Computing Similarity between Cultural Heritage Items using Multimodal Features

Nikolaos Aletras
Department of Computer Science
University of Sheffield
Regent Court, 211 Portobello
Sheffield, S1 4DP, UK
n.aletras@dcs.shef.ac.uk

Mark Stevenson
Department of Computer Science
University of Sheffield
Regent Court, 211 Portobello
Sheffield, S1 4DP, UK
m.stevenson@dcs.shef.ac.uk

Abstract

A significant amount of information about Cultural Heritage artefacts is now available in digital format and has been made available in digital libraries. Being able to identify items that are similar would be useful for search and navigation through these data sets. Information about items in these repositories is often multimodal, such as pictures of the artefact and an accompanying textual description. This paper explores the use of information from these various media for computing similarity between Cultural Heritage artefacts. Results show that combining information from images and text produces better estimates of similarity than when only a single medium is considered.

1 Introduction and Motivation

In recent years a vast amount of Cultural Heritage (CH) artefacts have been digitised and made available on-line. For example, the Louvre and the British Museum provide information about exhibits on their web pages[1]. In addition, information is also available via sites that aggregate CH information from multiple resources. A typical example is Europeana[2], a web-portal to collections from several European institutions that provides access to over 20 million items including paintings, films, books, archives and museum exhibits.

However, online information about CH artefacts is often unstructured and varies by collection. This makes it difficult to identify information of interest in sites that aggregate information from multiple sources, such as Europeana, or to compare information across multiple collections (such as the Louvre and British Museum). These problems form a significant barrier to accessing the information available in these online collections. A first step towards improving access would be to identify similar items in collections. This could assist with several applications that are of interest to those working in CH including recommendation of interesting items (Pechenizkzy and Calders, 2007; Wang et al., 2008), generation of virtual tours (Joachims et al., 1997; Wang et al., 2009), visualisation of collections (Kauppinen et al., 2009; Hornbaek and Hertzum, 2011) and exploratory search (Marchionini, 2006; Amin et al., 2008).

Information in digital CH collections often includes multiple types of media such as text, images and audio. It seems likely that information from all of these types would help humans to identify similar items and that it could help to identify them automatically. However, previous work on computing similarity in the CH domain has been limited and, in particular, has not made use of information from multiple types of media. For example, Grieser et al. (2011) computed similarity between exhibits in Melbourne Museum by applying a range of text similarity measures but did not make use of other media. Techniques for exploiting information from multimedia collections have been developed and are commonly applied to a wide range of problems such as Content-based Image Retrieval (Datta et al., 2008) and image annotation (Feng and Lapata, 2010).

[1] http://www.louvre.fr/, http://www.britishmuseum.org/

[2] http://www.europeana.eu

Proceedings of the 6th EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pages 85–93,
Avignon, France, 24 April 2012.

This paper makes use of information from two media (text and images) to compute the similarity between items in a large collection of CH items (Europeana). A range of similarity measures for text and images are compared and combined. Evaluation is carried out using a set of items from Europeana with similarity judgements that were obtained in a crowdsourcing experiment. We find that combining information from both media produces better results than when either is used alone.

The main contribution of this paper is to demonstrate the usefulness of applying information from more than one medium when comparing CH items. In addition, it explores the effectiveness of different similarity measures when applied to this domain and introduces a data set of similarity judgements that can be used as a benchmark.

The remainder of this paper is structured as follows. Section 2 describes some relevant previous work. Sections 3, 4 and 5 describe the text and image similarity measures applied in this paper and how they are combined. Sections 6 describes the experiments used in this paper and the results are reported in Section 7. Finally, Section 8 draws the conclusions and provides suggestions for future work.

2 Background

2.1 Text Similarity

Two main approaches for determining the similarity between two texts have been explored: *corpus-based* and *knowledge-based* methods. Corpus-based methods rely on statistics that they learn from corpora while knowledge-based methods make use of some external knowledge source, such as a thesaurus, dictionary or semantic network (Agirre et al., 2009; Gabrilovich and Markovitch, 2007).

