Walking Through Shanshui: Generating Chinese Shanshui Paintings via Real-time Tracking of Human Position

Aven Le Zhou
Paper ID 40

Abstract

Shanshui is a traditional East Asian style of ink brush painting that depicts natural landscapes in a semi-abstract fashion. To create a Shanshui painting, ancient Chinese scholar-artists rely heavily on their travel experiences as well as their movements in natural spaces. In this paper, we propose an interactive system - "Walking Through Shanshui" - based on AI using Generative Adversarial Networks and various computer vision techniques. The system is an interactive art installation that helps bring the original experience of creating Shanshui to participants by tracking their movement through walking in a virtual space. It uses position tracking as input to generate Shanshui from participant's movements and to paint with a custom generative Sketch-to-Shanshui translation model. The system detects the participant's position in real-time and automatically traces it to generate a Shanshui painting instantly.

1. Background

Chinese Shanshui, also known as literati painting, is a type of ink brush painting that depicts the natural landscape and is popular in East Asian but originated from China. Shanshui literally means mountain and water, and the style reflects the smoothness of water embodying the shape of the mountainous landscape in the painting. Unlike most realistic painting techniques, Shanshui describes nature in a semi-abstract fashion. Historically, Chinese scholars in ancient times never depict the scene on-site but instead recreate it from their memories after their travels, or even from their imagination or recollection of the landscape [5].

In ancient times, Chinese literati traveled around and immersed themselves in the nature. They walked around beautiful places and returned to their canvasses to capture their memories as a Shanshui painting. Due to these unique laborious production methods, long Shanshui scrolls may consist of multiple changing perspectives with different views. The contents in a typical Shanshui are also more a collage of many scenes depicting different legs of a journey. And the creation process is very much relevant to experiences of traveling, walking and moving in space. Our work seeks to honor these traditions and experiences to create Shanshui paintings that capture the graceful movements of travelers in natural spaces.

Figure 1. Participant Interacts with the System.

2. Introduction

In this project, we are dedicated to providing audiences with a similar experience of creating Shanshui as in ancient times with various technologies that enable us to recreate the experience of moving through nature. We attempt to mimic the traveling and walking experience with a simplified tracking model - walking in a room-size installation. Taking advantage of the advancement of Artificial Intelligence technologies, we use AI to help to generate a painting from a participant's walking experience. Looking at AI's social impacts, we aim to use this project to establish a creative and collaborative AI experience, where AI captivates the imagination to enhance creativity, especially for casual participants, including younger generations, that have a passion for art without a formal training in the area.

In this exhibition, participants are invited to walk inside a digitally interactive space that uses sophisticated motion tracking. Rather than painting with traditional brushes and ink, they walk and create art through their footsteps! There are two essential components to the system: the tracking and the AI art generative system. The tracking system can see and record the participant's movements and the posi-

tions will be continuously logged as white lines over a black background canvas (the sketch). After a participant walks in and out of the exhibit, the system will take the motion-captured input sketch and generate a Shanshui painting. We trained CycleGAN [7] with a custom data set [6], and optimized the trained model with several techniques. With the optimized Sketch-to-Shanshui translation model as a generative function, the system can translate the sketch into a Chinese Shanshui painting.

We call this system "Walking Through Shanshui," not only to describe the general interaction but also referring to the origins of ancient Shanshui painting practices. With the capability of AI, we walk, travel and present related memories and experiences in the form of Chinese traditional art.

Main Contribution: This paper presents a novel application which adopts multiple computer vision techniques to generate Eastern traditional visual art in a creative, interactive and engaging way. Deep cultural heritage meets rapidly changing forms of art powered by technological advancements allows us to rediscover traditions and allow their essence to evolve through emerging technology while maintaining elements of both familiar and new. This opens up a new form of expression by extending the form and finesse of the artist to the curiosity of the observer, in effect making thee audience an integral part of the art itself via an interactive display. This project helps us find the inner artist in ourselves and brings significant value by turning the observer into an artist for others to visualize their journey in real time. We demonstrate that AI can bring art to the masses and help them express themselves seamlessly through technology.

