

卒業研究論文

題目
深度を利用した2.5Dレリーフ
による視覚障害者のための触覚写真の
表現力拡張

知能メディア処理研究グループ

指導教員 岩村 雅一 教授

令和 7 年 (2025 年) 度 卒業

(No. AJG22102) 高石康世

大阪公立大学工学部情報工学科

目次

第1章 緒論	1
第2章 関連研究	5
2.1 写真の触図化技術	5
2.2 単一写真からの2.5Dレリーフ生成	6
2.3 触覚地図における2Dと3D表現の比較	6
第3章 深度と輝度に基づく2.5Dレリーフの表現拡張	9
3.1 手法	10
3.1.1 ベースライン手法: 深度情報に基づく2.5Dレリーフ (DR)	10
3.1.2 提案手法: 深度と輝度に基づく2.5Dレリーフ (DLR)	10
3.1.3 ベースライン手法: 触図 (TG)	11
3.2 ユーザスタディ	12
3.2.1 実験参加者	12
3.2.2 実験設計	12
3.3 結果	15
3.3.1 事前インタビュー	15
3.3.2 触覚モデルの評価	16
3.3.3 事後インタビュー	17
3.4 小括	19
第4章 深度情報に基づく前景除去による触察体験の向上	25
4.1 提案手法: 前景メッシュ除去	26
4.2 予備実験	27
4.2.1 実験条件	28
4.2.2 評価方法	28

4.2.3	結果 1: 生成された 2.5D レリーフの視覚的評価	28
4.2.4	結果 2: 生成された 2.5D レリーフの触覚的評価	29
4.3	小括	31
第 5 章	議論	33
5.1	視覚経験と情報取得方略	33
5.2	視覚的概念の触覚的翻訳における課題	33
5.2.1	色・見切れ・構図	34
5.2.2	遠近法の理解	34
5.3	写真鑑賞の目的と適応的なデザイン	35
5.4	本研究の限界	36
第 6 章	結言	37
	謝辞	39
	参考文献	41

目次

1.1	触覚モデルの例. 左側に元画像, 右側にそれぞれの触図 ((a) および (b)) および 2.5D レリーフ (c) を示す.	4
3.1	シマウマの画像 (左) と生成された 2.5D レリーフ (右). 深度情報のみを用いた 2.5D レリーフではシマウマの様子が触察できない.	9
3.2	提案手法の全体フロー. 入力画像から深度マップを推定し, 輝度値マップを重畳したのち, 2.5D レリーフを生成する.	11
3.3	ユーザスタディの様子	13
3.4	ユーザスタディでを使用した画像とモデル (1 / 3). 左から元画像, 触図 (TG), 深度に基づく 2.5D レリーフ (DR), 深度と輝度に基づく 2.5D レリーフ (DLR).	20
3.5	ユーザスタディでを使用した画像とモデル (2 / 3). 左から元画像, 触図 (TG), 深度に基づく 2.5D レリーフ (DR), 深度と輝度に基づく 2.5D レリーフ (DLR).	21
3.6	ユーザスタディでを使用した画像とモデル (3 / 3). 左から元画像, 触図 (TG), 深度に基づく 2.5D レリーフ (DR), 深度と輝度に基づく 2.5D レリーフ (DLR).	22
3.7	各触覚モデルに割り当てられた順位の箱ひげ図. 矢印はウィルコクソン符号付き順位検定に従って統計的に有意な差のあるペアを表す. 赤い点は各箱ひげ図の中央値を示す. <i>TG</i> , <i>DR</i> , <i>DLR</i> はそれぞれ触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する.	22

3.8	各被験者が各触覚モデルに割り当てた順位の箱ひげ図. 矢印はウィルコクソン符号付き順位検定に従って統計的に有意な差のあるペアを表す. 赤い点は各箱ひげ図の中央値を表す. <i>TG</i> , <i>DR</i> , <i>DLR</i> はそれぞれ触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する.	23
3.9	写真毎に各触覚モデルに割り当てられた順位の箱ひげ図. 矢印はウィルコクソン符号付き順位検定に従って統計的に有意な差のあるペアを表す. 赤い点は各箱ひげ図の中央値を示す. <i>TG</i> , <i>DR</i> , <i>DLR</i> はそれぞれ触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する.	24
4.1	前景が触察を妨げる例. ネコの画像と生成された 2.5D レリーフには前景のメッシュが存在し, 触察を妨げる.	25
4.2	提案手法のパイプラインの概要と生成例	26
4.3	実験で使用した画像および生成された 2.5D レリーフと, 各画像の前景除去後の画像および生成された 2.5D レリーフ.	27
4.4	予備実験の様子	30
5.1	遠近法を用いた写真の例とその 2.5D レリーフ.	34

表 目 次

3.1	参加者の詳細な属性情報. BLは全盲, LVは弱視を示す. 視覚障害の欄の数値は視力を表す. 「先天盲」は出生時または3歳以内の失明, 「視覚経験」は視覚的な記憶の有無を示す.	14
3.2	一対比較法の結果: <i>VD</i> , <i>TG</i> , <i>DR</i> , <i>DLR</i> はそれぞれ言語説明, 触図, 深度に基づく2.5Dレリーフ, 深度と輝度に基づく2.5Dレリーフを意味する. CIは一貫性指標を示す.	17
3.3	各基準に対する一対比較法の結果: <i>VD</i> , <i>TG</i> , <i>DR</i> , <i>DLR</i> はそれぞれ言語説明, 触図, 深度に基づく2.5Dレリーフ, 深度と輝度に基づく2.5Dレリーフを意味する. CIは一貫性指標を示す.	17
3.4	決定された基準の重み. CIは一貫性指標を示す.	17
3.5	AHPの結果(全体の優先度値): <i>VD</i> , <i>TG</i> , <i>DR</i> , <i>DLR</i> はそれぞれ言語説明, 触図, 深度に基づく2.5Dレリーフ, 深度と輝度に基づく2.5Dレリーフを意味する. CIは一貫性指標を示す.	17

第1章 緒論

写真家 Destin Sparks は, “Photography is the story I fail to put into words” (「写真は私が言葉にできなかった残りの物語だ」) と述べている [4]. この言葉は, 写真が言語を用いずに意味や感情を伝える, ユニークかつ強力な表現手段であることを端的に示している. 写真は単なる視覚的記録にとどまらず, 人が出来事や感情を共有するための重要な媒体として機能する. 私たちの日常生活においても, 写真は情報を伝え, 思い出を残し, 自己を表現し, 他者と結び付くための欠かせない手段である. さらに, 写真は時間の経過とともに記憶を呼び起こす手がかりとなり, 懐かしさを喚起するだけでなく, 個人が過去の出来事を思い返しながらか自分の人生を振り返る助けとなる. また, デジタル化の進展によって, 写真の社会的な役割はさらに大きなものとなっている. SNS やオンラインの交流の場を通じて写真を共有することは, 自己を表現し, 他者との関わりを築く上で重要になりつつある. 人々は写真を介して自分の経験を他者と共有し, 地理的な距離を越えて人とのつながりを広げ, 関係を深める. しかし, このように写真が重要な役割を果たす一方で, 写真が本来持つ豊かな情報を理解することは, 視覚障害者にとって容易ではない. 視覚障害者が写真にアクセスするための支援技術として, 近年では, Be My AI [5] に代表される, 生成 AI を用いて写真の内容を言葉で説明する技術が登場しているが, 十分とは言えない. はじめに触れた Destin Sparks の言葉の通り, 写真を言葉によって完全に説明することには難しい部分がある. 例えば, 写真の中に含まれる質感や空間の奥行き, 光の表情などの繊細な視覚情報を, 言葉による説明だけで十分に伝えることは難しい. さらに, 言葉での描写は情報を単純化してしまったり, 説明者の主観が入り込んだりする可能性があり, 視覚障害者が写真を深く理解し, 楽しむ上での妨げとなることがある.

言葉による説明だけで写真を理解することが難しいのであれば, それを触覚的な情報で補うことは自然な発想である. その代表的な方法として図 1.1(a) や図 1.1(b) のような「触図 (tactile graphics, TG)」が挙げられる¹. これは, 物体の輪郭など輝度の変化が大

¹本研究における触図 (TG) とは, 北米点字委員会 (BANA) の指針に定義された触図に必ずしも準拠しないものを含み, 触察のために表面の一部を特殊紙などを用いて離散的に隆起させた写真を指す.

きい領域を紙面上で隆起させることにより、触覚的に形状を認識できるようにしたものである [1,2,6-9]. しかし, このような紙状の触覚媒体は平面上で一定の高さしか表現できないため, 表現力に限界がある. その結果, これまでの研究では, 視覚情報を単純化し, 情報量を意図的に減らす方向に焦点が置かれてきた [1,2,8,9]. 実際, 北米点字委員会 (BANA) の指針 [10] や Edman の代表的著書「Tactile Graphics」 [11] においても, 情報量の多い写真は触覚表現の実用的範囲を超えると指摘されている.

そこで, TG に代わる手法として, 本研究では図 1.1(c) のような 2.5D レリーフに注目する. 2.5D レリーフとは, 各領域の高さを奥行きに応じて変化させたものであり, 観察者に近い部分を高く, 遠い部分を低く表現する. 例えば, 米国 25 セント硬貨に刻まれたジョージ・ワシントンの浮き彫りがその例である. この手法は, 二次元の TG と三次元モデルの中間に位置する表現形式である [12]. TG が形状を一定の高さの輪郭線などで表すのに対し, 2.5D レリーフは高さの連続的な変化によって空間的な奥行きや細部を表現できる [13]. 一方で, 完全な三次元復元モデルのように見えない部分を推定することはなく, 元の画像が持つ視覚的な制約を保ったまま立体化できる点が特徴である. これまで, 2.5D レリーフは視覚障害者による美術作品の理解支援などに利用されてきたが [3], 従来の制作方法は多大な労力と費用がかかるため, その利用は限られていた. また, 陰影の手がかりに基づく半自動生成手法 [14] も提案されているが, 照明条件が一定でない写真に対しては精度が低く, 信頼性に欠けるという課題がある. このため, 写真が社会的・個人的なコミュニケーションにおいて重要性を増しているにもかかわらず, 2.5D レリーフの活用は依然として限定的である.

近年のコンピュータビジョン技術の進展により, 写真から精度の高い深度マップを生成することが可能になった [15]. 本研究では, これを応用することで, 一枚の写真から 2.5D レリーフを作成できることに着目した. この応用によって, 従来に比べて 2.5D レリーフの制作の手間やコストが大幅に削減され, 視覚障害者が自分の写真に日常的に触れて楽しむことが, より現実的になりつつあると考えられる. しかしながら, 単に深度情報を高さに変換するだけでは, 写真鑑賞における実用上の課題が残る. 本研究では, 特に以下の 2 つの課題に着目し, それぞれの解決手法を提案する.

