

# 視覚障害者による写真鑑賞のための深度推定に基づく 2.5Dレリーフの生成と編集

高石 康世

tkosei2003@gmail.com  
大阪公立大学 工学部 情報工学科

南谷 和範

minatani@rd.dnc.ac.jp  
独立行政法人 大学入試センター  
研究開発部

岩村 雅一

masa.i@omu.ac.jp  
大阪公立大学 大学院情報学研究科

## 深度推定による2.5Dレリーフ生成手法（単純な手法）



## 前景削除のための深度ベースの分解手法（提案手法）



図 1: 深度推定に基づく2.5Dレリーフ生成のための単純な手法 ((a)-(c)) と、提案する前景の除去に適用された深度ベースの分解技術の概要 ((a)-(b)および(d)-(g)).

### 概要

コンピュータビジョン技術の発展により、自然画像1枚から高精度の深度推定を行うことが可能になった。この技術を応用することで、写真をもとにした2.5Dレリーフの自動生成が実現できる。しかし、このように生成されたレリーフは、必ずしも触察に向いているとは限らない。例えば、写真全体にわたる網などの物体（障害物）が前景にある場合、その背後の物体の触察が阻害されてしまい、十分に写真の情報を伝達することができないという問題が考えられる。そこで我々は、深度情報に基づく分解手法を提案する。これは自動生成されたレリーフを複数のレイヤーに分割し、前景の遮蔽物を取り除くことを可能にするものである。予備実験の結果、この障害物の除去によってレリーフの明瞭さと使いやすさが大幅に向上し、認知的負荷を軽減することが示された。

### キーワード

視覚障害者、写真、2.5Dレリーフ、奥行き推定、インペインティング、触察

### 1 はじめに

写真は、記憶を保存し、感情を共有し、個性を表現し、さらには社会的なつながりを育むための手段として、きわめて重要な価値を持っている。例えば、結婚式や子どもの成長の節目といった重要な瞬間を記録し、時間を経てもそれらの記憶を振り返ることを可能にする。さらに、写真をソーシャルメディアで共有することにより、人々は地理的な距離を越えてつながり、互いの経験を分かち合うことで関係をより深めることができる。これらの例は、写真が人間のコミュニケーションにおいて果たしている、独自かつ不可欠な役割を示している。

しかし、視覚障害者にとって、写真と関わることには「撮影」「整理」「閲覧」という三つの側面において課題

が存在する。支援技術の発展により、撮影 [5, 11, 12, 20, 21, 24] や整理 [3, 4, 12, 28] の支援は進展を見せているものの、閲覧に関しては依然として大きな困難が残っている。従来、視覚障害者が写真や図を触覚的に理解するための手段としては、触図が一般的に用いられてきた。触図とは、物体の輪郭や明暗の変化などを、スウェルペーパーやマイクロカプセル紙などの特殊な紙に隆起させて表現する技法である [2, 8, 17, 22, 23, 26]。触図は、平面上の凹凸を指先でなぞることで、形状や構造を触覚的に探索・理解することを目的としている。しかし、このような触図は、平面上で一定の高さしか表現できないという物理的な制約を持ち、複雑な写真の情報を十分に伝えることが難しい。実際に、北米点字委員会 (BANA: Braille Authority of North America) の指針 [6] や、Edman の代表的著作 *Tactile Graphics* [7] においても、詳細な写真表現は触図の実用的な範囲外であるとされている。その結果、これまでの研究の多くは、視覚情報を簡略化する方向に焦点を当ててきた。具体的には、顕著な輪郭のみを抽出する [8]、注目すべき前景オブジェクトを分離する [2, 26]、あるいはオブジェクトを単純な形のアイコンに置き換える [17] といった手法である。しかし、これらの「情報の単純化」に基づく手法では、豊かな情報を持つ写真の伝達手段としては、不十分であると考えられる。

