ROGER: Extracting Narratives Using Large Language Models from Robert Gerstmann's Historical Photo Archive of the Sacambaya Expedition in 1928

Mauricio Matus[1], Diego Urrutia[2], Claudio Meneses[2] and Brian Keith[2]

[1] *School of Journalism, Universidad Católica del Norte, Antofagasta, Chile*
[2] *Department of Computing & Systems Engineering, Universidad Católica del Norte, Antofagasta, Chile*

Abstract

This article presents the ongoing work on developing a methodology for the systematic analysis and narrative of heritage image archives, focusing on the photo archive by Robert Gerstmann of the Sacambaya Expedition in 1928. This work combines state-of-the-art artificial intelligence techniques, such as advanced algorithms used in computer visión like convolutional neural networks, with Large Language Models (LLM) for generation purposes. The intent is to establish a practical and accessible framework in this area for institutions and individuals. The proposed method incorporates human-generated image labels with LLMs to produce narratives that aid researchers and users in their sense-making process while they explore a large archive of images. Through this iterative process, we aim to contribute not only to the understanding of this specific historical photo collection but also to the broader development of scalable solutions for the exploration and interpretation of heritage image archives. We seek to achieve a deeper understanding of the contents and meanings of the analyzed files, suggesting and highlighting new clustering of these materials and thematic/narrative connections that may not have been considered by a human observer.

Keywords

Narrative Extraction, Heritage Image Archives, Sacambaya Expedition, Large Language Models, Image Labeling.

1. Introduction

The emergence of new technologies and the availability of a vast photographic archive have motivated a multidisciplinary project that explores the potential of these two elements in expanding the scope of multidisciplinary research. This ongoing research in the field of computational narrative extraction aims to develop a methodology for the analysis and semi-automatic construction of meaning and narratives from historical image archives using Large Language Models (LLMs). On the heritage side of this research, we aim to uncover and specify narratives inherent in large banks of photos for which there is limited information and dissemination [1].

1.1. The Sacambaya Expedition photo archive

Robert Gerstmann (1896-1964) a German engineer and photographer, arrived in Chile in 1925. His work has been preserved in its original physical format since 1964 in Antofagasta, Chile. This material consists of 43,475 negatives and 15,054 positives in different formats, representing a period of photographic capture of approximately 40 years [2]. The images detail landscapes from the beginning of the 20th century from the heights of Bolivia to Antarctica, the Pacific islands, and the Andes Mountain range [3]. In this context, we focus on the Sacambaya Expedition archive, a part of the photographic work of Robert Gerstmann.

In January 1928, Edgar Sanders, a Swiss engineer, established a company in London to search for an alleged Jesuit treasure hidden in an old monastery located in a ravine in the province of Inquisiví, Bolivia. The expedition comprised 21 individuals with diverse professional and military backgrounds, including 19 English citizens, 1 German, and 1 North American. The team scoured four different locations in the sector for five months, but their efforts proved fruitless, and they returned to Europe in November of that year [4]. We note that most of the records related to this expedition have not been digitized before the present project. It is estimated that only approximately 15% of these images are digitized and accessible online [5].

1.2   Proposed model

This research seeks to exploit the power of LLMs [6][7] and image processing techniques to extract narratives from a compendium of historical images. This endeavor aims to augment the corpus of historical knowledge by providing a narrative context to image archives that capture specific historical events. The overarching goal of this project is to establish a comprehensive framework/pipeline, named Robert Gerstmann Repository (ROGER), that enables efficient exploration, categorization, and semi-automated extraction of narratives implicit in previously unexamined heritage archives. The proposed methodology is designed to be iterative and incremental, ensuring clear documentation of progress through each phase. The efficacy of the methodology is illustrated through a case study centered on a historical event, utilizing the framework to narrate its story systematically.

The ROGER Narrative Pipeline unfolds in a structured, multi-phased approach, incorporating human expertise and AI in a collaborative narrative pipeline. Initially, images are labeled through a combination of AI-driven algorithms and human judgment, resulting in a curated and contextually enriched dataset. Subsequent phases involve the use of AI to generate descriptive narratives and cluster these into thematic groups. The proposed pipeline integrates the use of prompt engineering loops [8] — a process where human input is used to iteratively refine AI outputs — in the narrative extraction process, thus ensuring that the emerging narratives are not only accurate but also resonate with human interpretative frameworks. The final phase of the process is the drafting and construction of a coherent narrative with human feedback from the clustered image data. This interactive and iterative process between AI and human intelligence is instrumental in producing a polished and nuanced narrative output, ready for presentation and scholarly exploration.

