精密工学会 画像応用技術専門委員会
サマーセミナー2023 テキスト Vol.32(8.28-8.29)

Multi-view CNN を用いた３次元形状の印象推定モデル

杉山 幸音† 飛谷 謙介‡ 長田 典子† 橋本 学¶

†関西学院大学大学院理工学研究科 〒669-1330 兵庫県三田市学園上ヶ原 1
‡長崎県立大学情報システム学部 〒851-2130 長崎県西彼杵郡長与町まなび野 1-1
¶中京大学工学部 機械システム工学科 〒466-8666 愛知県名古屋市昭和区八事本町 101-2

E-mail: †{ggs53875, nagata}@kwansei.ac.jp, ‡tobitani@sun.ac.jp ¶mana@isl.sist.chukyo-u.ac.jp

あらまし 本研究では，3 次元物体の印象に基づく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，多様な形状特徴を持つ一般的な 3 次元物体の形の印象推定手法を提案する．印象推定には多視点画像群を入力とする Multi-view CNN (Convolutional Neural Network) を使用し，主観評価の分布をデータセットとすることにより，高い精度を持つ印象推定モデルを実現できた．検証実験では推定結果（推定印象分布）が主観評価分布と中程度以上の正の相関関係を示し，従来研究より高い推定精度を持つことが確認された．さらに人の印象評価のばらつきの傾向を捉えることも示唆された．

キーワード プロダクトデザイン, DNN, 印象推定モデル, 多視点画像.

1. はじめに

第５次産業革命では「人間中心」と「持続可能性」が重要なコンセプトとなっている．この技術革命の実現において，3D プリンタをはじめとするデジタルファブリケーション技術が重要な役割を担っている．例えばユーザニーズにあわせた多様な造形物を生成したり（カスタマイゼーション／ビスポーク），不要なモノをより良いモノに作り替えて付加価値を生み出したりする（アップサイクル）ことを可能にする．

一方で，モノのデザインには専門的な知識や技術が必要であるため，一般ユーザが，自身の持つ潜在的なニーズやイメージを具体的な物体形状に反映させることは容易ではない．そこで本研究では，ユーザのイメージ（印象）に基づく 3 次元物体のデザイン支援を最終的な目的とし，多様な形状特徴を持つ一般的な 3 次元物体の印象推定手法を提案する．

2. 先行研究

これまでに３次元物体に関する研究として，形状認識手法や形状特徴量に関する研究が盛んに行われている．特に，最近では Deep Neural Network（以下，DNN）を利用した認識手法や特徴量設計が主流となっている．大規模３次元物体認識においては Appearance 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．飛谷らは，Appearance ベースの手法の一つである Multi-view ベースと呼ばれる，３次元形状に対して仮想的な複数の視点から撮影した多視点画像群を用いる MVCNN を応用し，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．本研究では，飛谷らの印象推定モデルを発展させ，データ数を増やした上で精度検証実験を行い，Multi-view ベースの手法の優位性を示す[1]．さらに学習済みの印象推定モデルを未学習の別ドメイン（カテゴリ）に適用することを試みる．

3. 提案手法

本研究では，MVCNN を用いた 3 次元物体の印象推定手法を提案する．図 1 に提案手法の概要図を示す．まず，3 次元物体を対象に主観評価実験を行い，3 次元物体から喚起される印象を定量化する．これによって得られる印象分布をモデルの教師信号とする．次に，3 次元物体を複数の視点からレンダリングした多視点画像を作成し，これをモデルの入力信号として用いる．最後に，MVCNN を用いて前述の入力信号と教師信号の関係性をモデル化することで，3 次元物体の印象推定を実現する．以上の基本設計により，MVCNN を用いた 3 次元物体の印象推定タスクの検証が可能になる．

