National Endowment for the Humanities
Office of Digital Humanities

Application for a Level I Start-Up Grant
for
Tool for Topic Modeling of Moving Images (ToMMI)

by
Juan Llamas Rodriguez
University of California, Santa Barbara

Table of Contents

I. Abstract

This proposal is for a Level I start-up digital humanities grant to build a Tool for Topic Modeling of Moving Images (ToMMI), an application that, given a set of films, will process them to produce a number of topics associated with their visual style and content. ToMMI will help researchers in areas of moving image inquiry, such as film studies, to study large corpora of audiovisual texts by making their underlying structures and patterns visible and presenting them in manageable forms such as graphs, timelines, and visualizations. This tool will remain focused on the types of analyses that are specific to moving image studies; therefore, both during its processes and within the output visualizations, these visual texts will be compared, modeled, and represented as images, and not reduced to linguistic markers or numerical values. ToMMI is expected to become an aid in the research and analysis of large sets of visual texts, particularly those that have heretofore gone undertheorized.

Statement of Innovation

The most salient innovative feature of ToMMI is its treatment of moving images as images during the processing and output, allowing for characteristics that often go unnoticed in semantic search—such as framing, lighting, and depth—to also be considered in the topic modeling. This feature differentiates ToMMI from other image search and annotation tools and makes it indispensable to disciplines concerned with moving images since these types of characteristics are just as important for analysis. Furthermore, the outputs as “direct visualizations” ensure that the results from the topic modeling will not be reduced to word descriptors or numerical data but rather will be presented in their qualitative entirety for the researcher to analyze.

Statement of Humanities Significance

The recent influence of fields such as cultural studies, media industry studies, and media anthropology on disciplines of moving image inquiry like film studies has been most notable in regards to the increasing theorization of popular texts previously not considered worthy of study. This important disciplinary move allows us to get a better sense of the entire cinematographic canon, but it also presents several methodological obstacles, not least of which is the impossibility of analyzing textually hundreds of films for any one project. ToMMI will represent a significant aid in overcoming these obstacles since it will allow a researcher to automatically model patterns in large corpora of films at a time, as well as make these patterns visible in novel yet manageable forms.

II. Narrative

In his monograph *Shadow Economies of Cinema*, Ramon Lobato proposes that there is a slaughterhouse of cinema, which, like the slaughterhouse of literature that Franco Moretti identifies, metaphorically represents the place where thousands of films are left to be forgotten despite the fact that they comprise a significant portion of past and present cinematographic production.[1] The sorts of films that Lobato identifies—grindhouse horror and other niche genres, pirated and other informally-made movies, Christian and other religious-themed films, to name a few—are different in most regards. However, what makes these B-movies similarly condemned to the slaughterhouse of cinema is that they are stylistically modular and immensely popular, and quantitatively represent the majority of global audiovisual consumption.[2] Ramon Lobato urges that film studies should also attend to the exploration of these texts. However, since film studies remains heavily invested in the textual analysis of moving images as a foundational method in the development of scholarship—namely, a "close reading" type of analysis such the shot-by-shot breakdown—an epistemological move like the one Lobato suggests proves unfeasible when we are confronted with huge amounts of B-movies.

Therefore, the basic premise of this project will be to take Lobato's urging to heart by developing a tool that would enable textual analysis of large corpora of films on a practical level. On the one hand, this project will provide a technical solution for the development of the types of projects that Lobato suggests. On the other hand, this tool will help address research questions present in both film studies and digital humanities, allowing these two areas to come into further conversation. The theoretical question underlying this project is, how can film studies retain its focus on textual analysis when considering large corpora of films? In other words, how could such an analysis take place and what would it look like?

In this regard, the lessons learned from developments in digital humanities prove a generative starting point. Following on Franco Moretti's distinction between close reading and distant reading, I propose that the best way to tackle the study of large corpora of films is through a form of "distant watching", a form of analysis where the distance is a specific form of knowledge unavailable through close readings.[3] Elsewhere I have referred to this form of distant watching as telescoping in an attempt to emphasize a) the primacy of visual analysis over its linguistic analogue, b) the insurmountable distance both assumed and requisite for such analysis, and c) the figurative compression of incredibly large bodies of films into smaller, more manageable forms. When extended, this figurative sense of telescoping also emphasizes the fact that in distant watching, the film texts themselves disappear and the objects of study become worlds—that is, systems, themes, or patterns.

