様な特徴をもつ一般的な3次元形状の印象評価については未だ体系的な結果が報告されていない。そのため、本研究の目的であるデザイン支援を想定した場合、その応用は限定的になると予想される。

一方で、3次元物体から印象を推定するということは、目的変数を印象値、説明変数を3次元物体の物理的な特徴量とした分類・回帰による認識問題と言い換えることができる。3次元物体認識に関する研究は、当初、当該研究の重要な構成要素である3次元特徴量の設計に関する研究が盛んに行われていたが[7]、近年はDNNを利用することで、End-to-Endで学習し、認識を行う手法が主流となっている。このDNNを基にした認識手法は、主にRGBDベース[8],[9]、Point Cloudベース[10],[11]、Voxelベース[12],[13]、Multi-viewベース[14],[15]に大別でき、現状、大規模3次元物体認識においてMulti-viewベースの手法が他手法と比較して高精度だと言われている[16]。

これまでに我々は多様な抽象的形状を用いた印象評価実験を行い、様々な3次元形状に共通し得る印象評価の構造を明らかにした[17],[18]。本研究では、この印象評価構造に基づき、印象を定量化した値(印象値)を付与した、複数の形状カテゴリーの3次元モデルデータセットを作成し、学習データとして使用する。更に、Multi-viewベースのDNNを用い、当該データセットをEnd-to-Endで学習することで、対象とする印象や形状カテゴリーごとに特徴量設計を行う必要がな

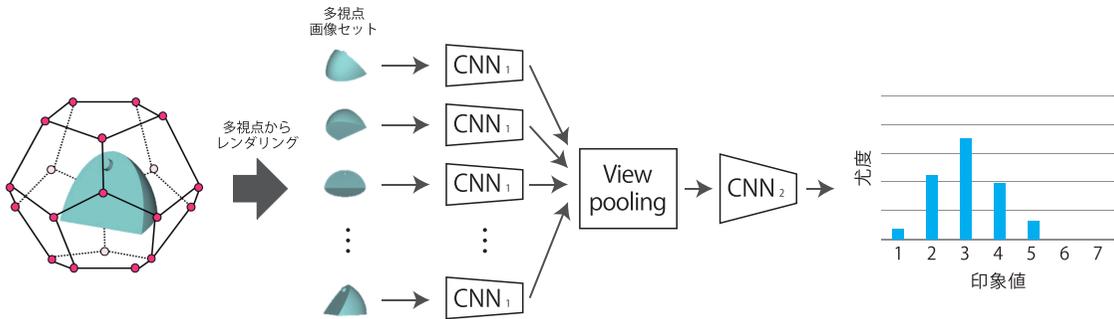


図1 提案手法の概念図

い、単一の枠組みでの多様な3次元物体の印象推定を実現する。

### 3. 印象推定モデルの構築

#### 3.1 基本設計

本研究では Multi-view ベースの MVCNN [14] を用いて3次元物体の印象を推定する。提案手法の概念図を図1、提案手法で用いる DNN の仕様を表1に示す。MVCNN は、多視点画像群を入力データとし、視点情報を統合する View pooling 層を介して出力層でクラス分類を行う。当該ネットワークにおいて教師データとして与えるクラスラベルの番号を、印象の強弱と対応させることで、クラス分類として印象推定を実現する。

また、ネットワークに入力する多視点画像群は複数の視点から3次元モデルをレンダリングすることで得られる。レンダリングには Phong の反射モデルを用い、カメラ位置は3次元モデルの重心を中心とした正十二面体の各頂点とする。各カメラ位置から、正十二面体に内包された3次元モデルの重心方向に向かってレンダリングすることにより、20枚1セットの画像セットを作成する。その際、3次元モデルの位置姿勢は統一した。本研究では個人によって評価のばらつきが生じる印象値を離散確率分布として扱うため、損失関数としてクロスエントロピー誤差を用いる。

#### 3.2 データセットの作成

3次元物体から印象を推定するには、印象値が付与された3次元モデルのデータセットが必要である。そこで先行研究[17]に基づき、大規模3次元モデルデータベースである ModelNet40 における三つの形状カテゴリー (car, vase, chair) に対し、クラウドソーシングを利用した SD 法による印象評価実験を行った。これにより各モデルに対し、20~40人分の18種類の形容詞対における7段階の評価データを取得した。得られた評価データに対し、回答内容や回答時間から評価

