PhotoGlas

Ramesh Srinivasan
MIT Media Laboratory
20 Ames St., Room E15 – 020d
Cambridge, MA 02139, USA
+1 – 617 – 253 – 5108
ramesh@media.mit.edu

Roderick Ferguson
MIT Media Laboratory
20 Ames St., Room E15 – 020d
Cambridge, MA 02139, USA
+1 – 617 – 253 – 5108
roderick@mit.edu

Befekadu Ayenew
MIT Media Laboratory
20 Ames St., Room E15 – 020d
Cambridge, MA 02139, USA
+1 – 617 – 253 – 5108
befekadu@mit.edu

ABSTRACT
In this paper we describe *PhotoGlas*, a visual browser for news. *PhotoGlas* uses a web crawler to retrieve photographs from a number of pre-determined international-news web sites. The system then categorizes the images it retrieves based on keywords extracted from the article in which they are found. The images are presented as a montage that reflects a news topic. We describe the process by which categorization occurs and explore the potential of this system to present meaningful juxtapositions of news events to users.

Keywords
Montage, juxtaposition, visual browser, semiology, visual model building, photograph, news, categorization

1. INTRODUCTION
Writing in 1964, Barthes declared that "semiology aims to take in any system of signs, whatever their substance and limits; photographs, gestures, musical sounds, objects, and the complex associations of all of these, which form the content of ritual, convention or public entertainment: these constitute, if not *languages*, at least systems of signification."[1] For Barthes, language was cross-functional. In other words, it is determined by a set of complex connections between objects and other forms of expression, and that correspondingly meaning could be decoded by a system of connections that respond and interact with forms that are independent entities themselves.

This idea is confirmed by semiology, which states that codification is the formation of organized connections or relations with other forms by which matter or energy are organized. Semiology, then, is a relational process of codification by which networks of relationships develop to transform energy into concrete information. While the domain of semiology is large, we focus on visual symbols. First, we introduce Lindley's hierarchy of codification for visual phenomena[2]:

perceptual level: the level at which visual phenomena become perceptually meaningful;

diegetic level: the means by which these phenomena are represented in two-dimensional space;

cinematic level: image compositions and their placement into an expressive artifact space;

connotative level: the level of metaphorical, analogical and associative meanings objects may have;

subtextual level: the specialized meanings of signifiers.

We believe that a key to unlocking the semiology of the Web is the development of systems that begin to represent symbols on both a diegetic and cinematic level based on the connotative and subtextual criteria of the information context around the photograph. This is particularly important because the diegetic role of an individual photograph as an expressive form is maintained along with its enhancement as a bridge to the information and concepts behind it (connotative). This can catalyze the development of systems that can start to represent the semantics, personalities, and meanings behind virtual events.

With the expansion of information available on the Web, there has not been a corresponding development of a structure by which the common ideas and topics among them can be effectively integrated. Specifically, in the area of news, while the presence of grassroots and international media has escalated on the Web, a thorough browsing of any single topic remains relatively difficult. There is no single place where one can find an exhaustive representation of all the perspectives, side-stories, and information related to a particular news topic. For example, while the Palestinian perspective toward recent confrontations in the Middle East can be found at the Electronic Intifada web site[3], a contrasting Israeli perspective can be found at the Israeli Defense Force's web site[4]. We hypothesize, however, that a juxtaposition of the two articles could serve as a symbol that tells a more complete story.

The belief that such juxtapositions are useful is based on the arguments made by theorists such as Manovich, in his "Languages of New Media." Manovich argues that such uses of symbolism will drive the media of the future. He describes this movement as compositing:

The computer era introduces a different paradigm: This paradigm is concerned not only with time but space. . .Thus, in order to qualify as an example of montage, a new media object should fulfill two conditions: the juxtapositions of elements should follow a particular system; and these juxtapositions should play a key role in how the work establishes its meaning. . .The borders between different worlds do not have to be erased, the individual layers can retain their separate identity rather than being merged into a single space; the different worlds can clash semantically rather than form a single universe.[5]

Manovich uses the terms juxtaposition and montage. We invoke these forms as appropriate symbols for representing news on the web. Our definitions of these terms are as follows:

juxtaposition: positioning objects close together;

montage: an image made by the juxtaposition of photographs.[6]

We refer to juxtaposition as a description of form, and refer to montage as the process of combining photographs. We believe that the application of montage to the photographs found on such disparate news web sites as the Electronic Intifada and the Israeli Defense Force can serve as an engaging and thought-provoking introduction to complex events in the news. Jean Luc Godard, the new-wave filmmaker, referred to the world as a "vague and complicated system that we are all continually entering and watching."[7] Montage would integrate space, through its juxtaposition of photographs that were previously separated from one another, and time, by representing these juxtapositions dynamically, as the photographs that constitute them change with the flow of news.

