The BildKlangPoet - a Framework for Designing Mixed Reality Painting Experiences Using Generative AI

Christian Geiger

Mixed Reality and Visualization Lab, University of Applied Sciences Düsseldorf, geiger@hs-duesseldorf.de

Ali Soren

Mixed Reality and Visualization Lab, University of Applied Sciences Düsseldorf, soren.ali@hs-duesseldorf.de

Philipp Dilchert

Mixed Reality and Visualization Lab, University of Applied Sciences Düsseldorf, philipp.dilchert@hs-duesseldorf.de

This project aims to bridge the gap between generative AI and the individual user experience by creating an interactive installation that combines self-created or selected content with real-time and AI-generated media. The primary aim is to actively involve visitors in the process of creating a digital painting. Their face, decisions and drawings or writings become the basis for an AI-generated painting. The installation is based on the principle of AI-supported media generation, in which users play an active role in the creative process. Personalized works of art are created from photos, text or drawings and combined with generated images, poems and music. The results are partially rendered in real time. The second goal is to support the creation of new media art experiences through remixing.

CCS CONCEPTS • Applied Computing~Arts and Humanities~Media Arts

1 INTRODUCTION

The concept of remixing, defined as the reuse of existing materials to create new cultural artefacts, has both historical depth in art and high significance in today's digital landscape. At the heart of remixing lies the notion that all creative works are based on prior material, supporting the idea that all forms of cultural expression incorporate elements of remixing. Originally rooted in avant-garde movements of the early 20th century — particularly the reuse of everyday objects, as well as montage and collage — digital technologies have made remixing more accessible, facilitating its dissemination and democratising the creation of new works, especially in new media art [2]. The diverse possibilities and ease of use of digital technologies are expanding the areas of application beyond mere aesthetic enjoyment, enabling critical engagement with cultural narratives and allowing more people to become active participants in meaning-making processes. L. Manovich describes the concept of 'deep remixability', whereby 'not only the content of different media is mixed, but also their fundamental techniques, working methods, and modes of representation' [3]. However, deep remix is not without complications, particularly concerning originality and copyright issues. The act of remixing often raises questions of authorship and the legal boundaries of fair use. Remixers often find themselves navigating the complex intersection of creativity and copyright law, where they may feel marginalised by the pressures imposed by corporations keen to protect their intellectual property [5]. The evolution of AI adds another layer to the remix discourse. AI, as a tool for generating new remixes, challenges traditional notions of authorship and originality. Machines can learn and reproduce artistic styles, thus blurring the lines between human- and machine-created content and enlarging the remixing dilemma and the tradeoff between generativity and originality [5]. This raises concerns about the value of human intent and interpretation in the subjective realm of artistic creation. Consequently, the rapid advancement of AI may produce remixes that lack the intentionality and emotional depth often associated with human-centred creative processes. We present BildKlangPoet, our work that addresses some of these topics in the context of mixed reality paintings.

Figure 1: Generating paintings and detecting objects and people in real-time with this look.

2 BACKGROUND & RELATED WORK

Mixed Reality offers a compelling intersection between traditional paintings and digital interactivity, enhancing viewer engagement and redefining art appreciation. The book "Augmented Reality Art" by V. Geroimenko provides a comprehensive introduction to AR as a new artistic medium. It explores the evolution of AR from a technical innovation to a platform for creative experimentation and features contributions from artists, designers and scholars. The Handbook of "Artificial Intelligence and the Arts" offers a comprehensive, interdisciplinary exploration of the intersection between AI and the arts. It critically reflects on issues such as authorship, creativity, materiality and the role of AI as a cultural and aesthetic agent. Directly comparable works deal with individual aspects of this approach. For example, Long et al. (2024) use voice-controlled generation to create game scene illustrations [10]. In "Visions of Destruction" (Sola and Guljajeva, 2024), gaze interaction is used to visually realise the destruction of symbolic images [11]. In contrast, the "Prompt to Anything" project [12] generates creative images from prompts but completely dispenses with any personal or poetic reference. A notable project in this field is Mario Klingemann's "Memories of Passers-By" [9], which uses a neural network to create evolving portraits based on audience interaction. This synthesis of AI and XR enables viewers to engage in a dynamic conversation with the artwork, whereby the portraits adapt in real time to their interaction, fostering a unique interpretation for each spectator. Mixed reality is used in many areas of the visual arts for augmentation, providing additional information, for example, or for reconstructing paintings [13] or as a multimodal extension [14]. For example, research by Yoo et al. describes an AR system that provides additional content to capture the historical context of paintings, which is often the domain of experts [15]. In summary, these MR/AI advancements illustrate transformative potential in the museum landscape by promoting innovative modes of interaction and engagement that are essential for new media art.

