Photo-Based 2.5D Reliefs for Teaching Perspective and Depth to the Visually Impaired

Kazunori Minatani, Kosei Takaishi, and Masakazu Iwamura

[1] National Center for University Entrance Examinations, Tokyo, Japan
`minatani@rd.dnc.ac.jp`
[2] Osaka Metropolitan University, Osaka, Japan
`tkosei2003@gmail.com`
[3] Osaka Metropolitan University, Osaka, Japan
`masa.i@omu.ac.jp`

Abstract. The visually impaired have limited access to understanding vision-based spatial phenomena such as linear perspective, depth cues, and occlusion, which sighted people routinely use to interpret images. While prior work has highlighted the importance of making perspectival images accessible to blind learners, it has provided few concrete, generalizable strategies for systematically teaching pictorial space. We propose a practical and scalable approach that uses photographs (rather than canonical paintings) as the source of tactile educational material. Our pipeline estimates depth from a photograph to generate a 2.5D relief height field, then fabricates tactile models via 3D printing. This photo-based route is (i) *general*, because photographs can be captured on-demand to exemplify specific spatial cues in a staged and progressively structured manner, and (ii) *practical*, because it reduces manual production burden compared to handcrafted tactile translations. We integrate lessons from tactile picture perception research and present a stepwise pedagogy that begins with monocular perspective cues and extends toward binocular depth concepts. We further propose a qualitative evaluation framework based on think-aloud protocols, semi-structured interviews, and thematic analysis to assess educational efficacy of tactile photographs.

Keywords: Visual impairment, tactile pictures, perspective education, 2.5D relief, depth estimation, 3D printing

1 Introduction

Understanding pictorial space—how a three-dimensional (3D) environment is projected into a two-dimensional (2D) image—is non-trivial even for sighted novices; for individuals who are congenitally blind, it is especially challenging because the relevant cues (foreshortening, convergence of parallel lines, instantaneous field-of-view occlusion) are inherently vision-based conventions rather than direct haptic invariants. Classic research on tactile picture perception shows that interpreting raised-line pictures is difficult for both blind and blindfolded sighted observers when the task is novel, and performance depends strongly on complexity, familiarity, and access to categorical information [4].

Importantly, congenitally blind participants may benefit from explicit instruction when pictures depict 3D objects with foreshortening [4], suggesting that perspective-related phenomena are learnable but require targeted pedagogy.

A key conceptual hurdle is that haptic exploration does not naturally instantiate geometric perspective: objects do not shrink with distance for touch, and edges do not converge in the distance in the same way they appear to in a picture. Consistent with this, mental imagery studies have reported that congenitally blind participants often imagine objects within reachable space and do not reliably follow the "laws of perspective" (e.g., decreasing image size with distance) in imagery-based tasks [2].

Similarly, studies of tactile drawing production found that congenitally blind participants did not spontaneously produce foreshortened representations when asked to depict slanted surfaces, although they could improve in matching tasks after exposure [5].

These results align with a broader developmental perspective: visual parsing and grouping mechanisms that support rapid interpretation of cluttered scenes are learned through early visual experience; evidence from late sight restoration indicates that even with recovered vision, scene segmentation and parsing can remain impaired for a period, highlighting how demanding these mechanisms are and how experience-dependent they can be [9].

At the same time, there is robust evidence that blind individuals can effectively use tactile pictures and learn pictorial conventions under appropriate instructional conditions [4, 6].

This motivates a pedagogical strategy: provide tactile representations that (i) are designed to reduce cognitive load for sequential exploration, (ii) expose learners to pictorial depth cues in a graded, systematic sequence, and (iii) enable repeated practice with many exemplars.

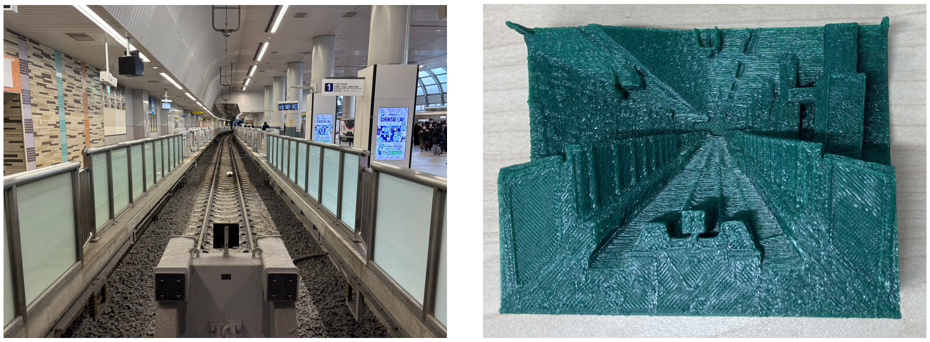
1.1 Formulating Problems Using Photographic Examples