A previous study (Aletras et al., 2012) compared the effectiveness of various methods for computing the similarity between items in a CH collection based on text extracted from their descriptions, including both corpus-based and knowledge-based approaches. The corpus-based approaches varied from simple word counting approaches (Manning and Schutze, 1999) to more complex ones based on techniques from Information Retrieval (Baeza-Yates and Ribeiro-Neto, 1999) and topic models (Blei et al., 2003). The knowledge-based approaches relied on Wikipedia (Milne, 2007). Aletras et al. (2012) concluded that corpus-based measures were more effective than knowledge-based ones for computing similarity between these items.

2.2 Image Similarity

Determining the similarity between images has been explored in the fields such as Computer Vision (Szeliski, 2010) and Content-based Image Retrieval (CBIR) (Datta et al., 2008). A first step in computing the similarity between images is to transform them into an appropriate set of features. Some major feature types which have been used are colour, shape, texture or salient points. Features are also commonly categorised into global and local features.

Global features characterise an entire image. For example, the average of the intensities of red, green and blue colours gives an estimation of the overall colour distribution in the image. The main advantages of global features are that they can be computed efficiently. However, they are unable to represent information about elements in an image (Datta et al., 2008). On the other hand, local features aim to identify interesting areas in the image, such as where significant differences in colour intensity between adjacent pixels is detected.

Colour is one of the most commonly used global features and has been applied in several fields including image retrieval (Jacobs et al., 1995; Sebe and Michael S. Lew, 2001; Yu et al., 2002), image clustering (Cai et al., 2004; Strong and Gong, 2009), database indexing (Swain and Ballard, 1991) and, object/scene recognition (Schiele and Crowley, 1996; Ndjiki-Nya et al., 2004; Sande et al., 2008). A common method for measuring similarity between images is to compare the colour distributions of their histograms. A histogram is a graphical representation of collected counts for predefined categories of data. To create a histogram we have to specify the range of the data values, the number of dimensions and the bins (intervals into which ranges of values are combined). A colour histogram records the number of the pixels that fall in the interval of each bin. Schiele and Crowley (1996) describe several common metrics for comparing colour histograms including χ^2, *correlation* and

intersection.

2.3 Combining Text and Image Features

The integration of information from text and image features has been explored in several fields. In Content-based Image Retrieval image features are combined together with words from captions to retrieve images relevant to a query (La Cascia et al., 1998; Srihari et al., 2000; Barnard and Forsyth, 2001; Westerveld, 2002; Zhou and Huang, 2002; Wang et al., 2004). Image clustering methods have been developed to combine information from images and text to create clusters of similar images (Loeff et al., 2006; Bekkerman and Jeon, 2007). Techniques for automatic image annotation that generate models as a mixture of word and image features have also been described (Jeon et al., 2003; Blei and Jordan, 2003; Feng and Lapata, 2010).

2.4 Similarity in Cultural Heritage

Despite the potential usefulness of similarity in CH, there has been little previous work on the area. An exception is the work of Grieser et al. (2011). They computed the similarity between a set of 40 exhibits from Melbourne Museum by analysing the museum's web pages and physical layout. They applied a range of text similarity techniques (see Section 2.1) to the web pages as well as similarity measures that made use of Wikipedia. However, the Wikipedia-based techniques relied on a manual mapping between the items and an appropriate Wikipedia article. Although the web pages often contained images of the exhibits, Grieser et al. (2011) did not make use of them.

3 Text Similarity

We make use of various corpus-based approaches for computing similarity between CH items since previous experiments (see Section 2.1) have shown that these outperformed knowledge-based methods in a comparison of text-based similarity methods for the CH domain.

We assume that we wish to compute the similarity between a pair of items, A and B, and that each item has both text and an image associated with it. The text is denoted as A_t and B_t while the images are denoted by A_i and B_i.

3.1 Word Overlap

A common approach to computing similarity is to count the number of common words (Lesk, 1986). The text associated with each item is compared and the similarity is computed as the number of words (tokens) they have in common normalised by the combined total:

$$sim_{WO}(A, B) = \frac{|A_t \cap B_t|}{|A_t \cup B_t|}$$

3.2 N-gram Overlap

The Word Overlap approach is a bag of words method that does not take account of the order in which words appear, despite the fact that this is potentially useful information for determining similarity. One way in which this information can be used is to compare n-grams derived from a text. Patwardhan et al. (2003) used this approach to extend the Word Overlap measure. This approach identifies n-grams in common between the two text and increases the score by n^2 for each one that is found, where n is the length of the n-gram. More formally,

$$sim_{ngram}(A, B) = \frac{\sum_{n \in n-gram(A_t, B_t)} n^2}{|A_t \cup B_t|}$$

where $n-gram(A_t, B_t)$ is the set of n-grams that occur in both A_t and B_t.

