Image Identification by Content Similarity: An Approach to Accurate the Caption generation for Article Images

K Sarika
M.Tech in Computer Science Engineering
Aurora's Technological & Research Institute,
Parvathapur, Uppal, Hyderabad-500039

Mr. K Ramana Reddy
Sr.Asst Prof in Computer Science Dept
Aurora's Technological & Research Institute,
Parvathapur, Uppal, Hyderabad-500039

Abstract: We present a holistic image content and description driven approach to image caption generation, exploiting the vast amount of (noisy) parallel image data and associated natural language descriptions given. More specifically, given a query image along with description, we recommend CBIR (Content Based Image Retrieval) approach to identify the image, and then selectively combine the phrases extracted from the description to generate a novel caption for the query image. We cast the generation process as constraint optimization problems, collectively incorporating multiple interconnected aspects of language composition for caption planning, surface realization and discourse structure.

I. Introduction

Recent years have witnessed an unprecedented growth in the amount of digital information available on the Internet. Flickr, one of the best known photo sharing websites, hosts more than three billion images, with approximately 2.5 million images being uploaded every day. Many on-line news sites like CNN, Yahoo!, and BBC publish images with their stories and even provide photo feeds related to current events. Browsing and finding pictures in large-scale and heterogeneous collections is an important problem that has attracted much interest within information retrieval.

Many of the search engines deployed on the web retrieve images without analyzing their content, simply by matching user queries against collocated textual information. Examples include meta-data (e.g., the image's file name and format), user-annotated tags, captions, and generally text surrounding the image. As this limits the applicability of search engines (images that do not coincide with textual data cannot be retrieved), a great deal of work has focused on the development of methods that generate description words for a picture automatically. The literature is littered with various attempts to learn the associations between image features and words using supervised classification (Vailaya et al., 2001; Smeulders et al., 2000), instantiations of the noisy- channel model (Duygulu et al., 2002), latent variable models (Blei and Jordan, 2003; Barnard et al., 2002; Wang et al., 2009), and models inspired by information retrieval (Lavrenko et al., 2003; Feng et al., 2004).

In this paper we go one step further and generate captions for images rather than individual keywords. Although image indexing techniques based on keywords are popular and the method of choice for image retrieval engines, there are good reasons for using more linguistically meaningful descriptions. A list of keywords is often ambiguous. An image annotated with the words blue, sky, car could depict a blue car or a blue sky, whereas the caption "car running under the blue sky" would make the relations between the words explicit. Automatic caption generation could improve image retrieval by supporting longer and more targeted queries. It could also assist journalists in creating descriptions for the images associated with their articles. Beyond image retrieval, it could increase the accessibility of the web for visually impaired (blind and partially sighted) users who cannot access the content of many sites in the same ways as sighted users can (Ferres et al., 2006).

II. Related Work

Although image understanding is a popular topic within computer vision, relatively little work has focused on the interplay between visual and linguistic information. A handful of approaches generate image descriptions automatically following a two-stage architecture. The picture is first analyzed using image processing techniques into an abstract representation, which is then rendered into a natural language description with a text generation engine. A common theme across different models is domain specificity, the use of hand- labeled data, and reliance on background ontological information.