This paper presents the results from the application of the ROGER Narrative Pipeline to the 1928 Sacambaya Expedition historical archive, uncovering the implicit narratives embedded within historical image archives. The proposed approach contributes to the interdisciplinary dialogue on narrative extraction, advancing our understanding of computational narrative construction in historical research. This represents a novel initiative to systematically decipher the stories enshrined in historical visual records. Finally, in the conclusion section, it is noted that this is the first work that has attempted to computationally unearth the stories contained in any of Robert Gerstmann's historical photo archives.

2.   Technological framework

The computational analysis and interpretation of historical image archives is an emerging interdisciplinary field that integrates computer vision [9], natural language processing [10], information retrieval [11], and historical research methods [12]. Several recent projects [13] have demonstrated the potential of computational techniques to aid in making sense of large-scale image archives and constructing historical narratives from them. From a technological point of view, the process described in the following Methodology section uses two key technologies, in order to implement a pipeline that takes a set of related patrimonial pictures

as input and generates a coherent set of narratives as output. These key technologies are LabelBox and LLMs (e.g., ChatGPT).

2.1.   Label Box

Labelbox is a machine learning annotation platform that simplifies the creation and management of annotated datasets, which are vital for AI development. It supports a variety of data types and provides tools for both manual and semi-automated annotation, aimed at increasing efficiency and accuracy. The platform encourages teamwork with its collaborative features and maintains high-quality annotations with robust quality control measures. Its user-friendly interface and integrative capabilities with machine learning workflows make it accessible for users of different skill levels, streamlining the annotation process from start to finish.

2.2   LLMs and ChatGPT

LLMs are advanced AI systems designed to process and generate human-like text. They have been trained on a massive amount of data and can understand and generate text in a wide variety of languages and styles [14][15][16][17]. ChatGPT, developed by OpenAI, is a specific example of an LLM designed for conversational interactions. It uses GPT (Generative Pretrained Transformer) architecture to understand and respond to user inputs. ChatGPT has been trained on a wide range of internet text to grasp and mimic human-like conversational patterns.

Sensemaking of archives requires synthesizing across individual images to construct a higher-level understanding. Computational techniques for visual storytelling aim to build such narratives from image sequences [18]. However, this typically relies on constrained domains with limited vocabularies [19]. Our work leverages pre-trained LLMs capable of open-domain generation to construct narratives for historical image archives. In summary, our methodology builds upon advances in image recognition and LLMs while innovating in integrating these techniques for computational sensemaking over historical image archives. We believe this approach can provide both a macro-level narrative as well as a detailed understanding grounded in the image contents.

3.   Methodology

We present the ROGER Narrative Pipeline for the extraction of computational narratives from visual datasets. The general methodology is presented in Figure 1. This process commences with a systematic labeling phase where a set of input images is semantically annotated using a combination of software tools and human oversight, producing a curated dataset. This curation involves the enrichment of the images with contextual metadata in the form of image labels. These labels enhance the depth and relevance of the information that will be used in narrative construction.
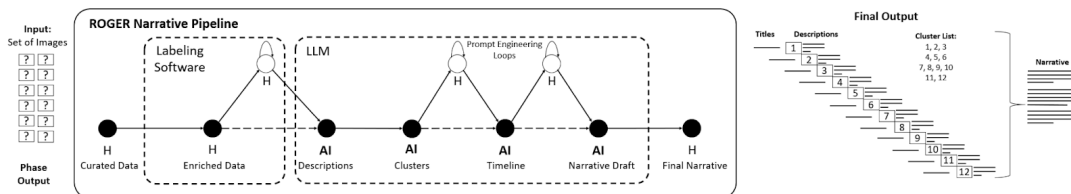

Figure 1: The ROGER Narrative Pipeline - A systematic framework illustrating the collaborative human-AI process for generating structured narratives from a dataset of images. The pipeline integrates human expertise with advanced AI algorithms for data curation, enrichment, and narrative construction, culminating in a coherent final narrative output

Central to our pipeline is the integration of an LLM, which generates textual descriptions and titles from the enriched image data. These descriptions are the bedrock upon which the narrative structure is built. Subsequently, we use the LLM to perform clustering on the textual descriptions to organize the images into clusters or coherent thematic groups (Figure 3), followed by the establishment of a timeline, ordering these clusters and images to create a draft narrative sequence. Integral to this process are the prompt engineering loops, where human operators iteratively refine the AI prompts based on the outputs to produce a final narrative (Figure 4). This iterative process is pivotal, allowing for the human operator's critical and creative inputs to sculpt the narrative, ensuring structural and thematic integrity.

