Journal of Universal Computer Science, vol. 17, no. 1 (2011), 3-15
submitted: 8/4/10, accepted: 12/12/10, appeared: 1/1/11

Document Retrieval Using SIFT Image Features

Dan Smith, Richard Harvey
(School of Computing Sciences
University of East Anglia
Norwich NR4 7TJ, UK
Dan.Smith@uea.ac.uk, R.W.Harvey@uea.ac.uk)

Abstract: This paper describes a new approach to document classification based on visual features alone. Text-based retrieval systems perform poorly on noisy text. We have conducted series of experiments using cosine distance as our similarity measure, selecting varying numbers local interest points per page, and varying numbers of nearest neighbour points in the similarity calculations. We have found that a distance-based measure of similarity outperforms a rank-based measure except when there are few interest points. We show that using visual features substantially outperforms text-based approaches for noisy text, giving average precision in the range 0.4-0.43 in several experiments retrieving scientific papers.
Key Words: document classification, SIFT
Category: H.3.1

1 Introduction

Information retrieval for text-based documents is a mature problem compared to image retrieval. Nevertheless many documents contain diagrams, images and other graphical elements which are not amenable to text-based retrieval. Furthermore text-based information retrieval and document classification systems rely heavily on having good quality machine-readable text and the quality of their results is significantly affected by noise.

The goal of the work reported here is to investigate the potential of document retrieval and classification using visual features. Our experimental results clearly demonstrate that visual features can provide good retrieval performance. The precision obtained shows that visual features are capable of capturing sufficient of the semantics of the documents to enable useful retrieval systems to be constructed.

The remainder of this paper is structured as follows. In section 2 we outline related work on scientific document retrieval; in section 3 we describe the SIFT algorithm and some of it applications. Following this, in section 4 we describe our corpus and experimental approach, in section 5 we describe the text-based experimental results, and in section 6 we describe our image-based experimental results. Section 8 contains our conclusions and suggestions for future work.

2 Scientific paper retrieval

The problems of finding papers with similar images, graphs and equations have attracted sporadic attention. The approach taken in most work is first to extract the objects of interest from the paper and then either to classify them by type, or - for restricted domains - to use a combination of image and text features to retrieve similar objects.

Lu et al [Lu et al., 2007] investigated the classification of figures in scientific papers by extracting the figures from approximately 2000 papers selected from the CiteSeer index, discriminating between photographs (23% of the total number of figures), 2D plots, 3D plots, diagrams, and others (54% of the total number of figures). Using Li and Gray's algorithm [Li and Gray, 2000] they obtained good performance in distinguishing photographs from other figures. To distinguish between the non-photograph classes they used a multi-class SVM-based classifier with mixed results because of the large disparity in class sizes.

Comparison of a collection of images manually extracted from papers in the 2005 volume of *Radiology* with results from the Medical Image task of ImageCLEF suggest that there is considerable scope for improving retrieval in this domain by incorporating image data into the retrieval [Deserno et al., 2007]. Work on medical image retrieval using combinations of texture-based image features and image caption or descriptive text from a controlled vocabulary and a small manually-classified corpus showed that the best results were obtained when image and text features were combined with a heavy weighting towards the image features [Névéol et al., 2009].

Work on diagram recognition has mostly used information extraction approaches to develop effective recognition or extraction methods for particular domains or applications. The work on engineering drawings, much of which has focused on recognising symbols and reconstructing views, has been reviewed by Ablameyko and Uchida [Ablameyko and Uchida, 2007]. Examples include work on on sketched diagrams and equations [Lank et al., 2001], data extraction from tables [Liu et al., 2007], 2-D plots [Brouwer et al., 2008], chemical information [Mitra et al., 2007].

3 The SIFT algorithm

The Scale Invariant Feature Transform (SIFT) algorithm was developed by Lowe [Lowe, 2004] for object recognition from images. The technique extracts features that are invariant to changes in scale and rotation and are robust to changes in illumination, viewpoint, noise and affine distortion. SIFT points are stable local grey-scale minima and maxima. SIFT features for a page from our corpus are shown in Figure 1. Each feature is described by its location, magnitude, orientation and a 128-dimensional feature descriptor.