3 DESIGNING EXPERIENCES WITH FUNCTIONAL BUILDING BLOCKS

With 'remix' as its underlying theme, 'The BildKlangPoet' explores the relationship between multimodal generative AI and personal aesthetic practice within the visual arts, examining how digital technologies can expand this field. This interactive installation focuses on classical painting techniques, while supporting active exploration through digital technologies such as XR and AI. During a creative prototyping phase, we developed functional building blocks that can be adapted and combined for the final installation.

Functional Mode 1 – Synesthetic Content Production: The user selects a painter's style and a poem. A set of digital paintings is then generated from the available data, visually corresponding to both the selected artist's style and the lyrics of the chosen poem. The composition and mood of the image match the selected poem and artist; for example, a panther might appear in a van Gogh painting. At the same time, the poem is sung and accompanied by AI-generated music in a genre that matches the chosen style. This mode allows the user to configure which styles/artists can be selected, which poems are visible, and how the lyrics are presented on the screens.

Functional Mode 2 – Prompt Selection and Meta Prompts: Depending on the selected configuration, different prompts are relevant, based on individual scenarios, the curated context or, if applicable, an artistic position. For example, the prompts could be that all the people in the painting should laugh, or that the people should be of different genders. To this end, we provide meta-prompts or employ a simple control structure that selects based on runtime parameters or enables individual prompt extensions, e.g. via direct input at runtime. Meta-prompts are also used to generate prompts.

Functional Mode 3 - Interactive and Personalized Mode: The system enables user-specific content to be incorporated into image production. Currently, this includes photos taken with a camera at the outset, as well as voice recordings or speech-to-text input for producing poetry or creating individual prompts. Character recognition enables users to write text or create sketches with a pen, which can then be used as image templates. Initially, the selection of content (e.g. artist, style or era, or a poem) is made via touchscreen, stylus or voice input.

Functional Mode 4 - Realtime, Print and eMail Output: We offer a variety of output options. The results are displayed on a Samsung "The Frame" with a custom picture frame. Users can choose to send paintings by email, which the recipient will then receive as a personal message. In addition, it is possible to output images on a professional photo printer, display them in a gallery on an additional large screen. A special output option is the third real-time display of the installation. Here, a camera captures the live image and interprets the recognized people and objects in real time based on a prompt generated by the system.

Now that the functional modules have been developed, different experiences can be easily generated. A typical application scenario can be seen in video [16]. Visitors have their picture taken, select a painting style and a poem, and the system generates a set of paintings presented in a picture frame, while a song with the poem's lyrics plays in the background. The user can select their favorite painting, which is then sent to their email address and displayed in a gallery on a large LED wall. The images produced during this session are analyzed, and a dedicated meta-prompt is generated and sent to the real-time screen. If a visitor chooses to sketch their own poem with two or three words and a rough image using a stylus on the touchscreen, a haiku is generated, and the visual sketch is transformed into a painting. Other scenarios include a remix of two selected painting styles (battle of isms) or to critique the motif of (female) muses for (male) artists.

4 THE MIXED REALITY TECHNOLOGY OF THE BILDKLANGPOET

The application runs entirely locally on a powerful system. Two central components control the technical implementation: Unity3D handles the entire user guidance and display, while ComfyUI — a node-based framework — connects all the AI modules — from text input and face integration to the final image output — in a modular fashion. Image generation is handled by Stable Diffusion XL (SDXL), supplemented by a refiner that improves resolution and depth of detail. Rather than being added retrospectively, the user's face is integrated directly into the generation process using PullID. This creates a new image that realistically reflects the style of the selected artist and the user's individual facial features, without the need for traditional face-swapping. The entire process — from the initial drawing to the final image — is managed by the ComfyUI pipeline, which incorporates text analysis, prompt logic, style interpretation, and SDXL image output. The camera image is then interpreted using StreamDiffusion (a real-time approach to Stable Diffusion implemented in TouchDesigner by Dotsimulate/Lyell Hintz) and converted into a painting in real time. StreamDiffusion uses a Stable Diffusion model that achieves the best possible results; in this case, SD-Turbo is used. More technical details of the concept and its implementation can be found in [17].