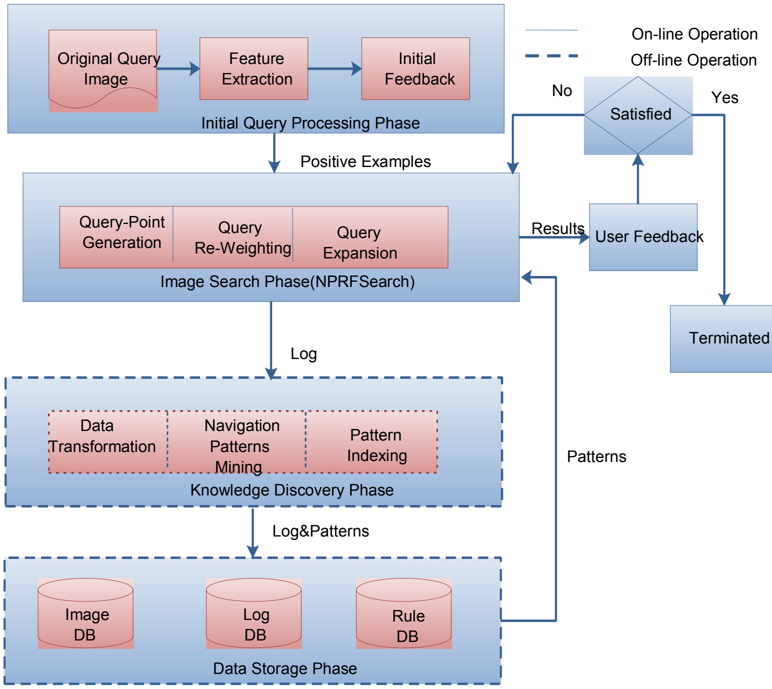
For example, Hede et al. (2004) generate descriptions for images of objects shot in uniform background. Their system relies on a manually created database of objects indexed by an image signature (e.g., color and texture) and two keywords (the object's name and category). Images are first segmented into objects, their signature is retrieved from the database, and a description is generated using templates. Kojima et al. (2002, 2008) create descriptions for human activities in office scenes. They extract features of human motion and interleave them with a concept hierarchy of actions to create a case frame from which a natural language sentence is generated. Yao et al. (2009) present a general framework for generating text descriptions of image and video content based on image parsing. Specifically, images are hierarchically decomposed into their constituent visual patterns which are subsequently converted into a semantic representation using WordNet. The image parser is trained on a corpus, manually annotated with graphs representing image structure.

III. PROPOSED APPROACH

As explained above, the crucial issue of RF can be chiefly concluded thus: how to get effective and efficient image retrieval. In order to deal with this issue, we explain how our suggested approach RRF merges the founded navigation patterns and 3 RF techniques for achieving efficient and effective exploration of images.

A. *Re-querying by Relevance Feedback*

The major difference between our suggested procedure and other modern procedures is that we approximate an optimal solution for resolving the problems prevailing in current RF like redundant browsing and exploration convergence. To this extent, the approximated solution takes benefit of exploited knowledge (navigation patterns) for assisting the suggested search strategy in efficiently hunting the required images. Usually, the work of the suggested procedure can be divided into 2 major operations. They are offline knowledge discovery and online image retrieval. As shown in Fig. 1, each operational phase has some crucial segments



Work Flow of NPRF

Fig: 1 Workflow of NPRF

in order to complete the specific process. Coming to online operation, after a query image is presented to this system, the system first detects the most similar images without taking into consideration any search strategy, and after that returns a set of the most similar images. The first query process is known as initial feedback. Now, the good examples chosen by the user give the important information to the image search phase, comprising new feature weights, new query point, and the user's intention. After that, by utilizing the navigation patterns, 3 search strategies, with respect to QPM, QR, and QAMP, are hybridized to find the desired images. Overall, at each feedback, the results are given to the user and the associated browsing information is saved in the log database. After gathering long-term users' browsing behaviors, offline operation for knowledge discovery is initiated for executing navigation pattern mining and pattern indexing. The framework of the suggested procedure is briefly explained below:

IV. Online Image Retrieval.

Initial Query Processing Phase: During this phase we don't consider the feature weight. This phase brings out the visual features from the original query image for finding out the similar images. Then, the good examples (also called as positive examples in this paper) chosen by the user are further scrutinized at the first feedback (also called as iteration 0 in this paper). Image Search Phase: Our intention behind the search phase is to extend the one search point to multiple search points by merging the navigation patterns and the suggested search algorithm RRF Search. In this manner, the diverse inclusion of the user's interest can be successfully put in. During this phase, a new query point at each and every feedback is produced by the preceding positive examples. Now, the k-nearest images to the new query point can be detected by augmenting the weighted query. The search process does not halt till the user is quenched with the retrieval results.

i. Offline Knowledge Discovery.

