Visual Exploration and Analysis of large scale multimedia Archives

Bocoum Ousmane[a], Yadong Wu[a,*]

[a] *Visualization and Virtual reality Lab, Southwest University of Science and Technology*

Abstract

This paper present a visual analytic framework for the exploration and analysis of large scale multimedia archives. By revealing different perspectives of a multimedia corpus, the framework gives high level overviews on each type of data and provides powerful mechanisms for detailed analysis and shallow exploration adapted to large scale audiovisual content. Using deep learning techniques applied to images, audio and text a pipeline is designed for the automatic indexation and classification of large archive of multimedia data. The applicability of our approach is demonstrated by practical case studies conducted on a real world archive dataset.

Keywords: Visual analytics, Large Multimedia, Image Visualization, Text Visualization

1. Introduction

Large multimedia archives[1] are maintained around the world by different services to preserve our audiovisual heritage. Although archivists have always built indexes and tools[2] to provide access to the archives, the quantities of the documents are such that a large part still remains difficult to access and explore[3]. Archivists therefore are facing a major challenge : how to make accessible to the public these millions of documents in a meaningful and user-friendly way.

To tackle this problem media owners such as The French Audio Visual Institute(INA)[2] and the U.S. National Archives and Records Administration[4] have designed web based platforms to give users access to the multimedia archives they are maintaining. While these platforms allow the user to browse the content, they are limited when it comes to exploration and interaction with the content therefore giving an overall bad user experience.

Considering these challenges we designed a web based visualization framework for large scale multimedia archives exploration and analysis. The system was developed in an iterative design process with continuous refinement guided by archivists. In collaboration with the archivists three major requirements were identified for the effective support of the analysis and exploration of a large multimedia corpora, these are:

*Corresponding author.
Email address: wyd028@163.com (Yadong Wu)

March 18, 2019

(1) Being able to visualize the corpus in a single interactive view
(2) Interact with the items displayed, query the corpus and get relevant answers
(3)Provide a pipeline to automatically index and classify the data in the corpus.

To meet these requirements we combined visualization techniques and deep learning techniques. Deep learning methods were used to classify automatically the data, witch include text, audio, image and video. The classified corpus was then visualized using multiple visual components, allowing the user to explore, query and interact with the corpus.

The main contributions of this paper are:

- A novel visual framework for the exploratory analysis of large scale multimedia archives is presented.
- We introduce a method to model a large archive corpus in a directed acyclic graph
- The relevance of the combination of deep learning based methods and visualization techniques is demonstrated.

This paper is organized as follows: Section 2 introduces the state of the art of information visualization and other related works, in section 3 we describe our framework and the data used. Section 4 presents the automatic data management pipeline.In the section 5 the applications scenarios and visualization components are described, we finally conclude in section 5 and provide perspectives for future improvements.

2. Related Works

Information visualization is a very active research area[5], numerous visualization systems has been designed to deal with large collections of data. Our framework is informed by related research from the fields of visual content analysis and deep learning techniques applied to text, audio, and image classification.

2.1. Visualization of large multimedia data collections

Large image collection data visualization has been increasingly used in fields such as medicine[6] , security, and personal album[7] management. Numerous works have successfully used visualization techniques for the exploration of large image collections,Photoland[8] is a system that visualizes hundreds of photos on a 2D grid space to help users manage their photos,Tan et al. presented imageHive[9] an Interactive content-aware image summarization system, tipiX[10] is a system that allows a rapid visualization of large medical image collections. Xie et al.[11] proposed a semantic based method for visualizing large image collections using Convolutionnal neural networks.

Different visualization techniques such as scatter plots[12] , tree[13] based methods[14] and directed graphs[15][16] has been used to facilitate large image collection analysis.

The main goal of text visualization is to allow the discovery of useful knowledge from large document collections effectively without completely going through the details of each document in the collection. To reach this goal different techniques[17] has been proposed

such as document similarity visualization[18] , text flowbased methods , Word-based methods, radial based methods, tree-based, and Semantic Oriented Techniques[17].

Using these methods researchers have designed numerous systems, the Text Visualization browser[19] present a survey of text visualization techniques, the Stanford Dissertation Browser[20] is a visualization system for document collections that enables richer interaction, it is an abstraction of Stanford's PhD dissertation from 1993-2008,the documents are presented through the lens of a text model that distills high-level similarity and word usage patterns in the data, they then present those patterns using visualization methods. Our system uses the same approach by extracting underlying patterns from raw documents and presenting them in a more clearly way using visual clues to the archivist.

The approaches presented above all focuses on a specific type of data collection or analysis task, although they allow the understanding of these single data types, such methods does not make use of the main structure of the collection as a whole. In contrast, our approach is more general and allows the user to visualize different data types in the same visualization component, and furthermore provides individual visual components for each type of document.

2.2. Deep Learning methods for Unstructured data classification

The recent advancements in deep learning has made deep convolutionnal networks based methods the go to for unstructured data classification[21] .Convolutionnal neural networks have shown tremendous results in classifying images[22] with human level accuracy. They have revolutionized computer vision, achieving state-of-the-art results in many fundamental tasks, as well as making strong progress in natural language processing, computer audition, reinforcement learning, and many other areas[23].

