Entangled Gazes: Reconstituting the Stereoscopic Box Set with AI and Virtual Reality

Emmanuelle Denove
emmanuelle.denove@epfl.ch
EPFL
Lausanne, Switzerland

Paul Heinrich Bethge
bethge@hkbu.edu.hk
HKBU
Hong Kong SAR, China

Dhruva Gowda-Storz
dhruva.gowdastorz@epfl.ch
EPFL
Lausanne, Switzerland

Adriano Viegas Milani
adriano.viegasmilani@epfl.ch
EPFL
Lausanne, Switzerland

Tai Leong Cheong
tlcheong2@cityu.edu.hk
CityUHK
Hong Kong SAR, China

Paul Bourke
paul.bourke@gmail.com

Sarah Kenderdine
sarah.kenderdine@epfl.ch
EPFL
Lausanne, Switzerland

Jeffrey Shaw
jeffreyshaw@hkbu.edu.hk
HKBU
Hong Kong SAR, China

Abstract

This paper presents a computational framework for critically reinterpreting colonial-era stereoscopic box sets. Using James Ricalton's China through the Stereoscope (1901) as a case study, we deconstruct its linear narrative using NLP methods for thematic modeling, semantic mapping, and colonial language identification. This analysis informs two complementary artworks in a large-scale 360° VR environment. Cross Eyed offers a guided experience, using the thematic models as lenses for inquiry and visually distinguishing identified rhetoric. Latent Cartographies, in contrast, empowers users to freely navigate the archive's raw semantic space, fostering diverse modes of experiential engagement with contested heritage.

CCS Concepts

• **Applied computing** → **Arts and humanities**; • **Computing methodologies** → *Natural language processing*; • **Human-centered computing** → Virtual reality.

Keywords

Archival Art, Stereoscopic Photography, NLP Methods for Art, Virtual Reality

ACM Reference Format:
Emmanuelle Denove, Paul Heinrich Bethge, Dhruva Gowda-Storz, Adriano Viegas Milani, Tai Leong Cheong, Paul Bourke, Sarah Kenderdine, and Jeffrey Shaw. 2025. Entangled Gazes: Reconstituting the Stereoscopic Box Set with AI and Virtual Reality. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (SIGGRAPH '25).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

SIGGRAPH '25, Hong Kong, HK

ACM ISBN 978-1-4503-XXXX-X/2025/06
https://doi.org/XXXXXXX.XXXXXXX

1 Stereoscopic Box Sets

At the turn of the 20th century, long before the advent of virtual reality, the most sophisticated form of immersive media was the "armchair travel" stereoscopic box set [4]. Pioneered by publishers such as Underwood & Underwood, these were not mere collections of images but carefully curated multimedia systems designed to simulate a virtual tour to distant lands. A typical set included a handheld viewer, up to 100 numbered stereographs, detailed maps charting the location of each photograph, and an expert guidebook. The technical core of this experience was the stereograph itself: a card with two photographs of the same scene captured from slightly different perspectives. When seen through a stereoscope, the brain fuses these paired images into a single view, creating a convincing and often startling illusion of three-dimensional depth.

For a public to whom international travel was largely inaccessible, these sets provided a powerful and educational "virtual tour" that fundamentally shaped how Western audiences perceived distant cultures and events. The combination of the immersive 3D illusion with a structured, first-person narrative allowed users to do more than just view images; they could follow a curated journey, witness historical moments, and even confront the battlefields of World War I, which were rigorously documented in this format [4]. This power and accessibility made the stereoscope a ubiquitous fixture in households and the dominant visual mass medium of its era.

Today, this powerful photographic medium has been largely forgotten by the public, though an estimated seven million stereographs were produced [4], and many complete box sets survive. These archives offer invaluable immersive windows into the past, and could allow us to revisit early paradigms of presence, virtuality, and the curated experience. Increasingly ubiquitous modern virtual reality are perfectly suited to reconstituting the stereo illusion and revive these collections for contemporary audiences. However, simply digitizing these collections is insufficient. The curated tours

Figure 1: An Altiscop stereoscope camera and Keystone library, c. 1930 (Heritage Auctions, 11 Nov. 2017, sale 5325, lot 65415). Courtesy of Heritage Auctions, HA.com.

they present are products of their time, with narratives deeply entangled with the imperialistic and civilizationist ideologies of the era. Reviving this material for modern audiences therefore requires a critical framework that can deconstruct the original narrative logic without censoring the challenging perspectives contained within.