The final stages of the pipeline revolve around the transformation of the AI-generated timeline into a narrative draft. This draft undergoes a human-led finalization process, where narrative/theme experts refine the storyline, ensuring linguistic precision, narrative flow, and overall coherence. The result is a final narrative that provides a textual representation of the visual data in a narrative format. The final output contains the ordered list of images, their titles, their description, and the associated narrative. Furthermore, the output is accompanied by a detailed cluster list that provides an overview of the narrative elements and their organization, thereby offering transparency into the narrative structure and content. Through this methodical and collaborative approach, the pipeline achieves a high-fidelity narrative extraction from visual inputs, demonstrating the potential for robust human-AI collaboration in this domain.

4.   Results

The main results associated with each stage of our pipeline, around a subset of 12 heritage images, are presented here.

4.1.   Data curation and sampling

To demonstrate the capabilities of our methodology, we present the results on a representative subset of 12 pictures from the photo archive of the Sacambaya Expedition in 1928 [20][21]. This archive consists of 545 negative originals in 10 x 15 format. From this collection, we discarded 45 photographs in the present analysis due to their defective nature and/or being over- or under-exposed, as they did not provide relevant information for labeling and subsequent categorization.

Prior to this study, this photographic archive was sustained by historical sources and the account of at least one of its members published in 1934 [22]. The historical narrative suggested by these sources coincides with what is seen in the heritage images found in the archive. These materials consist of a distinctive first phase of photos made on board a ship, another category of materials showing means of transportation and human displacement tasks, to finally settle into a last general category of images that exhibit the efforts of an excavation and the logistics that this semi-industrial human operation entails [5]. Therefore, these 500 photographs were pre-organized into the following thematic clusters:

- Journey by ship: Subgroup identified as "London to Arica" (LTA) with 69 images.
- Journey by land: Subgroup identified as "Arica to Sacambaya" (ATS) with 187 images
- Excavation sites: Subgroup identified as "Sacambaya" (SAC) with 244 images.

From each of these thematic groups we intentionally selected 4 images to exemplify the general progress of the expedition, the methodology employed, and suggest a temporal order in the narration of the journey.

4.2 Data enrichment through manual labeling

The enrichment of the original 500 images through the application of annotation tagging (both classification and object) by humans was facilitated by the use of the LabelBox [23] platform. The use of automatic classification algorithms was considered; however, the decision was made to prioritize establishing a baseline developed by humans. The procedure is comprised of several distinct phases.

4.2.1 Dataset generation

A corpus of 500 images was curated, encompassing approximately 2.1GB of data. These images were standardized to a width of 1920 pixels and subsequently compressed using the cjpeg software [24] to reduce their file size, culminating in a consolidated dataset of approximately 370MB.
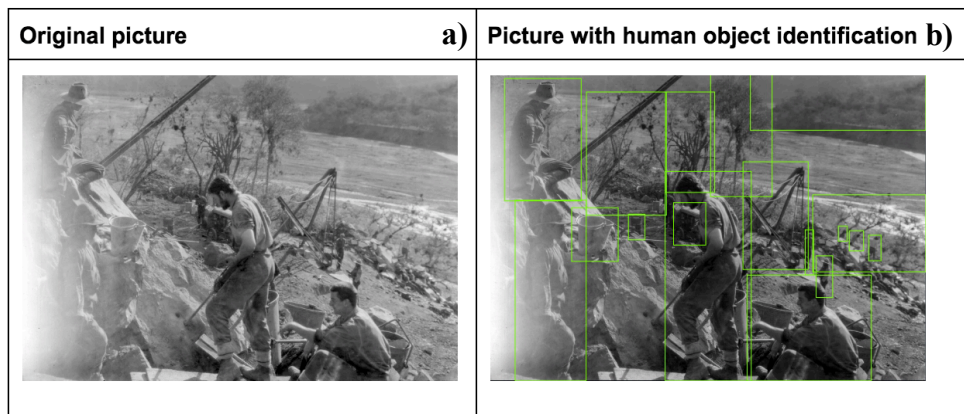

Figure 2: a) Original photo in the context of the Sacambaya expedition; b) The same photo with objects identified by humans (rectangular green box, indicating its type and name from a predefined set of categories).