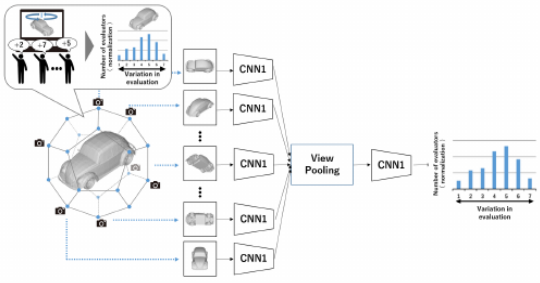


図 1 提案手法の概要図

4. 印象推定モデルの構築

4.1. データセット

本節では印象推定モデルの構築に必要となるデータセットの作成方法について示す．

4.1.1. 3D モデルデータの収集と選択

3 次元モデルデータの収集・選定を行う. 形状表現の多様性・網羅性を確保するため，大規模な 3 次元モデルのデータベースである ModelNet40, ShapeNet, CG DATA BANK から 3 次元モデルの収集を行った. その結果, Car カテゴリ（632 個）, Vase カテゴリ（575 個）, Chair カテゴリ（985 個）の 3 次元モデルデータを収集した.

4.1.2. 3D 形状のレンダリング

3 次元モデルを用いて, 実験刺激を作成する. 実験刺激は 2 種類あり，推定モデルの入力信号に用いる多視点画像と，主観評価実験用の映像がある．両実験刺激の作成において, 共通の前処理を示す. まず，各 3 次元モデルの大きさを取得し，その値を基にスケール変換することで 3 次元モデルの大きさをモデル間で統一した. 次に, Local Reference Frame （LRF）を用い，3 次元モデル間の姿勢を統一した.

レンダリングには，Phong の反射モデルを用いた. 多視点画像のレンダリングは，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, 正十二面体の各頂点から行った．実験用の映像のレンダリングは，カメラ位置を地面に対して垂直 18° の位置に設定し，水平方向に回転している 3 次元モデルの様子を捉えた．なお本研究では形状の印象のみを対象とし，色・質感の印象は別研究で扱う[4].

4.1.3. 主観評価値の付与

3 次元物体の印象を定量化するため，主観評価実験を行う．主観評価実験は，先行研究[2]の知見に基づき，SD 法を用いて行った．評価語には，3 つの形容詞対,”柔らかい-硬い”,”派手-地味”,”安定-不安定”を用いた．印象評価においては，前述した主観評価実験用の映像を提示し，様々な方向からの印象を総合的に評価してもらった．実験の結果，3 種類の形容詞対における 7 段階（-3~+3）の主観評価値を取得した．次に，主観評価値を 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 次元モデルの印象分布とした．この際，サンプル間で評価者数が異なるため，印象分布を各サンプルの評価者数で正規化した．正規化した印象分布を推定モデルの教師信号とする.

4.2. モデルの訓練

本研究では, モデルの学習と評価に 4.1.3 で示した印象分布を付与した 3 次元モデルデータベース（Car カテゴリ 632 個，Vase カテゴリ 575 個，Chair カテゴリ 985 個）を用いる．モデルの構築は，物体カテゴリと形容詞対の組み合わせの合計 9 条件行った．交差検証は，データセットを train，test，validation（8:1:1）に分割し，9 分割の交差検証を採用した.

次に，モデルに用いた DNN の構造について説明する（表 1）. CNN1 層の構造は AlexNet を用いた. CNN1 層は最適化の対象である CNN 内の重みを各視点で共有している. View-pooling 層は CNN1 層から出力される各視点の画像特徴量を 1 次元に平滑化し，行方向に結合した後，1 列ずつ最大値を持つ視点の値を抽出する．すなわち，View-Pooling 層は印象推定に有効な視点を選択する役割を担っていると言える．出力層の活性化関数は softmax 関数を使用した. 学習の最適化アルゴリズムは Adam を用いた. また, DNN において懸念される勾配消失問題を回避するため, 活性化関数は Rectified Linear unit を用いた．損失関数はクロスエントロピー誤差を用いた．なお，学習率は 0.001，エポック数は Car・Vase カテゴリを 300，Chair カテゴリを 200 とした．CNN1 のドロップアウト率は 0.5 とした.