Knowledge Discovery Phase: gaining knowledge from users' behaviors in image retrieval can be seen as one type of knowledge discovery. As a result of this, this phase primarily regards the development of the navigation model by finding the implied navigation patterns from users' browsing behaviors. This model can furnish image search with a good support to foretell optimal image browsing ways. Data Storage Phase: The databases in this phase can be considered as the knowledge marts of a knowledge warehouse that store integrated, time-variant, and nonvolatile collection of useful data comprising images, navigation patterns, log files, and image features. The knowledge warehouse is very helpful for ameliorating the quality of image retrieval. Observe that the process of developing rule base from the image databases can be done periodically for maintaining the validity of the suggested approach.

The thing is that usage mining has been developed on how to produce users' browsing patterns for facilitating the web pages retrieval. In a similar fashion, for web image retrieval, the user has to give a query term to the search engine, so-called textual-based image search. Now, the user can get a set of most relevant web images in accordance with the metadata or the browsing log. But, in the case of the result not satisfying the user, the query refinement can be put into the query procedure without any difficulty. This is the reason why CBIR utilizing RF has been the focus of the researchers in the field of image retrieval. To the extent the usage log of CBIR is taken into consideration, the challenge mainly depends on: how to produce and get benefit of the discovered patterns. In this paper, we construct a navigation-pattern based data structure penetrated by the Query Flow Structure aspect that has never been suggested by earlier studies. By using the special data structure, the user's intention can be caught more quickly and precisely.

Going into the details, the data structure can be seen as a hierarchy, comprising positive images, query points, and clusters. A query session has a set of iterative feedbacks (iterations) that is referred to a navigation path. After getting a feedback, the positive examples that show the results chosen up by the user, are utilized for deriving a referred visual query point by averaging the positive visual features. Lastly, the query sessions, iterations, positive examples, and visual query points are saved into the original log database, as indicated In the case of the original log data being ready, the next task is finding navigation patterns from the original log data. Primarily, navigation pattern discovery comprises two stages: data transformation and navigation patterns mining. For data transformation, as indicated in lines 1-6 of Fig. 2, the visual query points of the $i\ th$ iteration are made into groups of n clusters, where n indicates the maximum session length and $0 \leq i \leq n$. After that, the visual query points in each cluster are changed into a specific symbol, known as item #.The transformed log table is also divided into various sub tables at this time. In the case of navigation patterns mining, as indicated in lines 8-21 of Fig.2, the frequent item sets are mined from the navigation-transaction table.

```
Input: The original log database (referred to Figure 8) and the minimum support minsup;

Output: A set of navigation patterns NP;

Procedure Gen NP

1. generate the transformed log table by the original log data;
2. for $0 \leq i \leq n$ do
3. group query-point of the ith Iteration Into $m$ clusters; $I*m$ in the number of the clusters, which is determined by the system manager*/
4. symbolize each query-point Into a Item number by its belonging cluster number;
5. store Item # Into the transformed table;

6. end for
7. partition the transformed table Into Navigation-Transaction Table, QP Table and Partitioned Log Table; (referred to Figure 8)
8. let $F_e$ be the net of the frequent e-itemsets and find $F_1$ by scanning the navigation-transaction table;
```

```
let NP be the set of navigation patterns and initialize NP-0;
let $CA_e$ be the eth candidate Item set and initialize $e=2$;
while $|F_{c-1}| \neq 0$ do
    for each $p,q \in F_{c-1}$ do
        if $p_1 = q^1 \& p^2 = q^2 \& ... \& p_{c-1} = q_{c-1}$ then
            $fr$ concatenate itemset $\{p_1, p_2, .., p_{c-1}, p_c, q_c\}$;
            $CA_c = fr \cup CA_c$;
        end if
    end for
    $F_e = \{X | \sup port(X) \geq \min\sup, X \subseteq CA_e\}$;
    $NF - NF \cup F_e, /* NF$ denotes the set of the frequent Item-sets*/
    $e = e+1$;
end while
return NP;
```