第一の課題は, 表面情報の欠落である. 深度情報のみに基づくレリーフ (Depth-based 2.5D Relief: DR) では, 物体の形状は再現されるものの, 文字, 模様, テクスチャといった深度差を持たない表面の情報は凹凸として表現されず, 触覚的に把握できない. 例え

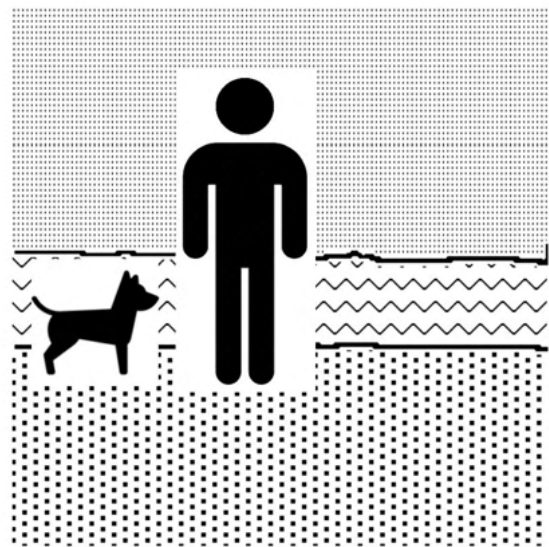
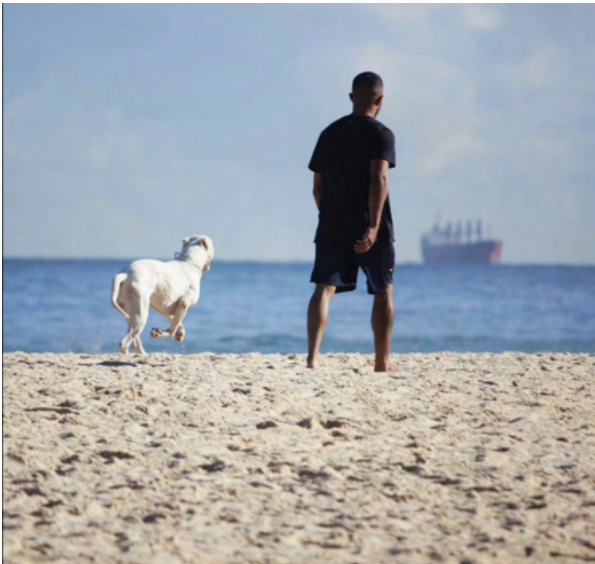
ば、服の柄や看板の文字などは、視覚的には重要であっても深度情報には反映されない。そこで本論文では、深度情報に加えて画像の輝度情報を統合することで、形状だけでなく表面の視覚的特徴も触覚的に提示可能な深度・輝度に基づいた2.5Dレリーフ（Depth and Luminance-based 2.5D Relief: DLR）の生成手法を3章で提案する。

第二の課題は、遮蔽物による触察の阻害である。写真の全体にわたって存在する前景の構造物（例えば動物を囲う檻など）は、被写体の触察を物理的に阻害し、対象の全体像を把握する妨げとなる。視覚的には透かして見るのが可能であっても、触覚的には手前の物体が邪魔をして奥に触れることができないという問題がある。そこで本論文では、深度ヒストグラムを用いて画像を層状に分解し、前景を除去した上で背景を補完する編集手法を4章で提案する。

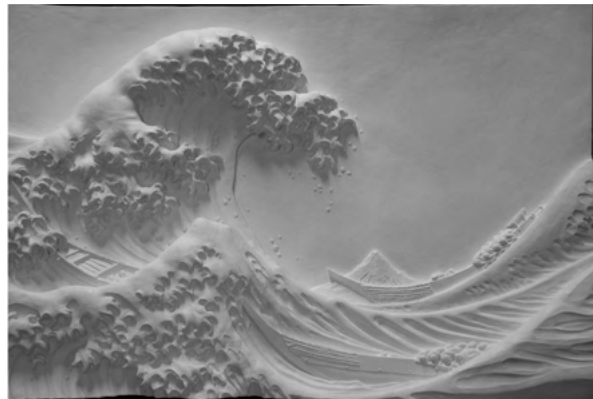
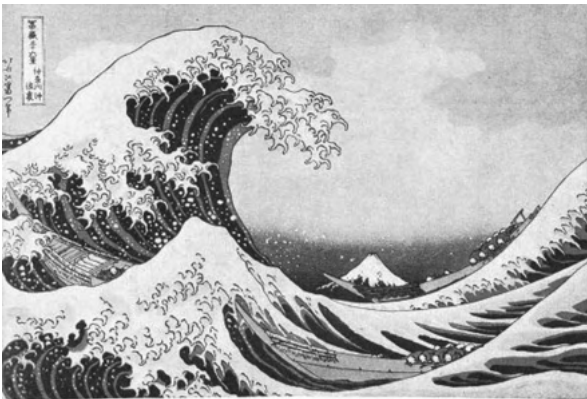
本論文の構成は以下の通りである。第2章では、触図生成や2.5Dレリーフ生成に関する関連研究について述べる。第3章では、深度と輝度を統合した表現拡張手法について詳述し、視覚障害者を対象とした評価実験の結果を示す。第4章では、触察を阻害する前景物体の除去手法について提案し、その有効性を予備実験によって検証した結果を示す。第5章では、これら2つのアプローチから得られた知見を統合し、視覚障害者が写真の空間的・視覚的情報をどのように受容・解釈するかについて、視覚経験の有無や認知的負荷の観点から考察を行う。最後に第6章で本研究の結論を述べる。



(a) Yoon らの手法による写真の触図 [1]



(b) Pakenaite らの手法による写真の触図 [2]



(c) 大内らによる絵画の 2.5D レリーフ [3]

図 1.1: 触覚モデルの例. 左側に元画像, 右側にそれぞれの触図 (a) および (b) および 2.5D レリーフ (c) を示す.

第2章 関連研究

本章では、視覚障害者のための画像提示手法、特に写真の触覚化に関する既存技術とその課題について概観する。はじめに、写真を触図（Tactile Graphics）として表現する従来のアプローチについて述べ、次に単一写真からの2.5Dレリーフ生成技術について整理する。最後に、地図などの分野における2次元（2D）と2.5次元（2.5D）表現の比較研究を参照し、本研究の着想における理論的背景を明らかにする。

2.1 写真の触図化技術

視覚障害者が写真を理解するための最も一般的な手法の一つが、輪郭線や点字パターンを用いた触図である。ここでは、絵画などの芸術作品を対象とした研究を除き、写真表現に特化した技術を中心に述べる。より広範な触図技術のレビューについては、Krufkaら [16] の調査が詳しい。WayとBarner [6,7] は、写真を立体コピー（カプセルペーパー）に出力する際、識別性を高めるための画像処理による単純化手法を提案した。また、FerroとPawluk [8] は写真から触図を自動生成する手法を提案したが、視覚障害当事者による評価は行われていない。より高度なアプローチとして、画像の顕著性（Saliency）に着目した研究がある。Yoonら [1] や Abdusalomovら [9] は、顕著性マップを用いて主要な被写体（前景）を分割し、その境界線や内部エッジを抽出する手法を提案した。しかし、これらの手法もユーザスタディによる有効性の検証は不十分である。Pakenaiteら [2] は、写真内の意味的に重要な前景物体を単純なアイコンなどの「触覚的な単語（tactile words）」として置き換えて表現し、背景領域を輪郭線ではなくテクスチャパターンで埋めるというユニークなアプローチを提案している。これらの既存手法に共通しているのは、「写真をそのまま触覚化すると複雑すぎて理解できない」という前提に立ち、輪郭線の抽出や情報の単純化を行っている点である。しかし、この単純化プロセスにおいて、写真が本来持っている奥行き情報や微妙な質感は失われてしまうという課題がある。

2.2 単一写真からの2.5Dレリーフ生成

写真の奥行き情報を保持するためのアプローチとして、2.5Dレリーフ（半立体）の生成技術がある。従来、単一写真から2.5Dレリーフを作成する主流の手法は、Shape from Shading (SFS) [17]に基づいている。SFSは、画像の陰影（シェーディング）を手がかりにして物体の3次元形状を推定する手法である。Wuら[18]やToら[19]は、SFSを用いて単一画像から自動的に2.5Dレリーフを生成する手法を提案したが、その適用範囲は人間の顔画像に限定されており、一般的な写真への応用については検証されていない。また、これらは視覚障害者の支援を目的としたものではない。視覚障害者支援の文脈では、Governiら[20]やFurferiら[14]の研究がある。Governiらは多段階のSFSを用いて画像から立体的な特徴を持つレリーフを生成する手法を開発し、Furferiらはこれを芸術作品に応用して視覚障害者の芸術鑑賞を支援する半自動システムを構築した。しかし、これらの手法は特定の芸術作品や単純な画像を対象としており、日常的なスナップ写真を処理することは想定されていない。SFSに基づく手法の最大の限界は、陰影を形状の手がかりとするため、元の写真の照明条件に品質が大きく依存することである。例えば、フレーム外の物体が落とす影や、遠くの背景に落ちる影などは正しく解釈できず、不自然な形状推定につながる。日常的に撮影される写真は照明条件が多様であり、SFSによるアプローチには限界がある。

これに対し、本研究では深層学習に基づく単眼深度推定技術を採用する。このアプローチは、照明や影の影響を受けにくく、一般的な写真からでもロバストに3次元形状を復元できる。なお、深層学習を用いた深度推定により写真からレリーフを再構成する研究としてはCuiら[21]のものがあるが、これは「既存のレリーフの写真を撮って、元のレリーフを復元する」という特殊なタスクを扱っており、視覚障害者に対する写真のアクセシビリティ向上のために、一般的な写真に対して深度推定を利用してレリーフを作成することを目的とする本研究とは異なる。

2.3 触覚地図における2Dと3D表現の比較

本節では、写真以外の領域、特に触覚地図における2D表現と3D（2.5D）表現の比較研究について触れる。これらの知見は、写真の触覚化においても重要な示唆を与える。Guel

ら [22, 23] は、触覚地図上のシンボル（点，線，領域）について 2D と 3D の効果を比較し、3 次元的なボリュームを持つシンボルが、2D の形状デザインと組み合わせられることで、ユーザの探索効率を向上させることを明らかにした。また、Holloway ら [24] は、3D プリントされた地図モデルが従来の 2D 触覚地図に比べて、情報の記憶保持や相対的な高さ関係の理解において優れていることを示している。これらの研究は、触覚提示において「高さ情報（3 次元性）」が加わることで、空間的な理解や情報の区別を容易にするという本研究の仮説を支持するものである。本研究では、この知見を写真という複雑なメディアに応用し、深度情報を高さとして物理的にマッピングすることの有効性を検証する。

第3章 深度と輝度に基づく2.5Dレリーフの表現拡張

単純に深度を用いた2.5Dレリーフの作成では、模様や文字といった視覚情報が失われてしまうという課題がある。例えば図3.1の深度情報に基づくレリーフ（DR）では、シマウマの模様は凹凸として表現されておらず、触知できない。この問題に取り組むため、我々は画像の輝度値に注目した。輝度情報は物体の表面特性やテクスチャを反映しており、これを深度情報に加えることで深度情報のみでは十分に表現できない表面の特徴を触覚的に提示可能になると考えた。そこで本章では、深度推定に基づく形状情報に、画像の輝度情報を統合することで表現力を拡張した深度と輝度に基づく2.5Dレリーフ（Depth and Luminance-based 2.5D Relief: DLR）を提案する。また、この提案手法について、従来写真理解に用いられてきた触図（Tactile Graphics: TG）や、単に深度情報のみに基づく2.5Dレリーフ（Depth-based 2.5D Relief: DR）と比較し、その有効性を評価する。