3. System Overview

Interaction Overview: We set up the system as a room-size enclosed interactive exhibit. On-site staff/assistants provide guidance during the live show. Only be one participant at a time can control the digital canvas and she or he will receive a yellow tracking hat from the assisting staff. Participants will first put on the hat and then walk into the interactive space to start engaging with the system. Only the hat bearer may enter the interactive stage. As the bearer moves, the system displays the real-time position-log sketch on the tracking monitor. When the participant feels ready, she or he can walk outside of the stage and hand the hat back to assisting staff for the next participant.

Installation Setup: Hardware requirements include three main parts, which are a 140 degree wide-angle lens webcam installed approximately 2.4 meters above and point to the floor, a XIAOMI 65" 4K TV display with wall-mount mounted on the wall horizontally with height of 1.4 meters (measured from bottom of the TV to floor) and a 2015 MacBook Pro 13" laptop with OS X Yosemite placed aside the TV display. Main software setup is a custom client-server system runs locally. Client web page runs on Chrome with full screen display. Back-end python flask server and inference engine runs on the local CPU.

Figure 2. Installation Setup.

System Logic: The system detects the human participant's position with real-time webcam monitoring, and switches between tracking and generating mode. When the participant steps into the interactive space, the tracking system will be activated and record her or his position history. At the moment the participant walks out, the AI system will be triggered and invoke the generative machine learning model to process the sketch. The Sketch-to-Shanshui translation algorithm generates a realistic Shanshui painting from the sketch input. The generated painting displays on the TV and presents to audiences in real-time.

3.1. Tracking System

The tracking system consists of a webcam and a custom color tracking algorithm. Mounted to the ceiling and face towards the floor, the camera captures the top view of the entire interactive zone, which is a 2.1×2.1 meter space covered by a gray carpet. When a participant puts on the yellow hat and steps in, the system tracks the location.

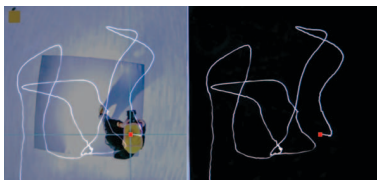

Figure 3. From Camera Capture to Sketch.

The camera captures a raw image as input I (adopted to 'RGB' standard) and the color value of the hat is C. The custom algorithm will compare each pixel in I to C by equation:

$$\mid I^{i}_{(x,y)} - C \mid \leq \mu \tag{1}$$

where i is frame number, μ is the threshold to determines if a pixel matches ($\mu = 5$ in this experiment), and (X, Y) describes pixel's location. The system counts all matched pixels in image I as N and calculates the average position value L, we use L to represent approximate point of participant:

$$L = \frac{\sum_{n=1}^{N}(X, Y)}{N} \tag{2}$$

The system sketches the participant's position by connecting the previous position to the current position with short line segments. We also utilize N to monitor the system and calculate the system state with equation $N < \alpha$, if N is lower than α, we consider that system is stand-by state (0), otherwise active state (1). When the state transitions from $0 \rightarrow 1$, it means that a participant steps in, the system starts tracking. And if the state changes from: $1 \rightarrow 0$ and participant moves out, it signals the AI system to sketch a Shanshui painting using the generative system.

3.2. Generative System

The generative system aims to translate a sketch to a Shanshui painting using AI. We employ the CycleGAN method to solve this as an image-to-image translation problem and we created a custom Sketch-to-Shanshui data set to train this model. CycleGAN utilizes two pairs of generators and discriminators in a cycle to constrain each other to translate back and forth between a pair of image domains. We train two generators and discriminators in the cycle but only utilize the Sketch-to-Shanshui generator in this system for this exhibit. With further engineering and experiments, we optimized the trained generator with Intel OpenVINO [3] to improve inference performance in the live exhibit. OpenVINO is a highly optimized AI deployment solution which affords us to run the model in real-time, whereas the original PyTorch model took several seconds per frame and wasn't suitable for live deployment.