PhotoGlas, a web browser for news photographs, is our first attempt at building such a system. Our goal in this project is to categorize photographs retrieved in real-time from a multitude of news-related web sites and in turn present these photographs in meaningful juxtapositions for users.

2. VISUAL MODEL BUILDING

An important subgoal of the *PhotoGlas* project is to create a tool for encouraging critical thinking and reflection among its users. We believe this can happen through the process of visual model building. (Smith defines visual model building as the use of photography to construct a theory related to the information it conveys.[8])

Smith's *Animal Landlord* project tests the ability of nature films to engage students in the development of reflective models that qualitatively describe the behaviors they observe. The results were contrasted with text-based (textbook) traditional learning methods. The experimenter concludes:

Visual events are rich with opportunities for students to pose questions and reflect on behaviors and processes. . . As a result, imagery establishes a context for problem solving, for generalizing explanations from pictorial evidence [8]

Smith describes the visual model building process in four major steps: (1) decomposition: identification of actors and in general central personalities/subcomponents of the presented photograph; (2) comparison: looking for similarities and differences between any given photograph and the others present in the film (or in our own conceptions/visual models); (3) linkage conjecture: analyses (based on previous knowledge and the appearance of the photographs) of the factors behind the choice and relative distances of the photographs displayed; and (4) model construction: generalization of causal models that explain the numerous conditions (time, setting, relationship between actors, etc.) that have made the photographs a reality.

Visual model building has motivated the design of *PhotoGlas*. We hypothesize that providing users with a juxtaposition of topically-related photographs will engage them in processes similar to those described by Smith. Through montage, users make comparisons between the setting, actors, emotions, and actions. These comparisons are used to discover links and common elements. It is from this perspective that Smith's final step, model construction, can occur. The commonalities and discrepancies that a user has perceived among the photographs lead the user to construct a story that captures the entirety of a news event.

3. PRIOR WORK

There have been other significant efforts in the use of montage for interactive storytelling that inform this effort. Murtaugh's *ConTour* system[9] interactively suggests movies and pictures based on a user's choices of how the story should proceed. *ConTour* provides the user with the power to make these decisions by suggesting certain keywords based on the user's previous choices. These keywords are in turn linked to the movies and pictures from the system's database. *ConTour* is able to categorize its media objects with keywords through an analysis of what the context behind these objects, while still empowering the user to determine the story received. Tsarkova's *North End Chronicles* project[10] (based on Boston's North End neighborhood) generates a number of potential story playouts using movie clips, then lets the user decide between three playouts that the system has decided are best for the story the user has experienced thus far. These playouts are juxtaposed with the movie clip currently playing to inform the user of the related branches his or her experience could follow. Together, this montage presents a picture of the user's immediate story environment.

4. CATEGORIZATION

We have constructed a simple web crawler to periodically check a number of pre-determined international-news web sites. Our crawler integrates, for example, the photographs it retrieves from a range of web sites that include *The People's Daily News* from Beijing, China, *The BBC News*, *The Moscow Daily News*, among others. We believe that combining these international news sources provides *PhotoGlas* with a more inclusive base from which it can depict news. We decided to build our own web crawler rather than use one publicly available because we wanted to optimize it for the domain of news websites, which often exhibit common patterns such as the format by which they present information. We constructed our crawler to account for the differences and intricacies between specific sites as well. For example, the crawler can recognize a photograph and its caption as existing in separate tables on the BBC site, while expecting them to be placed together on *The People's Daily News* site.

The web crawler is programmed to capture the associated caption and headline of an article in which a photograph is contained, along with the photograph itself. It is from this text that

PhotoGlas is able to extract keywords that are used to categorize a given photograph.

We make two refinements to the keyword list before initiating the categorization process. First, we have implemented a stoplist. (Stoplists contain words that are structural and inherently provide little meaning to a given text.) We are using a pre-defined stoplist that was statistically generated from a *Time Magazine* corpus to filter our keywords[11]. We have taken this abbreviated set of keywords and applied a simple Porter stemmer[11] to convert words with the same morphological form into the same "stem." For example, "aviator," "aviating," and "aviation" can be converted into "aviat."[12] These two steps are illustrated in Figure 1.

