Learning Visual Entities and their Visual Attributes from Text Corpora

Erik Boiy
Dept. of Computer Science
K.U.Leuven, Belgium
erik.boiy@cs.kuleuven.be

Koen Deschacht
Dept. of Computer Science
K.U.Leuven, Belgium
koen.deschacht@cs.kuleuven.be

Marie-Francine Moens
Dept. of Computer Science
K.U.Leuven, Belgium
sien.moens@cs.kuleuven.be

Abstract

We automatically construct a dictionary of visual (possible to perceive on a picture) or non-visual (impossible to perceive directly on a picture) entities and attributes, based on statistical association techniques used in data mining. We compute whether certain words that could function as entities or attributes of an entity are correlated with texts that describe images and use these words for the detection of visual nouns and visual adjectives. We compare our corpus-based approach with a knowledge-rich approach based on WordNet, and with a combination of both approaches.

1 Introduction

Text-based image retrieval is the search for images using their textual descriptions. Very often, people manually annotate images with key terms, however this is a tedious task. Furthermore, images often are accompanied by descriptive texts. Not all terms in these descriptions contribute to the image content. Many terms cannot contribute because they represent non-visual entities, or non-visual properties or attributes. If a resource that defines the visualness of a word exists, more refined index descriptors can automatically be assigned to the images. In addition, the textual descriptions filtered with regard to their visualness, automatically annotate the images and can be used for training image recognizers. Moreover, there is an increasing need for aligning content across different media, to which the dictionary with visual words contributes.

In this paper we learn visual and non-visual terms from corpora based on statistical association techniques. We compare our corpus-based approach with WordNet knowledge and integrate both approaches. We evaluate our methods on unseen texts that describe images and place our research in the context of related research that finds correlations between information objects across different media and in the context of text-based image retrieval.

2 Methodology

Entities (objects or persons) are usually expressed by the syntactic category of nouns (e.g., "cat"). Their qualification by attributes is usually signaled by adjectives or by participles of verbs used as adjectives, found in the following syntactic templates: 1) Adjective as a modifier of a noun (e.g., "the striped cat"); 2) Adjective as a predicate of a noun (e.g., "The car is red."). Not all nouns refer to a visual object or person, and not all adjectives refer to a physical and visual qualification of the entity. So, the task is given the above syntactic templates, finding these words that signal a visual entity or visual property of an entity.

The first step is to create a dictionary of words that are considered visual, using corpus-based association techniques. During this training phase no distinction is made between part-of-speech classes of words. Only in the testing phase, particular syntactic templates are considered. The dictionary list is obtained by mining two corpora: one corpus that is considered to be mostly visual and one that contains very little visual information. In a next step we build a visual dictionary based on the existing lexical resource, WordNet. Finally, we combine the two approaches.

2.1 Corpus-based association techniques

Hypothesis testing can be used to determine which words are related to a specific domain (in our case the visual domain), represented by a target corpus, compared to a general reference corpus [9] (henceforth assumed to be non-visual). We use here a likelihood ratio for a binomial distribution, and a chi-square metric.

2.1.1 Likelihood ratio

Following [12], we test the hypothesis of independence between a term (in our experiments below single words) and a certain class (here visual or non-visual) with the likelihood ratio for a binomial distribution. If we can reject the hypothesis, there is a strong indication that the word is correlated

	visual	!visual	
term = t	c_{12}	$c_2 - c_{12}$	c_2
term != t	$c_1 - c_{12}$	$N + c_{12} - c_1 - c_2$	$N - c_2$
	c_1	$N - c_1$	N

Table 1. Contingency table for words appearing in the visual and non-visual corpus.

with either the visual (if the words appears more often in the visual corpus) or the non-visual class (if the word appears more often in the non-visual corpus).
We first define p_1 and p_2:

$$p_1 = P(term = t|visual) \quad p_2 = P(term = t|!visual)$$

The hypotheses of independence (H_1) assumes that p_1 and p_2 are equal, i.e.:

$$H_1 : p_1 = p_2 = p$$

whereas H_2 allows all possible values of p_1 and p_2. We define the likelihood functions given a contingency table (Table 1). The cells of this table indicate the number of occurrences; c_{12} is the number of occurrences of term t in the visual corpus, c_2 is the total number of occurrences of t and c_1 is the number of terms in the visual corpus. N is the total number of tokens in both corpora. The probability that the visual class generates a term and the probability that the non-visual class generates a term are assumed to follow a binomial distribution.