表 1 DNN アーキテクチャの詳細

Layer		Input size	Output size	kenel	stride
CNN1	conv1	227 × 227 × 3	55 × 55 × 96	11 × 11	4
	maxpool1	55 × 55 × 96	27 × 27 × 96	3 × 3	2
	conv2	27 × 27 × 96	27 × 27 × 256	5 × 5	1
	maxpool2	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
	maxpool3	13 × 13 × 256	6 × 6 × 256	3 × 3	2
View pooling		20 × 9216	9216	-	-
CNN2	FC1	9216	4608	-	-
	FC2	4608	4608	-	-
	FC3	4608	7	-	-

5. 結果

本研究では提案手法の有効性を検証するため，複数の比較手法を用いた検証実験を行った．比較手法は，Voxel ベースの手法である 3D ShapeNets と，提案手法に単視点の画像を入力する Single-view CNN（以下, SVCNN）を用いた. 推定精度の評価指標は，相関係数と平均二乗誤差を用いた．相関係数は，人の印象分布及び推定結果（推定印象分布）それぞれを期待値に変換して算出した．平均二乗誤差は，人の印象価分布と推定印象分布における各クラスの誤差の総和から算出した.

5.1. 総合評価

各検証で得られた相関係数及び平均二乗誤差の平均を表 2 に示す．表 2 より，提案手法は 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．このことから，提案手法の実用上の有効性を確認した.

次に，比較手法に対する提案手法の優位性を確認する. 提案手法は 3D ShapeNets と SVCNN の両モデルと比較し推定精度が向上したことを確認した．特に，Car カテゴリの”安定-不安定”条件や，Vase カテゴリの”柔らかい-硬い”条件では，相関係数が SNCNN よりも 0.1 程度向上した．次節以降では，

手法間の比較に関する考察と物体カテゴリ間の比較に関する考察を示す．

5.2. 提案手法と各比較手法との比較

手法間の精度比較においては，提案手法（MVCNN）をコントロール群，各比較手法を処理群とした

表 2 各手法における相関係数と平均二乗誤差

Car

手法	入力信号	柔らかい-硬い		派手-地味		安定-不安定	
指標	-	LCC	MSE	LCC	MSE	LCC	MSE
3D ShapeNets	ボクセル	0.42	0.07	0.62	0.09	0.01	0.08
SVCNN	単一画像	0.50	0.05	0.69	0.06	0.23	0.05
MVCNN（提案手法）	多視点画像	0.56	0.05	0.73	0.06	0.26	0.05

Vase

手法	入力信号	柔らかい-硬い		派手-地味		安定-不安定	
指標	-	LCC	MSE	LCC	MSE	LCC	MSE
3D ShapeNets	ボクセル	0.12	0.08	0.18	0.09	0.59	0.12
SVCNN	単一画像	0.27	0.05	0.49	0.05	0.60	0.06
MVCNN（提案手法）	多視点画像	0.42	0.04	0.58	0.04	0.73	0.05

Chair

手法	入力信号	柔らかい-硬い		派手-地味		安定-不安定	
指標	-	LCC	MSE	LCC	MSE	LCC	MSE
3D ShapeNets	ボクセル	0.42	0.08	0.38	0.09	0.44	0.09
SVCNN	単一画像	0.56	0.05	0.58	0.06	0.64	0.05
MVCNN（提案手法）	多視点画像	0.59	0.05	0.59	0.05	0.66	0.05

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(μc：コントロール群の相関係数又は平均二乗誤差の平均値，μi：処理群の相関係数又は平均二乗誤差の平均値)．

まず，MVCNN と 3D ShapeNets の比較を行う．提案手法（MVCNN）と 3D ShapeNets の有意差検定の結果は，両指標とも全 9 条件中 9 条件で有意差($p < 0.01$)が認められた．これは，MVCNN の入力信号が 3D ShapeNets の入力信号よりも高解像度であることに起因すると考えられる．一方で，3D ShapeNets は 3 次元物体認識タスクにおいては約 77%の高い認識率を示す手法であった．この手法を印象推定タスクに適応した場合，その推定精度は著しく低下した．このことから，印象推定タスクには 3 次元物体認識タスクよりも物体の形状をより高解像度で表現する必要があることが示唆された．