2.2.1. Image classification using CNNs

Since the 2012 milestone, researchers showcased various cnn based architectures at the ImageNet Challenge[24].

The most notable models presented at imageNet are presented in the following table:

Table 1: Principal image classification models presented at Imagenet

Model	Year	Error rate
AlexNet[25]	imageNet 2012	top-5 error rate of 15.3%
VGG16[26]	imageNet 2014	top-5 error rate of 7.3%
GoogLeNet[27]	imageNet 2014	top-5 error rate of 6.7%
Inception V2[28]	imageNet 2015	top-5 error rate of 3.58%
Inception V4[29]	imageNet 2016	top-5 error rate of 3.08%
SE-Resnet[30]	imageNet 2017	top-5 error rate of 2.25%

These models are open source and can be used freely to perform custom image classification tasks. We used the inception V4 model as our image classifier.

2.2.2. Audio documents classification using CNNs

With the success of deep neural networks, a number of studies applied them to speech and other forms of audio data[31][32]. Representing audio in time is a challenging task, however Van Den Oord et al[33]. Addressed this challenge by introducing wavenet a deep neural network that generates waveforms from raw audio data to train custom classifiers. An alternative to their method is the spectogram of a signal which can represent both time and frequency[34][35] .Spectograms are images and can be used to extract features from audio data and train a convolutionnal neural net.We used Spectograms to represent our audio data and trained a convolutionnal neural net to classify the audio files.

3. Data and Framework description

3.1. Dataset

The dataset used in this work is the compilation of multimedia archives data from the Malian National Archives service. The dataset comports 400 000 images,130000 hours of video records,120000 audios hours of radio digitized records, 36000 text documents.The dataset was collected from 1960 to 2018.

3.2. Tasks analysis

To fully understand the requirements and the tasks of our system we conducted multiple brainstorming sessions with professional archivists, In collaboration with them we identified four major tasks for the effective support of the analysis and exploration process of a large multimedia corpora, these tasks are:

T1. **Summarize a large corpus of multimedia data**

A corpus of archives can contain a large number of items making it difficult to explore and analyze. Thus being able to visualize the corpus in a single interactive view is critical to conduct a smooth and meaningful exploration of its content.

T2. **Interact with the corpus**

Interaction is an important task in a visualization system. The user should be able to interact with the content and operate analysis operations such as filtering or searching.

T3. **Query the corpus**

The system should support various query methods. The user should be able to query the corpus and get relevant answers.

T4. **Automatic Indexation and classification**

Indexing and classifying a large database of multimedia is a tedious task. Therefore, the necessity of an automatic indexation and classification pipeline.

3.3. System Design

To cover the tasks described above we adopted the following design principles.

- Modeling a large archive corpus in a directed acyclic graph:

A large archive corpus can be represented in an acyclic graph to abstract the hierarchy and the relations between the documents. The graph representation offers a multi-level overview(T1) of the corpus allowing the users to identify groups of interrelated documents. Such representation make it easy to implement various interaction and filtering methods(T2)