表1 提案手法における DNN の仕様

Layer	Input size	Output size	kernel	stride
conv1	221x221x3	55x55x96	11x11	4
maxpool1	55x55x96	27x27x96	3x3	2
conv2	27x27x96	27x27x256	5x5	1
maxpool2	27x27x256	13x13x256	3x3	2
conv3	13x13x256	13x13x384	3x3	1
conv4	13x13x384	13x13x384	3x3	1
conv5	13x13x384	13x13x256	3x3	1
maxpool3	13x13x256	6x6x256	3x3	2
View-pooling	20x9216	9216	-	-
CNN2	FC1	9216	4096	-
	FC2	4096	4096	-
	FC3	4096	7	-

データのクリーニングを行った。クリーニングの手順を以下に示す。

(1) 回答システムの誤作動やネットワークの状況により評価が正常に終了できなかった実験参加者を未評価回答者とし、その全ての評価データを排除

(2) 1個以上のモデルに、全て0(7段階評価の真中：どちらでもない)と回答した実験参加者は不誠実回答者とし、その全ての評価データを排除

(3) 不誠実回答者の回答時間分布を確認し、その最頻値である156秒以下で回答を終了した実験参加者を短時間回答者とし、その全ての評価データを排除

クリーニングの結果、総回答者6101人の内、未評価回答者388名、不誠実回答者628名、短時間回答者307人をそれぞれ排除し、最終的に有効回答者は4778名となった。データのクリーニング後、7段階の評価データを1から7のクラスラベルに数値化し、その値を確率変数とした離散確率分布を3次元モデルの印象値(厳密には印象分布)とした。表2に使用した作成したデータセットの仕様を、表3に使用した18種類の形容詞対をそれぞれ示す。

表2 作成したデータセット仕様

形状カテゴリー	car	vase	chair
モデル数	297	575	989
データベース	ModelNet40		
1モデルあたりの評価数	20~40名		
対象実験参加者	20~60代男女(各年代約30名)		
3Dモデル提示形式	モデル回転動画		
評価語	18形容詞対		
評価尺度	7段階のSD法		

表3 印象評価実験に使用した形容詞対

形容詞対
規則的な-不規則的な
健康的な-不健康な
安定した-不安定な
まとまった-ばらばらな
はげしい-おだやかな
派手な-地味な
活発な-不活発な
賑やかな-静かな
動的な-静的な
陽気な-陰気な
繊細な-武骨な
柔らかい-硬い
緩んだ-緊張した
弱い-強い
はっきりした-ぼんやりした
なめらかな-粗い
鈍い-鋭い
重い-軽い

#### 4. 精度検証実験

提案手法の有効性を検証するために、Voxelベースの手法であるShapeNets [12] と、提案手法に単視点の画像を入力した場合、及び提案手法の各手法における推定精度を比較する。

##### 4.1 実験手続き

精度比較を行う上で、本研究では9分割の交差検証を採用した。その際、分割されたデータセットは手法間で共通とした。ネットワークに教師データとして印象値を与える際は、離散確率分布ではなく、評価データ数分のOne-hotベクトルを用いる。具体的には、実験参加者が評価したクラスラベルには1、それ以外のクラスラベルは0とした7次元ベクトルに変換し教師データとした。対象とする印象は表3に示した18種類の形容詞対の内、SD法研究で頻出する3因子(評価性・活動性・力量性因子) [19] が3次元形状の印象においても主要な基準を構成するという先行研究 [17] の知見に基づき選定した。先行研究において、この3因子への負荷量大きい形容詞対は“安定な-不安定な(評価性)”, “派手な-地味な(活動性)”, “柔らかい-硬

い(力量性)”であるため、これらの形容詞対に対する印象値を推定対象とした。

##### 4.2 結果

本研究では、実際の人の評価データと推定結果との相関関係を推定精度の比較指標とする。具体的には、検証用データにおける人の評価データの離散確率分布及び推定された離散確率分布それぞれにおいて期待値を算出し、期待値間のピアソン積率相関係数を用いた。各検証で得られた相関係数の平均を表4に示す。

