

DRIFT: A visual analytic tool for scientific literature exploration based on textual and image content

Ximena Pocco[†], Jorge Poco[‡], Matheus Viana^{§*}, Rogerio de Paula[§], Luis G. Nonato^{||} and Erick Gomez-Nieto[†]

[†]Department of Computer Science, Universidad Catolica San Pablo, Arequipa, Peru

[‡]School of Applied Mathematics. Getulio Vargas Foundation, Rio de Janeiro, Brazil

[§]IBM Research, Sao Paulo, Brazil

^{||}ICMC, University of Sao Paulo, Sao Carlos, Brazil

Abstract—Exploring digital libraries of scientific articles is an essential task for any research community. The typical approach is to query the articles’ data based on keywords and manually inspect the resulting list of documents to identify which papers are of interest. Besides being time-consuming, such a manual inspection is quite limited, as it can hardly provide an overview of articles with similar topics or subjects. Moreover, accomplishing queries based on content other than keywords is rarely doable, impairing finding documents with similar images. In this paper, we propose a visual analytic methodology for exploring and analyzing scientific document collections that consider the content of scientific documents, including images. The proposed approach relies on a combination of Content-Based Image Retrieval (CBIR) and multidimensional projection to map the documents to a visual space based on their similarity, thus enabling an interactive exploration. Additionally, we enable visual resources to display complementary information on selected documents that uncover hidden patterns and semantic relations. We show the effectiveness of our methodology through two case studies and a user evaluation, which attest to the usefulness of the proposed framework in exploring scientific document collections.

I. INTRODUCTION

A major component in scientific research is the compilation of pertinent literature, a task typically performed by querying various academic sources, e.g., articles, surveys, reviews, books, and thesis/dissertation stored in digital libraries. For instance, well-known repositories such as IEEE Xplore, ACM DL, and ArXiv enable the traditional searching paradigm where users perform queries based on keywords, resulting in a list of textual snippets containing the title, authors, and other information summarizing the content of each document. Users must manually inspect the snippets to find documents of interest; digital libraries do not provide resources to gather documents based on their content, making the literature compilation a tedious and time-consuming task. Moreover, resources to perform queries from images, tables, and charts are not available, impairing the search for content other than text. The image-based query has been widely used in Content-based Image Retrieval (CBIR) systems and could also be employed to support the exploration of scientific literature libraries. Performing queries based on images and other non-textual content can make it possible to answer questions such as: “Which are the typical images in papers from this author?”,

“Which articles have images similar to this one?” or even “Is this image similar to any other published?”.

Another critical issue in exploring scientific literature is how to enable visual resources that render the analysis of multiple document collections an easier task. Some academic search engines such as Microsoft Academic Visual Explorer and Google Scholar enable visual representations for co-authorship analysis and citation evolution over time. Besides being quite limited, the visual resources enabled in those tools are not linked to queries mechanism, what considerably restricts the scope of any exploratory analysis. There are also alternatives to replace the regular list of textual snippets with some visualization-oriented representations, mainly in the context of web search result analysis [1], [2]. However, despite the effectiveness demonstrated by these methods, they have not been introduced into digital libraries for exploring articles yet.

In this work, we propose an interactive visualization tool for exploring large collections of scientific documents. Called *DRIFT* (Document exploration based on Image and textual Features), the proposed methodology combines core CBIR functionalities with an interactive multidimensional projection mechanism that identifies documents with similar content, including images. In contrast to existing systems, our approach enables several exploratory and visualization resources that make complex analysis doable, increasing the user’s ability to perform complex searches and analyses.

In summary, the main contributions of this work are:

- A methodology that combines content-based image retrieval mechanisms, multidimensional projection, and visual analytic tools into a single framework that handles documents based on their textual and image content.
- A visual analytic tool called DRIFT, which implements the proposed framework to enable customized exploration of large collections of scientific documents.
- Two case studies and a user evaluation that demonstrate the utility and effectiveness of our methodology.

II. RELATED WORK

We organize our review on methods that explore scientific publication collections in three main categories:

Citation based methods focus on uncover citation and research collaboration patterns [3]–[5]. Liu *et al.* [6], for instance, search for citations in a specific paper and build

*Matheus Viana is no longer affiliated to IBM Research.

a tag cloud to intuitively conveying which part of the paper each citation refers to. Yan and Ding [7] analyze six different types of scholar networks (coupling, co-citation, topic, co-authorship, co-word) aiming a better understanding how they are related. Cite2vec [8] uses word embeddings for document exploration based on the context in which they are cited.

Textual content-based methods employ text processing strategies to establish similarities among documents. For instance, Action Science Explorer (ASE) [9] is a system that enables interactive analysis of a paper collection through linked-views, identifying key papers, topics, and research groups. The integration between text analysis and citation context turns ASE an informative representation. However, the visualization suffers from the problem of occlusion, as textual labels can overlap. Survis [10] is a visual analytic system designed to analyze and disseminate literature databases. A set of linked-views allows users to explore citation relations over time. One remarkable feature is the use of an interactive selector for enriching visualizations, providing a visual mechanism for ordering and filtering publications. Literature Explorer [11] uses standard components such as trees and theme river to detect thematic topics to support document retrieval, avoiding that the number of topics has to be pre-defined.