この課題を解決するために、本研究では触図に代わるアプローチとして、「2.5Dレリーフ (bas-relief)」に着目する。この手法では、各領域の高さを奥行きに応じて変化させ、観察者に近い部分を高く、遠い部分を低く表現することで、立体的な構造を再現する。このような写真の触覚表現は、二次元 (2D) と三次元 (3D) の中間に位置することから、「2.5D」と呼ばれている [14]。完全な3Dモデルとは異なり、2.5Dレリーフは写真に写っていない部分 (例えばオブジェクトの裏側など) の情報を保持しないが、輪郭線を隆起させた従来の触図とは異なり、高さの連続的な変化を表現できるため、より豊かな空間的表現が可能である [16]。しかし、従来の2.5Dレリーフ制作では、熟練した職人が手作業で彫刻を行う必要があり、制作の負担が大きい。近年は3Dプリント技術の発展により、デジタルデータから容易に立体物を生成できるようになったが、その前提となる写真の2.5Dレリーフのデータの作成には依然として課題が残る。既存の自動化・半自動化手法 [9, 10, 19, 25, 27] も提案されているが、人間の顔や絵画など特定の画像種に限定され、汎用的な自然写真には適用が難しい。

こうした背景から、本研究では、任意の自然写真1枚から奥行きを推定できる、ニューラルネットワークベースの単眼深度推定技術 [13] に注目した。近年のコンピュータビジョン技術の発展により、単眼深度推定の精度は大きく向上している。そのため、本研究では、この技術を活用することで、一般的な写真から奥行き情報を自動的に取得し、半自動的に2.5Dレリーフを生成できると考えた。我々はこの手法を基盤とする生成パイプラインを構築し、その出力を、図1に示す「単純な手法」として位置づける。さらに、このモデルを3Dプリンタ等で出力することで、視覚障害者は自分の好きな写真を触覚的な2.5Dレリーフとして鑑賞できるようになる。

しかし、単純な手法には依然として課題が残されている。写真の全体にわたって存在する前景の構造物 (たと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。この課題に対する着想として、我々は人体の構造理解におけるアプローチに注目した。人体のような複雑な構造は、解剖模型のように構成要素ごとに分解することで理解しやすくなることが知られている [1]。この発想を応用し、本研究では図1に示すように、深度情報に基づいて画像を層状に分解し、前景要素を除去する手法を提案する。本手法では、深度情報をもとに除去領域を特定するマスキング処理 (図1(d) 参照) と、除去後の領域を補完するインペインティング処理 (図1(e) 参照) を組み合わせる。インペインティングは、周囲の画素情報をもとにマスク領域を予測・再構成するコンピュータビジョン技術であり、物体除去などの用途で広く利用されている。

提案手法の有効性を検証するため、予備実験を実施した。実験では、この分解手法を図1に示すような前景の檻を除去する問題に適用し、その有用性を評価した。視覚評価としては、晴眼者の著者が各手法の出力の妥当性を確認した。さらに、先天性視覚障害をもつ著者が、単純な手法と提案手法の双方について触覚による評価を行った。

## 2 手法

### 2.1 単純な手法: 深度推定を用いた2.5Dレリーフ生成

図1に示すように、単純な手法では、高精度な深度推定モデルである Marigold [13]<sup>1</sup> を用いて、写真から2.5Dレリーフを生成する。以下では、図1の例を用いてその処理過程を説明する。まず、入力となる自然画像 (図1(a)) に対して深度推定を行い、画素ごとの深度情報を生成する (図1(b))。この深度情報に基づき、2.5Dレリーフの3Dモデル (図1(c)) を自動的に生成することができる。

### 2.2 提案手法: 前景の除去

図1の例を用いて、本手法の処理過程を示す。まず、図1(d)に示すように、檻の領域が含まれる深度範囲を閾値を設定することで指定し、そこに含まれる深度をもつ画素をマスクした画像を生成する。次に、このマスク画像をインペインティング (LaMa [18]) することで、檻部分が除去された画像 (図1(e)) を得ることができる。この画像を用いて新たに深度推定を行い、その結果 (図1(f)) をもとに、最終的な2.5Dレリーフ (図1(g)) を生成する。