When describing photographs verbally, rather than *merely listing shapes of objects in the frame* (Description Type A.), *the actual state of the photographed subject is described (as inferred)* (Description Type B.).

The photograph in Fig. 1(a) shows tracks within a large Japanese terminal station. The description of this photograph's main elements would typically be as follows.

Description of the Photograph in Fig. 1(a) Using Type B.

> A photograph taken within a Japanese terminal station, showing tracks extending from the foreground toward the direction of travel. A buffer stop is visible in the foreground, with two rails extending parallel toward the front. Platforms rise up parallel to the tracks on both sides. Fencings (platform doors) are installed at the boundary between the tracks and



(a) A photograph of tracks within a large Japanese terminal station.

(b) A relief printed by the Original Prusa MINI 3D printer [12] after converting the photo in Fig. 1(a) using our system.

Fig. 1: A photograph used as an example of description and its relief generated with the authors' technique

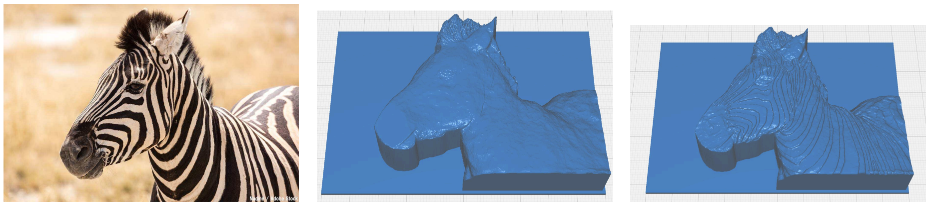
> the platforms on either side to prevent passengers from falling. The tracks curve gently to the right ahead.

This text describes the actual appearance of the photographed subject, following description Type B. In contrast, the literal description of the photographic graphics itself, i.e., description Type A., can be verbalized as follows:

Description of the Photograph in Fig. 1(a) Using Type A.

> A buffer stop is visible at the bottom of the frame, above which two rails extend toward the top, narrowing in spacing as they approach. Platforms are located on both sides of the rails, with fencing (platform doors) running along the outer edges of each platform. These platforms and fencing on both sides approach the rails at a steep angle, rising from bottom. The two rails and the platforms converge at a single point at center top in the image. They curve gently to the right at the top.

The two examples of expression above present a gap in how they express the shape and arrangement of subjects like rails and platforms. As mentioned above, description Type B. is the standard method for describing photographs in text, and descriptions for the visually impaired are also provided in this manner. For the visually impaired, who lack visual experience, the interpreted description in Type B. is extremely convenient. Conversely, this convenience highlights that for the visually impaired, imagining the actual scene based on the literal observation of the photograph, as seen in description Type A., is extremely difficult. Sighted individuals have empirically acquired the ability to derive an understanding of description Type A. from the perception corresponding to description Type B. The visually impaired, unable to rely on such experience, must bridge this gap



(a) The source photograph: a close-up of a zebra's face with its upper body.

(b) The generated relief without luminance representation: the general shape of the zebra resembling a horse.

(c) The generated relief with luminance representation: the general shape of the zebra.

Fig. 2: An examples of generation of 2.5D relief with luminance representation (reused from our paper [13]).

through intellectual operations, and this additional effort is substantial. Systematic teaching materials that enable the acquisition of skills to facilitate this intellectual operation are desired.