Image-based methods aim to extract and processing images from scientific documents, which are then employed to query and compare scientific documents. One of the few image-based approaches described in the literature is the work by Deserno *et al.* [12], which makes use of images with annotated words to query and group medical documents, reporting a gain in the quality of the query due to the use of images. In fact, the benefit of using images to enrich the querying process has also been reported by Muller *et al.* [13], showing that the relevance of documents retrieved from text and images is higher than using only textual queries. Commercial tools also are part of this group, as the case of Pinterest [14], and Alibaba [15], which faces the big challenge of searching into large image collections by using deep neural network models.

In a lower number, some approaches are devoted to combining two or more methods. For instance, Felizardo *et al.* in [16] uses graphs and edge bundles to understand how a network of articles references each other in the collection while examines the textual content by using a multidimensional projection method. However, it falls in a visual occlusion when it scales in number of documents, especially in its citation map, impairing the exploration of large datasets. PaperVis [17] presents a mixed representation based on keywords and citations for exploring scientific papers. Despite that, none of them allow us to manipulate these interactions from the user activity to generate new insights and supporting the exploration task.

The method proposed in this work combines the last two approaches discussed above, enabling interactive linked-components to efficiently uncover hidden relation patterns in scientific document collections. Moreover, DRIFT allows analysts to restore and compare previous states of their interaction, helping the construction of insights from different selections. DRIFT turns out to be useful in several tasks, as the quick

identification of papers of interest and analysis of their content.

III. GOALS AND ANALYTICAL TASKS

To define our goals, we had a series of meetings with multiple researchers with 5 to 15 years of experience. Also, we conducted an exhaustive literature review to evaluate available systems for scientific literature exploration. As a result of this, we came up with a set of goals and analytical tasks that guided our tool design.

A. Goals

Below we describe the four objectives that lead to the development of our tool.

G1. Support exploration of scientific documents collections.

Available digital libraries offer limited tools for analyzing scientific documents since their exploration relies on the accuracy of its search engine for retrieving relevant documents. However, researchers could not have exact inputs for performing accurate queries, requiring an exploratory analysis to know about the collection and extract significant insights. Our goal is to build a visual analytic tool that enables scientific document collection exploration by combining a set of interactive resources and allowing the analyst to identify documents of interest. In this way, digital libraries might benefit from this proposal.

G2. Integrate image and textual content. Most scientific literature exploration tools focus on text to organize documents—*e.g.*, text matching, citation networks—preventing the exploitation of several features available in scientific papers. Thus, one main goal for our project is to build a tool to perform a multimodal exploration. For that, we wish the analyst to query for both image and textual content to led the exploration process.

G3. Understand metadata and topics in documents groups.

Researchers are quite familiar with reviewing document metadata—*e.g.*, authors, publisher, and publication date—since it provides additional information to decide about document relevance. Likewise, recognizing topics rapidly from document groups enhances analysts' capabilities to review more literature. We identify this goal as an opportunity for improving the manner how researchers can effectively extract insights from document collections while interactively refine their search criteria.

G4. Support literature review task. Exploring and analyzing scientific literature end up in customized collections containing relevant documents for the analyst. These collections organize references according to specific interests and motivations. For instance, support the writing of the Related Work section for an article or prepare a bibliography for a curricular syllabus.

B. Analytical Tasks

After understanding the goals of the project, we define the set of analytical tasks that our tool must support.

T1. Image similarity queries. Given a query image, we want our tool to be able to retrieve a set of images ranked by similarity. These results allow the analyst to discover

documents associated with the retrieved images. This task supports goals **G1** and **G2**.

T2. Group documents based on image and textual content. Enable analysts to group documents and create collections considering both textual and image content. This task allow us to achieve goals **G1** and **G2**.

T3. Selecting and filtering collections. Allow the analyst to select a document collection and filter its content according to his/her search criteria. This task allow us to achieve goals **G1** and **G2**.

T4. Compare document collections. Enable topics and meta-data analysis in document collections created by the analysts. Moreover, our tool must facilitate comparison between document collections to identify similarities among them. This task gives support to goal **G3**.

T5. Storing and managing document collections. Each time an analyst creates a document collection, he/she must be able to save it. Moreover, users should have access to the stored collections to perform operations such as querying, retrieving, and merging. This task allow us to achieve goals **G3** and **G4**.

T6. Exporting customized document collections. After the exploration process, the analyst should be able to export his/her results into a human and machine-readable format. This task allows us to achieve goal **G4**.

IV. DRIFT

DRIFT is a visualization tool designed to support analysis and exploration of large scientific document collections, revealing the similarity between document contents while enabling interactive resources to store and recover intermediate steps of the exploratory analysis. DRIFT’s methodology, illustrated in Fig. 1, comprises three main steps: (i) extraction and processing of each document, (ii) interactive exploration using both image and textual information, and (iii) visual analysis of selected document subsets.