3.2.1 Data Set

We collected and processed a collection of scanned Shanshui to make the Sketch-to-Shanshui data set. Shanshui paintings used in this project are all high-resolution scans with frames. In total, we have collected 108 Chinese Shanshui paintings from the open data platform of the National Palace Museum [4] and purchased 96 Chinese Shanshui scans from Chinese Taobao platform [1].

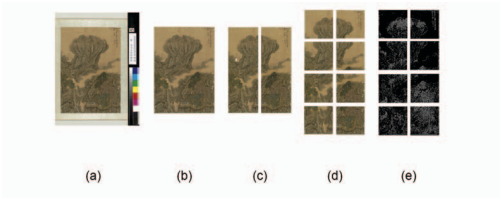

Figure 4. Creating Sketch-to-Shanshui Data Set.

We applied several computer vision techniques to create Shanshui and sketches. The processes include detecting and removing the frame (a); dividing the Shanshui into smaller square patches (c, d) and saving as Shanshui data; applying the Canny filter to all Shanshui data images and extracting edges as sketch data (e).

All 214 raw images are processed into 1770 sketch images and 1770 Shanshui paintings. All the images are downsized to the size of 512×512 pixels.

3.2.2 Train the Sketch-to-Shanshui Model

We implemented CycleGAN with PyTorch to train the network with a learning rate of 0.0002 and batch size of 1. The training task runs on a Linux server with Ubuntu 64-bit OS and six GPUs and took around 30 hours to complete. We save the Sketch-to-Shanshui generator into ONNX-format and then optimize it for inference with OpenVINO.

Inference with PyTorch on CPU is slow and we have applied two different methods to improve performance: First, we optimize the network architecture and replace repeatedly-used layers, such as instance normalization, and use OpenVINO to split and fuse multiply and addition operation into one operator. This method saves computational resources in terms of reads and writes from memory. Second, we apply the Intel MKL-DNN library which can accelerate CPU inference. After all these optimizations, the model with OpenVINO performs almost 8 times faster than the original model with standard PyTorch.

When using inference with consumer devices, especially laptops without a dedicated GPU, optimization is important for real-time experiences. When using inference with a CPU, the average time to generate one image is 306.42 ms. All the performance tests are performed on a 2018 MacBook Pro 15" with a 2.6 GHz Intel Core i7 processor.

3.3. Design and Implementation

We designed a client-server architecture to implement the system, the tracking system is on the front-end and a server processes the data query sent from the client. For the generative system, the optimized Sketch-to-Shanshui neural network serves via an API and is deployed on the back-end. All the custom software runs on a 2015 MacBook Pro 13" with OS X Yosemite. The webcam and TV display are connected via USB and HDMI.

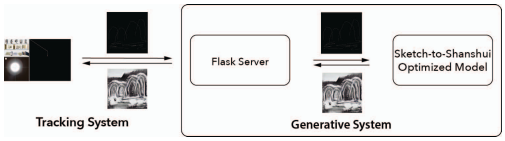

Figure 5. System Design.

The tracking system was programmed with a client-side JavaScript library called p5.js and it queries the Sketch-to-Shanshui API with a sketch input. The back-end generates a painting and posts to the front-end. The client-server system performs locally, and no network problem is required.