While several projects have explored the visualization of individual stereographs in VR [10, 1, 9], to our knowledge none have attempted to revive the box set as a complete narrative system. This paper presents such a framework, leveraging Natural Language Processing methods to reconfigure the relationships between text and image to create new digital representations suitable for exhibition in virtual reality. We demonstrate this framework through two artworks that reinterpret a particularly compelling case study: James Ricalton's China through the Stereoscope

2 The Archive: Ricalton's China in 3D

2.1 Content and Ideological Context

James Ricalton's *China Through the Stereoscope* was published in 1901 by Underwood & Underwood. It begins as a typical stereo travelogue discussing Chinese commerce, daily life, architecture, local customs, and natural resources. However, Ricalton's itinerary was abruptly interrupted by the Boxer Uprising, a violent anti-foreign and anti-Christian rebellion, thrusting the reader into the midst of the conflict. As both photographer and author, Ricalton acts as a war correspondent and personal guide. His 100 stereographs transcend typical travel scenery, showing viewers the 'horrors of war': scenes of destruction, dead soldiers, 'Boxer' prisoners, and gruesome warnings erected by occupying forces, such as the heads of beheaded criminals. Ricalton's accompanying book framed the experience as a guided tour, with 100 chapters each contextualizing one of the stereographs. It positioned the stereoscope as a "miracle of realism" that allowed audiences to feel as though they were truly witnessing events they had only read about in newspapers [8].

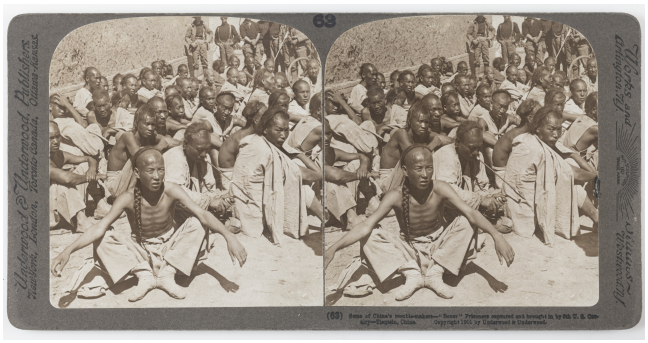

Figure 2: "Some of China's troublemakers-"Boxer" prisoners captured and brought in by 6th U.S. Cavalry - Tientain, China", 1901. Image 63/100. Courtesy of [Anonymized]

Yet, Ricalton's work is a product of its time, and his commentary frequently reveals a perspective deeply entangled with the imperialistic and civilizationist ideologies of the era. His practice of "stereoscopic ethnography" consistently frames his subjects through a lens of American exceptionalism and social Darwinism [12]. Such perspectives, which stand out to modern audiences, are representative of most other stereo commentaries of the time. Reviving and exhibiting this material requires acknowledging and critically address (without censoring or altering) such perspectives for modern audiences, which in itself represents an interesting scholarly exercise.

2.2 From Artifact to Dataset: Digitization and Preparation

2.2.1 The Book. A scan of the book was acquired from the UCLA California Public Library via the Internet Archive. Existing OCR of the source guidebook was unreliable. We generated a clean text corpus for NLP analysis by re-transcribing the book using Google Cloud's Vision API. The text was then subdivided into chapters, paragraphs, and individual sentences, forming the textual basis of our dataset.

2.2.2 The Stereographs. A complete 100-stereograph set generously donated by [Anonymized] was digitized with a medium-format camera at 3500x7000 resolution. A restoration workflow, including cropping, alignment, and damage repair, produced artifact-free image pairs suitable for comfortable stereoscopic display.

3 Methodology

Our artworks require transforming and challenging the linear narrative of the book and its accompanying images to present the corpus to a modern, non-expert audience. This was done using different computational and AI-assisted methods to extract, process, analyze, and reframe its different elements.

3.1 Reconfiguring text structures through LLMs

To create an alternative to the book's chronological organization, we implemented the topic modeling framework of van Wanrooij et al. [11]. The corpus was segmented into paragraphs as optimal

thematic units, with an LLM generating a comprehensive topic list and assigning a single topic per paragraph. To enable combinatorial queries (e.g., 'Food' + 'War' = nutrition during wartime), we developed a multi-label sentence-level assignment system leveraging LLM response probabilities: For each sentence-topic pair, the model determined relevance while logarithmic probabilities quantified assignment confidence, thereby establishing a probabilistic framework for polythematic categorization.