DRIFT allows the analysts to choose the number of images to be used for querying as well as the number of images to be retrieved by the CBIR components. Each component brings a set of documents associated with the images, *i.e.*, the documents that contain the retrieved images as part of their content. The associated documents are considered as control points to guide the multidimensional projection process, which is responsible for mapping the documents based on their similarity to 2D visual space. Textual features are only used to accomplish the projection. Thus, using images of interest users can find relevant documents that are then used to drive an exploratory analysis. Additionally, we implement visual resources to support analytical tasks, *i.e.*, author-frequency, and year-frequency histograms as well as a topic-based wordcloud. A streamgraph component, named as *Selection Visual Manager*, helps to save and display intermediate steps of the exploratory analysis. Such intermediate steps, called *states*, can be recovered, compared, and employed to generate new states, which can be downloaded as subsets of documents.

We design these visual components to achieve the identified goals. All components address at least one analytical tasks

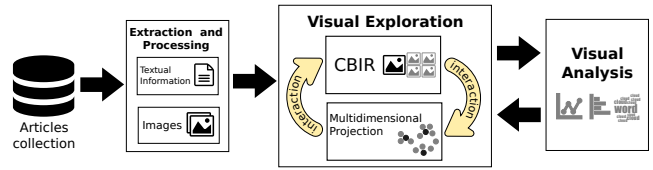


Fig. 1: An overview of our proposed three-step methodology.

TABLE I: Methodological and analytical properties and their related tools.

	T1	T2	T3	T4	T5	T6
Multi-CBIR View	•					
Multidimensional Projection View		•	•			
List-based Selection Refinement View		•	•	•		
Selection Content Summarization View			•	•		
Selection Inspector View					•	•
State Manager View					•	•

described in Section III-B. Table I details the relation between the visual resources and analytical tasks (T1-T6 columns).

A. Content Extraction and Processing

We used ArXiv® digital library and the e-proceedings of a well-known conference in visual computing from 2011 to 2014 as the document collections to be handled by DRIFT. The keywords used as well as the number of images contained in each category is described in Table II.

Content Extraction. The tool *pdf2text* from Poppler library is applied to convert the textual content of PDF files into ASCII text files. The tool *pdfimage* also from Poppler library is used to convert pages of PDF documents into 8-bit PNG image files. The PNG files are input into an image processing pipeline to extract figures contained on each page. We use the global Otsu method [18] to binarize the PNG files, searching for components in the resulting binary images.

Text Processing. We adopt simple but robust textual processing to perform this task. In summary, ASCII files associated with each document are processed to extract term frequency vectors. Additionally, we performed a conventional text processing filtering—*i.e.*, stemming, stop word removal, and definition of Luhn’s lower and upper cuts—ending with a tf-idf vector representation of each document. We process only the abstract rather than the full content of each document to reduce the computational burden.

Image Features Extraction. We relied on AlexNet [19] architecture pre-trained on the ImageNet dataset. Unlike the original network architecture, we changed the last layer from 1000 to 5 neurons. It was done to fine-tune on our first dataset (DT1) which consisted of 5 categories. Then, we modified the learning rate of the last layer by a factor of 10; this allows the back-propagation to have a high effect on the last layer and a slight impact on the previous ones. Finally, it run 50,000 iterations with a momentum of 0.9 and a base learning rate of 0.001. The features extracted from the last fully connected

TABLE II: Datasets used for our study

ID	Query	# Articles	# Images	Source
DT1	<i>seismic</i>	274	5,002	ArXiv
	<i>market</i>	273	2,772	
	<i>gravitational</i>	274	3,082	
	<i>disease</i>	274	3,795	
	<i>gene</i>	274	3,731	
DT2	Proceeding 2011	45	1,010	IEEE Xplore
	Proceeding 2012	45	1,195	
	Proceeding 2013	36	869	
	Proceeding 2014	45	1,033	
DT3	computer graphics	95	2,010	ArXiv
	image processing	93	1,922	
	computer vision	96	2,732	

layer is a vector of 4,096 elements. For the other two datasets (DT2-3) we did not fine-tune the CNN because the images did not contain classes, that is why we extract the feature of these two datasets using the model already trained with DT1. Note that although we are doing fine-tuning in a classifier, our goal is not to use the classifier output, but to refine and extract the characteristics for our problem. To avoid the curse of dimensionality, we additionally reduce the feature vector dimension to 50 using PCA.

B. Multi-CBIR and 2D Mapping

In the following, we describe the two main views that lead the exploration process and how they are integrated for interactively finding key documents.

Multi-CBIR view. Traditionally, a CBIR mechanism returns a similarity-based ordered list of images by querying one specific image. The similarity can be defined as a distance measure between feature vectors. In our implementation the CBIR retrieves a user-defined number of similar images, which are displayed next to the query image in an image board, as illustrated in Fig. 3a. Up to five queries can be performed simultaneously using different input images. On the image board, images belonging to the same document are highlighted when the user hovers a specific image, fading out the remaining images. The exploration starts in this view.

Multidimensional Projection view. Textual information from each document gives rise to a high-dimensional vector that represents the document. In order to interactively explore the documents based on their similarity, we map the high-dimensional vectors to a 2D visual space using a multidimensional projection method. Specifically, we use Local Affine Multidimensional Projection (LAMP) [20] due to its interactive capability and good performance in terms of accuracy. This method uses a reduced number of sample points (called control points) to drive the mapping of the remaining data instances into the visual space. LAMP makes it possible to interactively positioning control points on the visual space, updating the projection layout according to user intervention. This main feature is decisive for our choice of using LAMP over any other local multidimensional projection method. This view is located in the middle of the interface, as shown in Fig. 3b.

