15

Creating Accessible Online Floor Plans for Visually Impaired Readers

ANURADHA MADUGALLA and KIM MARRIOTT, Monash University
SIMONE MARINAI and SAMUELE CAPOBIANCO, University of Florence
CAGATAY GONCU, Monash University

We present a generic model for providing blind and severely vision-impaired readers with access to online information graphics. The model supports fully and semi-automatic transcription and allows the reader a choice of presentation mediums. We evaluate the model through a case study: online house floor plans. To do so, we conducted a formative user study with severely vision impaired users to determine what information they would like from an online floor plan and how to present the floor plan as a text-only description, tactile graphic, and on a touchscreen with audio feedback. We then built an automatic transcription tool using specialized graphics recognition algorithms. Finally, we measured the quality of system recognition as well as conducted a second user study to evaluate the usefulness of the accessible graphics produced by the tool for each of the three formats. The results generally support the design of the generic model and the usefulness of the tool we have produced. However, they also reveal the inability of current graphics recognition algorithms to handle unforeseen graphical conventions. This highlights the need for automatic transcription systems to return a level of confidence in the recognized components and to present this to the end-user so they can have an appropriate level of trust.

CCS Concepts: • **Human-centered computing** → **Accessibility systems and tools**; • **Computing methodologies** → Computer vision;

Additional Key Words and Phrases: Floor plans, visual impairment, navigation, trust

ACM Reference format:
Anuradha Madugalla, Kim Marriott, Simone Marinai, Samuele Capobianco, and Cagatay Goncu. 2020. Creating Accessible Online Floor Plans for Visually Impaired Readers. *ACM Trans. Access. Comput.* 13, 4, Article 15 (October 2020), 37 pages.
https://doi.org/10.1145/3410446

We acknowledge the support of Data 61 and CSIRO, which is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre for Excellence Program, and of the Australian Research Council through Linkage Project LP170100026, and Discovery Early Career Researcher Award DE180100057.

Authors' addresses: A. Madugalla, K. Marriott, and C. Goncu, Monash University, Australia; emails: {anu.madugalla, kim.marriott, cagatay.goncu}@monash.edu; S. Marinai and S. Capobianco, University of Florence, Italy; emails: {simone.marinai, samuele.capobianco}@unifi.it.

1936-7228/2020/10-ART15 $15.00
https://doi.org/10.1145/3410446

1 INTRODUCTION

Information graphics such as charts, diagrams, maps and plans are increasingly available online. At present, however, such web graphics are not easily accessed by visually impaired readers, as screenreaders are designed for textual content and fail miserably when graphics are involved [34]. While W3C guidelines [75] recommend that all online graphics are provided with an alternative textual description that can be read with a screenreader, this requires web authors to provide such a description. Studies measuring the accessibility of web graphics show that many images lack descriptions and those images that do have descriptions lack useful information [10, 59, 74]. Clearly, a better approach is required.

A number of researchers have addressed this problem. One of the more successful approaches is to use image processing techniques to automatically generate textual descriptions of simple online statistical graphics such as bar or line charts [15, 20]. However, this approach has limitations when used with other kinds of graphics such as plans or maps. First, such graphics are considerably more complex than simple charts, and so current automatic recognition techniques are error prone. Second, no matter how well-crafted, a textual description cannot capture the richness of spatial information present in a plan or map. For such graphics, accessibility guidelines recommend the use of representations such as raised line drawings called tactile graphics [51]. Therefore, in this article, we present a generic model to convert web graphics to accessible versions that are designed to address these two limitations. We then evaluate it through a case study: accessible presentation of online house floor plans.

Generic Model. Our approach to generating accessible online graphics uses the generic model shown in Figure 1. The model has two novel features. The first is that while the model is primarily designed to support fully automatic transcription, at the present time automated recognition is not totally accurate: it therefore supports optional semi-automatic transcription. In particular, the model allows a sighted friend, family member, or a colleague of a reader with severe visual impairment to correct errors in the automatic recognition of graphical components. Importantly, this is achieved by correcting the labeling of components in the original graphic rather than by directly modifying the generated accessible graphic. This obviates the need for the human editor to have expertise in transcription or the ability to read the accessible graphic: They simply need to understand the original graphic.

The second novelty in our model is that it supports presentation in different mediums. These mediums might include: a text-only description suitable for presentation on a screen or braille reader; a tactile graphic presentation, which requires access to a tactile graphic printer and paper; or a presentation suitable for exploration on a touchscreen device such as a smart phone or tablet. The users can choose the medium depending upon their preferences, cost, and availability of presentation devices.

The model has three main components that correspond to three main modules of the semi-automatic transcription tool for online floor plans described in Section 4:

(1) *Graphic Recognition System:* This identifies the graphical components in a raster graphic using image recognition techniques and produces a high-level semantic description of the graphic's components and their location.
(2) *Editor:* To support semi-automatic transcription, the model provides an editor that allows a sighted user to correct the recognition of the high-level components by identifying and labeling components in the original image.
(3) *Accessible Graphic Generation System:* This system takes the high-level description and generates an accessible graphic in one of the supported mediums.

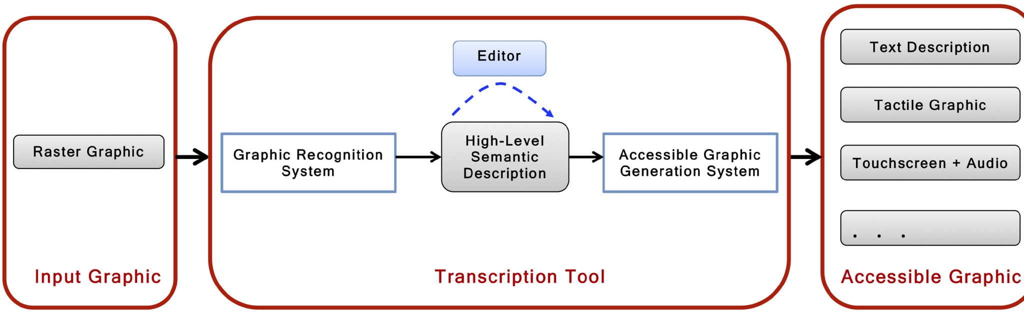


Fig. 1. Generic model for semi-automatic transcription of online graphics.

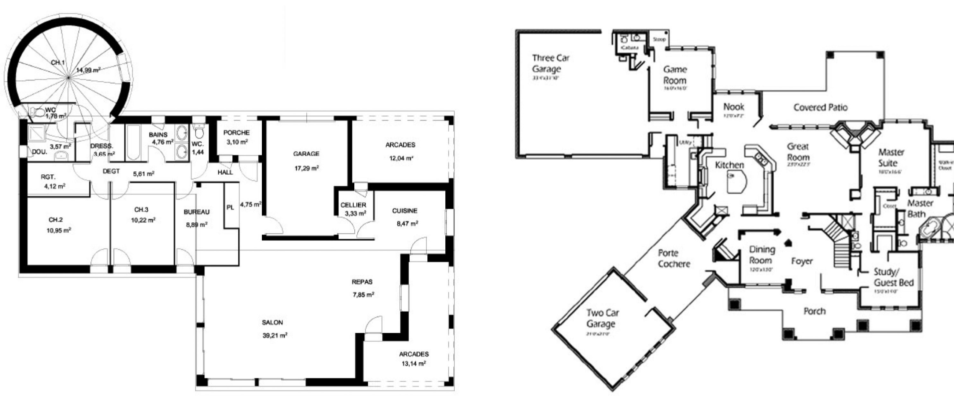


Fig. 2. As these two online floor plans show, floor plans are graphically complex and utilize a variety of different drawing conventions.

Case Study. While our generic model is applicable to any kind of online graphic including charts, maps, plans, or diagrams, at present there is no universal technique for recognition of all of these different kinds of graphics. We therefore chose to focus on a single type of online graphic as a case study to validate our model. We chose online floor plans. This was for two main reasons. First, floor plans are graphically complex, as shown in Figure 2, and so provide a good domain to test whether current image recognition techniques are efficient and accurate enough to support our model. Second, while there are currently tools supporting access by blind people to online charts and maps, as detailed in Section 2.3, there is a shortage of tools to support access to online floor plans. This is despite the usefulness of such plans for indoor navigation and for informing home rental and purchase. We chose to focus on house and apartment floor plans, as these are simpler. However, we believe our tool provides a good basis for providing accessible presentation of online floor plans of public spaces such as galleries, train stations, or shopping centers.

Contributions. The current article has four contributions: The first is the generic model for the semi-automatic transcription of accessible online graphics introduced above.

The second, and main contribution, is an in-depth evaluation of the model by application to a case study: online house plan transcription. To do so, we created a prototype tool using specialized graphics recognition techniques. This can automatically translate a raster house floor plan into three different presentation mediums: textual description, tactile graphic, and touch-controlled audio presentation on a touchscreen using the GraVVITAS app [26, 28]. We evaluated both the quality of graphics recognition using standard image processing benchmarks as well as conducted a user study with eight blind participants to evaluate the usefulness of the accessible presentations.

The presentation design was informed by an initial formative user study in which we asked five blind participants what information they would like to obtain from an accessible floor plan and asked for feedback on manually created floor plans in each of the three presentation mediums. Thus, a third contribution is to provide a better understanding of how to present floor plans to visually impaired users and the relative advantages and disadvantages of these three presentation mediums.

Our evaluation supports the utility of the generic model and its applicability to online floor plans. However, it also reveals the limitations of current graphics recognition techniques, in particular their brittleness and inability to handle unforeseen graphical conventions. An important finding of the research is the need to handle such uncertainty. Our fourth contribution is an initial exploration of the use of recognition techniques that return a level of confidence in the recognized graphic elements, and then appropriately present this uncertainty to the readers. This means that users are aware of the limitations of the recognition system and our user study suggests that at least to a limited extent they can use contextual information to deduce what the original elements might be.

A preliminary version of this article appeared in Reference [27]. The current version extends this in five main ways. First, the tool presented in Reference [27] only created presentations suited for GraVVITAS: text and tactile presentation were not considered. Second, the graphics recognition in the original tool was poor. The graphics recognition module has been completely rewritten and has considerably better recognition of windows, walls, and stairs as well as identification of functional sub-areas in open-plan areas. Third, the two user studies described here are new. Fourth, the evaluation of recognition accuracy is much more detailed and extensive, considering a second more difficult dataset. Finally, we consider the need to return levels of uncertainty.

2 BACKGROUND

We first review presentation methods for accessible graphics, methods for creating accessible graphics, and finally the current approaches to graphics recognition of floor plans.

2.1 Accessible Presentation Methods

Textual descriptions are the most common way of presenting accessible graphics. Crafting such a description is not straightforward because of the need to ensure that the descriptions are precise and complete yet do not overwhelm the user with too much information. Expert guidelines such as those of the American Council for the Blind [52] help. Textual descriptions can be either read with a screenreader or presented as braille text on a static or refreshable display.

Sonification, i.e., non-speech auditory presentation, is another approach. This is typically invoked as the user explores the graphic using a keyboard or touchscreen. Examples include: iSonic [81], which allows blind users to explore a table, a geographical map, or a scatterplot; Earth plus (Earth+) [50] for exploring satellite images and maps; TimbreMap [69]; and AudioGraf [39], which presents accessible graphs.

Tactile graphics are a third approach. The most common production methods are embossing with raised dots using a braille embosser, printing onto swell paper, or thermoforming. There is also interest in developing refreshable displays large enough to show graphical material. These include electromechanical displays HyperBraille [35] and Orbit [32], smart liquid displays Blitab [33] and electrodynamic surfaces [40, 72]. Directed kinesthetic presentation methods may also be used. PHANTOM [67] is one such solution that guides exploration in a 3D space using a stylus.

Hybrid touch-controlled audio approaches are also common. NOMAD [57] and Talking Tactile Tablet (TTT) [42] use a tactile graphic on top of a touchpad. As the user touches elements they receive speech and non-speech feedback. TeDUB [61] also combined tactile, audio, and haptic feedback to present technical diagrams such as UML diagrams. GraVVITAS [26, 28] is another solution. This supports multi-modal exploration of graphics on a touchscreen. It tracks finger movement providing both speech and non-speech feedback as elements are touched. An optional data glove provides vibration feedback.

2.2 Transcription

Transcription of accessible graphics can be manual, semi-automatic, or fully automatic. Creation of a tactile diagram usually requires an expert transcriber who has knowledge of transcription standards.

It is time-consuming and expensive, as manual transcription requires initial planning, drawing or tracing, image texturing, adding text, proofing at each stage, and then rendering the graphic to accessible presentation medium. For example, transcription of a mathematics textbook costs around AU$120,000 and takes several months [26]. Therefore, it is impractical for most online graphics. However, we note that crowdsourcing has been used to provide accessible descriptions of images [7, 73, 74].

In response to the high cost and length of time required for manual transcription, researchers have looked at how image processing can be used to semi-automate this process. Tactile Graphic Assistant (TGA) [36] and Infty [38] automate conversion of text to braille by using OCR (Optical Character Recognition) but graphical components of the image must be simplified manually with image editing software. In the case of maps, Bouhlel and Rojobi [8] use image processing techniques to simplify the map regions but still require an expert transcriber to assign textures to different image areas. While the system of Lareau and Lang [43] still requires manual intervention, this need not be by an expert. Text conversion to braille and some graphic simplification is performed automatically. However, for complex graphics, a sighted user is required to place landmarks on the graphic to guide graphic simplification.

Some researchers have tried to fully automate transcription of graphics. This can either be from the image itself or from the data underlying the image.

Wang and Barner [77] simplify the graphic using image processing methods, such as edge detection and blurring. This removes the complex features of the image, such as textures, filled areas, and gradients, and leaves a simple set of lines that can be presented as a tactile graphic. Ferro and Pawluk [21] suggest similar image processing methods to simplify graphics but do not present a complete system. Krufka et al. [41] explored automatic transcription for SVG, a vector graphic format. They simplify the graphic by removing fill and textures from the elements and use hierarchical information in SVG to identify child and parent elements, using this to determine line thickness in the final image. It is fair to say that while the foregoing generic approaches may work for simple graphics, they do not work for more complex graphics such as floor plans.