How, then, do we make these patterns visible? If, by definition, B-movies are derivative—that is, we classify them as such because we assume that all their stylistic choices are only in response

[1] Ramon Lobato. *Shadow Economies of Cinema*. London: British Film Institute, 2012. 32-33.
[2] For a practical sense, I will refer to these types of films as B-movies, not in attempt to group them all as a genre, but simply to reiterate their status as secondary within scholarly research.
[3] Franco Moretti. *Graphs, Maps, Trees*. London: Verso, 2005. 1.

to genre conventions—then in order to notice these choices, we must draw patterns across the hundreds of films out there. In order to begin to do this, borrowing from another digital humanities practice proves instructive. Specifically, the use of topic modeling to examine texts and discover thematic patterns has already been shown to be an important tool. Topic modeling uses algorithms to discover hidden thematic structures in large collections of texts and provides results that then allow summarizing, exploring, and theorizing about the corpus.[4] As a process, topic modeling relies on two assumptions: a) there are a fixed number of patterns of word use that tend to occur together in documents, which are called topics, and b) each document in the corpus exhibits these topics to varying degrees.[5]

This project's aim is to produce a digital tool that provides similar analytical benefits to those of topic modeling to the study of moving images. The outcome of this project will be a *Tool for topic Modeling of Moving Images (**ToMMI**)*, which departs from the theoretical bases for statistical topic modeling but is adapted to the specific research needs of disciplines concerned with the textual analysis of moving images in general, and film studies in particular. ToMMI will depart from the same theoretical assumption that there are a set of patterns occurring together in these films and that these patterns are present to varying degrees. However, whereas topic modeling focuses on patterns of words, ToMMI will work with patterns of images. The film texts to be analyzed will be broken down into their constitutive frames, and these frames will be run through a search algorithm in order to extract patterns among them as well as values on how often these patterns are present in the set of films being considered. As with topic modeling's use of "stop words", ToMMI will depend on the separation of constitutive elements from their syntagmatic relations; namely, it may obfuscate several forms of editing. This is not an insignificant simplification, but I would argue, as Ted Underwood has done for words in a text, that a film's frames are already features of a very high-level kind.[6] But, to be sure, the analogy between words in topic modeling and frames in ToMMI ends there, since it is also true that an image is not reducible to a word.

Therefore, it is important to note that at no point during the process are these frames translated to single or multiple linguistic terms. Unlike other tools for search and annotation of videos and images (see the examples in the Environmental Scan), ToMMI is not concerned with describing what is in a frame but with the frame itself. In other words, the search algorithm performed on all the frames will focus on visual similarities between images qua images, and the results will be presented in image form so as to never translate visual characteristics to linguistic markers. Given that current semantic search algorithms for images are predicated on reducing an image to a concept (e.g. tiger, house, meal), they obfuscate important aspects of an image such as framing, shot angle, depth or lack thereof—crucial aspects for film analysis. ToMMI's visual similarity search will pair up images that have these characteristics in common, producing topics based on image composition which may include, but will not be limited to, semantic content. The

[4] David M. Blei. "Topic Modeling and Digital Humanities". *Journal of Digital Humanities* 2.1 (Winter 2012)
[5] Ibid.
[6] Ted Underwood. "Wordcounts are amazing". *The Stone and the Shell* <http://tedunderwood.com/2013/02/20/wordcounts-are-amazing/> Accessed 1 Dec. 2013.

reason for this is that, for film studies, any set of shots that are similarly framed, lit, and focused represents a visual topic as much as any set of shots with the same semantic content.

The lack of translation between visual characteristics and linguistic descriptors extends to the outputs of the ToMMI process. Topics are not labeled or described but rather represented by a visualization of the frames that constitute each topic. Informed by Lev Manovich's notion of "direct visualization", these results are reorganized into a visual representation that preserves the frames' original form.[7] The frames may be reduced quantitatively in size but not qualitatively by translating them to a linguistic marker or to numerical data. Therefore, the researcher's analysis and interpretation of topics occurs after the processing of the films and is not guided by previous assumptions about the characteristics to be found therein. In this way, a researcher is able to both confirm or disprove her assumptions about a set of films as well as discover new patterns she may not have previously considered. It is only then that the promise of a distant form of watching that produces a specific form of knowledge can be achieved.