## 3 予備実験

### 3.1 生成された2.5Dレリーフの視覚による評価

単純な手法と提案手法によって生成された2.5Dレリーフの品質は、ユーザによる触覚的理解に直接影響を与える。本研究では、定量的評価ではなく、晴眼者の著者による主観的評価を通じて生成結果の品質を確認した。いくつかの例を図3(a)に示す。これらの2.5Dレリーフは、最大辺長が10cm、奥行きが5cmとなるようにスケールングされ、3Dプリンターによって出力された。

まず、単純な手法による結果を評価した。図2(a)に示すように、この手法では猫の顔およびその周囲の檻の深度を良好に再現できたが、ケージの床と猫の胴体を区別できず、ケージの壁および天井は単一の平面であるかのように

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

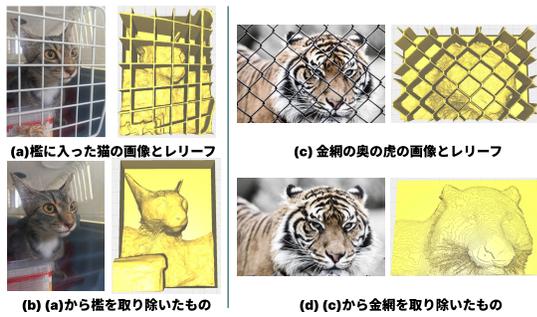


図 2: 予備実験で使用した画像（各小図の左側）と2.5Dレリーフ（右側）。



図 3: 予備実験で使用した出力済みの2.5Dレリーフと、触察の様子

図 4: 遠近法を用いた写真と2.5Dレリーフの一例。

深度が推定されていた。一方、図2(c)では、トラの顔のレリーフにおいて一部の深度情報が失われたものの、金網の交差部分は正確に再現されており、全体としておおむね良好な結果が得られた。

次に、提案手法の結果を評価した。図2(a)(b)および図2(c)(d)に示すように、インペインティングモデルは全体的に良好に機能していた。図2(b)では、元の白い檻からの反射によるわずかな白い霞が確認されるが、最終的な2.5Dレリーフには目立った影響を与えていない。また、トラの画像においては、特に顔の部分で、図2(c)の金網を除去していない場合に比べ、図2(d)の金網を除去した画像の方が、より現実的で自然な造形のように観察された。

全体として、両手法はいずれも一定の成果を達成し、視覚的に満足のいく結果を得ることができた。一方で、深度推定およびインペインティングの精度には、さらなる改善の余地が残されている。

### 3.2 生成された2.5Dレリーフの触覚的評価

先天性視覚障害をもつ著者（以下、「ユーザ」と呼ぶ）は、2.5Dレリーフの触覚的アクセシビリティを評価した。図3(b)は、評価中の様子を示した図である。檻を除去せずに単純な手法で生成したレリーフでは、背後の物体を理解しようとする際に、檻の存在が大きな妨げとなることが報告された。特に、図2(a)に示すレリーフでは、右側に向かうにつれて格子のサイズが小さくなっている。この現象は、右側の格子がカメラからより遠くに位置し、遠近法による歪みを受けるために生じる。その結果、触察のために指を挿入することが右の格子では困難になる。また、指を差し込める程度に十分な大きさの格子であっても、[16]で指摘されているような、指先を表面上で滑らせる一般的な触察動作が妨げられた。その代わりに、ユーザは網目の

隙間に指を差し込み、内部の物体を探索しながら、それぞれの相対的位置や形状を考慮して統合する必要があった。この作業は認知的負荷が高く、非常に困難であることが報告された。たとえば、図2(a)においては、レリーフの説明を受けた上で触察を行うことで、檻の内部に猫が存在することや、最も特徴的であった猫の耳の位置を把握するにとどまった。また、図2(c)では、檻の背後にある物体が画像の右側に位置していることを理解するにとどまった。