Finding Key Documents. DRIFT combines multi-CBIR and multidimensional projection views to enable an interactive

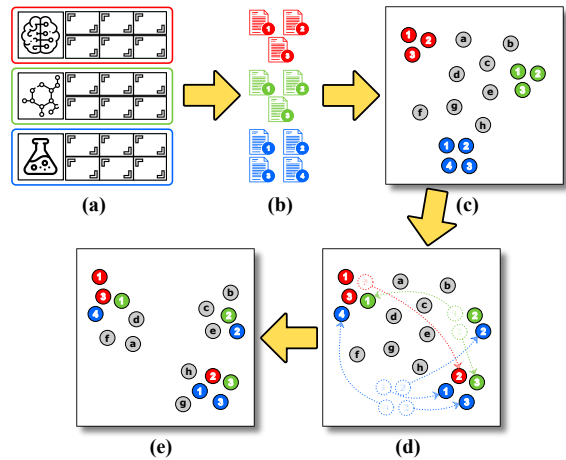


Fig. 2: Finding key documents by the combination of both Multi-CBIR and Multidimensional Projection views.

exploration of document collections. As illustrated in Fig. 2 a user query for images using the CBIR component and (a) our system retrieves multiple set of images based on image queries, (b) automatically it selects the documents where retrieved images are contained, (c) these documents are used as control points by our multidimensional projection view for mapping the entire document collection where the parameter α can be tuned between zero and one forcing LAMP to perform a more local or global mapping, respectively, (d) according to his/her interests the analyst repositions the control points to customize groups, and finally (e) the entire collection is reprojected based on such repositioning.

C. Visual Analysis

A main functionality of our methodology is the interactive selection of subsets of scientific documents. The user can select a subset of articles by drawing a polygon around points (documents) of the projection. The borders of the points selected will be colored in red. Each time the analyst selects a subset of documents, linked-views are updated showing relevant information from the selected documents. Relevant information is depicted in the following visual components:

List-based Selection Refinement view. This component shows the list of selected documents, depicting the title and DOI, where the later is linked to the original publication page. Particular documents can be removed from the list by clicking in the trash button, as shown in Fig. 3c.

Selection Content Summarization view. Once a group of documents is selected, three visual summarization widgets are updated to show relevant content from the selected subset. Specifically, the visual summarization widgets show author-frequency and publication-year frequency histograms, and topic wordcloud, as shown in Fig. 3d. Those widgets provide an overall view of most-cited authors, topics discussed, or which period comprises the larger number of publications.

Selection Inspector view. The process of selecting documents, inspect their summary to choose the most relevant ones is typically accomplished several times. The final subset is the merging of documents that remains at the end of each

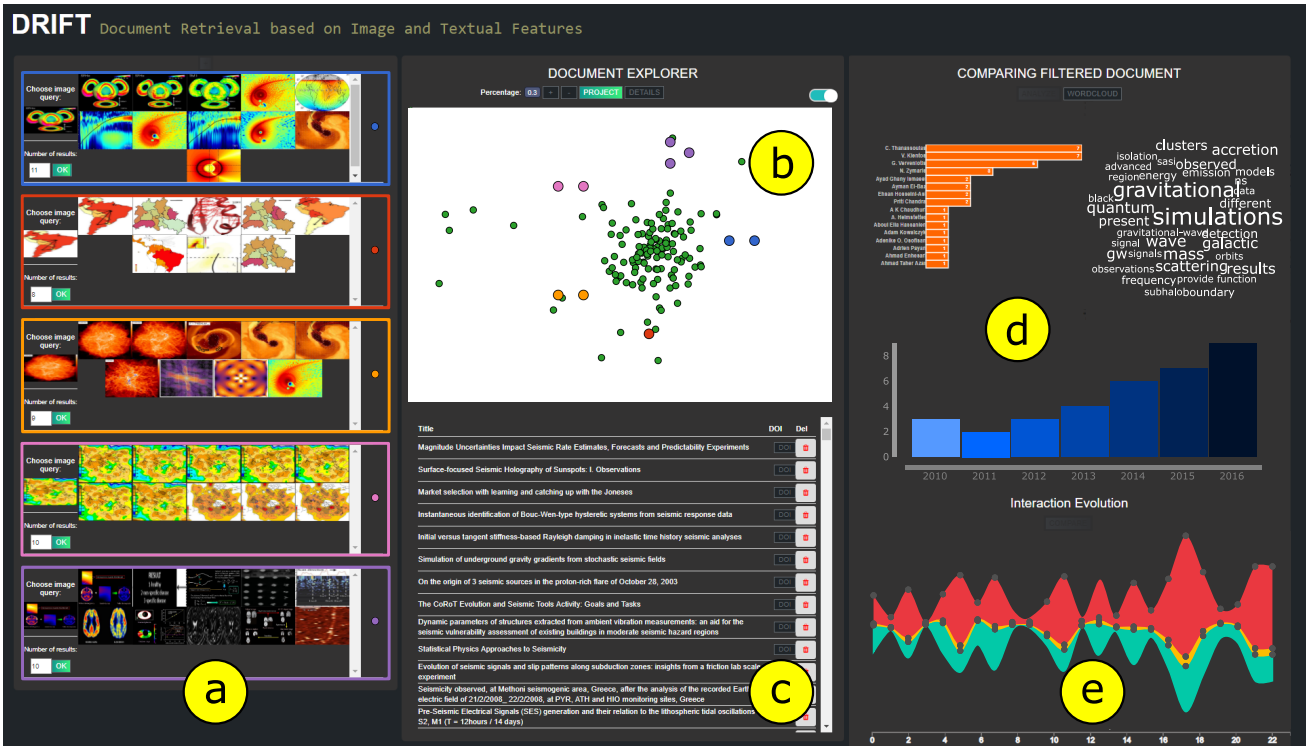


Fig. 3: An overview of DRIFT views: (a) Multi-CBIR, (b) Multidimensional Projection, (c) List-based Selection Refinement, (d) Selection Content Summarization, and (e) Selection Inspector.