There has been more success in automatically generating textual descriptions for simple graphs and chemical formulas. iGraph-Lite [20] uses data values from an accompanying Excel sheet to identify features of the graph such as the key values and axis and these are then presented to visually impaired users as a summary or by using interactive text-based exploration. IGR [62] and Interactive SIGHT [15] use image processing methods to extract data from simple bar chart graphs that are presented to users as a textual description. Kekule [9] allows speech-based exploration of chemical molecules by visually impaired users based on information from a CML (Chemical Markup Language) file. MolRec [66, 68] provides similar functionality but takes an image of a structural formula as its input.

A number of systems generate accessible maps from GIS data. Wang et al. [76] present a static map and navigation instructions with IVEO [24] when users enter an origin and destination using data from MapQuest. The annotated tactile map system (ATMap) [78, 79] and that of Taylor et al. [71] provide similar functionality. SmartTactMaps [30, 31] uses GIS data to generate a 3D printed map with text labels as barcodes. The users are then able to explore the 3D map using a mobile phone that reads the barcodes. Maps are transcribed by Google to allow users to follow a route with applications such as Google's Walky-Talky [44] and Intersection explorer [45].

2.3 Accessible Floorplans

There has been considerable research into navigation aids for indoor navigation, e.g., Reference [23], including the use of tactile-audio maps for indoor navigation [55]. This has focused on larger public spaces, such as university campuses or shopping centers, rather than residential spaces.

Vision-impaired people have been surveyed to determine what information should be provided on outdoor [5, 64] and indoor maps, i.e., floor plans [56]. These reveal that entrances, stairs, elevators, toilets, doors, room numbers, steps, and hazards should be shown.

Generation of accessible floor plans was previously explored in the TeDUB project [60, 61]. In an initial requirements analysis [60] participants were asked what tasks they would use a floor plan for. They said that that they wanted an overview of the building layout, to understand the spatial relationship between rooms, an overview of the spatial layout of a particular room, and to plan a route through a building. The researchers also consulted architects on how to describe a floorplan. They recommended providing an overview including shape of the building, its orientation, and then describe the layout from the circulation area or main entrance. TeDB used image processing and OCR to generate a map view, spatial navigation view, walkthrough, route planner, and text view. These allowed the user to explore a kind of schematic network of the rooms in the house. It did not provide information about the shapes of rooms, location of windows, stairs, and so on.

Participants found it difficult to use the tool without help and asked for more geometric information, such as the shape of the building.

Paladugu et al. [53, 54] developed a system to automatically generate a textual description of a floor plan based on the position of text-labeled landmarks such as rooms. Based on an initial survey of vision-impaired people, this described the shape of the building, the position of the main entrance, and the position of landmarks relative to the main entrance. However, they also do not provide detailed spatial or geometric information and relied on text labels for room recognition.

Tang et al. [70] automatically generate 3D floor plans from floor plan images drawn with AutoCAD. They use layer information in AutoCAD to identify floor plan components and also use image processing methods to aid this identification. These are presented using a 3D model. The restriction to AutoCAD floor plans means that only a small percentage of online floor plans can be transcribed. Furthermore, they do not identify windows, stairs, or furniture.

Therefore, we can see that there is currently no tool that can handle common online raster graphic formats and generate an accessible version of a floor plan with sufficient spatial and geometric information about room shapes and location of windows, doors, stairs, and so on.

2.4 Automatic Floor Plan Recognition

The automatic conversion of floor plan raster images into a vectorial format has been the subject of research for many years now. The first approaches to floor plan analysis were proposed in the 1990s [4, 18, 65] with techniques closely related to the broader area of graphics recognition that include the recognition of technical drawings and other diagrams. More recently, a new wave of research on floor plan analysis has started with several approaches addressing the conversion of 2D floor plans into 3D models [1, 3, 6, 12, 13, 16, 46, 47].

Overall, different inputs are used, such as hand-drawn sketches, scanned input, or documents created in digital form. Floor plan analysis has been conducted for different aims such as robot navigation, 3D model generation [25], content based image retrieval, emergency evacuation simulation, Wi-Fi positioning systems [80], and pre-journey planning for visually impaired users using CAD floor plans [70]. The outputs of these systems have been also different, including JSON or text files, ontologies, and graphs for content-based image retrieval and 3D printable files for 3D model generation.

Despite this research, systems able to automatically recognize generic floor plans with high precision are still under investigation. Among other reasons for these difficulties are the lack of standard notation [12], the extreme variability of floor plan representation due also to the observation that architectural design is somehow a mix of both engineering and art [18].

Most systems for the automatic interpretation of floor plans (including the one described in this article) follow a single pipeline that has three main steps. (1) in the lexical level some pre-processing is performed on the input image, followed by primitive detection (e.g., connected components detection and vectorization); (2) in the syntactic level text/graphics separation (often addressed by relying on connected components in the image [22]) splits textual content from graphical content, allowing the system to recognize room labels and other textual information; (3) in the semantic level basic items are identified and recognized (text, wall, door, window, stairs) and at the end complete rooms are found.

A contribution of this research is to investigate to what extent pipelined based graphic recognition techniques can be used to automate the generation of accessible floor plans.

3 FORMATIVE USER STUDY

Our case study for the use of a generic semi-automatic transcription model was therefore online floor plans. We chose to explore three different presentation mediums: text, tactile graphic, and touch-controlled audio presentation on a touchscreen, as we felt each would be useful in some context.

While previous studies had surveyed blind people to ascertain what information they would like shown on outdoor and indoor maps [5, 56, 64], this had not considered the specific case of residential floor plans. And while Reference [60] had asked what purposes an accessible floor plan might be used for, they had not explicitly asked which elements should be shown. Therefore, as a first step, we conducted a formative study with five blind participants to:

(1) Ascertain the information that should be included in an accessible floor plan.
(2) Obtain feedback on initial designs of accessible graphics presented in the three targeted presentation mediums.

3.1 Materials

We manually created accessible versions of six different floor plans, one simple and one complex floor plan for each of the three presentation mediums. These were based on tactile guidelines and advice from an expert transcriber who is a member of the research team. The accessible versions of the simple floor plans are shown in Figures 3, 4, and 5.

Based on guidelines and discussions with the transcriber, and similarly to TeDUB [60], we decided to present each floor plan on two levels: an overview of the entire floor plan and detailed views of individual rooms. We created both levels for each of the three presentation mediums. The entire floor plan view contained information about the walls, doors, windows, stairs, and rooms but did not contain information about furniture and other objects, such as sinks, toilets, showers, and bathtubs. Instead, this was presented in individual room views. We also decided

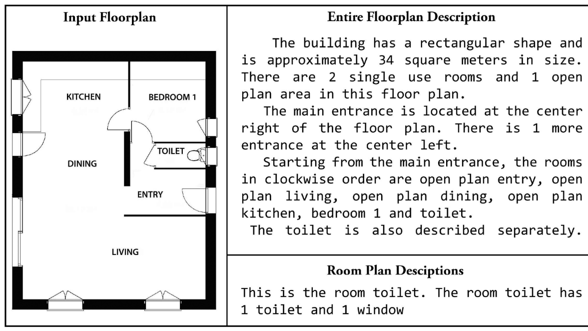


Fig. 3. A simple floor plan and its text description.

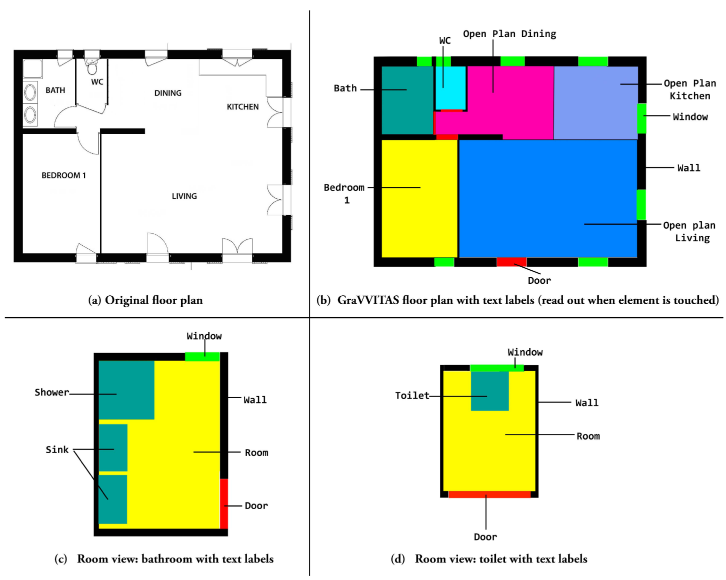


Fig. 4. GraVVITAS presentation of a second simple floor plan.

to provide individual room views only for those rooms that contain objects, such as toilets and sinks, to reduce information overload.

The text-only description was based on guidelines from the American Council for the Blind [52] and feedback from the expert transcriber. For the entire floor plan view, we started with general information and moved on to more specific details. We added building size, shape, and orientation information, followed by the number of rooms. The size was calculated using the same formula as described in Section 4.3. Then, we detailed how the entrance doors and rooms are structured. We described the layout of the rooms in the floor plan in clockwise order starting from the main entrance. Using the main entrance as the reference point was supported by the study of Petrie et al. [60] and by the expert transcriber. For the individual room view, we provided information about the number of doors, windows, and objects inside the rooms.

We used the North American Braille Authority guidelines [51] and feedback from the transcriber to guide the design of the tactile graphic versions of the floor plans. We provided a legend to enable the touch reader to understand the meaning of floor plan symbols and, as the guidelines

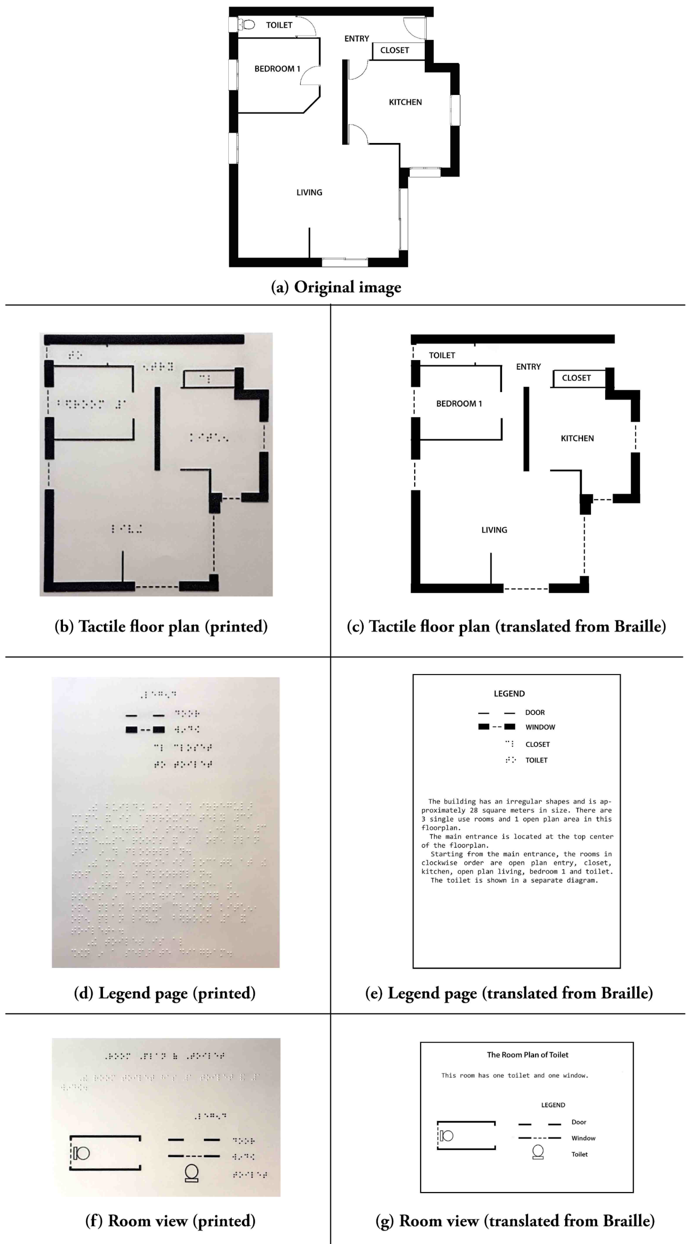


Fig. 5. Tactile presentation of a third simple floor plan.

recommend, added a textual description with an overview of the floor plan. The description was based on the textual description discussed above and converted to BRF (Braille Ready Format), then printed with a braille embosser.

We used an existing iPad app for touch-controlled audio presentation on a touchscreen, the GraVVITAS Graphics Reader App [29]. This supports speech and non-speech audio presentation of SVG graphics on iOS devices. It is integrated with Apple's VoiceOver technology. When the user first opens a graphic, an associated textual summary is read aloud; this summary can be accessed at any time using a two-finger swipe gesture. The graphic can then be explored using a single finger touch. When a graphic component is touched its associated metadata label is read, such as “wall” or “window.” Non-speech audio, such as a musical note, is also associated with each element. This continues to play while the reader touches the element and is louder near the boundary to signal that the edge is close by. In the GraVVITAS presentation, we used a summary for the entire floor plan and for each individual room based on the textual description described above.

A major decision in generating content for GraVVITAS was how to represent open plan areas. One approach was to represent them as a single polygon with a single room label for the entire open plan. However, this would not allow the users to understand the location and extent of the different functional sub-areas. In the tactile floor plans, this information was implicit because of the braille label positions. We, therefore, decided to partition each open plan into its functional sub-areas and associate a different color, sound, and audio label with each sub-area. The label gave each functional sub-area a name as well as an indication that it was part of the open plan.

3.2 Participants

We had five participants in this study. Three were totally blind and two legally blind. One had congenital vision loss and four had acquired vision loss. All were aged between 25 and 65. Two had interacted with tactile graphics before, while three had never used them. Two used touch reading in their daily life, while two preferred audio, and one used both touch and audio with no preference for either.

Participants were recruited using personal contacts of the research team and by advertising through social media and email groups. As discussed by Petrie et al. [58], finding participants for studies is difficult, in part because of the difficulty that blind people face in traveling.

3.3 Procedure

Due to the limited availability of participants, we conducted our study using phone calls as well as face-to-face. There were four face-to-face interviews and one phone interview. The study took about one hour per participant and had three parts.

