Current Literature Review of Approaches to Image Indexing

Leslie A. Walbolt

LIS6711 – Organization of Knowledge 1

University of South Florida

Current Literature Review of Approaches to Image Indexing

Introduction

Cleveland and Cleveland (2013, p. 224) end their chapter on image indexing by looking to the future given the complicated process of indexing and retrieving images. The authors recognize that the exponential increase in the number of digital images, joined with the semantic ambiguities of the language required to describe the images, creates an automation problem that computer scientists are working to overcome (Cleveland & Cleveland, 2013, p. 225). At the same time, human indexers work towards creating thesauri and indexes to help online users discern the search attributes needed for access to information (Cleveland & Cleveland, 2013, p. 225). Six years after the textbook's most recent publication, researchers in the field of image indexing continue to struggle with similar access issues. This paper intends to pursue a literature review of current approaches to image indexing with a particular eye toward the art world.

Shatford/Panofsky Model and Image Indexing

The Shatford/Panofsky model serves as the framework for understanding the subject in image indexing (Stewart, 2015, p. 8). In 1939, Panofsky created his three-level model as a way of interpreting the subject matter or meaning of Renaissance works of art (Dam Christensen, 2017, p. 1782; Rafferty, 2019, p. 7). The three levels are distinguished as "pre-iconographical description (objects or events), iconographical analysis (themes), and iconographical interpretation (meaning)" (Stewart, 2015, p. 8).

Shatford Layne modifies Panofsky's model to generalize classification to all images for indexing (Dam Christensen, 2017, p. 1782; Rafferty, 2019, p. 7; Stewart, 2015, p. 8). Shatford Layne's levels consist of primary ("generic of"), secondary ("specific of"), and intrinsic ("about"), and roughly correlate to Panofsky's pre-iconographic, iconographic, and iconological

levels (Rafferty, 2019, p. 7). These levels, when combined with four subject facets – who, what, where, and when, create a matrix for understanding the meaning of images (Rafferty, 2019, p. 7; Stewart, 2015, p. 9).

New Approaches Using the Shatford/Panofsky Model in Image Indexing

Other researchers further adapted the Shatford/Panofsky model from the pre-digital to the modern world (Rafferty, 2019, p. 8). Rafferty argues that these image indexing approaches highlight the intricacies of finding the meaning of a subject without textual information; this issue occurs whether the image is analog or digital (2019, p. 8). Huang, Soergel, and Klavans (2015, p. 1616-17) formulated a method for integrating topical relevance with the complexities of (art) image-text relationships.

Huang et al.'s (2015, p. 1617-18) framework is based on three layers: (1) visual features of the image, (2) meaning and interpretation of the image (image subject), and (3) function of the image. These three layers form Facet 1; Facet 2 occurs at the Layer level where the meaning of the image can be discerned using two different perspectives (one from the artist's intent and the other from the viewer's response) (Huang et al., 2015, p. 1617). Facet 3 focuses on the relationships in how the "[i]mages relate to text, [] tags, or information needs" (Huang et al., 2015, p. 1617). Table 1, located in Huang et al.'s (2015, p. 1617) article, more easily describes the comprehensive matrix with Facet 1, Layer 1 being best for content-based image retrieval, Facet 1, Layer 2 being an extension of Panofsky's three strata, and Facet 1, Layer 3 describing the effect on the viewer and the image's use. Its result provides improved image tagging and semantic understanding of images (Huang et al. 2015, p. 1617). Huang et al.'s (2015, p. 1637) focus on topical relationship between an image and its tags may generate more precise search queries.

Huang et al.'s framework is not without criticism; Dam Christensen (2017, p. 1782) notes that certain art styles could also be construed as techniques, which confuse all using this framework. Further, Dam Christensen (2017, p. 1783) questions Huang et al.'s use of Panofsky's model solely in Layer 2. The advent of automation alone does not erase Panofsky's knowledge of the formal visual aspects of art; Panofsky made "'history of style'…a controlling principle at the level of 'pre-iconography'" (Dam Christensen, 2017, p. 1783). Thus, Dam Christensen (2017, p. 1783) argues that Huang et al.'s topical relational purpose underlying Panofsky's model subverts his original interpretative process.

Automation and Image Indexing

Using computers to automatically recognize subject matter in images seems ideal, but some researchers maintain that deeper meaning can only be identified through human review (Harpring, 2019b; Klenczon & Rygiel, 2014, p. 43). This content-based image indexing problem is defined as the 'semantic gap,' which "refers to the difference between the low-level information retrieved by a computer and the high-level semantic information understood by the user" (Rafferty, 2019, p. 6-7). Rafferty (2019, p. 7) emphasizes that this gap creates such a disruption in the computer automation process that its efficiencies are somewhat negated when it comes to meaning.

Modern methods to reduce the semantic gap include object ontology, machine learning algorithms, relevance feedback, semantic template generation, and fusion of textual and visual content of images (Pandey, Khanna, & Yokota, 2016, p. 573). Automated systems like content-based image retrieval and automatic image annotation use these techniques to decrease the semantic gap and become human-like in perception (Pandey et al., 2016, p. 571, 573). Pandey et

al. (2016, p. 572, 574) propose a fully automatic system to allow for complete semantic and image retrieval within a large image database.