$$L(H_1) = b(c_{12}; c_1, p) b(c_2 - c_{12}; N - c_1, p)$$

$$L(H_2) = b(c_{12}; c_1, p_1) b(c_2 - c_{12}; N - c_1, p_2)$$

where $b(k; n, x)$ is the binomial distribution. The maximum value of the likelihood functions above is obtained by setting their first derivative to zero, which yields for p, p_1 and p_2 the following values:

$$p = \frac{c_2}{N} \qquad p_1 = \frac{c_{12}}{c_1} \qquad p_2 = \frac{c_2 - c_{12}}{N - c_1}$$

We then determine the likelihood ratio λ

$$\begin{aligned} \lambda &= \frac{L(H_1)}{L(H_2)} \\ &= \frac{L(p, c_{12}, c_1) L(p, c_2 - c_{12}, N - c_1)}{L(p_1, c_{12}, c_1) L(p_2, c_2 - c_{12}, N - c_1)} \end{aligned}$$

with

$$L(p, k, n) = p^k (1 - p)^{n-k}$$

We take

$$\begin{aligned} \log \lambda = \; & -\log L(p_1, c_{12}, c_1) - \log L(p_2, c_2 - c_{12}, N - c_1) \\ & + \log L(p, c_{12}, c_1) + \log L(p, c_2 - c_{12}, N - c_1) \end{aligned}$$

which is a value that is asymptotically χ^2 distributed. By selecting a confidence level from the χ^2 distribution table (1 degree of freedom), the obtained likelihood ratio value allows us to accept or to reject the H_1 hypothesis.

2.1.2 Chi-square

Given the above contingency table we can test the hypothesis of independence of the occurrence of a term in the visual and non-visual classes by means of a χ^2 test. The χ^2 statistic is defined as:

$$\chi^2 = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

with $E_{i,j}$ the expected value for $O_{i,j}$ (the observed value, found at position i, j in the cells of the contingency table). The expected frequencies $E_{i,j}$ are calculated from the marginal probabilities:

$$E_{i,j} = \frac{O_{i,1} + O_{i,2}}{N} \times \frac{O_{1,j} + O_{2,j}}{N} \times N$$

The resulting value is χ^2 distributed and allows us to accept or reject independence as was the case for the likelihood ratio.

2.2 Use of WordNet

The lexical resource of WordNet does not contain information on the visualness of a term. Therefore, we infer this visualness from the information present in WordNet and additional world knowledge. The method we implemented is described and evaluated in detail in [10] and is summarized here. The visualness is defined as the degree that a noun or an adjective is considered visual. Inspired by [13] who identify the polarity (i.e., expressing of positive or negative feeling) of adjectives, we manually identify seed synsets in WordNet that are certainly visual (e.g. “person”, “red”) or completely invisible (e.g. “power”, “confidential”). This is done separately for nouns and adjectives.

We set the visualness of these seed synsets to either 1 (visual) or 0 (non-visual). We determine the visualness of all other synsets using these seed synsets. A synset that is close to a visual seed synset gets a high visualness (vis) and vice versa. We choose a linear weighting:

$$vis(s) = \sum_i vis(s_i) \frac{sim(s, s_i)}{C(s)}$$

where $vis(s)$ returns a number between 0 and 1[1]; s, s_i are the seed synsets; $sim(s, t)$ returns a number between 0 and

[1] A cutoff value is selected to determine which nouns and adjectives we consider visual for our classification.

1 denoting the similarity between synsets s and t; and $C(s)$ is constant given a synset s:

$$C(s) = \sum_i sim(s, s_i)$$

For nouns we use the similarity $sim(s_i, s_j)$ as defined by [15]. Alternative measures that compute the similarity between synsets can be found in [5]. For adjectives WordNet does not define a hierarchical relationship, thus we use the semantic distances as defined in [14]. In this approach we always manually choose the seed set for nouns and adjectives. The computation of visualness of proper nouns is out of scope of this paper. Here a named entity recognizer [18] could detect visual categories of proper names (e.g., person names).