In the first part of the study, we asked questions to find out what information participants would expect from an accessible floor plan. First, we evaluated their awareness and opinion of floor plans by asking if they had used accessible floor plans before and if they would like to use them in the future. Then, we asked a general question about what information they expected to find from a floor plan. Since most participants had not seen a floor plan before, they found this hard to answer. So, we prompted the discussion by asking specific questions, such as whether they would like to know about the shape, size, or orientation of the building.

In the second part of the study, we showed the four participants interviewed face-to-face the accessible floor plans described previously. As discussed, a simple and a complex floor plan were shown for each of the three presentation methods: text description, tactile floor plans, and GraVVITAS presentation. The order in which the presentation methods were shown was counterbalanced, but for each method the simple floor plan was presented first to allow participants to gain experience. Finally, we requested feedback on each presentation method, such as their preferences and how to improve the presentation.

In the case of the phone interview, we provided the textual description and explained the other two presentation methods in detail. This participant had prior experience in tactile graphics and was therefore able to imagine the tactile graphic. As for GraVVITAS, he was able to build an understanding, since he was familiar with iPad technology. We limited his feedback by not asking detailed questions about the presentation mediums he did not explore.

3.4 Results and Discussion

In the first part of our user study, we found that:

- While only one participant had seen a floor plan before, all thought they would be useful.
- All participants wanted to know about the orientation of the house, and most wanted to know house size and shape.
- All participants wanted to know about the number and type of rooms in the house, the boundaries of functional sub-areas in the open-plans, the spatial arrangement of the rooms and door, stair position. While some wanted to know the size of rooms and position of windows in the overview floor plan, others felt this amount of detail might be overwhelming.
- In the individual room view, all participants wanted to know about objects in the room, such as sinks, toilets, as well as connections to the other room. Most also wanted to know the room size and placement of windows.

In general these findings accord with Reference [60] about the usefulness of accessible floor plans and the need to provide information about spatial layout, entrances, orientation, and room connections as well as objects in rooms. They provide more detail about what objects to include, such as windows and stairs, and whether to include these in the entire floor view or in the individual room view.

In the second and third parts of the study, we discovered that the textual description was found to be useful by all participants. However, for descriptions such as room layout, they had to go through the same content several times to obtain an understanding. Some participants believed that more information such as room connectivity could be added to the entire floor plan description. Overall, all participants thought that even though the textual description was useful, it was best used only when no other method was available.

Tactile floor plans were explored by four users. Out of these, two participants could not read braille. Therefore, they were not able to use the braille text summary in the tactile floor plans. One of them had some vision, so she used vision combined with touch to read the tactile floor plan. The other participant who could not read braille explored the floor plan with his fingers. When his finger came across a braille room label, we aided him by reading out the content of the label. The two participants who could read braille felt the summary to be useful. They first briefly touch-scanned the floor plan to get an overview. Then they went through the summary to understand what content to find in the floor plan. Then they conducted a detailed exploration by referring to the text description from time to time, to verify their findings. They explained that having the summary helped them to "know what to search for." All participants were able to match legend symbols with floor plan content and they all believed having individual room views for rooms with objects was useful. Two participants believed that adding room size and room orientation could be useful. One participant commented:

> "Tactile floor plans are very clear. It's very cool to be able to see this much at once. I feel like I'll be able to walk into this house and find my way pretty easily."

GraVVITAS was used by four participants, and they all preferred to first listen to the summary and then to explore the floor plan. In the exploration, two participants first followed the walls to obtain an overview of the building structure and shape, then went on to explore the floor plan in detail. One participant explored randomly and after some time started following walls. The other participant had some vision but was color-blind. Therefore, she focused on each floor plan component by sight and listened to what she could hear when she touched it. All these participants found the summary to be useful, and they also found the partitioning of the open plan areas into their functional sub-areas to be useful. However, they had divided opinions on adding room size information to the sub-areas. One participant found it hard to coordinate audio cues with finger movements, but the other three were happy with the GraVVITAS presentation.

After participants explored the three presentation mediums and we had described the tactile and GraVVITAS mediums to the phone interview participant, we asked them to rank the methods. Out of the four face-to-face interviewed participants, two preferred tactile floor plans, one text description, and one GraVVITAS. Based on our descriptions of the methods, the participant who took part in the phone interview preferred tactile floor plans. In terms of their second preference, from the four face-to-face interview participants, two preferred tactile floor plans and two GraVVITAS. The phone interview participant chose text description. These preferences appeared to be based on the participants (lack of) familiarity with GraVVITAS and/or tactile graphics, their willingness to try new technologies, and a general preference by four of the five participants for more "graphical" representations, such as tactile graphics and GraVVITAS, over a textual description.

Therefore, as suggested in our general model in Chapter 3, providing them with the freedom to select their preferred medium was best. None of the presentation mediums was, however, regarded as useless, and everyone was able to find a medium that was most comfortable for them.

Our final question was when they thought accessible floor plans would be useful to them. All believed they could use them to understand floor plans in real estate websites when they are home hunting, for wayfinding when they go for inspections, and also when they are renovating their house and they wish to understand proposed plans.

4 TRANSCRIPTION TOOL

In this section, we briefly describe the semi-automatic transcription tool for online floor plans that was implemented as a case study for our generic model. In accord with the generic model, it has three components: the floor plan recognition system, the floor plan editing tool, and the accessible floor plan generation system.

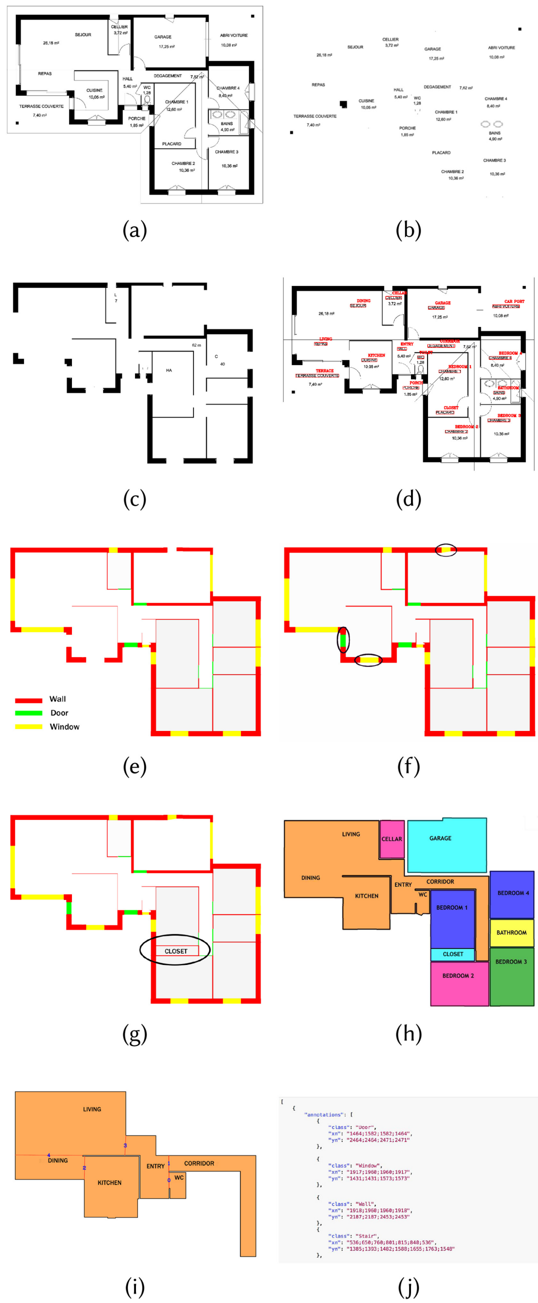
4.1 Floor Plan Recognition System

It is clear from the formative user study that vision-impaired users of accessible floor plans want detailed information about layout and location of virtually all of the components in the original plan. It is the role of recognition system to identify these components. The system is based on the standard pipeline approach described in Section 2.4. It is written in OpenCV and Python and consists of 21,500 lines of code. An example illustrating the nine main steps in recognition is listed below and shown in Figure 6.

(1) *Pre-processing:* The image is first cleaned by removing thin lines, as these typically represent texture or the roof position. The system then finds the connected components[1] (CC) in the floor plan image that are classified as either textual or graphical CCs considering their aspect-ratio and area. The textual and graphical CCs are shown in Figures 6(b) and 6(c), respectively. As we can notice some CCs are wrongly considered as text. Most of these are subsequently filtered out in text recognition.

(2) *Text identification and recognition:* Optical character recognition (OCR) is performed on the textual CCs to identify and recognize text labels. For the actual text recognition, the opensource Tesseract 3.03 is used. Labels are combined if they are sufficiently close to each other, e.g., "living" and "room" are combined to "living room." Symbol filtering and dictionary matching are used to check, correct, or ignore labels. The result of text

[1]A set of connected pixels in the image corresponding, e.g., to single characters.



(a) Input floor plan (b) Textual CCs (c) Graphical CCs (d) Text identification (e) Wall, door and window recognition (f) Closing external walls (g) Closet detection (h) Room identification (i) Open plan partitioning (j) Extract from the JSON file

Fig. 6. Main stages in floor plan recognition.

recognition is shown in Figure 6(d) with red text corresponding to recognized (and translated) labels.

(3) *Wall, door, window, and stair recognition:* This step classifies the graphical CCs. The largest CCs are labeled as containing walls. Those sub-parts of these CCs corresponding to walls are segmented and categorized as internal or external based on their average thickness (similarly to Reference [11]). Arc-shaped CCs of appropriate size are recognized as doors. Windows are found by looking for gaps between contiguous outer walls and stairs are found by searching for neighboring rectangles of similar size (corresponding to the stepped stair pattern). The identified walls, doors, and windows are shown in Figure 6(e).
(4) *Closing external walls:* Wall, door, and window recognition is not always perfect, and so the external wall of the building may contain gaps. These must be closed to properly detect rooms. We extended the method of Feltes et al. [19]. We identify gaps in the external wall and the disconnected wall ends are then connected using heuristics based on distance and geometry. The result is shown in Figure 6(f).
(5) *Closet detection:* The preprocessing step removes thin lines, which unfortunately means that objects like closets (PLACARD in Figure 6) may not be detected, as their boundaries are often drawn with thin lines. These objects are therefore recognized in a separate step. When recognizing text, labels such as "closet," 'cupboard," and "wardrobe" are identified and their bounding box computed. The boundary for these closet-like elements is computed by finding the smallest surrounding contour[2] in the original image for the element's bounding box. Using this process the closet is correctly recognized in Figure 6(g).
(6) *Room identification:* We used a similar approach to Reference [1] to recognize rooms. The OpenCV contour detection function is applied to an image containing the CCs of the walls, doors and windows. The result is filtered to remove contours that are too big or too small. The remaining contours are classed as room boundaries. Text labels are assigned to a room based on their location. A room is classified as a single-use room if it contains a single label and as an open plan area if it has multiple labels. The detected rooms are shown in Figure 6(h).
(7) *Open plan partitioning:* In the formative study, participants identified that they would like to know the extent of the functional sub-areas in the open plan regions. We developed a novel algorithm to perform this partitioning [49]. This generates candidate partitions for each subarea in the open plan. Candidate partition lines are scored based on their length and orientation and then the algorithm greedily selects the best partition line for each individual subarea iteratively. The result is shown in Figure 6(i).
(8) *Object/furniture identification:* To recognize common objects in floor plans, we implemented a simple adjacency graph of CCs based approach. We then created a dictionary of graphs for the objects and then searched each room graph for sub-graphs matching with a dictionary element [6]. Unfortunately, this approach was not successful and had very low accuracy. Thus, in the current version, the system does not use this step. A major focus of future research will be to improve object/furniture recognition by using deep-learning-based approaches [83].
(9) *High-level description generation:* The final step in floor plan recognition is to write a high-level description of the text, walls, doors, windows, stairs, single-use rooms, and open plan areas to a JSON file. A sample is shown in Figure 6(j). This file is the input for the accessible floor plan generation system and editing tool.

[2]A contour is a boundary of a connected component.

Even apart from object/furniture recognition, as we developed the floor plan recognition system it became obvious that the current graphics recognition software cannot provide 100% accurate, robust identification of floor plan components because of the widely varying conventions used in online floor plans.

We decided that the system should report its best guess at the components but also indicate a level of confidence to be reported to the vision-impaired. This ensures that they will have the appropriate level of trust in the tool's output and where necessary can use contextual information to check if the identification makes sense or deduce what an unknown element might be.

We decided to distinguish between three levels of certainty: high, medium, and low, rather than providing a numeric measure such as a percentage, as we did not feel that would be particularly meaningful. Elements with high certainty were reported as is, those with medium certainty had the label "maybe" appended to the identification, while those with low certainty were either ignored or identified as an "unknown element." This change affected the following steps:

(1) *Text identification:* Tesseract OCR provides a confidence level. If greater than 60% confidence was high, between 60% and 30%, it was medium and the text label had "maybe" appended to it, and less than 30% confidence was low and the text label was not reported.
(2) *Closing external walls:* When gaps are closed in the external wall, the area under the segment added to close the gap in the original floor plan was examined to see if it overlaid a wall, door, or window that was missed in the original identification step. Elements identified in this way with high confidence are reported as is, those with medium confidence have "'maybe" appended to the label, while those with low confidence are identified as an "unknown element."
(3) *Open plan partitioning:* Similarly after open plan areas are partitioned, the area under the partition lines is re-examined to see if this overlays a door missed in the original identification. Doors identified in this way with high confidence are reported as is, those with medium confidence have "maybe" appended to the label, while those with low confidence are not reported.
(4) *Room detection:* In room detection stage, the system checks if each room has a text label and, if not, adds the text label "Unknown element" to the center of the room.

4.2 Floor Plan Editing Tool

We have created a proof-of-concept editor for annotating floor plan images so a sighted user can correct recognition errors made by the floor plan recognition system. We decided to extend a predefined labeling system to describe images named Sloth (https://github.com/cvhciKIT/sloth). Sloth is a tool for labeling image and video data for computer vision research that makes it possible to define polygons or rectangular bounding boxes around the semantic parts contained in the image. We defined several semantic categories to label the floor plan image; in the end the user can save the created items in a structured JSON file. The tool can read the JSON file to initialize the annotation. In this use-case, the tool can take the JSON file produced by the recognition system and the original image and present the original image with graphic annotations showing the recognized components such as text doors, walls, windows, rooms, and so on. The user can then delete or add new annotations to correct the recognized components and the JSON file is updated accordingly. A screenshot of the tool is shown in Figure 7.

