2013 IEEE International Symposium on Multimedia

MultiFacet: A Faceted Interface for Browsing Large Multimedia Collections

Michael J. Henry, Shawn Hampton, Alex Endert, Ian Roberts, and Deborah Payne
Pacific Northwest National Laboratory
Richland, WA, USA
michael.j.henry@pnnl.gov

***Abstract*— Faceted browsing is a common technique for exploring collections where the data can be grouped into a number of pre-defined categories, most often generated from textual metadata. Historically, faceted browsing has been applied to a single data type such as text or image data. However, typical collections contain multiple data types, such as information from web pages that contain text, images, and video. Additionally, when browsing a collection of images and video, facets are often created based on the metadata which may be incomplete, inaccurate, or missing altogether instead of the actual visual content contained within those images and video. In this work we address these limitations by presenting MultiFacet, a faceted browsing interface that supports multiple data types. MultiFacet constructs facets for images and video in a collection from the visual content using computer vision techniques. These visual facets can then be browsed in conjunction with text facets within a single interface to reveal relationships and phenomena within multimedia collections. Additionally, we present a use case based on real-world data, demonstrating the utility of this approach towards browsing a large multimedia data collection.**

Keywords Visual Analytics, content-based image retrieval, image classification, faceted search, visualization, multimedia

I. Introduction

Analysts are often tasked with exploring a collection in search of a particular item of interest or a set of items that all match the given criteria. The criteria may not be known ahead of time; rather, users may benefit from having a system that can reveal the characteristics within the collection and allow them to uncover items of interest. That exploration process can be facilitated by a tool that enables analysts to filter the collection based on known criteria such as metadata and characteristics that can be discovered computationally such as visual similarity and classification of images and video.

To help analysts explore large, complex collections, systems can employ facets to navigate classifications. Within the context of search, facets are combinations of dimensions that align with the data, enabling filtering. These facets facilitate navigation and exploration of large collections of data, reducing the set of interesting items down to a manageable set, while allowing analysts to choose the pertinent relationships among the data in the collection.

Facets can be used for searching and browsing large text collections and have also been used for navigating large image collections [8, 15]. While each data modality can be explored on its own, a great deal of data contains a mix of multiple modalities—for example, a PDF file may contain both text and images. Providing a browsing interface for just one data type prevents an analyst from making connections across the multiple types present in a mixed-media collection.

Searching image collections typically falls under the category of content-based image retrieval, where low-level image features (for example, SIFT [13]) are used to determine similarity between images. Other image retrieval systems have also used facets to group images, but the facets in those systems were built using the image metadata [22]—which may be inaccurate, incomplete, or missing altogether—rather than the visual features of the images in the collection.

We have developed a faceted search system, MultiFacet, to extend existing faceted browsing systems. MultiFacet connects to an existing multimedia analytics tool, Canopy [5], that has been designed to facilitate the analysis of multimedia. We leveraged Canopy's data analysis framework to provide an interface for analysts navigating large collections of multimedia using facets built from the text content of the data, the low-level features computed from the image data, and the metadata associated with the text and images.

This paper makes the following two contributions:

1. A faceted search system that integrates facets from text, image, and video data
2. Image facets that are built based on the low-level visual features of the images.

The rest of the paper is organized as follows: Section 2 gives an overview of related research in this area, Section 3 provides a description of the technology presented in this paper, Section 4 describes the use cases to which this system can be applied, and Section 5 offers a discussion of future work that can be built off what has been presented here.

II. Related Research

Faceted browsing allows analysts to narrow a collection of information based on selections from a preset list of criteria. Faceted searching is a technique that can be applied to a number of different domains, including text, image, and video [9, 10, 16, 12].

Facets are often used to present a means for filtering or pivoting collections. For example, Shi et al. show how text data sets can be browsed using facets based on recognition of patterns and aggregations of raw extracted features [21]. Similarly, Rose et al. show how themes can be aggregated from similar extracted features across text data sets [18]. Techniques can be used to develop facets for image data sets, such as in Yee et al. [23] where researchers built a faceted search system for an image collection with facets constructed

978-0-7695-5140-1/13 $26.00 © 2013 IEEE
DOI 10.1109/ISM.2013.66

from text that accompanied each image in the collection. As such, facets represent an aggregation of extracted features or dimensions of data. Depending on the data and features extracted, facets may represent easily understood criteria by which to browse or filter the data. For example, the raw dimensions extracted from text documents can include unique keywords. Facets that describe a text corpus represent themes of topics within the data set and are created using an aggregate of the raw dimensions [18].