3.3 TF.IDF

The word and n-gram overlap measures assign the same importance to each word but some are more important for determining similarity between texts than others. A widely used approach to computing similarity between documents is to represent them as vectors in which each term is assigned a weighting based on its estimated importance (Manning and Schutze, 1999). The vectors can then be compared using the cosine metric. A widely used scheme for weighting terms is tf.idf, which takes account of the frequency of each term in individual documents and the number of documents in a corpus in which it occurs.

3.4 Latent Dirichlet Allocation

Topic models (Blei et al., 2003) are a useful technique for representing the underlying content of documents. LDA is a widely used topic model

that assumes each document is composed of a number of topics. For each document LDA returns a probability distribution over a set of topics that have been derived from an unlabeled corpus. Similarity between documents can be computed by converting these distributions into vectors and using the cosine metric.

4 Image Similarity

Two approaches are compared for computing the similarity between images. These are largely based on colour features and are more suitable for the images in the data set we use for evaluation (see Section 6).

4.1 Colour Similarity (RGB)

The first approach is based on comparison of colour histograms derived from images.

In the RGB (Red Green Blue) colour model, each pixel is represented as an integer in range of 0-255 in three dimensions (Red, Green and Blue). One histogram is created for each dimension. For grey-scale images it is assumed that the value of each dimension is the same in each pixel and a single histogram, called the luminosity histogram, is created. Similarity between the histograms in each colour channel is computed using the intersection metric. The *intersection* metric (Swain and Ballard, 1991) measures the number of corresponding pixels that have same colour in two images. It is defined as follows:

$$Inter(h_1, h_2) = \sum_I min(h_1(I), h_2(I))$$

where h_i is the histogram of image i, I is the set of histogram bins and $min(a, b)$ is the minimum between corresponding pixel colour values.

The final similarity score is computed as the average of the red, green and blue histogram similarity scores:

$$sim_{RGB}(A_i, B_i) = \frac{\sum_{i \in \{R,G,B\}} Inter(h_{A_i}, h_{B_i})}{3}$$

4.2 Image Querying Metric (imgSeek)

Jacobs et al. (1995) described an image similarity metric developed for Content-based Image Retrieval. It makes use of Haar wavelet decomposition (Beylkin et al., 1991) to create signatures of images that contain colour and basic shape information. Images are compared by determining the number of significant coefficients they have in common using the following function:

$$dist_{imgSeek}(A_i, B_i) = w_0 |C_{A_i}(0,0) - C_{B_i}(0,0)| + \sum_{i,j:\tilde{C}_{A_i}(i,j) \neq 0} w_{bin(i,j)} (\tilde{C}_{A_i}(i,j) \neq \tilde{C}_{B_i}(i,j))$$

where w_b are weights, C_I represents a single colour channel for an image I, $C_I(0,0)$ are scaling function coefficients of the overall average intensity of the colour channel and $\tilde{C}_I(i,j)$ is the (i,j)-th truncated, quantised wavelet coefficient of image I. For more details please refer to Jacobs et al. (1995).

Note that this function measures the distance between two images with low scores indicating similar images and high scores dis-similar ones. We assign the negative sign to this metric to assign high scores to similar images. It is converted into a similarity metric as follows:

$$sim_{imgSeek}(A_i, B_i) = -dist_{imgSeek}(A_i, B_i)$$

5 Combining Text and Image Similarity

A simple weighted linear combination is used to combine the results of the text and image similarities, sim_{img} and sim_t. The similarity between a pair of items is computed as follows

$$sim_{T+I}(A, B) = w_1 \cdot sim_t(A_t, B_t) + w_2 \cdot sim_{img}(A_i, B_i)$$

where w_i are weights learned using linear regression (see Section 6.4).

6 Evaluation

This section describes experiments used to evaluate the similarity measures described in the previous sections.