4.3 Accessible Floor Plan Generation System

This system generates an accessible floor plan from the information in the JSON file in one of three formats: text description, tactile floor plan, or GraVVITAS presentation. Based on the formative

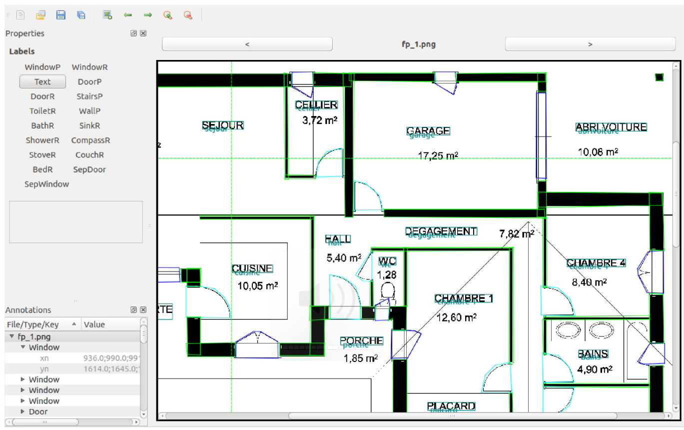


Fig. 7. The floor plan editing tool.

study, an entire floor plan view and individual room views are generated in the desired format. The entire floor plan view contains information about doors, windows, walls, stairs, rooms, text labels, and unknown elements in the floor plan. Information about furniture and other objects (sinks, toilets, showers, bathtubs) are presented in the individual room views. Individual room views are generated only for those rooms that contain objects.

Textual Presentation. A template slot-filling method [14, 63] is used to generate a natural language description of the floor plan components.[3] The "slots" in the templates are filled with values computed from the JSON file. The description is based on the American Council for the Blind guidelines [52]. Based on feedback from the formative study, we extended the textual presentation to include information about room connectivity and summary of the different number of rooms in each room type to the entire view and the connected rooms to the individual room views.

The textual description of the entire floor plan view details the following:

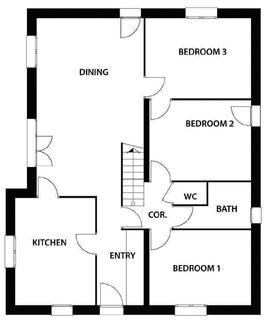
(1) Shape of the building: Rectangle, square, or irregular. Calculated using the ratio between building area and its bounding box area using threshold values.
(2) Approximate area of the building: Calculated using the formula:

$$Building\ Area(m^2) = \left(\frac{real\ world\ door\ width(m)}{floor\ plan\ door\ width(pixels)}\right)^2 \times floor\ plan\ area(pixels^2).$$

(3) Number of single-use rooms and open plan areas.
(4) Key room types and their count: e.g., number of bedrooms, toilets, bathrooms.
(5) Location of entrances: Main entrance location is followed by all other entrance locations. Entrances are detected by finding doors that are next to the building boundary. It is the main entrance if the door is connected to a room labeled "Entry" or "Living." The location of an entrance is determined by splitting the floor plan in to a 3 × 3 matrix and assigning position such as "Top left," "Top Center," "Top Right," and so on.
(6) Layout of the rooms: Starting from the main entrance, the external rooms are found by following the outer building boundary and the rest of the rooms that are not along the outer boundary are identified as internal rooms.

[3]Generation of human-like descriptions of spatial configurations is complex, e.g., Reference [17], but for our purposes a simple template approach was sufficient.

Input Floorplan



Floorplan Description

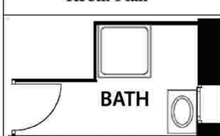
This is a house with 3 bedrooms, 1 bathroom and 1 toilet.

The building has a rectangular shape and is approximately 111 square meters in size. There are 7 single-use rooms and 1 open plan area in this floor plan. The main entrance is located at the bottom center of the floor plan. [There is 1 more entrance at the top center.]

Starting from the main entrance, the rooms in clockwise order are open plan entry, kitchen, open plan dining, bedroom 3, bedroom 2, bathroom and bedroom 1. [1 corridor and 1 toilet are situated internally.]

The open plan entry is connected to outside, corridor and kitchen. The kitchen is connected to open plan entry and open plan dining. The open plan dining is connected to outside, bedroom 3 and kitchen. The bedroom 3 is connected to open plan dining and bedroom 2. The bedroom 2 is connected to corridor and bedroom 3. The bathroom is connected to corridor. The toilet is connected to corridor. The bedroom 1 is connected to corridor.

Room Plan



Room Description

This is a bathroom plan. It has 1 shower, 1 sink and 1 window. The room door leads to the corridor.

Fig. 8. Example text description.

(7) Room connections: We identify the two rooms connected by each door. These connections are then processed to generate room connection information.

The textual description of the entire floor plan does not contain information about windows, as in our formative user study some participants feared it would be overwhelming. Instead, as most participants requested, information about windows was added to the individual room views.

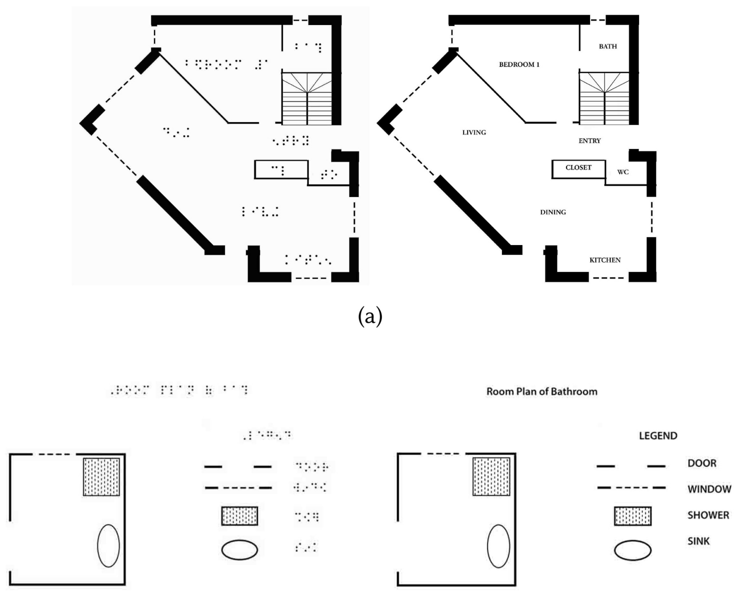
The individual room view contains:

(1) Name of the room
(2) Number and kind of objects in the room
(3) Number of windows
(4) Connected rooms

Examples are shown in Figure 8.

Tactile Graphic Presentation. The tactile graphic presentation was informed by the North American Braille Authority guidelines [51], discussions with an expert transcriber, as well as feedback from the formative study. Generation of the entire floor plan view is as follows:

(1) *Walls, doors, windows*: Add pre-defined tactile symbols to represent these components based on the location in the JSON file



(a)

(b)

(a) Entire floor plan (with original braille and text translation) (b) Single room plan (with original braille and text translation)

Fig. 9. Example tactile floor plans.

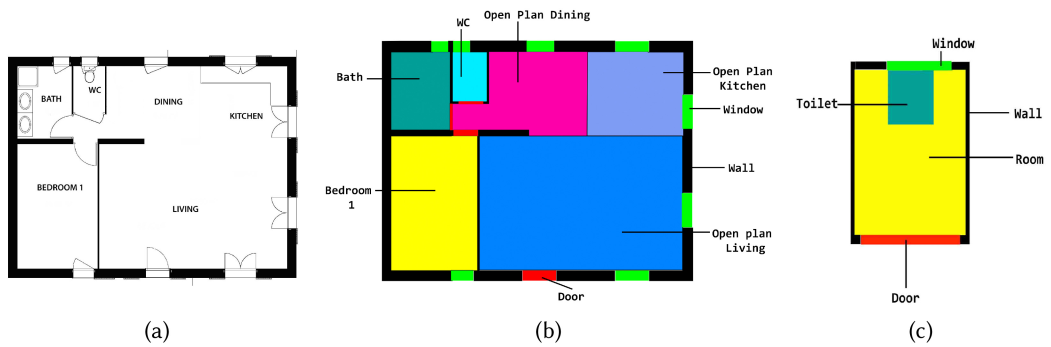
(2) *Text labels*: Find matching braille labels for room names and place them at a suitable position inside the room
(3) *Textual overview*: Simplify the textual description given above by removing information about room connections and convert it to Braille Ready Format (BRF).
(4) *Legend*: Based on symbols used in the plan, generate a legend.
(5) *Combine Content*: Map the content to fit on A4 tactile paper size.

Individual room view generation is similar. Graphical output is suitable for printing on a **PIAF (Pictures In A Flash)** machine with text printed using a braille embosser. An example is shown in Figure 9.

GraVVITAS Presentation. GraVVITAS displays SVG graphics with a specialized format. The entire floor plan view is generated as follows:

(1) *Walls, doors, windows, rooms, functional sub-areas in open-plan areas*: Add these to the SVG file as polygons.
(2) *Audio-feedback*: For each polygon add metadata specifying the text to be read out when the user touches it and an ID that is linked to a color and audio file to be played while the user's finger is on the element.
(3) *Description*: Add overview text description to the metadata.

The process is similar for individual room views. An example is shown in Figures 10(b) and 10(c).



(a) Input floor plan (b) GraVVITAS floor plan (c) GraVVITAS room plan

Fig. 10. Example GraVVITAS-based floor plans.

5 EVALUATION 1: FLOOR PLAN RECOGNITION ACCURACY

We evaluated our floor plan transcription system in two ways. In the first evaluation, reported in this section, we measured recognition accuracy of the floor plan recognition system. In the second evaluation, reported in the next section, we tested usability of the generated accessible floor plans with vision-impaired users.

5.1 Datasets

Floor plan recognition accuracy was evaluated with two datasets. Measurements of recognition accuracy rely on the existence of a *ground truth* in which floor plan components in the dataset images, such as walls or doors, have had their boundaries identified and been annotated with their correct classification. Fortunately, the floor plan editing tool can be used for this purpose. Initial identification can be performed by the recognition system, and then the floor plan editing tool can be used to manually correct the identification and so produce the ground truth.

UAB-CVC Dataset. The UAB-CVC corpus is the standard dataset used for evaluating floor plan recognition [2, 12, 47]. It contains 90 raster-based floor plans. Rooms have text labels and contain some furniture items. The plans use consistent graphical conventions, as they are from a single architectural office. The dataset comes with manually created ground truth for doors, windows, walls, and rooms. We added the missing ground truths for text labels and stairs using the floor plan editing tool.

RSVG Dataset. The recognition system was designed for the UAB-CVC dataset. We also wished to test the system with a dataset that is more representative of the variety found in real-world floor plans. To do so, we created the **Rasterized SVG (RSVG)** dataset. We collected 100 floor plans from the internet ensuring that the dataset was diverse by selecting floor plans from many different websites and manually checking that the floor plans used different color schemes, textures, notations, and words to label rooms. We downloaded the floor plans in SVG format and then converted them to high-resolution raster images. Ground truths were created using the floor plan editing tool. This new dataset and ground truth are freely available online [48].

5.2 Quantitative Evaluation

We evaluated the floor plan recognition system on both the UAB-CVC and RSVG datasets by comparing the detected elements with the ground truth. We did not evaluate the recognition of furniture and other objects because currently this step has very low accuracy and so is not used.

Table 1. Precision and Recall of Door, Window, Stair, and Text Recognition

Element	Total Number	Precision %		Recall %	
		High Confidence	Medium Confidence	High Confidence	Medium Confidence
UAB-CVC					
Door	768	91	30	79	1
Window	802	86	62	81	4
Stairs	49	76	NA	86	NA
Text Labels	1,123	98	0	89	0
RSVG					
Door	978	76	50	36	0
Window	1,011	56	43	53	1
Stairs	80	61	NA	58	NA
Text Labels	1,574	82	0	55	0

Total Number gives the number of instances in the dataset. Precision for High (Medium) Confidence is the percentage of elements identified with high (medium) confidence that are correctly identified, i.e., match ground truth elements. Recall for High (Medium) Confidence is the percentage of ground truth elements that are recognized with that confidence. Note that stairs are only ever recognized with high confidence.

Recognition was evaluated using two standard measures: *precision*, which is the proportion of recognized objects that have been correctly recognized, and *recall*, which is the proportion of objects in the original image that have been correctly recognized. In our case it is a little more complex, because there may be different confidence levels for the recognized objects. Thus, we separately report precision and recall for objects recognized with high confidence and with medium (maybe) confidence.

Doors, Windows, and Stairs. Doors, windows, and stairs in the ground truth were matched with the system detected components using the Jaccard Index (or "intersection over union"). This is widely used in image recognition to compare the similarity of two areas (e.g., the detected object A and the corresponding ground-truth B):

$$JI_{A,B} = \frac{A \cap B}{A \cup B}.$$

The Jaccard index is the ratio between the intersection and union of the two areas and we accept a match between a system-detected component A and a ground truth component B if $JI_{A,B} > thres_J$ where $thres_J$ is a threshold of 0.6.

Table 1 gives precision and recall for door, window, and stair recognition. For the UAB-CVC dataset precision and recall for door and window recognition is high. Precision for stair recognition is slightly lower, as tiled patterns are occasionally recognized as stairs. As shown in the table for the UAB-CVC dataset, total recall, that is, recognition with either high or medium confidence, for door recognition is 80% and for windows 85%. This same dataset was used by Gimenez et al. [25]. They report door recall of 60% and window recall of 61%, which shows that our system is more than comparable with prior art. They do not give figures for precision. Both precision and recall are much lower for the RSVG dataset. This is not surprising, given that the recognition software was not designed for the widely varying notational conventions used in this set.

Text Labels. We used position to match system-detected text with ground truth text. The distance from system-detected texts' centroid to ground truth texts' centroid was used to measure the similarity of the position. A text label was only considered as a match if it was identical to the ground truth text.

