

GUANGDONG AND HONG KONG UNIVERSITIES

“1+1+1” Joint Research Collaboration Scheme

粵港高校「1+1+1」聯合資助計劃

AI-Mediated Archives

A Revolutionary Approach To Interactive Narrative Generation

Led by Jeffrey Shaw (HKBU) & Dr. Xi Wang (BNBU)

Introduction

In the rapidly evolving landscape of media and technology, the way we interact with and consume media content is undergoing a significant transformation. One of the greatest challenges in this digital age is the difficulty of efficiently retrieving meaningful information from large, often sprawling, media archives.

This proposal aims to be at the forefront of addressing this challenge. We will develop an innovative system that leverages cutting-edge AI technologies to create user-driven narratives from cultural archives. This project will serve as a model for how AI technologies can be used to create immersive, interactive experiences with media archives in the future bringing a radical new dimension to public engagement with audiovisual archives more generally.

#digital archives #interactive narratives #large language models #computer-vision #information retrieval

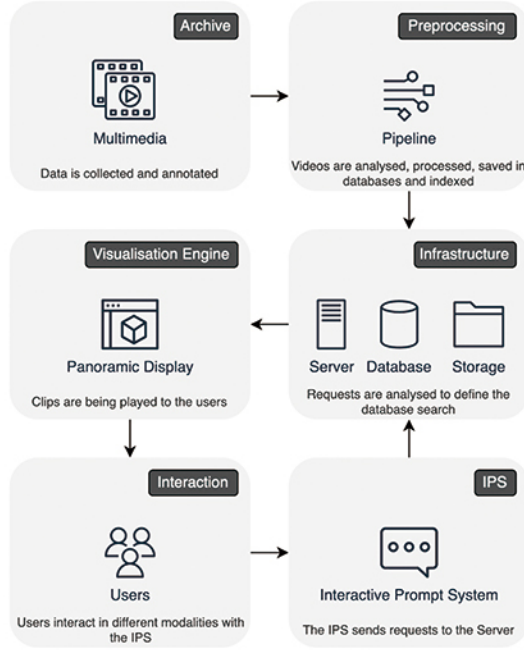


Chinese Shadow Puppetry

Our dataset focuses on Chinese shadow puppetry (皮影戲), a precursor to modern cinema. In live performances, intricately crafted figures are animated behind a cloth screen, their silhouettes cast by light and accompanied by live music and storytelling that often includes improvisation. With a history spanning more than 2,000 years, it is recognized by UNESCO as Intangible Cultural Heritage and warrants special attention and careful preservation.

BNBU's Learning Research Centre currently houses about 250 pieces. While valuable, this collection is limited in thematic scope. With the acquisition of Mr. Tong Min's collection, it has been significantly expanded, adding 1,195 historically significant pieces crafted from the now-rare donkey skin and dating to the 1960s-1970s.

As an initial step, we scanned the materials at ultra-high resolution and catalogued item properties such as material, style, role/function, and description, as well as associated data such as represented identity and appearances in plays. In a last step BNBU will invite professionals to perform and record traditional plays drawing materials solely from this new collection.