iteration. Aiming to facilitate such an iterative process, we employ an interactive streamgraph metaphor that stores the documents, images, and authors resulting from each iteration cycle. The number of documents, images, and authors are represented as streamgraph layers— orange, red, and turquoise, respectively— and each iteration cycle is marked with three vertically aligned dots in the layout, as illustrated in Fig. 3e. This widget allows to recovering a state saved during the analytical process, supporting to restore relevant articles identified during any cycle. Indeed, the widget enables a wide range of operations over, *e.g.*, compare, combine, or delete the result of any iteration cycle.

State Manager view. Suppose that during the exploratory analysis two states (S_A and S_B) are produced. After selecting these two states from the streamgraph and click on the “compare” button a modal window shows up, as illustrated in Fig. 4. To compare the content of two states, we implement this view (inspired in [21]), which performs set operations on states S_A and S_B : intersection (Fig. 4a) and difference (Fig. 4b-c). The result of a set operation (Fig. 4d) can be saved as a new state in the streamgraph. On the bottom part of the modal window, under title *Selected*, the title of chosen documents are displayed. Once the *New state* button is selected, a new state will be added to the streamgraph. The user can also export the filtered documents as *.json* file containing the selected article titles and their respective web links.

Our prototype is developed in JavaScript, using D3.js library what should make it possible to plug DRIFT into digital libraries running on Web. The preprocessing steps such as feature extraction are speeded up using C++ libraries.

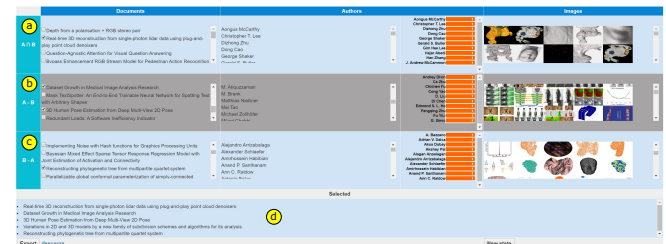


Fig. 4: Comparing two different states of user interaction by using the *State Manager View*.

V. CASE STUDIES

We propose two scenarios to assess DRIFT’s effectiveness in terms of exploration of a scientific document collection.

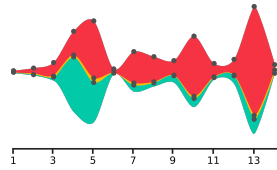
A. Exploring the DTI Dataset

For a fuller understanding of the step-by-step performed in this case study, support your reading by watching the supplementary material. Suppose we are looking for articles related to gravitational waves associated with supernovae. We start the exploratory analysis by performing queries from two images related to the topic of interest.

To emulate the behavior of an analyst, we selected two inputs knowing *a priori* that they appeared in articles that talk about the topic of search. The first input image is related to a novel gravitational-wave signature in supernovae and we decided to retrieve up to six images from CBIR. Images after the sixth do not belong within the domain we are looking for. We found three articles related to seismic features and gravitational waves. Using the second input image and setting

the number of retrieved images to nine the search results in five documents.

Eight documents, used as control points, drive the mapping of the document collection. The inline figure illustrates the entire process displaying the resulting streamgraph. In the first state, the projection displays two sets, the blue and the red control points on the right and the left side, respectively. Complementary components, such as the wordcloud, summarize the selection with the words “*gravitational*”, “*frequency*” and “*scattering*”. In the same way, the second state is the selection of points on the left side. In the state 5, the inner region of the projected point cloud is selected, revealing a broader range of topics published from 2007 to 2016.



States 6 to 9 comprise different document selections on the same projection. State 10 combines eight relevant articles from states 8 and 9. On the state 12, documents on the rightmost region of the projection layout are associated with topics of interest. However, some documents clutter the analysis, so we resort to the managing states tool to compare the current and the ninth state. The resulting analysis gives rise to a subset of nine documents saved in the thirteenth state, which are mostly related to “*gene*”, “*data*”, and “*model*”. Finally, we decided to combine the two states deemed most relevant for our analysis, the eleventh and thirteenth states. We use one last time the managing state tool, resulting in a set of articles closely related to the topics of interest, namely “*gravitational*”, “*wave*”, “*simulations*”, “*scattering*”, which have been published in 2011, 2013, 2015, and 2016, this later with a larger number of publications. The merged states export as a JSON file for future analysis and readings.