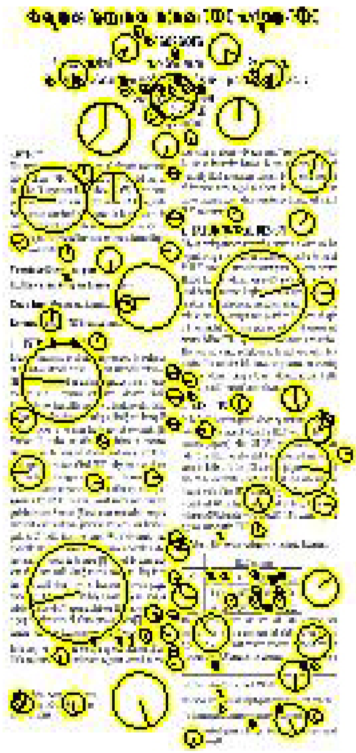

Figure 1: Example page showing SIFT features

The main steps in the SIFT algorithm are:

1. Scale-space extrema detection: Potential interest points are identified by using a difference of Gaussian function to identify points in the image that are invariant to scale and orientation. Each sample point in the image is compared with its eight immediate neighbours, and nine neighbours in the scale above and the scale below. Potential interest points are either larger or smaller than all their neighbours; each has a scale and location.

2. Keypoint localisation: Each potential interest point is examined and those which have low contrast or are poorly localised along an edge are discarded as they are not stable. At this stage there may also be an attempt to obtain sub-pixel resolution for the location for the position of the keypoints.

3. Orientation assignment: An orientation is calculated for each interest point based on local image gradients.

4. Descriptor generation: Histograms of the local image gradients are calcu-

lated, rotated relative to the feature orientation and normalised to reduce the effect of differences of illumination.

The algorithm provides one of the most successful methods of identifying and describing objects in images and has been applied to a wide range of problems. It is also computationally light in comparison to many alternative algorithms [Lowe, 2004] and, using current technology on a domestic PC it is possible to run at 10-15 frames per second on VGA video. However, SIFT performance can be an issue of on larger sets of images, where the matching step is a bottleneck because of the high dimensionality of the descriptor; this has led to many proposals to speed up matching based on approximate matching or – most notably – interest points based on Hessian matrices [Bay et al., 2008].

Several improvements have been proposed, such as the use of Gabor filters [Moreno et al., 2009], PCA [Ke and Sukthankar, 2004], and an implementation has been released as a Matlab toolbox [Vedaldi and Fulkerson, 2008]. An extension to deal with colour has been proposed by Li and Ma [Li and Ma, 2009].

SIFT feature magnitudes typically have a long tail distribution; Figure 2 shows the feature magnitude distribution for a typical page in the corpus. There is some evidence to suggest that the largest magnitude features may also not be good discriminators, although we found no evidence for this in a limited series of experiments.

Comparative evaluations showed that SIFT descriptors outperformed a large number of other interest-point techniques [Mikolajczyk and Schmid, 2003]. More recently Deselaers et al. [Deselaers et al., 2008] have compared a large number of approaches in content-based image retrieval (CBIR) and image classification tasks. The SIFT-based approach they used performed relatively poorly compared with other techniques that effectively exploited colour information. However, SIFT features have shown good performance in a wide range of applications where colour does not have a large impact on the retrieval.

3.1 SIFT applications

SIFT was originally developed for object recognition, matching features from an image against a database of features extracted from a training set of images. For this work Lowe [Lowe, 2004] accepted matches between features if the distance ratio between the nearest match and the second match was less than 0.8 and at least three features link the target object and an object in the database.

SIFT-based approaches have been used in a wide variety of image classification and matching applications. These include remote sensing to classify land use types [Yi and Newsam, 2008]; biometric identification [Ladoux et al., 2009]; CBIR in a museum context, where it substantially outperformed an existing

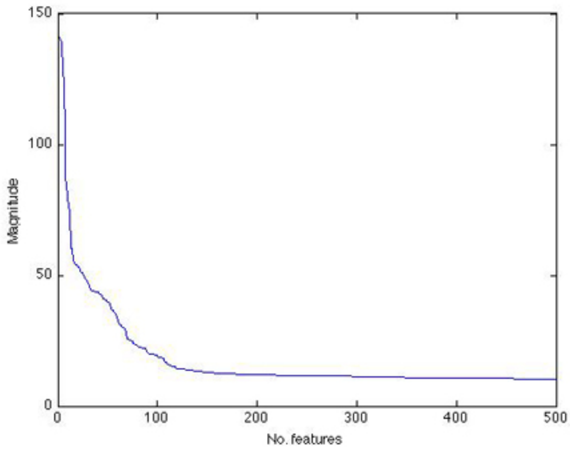

Figure 2: SIFT feature magnitude distribution

colour histogram-based system [Valle and Cord, 2006], retrieving images of buildings [Wangming et al., 2008], a near-duplicate detection system (using PCA-SIFT) which gave near-perfect results on a standard test set [Ke et al., 2004] and matching slides to presentation video [Fan et al., 2006].