Fig: 2 Procedure for offline knowledge discovery

i. Data Transformation

Till now, a very few important studies have been successful in semantic image retrieval or image recognition. The reason is the complicated visual contents. For handling the unclearness in image presentation, data transformation for visual content is a primary and significant operation as it can make simpler both the illustration of visual query points as well as the discovery of navigation patterns. In other terms, if we don't consider the data transformation, then we have to take into consideration all positive images of each query session in the log database. In the case of all positive images being considered for navigation pattern mining, many items make the frequent item sets (navigation patterns) difficult to find. Moreover, the mining cost is costly. Because of this, the goal of data transformation is producing Query Point Dictionary (QPD) for reducing the kinds of items on the transaction list.

As a query point is projected onto the QPD, the item number that is referred is saved into the changed log table. Then the changed log table has to be further divided into 3 tables for different needs in this paper, comprising QP table, Navigation-transaction table, and Partitioned Log table. Navigation-transaction table is utilized for navigation patterns mining. QP table and Partitioned Log table are required for image search talked about at length in further sections. In viewing the complete data, the jointed table can be deduced with connecting the joint attributes of distinct tables.

ii. Pattern Indexing

During this stage, we explain in detail how to construct the navigation pattern tree using the navigation patterns that are discovered. As given , the navigation patterns can be considered as the branches of the navigation pattern tree. After the navigation patterns are produced, the query item C1m 2 Qin each and every navigation pattern is utilized as a seed (called query seed) for planting the navigation pattern tree. So, in the case of the cardinality of the clusters being 7, there are 7 navigation trees produced during this stage. A tree possesses various navigation paths, and each node of the paths indicates an item comprising several visual query points. A visual query point shows a set of positive images. Particularly, to decline the complexities of pattern search and pattern storage, the unnecessary navigation patterns have to be curtailed to a larger extent. The redundancy that is seen between two patterns can be defined as follows:

Definition 1 (Pattern Redundancy). Take into consideration 2 navigation patterns, namely Fitemset1= $\{C_{ab}, ...C_{ij}\}$ and Fitemset2 = $\{C_{pg}, ...C_{xy}\}$. If C_{ab} = C_{pg} , C_{ij} = C_{xy} , and |Fitemset1| ≥ |jFitemset2|, Fitemset1 is called inessential navigation pattern. After removing the unnecessary patterns, the curtailed navigation pattern tree declines the search cost to a large extent. Depending on the navigation pattern tree, the required images can be captured more promptly without having repetition of the scan of the whole image database at each and every feedback, more specifically for the large-scale image data.

V. Algorithm RRF Search

Based on above discussion, RRF Search is suggested to get the high precision of image retrieval in a shorter query procedure by utilizing the highly useful navigation patterns. This section explains the details of RRF Search. As shown in Fig. 3, the RRF Search algorithm is induced by receiving:1) a set of positive examples G and negative examples N ascertained by the user at the preceding feedback, 2) a set of navigation patterns $\{tr_1, tr_2 ....., tr_h\}$, where each tr_h contains a query seed rt_h and many patterns an accuracy threshold thrd. Briefly, the iterative search process can be divided into different steps as given below:

