Towards AI-based Accessible Digital Media: Image Analysis Pipelines and Blind User Studies*

Mazen Salous, Daniel Lange, Timo Von Reeken, Wilko Heuten, Susanne Boll, Larbi Abdenebaoui
Society Department
OFFIS Institue for Information Technology
Oldenburg, Germany
fisrtname.lastname@offis.de

***Abstract*—We report from our work in progress within our project ABILITY, in which we implemented AI-based image analysis pipelines to improve the accessibility of digital media for Blind and Visually Impaired (BVI) users. In addition, we conducted two user studies with BVIs, the first one as a preliminary study, and the current one to evaluate different types of AI-based automatic description. Based on our current progress, multimodal assistant (namely speech, tactile representations and braille) will be transformed to a dedicated BVI-tablet.**

***Index Terms*—accessibility, AI, computer vision, Image analysis**

I. Introduction

Lifestyles are rapidly changing with the advent of digitalization. Almost every daily activity can now be supported by the Internet and digital media. However, blind or visually impaired (BVI) people still face challenges in adapting to these changes, despite efforts to make websites and applications accessible to this user group. AI technologies, particularly computer vision, could play a significant role in helping BVIs to use digital media and navigate the Web more efficiently. While there is some progress in developing computer vision apps dedicated for BVIs, such as Seeing AI [1] and Be My Eyes [2], they mainly assist BVIs using the smartphone camera in their everyday activities, and still have limited capabilities and do not offer end-to-end support for BVIs with digital media.

This paper presents our first two practical steps achieved towards developing an AI-based end-to-end assistant system for BVIs: 1) AI-based image analysis pipelines and 2) insights from user studies with BVIs regarding important aspects in image description in different scenarios.

II. Related Work

Apps For BVI users: There are several apps available to support BVI individuals in their daily activities. Two notable apps are Seeing AI [1], which uses smartphone cameras to identify objects and provide information, and Be My Eyes [2], which connects BVIs with sighted volunteers through live video calls. Currently, the developers of Be My Eyes have just announced their next step to add AI volunteers in order to enhance availability of assistance. While these apps represent valuable tools in promoting inclusivity and independence for BVIs in their everyday physical activities with the surroundings, we complement this advantage by focusing specifically on digital media accessibility support.

Visual Large Language Models: Nowadays we are witnessing a revolution of Visual Large Language Models, which combine language processing with visual understanding. One prominent recent example is the emergence of Visual ChatGPT [3], a cutting-edge model specifically designed to instruct AI in image editing tasks. Another notable and more recent advancement is LLava [4], a state-of-the-art model that excels at describing images using natural language. While LLava excels at generating textual descriptions of images, it may not capture the nuances and details that BVIs specifically rely on to comprehend visual content. For example, individuals with visual impairments may require explicit information about the spatial positioning and objects interaction within an image to form a complete mental representation. LLava represents an important step towards automated image description, however, further research and development are necessary to create dedicated models that address the unique needs of BVIs.

III. Towards Improved BVI-Computer Interaction: Understanding Image Contents and Context

We implemented several image analysis pipelines using state-of-the-art computer vision models. The pipelines vary in their order of services to offer richer interactive scenarios, i.e., we considered different cooperative situations between AI and BVIs, namely based on passing the lead of imagery model construction between AI and the BVIs themselves. Figure 1 shows several outputs of our pipelines e.g. automatic captioning (AI has the lead), question & answering (BVI has the lead) etc. All the outputs will be transformed as multimodal assistants for BVIs, namely speech and braille assistants (braille for both text and tactile simplified graphics).

IV. BVI User Studies

Preliminary Study: A preliminary workshop was conducted with three BVIs (age 33 - 76 years, 2 females) and one BVI-assistant to explore requirements for accessible digital media. The participants highlighted the need of different types of descriptions for different scenarios. When

This work is funded by the ABILITY project: https://www.ability-project.eu/.