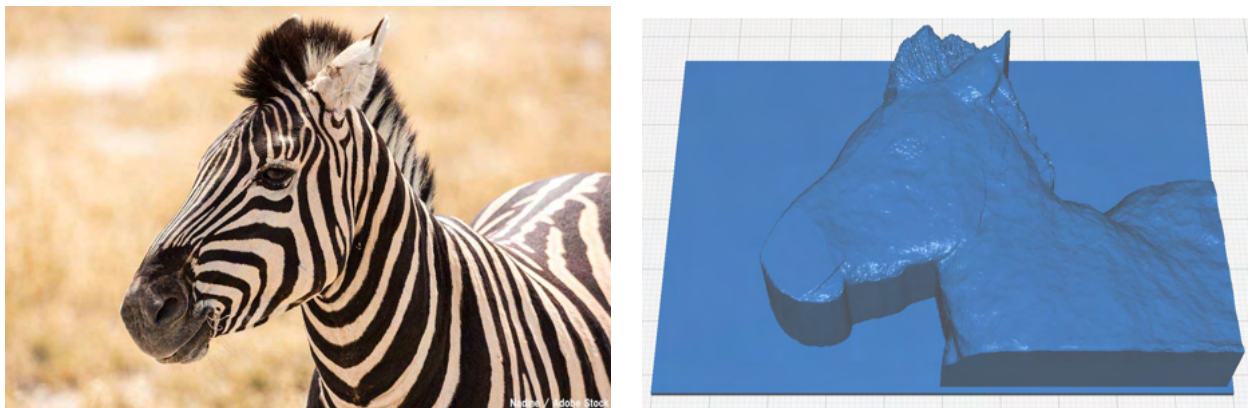


図3.1: シマウマの画像（左）と生成された2.5Dレリーフ（右）。深度情報のみを用いた2.5Dレリーフではシマウマの模様が触察できない。

3.1 手法

本節では、本研究で比較・検証を行う3種類のモデル (DR, DLR, TG) の生成プロセスについて詳述する。

3.1.1 ベースライン手法: 深度情報に基づく 2.5D レリーフ (DR)

本手法は、深度情報のみを用いて形状を生成するものであり、以下の手順で処理される。

(1) **ステップ 1: Marigold による深度推定** 入力画像に対して, Marigold [15]¹ を適用し, 深度マップ D を生成する. ここで, 各画素の値はシーン内におけるカメラからの推定距離を表している.

(2) **ステップ 2: 2.5D レリーフの生成** 得られた深度マップ D を, 指定された深度範囲 $[D_{\text{near}}, D_{\text{far}}]$ に収めるよう調整する. この範囲外の値を持つ画素については, 近い方の境界値にクリッピング (固定) を行う. この際, クリッピングされた領域 (深度範囲外の領域) をマスク領域 M として定義する. 調整後の深度マップは, Min-Max スケーリングを用いて $[0, 1]$ の範囲に正規化される. その後, ノイズを低減するためにメディアンフィルタを適用したものを D_{DR} とする. 最後に, D_{DR} に基づいて面を構成し, 2.5D レリーフモデルを生成する.

3.1.2 提案手法: 深度と輝度に基づく 2.5D レリーフ (DLR)

提案手法では, DR で得られた深度情報 D_{DR} に対し, 輝度情報を統合することで表現力を拡張する.

(1) **ステップ 1: 輝度情報の統合** 元の RGB 画像をグレースケール変換して輝度値 L を算出し, これを $[0, 1]$ の範囲に正規化したものを L_{norm} とする. マスクされていない画素 (対象領域) にのみ輝度情報を重畳するため, 以降の計算では $L_{\text{norm}} \cdot M$ (ここで M は前

¹<https://marigoldmonodepth.github.io/>

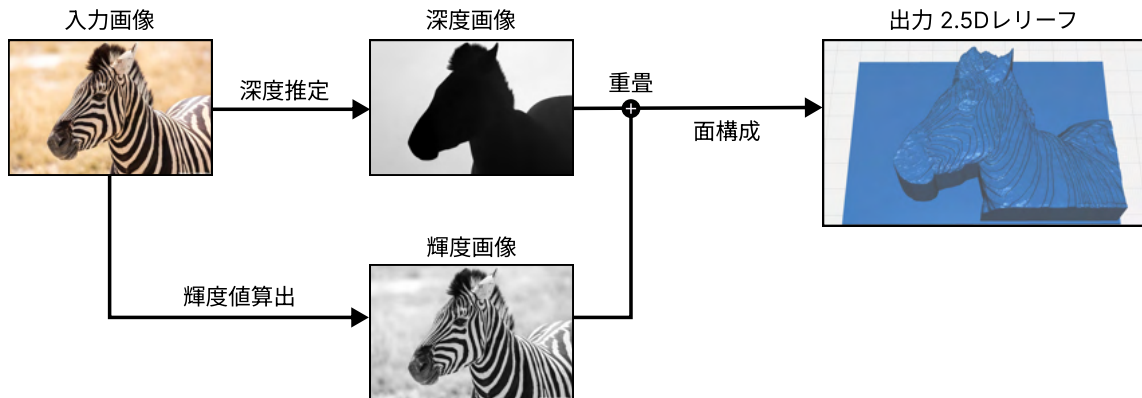


図 3.2: 提案手法の全体フロー. 入力画像から深度マップを推定し, 輝度値マップを重畳したのち, 2.5D レリーフを生成する.

節で定義したマスク領域)を用いる. 最終的な各画素の高さ D_{DLR} は, D_{DR} と $L_{\text{norm}} \cdot M$ の線形結合として以下の式により算出される.

$$D_{\text{DLR}} = (1 - \alpha)D_{\text{DR}} + \alpha(L_{\text{norm}} \cdot M) \quad (3.1)$$

ここで, 重み係数 α は輝度情報の寄与度を制御するパラメータである. この処理により, 輝度が高い (明るい) 領域は隆起し, 低い (暗い) 領域は沈み込むことになり, 視覚的なコントラストが触覚的な凹凸へと変換される.

(2) **ステップ 2: 2.5D レリーフの生成** 統合後のマップに対してメディアンフィルタを適用し, 局所的なノイズや過剰な表面のアーティファクトを低減する. その後, マップを再度 $[0, 1]$ の範囲に正規化し, 最終的な 2.5D レリーフを生成する.

3.1.3 ベースライン手法: 触図 (TG)

従来の触図との比較を行うため, 輪郭線に基づくモデル (TG) を作成する.

(1) **ステップ 1: 輪郭検出** Python の Pillow ライブラリを用いて画像の輪郭を検出し, グレースケールの輪郭画像 (画素値 0–255) を生成する. 得られた輪郭画像は, 閾値 T を

用いて二値化される。

(2) **ステップ2: マスキングと最終生成** 公平な比較を行うため、前述の手法 (DR, DLR) と同様に、マスク領域 M 内の画素値を 0 に設定する。得られた線画データを用いて、最終的な触図を生成する。

注釈: ガイドラインとの整合性について なお、本研究で使用した TG は、情報の簡略化を重視する北米点字委員会 (BANA) のガイドラインに必ずしも厳密には準拠していない点に留意されたい。ガイドラインに沿った適切な簡略化や触図化には専門的な知識と技術が必要であり、特に個人のが撮影した写真に対してこれらを適用した触図を作成することは困難である。したがって本研究では、専門知識を必要とせず、日常的に写真を触って楽しむための自動化された手法として、上記のアプローチを採用した。

3.2 ユーザスタディ

提案手法の有効性を検証するため、視覚障害者を対象としたユーザスタディを実施した。本実験では、DLR (提案手法)、DR (深度のみ)、TG (触図) の3種類のモデルを比較し、理解のしやすさや選好傾向を定量的および定性的に評価した。

3.2.1 実験参加者

実験には12名の視覚障害者が参加した (男性7名、女性5名)。平均年齢は49.8歳 (25歳~68歳) である。障害の内訳は、全盲が6名、弱視が6名であった。全盲の参加者のうち、先天盲 (あるいは3歳以前の失明) は6名中4名であり、残る2名は視覚経験を持つ中途失明者であった。詳細は表 3.1 に示す。

3.2.2 実験設計

本実験は、実験刺激の選定と作成、および評価タスクの実施から構成される。



図 3.3: ユーザスタディの様子

まず実験刺激として、図 3.4 から 3.6 に示す 10 枚の写真を使用した。写真は「単一かつ明確な被写体（壺，書籍，時計）」、「単一だが構造が複雑な被写体（シマウマ，トラ，若い女性，影，老女）」、「複数の複雑な被写体（部屋，壁画）」の 3 基準に基づいてインターネットから収集された。各写真に対し、前節で述べた 3 手法（TG，DR，DLR）を用いて触覚モデルを作成した。なお、比較の公平性を保つため、被写体と背景の距離が離れている 6 枚の写真については、深度推定結果に基づき手動で閾値（ D_{near} , D_{far} ）を設定し、背

表 3.1: 参加者の詳細な属性情報. BL は全盲, LV は弱視を示す. 視覚障害の欄の数値は視力を表す. 「先天盲」は出生時または3歳以内の失明, 「視覚経験」は視覚的な記憶の有無を示す.

ID	年齢	性別	視覚障害の状況		先天盲	視覚経験
VI-1	35	男性	BL		✓	
VI-2	39	男性	BL			✓
VI-3	35	女性	LV	左: 0, 右: 0.02		✓
VI-4	61	男性	LV	左右: 0.02, 強い視野狭窄		✓
VI-5	52	女性	BL		✓	
VI-6	68	女性	LV	左: 0, 右: 0.02, 強い視野狭窄		✓
VI-7	44	男性	LV	左: 0, 右: 0.04		✓
VI-8	25	男性	BL		✓	
VI-9	55	女性	BL			✓
VI-10	54	女性	LV	両眼 1.0 だが中心視野はほぼ欠損		✓
VI-11	63	男性	BL			✓
VI-12	66	男性	BL			✓

景領域を除去した. これは, 深度情報に基づくレリーフにおいて, 被写体と背景の距離が離れすぎている場合, 被写体内部の凹凸が圧縮されて触知困難になるのを防ぐためである. DLR においては輝度の重み α (通常 0.05 または 0.1) を手動で調整し, TG では二値化の閾値 $T = 235$ を用いた. DR および DLR には 5×5 のメディアンフィルタを適用した. すべての 2.5D レリーフは 3D プリンタで出力し (最大辺長 10cm, 奥行き 2cm), TG は同サイズで PIAF プリンタ [25] を用いて熱を加えると膨らむスウェルペーパーに印刷した. また, 各写真に対し, OpenAI が提供する ChatGPT (GPT-4o モデル) を利用し, 詳細な言語説明を生成した. プロンプトには画像と共に「視覚障害者に写真の内容を説明する専門家として, この写真の内容を文章で説明してください。」という指示を与えた.