1. Produce a new query point by averaging the visual features of positive examples.

2. Discover the suitable navigation pattern trees by ascertaining the nearest query seeds (root).

3. Discover the nearest leaf nodes (terminations of a path) from the corresponding navigation pattern trees.

4. Find the top s relevant visual query points from the set of the nearest leaf nodes.

5. Lastly, the top k relevant images are sent back to the user.

Considering RRF Search, step 1 can be considered as QPM and steps 2-5 can be considered as QAMP. For QR, the feature weights are iteratively updated depending on the positive examples at each and every feedback. In total, the suggested RRFSearch takes benefit of QPM, QAMP, QR, and navigation patterns for making RF more efficient and effective. Without using navigation patterns, RRF Search cannot get the high quality of RF. Considering the viewpoint of applicability; the aim of our approach is to quench each query efficiently in place of furnishing personalized functions for each and every user. Therefore, irrelevant queries from a user will not be cause a problem. By gathering a large number of query transactions, almost all the queries can be well answered for matching user's interests by RRFSearch. The aspects of the RRFSearch algorithm are explained as follows:

Input: A set of positive examples $G - \cup g$, picked up by the user, a set of negative examples $N - \cup n_a$, a set of navigation patterns TR-$\{tr_1, tr_2, ..., tr_3\}$ with referred query-seed set Q-$\{rt_1, rt_2, ..., rt_3\}$, and a accuracy threshold thrd;

Output: A set of the relevant images R;

Algorithm NPRFSearch

```
Generate a new query point qp… by G and compute the new feature weights by Equation 3;
Let NIMG be the accumulated set of negative examples, and $NIMG - NIMG \cup N;$
Store qp… and G into the log database;
Initialize each $tr_\lambda, rt_A chk = 0$ and CanPut= $\phi$;
For each g, ∈ G do
    Determine the special query-seed $rt$, with the shortest distance to g, where rt, ∈ $Q$;
    $rt, chk - 1$
End for
$if \frac{|G|}{|G \cup N|} < thrd$ then
    For each $n_u \in N$ do
        Determine the special seed $rt_a$ with the shortest distance to $n_{x,}$ where $rt_a$ ⊂ QandQ ⊆ TR;
        Count( $rt_a$ )++;
    End for
    Find the send $rt_a$ with max(count( $rt_a$ ));
    $rt_a chk - 0;$
End if
For each $rt_a$ do
    If $tr, rt, chk - 1$ then
        Find the set of the visual query points QPT within the leafnodes of pattern $tr_a$;
        CanPut-CanPut ∪ QPT; /*CanPut indicates the set of the accum ulated candidate query points*/
    End if
End for
Find the top s visual query points $SQPT = \{sqpt_1, sqpt_2, ..., sqpt_3\}$ similar to qp… for CanPut;
For i=1 to $\delta$ do
    Find the positive image set RIMG in the transformed log table, which is referred to $sqpts$;
    Canimg-Canimg ∪ RIMG; /* Canimg indicates the set of the relevant images*/
End for
Canimg={Canimg\NIMG}
Rank the images in Canimg;
Return the set of top k similar images R;
```

Fig:3 Algorithm for NPRF search

Query point generation. The primary idea of this operation is discovering the images not only with the specific similarity function. By recursively changing the query point, the search direction can move toward the targets progressively. Presume that a set of images is discovered by the query point qp_{old} at the preceding feedback. Now, the visual features of the positive examples G chosen up by the user are first averaged into a new query point qp_{new}. In the mean time, qp_{new} and the positive examples are saved into the log database to elevate the knowledge database. Now, the negative examples are attached to the gathered negative set NIMG. At each and every feedback, removing MING from the targets can augment the accuracy of image retrieval to a large extent. Apart from generating qp_{new} and MING, the weights of all features have to be calculated in order to keep searching the images that are similar to qp_{new}. In this paper, the feature weight for similarity computation is normalized as explained further. Feature reweighting.

VI. Experimental Setup

In this section we discuss our experimental design for assessing the performance of the caption generation models presented above. We give details on our training procedure, parameter estimation, and present the baseline methods used for comparison with our models.

Data All our experiments were conducted on the corpus created by Feng and Lapata (2008), following their original partition of the data (2,881 image-caption-document tuples for training, 240 tuples for development and 240 for testing). Documents and captions were parsed with the Stanford parser (Klein and Manning, 2003) in order to obtain dependencies for the phrase-based abstractive model.