4.2.2 Ontology definition

The ontology for the annotation process was crafted following an initial analysis of the photographs. This ontology comprised two sets of annotations: General Classification and Specific Objects. The development of this ontology was a critical step in ensuring that the annotations would be comprehensive and consistently applied across the entire image set by the research assistants. The detailed categorization was designed to facilitate nuanced analysis and to support the study's objectives by providing rich, structured data.

The general classification elements consisted of 22 labels (such as trees, road, city, boat, excavation, beach, square, etc.) that were defined based on a prior visual analysis by the authors. The specific elements were grouped into categories: person, animal, object, transportation, and landscape. Each of these categories encompassed between 5 to 7 labels. This structured approach to classification allows for a detailed and organized analysis of the photographs. By having both general and specific categories, the researchers could ensure that the labeling process was thorough and nuanced, capturing both broad and fine-grained details within the images.

4.2.3 Annotator recruitment and annotation process

Ten undergraduate students from the fields of journalism and computer engineering were recruited to serve as research assistants. These individuals were selected based on the criterion of having completed at least 50% of their academic program and having prior participation in various research projects. Over the course of two weeks, these individuals were tasked with the systematic manual delineation of bounding boxes and the subsequent

assignment of labels corresponding to the identified objects within the images. Each photograph was labeled by at least two students and subsequently reviewed by the researchers. The team identified a total of 4,868 objects, of which 44.1% were classified as people, 22.17% as landscape details, 15.14% as various objects, with the remaining categories including animals and transportation. The average number of annotations produced by each student was 54.4.

4.2.4 Annotation extraction

The annotations were subsequently extracted from LabelBox in JSON format, which facilitated their processing and analysis through the Python programming language. In a detailed analysis of the dataset, 12 images were selected as representative samples from the collection of 500, and their annotations were obtained for further investigation. This structured approach to annotation not only enhances the reliability of the data but also ensures a level of granularity that is conducive to subsequent computational analysis.

Table 1
An example of the prompt for AI-generated description.

|	
Prompt	You are an expert documentary filmmaker and you are exploring the historical archive of photos of the Sacambaya Expedition in Bolivia from 1928, taken by the German photographer Robert Gerstmann. The expedition sought to find the lost treasure of the Jesuits, hidden somewhere near Sacambaya. You have the following picture that was tagged by humans in Spanish. Repeated tags mean that multiple objects of the same type appear in the image. Here are the tags:
	Elementos Generales: - Quebrada / Valle - Ruinas / Escombros - Maquinaria - Personas - Árbol / Arbusto
	Elementos Especificos: - Paisaje Tipo Árboles / Arbustos - Persona Sexo Masculino Edad Adulto - Objeto Tipo Otro Nombre Cubeta - Objeto Tipo Otro Nombre Barril - Transporte Tipo Tren / Locomotora / Carro - Objeto Tipo Maquinaria - Paisaje Tipo Árboles / Arbustos
	Generate a short title for the photo and a description of the events depicted by the photo. The description must be one paragraph long. Please avoid flowery language, your description must be written in an academic style.
Output	**Title: "Excavation Effort in Sacambaya Valley, 1928"**
	Description: This historical photograph captures a scene from the 1928 Sacambaya Expedition in Bolivia, overseen by German photographer Robert Gerstmann. The image depicts a group of adult male workers engaged in what appears to be an excavation or construction activity within a quebrada or valley, surrounded by the natural landscape of trees and shrubs. The men are observed utilizing basic equipment and machinery of the era, including buckets and barrels, indicative of manual labor practices. A rudimentary transport system, possibly a narrow-gauge railway, is also visible, suggesting the movement of materials or debris. The terrain is littered with ruins and rubble, perhaps remnants of previous structures or the outcome of the ongoing excavation. This photo serves as a visual record of the intense human and mechanical effort invested in the search for the Jesuits' lost treasure near Sacambaya, reflecting both the arduous conditions and the determination of the expedition members.