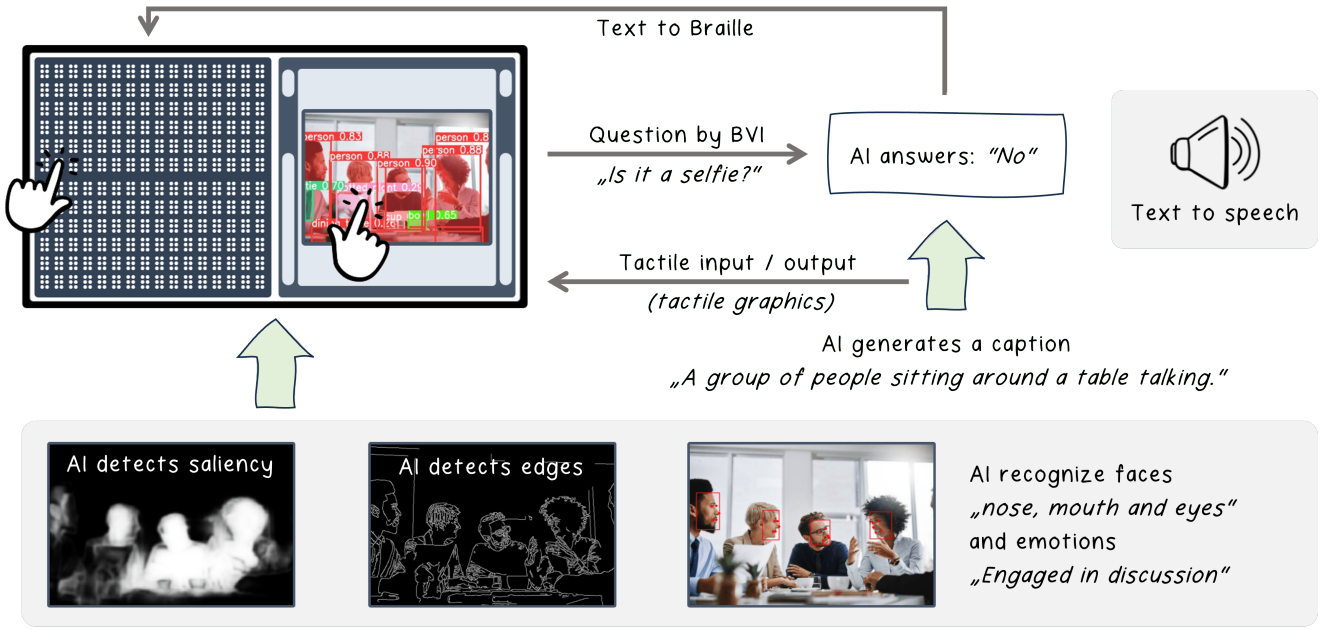

Fig. 1. AI-based Multimodal Support for BVIs: The surfed image is analyzed in multi-levels using state-of-the-art computer vision models. While auto-generated caption is agreed by BVIs to be the first supporting step to allow BVIs initializing their imagery model, BVIs argued that a dynamic order of the remaining explanations would best fit their needs.

searching for a specific image, participants would prefer concise and informative descriptions (caption) that allow them to quickly determine the image's relevance. In newspaper articles, descriptions should provide contextual information such as identifying the person shown. In a digital photo book, descriptions could be more artistic and story-based, complementing the aesthetic design of the pages. The idea of BVI-AI interaction through questions was highly encouraged by all participants to obtain additional information.

Automatic Description Study: An interview has been conducted with three BVIs (Age 47 - 63, 3 males), in which we evaluated a variety of automatic generated image descriptions: informative concise caption, and 3 detailed descriptions, each focuses on different aspect (objects' shapes, objects' spatial positioning and interaction between objects). Interviews were conducted individually with each participant, to avoid influences from one's on other's opinions. We evaluated their imagery model after each offered description (spoken, they could also read it in braille), by asking them to freely explain their imagery model about the image, thus, we compared their explanation to the ground truth image. Early stage findings: surprisingly, shapes were considered less relevant information, while BVIs were able to describe the image very well after the object interaction and the spatial positioning descriptions. These results mostly coincide with the literature of image descriptions for blind people [5].

V. Current Progress, Conclusion and Future Work

Following an agile process, we first collected requirements from BVIs for accessible digital media, then we implemented a variety of AI-based image analysis pipelines and we have already conducted a new user study with BVIs and evaluated various AI-based automatic image descriptions. One limitation is the inclusion of a small number of individuals with a wide range of ages in the preliminary blind user studies, which may introduce confounding factors. However, it is worth noting that our early stage results do align with existing literature on image description for blind individuals. In future studies, we plan to include a convenient larger number of participants to facilitate more robust statistical measurements while considering potential age effects and interactions.

Currently, our project partners are developing a dedicated BVI-tablet with braille display. Waiting for the tablet being developed and delivered, we are about developing a tablet-simulator, in which the tablet functionalities will be simulated as an App, then we can conduct a new user study with BVIs and BVI-assistants to further visualize and evaluate our AI-based analysis functions.

References

[1] Kelley, Steven. "Seeing AI: Artificial intelligence for blind and visually impaired users." (2018).
[2] N. Lakhani, H. Lakhotiya and N. Mulla, "Be My Eyes: An Aid for the Visually Impaired," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2022, pp. 1-6, doi: 10.1109/GCAT55367.2022.9972160
[3] Wu, Chenfei, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang, and Nan Duan. "Visual chatgpt: Talking, drawing and editing with visual foundation models." arXiv preprint arXiv:2303.04671 (2023).
[4] Liu, Haotian, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. "Visual instruction tuning." arXiv preprint arXiv:2304.08485 (2023).
[5] Stangl, Abigale, Meredith Ringel Morris, and Danna Gurari. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions." Proceedings of the 2020 chi conference on human factors in computing systems. (2020).