##### 4.3 考察

表4より、提案手法において9条件中、2条件で強い正の相関(0.7以上)、6条件で中程度の正の相関(0.4以上0.7未満)を示したことから、提案手法の実用上の有効性を確認した。また、比較手法に対する提案手法の有効性を統計的に検定するため、提案手法(Multi-view)をコントロール群、各比較手法を処理群としたDunnett法による多重比較検定を行った。その際、対立仮説は $\mu_c > \mu_i$ とした( $\mu_c$ :コントロール群の相関係数の平均値、 $\mu_i$ :処理群の相関係数の平均値)。その結果、提案手法とVoxelベースの手法 [12] との間では全9条件で有意差( $p < 0.05$ )が認められた。この結果は、Voxelベースの手法がその構造上、入力データを低解像度化せざるを得ないことに起因していると考えられる。3次元形状に対する人の印象は、大まかな形状だけでなくその微細な構造によって喚起される可能性があり、Voxelベースの手法を用いた場合、その情報が消失してしまう可能性があることが予想される。更に、提案手法とSingle-viewとの間では、Vaseの派手-地味条件、Chairの派手-地味条件、安定-不安定条件以外の6条件で有意差( $p < 0.05$ )が認められた。この結果は、印象評価実験の際、実験参加者には3次元物体を水平方向に回転させた動画を提示するという、多方向からの印象を総合的に評価するような実験デザインであったことに起因すると考えられ、提案手法はこのような印象の視角依存性を学習可能な枠組みであることが示唆される。しかしながら、幾つかの条件下では視点の増加による推定精度の向上が確認できなかった。これは、ある条件下では印象推定に寄与しない、若しくは精度の悪化に繋がる視点が存在することに起因していると考えられる。そのため、印象推定に寄与している視点の選択や重み付けなどの処理が精度向上に有効だと考えられる。

また、「安定-不安定」条件において、形状カテゴリーによって推定精度に差があることがわかる。この現象

表4 各手法における評価データと推定値の相関係数の平均

car				
手法	入力形式	柔らかい-硬い	派手-地味	安定-不安定
ShapeNets [12]	Voxel	0.42	0.62	0.01
Single-view	単一画像	0.43	0.71	0.08
Multi-view (提案手法)	多視点画像	0.52	0.78	0.18
vase				
手法	入力形式	柔らかい-硬い	派手-地味	安定-不安定
ShapeNets [12]	Voxel	0.05	0.31	0.61
Single-view	単一画像	0.36	0.51	0.66
Multi-view (提案手法)	多視点画像	0.47	0.53	0.73
chair				
手法	入力形式	柔らかい-硬い	派手-地味	安定-不安定
ShapeNets [12]	Voxel	0.44	0.30	0.48
Single-view	単一画像	0.57	0.58	0.62
Multi-view (提案手法)	多視点画像	0.65	0.55	0.63

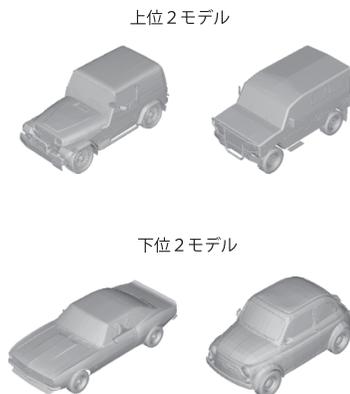


図2 Car カテゴリーにおける「安定-不安定」の評定値が高いモデルと低いモデル

は、同じ印象であってもその判断根拠が形状カテゴリーごとに異なることに起因すると思われる。そこで、Car カテゴリーの「安定-不安定」条件における評定値の高いモデルと低いモデルを視覚的に確認した。評定値上位2サンプルと下位2サンプルを図2に示す。図2より、上位2サンプルはオフロードカーで、下位2サンプルはスポーツカーとコンパクトカーであることがわかる。このことから、当該条件では物理的な特徴だけでなく、機能的側面も評価基準に含まれていると予想される。このように評価基準が混在した場合、学習データ数が少ないと、その影響が大きくなる。形状カテゴリーの内 Car カテゴリーが最もサンプル数が少ないため、当該条件で大きく精度が低下したと思われる。そのため、推定精度の更なる向上には学習データの拡充が必要である。

## 5. む す び

本研究では、3次元物体の直感的なデザイン支援を最終的な目標とし、その最初の試みとして、3次元物体の印象推定手法を提案した。推定には、Multi-viewベースのMVCNNを用い、独自に作成した印象値が付与された3次元モデルデータセットをEnd-to-Endで学習することで、単一の枠組みによる多様な3次元物体の印象推定を試みた。3種類の形状カテゴリーにおける検証用データ用い、「安定な-不安定な」、「派手な-地味な」、「柔らかい-硬い」の三つの印象について、推定結果と実際に人が付与した印象との相関係数を算出したところ、9条件中8条件で中程度以上の正の相関関係を示し、提案手法の実用上の有効性を確認した。また、従来手法と精度比較をした結果、Voxelベースの手法に対しては全ての条件で、単視点の画像を入力とした手法に対しては9条件中6条件で、提案手法の有効性を統計的に信頼できる形で確認した。