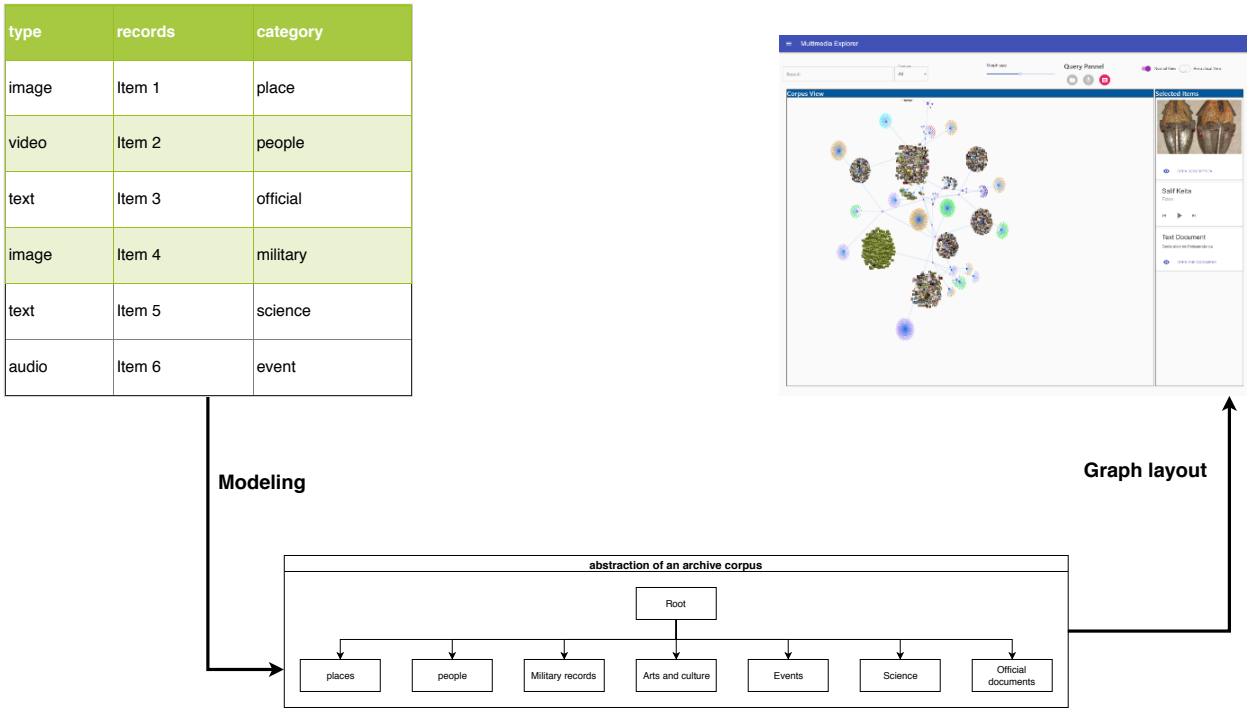

Figure 1: Process of modeling a large database of archive to an acyclic graph

- Multiple Querying methods

We implemented multiple querying methods such as filtering and searching, we furthermore implemented a document retrieval strategy allowing the user to query the corpus from an image or audio file(T3).

- Automatic indexation and classification pipeline

Using deep learning methods we trained various classifiers for each data type(T4), section 4 describe the details of each model.

3.4. System Architecture

Our system is a client-server application.We used Reactjs as front end framework,Flask as back end and MongoDB database. We trained our models using Tensorflow. The models are deployed on the server where the input data is prepossessed and classified. The classified data is then sent to the frontend for visualization.

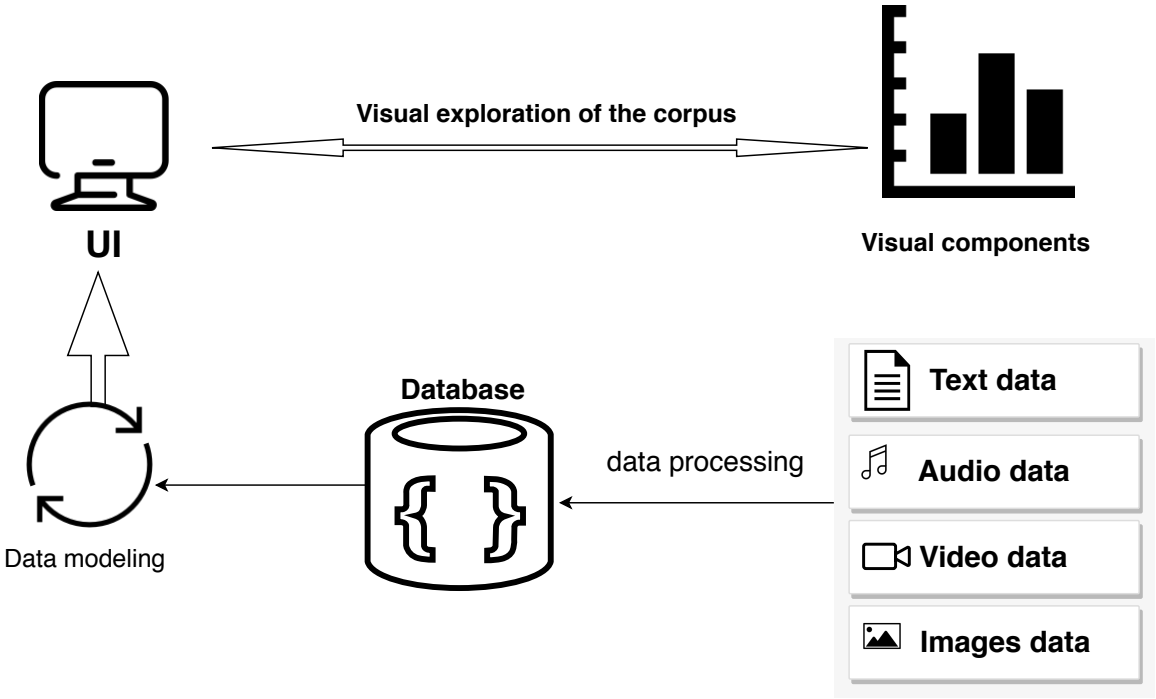

Figure 2: System architecture

4. Automatic data management

In this section we describe the models used for the automatic management of each type of data.

4.1. Automatic Image classification

Our image classification pipeline is a combination of the Inception V2 [28] trained on imagenet dataset[24] used as an entry classifier and a custom face recognition model. The Inception V2 takes the raw input images and classify the images to the 1000 classes of Imagenet.

We then feed the pre-classified images containing people to a custom face classifier to index images pertaining to notable people present in the dataset.

Our face classifier was trained using transfer learning on Facenet[36] , the classes are famous people and important people present in the corpus.