次に，MVCNN と SVCNN の比較を行う．提案手法（MVCNN）と SVCNN の有意差検定の結果は，評価指標で異なっていた．相関係数では 5 条件で有意差($p < 0.05$)が認められた．これに対し，平均二乗誤差では全 9 条件で有意差が認められなかった．これらの結果は 2 つの評価指標の性質の違いに起因していると考えられる．本研究では相関係数を離確率分布の期待値から算出している．期待値は，印象分布における各クラスの所属確率で重み付けされるため，平均二乗誤差に比べて，分布の大局的な傾向を評価することに適している．提案手法が相関係数でのみ有意差を示したことは，提案手法が印象分布の大局的な傾向，つまり人の印象評価のばらつきの傾向を SVCNN より良く捉えていると言える．すなわち，1 視点からの形状の情報より多視点からの形状の情報を用いる方が，人の印象評価のばらつきの傾向を捉えられることが示唆された．

5.3. カテゴリごとの印象推定値の比較

3 つの物体カテゴリに着目する．いずれの物体カテゴリにおいても，提案手法の推定精度は SVCNN より向上していることが確認できる．しかしながら，物体カテゴリの形状に関連する固有の性質・構造よって，MVCNN を用いる効果が強い条件と弱い条件が確認できる．具体的には，Car カテゴリと Vase カテゴリは MVCNN を用いることで推定精度が向上する傾向にあり，Chair カテゴリは推定精度が変化しない傾向が確認できる．これらのことから，提案手法は，特に視点によって形状の見えや印象が変化する物体カテゴリ条件では SVCNN より優位性を発揮することが確認できる．

6. 印象推定モデルの検証

提案手法で構築した印象推定モデルの有効性を検証するため，モデルの推定結果と 3 次元物体の形状との関係性を視覚的に確認する．

6.1. 結果

検証用データを対象に推定評価値を付与し，これを期待値に変換する．なお，推定には物体カテゴリと評価語の組み合せごとに，交差検証の際最も精度が高かったモデルを採用した．物体カテゴリの評価語ごとに，期待値を降順で並び替え，上位 4 サンプルと下位 4 サンプルを明らかにした．その結果を図 2 に示す．

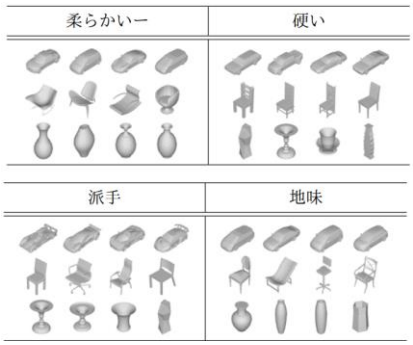


図 2 推定評価値の期待値が上位・下位 4 サンプル

6.2. 考察

図 2 の結果から，構築した印象推定モデル Car カテゴリの“柔らかい-硬い”条件を主に角・直線・曲線・曲面等の形状，“派手-地味”条件を主に面の重なり・エッジ等の形状，“安定-不安定”条件を主に高さ・長さ等の形状によって識別していることを確認した．次に，Vase カテゴリの“柔らかい一硬い”条件を主に角・直線・曲線・曲面等の形状，“派手―地味”条件を主に面の重なり・エッジ等の形状によって識別

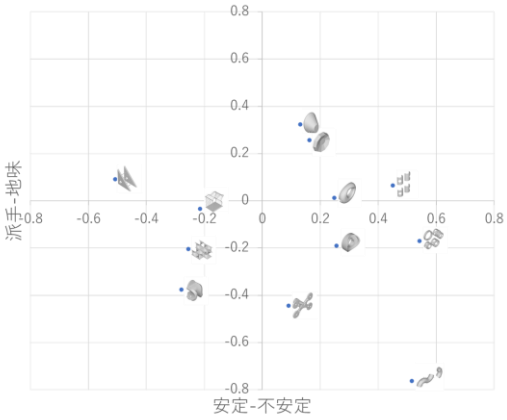
していることを確認した．最後に，Chair カテゴリの“柔らかい-硬い”条件を主に角・直線・曲線・曲面等の形状，“安定-不安定”条件を底面の形状によって識別していることを確認した．これらのことから，“柔らかい-硬い”条件は角・直線・曲線・曲面といった形状から識別されており，”派手-地味”条件は面の重なりやエッジ等の形状によって識別されている．これらの印象の評価基準が物体カテゴリ間で共通していることから，提案手法は評価語に関連する主たる物理的な特徴を識別可能であることが示唆された．これらの物理的な特徴と印象の関係性の解釈に関しては，３次元物体に対する人の印象評価の構造の定量化に関する研究[3]でも報告されており，我々の主張の妥当性を確認した．