今後は、提案手法の推定精度や汎化性能の更なる向上を目指し、学習データの拡充や他の形状カテゴリーでの精度検証を行いたい。また、本研究の最終的な目標である3次元物体の直感的なデザイン支援を考慮した場合、印象推定の判断根拠を可視化する必要がある。更に、提案手法は視点数が多ければ多いほど学習時間が増加してしまう構造であり、また、印象推定に寄与しない、若しくは精度の悪化に繋がる視点が存在することが予想されるため、視点の最適化についても検討する必要がある。

謝辞 本研究は、JSTCOI, JPMJCE1314の支援を受けたものである。

## 文 献

- [1] 栗田多喜夫, 加藤俊一, 福田郁美, 坂倉あゆみ, “印象語による絵画データベースの検索,” 情処学論, vol.33, no.11, pp.1373–1383, Nov. 1992.
- [2] 荻野晃大, 加藤俊一, “感性検索システムの設計手法: 感性システムモデリング,” 情処学論, vol.47, no.SIG4(TOD29), pp.28–39, March 2006.
- [3] 太田 茂, 竹之内宏, 徳丸正孝, “深層ニューラルネットワークによる特徴抽出を用いた衣服の感性検索,” 日本感性工学会論文誌, vol.16, no.3, pp.277–283, Dec. 2017.
- [4] Y.W. Chen, X. Huang, D. Chen, and X.H. Han, “Generic and specific impressions estimation and their application to KANSEI-based clothing fabric image retrieval,” Int. J. Pattern Recognit. and Artif. Intell., vol.32, no.10, 1854024, April 2018.
- [5] 向江亜紀, 加藤俊一, “3次元物体の形状と質感に対する感性のモデル化,” 情処学論, vol.47, no.SIG8(TOD30), pp.134–146, June 2006.
- [6] L.C. Ou, W.Y. Lee, and M. Luo, “Assessing the affective feelings of two- and three-dimensional objects,” Color Research and Application, vol.34, no.1, pp.75–83, Feb. 2009.
- [7] 橋本 学, 秋月秀一, 武井翔一, “物体認識のための3次元特徴量の研究動向,” 電学論 (C), vol.136, no.8, pp.1038–1046, Aug. 2016.
- [8] K. Lai, L. Bo, X. Ren, and D. Fox, “Sparse distance learning for object recognition combining RGB and depth information,” Proc. 2011 IEEE Int. Conf. Robotics and Automation, pp.4007–4013, Shanghai, China, May 2011.
- [9] X. Song, L. Herranz, and S. Jiang, “Depth CNNs for RGB-D scene recognition: Learning from scratch better than transferring from RGB-CNNs,” Proc. Thirty-First AAAI Conference on Artificial Intelligence, pp.4271–4277, California, USA, Feb. 2017.
- [10] J. Li, B.M. Chen, and G.H. Lee, “SO-Net: Self-organizing network for point cloud analysis,” Proc. 2018 IEEE Conference on Computer Vision and Pattern Recognition, pp.9397–9406, Utah, USA, June 2018.
- [11] H. Deng, T. Birdal, and S. Ilic, “PPFNet: Global context aware local features for robust 3D point matching,” Proc. 2018 IEEE Conference on Computer Vision and Pattern Recognition, pp.195–205, Utah, USA, June 2018.
- [12] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, “3D ShapeNets: A deep representation for volumetric shapes,” Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp.1912–1920, Massachusetts, USA, June 2015.
- [13] N. Sedaghat, M. Zolfaghari, E. Amiri, and T. Brox, “Orientation-boosted Voxel Nets for 3D Object Recognition,” Proc. 28th British Machine Vision Conference, London, United Kingdom, Sept. 2017.
- [14] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, “Multi-view convolutional neural networks for 3D shape recognition,” Proc. 2015 IEEE International Conference on Computer Vision, pp.945–953, Santiago, Chile, Dec. 2015.
- [15] A. Kanazaki, Y. Matsushita, and Y. Nishida, “RotationNet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints,” Proc. 2018 IEEE Conference on Computer Vision and Pattern Recognition, pp.5010–5019, Utah, USA, June 2018.
- [16] Large-Scale 3D shape retrieval from shapenet CORE55, ShapeNet, <https://shapenet.cs.stanford.edu/shrec17/>, April 2017.
- [17] 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次元造形物体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, Aug. 2016.
- [18] 片平建史, 武藤和仁, 飛谷謙介, 橋本 翔, 長田典子, “SD法を用いた感性の測定における評価の階層性-EPA構造の評価因子の多義性に注目して-,” 日本感性工学会論文誌, vol.17, no.4, pp.453–463, July 2018.
- [19] C.E. Osgood, D. Suci, and P.H. Tannenbaum, The Measurement of Meaning Stage of Design, University of Illinois Press, 1957.

(2020年1月20日受付, 6月30日再受付,  
8月3日早期公開)