4.2. Automatic audio classification model

An audio classification model was trained to recognize the genre of each file. The model is a convolutionnal neural network based on the VGG16 [26] architecture and trained with our own datasets.From each sample the features for each category was extracted using melspectograms.A spectrogram is a visual representation of the spectrum of frequencies of sound as they vary with time.The spectograms were then used to train an audio classification

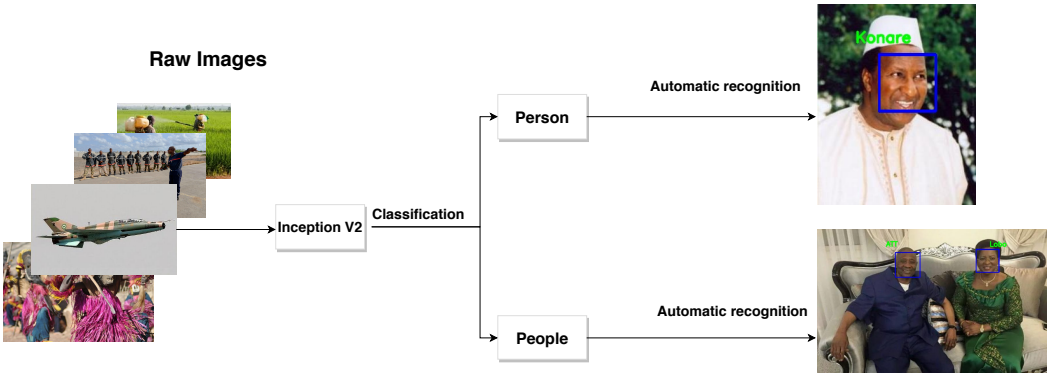

Figure 3: The image classification pipeline takes raw input images,and the Inception v2[28] model trained on imagenet[24] dataset gives a class to each image,the images containing people or a person are then sent to a custom face recognizer,the the face recognizer is trained to recognize famous and key people present in the corpus.

model to map each data to one of the following classes:Speech,chants traditionnels,chants mandingues,chants millitaires,discours,folklore,chants sonhrai,instruments,autres.

The goal with these models is to help the archivists to automatically classify and index raw multimedia data, however images and audio can be misclassified and have to be manually corrected by the archivists.

4.3. Text Similarity model

For each text document standard data processing such as cleaning,n-gram extraction was conducted. We then used the Multi-Perspective Sentence Similarity model proposed by He et al[37] to measure the similarity between documents. The method uses convolutionnal neural nets to extract features from sentences and compare sentence similarity.

The open source Stanford named entity recognizer[38] was used to extract the named entities present in each document. Named entities are very important in understanding the content of a text document, they help to abstract the context of a large document. The entities are displayed in a named entity graph[39].

5. Application scenarios

In this section we describe the visual components and the use scenarios of the framework. The system is composed of multiple views each of which displays an aspect of the corpus.

5.1. Description of the interfaces

The user interface provides the user with multiple visual components in two main mode:

- Exploration mode In the exploration mode,the complete corpus is displayed with entities highlighted in their respective colors. We choose to map each data type to a discriminative visual variable such as color or shape,images and videos are represented by thumnails.The user can search,explore or navigate through a given set of data.