Narrative Engine & Database

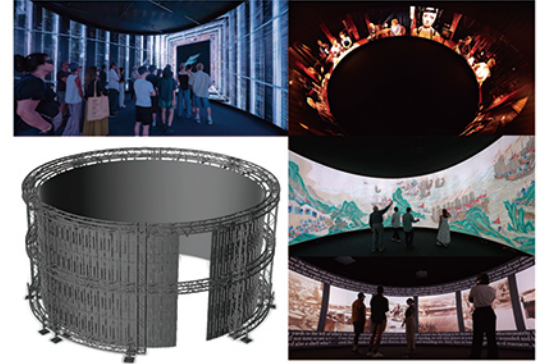
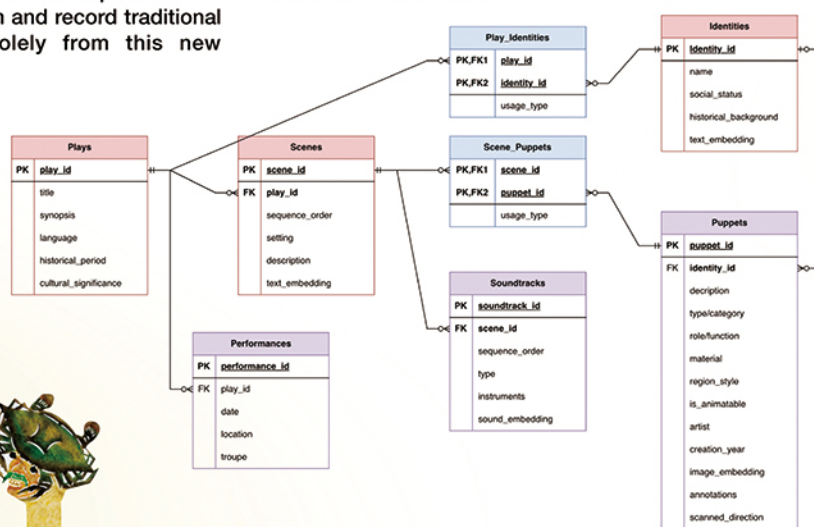
At the project's heart is the *Narrative Engine*, that processes cultural heritage data and interprets user input, and the *Narrative Database* which stores the digital twins, their meta data, related data and their relationship.

Our ingestion pipeline standardizes the puppet performance recordings and enriches them with script and asset data for semantic search. We transcode each video to a common format, extract audio, and align the provided play script to the timeline, then segment the performance into scenes and paragraph-level beats. For each segment we cut reference clips, sample frames at about 1 FPS, and compute image embeddings; in parallel we embed the corresponding script paragraphs.

High-resolution puppet scans are processed to generate image embeddings and are tagged with known puppet identities; identities specified in the script or detected in segments are attached as metadata.

Video frame and script embeddings are timestamped and linked to scene and puppet identities, then stored with the derived clips and metadata in a vector-ready index. This way we allow user prompts to function as semantic search over this multimodal archive. Additionally, we will reduce the high-dimensional frame embeddings using UMAP to organize video snippets in 3D virtual worlds. The reduction may extend to multi-modal embeddings to include other data such as sound.

Provisional Database Schematic for Chinese Shadow Puppetry Scans



Visualization Engine: nVis

The main visualization target for the newly developed applications is HKBU's nVis system located in the Visualization Research Centre. nVis is the world's first 360-degree 3D LED cinematic screen, providing an enveloping environment where audiences can immerse themselves in virtual worlds. Its cylindrical architecture is eight metres wide and four metres tall and consists of 400 LED panels providing a resolution of 2K x 13K totalling 26 million pixels. An integral aspect of nVis is its comprehensive instrumentarium of Human-Computer Interfaces.



Example Archive Demonstrator

To demonstrate the abilities of our *Narrative Engine*, we built an immersive VR browser that turns a multi-modal dataset into a navigable space of images, maps, and text. Inside nVis, the 100 stereographs float around the visitor; a handheld prompt lets users reorganize the gallery by meaning and foreground images to reveal their associated data. From there, a click jumps into a "semantic realm," where the book's passages are laid out as a spatial landscape, organized by reducing high-dimensional embeddings into the three dimensional virtual world. Speaking or typing a query steers a virtual camera to the most relevant excerpts, with nearby passages offering context and tangents to follow.

This representation makes new forms of exploration possible — nonlinear discovery by interest rather than chronology, cross-linking between photograph, place, and prose, rapid comparison of dispersed scenes and themes, and a more critical, reflective reading of the work's colonial framing — while preserving the immediacy and wonder of the original stereoscopic journey.

This archive demonstrator was accepted and presented in SIGGRAPH Asia 2025 and proven to be a solid foundation model for our video archive.