2.3 Combination of association techniques and WordNet

A third method combines the two previously discussed methods. We set the seed set of synsets to the best 25 visual and non-visual nouns and the 25 best visual and non-visual adjectives from the dictionary that is built using the statistical association techniques. We then employ the method using the WordNet database as discussed before.

2.4 Experiments

We first learn the dictionary of visual words from a visual and non-visual corpus. Then, we test the validity of the dictionary on an unseen corpus, where every head noun and adjective is manually annotated as visual or non-visual.

For the visual corpus used for training we use the following corpora (or combinations of):

- **Flowers corpus**: This collection of flower descriptions contains 15,226 word tokens including punctuation, and is obtained from http://www.uniqflowers.com.
- **Antiques corpus**: This collection of old picture descriptions contains 619,515 word tokens, and is obtained from the Oregon Archives found at http://photos.salemhistory.org/cdm4/browse.php.

Example items of the corpora can be found in Figure 1. For the non-visual corpus we use the full collection of English Wikipedia article texts (407,074,407 word tokens), or a subset consisting of the main articles on the major religions in China (Confucianism, Taoism, Buddhism; 16,965 word tokens).

From these corpora we extract the list of visual words by applying the association techniques discussed above. If the technique finds that a word is significantly more likely to appear in a visual context (corpus), it is added to the list. Examples of learned visual words are "building", "wooden", "flower" and "white". Examples of words strongly considered non-visual are "year", "great", "new" and "season".

"African violets (Saintpaulia ionantha) are small, flowering houseplants or greenhouse plants belonging to the Gesneriaceae family. They are perhaps the most popular and most widely grown houseplant. Their thick, fuzzy leaves and abundant blooms in soft tones of violet, purple, pink, and white make them very attractive. Numerous varieties and hybrids are available. African violets grow best in indirect sunlight."

"These small sculptures depict two identical human figures. The wooden bodies are weathered brown and the hair is faded blue. Both sculptures have a round base about one inch high. The feet are large and flat, with grooves cut into the front to distinguish toes. The legs are short, ..."

"A small girl looks up at a person dressed in the costume of an animal which could be "Woody Woodchuck" at the State Fair in Salem, Oregon."

Figure 1. Entries in the flowers corpus (top), art corpus (middle) and antiques corpus (bottom).

We evaluate on a separate ground-truth corpus, henceforth called the **Art corpus**. This is a collection of art comments and elaborate descriptions of the history of the object or artist in addition to those of the object itself. The corpus is found at http://tours.daytonartinstitute.org/accessart/tour.cfm?TT=ct. The texts of the test corpus are POS tagged, and all head nouns of noun chunks and adjectives of these chunks are identified, and their visualness is annotated. In total 8 art items are annotated (resulting in 1,337 word tokens).

Classification of the above adjectives and nouns found in the ground truth corpus is performed automatically by looking up the word in the dictionary of visual words. If a word is present, it is classified as visual, if not, the non-visual class is assigned. We compute the accuracy, precision and recall of the visual and non-visual classification. Table 2 shows the results based on the statistical corpus-based methods. Table 3 shows the results where the dictionary is built based on WordNet. Table 4 shows the results based on the combination of both approaches. Although all words can be labeled using the two WordNet-based approaches, we did not evaluate the words that were not known by the statistical technique for the purpose of comparison. In all tables, "T" stands for the technique used (λ = likelihood ratio, χ^2 = chi-square statistics) and α for the confidence level ("-" means that no confidence threshold was applied; all words that were positively correlated with the visual class are included). "A %" stands for accuracy, "? %" for percentage of words in the test corpus not seen in the training corpora[2]. Finally, "V" stands for visual, while "!V" for non-visual.

In a final experiment, we used the combination of the statistical and WordNet-based approach for building the visual dictionary (trained on the flowers, antiques and wiki corpora) and tested how well a head noun inherits the visualness of its adjective modifier, and vice versa (Table 5). We first identified noun groups by using LTChunk [17], then the relation between adjective and head noun within a chunk can be trivially extracted. When evaluating an adjective, the head noun it modifies is looked up in the dictionary and its visual and non-visual classification is transferred to the adjective. When evaluating a head noun, the classification is taken from the preceding adjective if present. Some nouns (65.39%) and adjectives (9.05%) did not receive a classification, when they occur respectively without modifier or head noun.