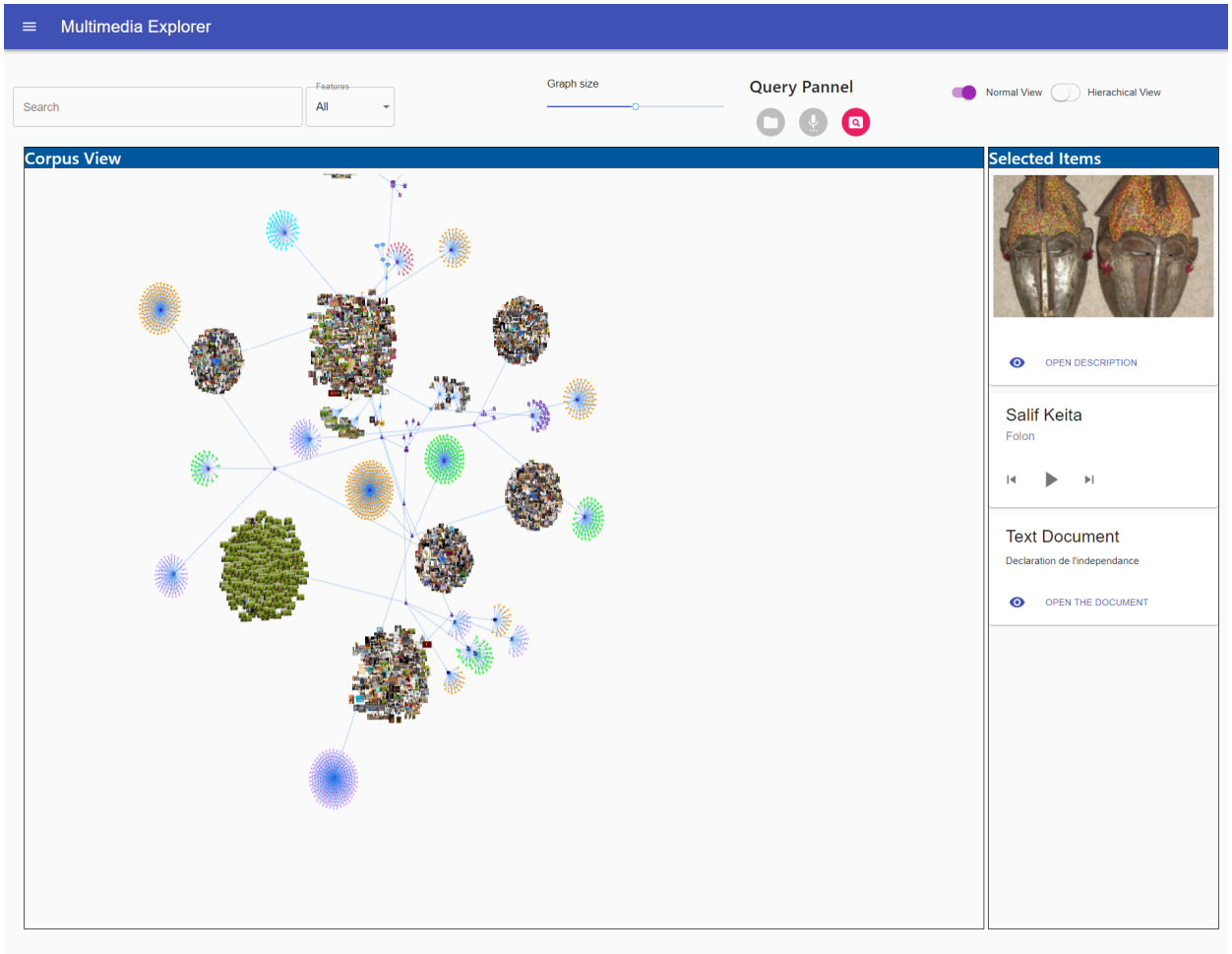

Figure 4: In the exploration mode the corpus is displayed in an interactive graph layout,the user can interact with the graph and select items for further analysis,the selected items are displayed in the selected item area.

- Analysis mode

In the analysis mode a specific data type is displayed for in depth analysis.The user can use searching and filtering operation to analyze the dataset.A query panel allow the user to perform queries on the corpus and get relevant results.

5.2. Analysis of video records

The main goal of this component is to provide an interactive visual summary of the videos in the corpus.The component comport a graph layout displaying thumbnails of the video records,the graph displays clusters of videos by category.The user can select a video and play it directly in the interactive view.Key frames of each selected video are displayed for allowing the user to grasp the overall content of the video without having to play the file in it entirety.

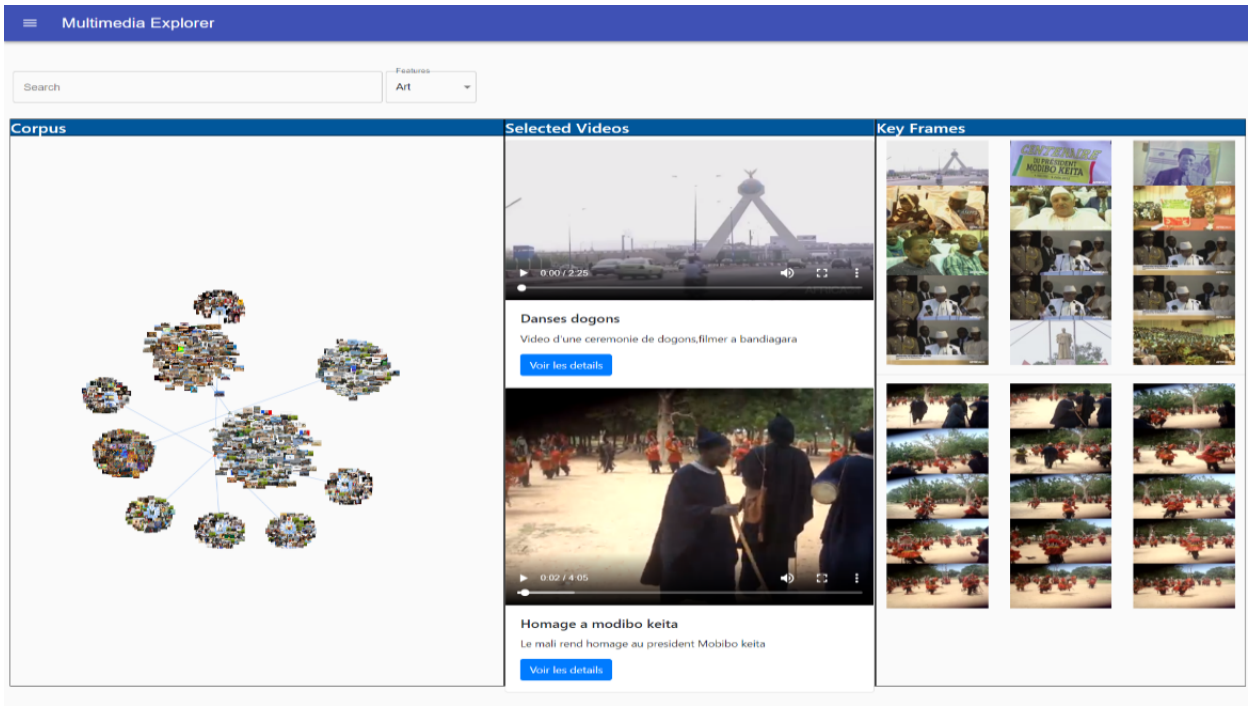

Figure 5: Video analysis component,the records are displayed in a graph and the user can select a video and display it,key frames of each selected video are displayed.