実験手順として, 参加者はまず事前インタビューにて触覚経験等の聴取を受けた. このインタビューでは, 参加者の触覚経験や写真理解に関する経験について, 簡単な質問を行った. 次に, 触察タスクを実施した. 本タスクでは, GPT-4o により生成された言語説明 (音声) を聴取した後, ランダムな順序で提示される 3 種類の触覚モデルを自由に触察した. 時間は無制限とし, 必要に応じて言語説明の再聴取を許可した. 触察後, 参加者は「分かりやすさ」に基づき 3 つのモデルに同率を許して順位をつけた. 全 10 枚の評価終了後, 事後インタビューを実施した. ここでは定性的なフィードバックの収集に加え, 階層分析法 (Analytical Hierarchy Process, AHP) および一対比較法を用いた評価を行った.

評価基準は「物体の理解（形状等の正確さ）」、「空間の理解（奥行き等）」、「場面の理解（文脈）」、「利用意欲」、「認知的負荷（労力）」の5項目を設定した。これら五つの基準は、著者によって事前に設定されたものであり、触覚モデルの評価において重要と考えられる要素を含んでいる。AHPにより、各モデルの優先度（重み係数）が算出される。これらの値が大きいほど、その選択肢が参加者に好まれたことを示す。AHPを用いることにより、参加者の主観的な評価を定量化し、モデル間の比較を容易にした。

3.3 結果

本節では、事前インタビュー、触覚モデルの評価（順位付け）、および事後インタビューの結果について述べる。

3.3.1 事前インタビュー

触覚タスクに先立ち、参加者の触覚および写真に関する経験について聴取した。まず、触図の使用頻度について尋ねたところ、66.7%（8名）が「月に1回以下」と回答しており、日常的な利用は限定的であることがわかった。また、写真の触覚経験に関しては、参加者全員（12名）が「ない」と回答した。その背景として、11名は写真を触覚化するためのツールが存在することを知らない、あるいは利用できる環境がないことを挙げている。一方で、日常生活において写真を理解・鑑賞したいという欲求は高く、50%（6名）が「頻繁にある」、33.3%（4名）が「時々ある」と回答した。「頻繁にある」・「時々ある」と回答をした10名の参加者のうち、90%（9名）は現状の手段としてAIや晴眼者による言語説明に依存しており、一部の弱視の参加者は拡大読書器などを併用していた。具体的な利用場面としては、60%（6名）が自身や友人のプライベートな写真を、40%（4名）が詳細情報の確認を要する場面を挙げている。しかし、既存の手段による写真の触覚化には課題も存在する。ある参加者は、ひいきの人物の顔立ちをより詳細に理解するために、触図プリンターのある施設に写真を持ち込んだ経験を語った。しかし、触覚によって理解可能な画像へと変換するには、情報の適切な取舍選択や単純化といった高度な編集作業が不可欠であり、専門知識を持つ人材の不足や手間を理由に、触覚化を断念せざるを得なかったという。触覚の意義について、数名の参加者は他者に依存することなく直感的な理解が得

られ、誤解を低減できる点を強調した。その一方で、言語説明が持つニュアンス、感情、色情報の伝達能力を評価する意見も見られた。触覚的手法と言語的手法の重要性を比較した結果、8名が触察を同等以上に重要であるとし、5名は相互補完的に両方の手法が必要であると述べている。最後に、触察を支援する写真理解ツールの将来的な利用意欲について尋ねたところ、66.7% (8名) が「使ってみたい」と回答し、33.3% (4名) が「コストや使い勝手次第で検討する」と回答した。全体として、現状のアクセス環境は限られているものの、参加者は写真の触察に対して強い関心を示した。

3.3.2 触覚モデルの評価

3種類の触覚モデルに対する順位付けの結果を報告する。同順位の場合は平均順位を割り当てた。順位データは順序尺度であるため、全体の分析には平均値ではなく中央値を用いた。

順位の分布を図3.7に示す。参加者による順位の中央値に基づくと、深度と輝度に基づく2.5Dレリーフ (DLR) が最も好まれ (中央値 = 1)、次いで深度に基づく2.5Dレリーフ (DR, 中央値 = 2)、触図 (TG) が最も低い評価 (中央値 = 3) となった。統計的有意性を確認するためにフリードマン検定を行った結果、 $p < 0.01$ となり、3条件間に有意差があることが示された。続いて、ホルムの補正を用いたウィルコクソン符号付き順位検定による多重比較を行ったところ、すべてのペア間において $p < 0.0033$ となり、1%水準での有意差が確認された。

さらに、参加者ごとおよび画像ごとでの追加分析を行った。箱ひげ図を図3.8と3.9に示す。参加者ごとの分析 (有意水準5%) では、6名がTGよりもDRを有意に好み、9名がTGよりもDLRを有意に好んだ。また、5名はDRよりもDLRを有意に好んだ。全参加者12名中11名が、DLRを他の手法と同等またはそれ以上に高く評価した。画像ごとの分析では、10枚すべての画像においてDLRが最も良い順位中央値を獲得した。特に「部屋」の写真では、参加者全員がこのモデルを1位とした。「トラ」、「若い女性」、「老婆」の写真では、TGと2種類の2.5Dレリーフ (DR, DLR) の間に有意差が認められた。「壁画」、「壺」、「本」では、DLRが他の手法よりも有意に優れていた。「時計」ではDRがTGよりも優れており、「影」ではDLRが好まれた。「シマウマ」については、フリードマン検定では有意傾向が示されたものの、その後のウィルコクソン検定では有意差は確認されな

表 3.2: 一対比較法の結果：VD, TG, DR, DLR はそれぞれ言語説明, 触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する. CI は一貫性指標を示す.

優先度				CI
VD	TG	DR	DLR	
0.588	0.058	0.108	0.247	0.0881

表 3.3: 各基準に対する一対比較法の結果：VD, TG, DR, DLR はそれぞれ言語説明, 触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する. CI は一貫性指標を示す.

基準	優先度				CI
	VD	TG	DR	DLR	
Object Comprehension	0.547	0.088	0.087	0.278	0.0304
Spatial Comprehension	0.592	0.061	0.105	0.243	0.0173
Scene Comprehension	0.592	0.061	0.105	0.243	0.0479
Willingness to Use	0.440	0.061	0.164	0.334	0.0296
Cognitive Load	0.375	0.075	0.208	0.341	0.0263

表 3.4: 決定された基準の重み. CI は一貫性指標を示す.

重み					CI
物体の理解	空間の理解	場面の理解	利用意欲	認知的負荷	
0.367	0.133	0.190	0.172	0.138	0.0135

表 3.5: AHP の結果 (全体の優先度値)：VD, TG, DR, DLR はそれぞれ言語説明, 触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する. CI は一貫性指標を示す.

総合優先度			
VD	TG	DR	DLR
0.438	0.075	0.168	0.319

かった.

3.3.3 事後インタビュー

事後インタビューの結果について報告する. まず, AHP および一対比較の結果を分析した. AHP 分析では, 全参加者の一対比較値の幾何平均を用いて一対比較行列を作成し,

それに基づいて基準の重みや各基準ごとの優先度を算出した。結果を表 3.2 から 3.5 に示す。この際、一対比較における判断の矛盾の程度（論理的整合性）を定量的に評価するため、整合度指数（Consistency Index: CI）を用いた。一般に $CI < 0.1$ であれば判断に整合性があるとみなされるが、本分析ではすべての行列で $CI < 0.1$ となり、十分な整合性が確認された。最後に、各評価基準の重みを算出した後、重み付き線形和として総合スコアを計算した。

一対比較による優先度スコアの順位は以下の通りとなった。

言語説明 > DLR（深度・輝度 2.5D） > DR（深度 2.5D） > TG（触図）

AHP の結果も、「物体の理解」を除くすべての基準においてこの順序と一致した。「物体の理解」においては、DR が TG を上回った。「空間の理解」および「認知的負荷」においては、DLR と言語説明のスコアが近接していた。すべての基準の中で、「物体の理解」が最も高い重みを持っていた。最終的な AHP の総合スコアは、一対比較と同様の順位を反映していた。

また、参加者からは得られた定性的なフィードバックについても報告する。まず、言語説明と触覚モデル（TG, DR, DLR）の役割の違いについてである。言語説明は、写真の雰囲気や詳細な描写、色情報を理解する上で有用であると評価された。対照的に、触覚モデルは、物体の具体的な形状、大きさ、位置関係、そして写真の構図を、言語説明以上に明確に伝える手段として評価された。次に、従来の触図（TG）と 2.5D レリーフ（DR, DLR）の比較において、レリーフは奥行きや空間配置の理解により効果的であると評価された。一方で、TG は微細な模様の識別において優れているという意見や、白地に黒で印刷されているため、弱視の参加者にとっては単色の 3D プリントであるレリーフよりも視覚的に解釈しやすいという意見もあった。提案手法である DLR（深度・輝度統合モデル）に関しては、大多数の参加者が DR（深度のみ）と比較して、追加された輝度情報が形状や空間配置の理解を向上させたと高く評価した。一部の参加者は、3D 構造と輝度テクスチャの組み合わせが個々の物体の区別を明確にすると述べている。しかしながら、詳細度が増したことでモデルが複雑になり解釈が難しくなったという意見や、表面の模様が粗く触り心地が悪い、モデルがかさばるといった課題も少数ながら報告された。

実験後の全体的な印象として、58%（7 名）の参加者が今後も触察によって写真を探索したいという意欲を示し、83%（10 名）が触図よりも 2.5D レリーフを好んだ。輝度情報の利用に関しては、67%（8 名）が形状および空間理解の向上を報告し、58%（7 名）が触

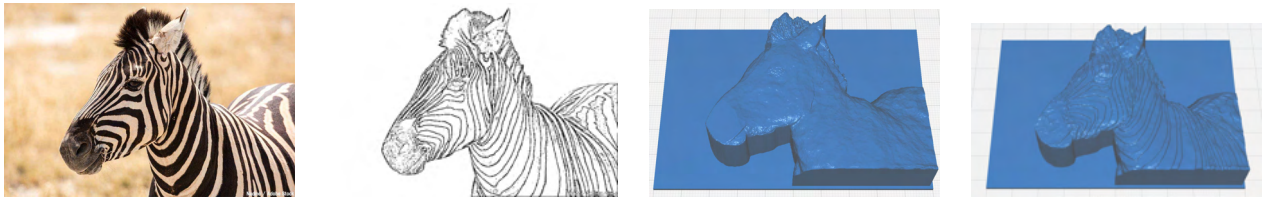
察が容易になったと感じ、83%（10名）が伝達される情報の豊かさが増したと回答した。

3.4 小括

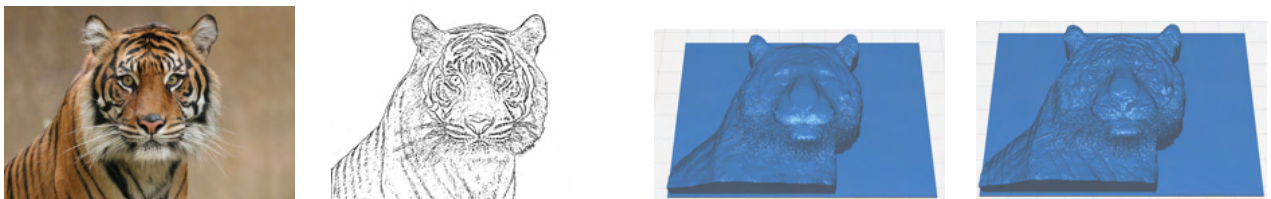
本章では、写真の深度情報と輝度情報を統合した2.5Dレリーフ（DLR）生成手法を提案し、その有効性を検証した。実験の結果、DLRは深度のみを用いた2.5Dレリーフ（DR）、従来の触図（TG）と比較して高い評価を得た。DLRは、従来のTGより優れた奥行き表現能力を有し、DRでは欠落する視覚的特徴を触覚化することで触覚写真の表現領域を拡張する、有効な手法であることが示唆された。