Table 2

Clustering prompt and corresponding output. The prompt was followed by the list of all the previously generated descriptions of the photos. Only the relevant parts of the output are shown due to space constraints.

Prompt	You are an expert documentary filmmaker and you are exploring the historical archive of photos of the Sacambaya Expedition in Bolivia from 1928, taken by the German photographer Robert Gerstmann. The expedition sought to find the lost treasure of the jesuits, hidden somewhere near Sacambaya. You have the following list of descriptions of photos. Your first goal is to identify clusters of the photos based on their content and their temporal order. Your second goal is to generate narrative descriptions of the identified clusters. Note that the provided descriptions might not be shown in the correct order and the IDs are not associated with any ordering. Please avoid the use of flowery language, the description of the clusters should be written in an academic style.
Output	Image IDs for each cluster. Cluster 1 "Maritime Prelude": 23, 65, 60, 30. Cluster 2 "Expedition Life and Challenges": 40, 91, 85, 71. Cluster 3 "Industrial and Excavation Efforts": 46, 37, 81.

4.3 Descriptions and clustering

Following the manual labeling process, we generated narrative descriptions of each picture in the dataset. These descriptions are generated using ChatGPT with GPT4. In particular, we prompt the model with minimal context about the Sacambaya Expedition, upload the image, and provide the human-generated labels described previously. Table 1 illustrates the prompt design used to generate a narrative description and title for the picture shown in Figure 2 and the output generated by GPT4. We note that our annotators created the tags in Spanish as this was their native language, they were left untranslated in the prompt. We note that there was not a significant difference in translating them beforehand.
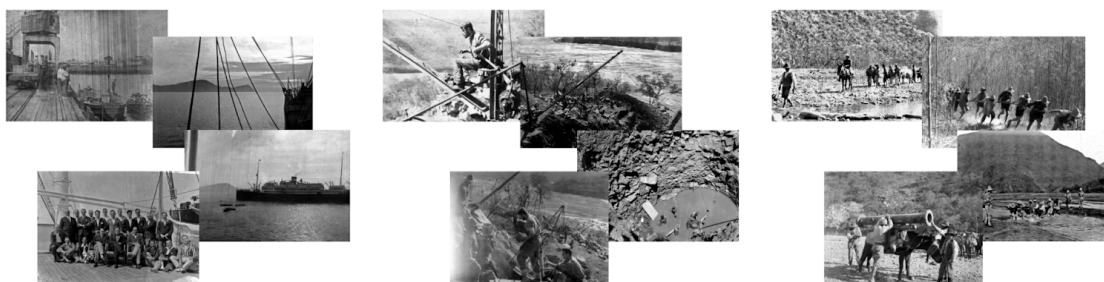
Cluster 1: "Maritime Prelude".   Cluster 2: ""Expedition Life and Challenges".   Cluster 3: "Industrial and Excavation Efforts".

Figure 3: ChatGPT proposal of coherent thematic groups clustered the images around common elements derived from manual labeling and descriptions of the images created with LLMs. We include the unsolicited titles created by GPT for the clusters.

The minimum context information supplied to the model was the name, year, and purpose of the expedition. The prompt included the names of some key places in the historical narrative (Bolivia and Sacambaya) and the name and nationality of the photographer (see Tables 1 to 3). Following the construction of all the narrative descriptions, we used another prompt to ask the LLM to cluster the images based on their content and reorganize them

chronologically. The prompt used all the image descriptions generated beforehand to generate these clusters. The prompt is shown in Table 2.

4.4 Timeline and narrative draft

Following this clustering process, we asked the LLM to generate a timeline of the photos followed by a draft of the final narrative. We note that our work required extensive prompt design to ensure that the generated descriptions, clusters, and final narratives were coherent. We show the final version of the timeline and narrative extraction prompt alongside the corresponding narrative summary output in Table 3. The final timeline is shown in Figure 4 with the corresponding images. These two elements (timeline and narrative summary) represent the final output of the proposed pipeline.