5.3. Analysis of Audio records

Audio records are a large part of the archives and most of the records are not correctly annotated making it dificult to explore.The main goal of this feature is to display a visual summary of the audio records.The records are classified and displayed in clusters according to their category.The user can select a given item and play it.The component support searching and filtering making it easy for the user to explore and analyze the records.

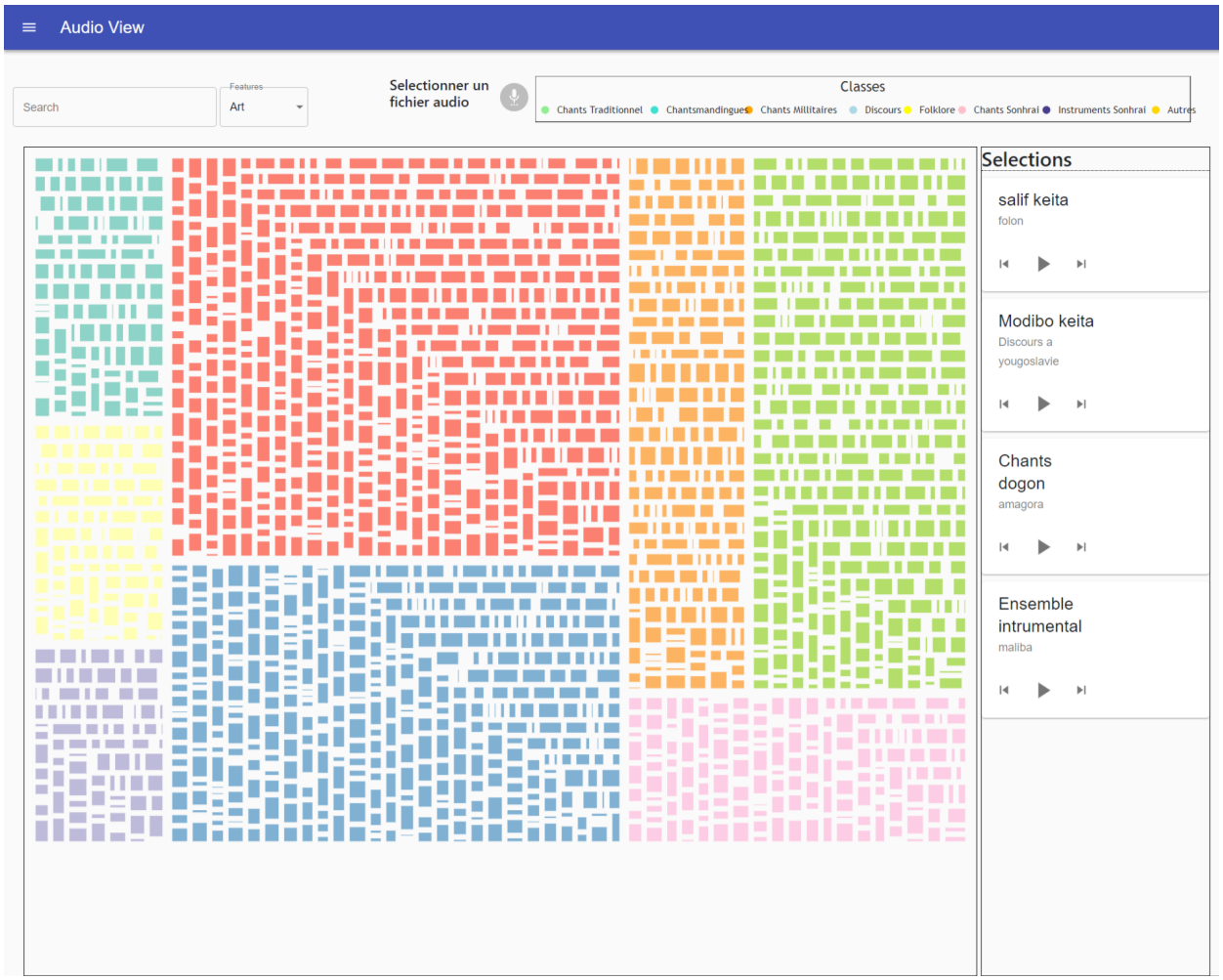

Figure 6: Audio analysis component,the records are classied and displayed in clusters according to their category.

5.4. Text documents exploration and analysis

The text component visualizes the text documents of the corpus. It is composed of:

- A bubble graphFigure (A) displaying the documents classified per category, the distance between each item is proportional to the similarity of the documents. The user clicks on a document and trigger the visualization of the content on the other components of the view.
- A named entity graphFigure (B) shows the named entities contained in the document
- The text viewFigure (C) displays the document with the named entities highlighted
- A Word cloud viewFigure (D) displays the important words in the document per frequency