6.1 Europeana

The similarity measures are evaluated using information from Europeana[3], a web-portal that provides access to information CH artefacts. Over 2,000 institutions through out Europe have contributed to Europeana and the portal provides access to information about over 20 million CH artefacts, making it one of the largest repositories

[3] http://www.europeana.eu

of digital information about CH currently available. It contains information about a wide variety of types of artefacts including paintings, photographs and newspaper archives. The information is in a range of European languages, with over 1 million items in English. The diverse nature of Europeana makes it an interesting resource for exploring similarity measures.

The Europeana portal provides various types of information about each artefact, including textual information, thumbnail images of the items and links to additional information available for the providing institution's web site. The textual information is derived from metadata obtained from the providing institution and includes title, description as well as details of the subject, medium and creator.

An example artefact from the Europeana portal is shown in Figure 1. This particular artefact is an image showing detail of an architect's office in Nottingham, United Kingdom. The information provided for this item is relatively rich compared to other items in Europeana since the title is informative and the textual description is of reasonable length. However, the amount of information associated with items in Europeana is quite varied and it is common for items to have short titles, which may be uninformative, or have very limited textual descriptions. In addition, the metadata associated with items in Europeana is potentially a valuable source of information that could be used for, among other things, computing similarity between items. However, the various providing institutions do not use consistent coding schemes to populate these fields which makes it difficult to compare items provided by different institutions. These differences in the information provided by the various institutions form a significant challenge in processing the Europeana data automatically.

6.2 Evaluation Data

A data set was created by selecting 300 pairs of items added to Europeana by two providers: Culture Grid[4] and Scran[5]. The items added to Europeana by these providers represent the majority that are in English and they contain different types of items such as objects, archives, videos and audio files. We removed five pairs that did not have any images associated with one of the items. (These items were audiofiles.) The resulting dataset consists of 295 pairs of items and is referred to as **Europeana295**.

Each item corresponds to a metadata record consisting of textual information together with a URI and a link to its thumbnail. Figure 1 shows an item taken from the Europeana website. The title, description and subject fields have been shown to be useful information for computing similarity (Aletras et al., 2012). These are extracted and concatenated to form the textual information associated with each item. In addition, the accompanying thumbnail image (or "preview") was also extracted to be used as the visual information. The size of these images varies from 7,000 to 10,000 pixels.

We have pre-processed the data by removing stop words and applying stemming. For the *tf.idf* and *LDA* the training corpus was a total of 759,896 Europeana items. We have filtered out all items that have no description and have a title shorter than 4 words, or have a title which has been repeated more than 100 times.

6.3 Human Judgements of Similarity

Crowdflower[6], a crowdsourcing platform, was used to obtain human judgements of the similarity between each pair of items. Participants were asked to rate each item pair using a 5 point scale where 4 indicated that the pair of items were highly similar or near-identical while 0 indicated that they were completely different. Participants were presented with a page containing 10 pairs of items and asked to rate all of them. Participants were free to rate as many pages as they wanted up to a maximum of 30 pages (i.e. the complete Europeana295 data set). To ensure that the annotators were not returning random answers each page contained a pair for which the similarity had been pre-identified as being at one end of the similarity scale (i.e. either near-identical or completely different). Annotations from participants that failed to answer correctly these questions or participants that have given same rating to all of their answers were removed. A total of 3,261 useful annotations were collected from 99 participants and each pair was rated by at least 10 participants.