The results show that we can quite successfully acquire the visualness of attributes (adjectives) from small visual (description of flowers) and non-visual (description of religions) corpora. When using a statistical confidence threshold of 0.99 the dictionary size is substantially reduced (from 2,504 words without applying a threshold to 102 words for the likelihood ratio metric and 229 words for the chi-square metric), possibly causing the drop in recall of the visual class.

Using larger, less specific corpora increases coverage at the cost of reduced F1 values. Here the difference between applying a confidence threshold or not has less impact on the dictionary size and thus the results. For adjectives the recall of the visual class drops by 32% compared to using specific corpora, while only 23% more data is evaluated. Given the fact that the dictionary is larger, this leads us to the conclusion that using less specific corpora in training has a negative impact on dictionary quality. For nouns this effect is not observed.

The superior results for adjectives (especially recall of the visual class), combined with the observation that an adjective is 1.7 times more likely to be visual in our ground truth corpus, indicates that the generated dictionary has a better coverage of visual adjectives than of visual nouns. This could be explained by the assumptions that visual adjectives are more generic than visual nouns, so applicable to more texts, whereas visual nouns tend to be more specifically related to subject domains. Looking at our corpora, 78% of the adjectives in the ground truth corpus were also present in either of the visual corpora, whereas for the head nouns this is only 57%.

The likelihood ratio and the chi-square metric give almost identical results, although we found the chi-square metric easier to implement.

Computing the visualness of nouns using semantic distances in WordNet based on manually selected seed sets of visual and non-visual words gives already satisfying results. Judging from the lines with high coverage in Table 4, we can improve this visual and non-visual classification by selecting machine generated seed sets for both adjectives and nouns.

Some errors could be due to the fuzziness in the concept of visualness, which could be to some degree context-dependent.

3 Relevance for information retrieval and related research

Since the early days of image retrieval, text-based approaches are common because users often express an information need in terms of a natural language utterance [7, 21]. Especially in a Web context, text-based image retrieval is important given that users are acquainted with keyword searches. [8] demonstrated the importance of content that surround the images on Web pages for their effective retrieval and investigated how multiple evidence from selected content fields of HTML Web pages (e.g. meta tags, description tags, passages) contribute to a better image annotation. Also [22] combined textual and visual evidence in Web image retrieval. Textual context is not only useful in retrieval, it can also assist in annotating the images or training tools for image content recognition. Recognizing content in images that relies on descriptions of surrounding texts is researched, for instance, by [3, 19, 20]. [1, 2, 4, 24] recognize image content based on textual and visual cues. In all these state of the art examples, the textual analysis is limited to a bag-of-words representation where terms are often weighted with a tf x idf scheme. In text based image retrieval where pictures or video images are searched that have elaborate accompanying texts, a better discrimination of the terms based on their visualness seems appropriate. Moreover, a sequence of visual terms can pinpoint a section in the text that describes the images. Term discrimination based on visualness can be integrated in a vector space retrieval model and in probabilistic models, such as language models (cf. [11]) and Bayesian networks.

Association techniques are popular to identify collocations of related terms in monolingual corpora [16] or for finding term correlations across languages as was done by [6] and [23]. Using association metrics to identify visual-

[2]It would be unfair to include such a word in the evaluation, because we did not learn whether it is visual or not.