This methodology constitutes Cross Eyed's core interaction model and represents a deliberate curatorial shift. Where Ricalton imposed a linear narrative, our LLM collaboration exposes latent thematic networks within his text, enabling polyvocal interpretations that transcend authorial intent.

3.2 Exploring the world of text embeddings

A different approach for deconstructing the text employs semantic embeddings via the BGE-M3 model, encoding textual elements into a 1024-dimensional latent space to quantify content relationships through vector operations [2]. User queries are similarly embedded, enabling semantic distance calculations for content retrieval and relevance visualization [7]. Under the premise that chapter text fragments intrinsically link to their corresponding images, an image ranking strategy is implemented: The 50 nearest text embeddings to a query (empirically determined) are normalized, inverted, and aggregated into a composite score reflecting proximity and frequency. This facilitates image search functionality. For spatial representation, Uniform Manifold Approximation and Projection (UMAP) reduces dimensionality to three axes while preserving global thematic structure [6].

Critically, this semantic landscape constitutes a cultural artifact shaped by the BGE-M3 model's architecture and training data. The resulting projection is not an objective textual representation but a technologically mediated interpretation. Latent Cartographies invites users to navigate this construct, simultaneously exposing Ricalton's original cartography and its AI-generated counterpart.

3.3 Automatic Decolonization

As established in the Introduction, this corpus features colonial language requiring systematic detection and contextualization. To identify optimal automated detection methods, we evaluated multiple techniques against a manually curated validation set of 140 sentences (70 offensive, 70 neutral). Two primary approaches were assessed: sentiment analysis models detecting affective polarity, and an LLM-based classifier inspired by Chiu et al. [3]. The latter was deployed in two configurations - with and without historical contextualization noting the text's 1900 Sino-centric provenance.

Performance results (Table 1) demonstrate that while no method achieves perfection, the contextualized LLM approach significantly outperforms sentiment analysis. This detector is implemented in the Cross Eyed installation for real-time textual flagging. Critically, these annotations function not as definitive judgments but as epistemic provocations: By rendering the AI's assessments visible, the work foregrounds the inherent challenges in operationalizing colonial discourse detection, directing attention to the constitutive ambiguities of such language itself.

Table 1: Colonial Language Detection Methods

Model	F1-Score	Comments
NLTK VADER	0.477	Sentiment Analysis
TextBlob	0.367	Sentiment Analysis
Transformers	0.537	Sentiment Analysis
LLM-based	0.816	w/-out context
LLM-based	0.867	w/ context

4 A Pair of Artworks

We presented two artworks borne from our computational reconfiguration of the archive. Both installations reframe the modern user's engagement with a historical stereograph box set, to reinterpret this narrative journey through 20th century China, challenging the colonial and imperialist discourse of the time.

Though unified by a shared experimentation phase and technical platform, the artworks adopt divergent strategies: The first artwork employs thematic "lenses" to deconstruct the archive's narrative, fostering critical analysis of Ricalton's dual voice as observer and ideologue. The second piece utilizes text-embedding manifolds to create nonlinear, user-driven explorations of semantic connections within the archive.

Both artworks are presented in an 8x4m, cylindrical VR multi-user theatre based on the AVIE system [5], using active shutter glasses, a 29.4-channel spatial audio system, and HTC Vive-tracked tablets for interaction. This stereoscopic system reconstitutes the stereo illusion of Ricalton's stereographs allowing them to be viewed as intended: in immersive 3Ds.

4.1 Cross Eyed: Decolonizing the Gaze of 19th Stereophotography

Cross Eyed, reimagines engagement with the historical archive by inviting users to view it through a series of conceptual lenses. Its dual aim is to make the dense material more accessible while creating a visual paradigm for critically decolonizing the author's gaze.

These lenses (such as 'Colonialism', 'Missionary Work', or 'Family Structures') are thematic categories generated by the LLM-based topic modeling strategy detailed in Section 3.2. Via a tablet interface, visitors can select a single lens or combine multiple lenses (e.g., 'War' and 'Food') to create combinatorial queries. This interaction model bypasses the book's linear narrative, instead revealing the latent, often unexpected, connections between different subjects in Ricalton's journey, and gives users accessible entry points to query the archive.