Previous work has analyzed the way users form themes and groups from information. For example, when asked to group text documents, users often create groups that are described and referenced at a higher-level semantic concept than the raw features [7, 1]. That is, user-generated groups of documents may more accurately describe facets than dimensions or features. Similarly, groups of multimedia information may also benefit from a high-level semantic understanding [21].

Given the scale of modern image collections, much research has been conducted in presenting images to analysts in a way that facilitates discovery. Rodden et al. [17] showed that users prefer interfaces that group images by similarity, as opposed to a random layout. They also showed that grouping images based on associated text captions was helpful. Other techniques have been explored, including iterative supervised grouping [8], projecting images from a high-dimensional feature space down to a 2D plane [19, 22, 17, 20, 14] or hierarchical clustering mapped into a space-filling curve [6].

In Villa et al. [24], researchers showed that users prefer browsing large image collections using facets to a conventional method based on tabbed browsing through web interfaces. Yee et al. [23] found that a categorical-based faceted search for images was more appealing to users than a keyword-based method.

Some research has been conducted in using visual features to construct facets when browsing an image collection. Müller [15] explored combining both semantic (text) features and visual features within a search tool. However, the visual features were basic, such as most frequent color or image coarseness, and not based on semantic categories. Bartolini and Ciaccia [2] built a system that clusters images based on visual similarity and allows users to interact with the cluster dimensions while browsing the photo collection.

III. Technology Description

MultiFacet relies on an underlying infrastructure that enables associations among elements of multiple types. Collections contain documents: files that are uploaded from a hard drive or network share [5]. Each document could be a simple document or a compound document. Simple documents are plain text, images, and frames extracted from video. Compound documents comprise a set of base objects. For example, Word and PDF documents may contain the base objects of images and text; videos contain multiple shots. We call the base type objects elements whether they come from simple documents or are derived from compound documents. We call the original compound documents “parent” elements and the text, images, and shots extracted from the parents “child” elements. Child elements such as images and text that are extracted from the same parent document are called “siblings.”

During extraction, metadata for all documents is collected and stored. Text data is analyzed with a text engine, while image and video data is analyzed with a content analysis engine. The content analysis engine process leverages the bag-of-features model for image content analysis and uses features such as color, SIFT [13], Dense SIFT [4], and Pyramid Histogram of Orientation Gradients [3]. When the content analysis engine compares two images in a collection, the features derived from these processes are combined in a method based on the work of Arevalilo-Herráez et al. [25].

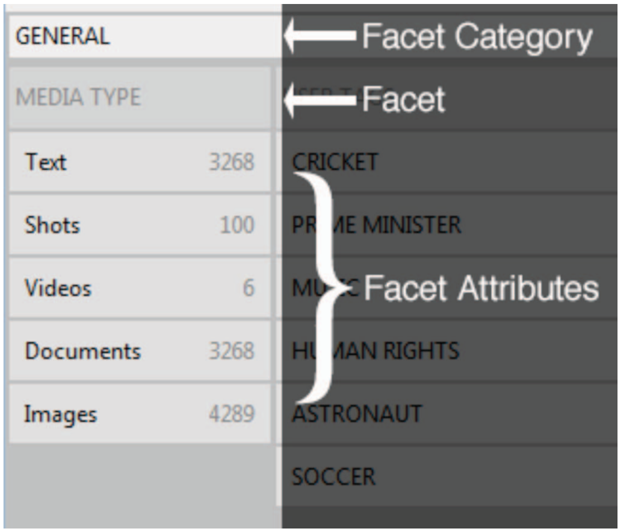


Fig. 1.Three-Level Hierarchy

Videos in the collection are analyzed by first determining shot breaks by computing per-frame similarities in the video. Key frames are extracted from each shot, and image features are then computed for each shot.

Instead of providing a flat interface to multiple facets as seen in previous work, MultiFacet is implemented with a three-level hierarchy: facet, facet category, and facet attribute (Figure 1).