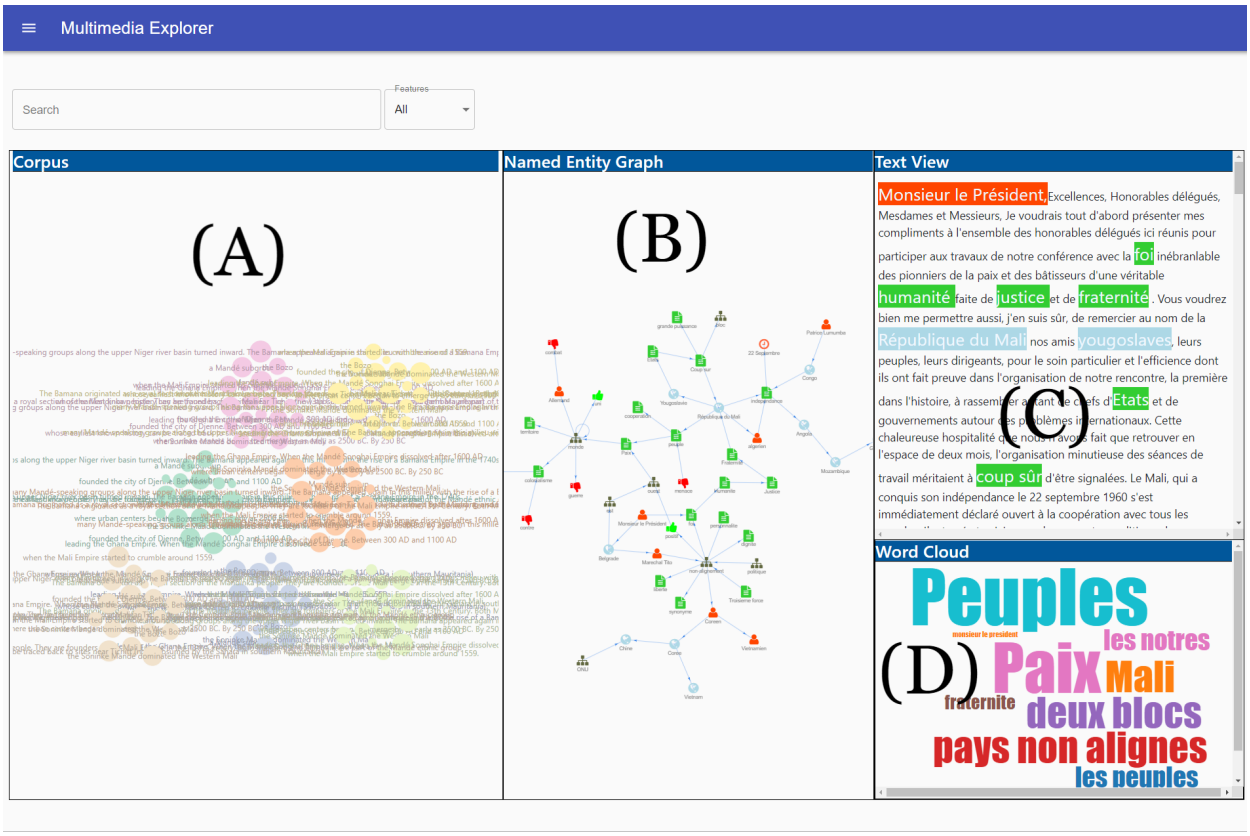

Figure 7: Text analysis component.

6. Evaluation

To assess in real context the impact of our tool,a comparative study has been made between the traditional web interfaces used by the archivists to display the archives and the visualization approach described above.The Iso 9241-11[37] usability criteria was used as the evaluation criteria. The analysis of the results shows an overall preference for the Interactive visualization approach.The archivists said the system added value to analysis and exploration of archives data,especially for previously unknown data.They also gave us insights to improve the system in future works.

7. Conclusion

We presented a visual analytic framework for large multimedia archives data visualization. Our approach represents a large corpus of archives in an acyclic directed graph. The graph model was then visualized using multiple interactive views, each view displayed an aspect of the corpus. We also used deep learning techniques to facilitate the tedious task of indexing and classifying a large corpus of multimedia data.

In future works we would like to improve the automatic indexation pipeline and make it more accurate.

March 18, 2019

References

[1] Cees GM Snoek, Marcel Worring,Jan C Van Gemert,Jan-Mark Geusebroek and Arnold WM Smeulders. *The challenge problem for automated detection of 101 semantic concepts in multimedia.* Proceedings of the 14th ACM international conference on Multimedia,421-430,2006.

[2] Marie-Luce Viaud. *Interactive components for visual exploration of multimedia archives.* 2008.

[3] Martha Rose Henley.
Method and system for managing movement of large multi-media data files from an archival storage to an active storage within a multi-media server computer system 1998

[4] Jay Atherton. *From life cycle to continuum: some thoughts on the records management–archives relationship..* Archivaria,volume 21,42-43.

[5] Chris North. *Information visualization.* Handbook of human factors and ergonomics,page 1209-1236,2012.

[6] William Plant,Gerald Schaefer,. *Visualisation and browsing of image databases.* Multimedia Analysis, Processing and Communications,page 3-57,2011.

[7] Erik Cambria,Amir Hussain. *Sentic album: content-, concept-, and context-based online personal photo management system.* Cognitive Computations,volume 4,page 477-496,2012.