We note that our proposed prompt structure forces the model to hypothesize a plausible chronological order before generating the narrative draft. The data set of historical photos does not contain explicit temporal or spatial information not sharing with the model the thematic grouping mentioned in point 4.1. Therefore the LLM must be able to infer the order of the photos based on the general context of the expedition. While a human could also help with the timeline generation process and provide further information via prompting, the amount of photos in a full archive can make this generally unfeasible and too complex to scale. Thus, for scalability purposes, we rely on the capabilities of the LLMs to make sense of the underlying temporal and spatial information.

In general, sorting an unordered set of photos in the correct temporal order is a complex problem. In this context, GPT4 had issues when generating the correct timeline with only limited contextual information, which is why we had to include additional context in our prompt design. In reviewing the results, we found that the LLM model presented historical events in the correct chronological order, as long as it was provided with minimal contextual information.
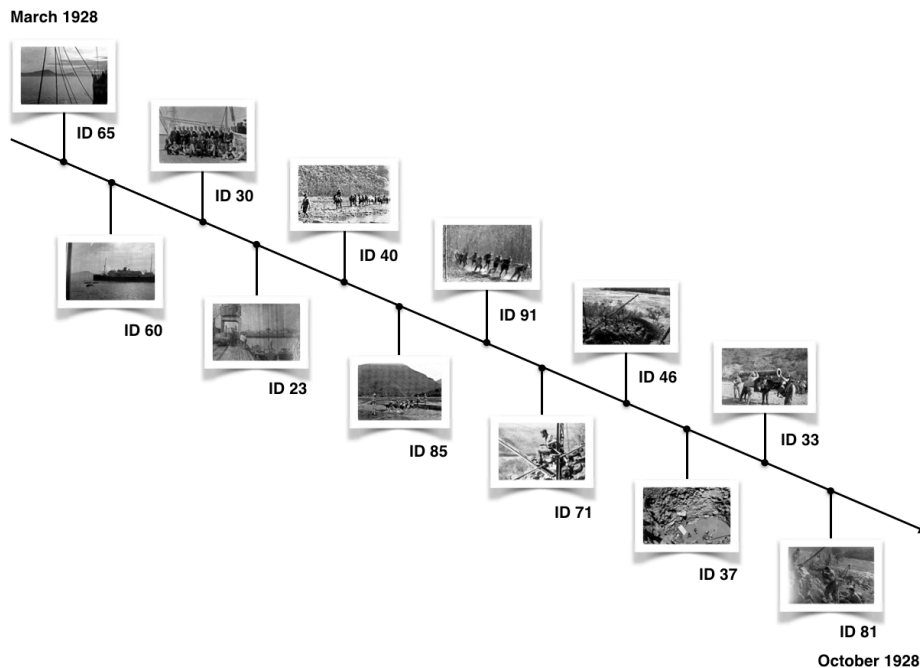

Figure 4: The timeline generated for the subset of 12 photographs orders and describes the European expedition in general terms, from its inception in Liverpool in March of that year to the depths of the Bolivian mountains between June and October 1928.

60

Table 3

Timeline extraction and narrative summarization prompt and final "Narrative Summary" output generated over the 12 photos subset of the historical archive. The timeline portion of the output is shown in Figure 4. The prompt was followed by the list of all the previously generated descriptions of the photos.

Prompt	You are an expert documentary filmmaker and you are exploring the historical archive of photos of the Sacambaya Expedition in Bolivia from 1928, taken by the German photographer Robert Gerstmann. The expedition sought to find the lost treasure of the jesuits, hidden somewhere near Sacambaya. You have the following list of descriptions of photos. Your goal is to generate two outputs: a timeline of these photos from the provided descriptions so that it forms a coherent narrative and a written summary of the narrative. The timeline should explicitly reference the IDs of the photos. The narrative summary should also reference the IDs of the photos as needed between brackets. Note that the provided descriptions might not be shown in the correct order and the IDs are not associated with any ordering. Please avoid the use of flowery language, your narrative should be written in an academic style. Do not explain your process.
Output	The Sacambaya Expedition of 1928, documented by German photographer Robert Gerstmann, embarked on a challenging quest to uncover the Jesuit's lost treasure in Bolivia's Sacambaya Valley. It began with maritime scenes (IDs 65 and 60), capturing the tranquil ocean and coastal topography as the British expedition approached its destination. Aboard the ship, the crew is depicted in a moment of camaraderie and preparation (ID 30), followed by a brief sojourn at a riverine port, indicating the integration of transportation modes (ID 23). Upon reaching the Sacambaya Valley, the expedition faced the valley's rugged terrain (ID 40), fording rivers and navigating hilly landscapes (ID 85), and traversing dense thickets (ID 91). The photographs depict the challenging nature of their journey, highlighting both the physical and logistical hurdles encountered. As the expedition progressed, the team's burden of transporting heavy machinery through the dry valley is evident (ID 71), reflecting the blend of hope and hardship. Moments of rest and mechanical endeavors are captured (ID 46), as well as industrious activities amidst natural ruins (ID 37), underscoring the expedition's reliance on technology of the time. The concluding images (IDs 33 and 81) focus on the excavation and manual labor involved in the treasure hunt, set against a backdrop of natural and historical landscapes. This narrative, woven through Gerstmann's photographs, tells a story of determination, challenge, and the enduring human spirit in the pursuit of historical treasures.