Model Parameters For the image annotation model we extracted 150 (on average) SIFT features which were quantized into 750 visual terms. The underlying topic model was trained with 1,000 topics using only content words (i.e., nouns, verbs, and adjectives) that appeared no less than five times in the corpus. For all models discussed here (extractive and abstractive) we report results with the 15 best annotation keywords. For the abstractive models, we used a trigram model trained with the SRI toolkit on a newswire corpus consisting of BBC and Yahoo! news documents (6.9 M words). The attachment probabilities (see equation (14)) were estimated from the same corpus. We tuned the caption length parameter on the development set using a range of [5,14] tokens for the word-based model and [2,5] phrases for the phrase-based model. Fol¬lowing Banko et al. (2000), we approximated the length distribution with a Gaussian. The scaling parameter P for the adaptive language model was also tuned on the development set using a range of [0.5,0.9]. We report results with P set to 0.5. For the abstractive models the beam size was set to 500 (with at least 50 states for the word-based model). For the phrase-based model, we also experimented with reducing the search scope, either by considering only the n most similar sentences to the keywords (range [2,10]), or simply the single most similar sentence and its neighbors (range [2,5]). The former method delivered better results with 10 sentences (and the KL divergence similarity function).

VII. Conclusions

We have presented extractive and abstractive models that generate image captions for news articles. A key aspect of our approach is to allow both the visual and textual modalities to influence the generation task. This is achieved through an image annotation model that characterizes pictures in terms of description keywords that are subsequently used to guide the caption generation process. Our results show that the visual information plays an important role in content selection. Simply extracting a sentence from the document often yields an inferior caption. Our experiments also show that a probabilistic abstractive model defined over phrases yields promising results. It generates captions that are more grammatical than a closely related word-based system and manages to capture the gist of the image (and document) as well as the captions written by journalists.

Future extensions are many and varied. Rather than adopting a two-stage approach, where the image processing and caption generation are carried out sequentially, a more general model should integrate the two steps in a unified framework. Indeed, an avenue for future work would be to define a phrase-based model for both image annotation and caption generation. We also believe that our approach would benefit from more detailed linguistic and non-linguistic information. For instance, we could experiment with features related to document structure such as titles, headings, and sections of articles and also exploit syntactic information more directly. The latter is currently used in the phrase-based model by taking attachment probabilities into account. We could, however, improve grammaticality more globally by generating a well-formed tree (or dependency graph).

References

Banko, Michel, Vibhu O. Mittal, and Micheael J. Witbrock. 2000. Headline generation based on statistical translation. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics.* Hong Kong, pages 318-325.

Barnard, Kobus, Pinar Duygulu, David Forsyth, Nando de Freitas, David Blei, and Michael Jordan. 2002. Matching words and pictures. *Journal of Machine Learning Research* 3:1107-1135.

Blei, David and Michael Jordan. 2003. Modeling annotated data. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Toronto, ON, pages 127-134.*

Blei, David, Andrew Ng, and Michael Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3:993-1022.

Corio, Marc and Guy Lapalme. 1999. Generation of texts for information graphics. In *Proceedings of the 7th European Workshop on Natural Language Generation*. Toulouse, France, pages 49-58.

Dorr, Bonnie, David Zajic, and Richard Schwartz. 2003. Hedge trimmer: A parse-and-trim approach to headline generation. In *Proceedings of the HLT-NAACL 2003 Workshop on Text Summarization.* Edmonton, Canada, pages 1-8.

Duygulu, Pinar, Kobus Barnard, Nando de Freitas, and David Forsyth. 2002. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the 7th European Conference on Computer Vision.* Copenhagen, Denmark, pages 97-112.

Elzer, Stephanie, Sandra Carberry, Ingrid Zukerman, Daniel Chester, Nancy Green, , and Seniz Demir. 2005. A probabilistic framework for recognizing intention in information graphics. In *Proceedings of the 19th International Conference on Artificial Intelligence.* Edinburgh, Scotland, pages 1042-1047.