A *facet* represents a single aspect or feature of an object and typically has multiple possible values e.g. author, location, and color. Facets are visualized in MultiFacet by a single column of values, each value being defined as a facet attribute. *Facet attributes* represent a set of unique values for the elements. The facets themselves are grouped into a *facet category*. MultiFacet has categories for both metadata and content-based image retrieval characteristics. Examples of metadata facets include author and file type.

In MultiFacet, the content from images and shots is used to construct several facets: Image Cluster, Color, and Classifiers. The facet attributes for the Image Cluster facet are the resulting clusters found by performing k-means on the collection of images and shots, facilitating the exploration of a collection when the content of that collection is not known ahead of time. The Color facet contains attributes based on the dominant color of each image and shot. The Classifiers facet contains attributes constructed from pre-defined image classifiers trained offline. Classifiers can be trained to detect a number of different objects in images and shots, such as cars, trains, and donuts. For the use case presented in Section 4, a support vector machine classifier was created to detect apples, Apple Inc. logos, Apple computers, and the

Australian flag. Facets built on object classifiers are helpful when analysts know ahead of time what type of information they are analyzing and can configure MultiFacet to use prebuilt facets that specifically target an object of interest.

IV. USE CASE

In the presented use case, an analyst is interested in apple recipes and will be exploring the collection to discover which elements contain information about apple recipes.

A. *Data and Display*

For the use case presented in this paper, a project was constructed that consisted of data gathered from the web such as news articles and Wikipedia entries as well as documents gathered from a hard drive. 3,298 files were uploaded and extracted into 10,931 base elements. The breakdown of the elements is as follows:

- 3,268 text
- 4,289 images (jpg, gif, etc.)
- 3,268 compound documents (Word, PDF, etc.)
- 6 videos (mov, rm, qt, etc.)
- 100 shots

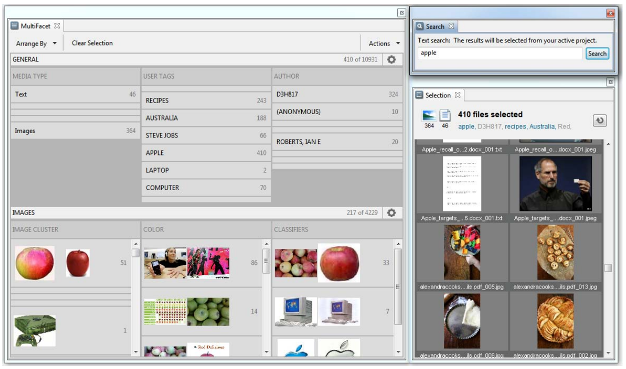
In MultiFacet, the facets are grouped into configurable categories in the display and divided into two rows. The top row contains facet categories based on the metadata of the multimedia elements. Figure 2 shows the categories of interest for this use case: *Media Type*, *User Tags*, and *Author*. The bottom row contains facet categories built from the visual content of the images and shots from videos in the collections. Again in Figure 2, the visual content categories of interest are *Image Clusters*, *Color*, and *Classifiers*. MultiFacet incorporates both metadata facets based on text information and visual facets based on content analysis of images and video.

B. *Use Case Walkthrough*

To start, the analyst creates a collection using the data described in Section 3.1. He suspects that this collection contains recipes. Using a text search, the analyst is able to find all text elements containing the word *apple* and selects each of those elements. He asks the system to select the image elements that co-occur within the parents of the text elements. The result is a selection of images found to co-occur in documents that contain the word *apple*. Figure 2 shows the result of these operations; the text search in the top right corner, the result of the request for siblings shown in the selection pane on the bottom right and in the MultiFacet pane on the left. Next, the analyst uses an additional feature to select images that are visually related to the images found with the text search of *apple* (Figure 3).

Because the analyst is interested in exploring the unknown pieces of the data set, new creative techniques are required to explore the data further. This type of guided exploration is ideally suited to faceted browsing, and MultiFacet can now be used to provide greater insight into the multimedia collection.