一方で、本手法は深度推定に基づいているため、撮影環境によっては課題も残る。特に、動物園の檻や窓の格子など、被写体の手前に遮蔽物が存在するシーンでは、遮蔽物がそのまま立体化されてしまい、奥にある主要な被写体の触察を物理的に阻害してしまう。このような複雑なシーンにおいて、2.5Dレリーフの表現力を十分に発揮させるためには、深度推定を行う前に遮蔽物を除去するなどの適切な画像処理を行うことが望ましい。

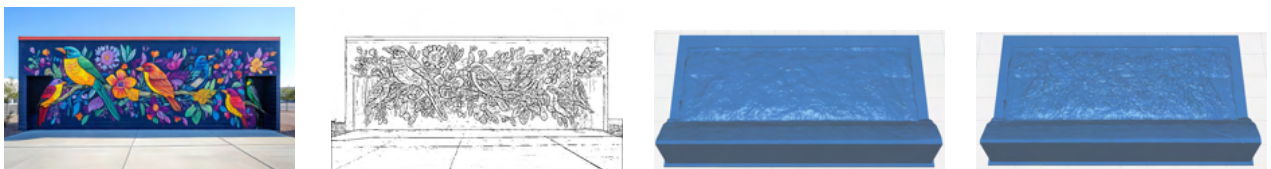
次章では、こうした遮蔽物の課題へのアプローチとして、深度情報を活用した前景除去手法について述べる。



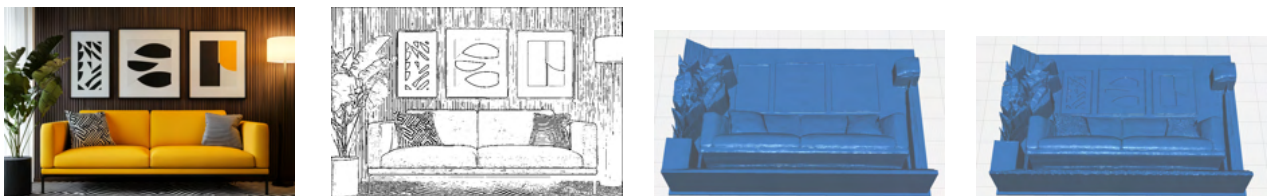
シマウマの写真



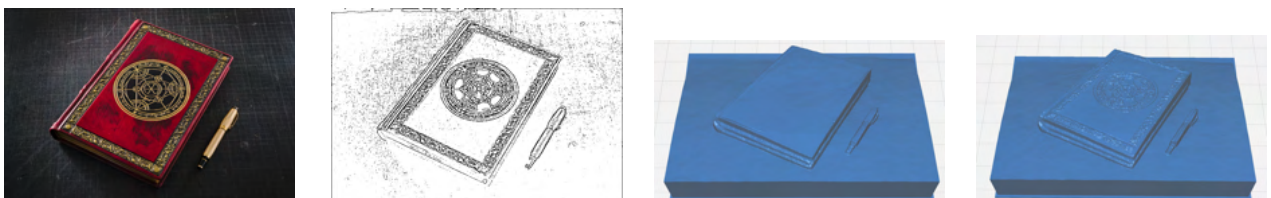
トラの写真



壁画の写真

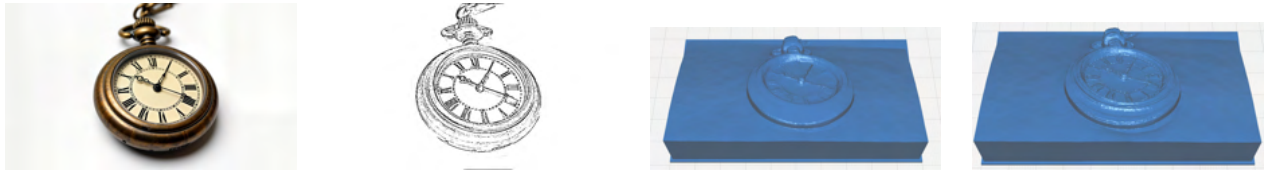


部屋の写真

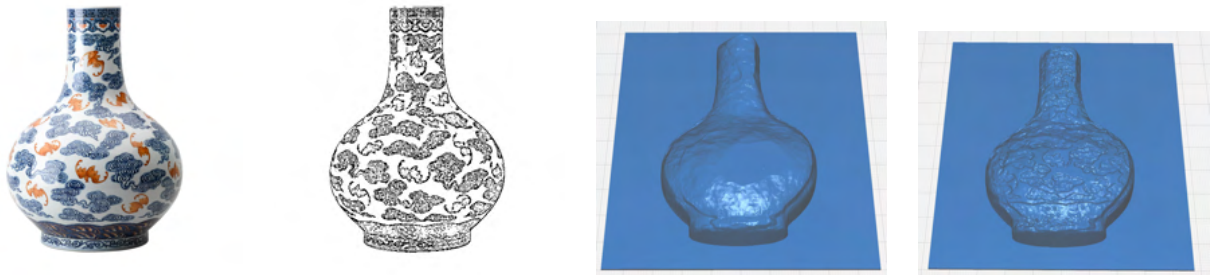


本の写真

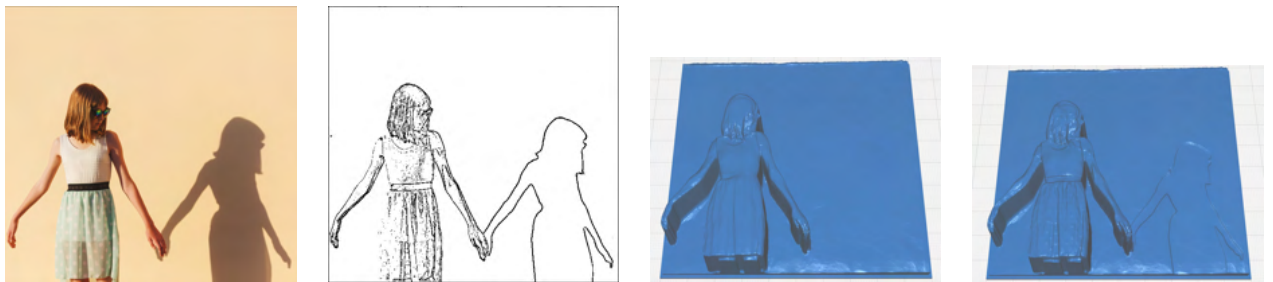
図 3.4: ユーザスタディで使った画像とモデル (1 / 3). 左から元画像, 触図 (TG), 深度に基づく 2.5D レリーフ (DR), 深度と輝度に基づく 2.5D レリーフ (DLR).



時計の写真



壺の写真

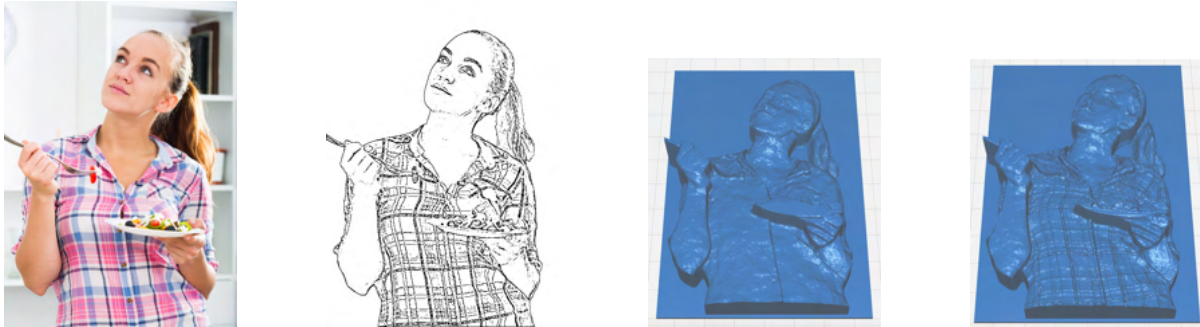


影の写真



年配の女性の写真

図 3.5: ユーザスタディで使用した画像とモデル (2 / 3). 左から元画像, 触図 (TG), 深度に基づく 2.5D レリーフ (DR), 深度と輝度に基づく 2.5D レリーフ (DLR).



若い女性の写真

図 3.6: ユーザスタディで使った画像とモデル (3 / 3). 左から元画像, 触図 (TG), 深度に基づく 2.5D レリーフ (DR), 深度と輝度に基づく 2.5D レリーフ (DLR).

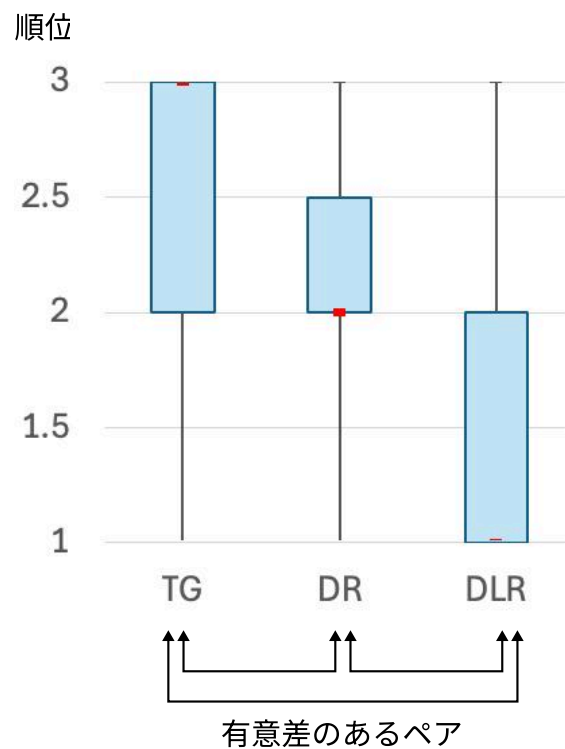


図 3.7: 各触覚モデルに割り当てられた順位の箱ひげ図。矢印はウィルコクソン符号付き順位検定に従って統計的に有意な差のあるペアを表す。赤い点は各箱ひげ図の中央値を示す。TG, DR, DLR はそれぞれ触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する。