5. Conclusions

This ongoing research represents an advancement in the field of narrative extraction from heritage image archives. Through the iterative and incremental development of our methodology, a framework is established that not only enhances the understanding of historical events but also contributes to the broader discourse on computational narratives and artificial intelligence [25].

This experimental model facilitates a utilitarian approach to solving issues prevalent in numerous audiovisual archives. The methodology proposed herein aims to achieve quality

control measures concurrently with narration. While beyond this work's scope, future initiatives will prioritize expanding the search range and implementing an integral quality control system to regulate labeling accuracy, clustering thoroughness, and narrative coherence levels. Additionally, we acknowledge the potential of Large Language Models (LLMs) to autonomously identify a broad spectrum of objects. Consequently, subsequent projects may explore creating narrations without human support, relying solely on object identification and computer vision capabilities. However, if pursuing this avenue, methodological and future discussion elements must be incorporated to ensure a minimum level of confidence in the results, rendering them relevant for anthropological, historical, and heritage discourse. Limitations of this case study and the need for a more comprehensive evaluation are identified, as its primary objective was to illustrate the methodology rather than validate it with an extensive data spectrum.

The successes of the proposed method in constructing coherent historical narratives suggest a potential paradigm shift in how narrative extraction from visual historical records can be approached. Thus, this ongoing research represents a significant contribution to the challenge of uncovering narratives concealed within historical image archives. Furthermore, our aim is to observe significant changes, trends, and prevalent elements in large groups of visual information that may not be readily apparent through individual observations [26]. This broader perspective facilitates the finding and construction of narratives that extend beyond individual images.

These experiments using easily available LLMs demonstrate the need to always maintain human control in the process, as shown by all the required prompt engineering. Future work will consist of applying the proposed pipeline to the entire collection of 500 images. We hope that our proposed methodology and technical pipeline streamline the work of expert catalogers, documentarians, and media creators, who can now have a minimal foundational basis to explore large, undisclosed photographic collections. Additionally, in the present case of the archive of the Sacambaya Expedition by Robert Gerstmann, we hope to have contributed to the historical and heritage enrichment of a part of this little explored collection.

In conclusion, we propose that the analysis of a specific photographic collection can be further enriched through the organization and utilization of information in narratives [13]. Theories on sensemaking emphasize that sensemaking and narrative are two inherently interconnected concepts about how people understand the world around them [27]. Given its replicability, we consider our proposed method to be a contribution to the discovery, enrichment, and dissemination of the worlds and narratives "hidden" inside photographic heritage archives.

Acknowledgements

References

[1]   Fornaro, Peter & Chiquet, Vera. (2020). Artificial Intelligence for Content and Context Metadata Retrieval in Photographs and Image Groups. Archiving Conference. 2020. 79-82. 10.2352/issn.2168-3204.2020.1.0.79.

[2]   M. Alvarado, Roberto Gerstmann: fotografías, paisajes y territorios latinoamericanos, 1st. ed.Pehuén, Santiago, Chile 2009.

[3]   M. Matus, Roberto Gerstmann's last photography, Video, 2022. URL: https://youtu.be/9nFvhoZd5Os.

[4]   E. Sanders, E. The Story of the Jesuit Gold Mines in Bolivia and of the Treasure Hidden by the Sacambaya River. (1928) Rauner Special Collections Library - Dartmouth College.

[5]   J. Pavez, Imágenes de la revolución industrial: Robert Gerstmann en las Minas de Bolivia (1925-1936), 1st ed. Plural, La Paz, Bolivia 2017.