Given this current selection, the analyst is able to use MultiFacet and explore the data. The analyst is interested in apples, so he selects the Apple facet from the "Classifiers" category (Figure 4), which brings up a list of all images in the current selection that were identified as apples by the image classifier. Examining the facets on the top row, the analyst discovers that two elements labeled "apple" are keyframes of



Text search: The results will be selected from your active project.

apple

Fig. 2. Images that co-occur with documents that contain the word *apple*

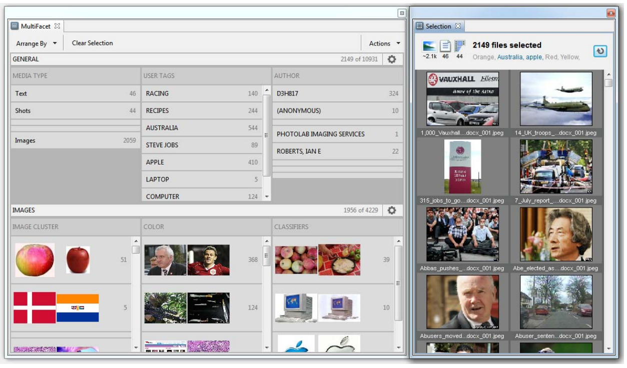


Fig. 3. Images visually related to the images in Figure 2

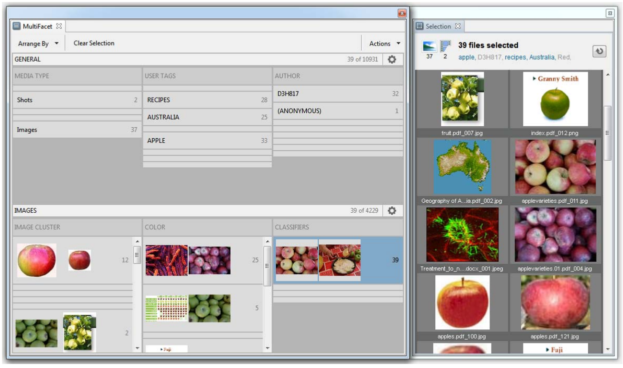


Fig. 4. Facet intersection with the images detected by the apple classifier

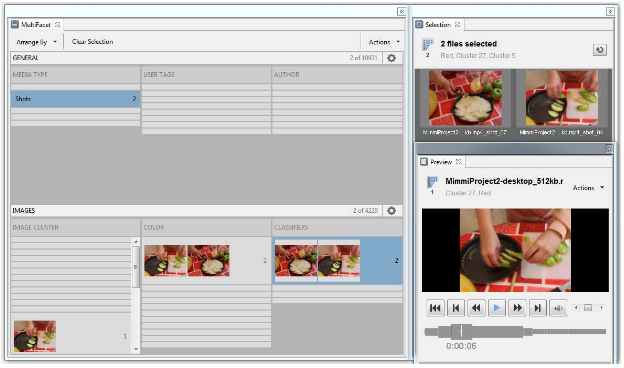


Fig. 5. Union of the shots facet with the apple classifier facet

video shots. Selecting the "Shots" facet (Figure 5), the analyst discovers a video containing a description about how to make an apple pie.

While the analyst in this case had a very specific target in mind—apples—the actual content of the multimedia collection was not known prior to exploration. Sometimes there may not be metadata available to make the connection. For instance, there is no metadata field available for making the connection that a picture of an apple and the text *apple* refer to the same thing. That connection was enabled by the parent-child-sibling relationship established in MultiFacet. The connection between the pictures of apples that were accompanied by text containing the word *apple* and those images that were not was enabled by computing the visual similarity between the two sets of images. The apple classifier facet was applied to the data reducing the set down to images of interest based on the visual content. Using these techniques, the MultiFacet view provides a means of bringing the data together in a meaningful way, allowing analysts to exploit the advantages provided by having both metadata facets calculated from text associated with the various media types and content-based facets calculated from visual similarity for images and videos within a single tool.

V. Discussion

In this work, we presented an interface, MultiFacet, for browsing a large multimedia collection. This paper made the following two contributions:

1. A faceted search system that integrates facets from text, image, and video data
2. Image facets that are built based on the low-level visual features of the images, rather than the image metadata.

We also presented a use case that illustrates the end user benefit gained from MultiFacet, demonstrating how the interface can facilitate connections across multiple media types when exploring a large multimedia data set.

MultiFacet facilitates insight into multimedia information independent of media-type. If these types were not combined a user would have to analyze the data with multiple tools and would have to make the cognitive connection of the relationships among the data independently. By presenting analytics built for each data type within the same view, the tool is able to assist in bridging those cognitive connections, both by presenting an interface where the analyst can create the connections themselves, but also by discovering those relationships and presenting them to the user, as shown in the use case.