Upon selection, the 360-degree screen is populated by corresponding stereographs and scrolling text fragments. The artwork's visual language is built around the ticker tape, a medium chosen for both its potent psychological effect and its rich cultural meaning. The primary goal is to compel engagement with historical text that viewers might otherwise ignore. Drawing on the ticker's innate "affective property," we leverage two fundamental cognitive principles. First, the continuous horizontal motion creates strong visual salience, involuntarily capturing attention in a way static

Figure 3: Visitors viewing stereographs and text through the 'War' lens within the artwork *Cross Eyed*.

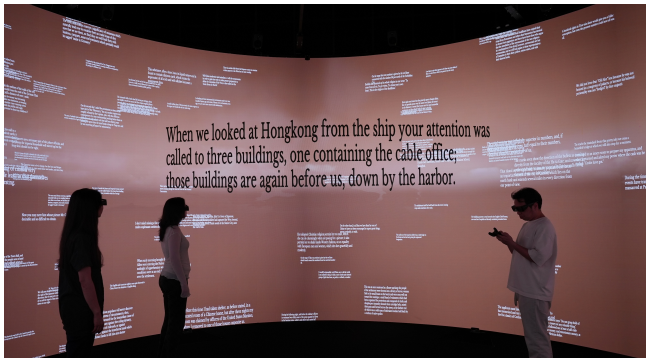

Figure 4: Visitors engaging with the Conscious Realm within *Latent Cartographies*

Figure 5: Visitors exploring the Physical Realm *Latent Cartographies*

text cannot. Second, the sequential revelation of text creates a powerful information gap. This forces the viewer into an anticipatory loop of prediction and fulfillment as they wait for each sentence to complete, effectively "hooking" them into a state of sustained reading.

Beyond this cognitive hook, our choice of the ticker is a critical intervention inspired by artists like Jenny Holzer, who subverted the medium's authority. The ticker is culturally coded; we associate its form with the objective, real-time authority of financial data and breaking news. By channeling Ricalton's subjective, often biased, 100-year-old voice through this modern medium of "truth," we create a deliberate cognitive dissonance. The artwork casts Ricalton as a contemporary war correspondent, but the clash between the authoritative format and the historically-contingent, imperialist content forces a critical re-evaluation. The viewer is no longer passively consuming a historical text; they are actively grappling with the mismatch between the medium and its message, which is the central goal of decolonizing Ricalton's gaze.

To make this critical grappling explicit, our system uses the algorithm developed in Section 3.4 to detect and visually flag sentences containing imperialist or racist perspectives. The text is not censored; instead, problematic passages are distinguished from neutral descriptions by their color and typography. This visual distinction transforms the act of reading into an act of critical inquiry, empowering viewers to identify the author's biases and form their own judgments about a historical remnant presented in its entirety.

4.2 Latent Cartographies

In contrast to the topic-based approach of the first artwork, this piece places the continuous latent space at the core of an interaction paradigm that emphasizes on user agency over algorithmic suggestion. Inspired by the author's foreword on people's physical and conscious realities, we created two interconnected scenarios:

Conscious Realm:

Users entering the cylindrical display encounter a three-dimensional semantic landscape where text fragments are spatially organized by UMAP-reduced text embeddings. Initially, this latent structure appears imperceptible, presenting a seemingly chaotic constellation. Sparse clusters of free-floating phrases within an apparently boundless space evoke cognitive associations, situating spectators within the author's mental universe. This reorganization enables novel exploration of semantically connected passages, circumventing both the original structure and authorial intent. Prompting the system either by text or speech input, propels the audience to a different location in that space. As the text slowly passes their gaze, visitors may catch inspiration for their next prompt. Upon arrival, the most relevant text fragment is displayed in front of them as depicted in Figure 4.

While cursor-based navigation remains possible, its functionality is deliberately constrained to privilege natural language interaction. To further deepen immersion, manually positioned harmonic phrases from piano, violin, and cello inhabit the same three-dimensional space. This ambient sonification enhances the illusion of infinite expanse while providing navigational cues during latent space traversal.

Physical Realm: Selecting a text fragment transitions users to a cylindrical arrangement of the complete image collection, with the corresponding image foregrounded (though not yet selected). Cursor hovering dynamically modulates image brightness and spatial positioning: targeted images advance toward the viewer while laterally displacing adjacent content. Users may alternatively employ natural language queries to rank images by semantic proximity to text inputs, as previously described.