This issue has not been sufficiently addressed even within the context of assistive technology research. For example, Buonamici et al. propose providing context-appropriate audio guidance by recognizing the hand movements of visually impaired users exploring reliefs with an RGB-D camera [3]. This approach is promising as a means of multimodal information presentation for visually impaired people. However, they do not address the following points:

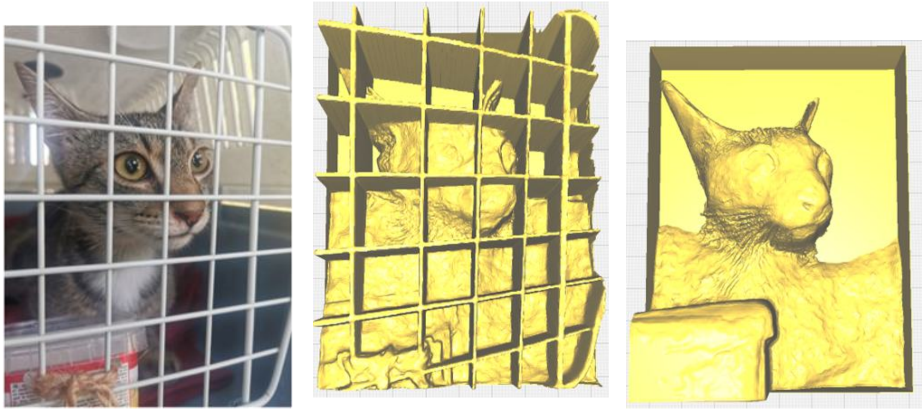
- that there are two types of audio guidance, both of which are forms of verbal explanation, as discussed here;
- that type (a) has largely been overlooked; and
- that if paintings or photographs are used as the source material for reliefs, type (a) must be explicitly taken into account.

1.2 Why Photographs (Not Paintings) for a General and Practical Route

Prior accessibility efforts in art and museums have often relied on tactile translations of paintings, including bas-relief renditions of perspectival artworks. Ansaldi articulated the importance of enabling blind learners to understand "painted spaces" and emphasized that faithful tactile copying is insufficient without an educational strategy that reveals the mechanisms of perspective [1].

However, the literature has offered limited guidance on a repeatable pipeline that can be used to generate large sets of pedagogically staged examples. Moreover, canonical paintings are not inherently optimized for teaching pictorial depth to novices, and they cannot be produced "on demand" to target specific cues in a curriculum.

We therefore propose using *photographs* as the source medium. Photographs can be staged and captured to exemplify particular spatial phenomena (e.g.,



(a) The source photograph: a cat inside a cage.

(b) The simply generated relief of a cat inside a cage. Tactile recognition of the cat is difficult.

(c) The generated relief of a cat inside a cage with our depth-based disassembling technique.

Fig. 3: An examples of generation of 2.5D relief with depth-based disassembling technique (reused from our paper [14]).

strong one-point perspective, controlled occlusions, repeated structures) at increasing levels of complexity. This supports *generality*: educators can construct a progressive learning set rather than relying on scarce, historically contingent artworks.

Photographs also support *practicality*: contemporary computer vision allows automated conversion from a photograph to a depth-based 2.5D relief, greatly reducing the manual labor typically required for tactile graphic production [8, 16]. The production bottleneck is a long-standing limitation in tactile materials; automated pipelines were explicitly motivated to reduce reliance on sighted experts and the time delays inherent in manual preparation [16].

1.3 Contributions

This paper contributes:

- A generalizable, photo-based pipeline that converts images to 2.5D tactile reliefs via depth estimation and 3D printing;
- A stepwise instructional design for teaching monocular perspective cues and extending toward binocular depth concepts;
- A related-work synthesis spanning tactile picture perception and photo-to-tactile conversion research;
- A mixed-method evaluation framework to assess educational efficacy and usability of tactile photographs.

2 Related Work

2.1 Tactile Picture Perception and Perspective Learning

Tactile picture perception research has established that raised-line pictures are useful for evaluating spatial cognition in congenitally blind people, while highlighting difficulty factors such as complexity and unfamiliarity [4].

Beyond recognition, perspective-related conventions are a special case: foreshortening is not a natural haptic invariant. Heller et al. reported that congenitally blind participants did not spontaneously produce foreshortened drawings of slanted panels, whereas late blind and sighted participants did [5]. Arditi et al. argued that congenital blindness is associated with imagery that does not incorporate perspective scaling with distance [2]. Taken together, these findings motivate explicit instruction and carefully designed representations that externalize the mechanisms of pictorial projection.