Training Corpora		T	α	? %	A %	Precision %		Recall %		F1 %	
V	!V					V	!V	V	!V	V	!V
flowers	religion	λ	-	52.55	**61.22**	81.71	**47.56**	**50.95**	79.59	**62.76**	**59.54**
flowers	religion	λ	.99	52.55	38.54	82.35	36.64	5.32	**97.96**	10.00	53.33
flowers	religion	χ^2	-	52.55	**61.22**	81.71	**47.56**	**50.95**	79.59	**62.76**	**59.54**
flowers	religion	χ^2	.99	52.55	40.49	80.65	37.20	9.51	95.92	17.01	53.61
flowers+antiques	wiki	λ	-	0.93	47.66	83.78	35.02	31.10	86.05	45.37	49.78
flowers+antiques	wiki	λ	.99	0.93	47.55	**84.02**	35.01	30.77	86.43	45.04	49.83
flowers+antiques	wiki	χ^2	-	0.93	47.66	83.78	35.02	31.10	86.05	45.37	49.78
flowers+antiques	wiki	χ^2	.99	0.93	44.63	82.63	33.78	26.25	87.21	39.85	48.70
flowers	religion	λ	-	24.69	**77.13**	88.26	**53.45**	**80.15**	68.13	**84.01**	**59.90**
flowers	religion	λ	.99	24.69	38.84	87.88	27.95	21.32	91.21	34.32	42.78
flowers	religion	χ^2	-	24.69	**77.13**	88.26	**53.45**	**80.15**	68.13	**84.01**	**59.90**
flowers	religion	χ^2	.99	24.69	47.38	**96.55**	31.88	30.88	**96.70**	46.80	47.96
flowers+antiques	wiki	λ	-	1.66	57.38	85.79	38.38	48.22	80.15	61.74	51.90
flowers+antiques	wiki	λ	.99	1.66	57.17	85.71	38.25	47.93	80.15	61.48	51.78
flowers+antiques	wiki	χ^2	-	1.66	57.38	85.79	38.38	48.22	80.15	61.74	51.90
flowers+antiques	wiki	χ^2	.99	1.66	56.75	86.34	38.14	46.75	81.62	60.65	51.99

Table 2. The results of detecting visual head nouns (first section) and adjectives (second section) in the Art corpus using a dictionary built with the corpus-based approach in terms of accuracy (A), precision, recall and F1 measure. The best results are in bold.

Cutoff	? %	A %	Precision %		Recall %		F1 %	
			V	!V	V	!V	V	!V
.3	0.00	64.00	82.02	43.87	62.02	68.58	70.63	53.51
.3	0.00	56.22	81.13	36.67	50.15	71.22	61.98	48.41

Table 3. The results of detecting visual head nouns (first section) and adjectives (second section) in the Art corpus using a dictionary derived from WordNet in terms of accuracy (A), precision, recall and F1 measure. Cutoff refers to the threshold of the visualness value vis used for defining the visualness of the term.

Training Corpora		Cutoff	? %	A %	Precision %		Recall %		F1 %	
V	!V				V	!V	V	!V	V	!V
flowers	religion	.5	52.55	60.24	78.74	**46.61**	52.09	**74.83**	62.70	**57.44**
flowers+antiques	wiki	.3	0.93	**66.71**	**82.67**	46.42	**66.22**	67.83	**73.54**	55.12
flowers	religion	.3	24.69	**73.00**	80.00	**45.21**	**85.29**	36.26	**82.56**	40.24
flowers+antiques	wiki	.3	1.66	61.39	**87.80**	41.26	53.25	**81.62**	66.30	**54.81**

Table 4. The results of detecting visual head nouns (first section) and adjectives (second section) in the Art corpus using a dictionary derived from WordNet, where the seed sets were annotated based on the corpus-based approach in terms of accuracy (A), precision, recall and F1 measure. The best results are in bold.

A %	Precision %		Recall %		F1 %	
	V	!V	V	!V	V	!V
56.79	81.82	32.97	53.73	65.59	64.86	43.88
58.86	90.23	33.73	52.17	81.16	66.12	47.66

Table 5. The visualness of the adjective modifiers is inherited by the its head noun (first line); the visualness of the head noun is inherited by its modifier (second line): Results in terms of accuracy (A), precision, recall and F1 measure.

ness of a term is novel, as well as the use of statistical association metrics in combination with WordNet knowledge.

4 Conclusions and future work

Hypothesis testing – as is done here using a likelihood ratio for a binomial distribution and a chi-square statistic – is valuable to determine the visualness of a term (in our case noun or adjective), when trained on texts that are presumed to contain visual information and texts that are presumed to contain non-visual information. An alternative way is deriving the visualness from a lexical resource such as WordNet based on the manual annotations of seed sets of visual and non-visual terms. The latter method could be improved by using the statistical corpus-based approach for seed set selection and yields the best results. In future work we will integrate our approaches in text-based image retrieval models.