これに対して、提案手法で生成されたレリーフでは、表面を連続的にたどることが可能であり、相対的な位置関係や形状の変化をより容易に把握できた。その結果、認知的負荷が軽減され、触察による理解が大きく向上した。ユーザは、物体そのものの形状や特徴についても、より詳細に理解することができた。場面の口頭説明を受けたうえで触察を行うことで、モデルの特定の部分を識別することが可能となった。たとえば、図2(b)では、元のレリーフでは檻に覆われていた画像左下の給餌箱を認識することができた。同様に、図2(d)では、画像がトラであると知らされた後、ユーザは目・鼻・口などの部位を特定できた。さらに、両画像の被写体がカメラに対してやや右向きであることなど、全体的なシーン理解も向上した。

## 4 議論

3Dプリンタの利用がより容易になれば、2.5Dレリーフは視覚障害者が視覚的メディアを理解するための身近な手段となることが期待される。しかし、その普及に伴い、新たな課題として遠近法の理解が問題となる可能性がある。先天性視覚障害者にとって、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概念は、直感的に理解にくいものである [15]。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、このことが視覚障害者にとって画像理解を困難にする要因となり得る。たとえば、図4(a)では、前景にカメラに近い位置で立つ女性が写り、背景には門と提灯が見えている。視覚的には、門までの距離やその実際の大きさを容易に把握できる。しかし、この写真を単純な手法で2.5Dレリーフに変換した場合（図4(b)参照）、前景の女性が相対的に大きく表現されるため、門の実際の大きさが過小に解釈されるおそれがある。また、遠近法の概念を理解している場合でも、触察の過程で遠近によるサイズ差を考慮しながら立体構造を把握することは、視覚障害者にとって認知的負荷が大きい作業となる。2.5Dレリーフの普及に向けては、これらの遠近法に起因する課題へ適切に対処することが重要である。このような写真に対しては、本研究で提案する深度情報に基づく分解手法を用い、前景と背景を分離したうえで個別に触察することで、認知的負荷の軽減が期待できる。

## 5 結論

本研究では、視覚障害者が写真を理解するための手段として、2.5Dレリーフの活用を検討した。近年のコンピュータビジョン技術の発展により、2.5Dレリーフを半自動的に生成することが可能となったが、その結果得られるレリーフは必ずしも触察に向いているものとは限らない。特に、写真の全体にわたる前景の構造物が存在する場合は、被写体の触察が物理的に阻害される。

そこで本研究では、深度情報に基づいて自動生成されたレリーフを層構造に分解する深度ベース分解手法を提案し、これを用いて前景の除去を実現できることを示した。視覚的な評価では、深度推定の品質にいくつかの課題が確認されたものの、全体として妥当な2.5Dレリーフが生成可能であることが示された。さらに、先天性視覚障害をもつ著者による触察による評価の結果、前景の檻を除去する本手法により、画像内の物体の理解が向上することが報告された。

今後の研究では、提案手法が視覚障害者の写真理解をどの程度向上させるかについて、定量的な評価を行う必要がある。また、2.5Dレリーフと二次元の触図との比較を通じて、画像理解支援における2.5Dレリーフの有効性を検証することも重要である。さらに、現時点では前景除去の要否やヒストグラム閾値の設定などに視覚健常者の介入を要する点が課題として残っている。今後は、これらのプロセスを自動化することが、視覚障害者の自立支援という観点からも重要な研究課題となるだろう。