Fasciano, Massimo and Guy Lapalme. 2000. Intentions in the coordinated generation of graphics and text from tabular data. *Knowledge Information Systems* 2(3):310-339.

Feiner, Steven and Kathleen McKeown. 1990. Coordinating text and graphics in explanation generation. In *Proceedings of National Conference on Artificial Intelligence*. Boston, MA, pages 442-449.

Feng, Shaolei Feng, Victor Lavrenko, and R Manmatha. 2004. Multiple Bernoulli relevance models for image and video annotation. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*. Washington, DC, pages 1002-1009.

Feng, Yansong and Mirella Lapata. 2008. Automatic image annotation using auxiliary text information. In *Proceedings of the 46th Annual Meeting of the Association of Computational Linguistics: Human Language Technologies*. Columbus, OH, pages 272-280.

Feng, Yansong and Mirella Lapata. 2010. Topic models for image annotation and text illustration. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles, LA.

Ferres, Leo, Avi Parush, Shelley Roberts, and Gitte Lindgaard. 2006. Helping people with visual impairments gain access to graphical information through natural language: The *graph* system. In *Proceedings of 11th International Conference on Computers Helping People with Special Needs.* Linz, Austria, pages 1122-1130.

Hede, Patrick, Pierre Allain Moellic, Joel Bourgeoys, Magali Joint, and Corinne Thomas. 2004. Automatic generation of natural language descriptions for images. In *Proceedings of Computer-Assisted Information Retrieval (Recherche d'Information et ses Applications Ordinateur) (RIAO).* Avignon, France.

Jin, Rong and Alexander G. Hauptmann. 2002. A new probabilistic model for title generation. In *Proceedings of the 19th International Conference on Computational linguistics.* Taipei, Taiwan, pages 1-7.

Klein, Dan and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics*. Sapporo, Japan, pages 423-430.

Kneser, Reinhard, Jochen Peters, and Dietrich Klakow. 1997. Language model adaptation using dynamic marginals. In *Proceedings of 5th European Conference on Speech Communication and Technology*. Rhodes, Greece, volume 4, pages 1971-1974.

Kojima, Atsuhiro, Mamoru Takaya, Shigeki Aoki, Takao Miyamoto, and Kunio Fukunaga. 2008. Recognition and textual description of human activities by mobile robot. In *Proceedings of the 3rd International Conference on Innovative Computing Information and Control*. IEEE Computer Society, Washington, DC, pages 5356.

Kojima, Atsuhiro, Takeshi Tamura, and Kunio Fukunaga. 2002. Natural language description of human activities from video images based on concept hierarchy of actions. *International Journal of Computer Vision* 50(2):171-184.

Lavrenko, Victor, R. Manmatha, and Jiwoon Jeon. 2003. A model for learning the semantics of pictures. In *Proceedings of the 16th Conference on Advances in Neural Information Processing Systems. Vancouver, BC.*

Lowe, David G. 1999. Object recognition from local scale-invariant features. In *Proceedings of International Conference on Computer Vision.* IEEE Computer Society, pages 1150-1157.

Mittal, Vibhu O., Johanna D. Moore, Giuseppe Carenini, and Steven Roth. 1998. Describing complex charts in natural language: A caption generation system. *Computational Linguistics* 24:431-468.

Monay, Florent and Daniel Gatica-Perez. 2007. Modeling semantic aspects for cross-media image indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(10):1802-1817.

Salton, Gerard and M.J. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York.

Smeulders, Arnols W.M., Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12):1349-1380.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*. Cambridge, pages 223231.

Steyvers, Mark and Tom Griffiths. 2007. Probabilistic topic models. In T. Landauer, D. McNamara, S Dennis, and W Kintsch, editors, *A Handbook of Latent Semantic Analysis,* Psychology Press.