The final similarity score for each pair was gen-

[4] http://www.culturegrid.org.uk/

[5] http://www.scran.ac.uk/

[6] http://crowdflower.com/

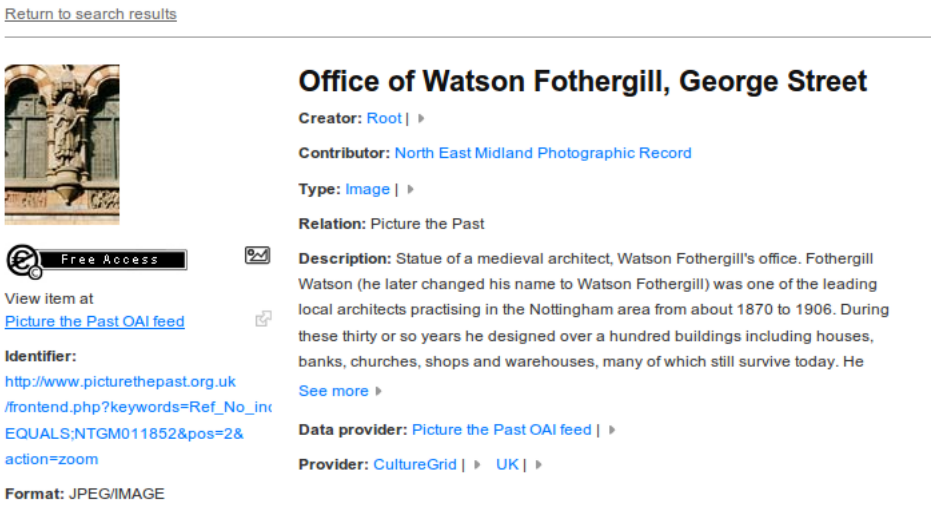

Figure 1: Example item from Europeana portal showing how both textual and image information are displayed. (Taken from `http://www.europeana.eu/portal/`)

erated by averaging the ratings. Inter-annotator agreement was computed as the average of the Pearson correlation between the ratings of each participant and the average ratings of the other participants, a methodology used by Grieser et al. (2011). The inter-annotator agreement for the data set was $\rho = +0.553$, which is comparable with the agreement score of $\rho = +0.507$ previously reported by Grieser et al. (2011).

6.4 Experiments

Experiments were carried out comparing the results of the various techniques for computing text and image similarity (Sections 3 and 4) and their combination (Section 5). Performance is measured as the Pearson's correlation coefficient with the gold-standard data.

The combination of text and image similarity (Section 5) relies on a linear combination of text and image similarities. The weights for this combination are obtained using a linear regression model. The input values were the results obtained for the individual text and similarity methods and the target value was the gold-standard score for each pair in the dataset. 10-fold cross-validation was used for evaluation.

7 Results

An overview of the results obtained is shown in Table 1. Results for the text and image similarity methods used alone are shown in the left and top part of the table while the results for their combination are in the main body.

		Image Similarity	
		RGB	imgSeek
Text Similarity		0.254	**0.370**
Word Overlap	**0.487**	0.450	**0.554**
tf.idf	0.437	0.426	0.520
N-gram overlap	0.399	0.384	0.504
LDA	0.442	0.419	0.517

Table 1: Performance of similarity measures applied to Europeana295 data set (Pearson's correlation coefficient).

The best performance for text similarity (0.487) is achieved by Word Overlap and the lowest by N-gram Overlap (0.399). The results are surprising since the simplest approach produces the best results. It is likely that the reason for these results is the nature of the textual data in Europeana. The documents are often short, in some cases the description missing or the subject information is identical to the title.

For image similarity, results using imgSeek are higher than RGB (0.370 and 0.254 respectively). There is also a clear difference between the performance of the text and image similarity methods and results obtained from both image similarity measures is lower than all four that are based on text. The reason for these results is the nature of the Europeana images. There are a large number of black-and-white image pairs which means that colour information cannot be obtained from

many of them. In addition, the images are low resolution, since they are thumbnails, which limits the amount of shape information that can be derived from them, restricting the effectiveness of imgSeek. However, the fact that performance is better for imgSeek and RGB suggests that it is still possible to obtain useful information about shape from these images.

When the image and text similarity measures are combined the highest performance is achieved by the combination of the Word Overlap and imgSeek (0.554), the best performing text and image similarity measures when applied individually. The performance of all text similarity measures improves when combined with imgSeek. All results are above 0.5 with the highest gain observed for N-gram Overlap (from 0.399 to 0.504), the worst performing text similarity measure when applied individually. On the other hand, combining text similarity measures with RGB consistently leads to performance that is lower than when the text similarity measure is used alone.

These results demonstrate that improvements in similarity scores can be obtained by making use of information from both text and images. In addition, better results are obtained for the text similarity methods and this is likely to be caused by the nature of the images which are associated with the items in our data set. It is also important to make use of an appropriate image similarity method since combing text similarity methods with RGB reduces performance.