[8] Dong-Sung Ryu,Woo-Keun Chung and Hwan-Gue Cho. *Photoland: a new image layout system using spatio-temporal information in digital photos.* Proceedings of the 2010 ACM Symposium on Applied Computing,page 1884-1891,2010.

[9] Li Tan,Yangqiu Song,Shixia Liu and Lexing Xie. *Imagehive: Interactive content-aware image summarization.* IEEE computer graphics and applications,volume 32,page 46-55,2012.

[10] Adrian V Dalca,Ramesh Sridharan and Natalia Rost. *tipiX: Rapid Visualization of Large Image Collections.*

[11] Xiao Xie,Xiwen Cai,Junpei Zhou,Nan Cao and Yingcai Wu. *A Semantic-based Method for Visualizing Large Image Collections.* IEEE Transactions on Visualization and Computer Graphics,2018

[12] Giang P Nguyen and Marcel Worring. *Interactive access to large image collections using similarity-based visualization.* Journal of Visual Languages Computing,volume 19,Issue 2,pages 203-224,2008

[13] , et al. *"." 33.09.* Journal of Visual Languages Computing,pages 2631-2635,2013

[14] Benjamin B Bederson. *PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps.* Proceedings of the 14th annual ACM symposium on User interface software and technology,pages 71-80,2001

[15] Yi Gu,Chaoli Wang,Jun Ma,Robert J Nemiroff and David L Kao. *iGraph: a graph-based technique for visual analytics of image and text collections.* Visualization and Data Analysis 2015,2015

[16] Yadong Wu. *A total variation-based hierarchical radial video visualization method.* Journal of Visualization,Volume 18,Issue 2,pages 255-267 ,2015

[17] Kostiantyn Kucher and Andreas Kerren. *Text visualization techniques: Taxonomy, visual survey, and community insights.* Visualization Symposium (PacificVis), 2015 IEEE Pacific,page 117-121

[18] Dylan Baker. *The Document Similarity Network: A Novel Technique for Visualizing Relationships in Text Corpora.* 2017

[19] Kostiantyn Kucher and Andreas Kerren. *Text visualization browser: A visual survey of text visualization techniques.* Poster Abstracts of IEEE VIS,2014

[20] Jason Chuang,Daniel Ramage,Christopher Manning and Jeffrey Heer. *Interpretation and trust: Designing model-driven visualizations for text analysis.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,pages 443-452,2012

[21] Md Zahangir Alom,Tarek M Taha,Christopher Yakopcic,Mahmudul Hasan,Brian C Van Esesn,Abdul A S Awwal and Vijayan K Asari. *The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches.* arXiv preprint arXiv:1803.01164s,2018

[22] Chen Wang and Yang Xi. *Convolutional Neural Network for Image Classification.* Johns Hopkins University Baltimore, MD

[23] Athanasios Voulodimos,Nikolaos Doulamis,George Bebis and Tania Stathaki. *Recent Developments in Deep Learning for Engineering Applications.* Computational intelligence and neuroscience, 2018

March 18, 2019

[24] *Imagenet large scale visual recognition challenge.*
[25] Alex Krizhevsky,Ilya Sutskever and Geoffrey E. Hinton. *"Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems.* Computational intelligence and neuroscience, 2018
[26] Karen Simonyan and Andrew Zisserman. *Very deep convolutional networks for large-scale image recognition..* arXiv preprint,volume 1409-1556, 2014
[27] Christian Szegedy. *Going deeper with convolutions..* 2015
[28] Christian Szegedy. *Rethinking the inception architecture for computer vision..* 2016
[29] Christian Szegedy. *Inception-v4, inception-resnet and the impact of residual connections on learning.* AAAI,Volume 4, 2017
[30] Jie Hu,Li Shen and Gang Sun. *Squeeze-and-Excitation Networks. arXiv.* page 1-11,2017
[31] Shawn Hershey. *"CNN architectures for large-scale audio classification." Acoustics, Speech* Signal Processing (ICASSP), 2017 IEEE International Conference on. IEEE,2017
[32] Jongpil Lee *Raw Waveform-based Audio Classification Using Sample-level CNN Architectures.* arXiv preprint,Volume 1712,2017
[33] Van Den Oord *Wavenet: A generative model for raw audio.* CoRR,volume 1609,2016
[34] Tan Jo Lynn, Ahmad Zuri Sha'ameri *Automatic analysis and classification of digital modulation signals using spectogram time frequency analysis.* Communications and Information Technologies,2007
[35] Honglak Lee *"Unsupervised feature learning for audio classification using convolutional deep belief networks." Advances in neural information processing systems* 2009
[36] Florian Schroff,Dmitry Kalenichenko and James Philbin *Facenet: A unified embedding for face recognition and clustering.* 2015
[37] Hua He,Kevin Gimpel and Jimmy Lin *Multi-perspective sentence similarity modeling with convolutional neural networks..* 2015
[38] Jenny Rose Finkel, Christopher D. Manning *Nested named entity recognition.* Association for Computational Linguistics,2009
[39] Mennatallah El-Assady *NEREx: Named-Entity Relationship Exploration in Multi-Party Conversations.* Computer Graphics Forum,volume 36,Issue 3,2017
[40] *https://www.sis.se/api/document/preview/611299/*