Selecting an image first moves a topographical map of China in the background to the chapter's region, reveals the photo's precise

location on the accompanied map and displays the full chapter, as seen in Figure 5. Based on the similarity to the user's previous prompts, each text fragment is highlighted to help the user finding their region of interest. Users may themselves mark a section as concerning, fostering critical discussion on spot and putting the human in the center of meaning making. Clicking a text fragment returns them to the Conscious Realm, revealing its associated cluster and closing the loop.

5 Conclusion

This research demonstrates how computational methods can transform colonial-era stereoscopic box sets into dynamic virtual spaces for critical engagement. Using James Ricalton's *China through the Stereoscope* as our case study, we first conducted extensive NLP experimentation to computationally deconstruct the archive. Building on these foundations, we developed two complementary artworks that reinterpret historical narratives through immersive technologies while addressing the complexities of colonial discourse.

Cross Eyed adopts a top-down analytical approach, leveraging LLM-generated thematic lenses to systematically deconstruct the archive. By enabling users to combine categories like "War" and "Commerce" through a directive interface, and by visually flagging imperialist rhetoric through typographic interventions, this work makes implicit biases explicit while training critical viewing practices. In contrast, Latent Cartographies employs a bottom-up exploratory paradigm, utilizing text-embedding manifolds to reveal emergent connections within the archive. Its immersion-driven, open-ended navigation through semantic and geographic spaces guides users on their own journey through latent relations.

Despite their contrasting methodologies—one decomposing the archive through algorithmic categorization, the other recomposing it through user-driven emergence — both works achieve critical shared objectives. They decentralize Ricalton's authoritative narrative through spatial-visual metaphors, materialize historical bias in tangible ways, and transform passive spectators into active participants through embodied interaction in our custom 360° VR environment. This dual approach demonstrates that meaningful engagement with contested heritage requires both analytical frameworks to expose power structures and exploratory spaces for personal meaning-making.

As critical media practice, this project offers new pathways for examining technological infrastructures of historical representation. Our integration of machine learning with immersive systems creates a living archive where past and present continuously reconfigure each other—not to simplify colonial history, but to complicate viewers' relationship to it. The artworks exemplify how media technologies can challenge dominant narratives by making archival interrogation an experiential process rather than a didactic lesson. Future applications of this framework could extend to other collections where technological recontextualization might foster restorative dialogues with contested pasts.

References

[1] Krystal Boehlert. 2016. Stereographs on your smartphone. *Visual Resources Association Bulletin*, 43, 1.

[2] Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. Bge m3-embedding: multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. (2024). arXiv: 2402.03216 [cs.CL].

[3] Ke-Li Chiu, Annie Collins, and Rohan Alexander. 2021. Detecting hate speech with gpt-3. *arXiv preprint arXiv:2103.12407*.

[4] William Culp Darrah. 1977. *The world of stereographs*. William Darrah Culp.

[5] Matthew McGinity, Jeffrey Shaw, Volker Kuchelmeister, Ardrian Hardjono, and Dennis Del Favero. 2007. Avie: a versatile multi-user stereo 360 interactive vr theatre. In *Proceedings of the 2007 workshop on Emerging displays technologies: images and beyond: the future of displays and interacton*, 2–es.

[6] Leland McInnes, John Healy, and James Melville. 2020. Umap: uniform manifold approximation and projection for dimension reduction. (2020). https://arxiv.org/abs/1802.03426 arXiv: 1802.03426 [stat.ML].

[7] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084. http://arxiv.org/abs/1908.10084 arXiv: 1908.10084.

[8] James Ricalton. 1901. *China Through the Stereoscope: A Journey Through the Dragon Empire at the Time of the Boxer Uprising...* Underwood & Underwood.

[9] Bryan Ricupero, Glory Dalton, and Amanda Lehman. 2019. Building and enhancing stereographic digitial collections: working across departments to augment reality. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 343–344.

[10] Daniel Taipina and Jorge CS Cardoso. 2022. Spectare: re-designing a stereoscope for a cultural heritage xr experience. *Electronics*, 11, 4, 620.

[11] Cascha van Wanrooij, Omendra Kumar Manhar, and Jie Yang. [n. d.] Topic modeling for small data using generative llms. ().

[12] Mitchell Arthur Winter. 2018. *Optics of American Empire: James Ricalton and Stereoscopic Ethnography in Early Twentieth Century India, 1888-1907*. University of California, Santa Cruz.