B. Exploring the DT2 Dataset

In the second case, we aim to find articles related to 3D models (see Fig. 5) starting with five query images. We chose four of them by their explicit relation to our target, and the last from another topic, *i.e.*, a common picture in the context of image processing. The main purpose of using such an image is to employ it to properly drive the multidimensional projection, pushing unrelated articles towards this control point.

The initial projection places most instances in the middle of the layout, as illustrated in Fig. 5a. The list of titles—shown with the interactive selector—allow us to identify that documents of peripheral regions belong to distinct topics, see Fig.s 5b and 5c, respectively.

The interaction between control points and images of CBIR enables the identification of relevant control points. Therefore, we can rearrange the projection by moving a blue control point to the bottom-left, and the red to the right, as shown in Fig. 5d. The content of the wordcloud reveals terms as “*face*”, “*reconstruction*”, “*skull*”. Also, in Fig. 5e we gather red, blue, and violet control points, and select them and their neighbors. Both selections reveal two configurations, in Fig. 5d

we group some documents about 3D mostly, while Fig. 5e depicts images and words conveying image processing context.

Then, we aim to explore the content in some different regions of the projection, so after one more interaction, we found two groups for analysis. In Fig. 5f we highlight (orange and purple borders) these selections which contain partially related articles to our search. Both of them were stored in *Selection Inspector View* as states 2 and 3 respectively, as illustrated in Fig. 5g. After interacting in our *State Manager View* we filtered a few articles to compose a new state, stored as state 4.

By simple inspection, we can notice that most of the retrieved articles depict similar textual content to the selected control points. We store this new selection as state 5. Then, we decide to analyze the contribution of one of the orange points, which talks about curves on surfaces, so we drag it towards the middle region, bringing with it the most similar documents, as illustrated in Fig. 5j. As can be noticed, the neighbors talk in general on geometry processing for surfaces, which is close, but not completely, related to our search. We store this selection as state 6. We opt to compare the states “0” and “1”, since they were not carefully explored yet. In Fig. 5k, the *State Manager* view shows two selections without intersections. We found four useful articles for our purposes inspecting titles, authors, and contained images We store them into state 7.

Finally, we compare the states 5 and 7. We found two selections without intersection but containing four articles highly relevant to our study, as illustrated in Fig. 5l. At the end of our exploration, we have produced three states containing scientific articles that allow us to extract related methods to 3D modeling in computer graphics, *i.e.*, fourth, sixth, and eighth states. As can be noticed, we successfully discriminate such articles, even in a highly related-topic collection, by using images and textual information included in each article.

VI. USER EVALUATION

We conducted a controlled user evaluation to assess whether DRIFT enables the discovery of documents of interest plausible time in comparison with the list-based traditional paradigm. The evaluation follows a four-step procedure: (1) *Introduction*, we gave a brief explanation of the purpose of the study to the participants. (2) *Tool exposure*, we show the participants the functionalities in DRIFT. (3) *User familiarization*, participants had 10 minutes to play with the tool, exploring a synthetic collection. And (4) *Evaluation*, we invited the participants to perform a specific search activity.

We set-up two search activities (named A1 and A2) by using two questions detailed in Table III. All activity involved an analytical procedure from the DT3 dataset, detailed in Table II.

For this study, we invited six users experienced in literature systematic review—four with a master’s degree and two with a doctoral degree—working on image processing, machine learning, or visualization from different research institutions. We split the participants into two groups, group GR1 containing users 1 to 3 (U1-U3), and group GR2 users 4 to 6 (U4-U6).