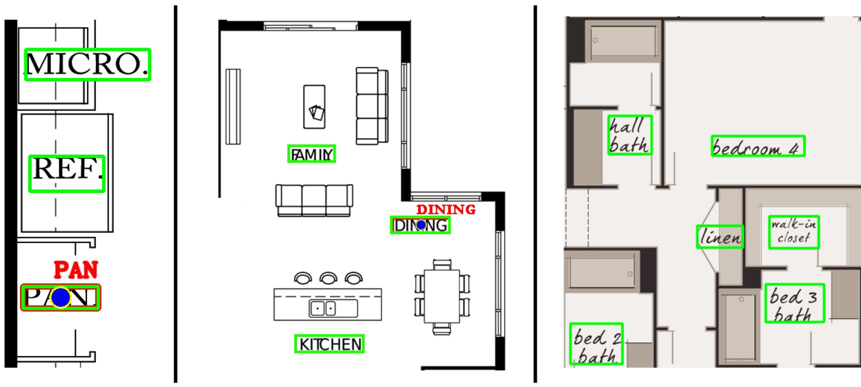


Fig. 11. Text recognition errors: (a) Dashed lines intersecting, (b) Tight letter spacing, (c) Unusual font.

Table 2. Precision and Recall of Wall Recognition

Wall	Precision %	Recall %
UAB-CVC	96	98
RSVG	85	86

See text for more details.

Table 1 gives also the precision and recall for text label recognition. For the UAB-CVC dataset, precision is almost perfect and recall is very high. For the RSVG dataset, precision remains high but recall drops markedly. This is due to the wide variability of text labels in the UAB-CVC corpus. Figure 11 shows representative reasons: text labels overlapping other elements such as dashed lines and unusual spacing of letters or font.

Walls. We measured precision and recall of wall recognition using the formulae:

$$Precision = \frac{W_{sys} \cap W_{gt}^{\star}}{W_{sys}}, \qquad Recall = \frac{W_{sys}^{\star} \cap W_{gt}}{W_{gt}},$$

where W_{sys} is the area of the image recognized as being a wall by the system, W_{gt} is the area of the image marked as a wall in the ground truth and the area dilation operator "$\star$" is used to allow for small differences between the two. We used a dilation kernel of 5×5 pixels. This was computed for each image and the average for each dataset is reported in Table 2. We see that precision and recall are almost perfect for the UAB-CVC dataset and remain high even for the RSVG dataset.

Rooms. Finally, we considered precision and recall of single-use rooms and functional subareas of open plan rooms. These matched if the corresponding room delimited by a polygon had an area of intersection above a threshold and the labels matched both position and text. For the figures for recall, we also considered matches in which the polygons matched but not the text or a single-use room was misidentified as being an open plan subarea or vice versa.

Table 3 shows precision and recall for single-use room recognition. We see that for the UAB-CVC dataset, precision is reasonable, and examination of the errors revealed that a common reason for incorrect identification was detecting the wrong text label or identifying a functional sub-area in an open plan area as a single-use room. In practice, recall was also reasonable, given that most single-use rooms were recognized though often as being part of an open plan rather than as a single-use room. Precision and recall for the RSVG dataset was poor.

Table 3. Precision and Recall of Single-use Room Recognition

Rooms	Total Number	Precision %			Recall %			
		Full match	Only polygon match	Total	Full match	Only polygon match	OP sub-area	Total
UAB-CVC	552	80	10	90	54	7	31	91
RSVG	509	12	15	27	10	13	7	30

In addition to full match in precision, we give the percentage of rooms identified with matching polygons but incorrect text labels, and in recall as well full match, we give figures for matching polygons but incorrect text label and incorrect identification as being part of an open-plan subarea.

Table 4. Precision and Recall of Functional Subarea of Open-plan Room Recognition

OP Sub-Areas	Total Number	Precision %	Recall %		
		Full match	Full match	Single use room	Total
UAB-CVC	314	54	76	20	96
RSVG	791	46	10	16	26

As well as full match in recall, we give the percentage of partial matches in which the functional subarea was incorrectly identified as a single-use room.

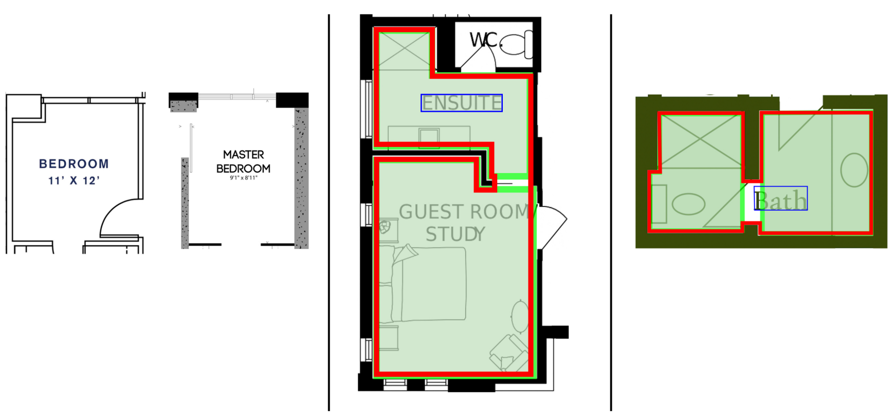


Fig. 12. Room recognition errors: (a) Different walls, (b) Undetected text label, (c) Missing text label.

Table 4 shows precision and recall for recognition of functional subareas in open plans. We see that for the UAB-CVC dataset, precision is quite low; this was because many single-use rooms were incorrectly identified as being a functional sub-area in an open plan area. In practice, recall was also reasonable, given that most single-use rooms were recognized though often as being part of an open plan rather than as a single-use room. Again precision and recall for the RSVG dataset was poor.

Poor recognition of single-use and open-plan rooms in the RSVG dataset is due to two main reasons. First is a low recognition rate for walls drawn with different filling and texturing methods, as shown in Figure 12(a). Accurate wall recognition is necessary for room identification. The second reason is missing or undetected text labels. During open-plan partitioning the system identifies and decides to partition an open plan only if a room has more than one text label. Therefore, when text labels are undetected 12(b) or missing in the original image itself 12(c), the system will not partition the open-plan area, treating it as a single-use room.

(a) Input floor plan (b) Textual description (c) Tactile floor plan (Braille text replaced with English) (d) GraVVITAS floor plan

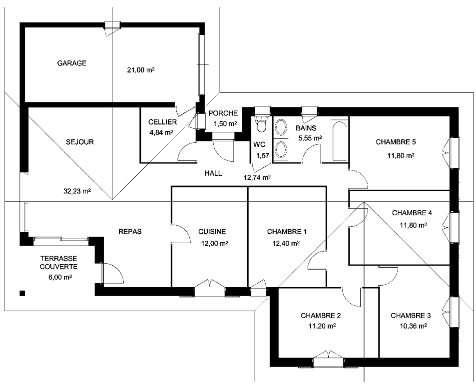
Fig. 13. Sample 1: CVC dataset.

5.3 Qualitative Evaluation

While the quantitative calculation of precision and accuracy for the basic graphic elements in a floor plan gives some idea of the usefulness of the tool, it does not really give a feel for how useful the tool might be in practice. To do so, we present a more qualitative study of the system output. We present four representative floor plans: two floor plans from each corpus, one showing good recognition, one showing poorer recognition. The results are shown in Figures 13, 14, 15, and 16.

Figure 13 is an example of perfect recognition from the UAB-CVC floor plan dataset. The graphic recognition module has recognized all floor plan components successfully and created an accurate text description, tactile floor plan, as well as GraVVITAS floor plan. All closets are recognized correctly and open-plan partitioning is effective as well. This will undoubtedly be useful for visually impaired users.

Figure 14 is an example of a floor plan from the UAB-CVC that has detection inaccuracies. A wall segment between dining and living is missed and this has led the system to inaccurately add



(a)

Floorplan Description

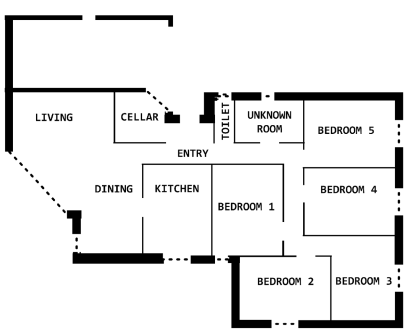
This is a house with 5 bedrooms 1 toilet, 1 bathroom and 1 unknown room.

The building has an irregular shape and is approximately 134 square meters in size. There are 9 single-use rooms and 1 open plan area in this floorplan. The main entrance is located at the top center of the floorplan.

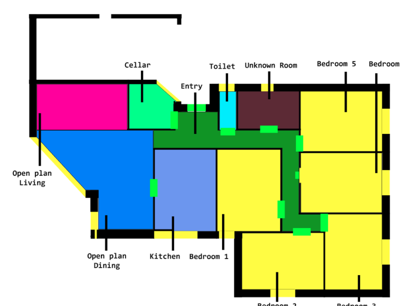
Starting from the main entrance, the rooms in clockwise order are open plan entry,toilet,unknown room,bedroom 5, bedroom 4, bedroom 3, bedroom 2, bedroom 1, kitchen, open plan dining, open plan living and cellar.

The open plan entry is connected to outside and to toilet, unknown room, bedroom 5, bedroom 4, bedroom 3, bedroom 2, bedroom 1 and cellar. The toilet is connected to open plan entry. The unknown room is connected to open plan entry. The bedroom 5 is connected to open plan entry. The bedroom 4 is connected to open plan entry. The bedroom 3 is connected to open plan entry. The bedroom 2 is connected to open plan entry. The bedroom 1 is connected to open plan entry. The kitchen is connected to open plan living. The open plan dining is connected to kitchen. The cellar is connected to open plan entry.

(b)



(c)



(d)

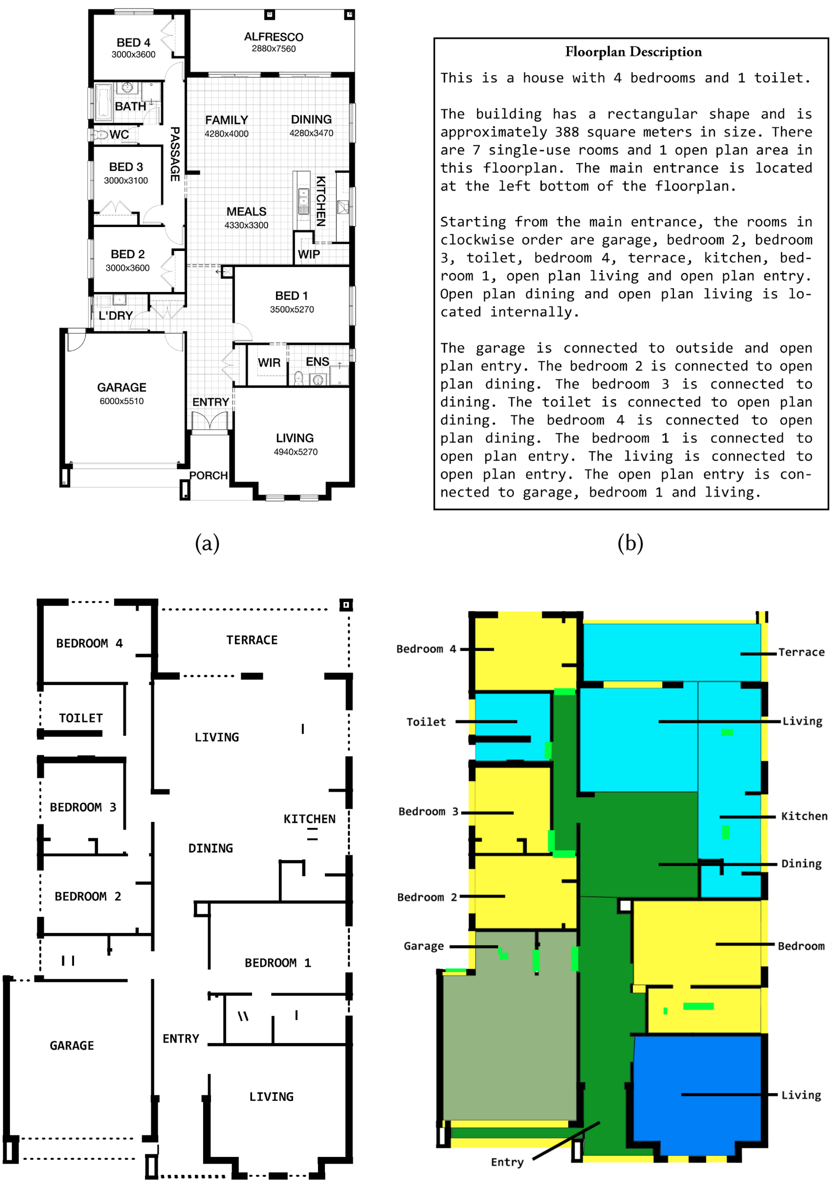
(a) Input floor plan (b) Textual description (c) Tactile floor plan (Braille text replaced with English) (d) GraVVITAS floor plan

Fig. 14. Sample 2: CVC dataset.

a window connecting the wall segments that were recognized. Similarly, a wall segment is missed next to cellar, causing the system to close the outer boundary by creating a connection between the identified wall segments. This connection is categorized as a window because of the length and the black pixel ratio in that connection area. The bathroom text label has not been recognized, and so this is marked as an "unknown" room. However, in spite of these inaccuracies, the main entrance is detected correctly, and the general shape of the building and most rooms and their layout and connections have been correctly recognized.

The accuracy of recognition of the other floor plans in the UAB-CVC dataset falls between these two examples, so it is fair to say that the recognition quality is high enough to provide useful information in the accessible versions for people who are vision-impaired.