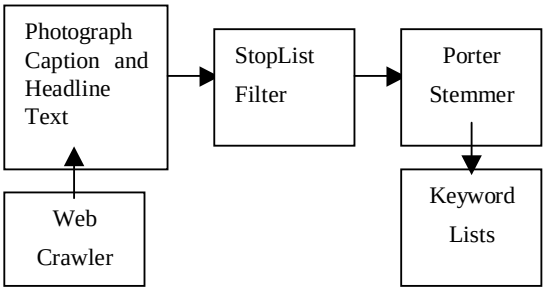

Figure 1. Flowchart detailing pre-categorization requirements

Having isolated relevant information associated with a given photograph, we then classify each photograph into a specific topic. The topics we have chosen are listed in Figure 2.

Topics
War
Terrorism
Poverty
Medicine
Business/Industry
Justice/Police
Entertainment/Media
Nationalism/Patriotism
Technology
Women/Feminism
Religion
Politics
Government/Leadership
Discrimination/Inequality
Nature/Environment

Figure 2. Topics chosen for categorization

To do the actual categorization, we pre-assembled a list of keywords associated with each topic. This was done using WordNet®[13], an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. For each of our topics we have assembled a set of WordNet-derived synonym and descriptive keywords.

To make the assignments between photograph text-derived keywords and topic-derived keywords, we implemented a simple matching algorithm.

For each photograph i in the database, let K_i be the set of keywords associated with the photograph. For each topic j, let T_j be the list of keywords associated with the topic. *PhotoGlas* calculates the correspondence between each photograph and topic using the ratio, R_{ij}.

$R_{ij} = (K_i \cap T_j)/T_j$, where $K_i \cap T_j$ refers to matching keywords found in the intersection of the two sets.

The correlation between a photograph and each of the 16 topics is expressed through an array of these ratios. We sort these arrays based on descending ratio values. The greatest ratio determines the topic to which a photograph is assigned. In the case of a tie, we assign the photograph to multiple topics.

We were interested in expressing relationships between topics as well. With this motivation, we asked a number of prospective users to try to express which topics listed in Figure 2 were most similar to each other. Our matrix of inter-topic similarities, detailed in Figure 3, is a result of this study.

Topic	**Related Topic**
War	Nationalism, Religion, Terrorism
Terrorism	Nationalism, Religion, Justice/Police, War
Poverty	Discrimination/Inequality, Nature/Environment, Medicine
Medicine	Technology, Business/Industry, Nature/Environment
Business/Industry	Technology, Justice/Police, Medicine, Nature/Environment
Justice/Police	Business/Industry, Discrimination/Inequality, Politics, Entertainment/Media
Entertainment/Media	Business/Industry, Technology, Politics
Nationalism/Patriotism	Government/Leadership, Politics, Discrimination/Inequality, War
Technology	Business/Industry, Medicine, Entertainment/Media
Women/Feminism	Discrimination/Inequality, Nationalism/Patriotism, Medicine
Religion	Nationalism/Patriotism, War, Politics, Terrorism
Politics	Government/Leadership, Justice/Police, Discrimination/Inequality, War
Government/Leadership	Nationalism/Patriotism, Justice/Police, Politics, Discrimination/Inequality

Discrimination/Inequality	Justice/Police, Politics, Government/Leadership, Women/Feminism, Poverty
Nature/Environment	Business/Industry, War, Technology, Politics

Figure 3. Inter-topic similarity ontology

5. INTERFACE

The *PhotoGlas* interface presents four topically-related spheres, a central one orbited by three satellites. Each sphere contains an identifying label for its topic. Photographs within each topic cluster are presented as a projection across the surface of a sphere. The shape of a sphere was chosen for its globe-like form. (We hope to remind users that when they are browsing in *PhotoGlas*, they are interacting with the world globally.) Selecting an individual photograph on any of the spheres takes the user to the web page from which that photograph was extracted. An example of the interface is displayed in Figure 4.

Figure 4. A screenshot of the *PhotoGlas* interface

The central sphere reflects the focus topic chosen by the user while the three spheres in the periphery reflect topics the system has identified as related to the focus. This identification is based on a hand-crafted ontology of topic similarities that we have created. The interface does constrain this ontology, as it allows a maximum of three satellite spheres to be displayed at once. We are investigating alternate interface forms that can better capture inter-topic relationships.

Selecting a satellite sphere causes that sphere to move to the center and for it to be surrounded by new satellites. This provides a means for the users to navigate the topic space and refine their browsing interests. For example, if the topic "war" shows photographs that do not quite reflect a user's interests but is still related to what he or she has conceptualized, the peripheral topics of "nationalism" or "terrorism" might suffice.