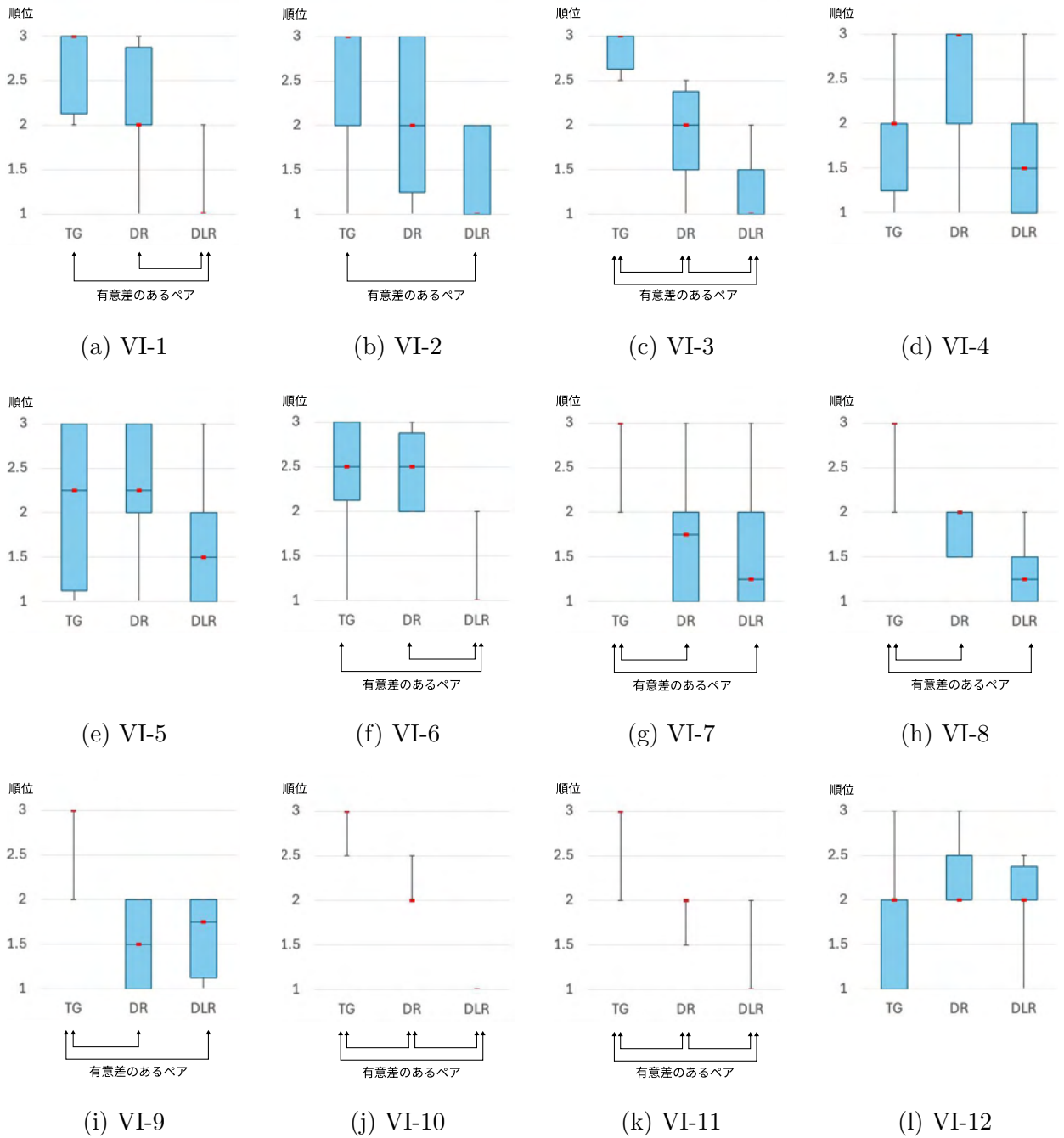


図 3.8: 各被験者が各触覚モデルに割り当てた順位の箱ひげ図。矢印はウィルコクソン符号付き順位検定に従って統計的に有意な差のあるペアを表す。赤い点は各箱ひげ図の中央値を表す。TG, DR, DLR はそれぞれ触図, 深度に基づく 2.5D レリーフ, 深度と輝度に基づく 2.5D レリーフを意味する。

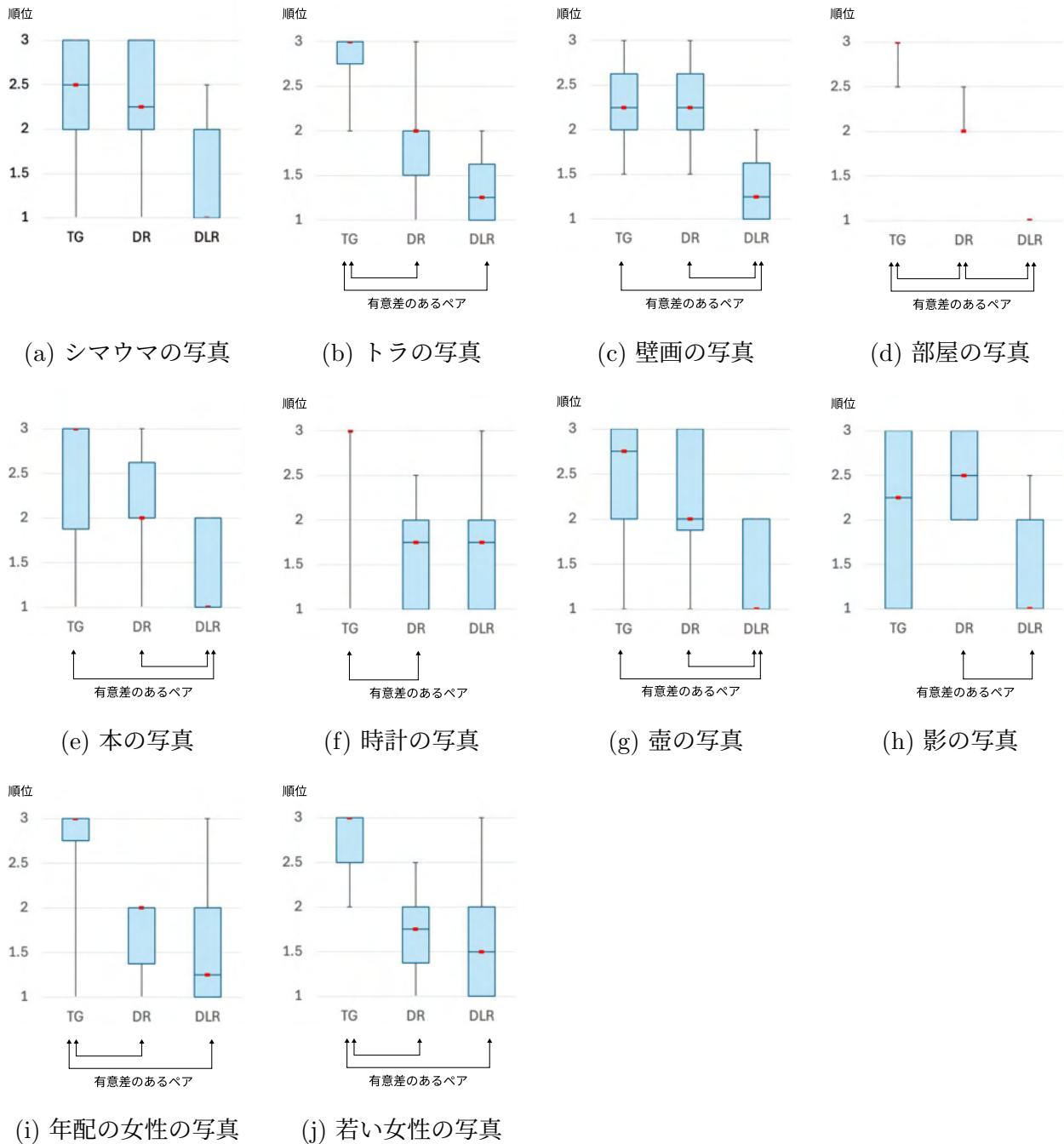


図 3.9: 写真毎に各触覚モデルに割り当てられた順位の箱ひげ図. 矢印はウィルコクソン符号付き順位検定に従って統計的に有意な差のあるペアを表す. 赤い点は各箱ひげ図の中央値を示す. TG, DR, DLRはそれぞれ触図, 深度に基づく2.5Dレリーフ, 深度と輝度に基づく2.5Dレリーフを意味する.

第4章 深度情報に基づく前景除去による 触察体験の向上

前章では、写真全体の情報を豊かに伝えるための汎用的な表現手法（DLR）について論じた。しかし、実環境における写真撮影では、意図せず手前に金網や柵などの遮蔽物が入り込む場合がある図 4.1。前章のような深度に基づくレリーフ生成においては、これら手前の物体が高い深度値を持つため、主要な被写体よりも手前に突出して造形される。このことは、奥にある主要な被写体の触察にとって大きな障壁となる。このようなケースで



図 4.1: 前景が触察を妨げる例。ネコの画像と生成された 2.5D レリーフには前景のメッシュが存在し、触察を妨げる。

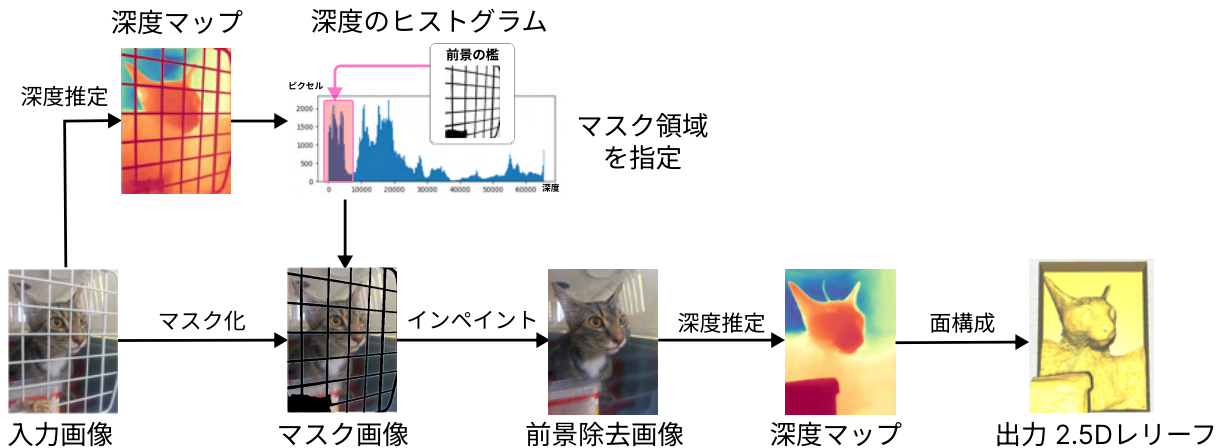


図 4.2: 提案手法のパイプラインの概要と生成例

は、レリーフ作成の前処理として画像から不要な遮蔽物を除去するアプローチが有効である。そこで本章では、深度情報を活用して手前の遮蔽物（前景）を指定し、画像修復技術を用いてこれを除去する手法について述べる。