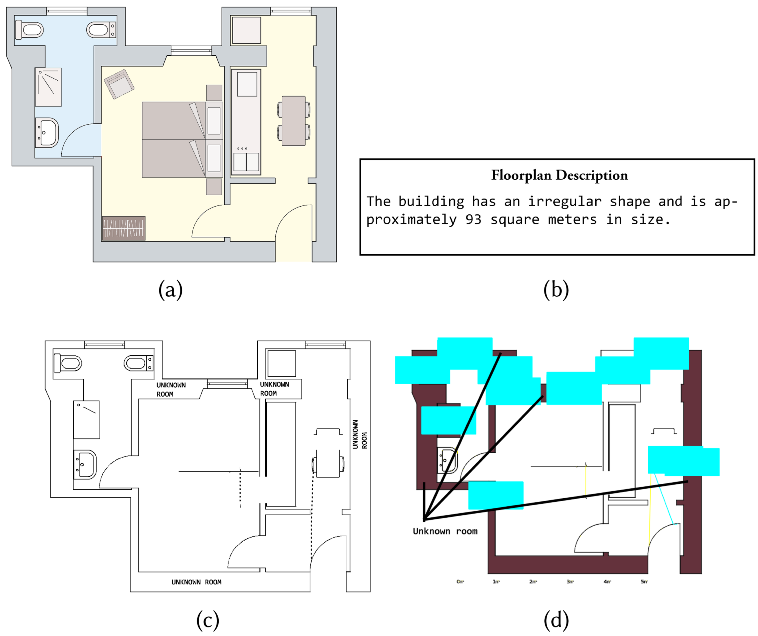
Figure 15 shows a sample from the RSVG dataset for which floor plan component recognition has performed fairly well. Most of the walls and windows are detected accurately. However, some floor plan areas are inaccurately assigned as windows. Some text has been missed, leading to subsequent errors. For instance, because "Bath" has been missed, the toilet and bathroom are recognized as one room "toilet." This means that the system does not realize that it should look for a door to



(a) Input floor plan (b) Textual description (c) Tactile floor plan (Braille text replaced with English) (d) GraVVITAS floor plan

Fig. 15. Sample 1: RSVG dataset.

the bathroom. Some components of the text labels have not been identified as textual connected components and so have been recognized as doors. However, despite these issues, the resulting accessible floor plans show most floor-plan components and the overall structure reasonably well and should be useful to a vision-impaired user.



(a) Input floor plan (b) Textual description (c) Tactile floor plan (Braille text replaced with English) (d) GraVVITAS floor plan

Fig. 16. Sample 2: RSVG dataset.

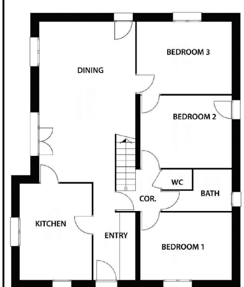
However, the example shown in Figure 16 is an example of a floor plan from the RSVG dataset where unusual graphical conventions mean that the system performs very badly. This floor plan doesn't have any text labels and has massively thicker walls than the other floor plans. The system has mistakenly identified the thick walls as rooms and has assigned them the text label "unknown room." However, even with this example, as you can see from the tactile floor plan, the detection achieved by the system can still help a person with vision impairment understand the general building shape and size. However, here it is clear that having a sighted person use the editing tool to correct the recognition errors would greatly improve the output from the system.

Recognition accuracy of the other floor plans in the RSVG dataset falls between these two examples, so it is fair to say that while the recognition quality definitely needs improvement, in most cases it is high enough to provide useful information in the accessible versions for people who are vision-impaired.

6 EVALUATION 2: USER STUDY

We evaluated the usability of accessible floor plans generated by our system. We did this by asking participants with severe vision impairment to use floor plans generated fully automatically by the system for route finding and answering some general questions about layout. In addition, we wished to obtain feedback on depicting uncertain information in accessible floor plans. As discussed above, the floor plan recognition system recognizes components such as walls, doors, windows, and text labels with different levels of confidence. In our formative user study, we had not considered this, as we had optimistically thought that recognition would be more accurate and robust. Therefore, in this second user study, we also showed floor plans in which some of

Input Floorplan



Floorplan Description

This is a house with 3 bedrooms, 1 bathroom and 1 toilet.

The building has a rectangular shape and is approximately 111 square meters in size. There are 7 single-use rooms and 1 open plan area in this floor plan. The main entrance is located at the bottom center of the floor plan. There is 1 more entrance at the top center.

Starting from the main entrance, the rooms in clockwise order are open plan entry, kitchen, open plan dining, bedroom 3, bedroom 2, bathroom and bedroom 1. 1 corridor and 1 toilet are situated internally.

The open plan entry is connected to outside, corridor and kitchen. The kitchen is connected to open plan entry and open plan dining. The open plan dining is connected to outside, bedroom 3 and kitchen. The bedroom 3 is connected to open plan dining and bedroom 2. The bedroom 2 is connected to corridor and bedroom 3. The bathroom is connected to corridor. The toilet is connected to corridor. The bedroom 1 is connected to corridor.

Fig. 17. Floor Plan 1 and its textual description.

the information was shown as uncertain to obtain feedback on the representations of uncertain information.

6.1 Participants

Eight participants completed the study: seven were totally blind and one was legally blind. Five had congenital vision loss and three had acquired vision loss. All were aged between 18 and 61, three were male while the other five were female. Seven had prior experience with tactile graphics, while one had never used them. All participants used audio to access text and four used braille as well. Two participants had taken part in the formative user study. The others were found using personal contacts of a research team members and through the participants in the formative study. We used different floor plans from the first user study to minimize the impact of prior experience.

6.2 Procedure and Materials

We carried out our study by conducting face-to-face interviews with the participants. The study had two parts and took approximately one hour per participant.

In the first part, we presented two accessible floor plans fully automatically generated by our system. Because of the system's limited object recognition capability, we did not generate individual room views. Floor Plan 1 was presented as a text description and Floor Plan 2 either as a tactile graphic or by using GraVVITAS—see Figures 17, 18, and 19. While we would have preferred to show each participant a floor plan in all three modalities, we chose to show only two modalities to allow the study to be completed in one hour. Participants used their own device to access the text description, such as a screenreader or braille reader. We presented the tactile floor plan to participants who were fluent in braille and GraVVITAS to the participants who were comfortable using an iPad. We ensured, however, that these two presentation modalities were presented to an equal number of participants.

Participants were presented with the textual description of Floor Plan 1 and then the tactile or GraVVITAS version of Floor Plan 2. They were asked to explore each floor plan. Then they were

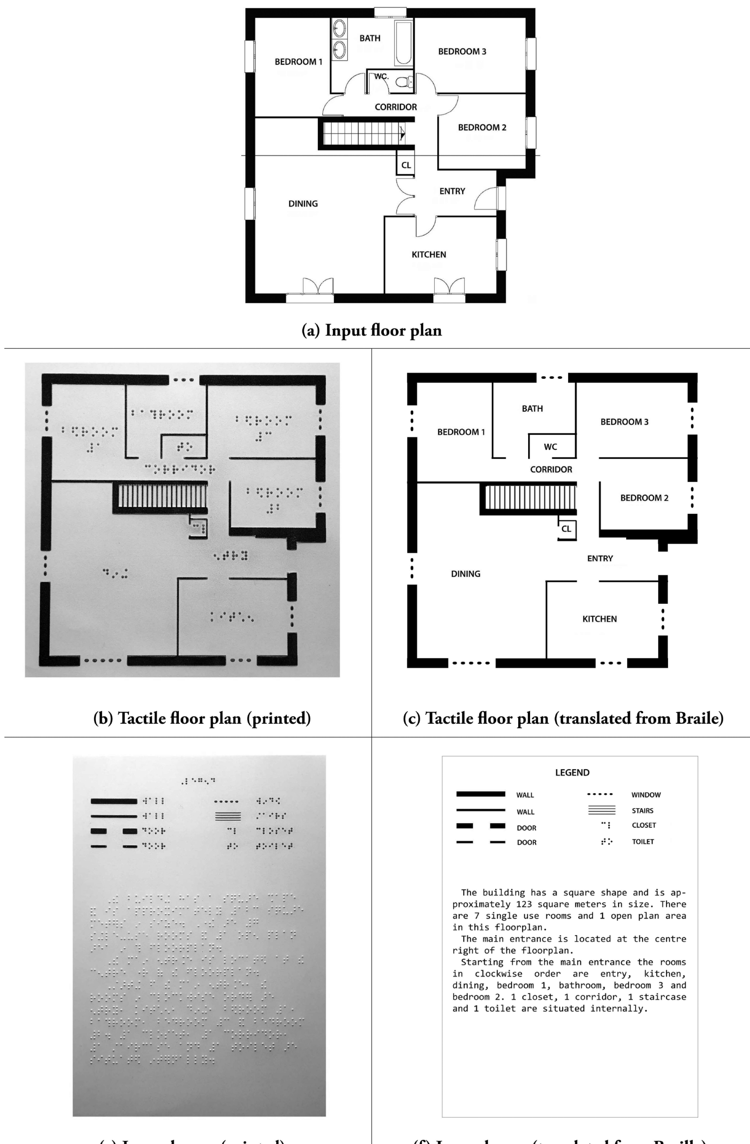
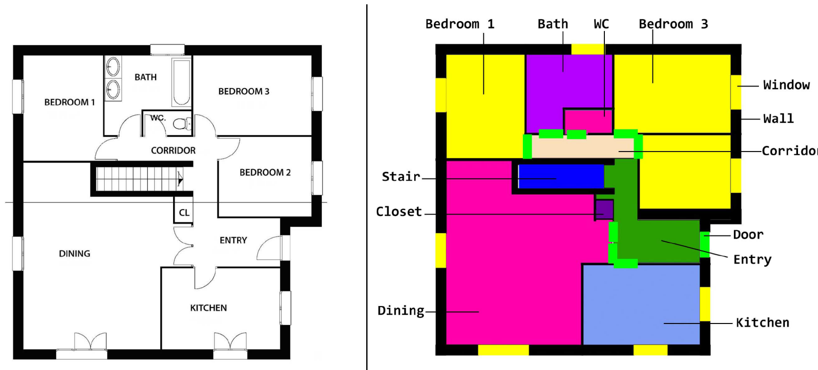


Fig. 18. Floor Plan 2: Tactile presentation.

asked a simple and a complex route finding question. We recorded response times and whether their answer was the shortest route. Then, we asked two more general questions about the floor plan layout, recording response time, and correctness. After the participants had seen and answered questions about both floor plans, we asked them to tell us which house they would prefer to live in and why. The actual questions are shown in Figure 20. Note that there was a choice of two simple and two complex route finding questions: these were counter balanced by presentation method.

In the second part of the study, we presented the participants with the same floor plans in the same mediums, but some information on the plans was shown as uncertain, either with



(a) Input floor plan **(b) GraVVITAS floor plan**

```
 The building has a square shape and is approximately 123 square meters in size.
There are 7 single use rooms and 1 open plan area in this floorplan.
 The main entrance is located at the centre right of the floorplan.
 Starting from the main entrance the rooms in clockwise order are entry, kitchen,
dining, bedroom 1, bathroom, bedroom 3 and bedroom 2. 1 closet, 1 corridor, 1
staircase and 1 toilet are situated internally.
```

(c) GraVVITAS text description

Fig. 19. Floor Plan 2: GraVVITAS presentation.

Questions for 2 Presentation Mediums	Simple Route Finding	Q1	Entry to Kitchen OR Entry to Dining
	Complex Route Finding	Q2	Bedroom 1 to Bedrom 3 OR Bedroom 1 to Dining
		Q3	Kitchen to Toilet
	General	Q4	Which side are the bedrooms?
		Q5	Is Dining connected to Kitchen?
Overall		Q6	Which house would you like to live in?

Fig. 20. Questions for Part 1 of the Final User Study.

medium confidence (e.g., Maybe bedroom, Maybe door) or low confidence (Unknown element). See Figure 21. The confidence is calculated for each floor plan component type using separate threshold values. Our choices for these thresholds were somewhat ad hoc: The need to further explore thresholds and presentation of uncertainty is discussed as part of future work in the conclusion. We asked the participants to explore these new floor plans and asked if they thought that in an unfamiliar floor plan they would still understand the content and be able guess what the unknown elements were. We also asked if they would like to represent uncertainty in some other way.

A limitation of the study design is that these were modifications of floor maps they had already seen and so participants could use their memory of the earlier floor plan to deduce the uncertain elements. However, this was because initial testing showed that the study was too long if we introduced completely new floor plans.

6.3 Results and Discussion

The results of the first part of the study are shown in Figure 22. While the number of participants was small, so one needs to be careful inferring too much from the study, it shows that for

Floorplan Description

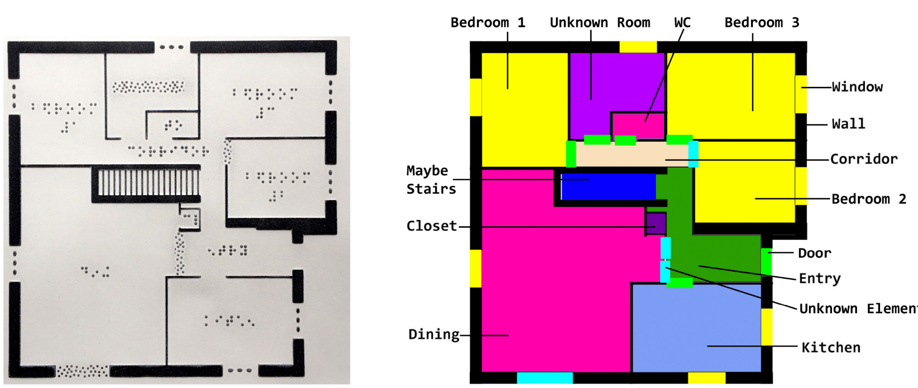
```
This is a house with 2 bedrooms, 1 toilet and
2 rooms with low level of certainty.

The building has an irregular shape and is ap-
proximately 111 square meters in size. There
are 5 single-use rooms and 1 open plan area in
this floorplan. The main entrance is located
at the bottom center of the floorplan. There
is 1 more entrance at the top center.

Starting from the main entrance, the rooms in
clockwise order are open plan entry, open plan
kitchen, open plan dining, maybe bedroom 3,
bedroom 2, unknown room and bedroom 1. 1 cor-
ridor and 1 toilet are situated internally.

The open plan entry is connected to outside
and corridor. The open plan dining is connect-
ed to outside and maybe bedroom 3. The maybe
bedroom 3 is connected to open plan dining and
bedroom 2. The bedroom 2 is connected to cor-
ridor and maybe bedroom 3. The unknown room is
connected to corridor. The toilet is connected
to corridor. The bedroom 1 is connected to
corridor.
```

(a) Text Described Floorplan



(b) Tactile Floorplan **(c) GraVVITAS Floorplan**

Fig. 21. Accessible floor plans with uncertainty.

all presentation modalities participants were able to answer the simple route-finding questions. As shown in Figure 22, we asked simple and complex route finding questions. When describing the route, participants answered by giving the order of rooms and provided general directions like "from dining I move up to living." If the participants chose the shortest path within four minutes, we scored it as "correct" and otherwise it was marked as "incorrect." All users gave correct answers to simple route finding questions with all presentations. For complex routes 3 of the 8 participants gave inaccurate answers for the textual description, while all answered correctly with the GraVVITAS and tactile presentations. Out of the three answers, which we have marked as incorrect, for two cases this was because their answer was not the shortest route. One participant (P1) could not find a route between the Kitchen and Toilet despite persevering for more than four minutes. Most of the other participants built a mental model from the text description and answered questions based on this model. However, this participant tried to find routes without a mental model, repeatedly referring back to the text.