2.2 Over-convergence on Drawing Tasks in Prior Studies

A methodological issue in prior studies is over-convergence on the practical task, *"Can blind participants draw perspective pictures?"*. In fact, this task is intrinsically two-layered. First, participants must understand perspective as a representational system. Second, they must externalize that understanding through drawing skills and conventions. Conflating these two layers risks underestimating conceptual understanding when drawing performance is weak. As discussed by Ansaldi, perspectival painting is historically tied to specific Western artistic developments and should not be treated as the only universal endpoint of visual-spatial understanding [1]. Even sighted people who are fully familiar with visual perspective do not necessarily produce perspective-correct drawings in ordinary contexts. Therefore, introducing advanced drawing production as the primary benchmark can obscure the foundational educational question: whether visually impaired learners can understand perspective relations. In this study, we prioritize this foundational question and treat drawing production as optional downstream work.

2.3 From Photographs to Tactile Representations

Research has long sought automated conversion of visual images into tactile graphics. Way and Barner developed TACTICS and related "Automatic Visual to Tactile Translation" methods to convert digitized images into tactile form using image processing, explicitly addressing the time and expertise required for manual tactile preparation [16, 17].

More recently, photo-to-tactile work has moved from edge-only depictions toward semantically informed representations. Pakěnaitě et al. demonstrated that tactile depictions constructed from semantic/icon information extracted from photographs can significantly improve users' ability to describe photographic content compared to edge maps [10]. In follow-on work, Pic2Tac proposed an

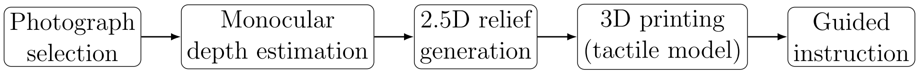


Fig. 4: Overview of our pipeline from a photograph to a tactile 2.5D relief used in guided instruction.

automated system translating photographs into accessible tactile images using semantic information from computer vision pipelines [11]. These approaches improve *what* information from a photo is communicated via touch; our work complements them by focusing on communicating *pictorial space* (depth and perspective) through a physically meaningful height field.

We build on our prior studies that explore photo-based tactile relief generation and tactile readability improvements. One study combined estimated depth with luminance-based modulation to expand expressive potential of tactile photography [13]. Another study investigated disassembling depth-estimation-based 2.5D reliefs to improve accessibility by reducing occlusion-related complexity [14].

We further note that remote education using tactile 3D models has been explored as a dissemination pathway, suggesting feasibility of distributing printed materials for guided instruction [7]. This practice is promising as an efficient and tangible means of delivering lectures on visual appearance to the visually impaired, who are a minority and often face mobility-related difficulties.

3 Method

3.1 Photo-to-2.5D Pipeline

Our pipeline comprises two main steps (Fig. 4):

Step 1: Depth estimation and 2.5D relief generation. Given a photograph, we compute a dense depth map using state-of-the-art monocular depth estimation models. The depth map is normalized and mapped to a relief height field with user-defined scaling (e.g., 0–5 mm). To improve tactile readability, we incorporate two techniques informed by prior work:

- **Depth + luminance relief:** add a small luminance-based height modulation (e.g., ±0.5 mm) on top of depth, to encode salient texture/contrast information without destroying depth ordering [13]. See Fig. 2.
- **Disassembly/occlusion reduction:** segment the scene into layers (foreground/midground/background) and optionally generate separate reliefs or remove obstructive foreground elements to reduce tactile clutter [14]. See Fig. 3.

The resulting height field is converted into a watertight mesh (e.g., STL) with a base plate and optional border.

Step 2: Fabrication via 3D printing. We fabricate the relief using consumer-grade FDM 3D printers.

The default footprint is chosen to support bimanual exploration (e.g., 120–160 mm) [15].

Production time is dominated by printing; the conversion process is largely automated. This improves scalability relative to handcrafted tactile translations, consistent with the long-standing motivation for automated tactile graphics [16, 8].

3.2 Instructional Design: Staged Learning from Monocular to Binocular Cues

We propose a staged curriculum:

- **Stage A (monocular cues):** size-distance relation, occlusion ordering, convergence and vanishing points using one-point perspective photos (roads, corridors), and repeated structures.
- **Stage B (composition under occlusion):** layered reliefs that separate foreground/background to explicitly teach how a 2D image can hide surfaces.
- **Stage C (toward binocular depth concepts):** conceptual introduction to disparity via pairs of tactile reliefs derived from two nearby viewpoints (or motion parallax sequences), emphasizing reasoning rather than sensory equivalence.