7. 別ドメインの３次元形状への適用

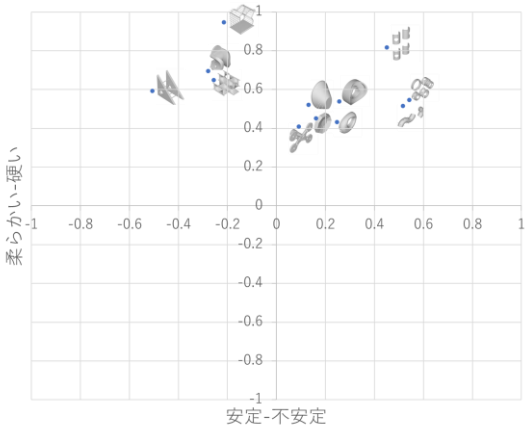
本節では学習済みの印象推定モデルを，未学習の別ドメイン（カテゴリ）の物体に適用し，印象の推定を試みる．ここではモデル構築に使用していないドメインである子供向け遊具の 3D モデル（12 個）をデータセットとして使用する．4 章で構築したモデルを使用し印象の推定を行った．評価語対，”柔らかい-硬い”，”派手-地味”，”安定-不安定”について-３点~３点の印象分布を予測し，その後期待値を算出した．印象推定の結果を以下の図 3 に示す．“柔らかい-硬い”ではエッジがないものを柔らかい，エッジのあるものを硬いと評価する傾向がある．“派手-地味”では，パーツが分散しているモデルを派手、パーツがまとまっているモデルを地味であると評価する傾向がある．“安定-不安定”では直線的な形状を安定、曲線的な形状が不安定であると出力する傾向がある．これらの結果から，別ドメインの 3D 形状においても，6 章で示された印象評価の傾向と同様の傾向がみられた．

8. まとめ

本研究では Multi-view CNN を用いた 3 次元物体の印象推定手法を提案した．推定モデルの学習と評価には，独自に作成した主観評価データを付与した 3 次元モデルデータを用いた．検証実験では，3 種類の物体カテゴリにおける 3 つの形容詞対の主観評価分布を推定した．提案手法の推定結果と人の主観評価分布は 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，提案手法は比較手法よりも推定精度が高く，人の評価のばらつきの傾向を捉えていることを確認した．今後の課題には別ドメインへの適用に関する検討等があげられる．



(a) “派手-地味”，”安定-不安定”



(b) “柔らかい-硬い”，”安定-不安定”

図 3 印象推定結果

謝　辞

本研究の一部は JST, COI JPMJCE1314 および COI-NEXT JPMJPF2111 の支援によって行われた．

文　献

[1] K. Sakashita, K. Tobitani, K. Taguchi, M. Hashimoto, I. Tani, S. Hashimoto, K. Katahira, and N. Nagata, “Impression estimation model of 3D objects using multi-view convolutional neural network,” IW-FCV2023, CCIS, vol.1578, pp.343-355, Springer, Cham (2022)

[2] 片平建史，武藤和仁，李奈栄，飛谷謙介，白岩史，中島加惠，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, pp.563‐570 (2016)

[3] S. Miyai, K. Katahira, M. Sugimoto, N. Nagata, K. Nikata, and K. Kawasaki, “Hierarchical structuring of the impressions of 3d shapes targeting for art and non-art university students,” HCII 2019, CCIS, vol.1032, pp.385–393, Springer, Cham (2019)

[4] Y. Sugiyama, N. Sunda, K. Tobitani, and N. Nagata, “Texture synthesis based on aesthetic texture perception using CNN style and content features,” IW-FCV2023, CCIS, vol.1857, pp.107-121, Springer, Singapore (2023)