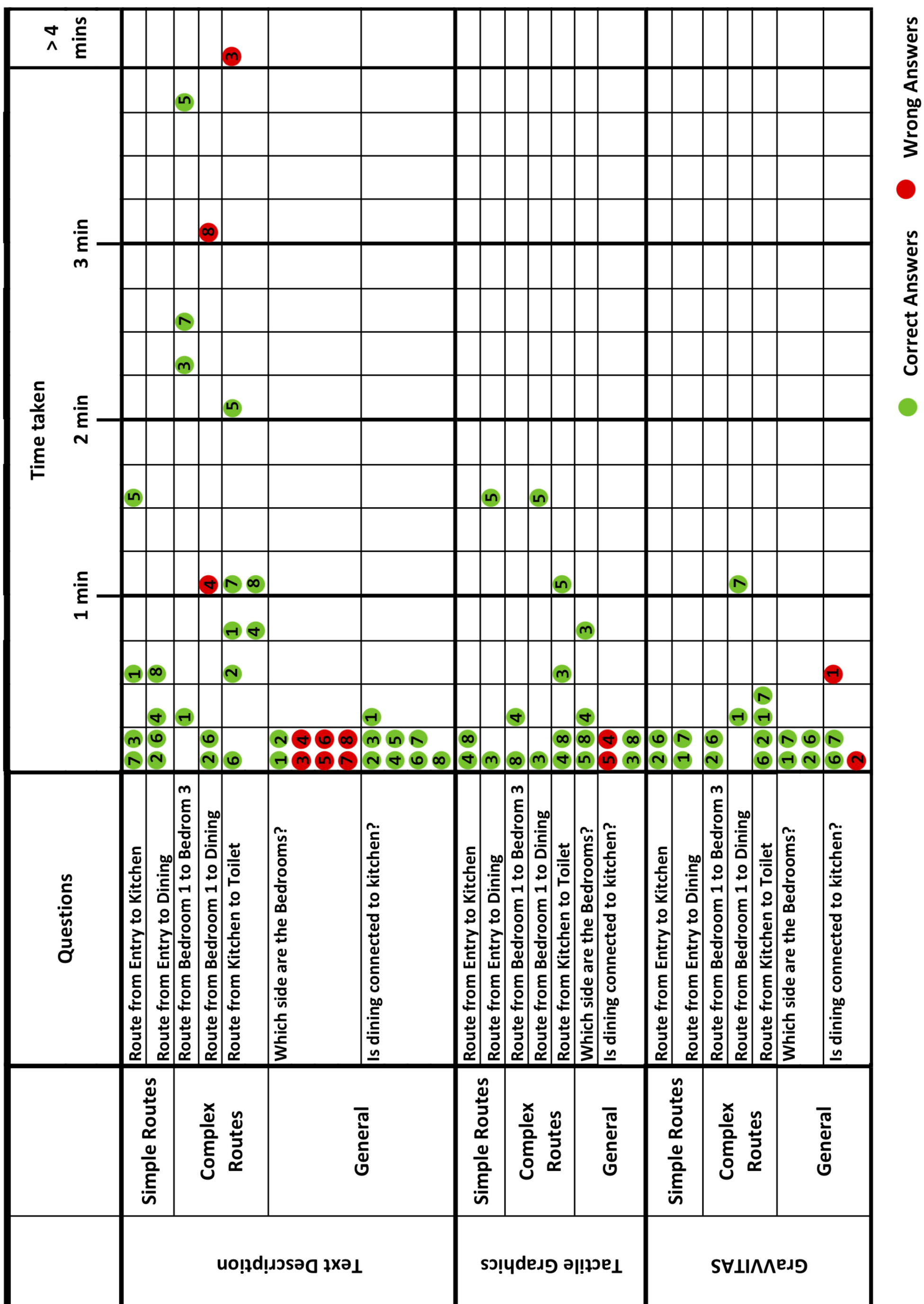


Fig. 22. User study results for Questions Part 1. Numbers identify participants.

With the text description, participants found the general question, "*which side are the Bedrooms?*" difficult with six wrong answers. They commented that even though the clockwise description of the rooms was useful to understand the layout of the rooms, it did not help them to understand which side of the house each room was located. These errors reflect a lack of spatial precision in the textual description rather than an inability by participants to understand it.

However, participants were more accurate in answering the question "*Is Dining connected to Kitchen?*" with the textual description than the tactile or GraVVITAS presentation. Examination of the tactile map in Figure 18 reveals that the doors from Dining to Entry and Entry to Kitchen are very close to each other. Our participants told us that this made them ignore the room entry and think that the two rooms were connected.

We also see that participants generally spent more time in answering questions with the textual description than with the tactile floor plan and GraVVITAS. Participants said that this was because, "text takes a lot of time to get used to. It's easier to use a graphical display." They also commented, "I like how [the] tactile floor plan allows me to explore and understand the layout" and, "GraVVITAS is fun to explore and I feel like it's talking to me. I like it and it's helpful."

The last question we asked the participants was which of the two houses did they like the most. Two participants preferred Floor Plan 1. Their reasons were, (a) "The house is much smaller and it's easier to get around," (b) "I like it that there are two doors connecting to outside" and (c) "Open plan dining and kitchen is better since it's easier to carry the plates." The other six participants preferred Floor Plan 2 because (a) "In the text-presented floor plan, the bedrooms were connected together. That affects privacy. So I like the other house better," (b) "It's nicer when dining and kitchen are separated" and (c) "Kitchen is bigger than in text-based floor plan." These quotes show that the participants had developed a good idea about the layout of the houses, and it also suggests that all three representations would be able to inform visually impaired people in making house purchasing or rental decisions.

In the second part of this user study, we presented participants with the modified floor plans containing uncertain elements. After exploring these floor plans, in the two presentation mediums they used earlier, all the participants felt that even with the uncertainty they could still understand the content. As they explained, they did not feel disturbed when they encountered unknown elements, rather, they made a guess about it and continued exploring. They approved of how we had presented uncertainty and only one participant suggested that maybe we could add percentages of confidence. However, all of the other participants thought that this would be too much information.

We also asked each participant to guess what each element was, and they were able to make good guesses with deductive reasoning. For example, the Bedroom 2 door in Figures 21(b), (c)—which is presented as an unknown element-was assumed to be a door, since, "that is the only entrance to the room" (P2, P3, P4, P7, P8), "it comes off the corridor" (P6), and "size is small and matches other door sizes" (P1). The same logic was applied to guess the unknown element between Entry and Dining in Figures 21(b),(c) as a door. In Figure 21(a), for the text-only presentation, seven participants guessed the unknown room to be a bathroom, since "it is located close to the toilet and because there is no other bathroom in the house" (P3). However, one participant thought it was a bedroom, as "it is located in the middle of other bedrooms" (P5). All of them guessed the unknown rooms correctly in the tactile and GraVVITAS presentations. They believed that they were able to make better inferences about the unknown room with the tactile plan and GraVVITAS than with the textual description, because they could explore and understand the dimensions. Therefore, they suggested adding room sizes for unknown rooms in the text description to help users make better deductions.

7 CONCLUSION

In this article, we have presented a new generic model for the provision of accessible online graphics to people with severe vision impairment. The model has two important properties. First, it caters for inaccurate recognition of the components in the original image by (a) allowing a sighted user to correct the recognition using a simple editing tool that allows them to annotate the original image with high-level semantic tags and (b) requiring the recognition system to provide different levels of confidence in the recognized elements and then incorporating this into the generated accessible graphic so the reader can have an appropriate level of trust in the transcribed graphic. Second, it supports generating accessible graphics in different presentation mediums such as a text-only description, tactile graphic, or a touch-controlled audio presentation on a touchscreen.

We have evaluated the model through a case study: transcription of online house floor plans. While this is only a single case study, we chose this domain because floor plans are challenging to recognize because of their notational complexity and variability. We are therefore confident that the model will perform at least as well with other kinds of graphical notation such as statistical graphics or chemical formulae. However, verifying this remains future work.

The case study supported the decision to provide the user with a choice of presentation modalities as users had different preferences and indicated that they would use different modalities depending upon the context.

Through two user studies, conducted as part of this case study, we have gained a better understanding of how to present floor plans to visually impaired users and what information to present.

However, the case study revealed that pipeline-based graphics recognition of floor plans is still quite brittle and that the resulting system, while accurate for the images they have been designed for, does not handle unforeseen variations very well. As a result, we explored how the tool might indicate that some recognition was uncertain. This was promising, as our user study provides evidence that at least to a limited extent vision impaired readers can use contextual information to deduce the nature of unknown elements. In our study, participants were able to identify doors, windows, and even room names using the contextual information by using the size, shape, and position of the unknown element to deduce its nature. However, more work is required to investigate this: what thresholds should be used for different levels of uncertainty and how users understand these different levels.

There are many ways in which we plan to extend the work. The most obvious is to improve the floor plan recognition system; in particular, recognition of furniture and other objects but also to investigate how we can improve robustness. Recently, deep-learning based image recognition approaches have been found to be superior to traditional pipeline-based approaches such as that used here (e.g., References [82, 83]) even if perfect automatic recognition is still not achieved. An important area of future work is to explore these approaches taking advantage of new datasets that have been proposed very recently (e.g., Reference [37]). Nonetheless, because of the huge variability in notational conventions, we believe that it will remain necessary to present levels of (un)certainty to the reader.

We would also like to extend the system to recognize floor plans of public spaces, such as airports, shopping complexes, and train stations. Much more ambitiously, we would like to develop a generic recognition model that is able to recognize most, if not all types, of online information graphics, such as statistical charts, maps, floor plans, and diagrams.

We also need to evaluate the floor plan editing tool to find out if the non-expert sighted users can use the tool to correct recognition errors. Additionally, we would like to experiment with other presentation mediums such as 3D printed floor plans. Finally, once the recognition system is improved, we would like to provide the tool as a web service or a browser plugin.

ACKNOWLEDGMENTS

We would especially like to thank our colleague Leona Holloway for providing advice on transcription and helping with recruitment of participants for the two user studies. We would also like to thank Stefano Quiriconi for the initial implementation of the floor plan recognition system.

REFERENCES

[1] Sheraz Ahmed, Marcus Liwicki, Markus Weber, and Andreas Dengel. 2011. Improved automatic analysis of architectural floor plans. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'11).* IEEE, 864–869.

[2] Sheraz Ahmed, Marcus Liwicki, Markus Weber, and Andreas Dengel. 2012. Automatic room detection and room labeling from architectural floor plans. In *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS'12).* IEEE, 339–343.

[3] Sheraz Ahmed, Markus Weber, Marcus Liwicki, Christoph Langenhan, Andreas Dengel, and Frank Petzold. 2014. Automatic analysis and sketch-based retrieval of architectural floor plans. *Pattern Recog. Lett.* 35 (2014), 91–100.

[4] Yasuhiro Aoki, Akio Shio, Hiroyuki Arai, and Kazumi Odaka. 1996. A prototype system for interpreting hand-sketched floor plans. In *Proceedings of the 13th International Conference on Pattern Recognition*, Vol. 3. IEEE, 747–751.

[5] Nikola Banovic, Rachel L. Franz, Khai N. Truong, Jennifer Mankoff, and Anind K. Dey. 2013. Uncovering information needs for independent spatial learning for users who are visually impaired. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, 24.

[6] Alessio Barducci and Simone Marinai. 2012. Object recognition in floor plans by graphs of white connected components. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR'12).* IEEE, 298–301.

[7] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. 2010. VizWiz: Nearly real-time answers to visual questions. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology (UIST'10).* ACM, 333–342.

[8] Nizar Bouhlel and Anis Rojbi. 2014. New tools for automating tactile geographic map translation. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, 313–314.

[9] Andy Brown, Steve Pettifer, and Robert Stevens. 2004. Evaluation of a non-visual molecule browser. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, 40–47.

[10] Timothy C. Craven. 2006. Some features of "alt" texts associated with images in web pages. *Inf. Res.: Int. Electron. J.* 11, 2 (2006), n2.

[11] Lluís-Pere de las Heras, David Fernández Mota, Alicia Fornés, Ernest Valveny, Gemma Sánchez, and Josep Lladós. 2013. Runlength histogram image signature for perceptual retrieval of architectural floor plans. In *Proceedings of the International Workshop on Graphics Recognition.* Springer, 135–146.

[12] Lluís-Pere de las Heras, Sheraz Ahmed, Marcus Liwicki, Ernest Valveny, and Gemma Sánchez. 2014. Statistical segmentation and structural recognition for floor plan interpretation. *Int. J. Doc. Anal. Recog.* 17, 3 (2014), 221–237.

[13] Lluís-Pere de las Heras, Oriol Ramos Terrades, Sergi Robles, and Gemma Sánchez. 2015. CVC-FP and SGT: A new database for structural floor plan analysis and its groundtruthing tool. *Int. J. Doc. Anal. Recog.* 18, 1 (2015), 15–30.

[14] Kees Van Deemter, Mariët Theune, and Emiel Krahmer. 2005. Real versus template-based natural language generation: A false opposition?*Comput. Ling.* 31, 1 (2005), 15–24.

[15] Seniz Demir, David Oliver, Edward Schwartz, Stephanie Elzer, Sandra Carberry, Kathleen F. Mccoy, and Daniel Chester. 2010. Interactive SIGHT: Textual access to simple bar charts. *New Rev. Hypermed. Multimed.* 16, 3 (2010), 245–279.