[6]   Ma, Wenchi, Xuemin Tu, Bo Luo, and Guanghui Wang. "Semantic clustering based deduction learning for image recognition and classification." *Pattern Recognition* 124 (2022): 108440.

[7]   Makridakis, Spyros, Fotios Petropoulos, and Yanfei Kang. "Large language models: Their success and impact." *Forecasting* 5, no. 3 (2023): 536-549.

[8]   Jiho Shin, Clark Tang, Tahmineh Mohati, Maleknaz Nayeb, Song Wang, and Hadi Hemmati. 2024. Prompt Engineering or Fine Tuning: An Empirical Assessment of Large Language Models in Automated Software Engineering Tasks. 1, 1 (October 2024), 22 pages.

[9]   Wevers, M., Vriend, N., & de Bruin, A. (2022). What to do with 2.000.000 historical press photos? The challenges and opportunities of applying a scene detection algorithm to a digitised press photo collection. TMG Journal for Media History, 25(1).

[10]   Witte, R., Kappler, T., Krestel, R., & Lockemann, P. C. (2011). Integrating wiki systems, natural language processing, and semantic technologies for cultural heritage data management. In Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series (pp. 213-230). Springer Berlin Heidelberg.

[11]   Wang, X., Ye, L., Keogh, E., & Shelton, C. (2008, June). Annotating historical archives of images. In Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries (pp. 341-350).

[12]   Jo, E. S. (2020). Foreign Relations of the United States Series, 1860-1980: A Study in New Archival History. Stanford University.

[13]   Lotfi, F.; Beheshti, A.; Farhood, H.; Pooshideh, M.; Jamzad, M.; Beigy, H. Storytelling with Image Data: A Systematic Review and Comparative Analysis of Methods and Tools. Algorithms 2023, 16, 135. https://doi.org/10.3390/a16030135.

[14]   Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language Models Are Few-shot learners. Advances In Neural Information Processing Systems,33: 1877−1901,2020.

[15]   Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., et al. PaLM: Scaling language modeling with pathways. arXivpreprint2204.02311,2022.

[16]   Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozi`ere, B., Goyal, N., Hambro, E., Azhar, F., et al. LLaMA: Open and efficient foundation language models. arXivpreprint2302.13971,2023.

[17]   OpenAI.GPT-4 technical report. arXivpreprint2303.08774, 2023. .URL: https://doi.org/10.48550/arX iv.2303.08774.

[18]   Kim, T., Heo, M. O., Son, S., Park, K. W., & Zhang, B. T. (2018). Glac net: Glocal attention cascading networks for multi-image cued story generation. arXiv preprint arXiv:1805.10973.

[19]   Huang, T. H., Ferraro, F., Mostafazadeh, N., Misra, I., Agrawal, A., Devlin, J., & Mitchell, M. (2016, June). Visual storytelling. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 1233-1239).

[20]   Maggiori, Emmanuel, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. "High-resolution aerial image labeling with convolutional neural networks." *IEEE Transactions on Geoscience and Remote Sensing* 55, no. 12 (2017): 7092-7103.

[21]   D. Buck, 2000 Tales of Glitter or Dust, accessed December 2023. URL: https://www.thefreelibrary.com/Tales of Glitter or Dust.-a073064246.

[22]   S. Jolly, "The Treasure Trail". John Long Limited, London 1934.

[23]   Labelbox, "Labelbox," 2024. URL: https://labelbox.com.

[24]   Wallace, Gregory K. "The JPEG Still Picture Compression Standard", Communications of the ACM, April 1991 (vol. 34, no. 4), pp. 30-44.

[25]   Keith Norambuena, Brian Felipe, Tanushree Mitra, and Chris North. "A survey on event-based news narrative extraction." *ACM Computing Surveys* 55, no. 14s (2023): 1-39.

[26]   Klingenstein, Sara & Hitchcock, Tim & DeDeo, Simon. (2014). The civilizing process in London's Old Bailey. Proceedings of the National Academy of Sciences of the United States of America. 111. 10.1073/pnas.1405984111.

[27]   Battad, Zev & Si, Mei. (2022). A System for Image Understanding using Sensemaking and Narrative. The Ninth Advances in Cognitive Systems (ACS) Conference 2021.