III. Technical Description of ToMMI

What follows is a general description of the technical process of ToMMI through which a set of films undergoes in order to create the topics, including what the researcher is in charge of defining and how the results can be presented. Initially, the researcher chooses the set of films that she wishes to topic model, and inputs each one as a digital file into ToMMI. Each film's digital file could contain information in their metadata that the researcher is interested in relating to the topics, such as film title, year, director, national precedence, etc. Once the files are inputted into ToMMI, the researcher can choose how many topics to output, and sets the tool to run.

The films are processed into topics in a two-part process. In the first part, each film is broken down into frames—one image per second of the film—and a document is created for each of the films (see *Figure 1*). A document ***D*** contains all the metadata information of the film from which it was made (year, director, etc.) as well as an index of all the frames that were extracted from that film. A frame ***f*** contains an image as well as two numerical values, the document number it belongs to and the order in which it appears in that document. The output from part one is a set of documents $\{D_1, D_2, D_3,..., D_n\}$ and a corpus of frames $\{f_1^1, f_1^2, f_1^3,..., f_n^m\}$.

[7] Lev Manovich. "What is Visualization?" *Visual Studies* 26.1 (2011): 41.

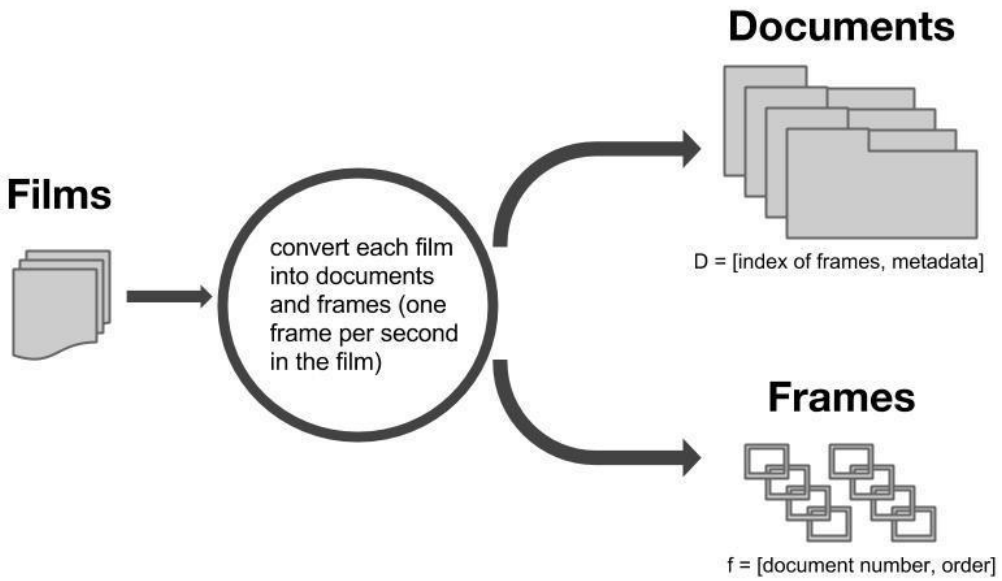

Figure 1: diagram of ToMMI process, part one

The second part of the process takes the corpus of frames and performs a search for each one, creating proto-topics (see *Figure 2*). A proto-topic ***t*** contains the frame that prompted that search and all the frame results, as well as a prevalence value $t_v = 1$. These proto-topics are then compared to each other. When two proto-topics have multiple frames in common, i.e. if 70% or more of the frames in the smaller proto-topic are in the larger proto-topic, then the smaller is combined into the larger. The larger proto-topic is given a prevalence value equivalent to its previous prevalence value plus the smaller proto-topic's prevalence value. All the frames in common are added 1 to their prevalence value ($f_v = f_v+1$). The output from this second part is a set of topics, however many the researcher initially chose. A topic ***T*** contains a prevalence value T_v which stands for how many proto-topics were combined to create said topic, and a frame matrix T_f which contains the frames in the topic with each frame's corresponding prevalence value f_v.