6. IMPROVEMENTS

What we have accomplished thus far only reflects our initial efforts to build an effective visual browser. There are a number of ways in which we plan to improve *PhotoGlas*:

- The assignment of photographs to "relevant" topics is fundamental to *PhotoGlas*. It is important that the topic areas we choose are indeed what users find most relevant, and that the photographs assigned are reflections of the topic. There has been significant research done on news ontologies and what users globally find newsworthy. We hope to define topics based on these results rather than just our intuitions.
- The decision to assign a photograph to a single topic merely because its ratio was highest may be shortsighted. We suspect that a photograph can potentially evoke a number of different topics and perspectives. We will to investigate the idea of linking photographs retrieved not only to multiple categories but also perhaps directly to each other. This could be achieved by using matches between the keywords themselves.
- Instead of creating groupings solely based around topics, groups can also derive from a specific keyword from a particular photograph. The choice of keyword could then generate another group of photographs, creating a juxtaposition of photographs that can represent a particular news keyword.
- We would like to develop a more rigorous means of connecting the components of our topical ontology. We plan to design an experiment to reveal which topics are indeed "closer" to others.
- Finally, we intend to investigate different interface forms that may provide a better sense of the context of a photograph. It may be that a static juxtaposition may not be an ideal form for the expression of context.

7. EVALUATION

We have articulated our goal of constructing a tool to present users with meaningful juxtapositions of news photographs. We also stated a subgoal was to encourage critical thinking and reflection among users.

In *PhotoGlas*, we have created a technology platform that allows us to experiment news montage. The question that we hope to focus on is which juxtapositions are actually meaningful to users? This project allows us to begin our search for the answer to this question.

In the spirit of Smith's Animal Landlord project, we believe that *PhotoGlas* can serve as an appropriate environment from which the four steps of decomposition, comparison, linkage conjecture,

and model construction can be taken. We intend to examine whether or not (and if so how) users go through the visual model-building process when interacting with *PhotoGlas*. Specifically, we hope to investigate whether or not Smith's theory of learning through reflecting can apply to a situation where there is no explicit task at hand.

We are also interested in determining what factors make a photograph evocative to its viewers, particularly when it is placed in a montage. We plan to address the problem of how to train *PhotoGlas* to select photographs that users find powerful.

Finally, *PhotoGlas* may benefit from incorporating advancements in the field of photograph recognition. The ability to capture certain descriptive aspects of a picture (such as the identity of people or a physical location) or aspects of its composition may further *PhotoGlas*'s goal of providing users with a more powerful combinations of photographs. This raises the possibility of enabling computers to associate a topic with a photograph without basing this association solely on the text accompanying the photograph.

8. ACKNOWLEDGMENTS

We would like to thank Walter Bender, for his guidance and active assistance on this project. This work is sponsored by the MIT Media Lab's Information: Organized Research Consortium and *eircom.*

9. REFERENCES

[1] Barthes, Roland, *Elements of Semiology* (trans. Annette Lavers & Colin Smith). London: Jonathan Cape ([1964] 1967).

[2] Lindley, Craig, "*A Computational Semiotic Framework for Interactive Cinematic Virtual Worlds*": http://wheelie.tees.ac.uk/users/p.c.fencott/

[3] Electronic Intifada site: http://electronicintifada.net/new.html

[4] Israeli Defense Fund site: http://www.idf.il/english/news/main.stm

[5] Manovich, Lev, *Languages of New Media*, MIT Press, March 2001, p.147

[6] Word Dict site: http://www.dict.org

[7] Godard, Jean-Luc, *Son + Image*, edited by Raymond Bellour (New York: Museum of Modern Art, 1992) p. 171

[8] Smith, Brian K. & Blankinship, Erik, *Imagery as data: Structures for visual model building* In Proceedings of Computer Support for Collaborative Learning 99, p.549-557

[9] Murtaugh, Michael L, *The Automatist Storytelling System* (MIT Media Lab, 1994)

[10] Tsarkova, Natalia, *Dynamic Collage: An Interface for Digital Stories* (MIT Media Lab, 1997)

[11] Marlow, Cameron A, *A Language-Based Approach to Categorical Analysis* (MIT Media Lab, 2001)

[12] Porter, M.F, An algorithm for suffix stripping. *Program 14*(3), (1980), p.130-137

[13] WordNet site: http://www.cogsci.princeton.edu/~wn/