## 謝辞

本研究は、キャノン財団の助成を一部受けて実施された。

## 参考文献

- [1] [n. d.]. *Disassembled Medical Model of Larynx*. Retrieved Jun. 26, 2025 from <https://www.turbosquid.com/3d-models/disassembled-medical-model-of-larynx-3d-model-2004706>
- [2] Akmalbek Abdusalomov, Mukhriddin Mukhiddinov, Oybek Djuraev, Utkir Khamdamov, and Taeg Keun Whangbo. 2020. Automatic Salient Object Extraction Based on Locally Adaptive Thresholding to Generate Tactile Graphics. *10, 10* (2020), 3350. <https://doi.org/10.3390/app10103350>
- [3] Dustin Adams, Sri Kurniawan, Cynthia Herrera, Veronica Kang, and Natalie Friedman. 2016. Blind Photographers and VizSnap: A Long-Term Study. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) (ASSETS '16). Association for Computing Machinery, New York, NY, USA, 201–208. <https://doi.org/10.1145/2982142.2982169>
- [4] Dustin Adams, Lourdes Morales, and Sri Kurniawan. 2013. A qualitative study to support a blind photography mobile application. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments* (Rhodes, Greece) (PETRA '13). Association for Computing Machinery, New York, NY, USA, Article 25, 8 pages. <https://doi.org/10.1145/2504335.2504360>
- [5] Jan Balata, Zdenek Mikovec, and Lukas Neoproud. 2015. BlindCamera: Central and Golden-ratio Composition for Blind Photographers. In *Proceedings of the Multimedia, Interaction, Design and Innovation* (Warsaw, Poland) (MIDI '15). Association for Computing Machinery, New York, NY, USA, Article 8, 8 pages. <https://doi.org/10.1145/2814464.2814472>
- [6] Braille Authority of North America (BANA). 2022. *Tactile Graphics Standards and Guidelines 2022*. Braille Authority of North America (BANA). [https://www.brailleauthority.org/sites/default/files/tg/Tactile%20Graphics%20Standards%20and%20Guidelines%202022\\_a11y.pdf](https://www.brailleauthority.org/sites/default/files/tg/Tactile%20Graphics%20Standards%20and%20Guidelines%202022_a11y.pdf) Accessed on 2025-06-24.
- [7] Polly Edman. [n. d.]. *Tactile graphics*. American Foundation for the Blind. Includes bibliographical references (p. 498-500) and index.
- [8] Tyler J. Ferro and Dianne T. V. Pawluk. 2013. Automatic image conversion to tactile graphic. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (2013-10) (ASSETS '13). ACM, 1–2. <https://doi.org/10.1145/2513383.2513406>
- [9] Rocco Furferi, Lapo Governi, Yary Volpe, Luca Puggelli, Niccolò Vanni, and Monica Carfagni. 2014. From 2D to 2.5D i.e. from painting to tactile model. *Graphical Models* 76, 6 (2014), 706–723. <https://doi.org/10.1016/j.gmod.2014.10.001>
- [10] Lapo Governi, Monica Carfagni, Rocco Furferi, Luca Puggelli, and Yary Volpe. 2014. Digital Bas-Relief Design: A Novel Shape from Shading-Based Method. *Computer-Aided Design and Applications* 11 (March 2014), 153–164. <https://doi.org/10.1080/16864360.2014.846073>
- [11] Naoki Hirabayashi, Masakazu Iwamura, Zheng Cheng, Kazunori Minatani, and Koichi Kise. 2023. VisPhoto: Photography for People with Visual Impairments via Post-Production of Omnidirectional Camera Imaging. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) (ASSETS '23). Association for Computing Machinery, New York, NY, USA, Article 6, 17 pages. <https://doi.org/10.1145/3597638.3608422>
- [12] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (Dundee, Scotland, UK) (ASSETS '11). Association for Computing Machinery, New York, NY, USA, 203–210. <https://doi.org/10.1145/2049536.2049573>
- [13] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caya Daudt, and Konrad Schindler. 2024. Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [14] Moritz Neumüller and Andreas Reichinger. 2013. From Stereoscapy to Tactile Photography. *19* (April 2013), 59–63. <https://www.vrvis.at/publications/PB-VRVis-2013-009>