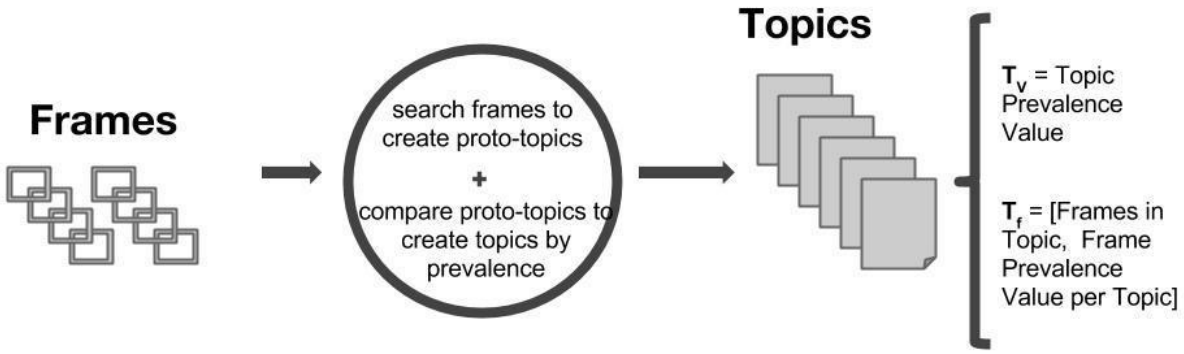

Figure 2: diagram of ToMMI process, part two

Given the outputs of documents, frames, and topics, ToMMI will allow the researcher to present the results in a variety of ways. Similarly to the types of visualizations that text topic modeling allows, ToMMI can also create graphs displaying how the topics are present in each of films in the set, or how a film is composed by different percentages of the topics. Other ways of displaying the presence of the topics—with timelines, for instance—can be chosen based on the metadata characteristics inputted with the films (year, director, etc). Most notably for this tool, each topic can be presented on its own as a direct visualization wherein the frames are shown in different sizes depending on their prevalence value within that topic (see *Figures 3 & 4*).

Figure 3: example of ToMMI output topic visualization as image spread

Figure 4: example of ToMMI output topic visualization as image collage

IV. Environmental Scan & Innovation

The innovation of ToMMI is best understood in comparison to the digital tools currently available for the study of films, particularly with regards to how ToMMI will expand on the search process, the sorting of content, and the visualization of the outcomes.

Tools for Annotation and Digital Analysis of Moving Images
CineMetrics is a tool that aids in recording data about film shot lengths, annotating these shots by type, and finally drawing statistics from this data. It has been used by scholars researching the history of film style by tracking, for instance, historical changes in editing patterns or across a filmmaker's corpus. Its database has steadily grown with the help of researchers' contributions, currently standing at over 13,000 entries. CineMetrics essentially provides a macro perspective of one of the constituent elements of a film, its shots, by presenting them as statistic data, and allows for comparative analysis at a distance.

The Digital Formalism project, a collaborative venture between the Austrian Film Museum, the University of Vienna, and the Vienna University of Technology that between 2007 and 2010 took the Vienna Vertov Collection and developed computational tools to aid in the analysis of Dziga Vertov's works. In order to make legible for computational analysis the multiplicity of semantic, compositional, and technical aspects of each specific shot, the Digital Formalism project developed algorithms for minor tasks such as detecting intertitles and depended human input for more complicated tasks such as motion tracking and image composition. Like CineMetrics, the tools developed by the Digital Formalism project depend heavily on the input of the researcher for each individual film, so this solution is only feasible with a very small number of films, and even then it can take years to complete. Depending on such processes seems unthinkable when the sets include hundreds or thousands of films.

A more automated tool, the Audio-visual Cinematic Toolbox for Interaction, Organization, and Navigation (ACTION) from the Bregman Lab at Dartmouth, which is currently in its initial phase, creates automatic analysis routines to extract raw data about the image and sound properties of the films, and then provides a workbench to study this data comparatively across films. However, the information that this tool can process is reduced to metadata about the image or sound, so compositional and semantic aspects of the moving images are ignored.

Tools for Visual (Re)presentation of Images
A simple yet popular form of film visualization is the movie barcode. These barcodes are created by taking each frame from the film and reducing it to a bar one pixel wide, which reduces the entire color distribution of that frame to its most predominant one. These individual bars are then put together to provide an overview of the color palette of the film in sequential order. However, this type of visualization is restricted to one sole characteristic, color, and then it only presents in in a reductive fashion.