8 Conclusion

This paper demonstrates how information from text and images describing CH artefacts can be combined to improve estimates of the similarity between them. Four corpus-based and two image-based similarity measures are explored and evaluated on a data set consisting of 295 manually-annotated pairs of items from Europeana. Results showed that combing information from text and image similarity improves performance and that imgSeek similarity method consistently improves performance of text similarity methods.

In future work we intend to make use of other types of image features including the low-level ones used by approaches such as Scale Invariant Feature Transformation (SIFT) (Lowe, 1999; Lowe, 2004) and the bag-of-visual words model (Szeliski, 2010). In addition we plan to apply these approaches to higher resolution images to determine how the quality and size of an image affects similarity algorithms.

Acknowledgments

The research leading to these results was carried out as part of the PATHS project (`http://paths-project.eu`) funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270082.

References

Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Pasca, and Aitor Soroa. 2009. A study on similarity and relatedness using distributional and wordnet-based approaches. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL '09)*, pages 19–27, Boulder, Colorado.

Nikolaos Aletras, Mark Stevenson, and Paul Clough. 2012. Computing similarity between items in a digital library of cultural heritage. *Submitted*.

Alia Amin, Jacco van Ossenbruggen, Lynda Hardman, and Annelies van Nispen. 2008. Understanding Cultural Heritage Experts' Information Seeking Needs. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 39–47, Pittsburgh, PA.

Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*. Addison Wesley Longman Limited, Essex.

Kobus Barnard and David Forsyth. 2001. Learning the Semantics of Words and Pictures. *Proceedings Eighth IEEE International Conference on Computer Vision (ICCV '01)*, 2:408–415.

Ron Bekkerman and Jiwoon Jeon. 2007. Multi-modal clustering for multimedia collections. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pages 1–8.

Gregory Beylkin, Ronald Coifman, and Vladimir Rokhlin. 1991. Fast Wavelet Transforms and Numerical Algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183.

David M. Blei and Michael I. Jordan. 2003. Modeling Annotated Data. *Proceedings of the 26th annual international ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR '03)*, pages 127–134.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Deng Cai, Xiaofei He, Zhiwei Li, Wei-Ying Ma, and Ji-Rong Wen. 2004. Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information. *Proceedings of the 12th annual ACM international conference on Multimedia (MULTIMEDIA '04)*, pages 952–959.

Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. 2008. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, 40(2):1–60.

Yansong Feng and Mirella Lapata. 2010. Topic Models for Image Annotation and Text Illustration. In *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 831–839, Los Angeles, California, June.

Evgeniy Gabrilovich and Shaul Markovitch. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '07)*, pages 1606–1611.

Karl Grieser, Timothy Baldwin, Fabian Bohnert, and Liz Sonenberg. 2011. Using Ontological and Document Similarity to Estimate Museum Exhibit Relatedness. *Journal on Computing and Cultural Heritage (JOCCH)*, 3(3):10:1–10:20.

Kasper Hornbaek and Morten Hertzum. 2011. The notion of overview in information visualization. *International Journal of Human-Computer Studies*, 69:509–525.

Charles E. Jacobs, Adam Finkelstein, and David H. Salesin. 1995. Fast multiresolution image querying. In *Proceedings of the 22nd annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*, pages 277–286, New York, NY, USA.

Jiwoon Jeon, Victor Lavrenko, and Raghavan Manmatha. 2003. Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th annual international ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '03)*, pages 119–126, New York, NY, USA.

Thorsten Joachims, Dayne Freitag, and Tom Mitchell. 1997. Webwatcher: A tour guide for the world wide web. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '97)*, pages 770–777.

Tomi Kauppinen, Kimmo Puputti, Panu Paakkarinen, Heini Kuittinen, Jari Väätäinen, and Eero Hyvönen. 2009. Learning and visualizing cultural heritage connections between places on the semantic web. In *Proceedings of the Workshop on Inductive Reasoning and Machine Learning on the Semantic Web (IRMLeS2009) and the 6th Annual European Semantic Web Conference (ESWC2009)*, Heraklion, Crete, Greece.

Marco La Cascia, Sarathendu Sethi, and Stan Sclaroff. 1998. Combining textual and visual cues for content-based image retrieval on the world wide web. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 24–28.