5 Acknowledgment

Our research is sponsored by the EU-FP6 027978 project CLASS and by the Nederlandse TaalUnie -NWO STE05039 CORNETTO projects.

References

[1] A. Amir, J. Argillander, M. Campbell, A. Haubold, G. Iyengar, S. Ebadollahi, F. Kang, M. R. Naphade, A. Natsev, J. R. Smith, J. Tešió, and T. Volkmer. IBM Research TRECVID-2005 Video Retrieval System. In *Proceedings of TRECVID 2005*, Gaithersburg, MD, 2005.

[2] S. Ayache, G. M. Qunot, J. Gensel, and S. Satoh. CLIPS-LRS-NII Experiments at TRECVID 2005. In *Proceedings of TRECVID 2005*, Gaithersburg, MD, 2005.

[3] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan. Matching Words and Pictures. *Journal of Machine Learning Research*, 3(6):1107–1135, 2003.

[4] T. L. Berg, A. C. Berg, J. Edwards, and D. Forsyth. Who's in the Picture? In *Neural Information Processing Systems*, pages 137–144, 2004.

[5] A. Budanitsky and G. Hirst. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1):13–47, 2006.

[6] J. G. Carbonell, Y. Yang, R. E. Frederking, R. D. Brown, Y. Geng, and D. Lee. Translingual information retrieval: A comparative evaluation. In *IJCAI (1)*, pages 708–715, 1997.

[7] S. Chang and A. Hsu. Image information systems: where do we go from here? *IEEE Transactions on Knowledge and Data Engineering*, 4(5):431–442, 1992.

[8] T. Coelho, P. Calado, L. Souza, and B. Ribeiro-Neto. Image Retrieval Using Multiple Evidence Ranking. *Image*, 16(4):408–417, 2004.

[9] D. Damle and V. Uren. Extracting significant words from corpora for ontology extraction. In *K-CAP '05: Proceedings of the 3rd International Conference on Knowledge Capture*, pages 187–188, New York, NY, USA, 2005. ACM.

[10] K. Deschacht and M.-F. Moens. Text Analysis for Automatic Image Annotation. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 1000–1007, 2007.

[11] K. Deschacht and M.-F. Moens. Finding the best picture: Cross-media retrieval of content. In *ECIR*, pages 539–546, 2008.

[12] T. E. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, 1993.

[13] J. Kamps and M. Marx. Words with Attitude. In *Proceedings of the 1st International Conference on Global WordNet*, pages 332–341, India, 2002.

[14] M. Lesk. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26. ACM Press New York, NY, USA, 1986.

[15] D. Lin. An Information-Theoretic Definition of Similarity. In *Proc. 15th International Conf. on Machine Learning*, 1998.

[16] C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, 1999.

[17] A. Mikheev and S. Finch. A workbench for finding structure in texts. In *Proceedings of the fifth conference on Applied natural language processing*, pages 372–379, San Francisco, CA, USA, 1997. Morgan Kaufmann Publishers Inc.

[18] M.-F. Moens. *Information Extraction: Algorithms and Prospects in a Retrieval Context*. Springer, Boston, MA, 2006.

[19] Y. Mori, H. Takahashi, and R. Oka. Automatic Word Assignment to Images Based on Image Division and Vector Quantization. In *RIAO-2000 Content-Based Multimedia Information Access*, Paris, April 12-14 2000.

[20] S. Satoh, Y. Nakamura, and T. Kanade. Name-It: Naming and Detecting Faces in News Videos. *IEEE MultiMedia*, 6(1):22–35, January-March 1999.

[21] H. Tamura and N. Yokoya. Image database systems: A survey. *Pattern Recognition*, 17(1):29–43, 1984.

[22] S. Tollari and H. Glotin. Web image retrieval on IMAGEVAL: Evidences on visualness and textualness concept dependency in fusion model. In *ACM International Conference on Image and Video Retrieval (CIVR)*, July 2007.

[23] M. Wang, C. A. Perfetti, and Y. Liu. Chinese-english biliteracy acquisition: cross-language and writing system transfer. *Cognition*, 97(1):67–88, August 2005.

[24] T. Westerveld. Image Retrieval: Content versus Context. In *Content-Based Multimedia Information Access, RIAO 2000 Conference Proceedings*, pages 276–284, April 2000.