[16] Samuel Dodge, Jiu Xu, and Björn Stenger. 2017. Parsing floor plan images. In *Proceedings of the 15th IAPR International Conference on Machine Vision Applications (MVA'17).* IEEE, 358–361.

[17] Stacy Doore, Kate Beard, and Nicholas Giudice. 2016. Spatial preposition use in indoor scene descriptions. In *Proceedings of the International Conference on GIScience.* UC Berkeley, 84–87.

[18] Philippe Dosch, Karl Tombre, Christian Ah-Soon, and Gérald Masini. 2000. A complete system for the analysis of architectural drawings. *Int. J. Doc. Anal. Recog.* 3, 2 (2000), 102–116.

[19] Max Feltes, Sheraz Ahmed, Andreas Dengel, and Marcus Liwicki. 2013. Improved contour-based corner detection for architectural floor plans. In *Proceedings of the International Workshop on Graphics Recognition.* Springer, 191–203.

[20] Leo Ferres, Petro Verkhogliad, Gitte Lindgaard, Louis Boucher, Antoine Chretien, and Martin Lachance. 2007. Improving accessibility to statistical graphs: The iGraph-lite system. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers & Accessibility.* ACM, 67–74.

[21] 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.* ACM, 39.

[22] Lloyd A. Fletcher and Rangachar Kasturi. 1988. A robust algorithm for text string separation from mixed text/graphics images. *IEEE Trans. Pattern Anal. Mach. Intell.* 10, 6 (1988), 910–918.

[23] Thomas Gallagher, Elyse Wise, Hoe Chee Yam, Binghao Li, Euan Ramsey-Stewart, Andrew G. Dempster, and Chris Rizos. 2014. Indoor navigation for people who are blind or vision impaired: Where are we and where are we going? *J. Loc. Based Serv.* 8, 1 (2014), 54–73.

[24] John A. Gardner and Vladimir Bulatov. 2006. Scientific diagrams made easy with IVEO. In *ICCHP: Computers Helping People with Special Needs*. Springer, 1243–1250.

[25] Lucile Gimenez, Sylvain Robert, Frédéric Suard, and Khaldoun Zreik. 2016. Automatic reconstruction of 3D building models from scanned 2D floor plans. *Autom. Constr.* 63 (2016), 48–56.

[26] Cagatay Goncu. 2012. *GraVVITAS: Accessible Graphics for Visually Impaired People*. Ph.D. Dissertation. Monash University.

[27] Cagatay Goncu, Anuradha Madugalla, Simone Marinai, and Kim Marriott. 2015. Accessible on-line floor plans. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 388–398.

[28] Cagatay Goncu and Kim Marriott. 2011. GraVVITAS: Generic multi-touch presentation of accessible graphics. In *Human-computer Interaction–INTERACT*. Springer, 30–48.

[29] Cagatay Goncu and Kim Marriott. 2015. Creating ebooks with accessible graphics content. In *Proceedings of the ACM Symposium on Document Engineering*. ACM, 89–92.

[30] Timo Götzelmann and Aleksander Pavkovic. 2014. Towards automatically generated tactile detail maps by 3D printers for blind persons. In *Proceedings of the International Conference on Computers for Handicapped Persons*. Springer, 1–7.

[31] Timo Götzelmann and Klaus Winkler. 2015. SmartTactMaps: A smartphone-based approach to support blind persons in exploring tactile maps. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2.

[32] HandyTech. 2018. Active Braille. Retrieved from https://handytech.de/en/products/mobile-braille-displays/active-braille.

[33] Caroline Hayes. 2017. Electronics lend a helping hand to young and old. *Eng. Technol.* 12, 1 (2017), 40–41.

[34] Marion A. Hersh and Michael A. Johnson. 2008. Accessible information: An overview. In *Assistive Technology for Visually Impaired and Blind People*. Springer, 385–448.

[35] HyperBraille. 2007. The graphics-enabled display for blind computer users. Retrieved from http://www.hyperbraille.de/?lang=en.

[36] Chandrika Jayant, Matt Renzelmann, Dana Wen, Satria Krisnandi, Richard Ladner, and Dan Comden. 2007. Automated tactile graphics translation: In the field. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers & Accessibility*. ACM, 75–82.

[37] Ahti Kalervo, Juha Ylioinas, Markus Häikiö, Antti Karhu, and Juho Kannala. 2019. CubiCasa5k: A dataset and an improved multi-task model for floorplan image analysis. In *Proceedings of the Scandinavian Conference on Image Analysis*. Springer, 28–40.

[38] Toshihiro Kanahori, Masayuki Naka, and Masakazu Suzuki. 2008. Braille-embedded tactile graphics editor with Infty system. In *Proceedings of the International Conference on Computers for Handicapped Persons*. Springer, 919–925.

[39] Andrea R. Kennel. 1996. Audiograf: A diagram-reader for the blind. In *Proceedings of the 2nd Annual ACM Conference on Assistive Technologies*. ACM, 51–56.

[40] Ig Mo Koo, Kwangmok Jung, Ja Choon Koo, Jae-Do Nam, Young Kwan Lee, and Hyouk Ryeol Choi. 2008. Development of soft-actuator-based wearable tactile display. *IEEE Trans. Robot.* 24, 3 (2008), 549–558.

[41] Stephen E. Krufka, Kenneth E. Barner, and Tuncer Can Aysal. 2007. Visual to tactile conversion of vector graphics. *IEEE Trans. Neural Syst. Rehab. Eng.* 15, 2 (2007), 310–321.

[42] Steven Landau and Lesley Wells. 2003. Merging tactile sensory input and audio data by means of the talking tactile tablet. In *Proc. Eurohaptics*, Vol. 3. 414–418.

[43] David Lareau and Jochen Lang. 2014. Instrument for haptic image exploration. *IEEE Trans. Instrument. Meas.* 63, 1 (2014), 35–45.

[44] Google LLC. 2011. WalkyTalky—Android Apps on Google Play. Retrieved from https://play.google.com/store/apps/details?id=com.googlecode.eyesfree.walkytalky&hl=en.

[45] Google LLC. 2013. Intersection Explorer—Android Apps on Google Play. Retrieved from https://play.google.com/store/apps/details?id=com.google.android.marvin.intersectionexplorer&hl=en_US.

[46] Tong Lu, Huafei Yang, Ruoyu Yang, and Shijie Cai. 2007. Automatic analysis and integration of architectural drawings. *Int. J. Doc. Anal. Recog.* 9, 1 (2007), 31–47.

[47] Sébastien Macé, Hervé Locteau, Ernest Valveny, and Salvatore Tabbone. 2010. A system to detect rooms in architectural floor plan images. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. ACM, 167–174.

[48] Anuradha Madugalla. 2019. Floorplan Corpus. https://bitbucket.org/AnuMd/taccess_material/src/master/Datasets/. (Accessed on 02/19/2019).

[49] Anuradha Madugalla, Kim Marriott, and Simone Marinai. 2017. Partitioning open plan areas in floor plans. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'17), 2017 14th IAPR International Conference on*, Vol. 1. IEEE, 47–52.

[50] NASA. 2009. Earth+ Earth Science. Retrieved from https://prime.jsc.nasa.gov/earthplus/.

[51] The Braille Authority of North America. 2010. Guidelines and Standards for Tactile Graphics. Retrieved from http://www.brailleauthority.org/tg/web-manual/index.html.

[52] American Council of the Blind. 2018. What is Audio Description. Retrieved from http://www.acb.org/adp/ad.html.

[53] Devi A. Paladugu, Hima Bindu Maguluri, Qiongjie Tian, and Baoxin Li. 2012. Automated description generation for indoor floor maps. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers & Accessibility.* ACM, 211–212.

[54] Devi Archana Paladugu, Qiongjie Tian, Hima Bindu Maguluri, and Baoxin Li. 2015. Towards building an automated system for describing indoor floor maps for individuals with visual impairment. *Cyber-phys. Syst.* 1, 2–4 (2015), 132–159.

[55] Konstantinos Papadopoulos, Marialena Barouti, and Konstantinos Charitakis. 2014. A university indoors audio-tactile mobility aid for individuals with blindness. In *Proceedings of the International Conference on Computers for Handicapped Persons.* Springer, 108–115.

[56] Konstantinos Papadopoulos, Konstantinos Charitakis, Lefkothea Kartasidou, Georgios Kouroupetroglou, Suad Sakalli Gumus, Efstratios Stylianidis, Rainer Stiefelhagen, Karin Müller, Engin Yilmaz, Gerhard Jaworek et al. 2016. User requirements regarding information included in audio-tactile maps for individuals with blindness. In *Proceedings of the International Conference on Computers Helping People with Special Needs.* Springer, 168–175.

[57] Don Parkes. 1994. Audio tactile systems for designing and learning complex environments as a vision impaired person: Static and dynamic spatial information access. *Learning Environment Technology: Selected Papers from LETA* 94 (1994), 219–223. Retrieved from http://www.ascilite.org/conferences/aset-archives/confs/edtech94/mp/parkes.html.

[58] Helen Petrie, Fraser Hamilton, Neil King, and Pete Pavan. 2006. Remote usability evaluations with disabled people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1133–1141.

[59] Helen Petrie, Chandra Harrison, and Sundeep Dev. 2005. Describing images on the web: A survey of current practice and prospects for the future. In *Proceedings of Human-computer Interaction International Conference (HC'05).*

[60] Helen Petrie, Neil King, Anne-Marie Burn, and Peter Pavan. 2006. Providing interactive access to architectural floorplans for blind people. *Brit. J. Vis. Impair.* 24, 1 (2006), 4–11.

[61] Helen Petrie, Christoph Schlieder, Paul Blenkhorn, Gareth Evans, Alasdair King, Anne-Marie O'Neill, George T. Ioannidis, Blaithin Gallagher, David Crombie, Rolf Mager et al. 2002. TeDUB: A system for presenting and exploring technical drawings for blind people. In *Computers Helping People with Special Needs.* Springer, 537–539.

[62] Ingo Redeke. 2001. Image & graphic reader. In *Proceedings of the International Conference on Image Processing*, Vol. 1. IEEE, 806–809.

[63] Ehud Reiter. 1995. NLG vs. templates. *Arxiv Preprint Cmp-lg/9504013* (1995).

[64] Jonathan Rowell and Simon Ungar. 2005. Feeling our way: Tactile map user requirements—A survey. In *Proceedings of the International Cartographic Conference.*

[65] Kathy Ryall, Stuart Shieber, Joe Marks, and Murray Mazer. 1995. Semi-automatic delineation of regions in floor plans. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Vol. 2. IEEE, 964–969.

[66] Noureddin M. Sadawi, Alan P. Sexton, and Volker Sorge. 2012. Chemical structure recognition: A rule-based approach. In *Document Recognition and Retrieval XIX*, Vol. 8297. International Society for Optics and Photonics, 82970E.

[67] J. Kenneth Salisbury and Mandayam A. Srinivasan. 1997. Phantom-based haptic interaction with virtual objects. *IEEE Comput. Graph. Applic.* 17, 5 (1997), 6–10.

[68] Volker Sorge, Mark Lee, and Sandy Wilkinson. 2015. End-to-end solution for accessible chemical diagrams. In *Proceedings of the 12th Web for All Conference.* ACM, 6.

[69] Jing Su, Alyssa Rosenzweig, Ashvin Goel, Eyal de Lara, and Khai N. Truong. 2010. Timbremap: Enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th International Conference on Human-computer Interaction with Mobile Devices and Services.* ACM, 17–26.

[70] Hao Tang, Norbu Tsering, Feng Hu, and Zhigang Zhu. 2016. Automatic pre-journey indoor map generation using autoCAD floor plan. *J. Technol. Pers. Disab.* 4 (2016), 176–191.

[71] Brandon Taylor, Anind Dey, Dan Siewiorek, and Asim Smailagic. 2016. Customizable 3D printed tactile maps as interactive overlays. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, 71–79.

[72] M. Tezuka, K. Ishimaru, and N. Miki. 2017. Electrotactile display composed of two-dimensionally and densely distributed microneedle electrodes. *Sens. Actuat. A: Phys.* 258 (2017), 32–38.

[73] Luis Von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 319–326.

[74] Luis Von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum. 2006. Improving accessibility of the web with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 79–82.

[75] W3C. 2008. Web Content Accessibility Guidelines 2.0. Retrieved from https://www.w3.org/TR/WCAG20/.

[76] Zheshen Wang, Nan Li, and Baoxin Li. 2012. Fast and independent access to map directions for people who are blind. *Interact. Comput.* 24, 2 (2012), 91–106.

[77] Thomas P. Way and Kenneth E. Barner. 1997. Automatic visual to tactile translation. II. Evaluation of the TACTile image creation system. *IEEE Trans. Rehab. Eng.* 5, 1 (1997), 95–105.

[78] Limin Zeng and Gerhard Weber. 2010. Audio-haptic browser for a geographical information system. In *Proceedings of the International Conference on Computers for Handicapped Persons*. Springer, 466–473.

[79] Limin Zeng and Gerhard Weber. 2012. ATMap: Annotated tactile maps for the visually impaired. In *Cognitive Behavioural Systems*. Springer, 290–298.

[80] Xuning Zhang, Albert Kai-Sun Wong, and Chin-Tau Lea. 2016. Automatic floor plan analysis for adaptive indoor Wi-Fi positioning system. In *Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI'16)*. IEEE, 869–874.

[81] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data sonification for users with visual impairment: A case study with georeferenced data. *ACM Trans. Computer-hum. Interact.* 15, 1 (2008), 4.

[82] Ruiyun Zhu, Jingcheng Shen, Xiangtian Deng, Marcus Walldén, and Fumihiko Ino. 2020. Training strategies for CNN-based models to parse complex floor plans. In *Proceedings of the 9th International Conference on Software and Computer Applications (ICSCA'20).* Association for Computing Machinery, New York, NY, 11–16. DOI: https://doi.org/10.1145/3384544.3384566

[83] Zahra Ziran and Simone Marinai. 2018. Object detection in floor plan images. In *Proceedings of the 8th IAPR Workshop on Artificial Neural Networks in Pattern Recognition*. Springer Nature, 383–394.

Received March 2019; revised May 2020; accepted July 2020