4.1 提案手法: 前景メッシュ除去

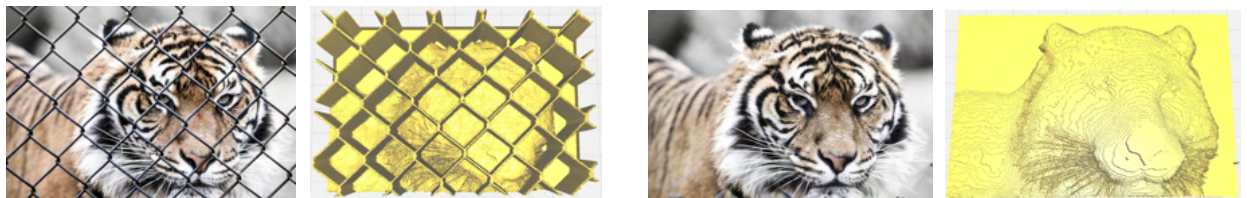
本章の手法の手順は以下の通りである。

1. **深度情報によるマスク画像の作成:** 入力画像を Marigold [15] に入力して深度マップを得る。図 4.2 の深度のヒストグラムに示すように、手前の遮蔽物はカメラに近い位置に突出したクラスタとして現れる。このクラスタを分離するための閾値を設定し、二値化処理を行うことで、除去すべき領域を示すマスク画像を生成する。
2. **画像修復:** 次に、遮蔽物の除去を行う。本手法では、画像のマスク部分を補完するためのコンピュータビジョン技術として、画像修復に注目した。画像修復では、マスク周囲の画素情報を元に、マスク部分の画像を推測し補完する。いくつかの画像修復モデルが提案されているが、本研究では大規模マスクに対応した画像修復モデルである Large Mask Inpainting (LaMa) [26] を採用した。元画像とステップ 1 で作成したマスク画像を LaMa に入力することで、前景の遮蔽物が除去され、背景が補完された画像（例：図 4.2 前景除去画像）が出力される。



(a) ネコの画像と 2.5D レリーフ

(b) ネコの画像の前景除去後と 2.5D レリーフ



(c) トラの画像と 2.5D レリーフ

(d) トラの画像の前景除去後と 2.5D レリーフ

図 4.3: 実験で使った画像および生成された 2.5D レリーフと、各画像の前景除去後の画像および生成された 2.5D レリーフ。

3. **2.5D レリーフの生成:** 修復された画像に対して深度推定を行い、正規化およびメッシュ化処理を適用することで、遮蔽物のない 2.5D レリーフ（例：図 4.2 の出力）を生成する。

4.2 予備実験

提案手法（前景除去）が視覚障害者の触察に与える影響を検証するため、予備的な評価実験を行った。

4.2.1 実験条件

本実験では、以下の2つの条件で作成された2.5D レリーフを比較対象とした。

(1) **条件 1: 基本手法 (前景除去なし)** 比較対象として、前章で述べた「深度情報に基づく2.5D レリーフ (DR)」の手法 (3.1.1 節) を適用したモデルを用いる。具体的には、Marigold による推定深度を $[0, 1]$ に正規化して立体化する。この条件では、手前の遮蔽物がそのまま立体化され、奥にある被写体の触察を妨げる状態 (例: 図 4.2(c)) となることが想定される。

(2) **条件 2: 提案手法 (前景除去あり)** 前節 (4.1 節) で述べた手順により、深度情報に基づいて遮蔽物を除去し、背景を補完した上で立体化したモデルを用いる。

4.2.2 評価方法

評価は、著者 (晴眼者) による生成結果の視覚的な整合性の評価と、全盲の共著者による触覚的な評価の2つの観点から実施した。使用した画像と生成結果の例を図 4.3 に示す。すべてのモデルは最大辺長 10 cm、奥行き 5 cm にスケールリングして出力した。

4.2.3 結果 1: 生成された2.5D レリーフの視覚的評価

まず、生成されたレリーフの品質を視覚的に評価した。

条件 1 (基本手法) の結果を見ると、図 4.3(a) の「ネコの画像」の例では、ネコの顔と手前の網の深度関係は比較的正確に推定されていた。しかし、ケージの底部とネコの体が深度的に区別できていない箇所や、ケージの上部・側面・背面が均一な背景として扱われてしまっている箇所があるなどの課題も見られた。図 4.3(c) の「トラの画像」の例では、フェンスの交差部分も含めて深度推定は概ね正確であったが、トラの顔の微細な凹凸情報は一部失われていた。

次に、条件 2 (提案手法) の結果を評価する。図 4.3(b) および (d) を見ると、LaMa による画像修復は概ね良好に機能しており、視覚的に自然な結果が得られている。ただし、

図 4.3(b) では、網を除去した跡に白い靄（もや）のようなアーティファクトが残る現象が見られた。これは元の白い網の反射光が周囲に散乱していた影響と考えられ、マスク処理で完全に取り切れなかったためである。しかし、これらのアーティファクトが深度推定の精度や生成されたレリーフの形状に対して影響を与えている様子は確認できなかった。特筆すべき点として、図 4.3(c) のレリーフではフェンスに遮られて認識が難しかったトラの顔が、除去後の図 4.3(d) のレリーフでは明確になり、深度推定の精度も向上していることが確認できた。

全体として、深度推定および画像修復の精度には改善の余地があるものの、視覚的には概ね満足のいく結果が得られたと言える。

4.2.4 結果 2: 生成された 2.5D レリーフの触覚的評価

続いて、先天盲の共著者（以下、評価者と呼ぶ）が、生成された 2.5D レリーフの触覚的なアクセシビリティを評価した。評価中の様子を図 4.4(b) に示す。

条件 1（基本手法）で生成されたレリーフについて、評価者は「手前の網が邪魔で、奥にある物体を理解するのが困難である」と報告した。特に図 4.3(a) の場合、遠近法の影響で画像の右側に行くほど網の目が小さくなっており、指を入れることが困難であった。指が入るサイズの網目であっても、一般的な触察動作である指でなでるような動き [13] が阻害されるため、評価者は網の隙間に指を差し込み、断片的な情報を脳内で統合して全体像を推測する必要があった。この作業は認知的負荷が非常に高いものであった。例えば、図 4.3(a) では「網の向こうにネコがいる」という事前情報を与えられれば耳などを識別できたが、図 4.3(c) では「何かが右側にある」程度しか認識できなかった。

対照的に、条件 2（提案手法）で生成されたレリーフでは、表面を連続的に指でなでることが可能となり、相対的な位置関係や特徴の変化を低い認知的負荷で把握できた。また、被写体そのものの理解度も大幅に向上した。口頭での説明と組み合わせることで、評価者はモデルの具体的な部位を特定することができた。例えば、図 4.3(d) では「これはトラである」と伝えられた後、目・鼻・口といったパーツを正確に識別できた。同様に図 4.3(b) では、元のレリーフでは網に隠れて分からなかった左下の餌箱を認識することができた。さらに、両方の画像において、被写体がカメラに対してわずかに右を向いていることなど、空間的な向きまで理解することができた。



図 4.4: 予備実験の様子

これらの結果は、手前の遮蔽物を除去することが、視覚障害者の触察による対象理解において有効である可能性を示唆している。

4.3 小括

本章では、深度情報に基づくレリーフ生成において、手前の遮蔽物が奥にある被写体の認識を物理的に阻害するという課題に対処するため、深度情報を活用して前景領域を特定し、画像修復技術を用いてこれを除去する手法を提案した。予備実験の結果、本手法を用いて金網などの遮蔽物を除去したモデルでは、ユーザの触察動作が改善されることが確認された。具体的には、網目の隙間に指を差し込んで断片的な情報を集めるという困難な動作が不要となり、表面を連続的になぞることで形状を把握できるようになった。この変化は、ユーザの認知的負荷を大幅に低減させると同時に、対象物の形状認識を容易にするものであると考えられる。

第5章 議論

本章では、第3章および第4章で得られた実験結果や、実験中に得られた観察から、視覚障害者が写真という視覚メディアを触覚的に受容する際の認知プロセス、視覚特有の概念の理解における課題、およびユーザの特性に応じたパーソナライズの必要性について総合的に議論する。また、本研究の限界についても述べる。

5.1 視覚経験と情報取得方略

本研究のユーザスタディから、視覚障害者が写真の触覚表現にどのように関わるかについて、いくつかの重要な傾向が示唆された。特に、参加者が情報を得る際の方略は、過去の視覚経験の有無によって異なる傾向が見られた。

参加者が情報を得る際の方略は、過去の視覚経験によって異なる傾向が見られた。例えば、中途失明の参加者の中には、言葉による説明を聞いた上で、得られた触覚と過去の視覚記憶とを結びつけて理解する者がいた。一方で、先天性視覚障害の参加者は、これまでの触覚経験に基づいて場面を想像すると述べた。この差異は、触覚モデルの設計において、ユーザの背景に応じた情報の提示方法を考慮する必要性を示唆している。

5.2 視覚的概念の触覚的翻訳における課題

本研究では、DLRや前景除去手法の有効性が示された一方で、触察による表現の限界、特に「視覚特有の要素」を伝えることの難しさも浮き彫りとなった。

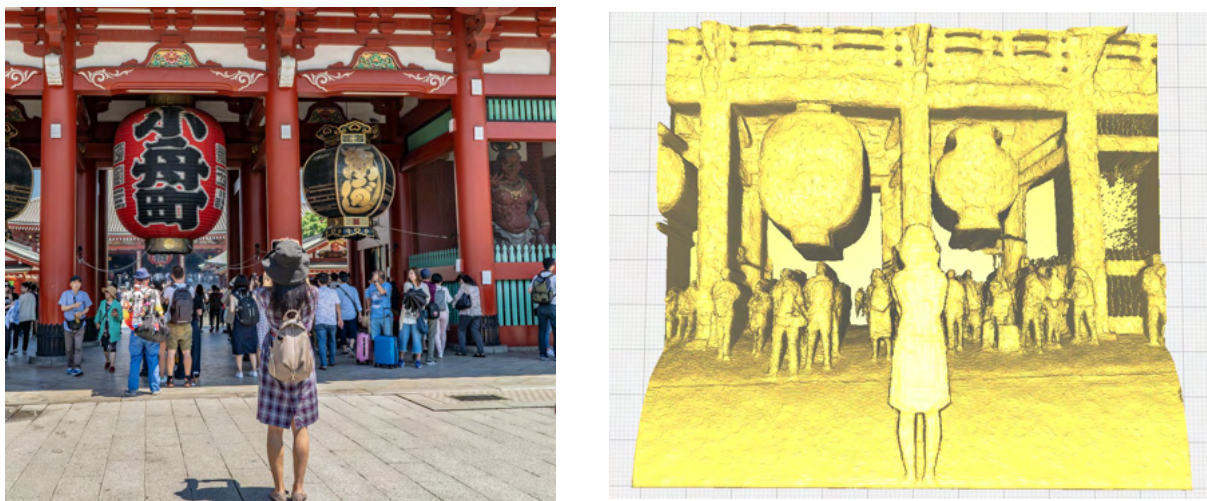


図 5.1: 遠近法を用いた写真の例とその 2.5D レリーフ。

5.2.1 色・見切れ・構図

多くの参加者が、色、見切れ、固定された画角といった写真特有の要素について、感覚的な理解の難しさを訴えた。中途失明の参加者の中には、色を感じ取ることができないことに対してもどかしさを表明する者がいた一方で、先天盲の参加者は、そもそも「色」という概念そのものを理解すること自体が困難であると述べた。また、写真の構図によって物体の一部が切り取られる「見切れ」についても、なぜそこで情報が途絶えるのかを視覚経験のないユーザが理解することは難しい。また、触覚による鑑賞では、それが見切れによる形状変化なのか、実際に被写体がそのような形状をしているのかを判断する必要性が生じる。これらのことは、触覚による写真鑑賞における認知的負荷を高める要因となる可能性がある。

5.2.2 遠近法の理解

2.5D レリーフは視覚障害者が画像を理解するための身近な手段となる可能性があるが、そこで大きな障壁となるのが「遠近法」の理解であろうと考えられる。先天盲の参加者にとって、「近くの物体が大きく、遠くの物体が小さく見える」という遠近法の内容は、直感的に理解しにくいものである。視覚メディアや写真はしばしば遠近法を構図に取り入れて