This design directly responds to the need for methodical, educationally oriented representations (rather than single-object tactile replicas) in perspective accessibility [1, 4].

4 Results

We report preliminary results that demonstrate feasibility of the proposed photo-based 2.5D relief pipeline. Because our primary goal in this paper is to establish a general and practical approach and its pedagogical rationale, we present these results as a formative demonstration rather than as a controlled user study.

4.1 Prototype Generation

We generated 2.5D tactile reliefs from photographs using monocular depth estimation followed by 3D printing (Fig. 4). We successfully produced reliefs for scenes with strong one-point perspective (e.g., railway tracks and platforms in Fig. 1(b)), layered occlusions (foreground objects partially covering background structures), and repeated structures (e.g., pillars). The resulting prints were robust to handling and suitable for bimanual exploration.

4.2 Readability-Oriented Variants

To reduce tactile clutter and support sequential exploration, we produced two variants described in Section 3. First, we generated depth+luminance reliefs to preserve salient photometric boundaries without disturbing the global depth ordering (Fig. 2). Second, we generated disassembled reliefs in which foreground obstacles are removed or separated to reveal occluded background structures (Fig. 3). These variants are intended to be used as paired materials within a lesson (e.g., background-only first, then full scene), matching the progressive learning strategy advocated in tactile picture perception research [4].

4.3 Formative Classroom Observations

We conducted formative, instructor-led trials in which a facilitator verbally guided tactile exploration while the learner handled the reliefs. Across these trials, we observed that (i) learners tended to begin by tracing large contours, then shifted to probing depth differences and repeated patterns, and (ii) layered/disassembled variants helped learners articulate “in front of / behind” relationships with fewer detours. These observations are consistent with the known sequential nature of haptic exploration and the benefit of reducing initial stimulus complexity [4, 16].

4.4 Results Summary and Limitations

Our preliminary results indicate that the proposed pipeline can reliably produce tactile materials that externalize pictorial depth cues in a form suitable for guided instruction. At the same time, we note limitations that motivate the qualitative evaluation described in Section 5. Depth estimation errors, depth compression inherent to 2.5D relief, and individual differences in tactile graphic experience can all affect interpretability. Accordingly, we focus our evaluation plan on qualitative evidence (think-aloud and interviews) to understand how learners construct meaning from tactile photographs and which design choices most strongly support comprehension.

5 Proposed Qualitative Evaluation Framework (Conceptual)

To evaluate the usefulness of 2.5D tactile photographs for visually impaired learners, we propose a qualitative framework. The goal is to capture interpretability, conceptual change, and learning processes that are difficult to reduce to a single numeric score.

5.1 Data Collection

We propose three complementary sources. (1) Think-aloud exploration sessions while participants read tactile photographs. (2) Semi-structured interviews immediately after each session. (3) Follow-up reflective interviews after a short interval to examine retention and transfer in daily explanation contexts.

5.2 Core Analytic Lenses

The analysis focuses on four lenses. **Perspective concept acquisition**: whether participants can verbally explain near-far relations, convergence, and occlusion logic. **Interpretation strategy**: how participants scan, segment, and re-check tactile evidence. **Breakdown points**: where confusion occurs (e.g., clutter, weak depth contrast, ambiguous boundaries). **Perceived educational value**: whether participants report improved confidence in understanding described photographs and paintings.

5.3 Analysis Procedure

Interview and think-aloud transcripts are coded using thematic analysis. We recommend two independent coders, iterative codebook refinement, and agreement checks. For design iteration, coded findings are mapped back to concrete model parameters (depth scaling, layer decomposition, luminance modulation, and scene selection).

5.4 Reporting outcomes

Instead of reporting only aggregate scores, we report representative learning trajectories, typical failure patterns, and design implications. This qualitative-first approach is consistent with our objective: establishing feasible pedagogy for perspective understanding before imposing advanced output tasks such as perspective drawing.

6 Discussion

Our proposal is motivated by two realities. First, congenitally blind learners may not spontaneously acquire pictorial conventions such as foreshortening and perspective scaling [2, 5]; second, tactile picture perception is feasible but demands careful design and instruction, especially for complex depictions [4]. A photo-based pipeline can generate many staged exemplars, enabling the repeated practice needed for concept acquisition. Compared to relying on fixed sets of tactile paintings, photographs support curriculum design (control over cues, progressive complexity) and scalable production through automation [16, 8].