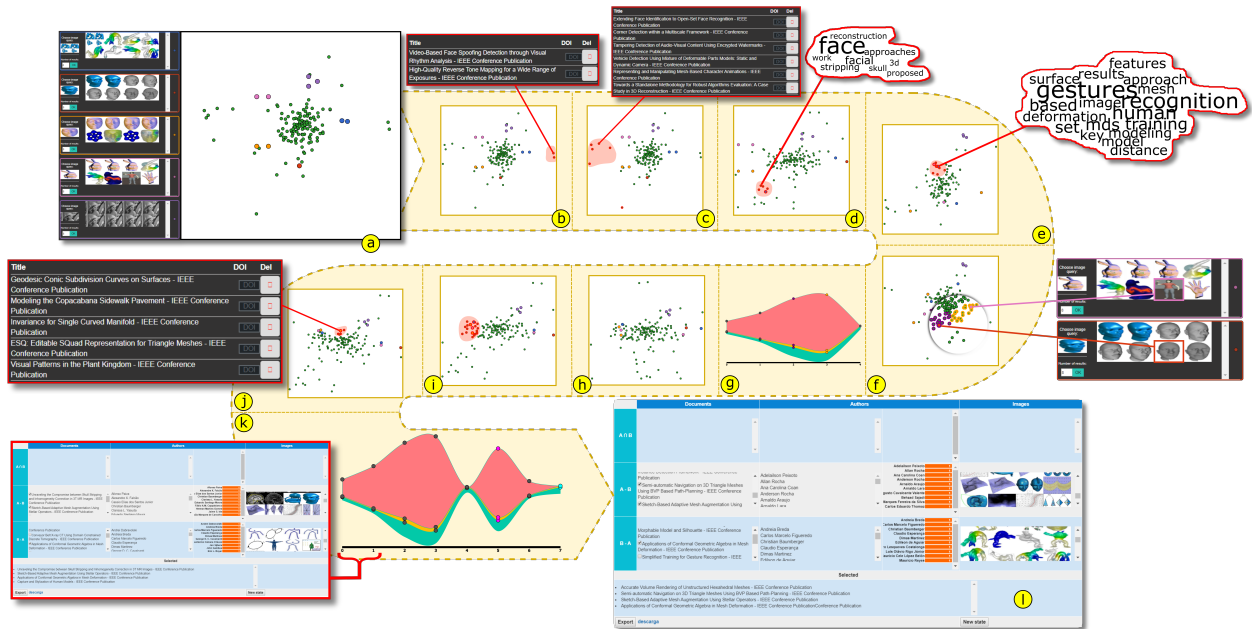


Fig. 5: Summarizing interactions in the case study on DT2: (a) input images and initial projection, (b-k) multiple interactions that include point reposition, comparison among states and inspection of visual resources, and (l) final selection.

Both groups performed the search activity (A1) using DRIFT and list-based paradigm respectively. Later, they performed the second activity (A2) inversely. We implement a list-based interface emulating traditional scientific repositories. Before our interface display all documents from DT3, we ranked them extracting the content from their abstracts and performing a string matching algorithm with the terms “*face rendering*”, and “*volumetric human people*”.

TABLE III: Proposed activities, questions, and group distribution to user evaluation.

	Activity Target	Question	DRIFT	List
A1	Identify a particular group of documents	How many and which documents use human faces for rendering?	GR1	GR2
A2		How many and which documents address volumetric representations of human body?	GR2	GR1

This study verified the following hypothesis:

- *Users of DRIFT will spend less time to answer questions that require a global analysis of the corpus, with no significant loss in precision.*

We computed Precision and Recall to evaluate the relevance of document retrieved. Additionally, we stored the elapsed times taken to accomplish A1 and A2 activities. Results are shown in Table IV. In A1, by the list-based approach, researchers obtained a significantly lower performance in terms of all measures. The difference between the best precision value for list-based and the worst for DRIFT is close to 0.22, and even that U1 performs perfectly the test, obtaining a precision of 1. However, the elapsed times show the users spent almost the same time (close to the average value) to perform this task.

TABLE IV: Precision (Pr), Recall (Rc) values and spent times (t) in minutes obtained by the six participants in A1 and A2.

		list-based			DRIFT			
		Pr	Rc	t	Pr	Rc	t	
A1	U4	0.20	0.20	12:08	U1	1.00	0.60	7:38
	U5	0.28	0.40	12:39	U2	0.80	0.80	7:34
	U6	0.33	0.20	12:02	U3	0.55	1.00	16:11
A2	U1	0.75	0.37	13:55	U4	0.75	0.09	6:27
	U2	0.28	0.63	10:28	U5	0.75	0.09	9:08
	U3	0.25	0.63	23:12	U6	0.70	0.09	9:29
Avg		0.35	0.41		0.76	0.45		

On the other hand, U1 and U2 of DRIFT obtained the lowest times for this experiment, except by U3 that spent much more time. In A2 activity, GR1’ precision average was decreased while GR2 was increased considerably. For instance, U4—who obtained the poorest precision in A1—improves its performance obtaining 0.75 of precision in A2. Moreover, inspecting the elapsed times, DRIFT’ users obtained the three lowest times. The lowest row in Table IV summarizes the activities by average.

A t-test comparison was performed for each group, showing 5 percent level ($\alpha = 0.05$), a statistical difference between DRIFT and the list-based approach. For precision values, we obtained a two-tailed p-value equals to 0.0023, which is considered to be statistically significant. These results confirm our initial hypothesis.

VII. DISCUSSION AND LIMITATIONS

The described design and case studies clearly show that our approach provides an efficient alternative for exploring and analyzing extensive collections of scientific documents. Our implementation into the web context aims to introduce a new paradigm into digital libraries’ exploration. In that way, we allow the analysts to extract insights while mitigating

the overwork for establishing mental relationships among documents, as digital libraries currently have us accustomed.

The novel combination of multiple-CBIR and multidimensional projection represents a flexible and powerful mechanism to gathers image and textual features in a methodology for document exploration. However, the feature extraction step dramatically impacts the whole process of analysis. It is crucial to have a valuable set of features describing images and texts to help us improve the accuracy of searches. Particularly, some articles do not contain images, in that case the interaction with any CBIR component is limited because DRIFT can only use their textual information.

Our implementation visually illustrates the states to be queried, filtered, and combined. It relies on a streamgraph-based plot that performs an advisory role. However, it is not entirely appropriate when document selections are unbalanced, *i.e.*, collections with few elements can be challenging to visualize. On the other side, lower sections of the graph can help reveal outliers. Moreover, it allows us to stack more attributes and visualize them simultaneously, *e.g.*, the number of reads/downloads or average h-index from the entire collection.

VIII. CONCLUSION

In this work, we propose DRIFT, a novel visual analytic tool for analyzing scientific literature collection. It comprises multiple linked components such as content-based image retrieval, multidimensional projection, frequency histograms, word clouds, and a streamgraph. The proposed method is fully interactive, intuitive for analysts aiming to extract subsets of documents according to its requirements. Moreover, it proposes a new paradigm that conciliates both image and textual features into a continuous feedback process. Furthermore, we implemented DRIFT in a web-based environment with the future envision to plug it into a digital library. We demonstrate the usefulness of our methodology in two detailed case studies and user evaluation. Results show that our approach is an attractive method for analyzing multiple types of documents.