The Software Studies Initiative’s project “How to Compare One Million Images?”, wherein the researchers set out to compare 1,074,790 manga pages, rightfully point out that the only way to

do so is through automated computer analysis. In essence, the project consisted of running digital image processing software on supercomputers to measure a number of visual features for each one of the images, and then use the data from this process to produce visualizations that presented the patterns shown across time and among different editions and manga series. The results from this project demonstrated how to work with large number of visual artifacts and present these results in manageable forms, but like ACTION, this project was focused on metadata primarily, not content and composition.

V. Work Plan & Evaluation

The workplan for this project consists of two main steps: a) programming of the code for each of the parts in the ToMMI process, and b) development of the interfaces for the tool and for the output visualizations. After a prototype is built, it will be tested with a "control set" of films to account for consistency of results, ease of use, and fulfillment of project goals.

1.- Programming of Code
This step will consist of two tasks: programming for Part One of the process, which will require an analyst and a programmer working under the supervision of the PI for a period of four weeks; and programming for Part Two, which will require an analyst and two programmers (ideally, the same analyst and programmer from the first task plus an auxiliary programmer) working under the supervision of the PI for a period of six weeks.

2.- Development of Interfaces
The main task in this step will be developing the interfaces for the various output visualizations: graphs, timelines, and particularly the visualizations of individual topics. This step will require a three-week work period with a graphic designer and a programmer under the supervision of the PI.

3.- Testing and Evaluation
Once a working prototype of ToMMI has been developed, it will be tested with a "control set" or sets of films. These sets would ideally be the corpus of one or more recognized film auteurs whose work has been extensively and minutely analyzed. Since the tool is designed to reveal patterns in bodies of films that have so far gone undertheorized, its testing should occur with those films that have plenty of scholarship dedicated to them in order to compare the results. The purpose is not to adapt the tool's code to replicate the expected results purported by existing scholarship, but rather to use this scholarship as a barometer of the types of patterns that the tool is (or is not) uncovering. The ideal testing audience for the prototype should be a group of students in a class on the selected auteur since they will represent an informed audience attuned to predictable and newfound patterns in the results as well as comprise a group that is unfamiliar with the tool, and could provide feedback on its ease of use. This step should take up to two weeks provided that the modifications following the tests are minimal.

VI. Final Product & Dissemination

The final product of this project will be the ToMMI application, which will be free to download from its home website. Given the large amounts of data produced and processed, the tool will not be able to have a web browser application and all the data will have to be locally hosted and stored. The website will serve as a hub for results obtained with the tool—for instance, jpegs of a topic's visualization—as well as a center for feedback and troubleshooting. The purpose of this public forum is both to encourage use and promotion of the tool and to foster innovation, as researchers who use it can voice ideas about how it could be better suited to different types of projects.

VII. Budget

NATIONAL ENDOWMENT FOR THE Humanities

Budget Form

OMB No 3136-0134
Expires 7/31/2015

Applicant Institution: *University of California, Santa Barbara*
Project Director: *Juan Llamas Rodriguez*
Project Grant Period: *09/01/2014-12/31/2014*

	Computational Details/Notes	Hourly Rate / Unit Price	Subtotal	Project Total
1. Salaries & Wages				
RA Analyst	200 hours (10 weeks x 20 hours per week)	29	$5,800	$5,800
RA Programmer	200 hours (10 weeks x 20 hours per week)	29	$5,800	$5,800
RA Auxiliary Programmer	120 hours (6 weeks x 20 hours per week)	29	$3,480	$3,480
RA Graphic Designer	60 hours (3 weeks x 20 hours per week)	15	$900	$900
RA Interface Programmer	60 hours (3 weeks x 20 hours per week)	29	$1,740	$1,740
				$17,720
5. Supplies & Materials				
Computers	2 desktops	1500	$3,000	$3,000
				$3,000
7. Other Costs				
Films to be Digitized for Trial Runs	200 films	20	$4,000	$4,000
				$4,000
8. Total Direct Costs				
			$24,720	**$24,720**
9. Total Indirect Costs				
			$4,500	**$4,500**
10. Total Project Costs			(Direct and Indirect costs for entire project)	**$29,220**
11. Project Funding		**a. Requested from NEH**		$29,220
				$0
				$29,220
		b. Cost Sharing		$0
				$0
12. Total Project Funding				**$29,220**