Limitations include depth-estimation errors, depth compression constraints in 2.5D relief, and the intrinsic sequential nature of touch. Our disassembly and luminance-modulation techniques address some of these issues [14, 13], but further work is needed to standardize best practices for tactile photo readability and pedagogy. Finally, dissemination can be supported by remote instruction models where printed materials are shipped or locally fabricated [7].

7 Conclusion

We presented a general and practical route for teaching pictorial space to visually impaired learners using photo-based 2.5D tactile reliefs generated from depth estimation and fabricated via 3D printing. By shifting from canonical paintings to staged photographs, the approach supports systematic curriculum construction and scalable material production. We also proposed a qualitative evaluation framework to assess educational efficacy and usability. Future work will implement and validate the framework with congenitally blind participants and refine design guidelines for tactile photographs targeting perspective understanding.

Acknowledgments

This work was supported by KAKENHI (21H04419), JST RISTEX Grant Number JPMJRX21I5 and Canon Foundation.

References

1. Ansaldi, B.: Perspective and the blind: Towards a communication of painted spaces to the visually impaired. In: Cocchiarella, L. (ed.) ICGG 2018 – Proceedings of the 18th International Conference on Geometry and Graphics, Advances in Intelligent Systems and Computing, vol. 809, pp. 541–552. Springer, Cham (2019)
2. Arditi, A., Holtzman, J.D., Kosslyn, S.M.: Mental imagery and sensory experience in congenital blindness. Neuropsychologia 26(1), 1–12 (1988)
3. Buonamici, F., Carfagni, M., Furferi, R., Governi, L., Volpe, Y.: Are we ready to build a system for assisting blind people in tactile exploration of bas-reliefs? Sensors 16(9) (2016), `https://www.mdpi.com/1424-8220/16/9/1361`
4. Heller, M.A.: Tactile picture perception in sighted and blind people. Behavioural Brain Research 135(1–2), 65–68 (2002)
5. Heller, M.A., Calcaterra, J.A., Tyler, L.A., Burson, L.L.: Production and interpretation of perspective drawings by blind and sighted people. Perception 25(3), 321–334 (1996)
6. Kennedy, J.M.: Drawing and the Blind: Pictures to Touch. Yale University Press, New Haven (1993)
7. Minatani, K., Watanabe, T., Iwamura, M.: Online symposium with tactile 3d model: Its general efficacy as a media for the visually impaired. In: Proceedings of the 2025 PErvasive Technologies Related to Assistive Environments (PETRA) Conference (2025)
8. Mukhiddinov, M., Kim, S.Y.: A systematic literature review on the automatic creation of tactile graphics for the blind and visually impaired. Processes 9(10), 1726 (2021)

9. Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., Sinha, P.: Visual parsing after recovery from blindness. Psychological Science 20(12), 1484–1491 (2009)
10. Pakėnaitė, K., Hall, P., Rivas, P., Parraman, C.: Communicating photograph content through tactile images to people with visual impairments. Frontiers in Computer Science 4, 787735 (2022)
11. Pakenaite, K., Hall, P., Rivas, P., Parraman, C.: 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) (2024)
12. Prusa Research: Original prusa mini+. `https://www.prusa3d.com/product/original-prusa-mini-semi-assembled-3d-printer-enclosure-bundle-5/`, accessed 2026-02-12
13. Takaishi, K., Minatani, K., Iwamura, M.: A depth- and luminance-based 2.5d relief to expand the expressive potential of tactile photography for people with visual impairments. In: Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25) (Oct 2025)
14. Takaishi, K., Minatani, K., Iwamura, M.: Disassembling depth-estimation-based 2.5d photographic reliefs to improve accessibility for people with visual impairments. In: Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25) (Oct 2025)
15. Watanabe, T., Sato, K.: Suitable size of 3D printing architecture models for tactile exploration. The Journal on Technology and Persons with Disabilities 7, 45–53 (2019)
16. Way, T.P., Barner, K.E.: Automatic visual to tactile translation—part I: Human factors, access methods and image manipulation. IEEE Transactions on Rehabilitation Engineering 5(1), 81–94 (1997)
17. Way, T.P., Barner, K.E.: Automatic visual to tactile translation—part II: Evaluation of the TACTile image creation system. IEEE Transactions on Rehabilitation Engineering 5(1), 95–105 (1997)