おり、このことが画像理解を困難にすることが考えられる。本研究のユーザスタディや予備実験においても、遠近法に起因する誤解や混乱が観察された。

例として、前景にカメラに近い人物が立ち、背景に遠くの建物が写っている写真を考える（図 5.1 参照）。視覚的には、人物と建物の距離関係や、それぞれの実際の大きさを容易に把握できる。しかし、この写真を DLR のような手法で単純に 2.5D レリーフに変換した場合、手前の人物が相対的に大きく造形されるため、触察者は背景の建物の大きさを実際より過小に評価するおそれがある。仮に遠近法の内容を知識として理解していたとしても、触察の過程で常に距離によるサイズ変化を補正しながら立体構造を把握することは、極めて認知的負荷が高い作業である。

第 4 章で提案した前景除去（分離）手法は、この課題に対する一つの解となり得る。遠近感が強く作用する写真に対しては、深度情報に基づいて前景と背景を分離し、個別に、あるいは前景を除去して触察することで、遠近法に起因する混乱を避け、認知的負荷を軽減できることが期待される。

5.3 写真鑑賞の目的と適応的なデザイン

参加者が写真を鑑賞する動機もまた、触覚モデルの好みに影響を与えることが明らかになった。参加者の一部は、物体認識や被写体の同定を重視し、ノイズの少ない明確な形状を好んだ。彼らは見切れや遠近法などの視覚的特徴を不要な情報と捉える傾向にあり、第 4 章のような遮蔽物除去や単純化されたモデルを支持した。一方で、他の参加者は写真全体の雰囲気や原本への忠実性を重視し、構図全体の印象や、元の写真に対する 2.5D レリーフの再現度を求めた。彼らにとっては、遠近法や見切れといった要素も写真らしさを構成する重要な要素であり、認知的負荷が高まってもそれを含めて鑑賞したいと考える傾向が見られた。

これらの結果は、万人に最適な単一の触覚モデルは存在せず、利用者の背景、目的、および理解の方略に応じて、情報の密度や加工の有無といったモデルの設計を柔軟に調整することが有益であることを示唆している。また、写真の忠実な表現を求める先天盲の参加者に対しては、モデルの提供だけでなく、遠近法などの視覚的表現についてのリテラシーを身につけるための触覚教材やガイダンスの併用も重要になると考えられる。

5.4 本研究の限界

本研究にはいくつかの限界が存在する。

第一に、評価の範囲である。本研究のユーザスタディにおける評価は、専門的な知識を必要としない自動的、あるいは半自動的で低コストな手法との比較のみに基づいており、専門家の手作業による触図、レリーフとの比較検証は行われていない。また、AHP分析で用いた評価基準は著者側で定義したものであり、今後はユーザ主体での指標設計（ユーザ自身が何を重要視するか）を検討する必要がある。

第二に、提案手法の技術的課題である。本手法では、輝度の重みパラメータや深度の閾値を手動で調整する必要があり、完全な自動化には至っていない。また、DLRにおいて輝度情報を高さの変化として加えるアプローチは、ユーザが実際の物体の形状としての凹凸と、輝度に由来するテクスチャとしての凹凸を区別しづらくなるという原理的な曖昧さを抱えている。

第6章 結言

写真は日常生活において重要な役割を果たしているが、視覚障害者はその内容に十分にアクセスすることが難しい。本研究は、この情報の格差を埋めることを目的とし、新たな触覚提示手法の提案および評価を行った。

まず、単一の写真から深度推定と輝度情報を組み合わせて2.5D レリーフを生成する新たな手法（DLR）を提案した。このモデルは、従来の触図（TP）と比較して奥行き情報の表現力に優れ、深度情報のみを用いる手法（DR）では失われがちな模様などの視覚的特徴を保持することができる。視覚障害者12名（全盲6名、弱視6名）を対象としたユーザスタディの結果、DLRはTPおよびDRと比較して全体的に高く評価された。

次に、2.5D レリーフにおける課題として、写真の手前に金網などの遮蔽物が存在する場合、被写体の触察が物理的に阻害される点に着目した。これに対し、深度情報に基づいて前景と背景を分離・除去する手法を提案した。視覚的な評価および全盲の共著者による触察評価の結果、本手法によって物理的な障壁を取り除くことで、画像内の物体の理解が向上することが確認された。

また、一連の調査を通じて、視覚障害者が情報を取得する際の方略が過去の視覚経験によって異なることが考察された。さらに、色や遠近法、見切れといった写真特有の視覚的概念は、触覚のみでの理解に認知的な困難を伴うことが示唆された。利用者が写真を鑑賞する際に分かりやすさを重視するか、忠実性を重視するかといった動機づけも触覚モデルの好みに影響を与えることから、個々の背景や目的に応じたパーソナライズされたモデル設計、および視覚的概念への理解を促す手引きの提供が、アクセシビリティ向上のために重要であると考えられる。

今後の課題としては、現時点ではパラメータ調整や前景除去の判断に視覚健常者の介入を必要とするプロセスの完全自動化が挙げられる。また、前景除去手法については、より多人数での定量的な評価を行う必要がある。将来的には、集合写真や動画のキーフレームといった多様なコンテンツへの適用を進め、視覚障害者の自立的な写真鑑賞を支援するシ

システムの構築を目指す。

謝辞

本研究を進めるにあたって、直接御指導いただいた岩村雅一教授には、研究内容や論文の書き方、発表方法、資料の作成などにおいて、多くの御指導、御助言を賜りましたことを深く感謝致します。また、黄瀬浩一教授をはじめ、岩田基準教授、内海ゆづ子准教授には研究発表会等で様々な御指摘、及び御助言を賜りましたことを深く感謝致します。さらに、大学入試センターの南谷 和範教授には、多くの御助言を賜りましたことを深く感謝致します。そして、研究開始当初からミーティングや発表練習等の場で御指導、御支援を頂いた視覚障害者支援班の方々をはじめ、多くの御協力を頂いた事務の方や様々な御助言を頂いた知能メディア処理研究グループの皆様に深く感謝致します。

2026年2月17日

参考文献

- [1] Hongchan Yoon, Baek-Hyun Kim, Mukhiddinov Mukhriddin, and Jinsoo Cho. Salient region extraction based on global contrast enhancement and saliency cut for image information recognition of the visually impaired. Vol. 12, No. 5, May 2018.
- [2] Karolina Pakenaite, Eirini Kamperou, Michael J Proulx, Adwait Sharma, and Peter Hall. Pic2tac: Creating accessible tactile images using semantic information from photographs. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '24, pp. 1–12. ACM, 2024.
- [3] 大内進, 国立特別支援教育総合研究所. 視覚障害教育のための3次元CADを活用した2次元画像の立体的翻案に関する研究 —研究成果報告書—. 研究成果報告書 特教研 F-149, 国立特別支援教育総合研究所, 神奈川県横須賀市, 2008. 参照日: 2026-02-11.
- [4] Destin Sparks. Destin sparks — quotes “photography is the story i fail to put into words”. <https://www.destinsparks.com/links/photography-quotes/>. Accessed: 2025-08-13.
- [5] Be My Eyes. Introducing: Be my ai. <https://www.bemyeyes.com/blog/introducing-be-my-ai/>, 2023.
- [6] Thomas Way and Kenneth Barner. Automatic visual to tactile translation - part i: Human factors, access methods, and image manipulation. *Rehabilitation Engineering, IEEE Transactions on*, Vol. 5, pp. 81 – 94, 04 1997.
- [7] Thomas Way and Kenneth Barner. Automatic visual to tactile translation. ii. evaluation of the tactile image creation system. *Rehabilitation Engineering, IEEE Transactions on*, Vol. 5, pp. 95 – 105, 04 1997.

-
- [8] Tyler J. Ferro and Dianne T. V. Pawluk. Automatic image conversion to tactile graphic. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '13*, pp. 1–2. ACM, 2013.
- [9] Akmalbek Abdusalomov, Mukhriddin Mukhiddinov, Oybek Djuraev, Utkir Khamdamov, and Taeg Keun Whangbo. Automatic salient object extraction based on locally adaptive thresholding to generate tactile graphics. Vol. 10, No. 10, p. 3350, 2020.
- [10] Braille Authority of North America (BANA). *Tactile Graphics Standards and Guidelines 2022*. Braille Authority of North America (BANA), 2022. Accessed on 2025-06-24.
- [11] Polly Edman. *Tactile graphics*. American Foundation for the Blind.
- [12] Moritz Neumüller and Andreas Reichinger. From stereoscopy to tactile photography. Vol. 19, pp. 59–63, April 2013.
- [13] 大内進, 国立特別支援教育総合研究所. 3D造形装置による視覚障害教育用立体教材の評価に関する実際的研究 —共同研究 研究成果報告書— 平成25年度～26年度, 2015. 参照日: 2026-02-11.
- [14] Rocco Furferi, Lapo Governi, Yary Volpe, Luca Puggelli, Niccolò Vanni, and Monica Carfagni. From 2D to 2.5D i.e. from painting to tactile model. *Graphical Models*, Vol. 76, No. 6, pp. 706–723, 2014.
- [15] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [16] S. E. Krufka and K. E. Barner. A user study on tactile graphic generation methods. Vol. 25, No. 4, pp. 297–311, 2006.

-
- [17] Ruo Zhang, Ping-Sing Tsai, J.E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 8, pp. 690–706, 1999.
- [18] J. Wu, R.R. Martin, P.L. Rosin, X.-F. Sun, F.C. Langbein, Y.-K. Lai, A.D. Marshall, and Y.-H. Liu. Making bas-reliefs from photographs of human faces. *Computer-Aided Design*, Vol. 45, No. 3, pp. 671–682, 2013.
- [19] Hai Thien To and Bong-Soo Sohn. Bas-relief generation from face photograph based on facial feature enhancement. *Multimedia Tools and Applications*, Vol. 76, No. 8, pp. 10407–10423, April 2017.
- [20] Lapo Governi, Monica Carfagni, Rocco Furferi, Luca Puggelli, and Yary Volpe. Digital bas-relief design: A novel shape from shading-based method. *Computer-Aided Design and Applications*, Vol. 11, pp. 153–164, March 2014.
- [21] MingYu Cui, Chao Yuan, Yujia Liu, and Yingying Zheng. More than shapes: Exploring the tactile parameters of art appreciation for the visually impaired. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '24, pp. 786–790. ACM, 2024.
- [22] Jaume Gual, Marina Puyuelo, and Joaquim Lloveras. Three-dimensional tactile symbols produced by 3D printing: Improving the process of memorizing a tactile map key. Vol. 32, No. 3, pp. 263–278, 2014.
- [23] Jaume Gual, Marina Puyuelo, and Joaquim Lloveras. The effect of volumetric (3D) tactile symbols within inclusive tactile maps. Vol. 48, pp. 1–10, 2015.
- [24] Leona Holloway, Kim Marriott, and Matthew Butler. Accessible maps for the blind: Comparing 3D printed models with tactile graphics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018.
- [25] HARPO. PIAF (pictures in a flash) —PIAF tactile by harpo. <https://piaf-tactile.com/piaf/>. Accessed: 2025-08-13.

-
- [26] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2149–2159, January 2022.