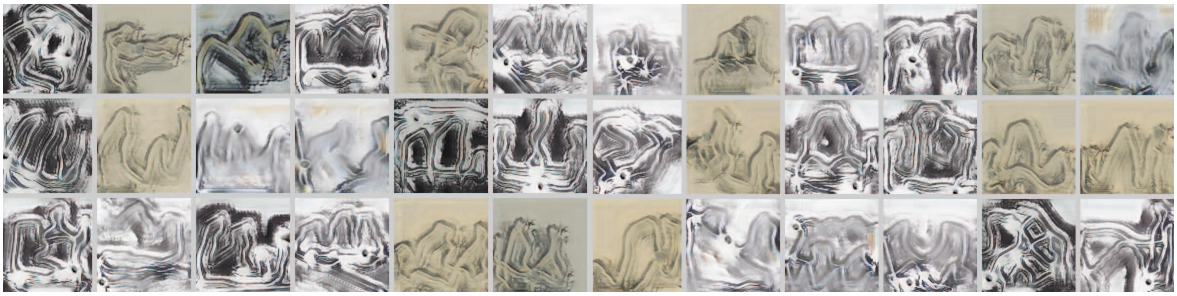

Figure 6. Selected Generated Shanshui from Day One.

The inference performance is satisfactory given the use case and we look forward to further optimizations that could make it even better. The generated result can be displayed on the TV live when a participant walks out. This liveness encourages a positive user experience to the observing crowd. When interacting with the system, participants can see a real-time sketch. And when they are satisfied with the sketch, they walk out of the installation and submit the sketch to our AI. The fast inference speed allows them see the immediate collaborative artwork in real-time.

4. Social Impact Discussion

This project has been exhibited at Shanghai DUOLUN Museum of Modern Art for 45 days. We were documenting on-site interactions on the first day. There were more than 300 engagements and generated paintings. There were long lines of excited observers waiting, but the experience was worth the wait and they spoke highly of the interaction and overall results. In figure 6, we selected some highlights to describe what participants created with our AI. Overall, most observer reacted very positively. There were 9642 participants who generated paintings during the exhibition.

Shanshui is a quintessential part of Chinese intellectual and spiritual life that has a very long and vivid history. But in modern times, the tradition is vanishing due to modernization. In this exhibit, we use the latest in technology to revive this tradition with cutting-edge AI that brings out the new while respecting old traditions. The seamless nature of the exhibit makes the technology fade into the background to help the user rediscover the grace of nature.

Here are some comments quote from participants (translated from Chinese):

This is interesting, never thought AI can do this! And it's fun, see, my son loves it! - A young father told the artist.

It's all right you don't know Shanshui. This is AI, AI can help you paint. - Grandmother told her granddaughter.

I just walked around and painted a Shanshui! It's magic, Shanshui Painting is fun! - Granddaughter responded.

With this live installation, we demonstrate some possibilities to seamlessly merge AI techniques with culture heritage. Participants, both old and young are mesmerized by this form of art, and it brings generations of families together to celebrate traditional values with the promise of the future heralded by younger generations.

5. Future Work

We would like to improve the tracking system in the future through the use of a depth- tracking instead of relying on the use of color. We plan to use an Intel RealSense [2] camera, so that participants can walk in and start playing directly(without the hat). We also plan to add the discriminator as a critic in the system. The AI critique can evaluate the quality of the generated Shanshui and give immediate feedback, to encourage participants to create more interesting and entertaining interactions between them and the AI.

Acknowledgement: We want to thank the Intel AI Academy for technical engineering and encouragement and the Intel Software Innovator Program for software and hardware support.

References

[1] Alibaba Corporation. Online shopping platform.
[2] Intel Corporation. Intel realsense technology, 2019.
[3] Intel Corporation. Intel distribution of openvino toolkit, 2019.
[4] National Palace Museum. Open data platform.
[5] Weili Shi. Shan shui in the world: A generative approach to traditional chinese landscape painting. 2016.
[6] Le Zhou, Qiu-Feng Wang, Kaizhu Huang, and Cheng-Hung Lo. An interactive and generative approach for chinese shanshui painting document. In *The 15th International Conference on Document Analysis and Recognition*, 2019.
[7] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.