This prototype begins with the image database mapping images for feature extraction (color, texture and shape) (Pandey et al., 2016, p. 574). It clusters these images at every node in its image tree, and the dominant clusters are assigned visual signatures at the node level (Pandey et al., 2016, p. 575-577). These visual signatures are indexed using three vantage-point trees, one each for color, texture, and shape (Pandey et al., 2016, p. 578). Not all dominant clusters are identified for a node; adaptive selection of tree branches are selected (Pandey et al., 2016, p. 578-80). Nodes that survive the pruning process are given general and specific semantics to locate an image (Pandey et al., 2016, p. 580). Pandey et al. (2016, p. 589) have created a fully automatic model that uses clustering to address semantic complexities inherent in image indexing. User feedback would be a worthwhile addition to enhance the system's performance, but then it would not be as automatic as its current iteration (Pandey et al., 2016, p. 589).

Democratization Methods of Image Indexing

Image retrieval using social subject tags or descriptors has increasing importance in image indexing (Huang et al., 2015, p. 1616). Particularly in the art world, a gap exists between expert, art historian, knowledge and how users describe the meaning of art images; democratizing tagging to allow indexing by users is one way to reduce this gap (Stewart, 2015, p. 17). As an alternative to traditional indexing, social tagging allows users to collaborate and classify subject 'ofness' or 'aboutness' (Stewart, 2015, p. 17). Rafferty (2019, p. 6) states that approaches to ad hoc, democratized systems of tagging involve "incorporating user behavior into image retrieval systems [through] browsing behavior (Treviosiol et al. 2012), sentiment analysis (Chen et al. 2014), and click-through behavior (Pan et al. 2014)." Rafferty (2019, p. 6) further

states computational approaches to social tagging involve creating algorithms around viewable aesthetics of artworks to discern their true artistic value as determined by experts and viewers. Thus, a hybrid model of social tagging, using traditional and user-distributed indexing, may be the most viable to tackle the large volume of images in the analog and digital world (Stewart, 2015, p. 17).

Conclusion

Image indexing continues to be complicated for both humans and computers. Traditional approaches are inadequate in today's 21st century landscape (Rafferty, 2019, p. 11). The underlying issue appears to be how can we precisely describe the relationships within and between images when language differs from user to user. Dam Christensen (2017, p. 1784) proposes a choice between a top-down controlled vocabulary system with expert indexers using theoretical models or a bottom-up democratized metadata approach. Controlled vocabularies in art indexing tend to be structured to emphasize hierarchical, associative, or equivalent relationships (Harpring, 2019a, p. 18). Harpring (2019b) at the Getty allows that crowdsourcing could be used among specialists for subject indexing, so long as proper training and rules are in place, but it is difficult to train experts to see an artistic subject the way a viewer does. That is where social tagging comes in to include users and viewers' viewpoints in determining meaning. Therefore, a combination of models and methods must be used for image indexing today. Automation can handle some of the workload of indexing images, but humans are still needed to review and enhance the results because they understand relationships in ways computers cannot yet fathom.

References

Cleveland, D. B., & Cleveland, A. D. (2013). *Introduction to indexing and abstracting* (4th ed.), 217-226. Retrieved from http://ebookcentral.proquest.com/lib/usf/detail.action?docID=1682964

Dam Christensen, H. (2017). Rethinking image indexing? *Journal of the Association for Information Science & Technology*, *68*(7), 1782-1785. Retrieved from https://doi-org.ezproxy.lib.usf.edu/10.1002/asi.23812

Harpring, P. (2019). Indexing art information: Using the Getty vocabularies. Retrieved from https://www.getty.edu/research/tools/vocabularies/vocab_brief_intro_and_indexing.pdf

Harpring, P. (2019). Subject access to art works: Overview using the Getty vocabularies. Retrieved from https://www.getty.edu/research/tools/vocabularies/subject_access_for_art.pdf

Huang, X., Soergel, D., & Klavans, J. L. (2015). Modeling and analyzing the topicality of art images. *Association for Information Science and Technology*, *66*(8), 1616-1644. Retrieved from http://search.ebscohost.com.ezproxy.lib.usf.edu/login.aspx?direct=true&db=edsbl&AN=RN375016861&site=eds-live

Klenczon, W., & Rygiel, P. (2014). Librarian cornered by images, or how to index visual resources. *Cataloging & Classification Quarterly*, *52*(1), 42-61. Retrieved from https://doi-org.ezproxy.lib.usf.edu/10.1080/01639374.2013.848123

Pandey, S., Khanna, P., & Yokota, H. (2016). A semantics and image retrieval system for hierarchical image databases. *Information Processing & Management*, *52*(4), 571-591. Retrieved from https://doi-org.ezproxy.lib.usf.edu/10.1016/j.ipm.2015.12.005

Rafferty, P. (2019). Disrupting the metanarrative: A little history of image indexing and retrieval. *Knowledge Organization*, *46*(1), 4-14. Retrieved from https://doi-org.ezproxy.lib.usf.edu/10.5771/0943-7444-2019-1-4

Stewart, B. (2015). Pictures into words. *The Indexer: The International Journal of Indexing*, *33*(1), 8-25. Retrieved from https://doiorg.ezproxy.lib.usf.edu/10.3828/indexer.2015.3