Michael Lesk. 1986. Automatic Sense Disambiguation using Machine Readable Dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the ACM Special Interest Group on the Design of Communication Conference (SIGDOC '86)*, pages 24–26, Toronto, Canada.

Nicolas Loeff, Cecilia Ovesdotter Alm, and David A. Forsyth. 2006. Discriminating image senses by clustering with multimodal features. In *Proceedings of the COLING/ACL on Main Conference Poster Sessions (COLING-ACL '06)*, pages 547–554, Stroudsburg, PA, USA.

David G. Lowe. 1999. Object Recognition from Local Scale-invariant Features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pages 1150–1157.

David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Christopher D. Manning and Hinrich Schutze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press.

Gary Marchionini. 2006. Exploratory Search: from Finding to Understanding. *Communications of the ACM*, 49(1):41–46.

David Milne. 2007. Computing Semantic Relatedness using Wikipedia Link Structure. In *Proceedings of the New Zealand Computer Science Research Student Conference*.

Patrick Ndjiki-Nya, Oleg Novychny, and Thomas Wiegand. 2004. Merging MPEG 7 Descriptors for Image Content Analysis. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, pages 5–8.

Siddhard Patwardhan, Satanjeev Banerjee, and Ted Pedersen. 2003. Using Measures of Semantic Relatedness for Word Sense Disambiguation. In *Proceedings of the 4th International Conference on Intelligent Text Processing and Computational Linguistics*, pages 241–257.

Mykola Pechenizkzy and Toon Calders. 2007. A framework for guiding the museum tours personalization. In *Proceedings of the Workshop on Personalised Access to Cultural Heritage (PATCH '07)*, pages 11–28.

Koen E.A. Sande, Theo Gevers, and Cees G. M. Snoek. 2008. Evaluation of Color Descriptors for Object and Scene Recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pages 1–8.

Bernt Schiele and James L. Crowley. 1996. Object recognition using multidimensional receptive field histograms. In *Proceedings of the 4th European Conference on Computer Vision (ECCV '96)*, pages 610–619, London, UK.

Nicu Sebe and Michael S. Lew. 2001. Color-based Retrieval. *Pattern Recognition Letters*, 22:223–230, February.

Rohini K. Srihari, Aibing Rao, Benjamin Han, Srikanth Munirathnam, and Xiaoyun Wu. 2000. A model for multimodal information retrieval. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '00)*, pages 701–704.

Grant Strong and Minglun Gong. 2009. Organizing and Browsing Photos using Different Feature Vectors and their Evaluations. *Proceedings of the ACM International Conference on Image and Video Retrieval (CIVR '09)*, pages 3:1–3:8.

Michael J. Swain and Dana H. Ballard. 1991. Color indexing. *International Journal of Computer Vision*, 7:11–32.

Richard Szeliski. 2010. *Computer Vision: Algorithms and Applications*. Springer-Verlag Inc. New York.

Xin-Jing Wang, Wei-Ying Ma, Gui-Rong Xue, and Xing Li. 2004. Multi-model similarity propagation and its application for web image retrieval. In *Proceedings of the 12th annual ACM International Conference on Multimedia (MULTIMEDIA '04)*, pages 944–951, New York, NY, USA.

Yiwen Wang, Natalia Stash, Lora Aroyo, Peter Gorgels, Lloyd Rutledge, and Guus Schreiber. 2008. Recommendations based on semantically-enriched museum collections. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):43–50.

Yiwen Wang, Lora Aroyo, Natalia Stash, Rody Sambeek, Schuurmans Yuri, Guus Schreiber, and Peter Gorgels. 2009. Cultivating personalized museum tours online and on-site. *Journal of Interdisciplinary Science Reviews*, 34(2):141–156.

Thijs Westerveld. 2002. Probabilistic multimedia retrieval. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02)*, pages 437–438, New York, NY, USA.

Hui Yu, Mingjing Li, Hong-Jiang Zhang, and Jufu Feng. 2002. Color Texture Moments for Content-based Image Retrieval. In *Proceedings of the IEEE International Conference on Image Processing (ICIP '02)*, pages 929–932.

Xiang Sean Zhou and Thomas S. Huang. 2002. Unifying keywords and visual contents in image retrieval. *IEEE Multimedia*, 9(2):23 –33.