Additionally, we have presented a use case based on real-world data to illustrate the utility of MultiFacet in providing insight when exploring an unfamiliar data set. Future work should involve the design of a user study to better quantify the efficacy of the MultiFacet interface.

Acknowledgment

This work was supported by a government client and performed at the Pacific Northwest National Laboratory (PNNL). PNNL is managed for the US Department of Energy by Battelle under Contract DE-AC05-76RL01830.

References

[1] C. Andrews, A. Endert, and C. North. Space to think: Large, high-resolution displays for sensemaking. In Computer-Human Interaction, 2010.

[2] I. Bartolini and P. Ciaccia. Integrating semantic and visual facets for browsing digital photo collections. In 17th Italian Symposium on Advanced Database Systems, 2009.

[3] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spa- tial pyramid kernel. In Proceedings of the 6th ACM International Con- ference on Image and Video Retrieval, 2007.

[4] A. Bosch, A. Zisserman, and X. Munoz. Image classification using ran- dom forests and ferns. In Proceedings of the 11th International Conference on Computer Vision, 207.

[5] R. Burtner IV, S. Bohn, and D. A. Payne. Interactive visual comparison of multimedia data through type-specific views. In Visualization and Data Analysis, 2013.

[6] S.E.Dillard, M.J.Henry, S.Bohn, and L.J. Gosink.Coherent imagel ayout using an adaptive visual vocabulary. In Electronic Imaging: Machine Vision and Applications, 2013.

[7] A. Endert, S. Fox, D. Maiti, C. Leman, and C. North. The semantics of clustering: Analysis of user-generated spatializations of text documents. In AVI, 2012.

[8] A. Gilbert and R. Bowden. igroup: Weakly supervised image and video grouping. In IEEE International Conference on Computer Vision, 2011.

[9] M. Hearst. Information visualization for peer patent examination. UC Berkeley.

[10] J.Kramer-Smyth, M.Nishigaki, and T.Anglade.http://archivesz.com/.

[11] B. Kules and R. Capra. Mapping the design space of faceted search interfaces. In Fifth Annual Workshop on Human-Computer Interaction and Information Retrieval, 2011.

[12] D. Laniado, D. Eynard, and M. Colombetti. Using wordnet to turn a folksonomy into a hierarchy of concepts. In Semantic Web Applications and Perspectives, 2007.

[13] D. G. Lowe. "distinctive image features from scale-invariant keypoints. In International Journal of Computer Vision, 2004.

[14] B. Moghaddam, Q. Tian, N. Lesh, C. Shen, and T. Huang. Visualization and user-modeling for browsing personal photo libraries. International Journal of Computer Vision, 56(1), 2004.

[15] W. Muller. Visualflamenco: Faceted browsing for visual features. In Multimedia Workshops, 2007.

[16] P. Pirolli and S. K. Card. Information foraging. Technical report, UIR, 1999.

[17] K.Rodden, W.Basalaj, D.Sinclair, and K.Wood. Evaluating a visualization of image similarity as a tool for image browsing. In IEEE Symposium on Information Visualization, 1999.

[18] S. Rose, I. Roberts, and N. Cramer. Facets for discovery and exploration in text collections. In Workshop on Interactive Visual Text Analytics for Decision Making, 2011.

[19] Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. In Sixth International Conference on Computer Vision, 1998.

[20] G. Schaefer and S. Ruszala. Image database navigation on a hierarchical mds grid. In Pattern Recognition, 2006.

[21] L.Shi, F.Wei, S.X.Liu, L.Tan, and X.Lian. Understanding text corpora with multiple facets. In Visual Analytics Science and Technology (VAST), 2010.

[22] J.Yang, J.Fan, D.Hubball, Y.Gao, H.Luo, W.Ribarsky, and M.Ward. Semantic image browser: Bridging information visualization with auto- mated intelligent image analysis. In IEEE Symposium on Visual Analytics Science and Technology, 2006.

[23] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In Computer-Human Interaction, 2003.

[24] Villa, R., Gildea, N. and Jose, J.M. A faceted interface for multimedia search Proceedings of the 31st International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, Singapore, Singapore, 2008

[25] Arevalilo-Herráez, Domingo, Ferri, Combining Similarity Measures in Content-Based Image Retrieval Pattern Recognition Letters 2008