5 CONCLUSION

Visitors will mainly be attracted to this project by the opportunity to integrate their own image into a painting and produce their own work, even if this is greatly reduced to selecting a painting style and, if necessary, sketching images and text. The system is versatile due to its modular internal structure and the ability to use and easily combine different LLMs. The system has already been used to create several different topics, including a mixture of painting styles, the generation of haiku poems from user-generated paper sketches and the production of paintings without domination by male artist data.

This paper was created using chatGPT, claude.ai, scite.ai and deepl, for related work research and translation support.

This research and development project is partially funded by the German Federal Ministry of Research, Technology and Space (BMFTR) within the "The Future of Value Creation – Research on Production, Services and Work" program (funding code: 02L19C402). The author is / authors are responsible for the content of this publication.

6 REFERENCES

[1] Knobel, M. and Lankshear, C. (2008). Remix: the art and craft of endless hybridization. Journal of Adolescent &Amp; Adult Literacy, 52(1), 22-33.

[2] Navas, E., Gallagher, O., & burrough, x. (Eds.). (2021). Forking Paths in New Media Art Practices: Investigating Remix [Special Issue]. *Media-N, Journal of the New Media Caucus*, 17(1). https://iopn.library.illinois.edu/journals/median/issue/view/88

[3] Manovich, L. (2007). Deep Remixability. *Artifact: Journal of Design Practice*, 1(2), 76-84. https://doi.org/10.1080/17493460701206751

[4] Freund, K. (2014). "Fair use is legal use": Copyright negotiations and strategies in the fan-vidding community. *New Media & Society*, *18*(7), 1347-1363. https://doi.org/10.1177/1461444814555952 (Original work published 2016)

[5] Hill, B. M. and Monroy-Hernández, A. (2012). The remixing dilemma. American Behavioral Scientist, 57(5), 643-663.

[6] Delfanti, A., & Phan, M. (2024). Rip It Up and Start Again: Creative Labor and the Industrialization of Remix. *Television & New Media*, *25*(7), 720-735. https://doi.org/10.1177/15274764241227613

[7] Geroimenko, V. (Ed.). 2022. *Augmented Reality Art: From an Emerging Technology to a Novel Creative Medium (3rd edition)*. Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-319-06203-7

[8] Catani, S., ed. 2023. *Handbuch Künstliche Intelligenz und die Künste*. Berlin/Boston: Walter de Gruyter (De Gruyter Reference). ISBN 978-3-11-065560-5. DOI: 10.1515/9783110656978

[9] https://onkaos.com/work/memories-of-passersby-i-companion-version/

[10] Long, G., Kumar, A., Cheruvu, R., Ramos, P., Pastushenkov, D., Wu, Z., & Lo, R. (2024). Bringing Adventure Gaming to Life Using Real-Time Generative AI on Your PC. ACM SIGGRAPH 2024 Talks, 1–2. https://doi.org/10.1145/3641233.3664354

[11] Sola, M. C., & Guljajeva, V. (2024). Visions of Destruction: Exploring a Potential of Generative AI in Interactive Art. https://doi.org/10.1145/3678698.3687185

[12] Borgomano, G. (2024). Prompt to Anything?: Exploring Generative AI's Iterative Potential in Mapping Show Production. ACM SIGGRAPH 2024 Talks, 1–2. https://doi.org/10. 1145/3641233.3664729

[13] Bernik, A., Vusić, D., & Kober, D. (2019). Implementation of augmented reality application and computer graphics: the case of the stolen paintings. Tehnicki Vjesnik - Technical Gazette, 26(6). https://doi.org/10.17559/tv-20181015160248

[14] T. Chen, "Revitalization of Ancient Painting: An Interactive Device Based on Augmented Reality and Human-Computer Interaction Technology Empowers Ancient Chinese Painting Exhibitions," *2023 2nd International Conference on Artificial Intelligence, Human-Computer Interaction and Robotics (AIHCIR)*, Tianjin, China, 2023, pp. 13-18, doi: 10.1109/AIHCIR61661.2023.00009.

[15] Yoo, E., Kwon, O., & Yu, J. (2023). Evaluation of an augmented reality for historical context experiences of 3d restored court paintings. IEEE Access, 11, 39197-39212. https://doi.org/10.1109/access.2023.3268528

[16] The BildKlangPoet. Video of a simple session https://vimeo.com/1103249331/bfd3b4f20b

[17] Soren Ali. Generative KI in Theorie und Anwendungim Kontext interaktiver und multimodaler Medieninhalte. Bachelor thesis, HS Düsseldorf, 2025

Figure 2: Generating mixed reality paintings and detecting objects and people in real-time with this look.