ACKNOWLEDGMENT

This work was supported by Universidad Católica San Pablo (Project #UCSP-2018-PI-47), CNPq-Brazil (grants #303552/2017-4, #312483/2018-0), São Paulo Research Foundation (FAPESP)-Brazil (grant #2013/07375-0) and Getulio Vargas Foundation. The views expressed are those of the authors and do not reflect the official policy or position of the São Paulo Research Foundation.

REFERENCES

- [1] E. Gomez-Nieto, W. Casaca, D. Motta, I. Hartmann, G. Taubin, and L. G. Nonato, "Dealing with multiple requirements in geometric arrangements," *IEEE TVCG*, vol. 22, no. 3, pp. 1223–1235, March 2016.
- [2] M. Dörk, S. Carpendale, C. Collins, and C. Williamson, "Visgets: Coordinated visualizations for web-based information exploration and discovery," *IEEE TVCG*, vol. 14, no. 6, pp. 1205–1212, 2008.
- [3] A. Abbasi and J. Altmann, "On the correlation between research performance and social network analysis measures applied to research collaboration networks," in *System Sciences (HICSS)*, 2011 44th Hawaii International Conference on, Jan 2011, pp. 1–10.
- [4] M. A. Rodriguez and A. Pepe, "On the relationship between the structural and socioacademic communities of a coauthorship network," *Journal of Informetrics*, vol. 2, no. 3, pp. 195 – 201, 2008.
- [5] C. M. Morel, S. J. Serruya, G. O. Penna, and R. Guimarães, "Co-authorship network analysis: a powerful tool for strategic planning of research, development and capacity building programs on neglected diseases," *PLoS Negl Trop Dis*, vol. 3, no. 8, p. e501, 2009.
- [6] S. Liu, C. Chen, K. Ding, B. Wang, K. Xu, and Y. Lin, "Literature retrieval based on citation context," *Scientometrics*, vol. 101, no. 2, pp. 1293–1307, 2014.
- [7] E. Yan and Y. Ding, "Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 7, pp. 1313–1326, 2012.
- [8] M. Berger, K. McDonough, and L. M. Seversky, "cite2vec: Citation-driven document exploration via word embeddings," *IEEE TVCG*, vol. 23, no. 1, pp. 691–700, 2016.
- [9] C. Dunne, B. Shneiderman, R. Gove, J. Klavans, and B. Dorr, "Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization," *J. Am. Soc. Inf. Technol.*, vol. 63, no. 12, pp. 2351–2369, Dec. 2012.
- [10] F. Beck, S. Koch, and D. Weiskopf, "Visual analysis and dissemination of scientific literature collections with surviS," *IEEE TVCG*, vol. 22, no. 1, pp. 180–189, Jan 2016.
- [11] S. Wu, Y. Zhao, F. Parvinzamin, N. T. Ersotelos, H. Wei, and F. Dong, "Literature explorer: effective retrieval of scientific documents through nonparametric thematic topic detection," *The Visual Computer*, pp. 1–18, 2019.
- [12] T. M. Deserno, S. Antani, and L. Rodney Long, "Content-based image retrieval for scientific literature access," *Methods of Information in Medicine*, vol. 48, no. 4, pp. 371–380, 2009.
- [13] H. Müller, A. Foncubierta-Rodríguez, C. Lin, and I. Eggel, "Determining the relative importance of figures in journal articles to find representative images," in *Medical Imaging 2013: Advanced PACS-based Imaging Informatics and Therapeutic Applications*, vol. 8674. International Society for Optics and Photonics, 2013, p. 86740I.
- [14] A. Zhai, D. Kislyuk, Y. Jing, M. Feng, E. Tzeng, J. Donahue, Y. L. Du, and T. Darrell, "Visual discovery at pinterest," in *Proceedings of the 26th International Conference on World Wide Web Companion*, ser. WWW '17 Companion. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017, p. 515–524.
- [15] Y. Zhang, P. Pan, Y. Zheng, K. Zhao, Y. Zhang, X. Ren, and R. Jin, "Visual search at alibaba," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 993–1001.
- [16] K. R. Felizardo, G. F. Andery, F. V. Paulovich, R. Minghim, and J. C. Maldonado, "A visual analysis approach to validate the selection review of primary studies in systematic reviews," *Information and Software Technology*, vol. 54, no. 10, pp. 1079–1091, 2012.
- [17] J.-K. Chou and C.-K. Yang, "Papervis: Literature review made easy," in *Computer Graphics Forum*, vol. 30, no. 3. Wiley Online Library, 2011, pp. 721–730.
- [18] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [20] P. Joia, D. Coimbra, J. A. Cuminato, F. V. Paulovich, and L. G. Nonato, "Local affine multidimensional projection," *IEEE TVCG*, vol. 17, no. 12, pp. 2563–2571, Dec 2011.
- [21] G. G. Zanabria, L. G. Nonato, and E. Gomez-Nieto, "istar (*): An interactive star coordinates approach for high-dimensional data exploration," *Computers & Graphics*, vol. 60, pp. 107–118, 2016.