- [15] Susumu Ouchi, Hideyuki Doi, and Loretta Sechi. 2006. [Initiatives for Art Appreciation for Children and Adults with Visual Impairments in Italy] (Itaria ni okeru Shikaku Shougai Jisha no tame no Kaiga Kanshou no Torikumi). *Sekai no Tokushu Kyouiku [Special Needs Education in the World]* 20 (2006), 83–100. <https://cir.nii.ac.jp/crid/1574231877296538624> In Japanese.
- [16] Susumu Ouchi and National Institute of Special Needs Education (NISE). 2014. Evaluation of Tactile Teaching Materials for Visual Impairment Education Using 3D Printing Technology (3D Zokei Sochi o Mochiita Shikaku Shogai Kyoiku-yo Rittai Kyozaai no Hyoka ni Kansuru Kenkyu). [https://www.nise.go.jp/cms/8\\_9313\\_52\\_293.html](https://www.nise.go.jp/cms/8_9313_52_293.html) In Japanese.
- [17] Karolina Pakenaite, Eirini Kamperou, Michael J Proulx, Adwait Sharma, and Peter Hall. 2024. Pic2Tac: Creating Accessible Tactile Images using Semantic Information from Photographs. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction* (2024-02) (TEI '24). ACM, 1–12. <https://doi.org/10.1145/3623509.3633377>
- [18] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. 2022. Resolution-Robust Large Mask inpainting With Fourier Convolutions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2149–2159.
- [19] Hai Thien To and Bong-Soo Sohn. 2017. Bas-relief generation from face photograph based on facial feature enhancement. *Multimedia Tools and Applications* 76, 8 (April 2017), 10407–10423. <https://doi.org/10.1007/s11042-016-3924-y>
- [20] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (Boulder, Colorado, USA) (ASSETS '12). Association for Computing Machinery, New York, NY, USA, 95–102. <https://doi.org/10.1145/2384916.2384934>
- [21] Marynel Vázquez and Aaron Steinfeld. 2014. An Assisted Photography Framework to Help Visually Impaired Users Properly Aim a Camera. *ACM Trans. Comput.-Hum. Interact.* 21, 5, Article 25 (Nov. 2014), 29 pages. <https://doi.org/10.1145/2651380>
- [22] Thomas Way and Kenneth Barner. 1997. Automatic visual to tactile translation - Part I: Human factors, access methods, and image manipulation. *Rehabilitation Engineering, IEEE Transactions on* 5 (04 1997), 81 – 94. <https://doi.org/10.1109/86.559353>
- [23] Thomas Way and Kenneth Barner. 1997. Automatic visual to tactile translation. II. Evaluation of the TACTile image creation system. *Rehabilitation Engineering, IEEE Transactions on* 5 (04 1997), 95 – 105. <https://doi.org/10.1109/86.559354>
- [24] Samuel White, Hanjie Ji, and Jeffrey P. Bigham. 2010. EasySnap: real-time audio feedback for blind photography. In *Adjunct Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) (UIST '10). Association for Computing Machinery, New York, NY, USA, 409–410. <https://doi.org/10.1145/1866218.1866244>
- [25] J. Wu, R.R. Martin, P.L. Rosin, X.-F. Sun, F.C. Langbein, Y.-K. Lai, A.D. Marshall, and Y.-H. Liu. 2013. Making bas-reliefs from photographs of human faces. *Computer-Aided Design* 45, 3 (2013), 671–682. <https://doi.org/10.1016/j.cad.2012.11.002>
- [26] Hongchan Yoon, Baek-Hyun Kim, Mukhiddinov Mukhriddin, and Jinsoo Cho. 2018. Salient Region Extraction based on Global Contrast Enhancement and Saliency Cut for Image Information Recognition of the Visually Impaired. *12, 5* (May 2018). <https://doi.org/10.3837/tiis.2018.05.021>
- [27] Ruo Zhang, Ping-Sing Tsai, J.E. Cryer, and M. Shah. 1999. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 8 (1999), 690–706. <https://doi.org/10.1109/34.784284>
- [28] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2017. The Effect of Computer-Generated Descriptions on Photo-Sharing Experiences of People with Visual Impairments. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 121 (Dec. 2017), 22 pages. <https://doi.org/10.1145/3134756>