Overall, these show clearly that SIFT descriptors perform well for a wide variety of image retrieval and matching tasks.

4 The corpus

The corpus used for these experiments is designed to provide (i) some clearly distinct groups of documents which should be easily recognisable, (ii) some similar documents from different sources, and (iii) other documents of similar appearance and varying relevance to the query topics. It is important the corpus should be small enough to perform image-based experiments within a reasonable time, as our initial aim is to demonstrate the potential of retrieval using visual features.

The corpus consists of computer vision and information retrieval conference papers from ACM SIGIR 2008 (207 papers, 898 pages), BMVC 2003 (51 papers, 511 pages), BMVC 2002 (19 papers, 218 pages) and Fisheries and Conservation (FC) 2006 (31 papers, 194 pages). The papers range from 1 to 42 pages; 142 papers have 8 or 10 pages and 147 (almost all from SIGIR) have 1 or 2 pages. There are 1821 pages from 308 papers in the corpus.

Ten standard papers were selected at random as query papers from the standard length (8–10 pages) papers: five from SIGIR, three from BMVC and two from FC to ensure coverage of the whole corpus.

Three relevance judgements for each remaining paper were obtained from members of an undergraduate information retrieval class working independently. Papers were scored as relevant (1.0), somewhat relevant (0.5) or not relevant (0.0). The individual judgements were averaged and to arrive at binary relevance judgements for the corpus on the query papers a minimum threshold of 0.5 was applied. This threshold ensures that the three judgements were either all "somewhat relevant" or were a mixture of"relevan" and "somewhat relevant". This gave a total of 169 papers that are relevant to one of the query papers.

A random selection of documents retrieved in response to the queries should have an average precision of 0.04, which we estimate by averaging the expected precision of random retrievals for each query (so accounting for differences in the numbers of documents relevant to each query).

The papers were prepared for the image experiments by first splitting the PDF files into their individual pages and converting them to portable graphics metafile (PGM) format, giving an equivalence between a page and an image. Second, each page image was resized so that the maximum dimension is no more than 600 pixels to reduce the number of features generated; a typical image used for these experiments is approximately 600 by 450 pixels. Third, the local interest features were extracted using Lowe's SIFT code (the SIFT extraction code in the Matlab toolbox developed by Vedali and Fulkerson gives equivalent results [Vedaldi and Fulkerson, 2008]), which gives an average of 1888 features per page image.

From this, data sets of the features with the largest magnitudes were created with maxima of 25, 50, 100, 150, 200, 250, 350, and 500 features per page. There are very few duplicated features (under 0.2%) in any of the datasets; this makes any attempt to reduce the computation using a feature-to-feature lookup table infeasible.

The performance measures we use are R-precision and average precision. These measures have been shown to be good indicators of overall retrieval performance [Buckley and Voorhees, 2005].

5 Text experiments

The baseline text experiments were undertaken using the Terrier information retrieval system [Ounis et al., 2007] with its default model parameters. The results of these experiments do not necessarily represent the best performance obtainable, but form a reasonable baseline. We ran experiments with several combinations of retrieval model and preprocessing steps, but the differences between them were small.

Several sets of experiments were conducted to establish appropriate queries, using varying combinations and selections of terms from the title and abstract.

Table 1: Text experiment results

Query terms	Av. Precision	R-precision
Title only	0.0857	0.1294
5 words and title	0.0874	0.1088
10 words and title	0.1019	0.1093
20 words and title	0.1096	0.0904
50 words and title	0.1233	0.1286
100 words and title	0.1300	0.1271
Whole abstract + title	0.1324	0.1333

The abstract of each query paper was tagged for parts of speech using the Stanford PoS tagger [Schütze, 1995] and all the closed-class words were removed, leaving nouns, verbs, adjectives and adverbs – which should be more descriptive of the content and are a better approximation to human-generated queries than fragments from the papers [Jansen and Spink, 2006]. We conducted experiments with the first n words from the abstract, with and without the title. These experiments showed that there is a steady increase in both average precision and R-precision as the query size increases (Table 1) and a small, but consistent, improvement when the title is added to the query.

A further set of experiments, using either the full text of the query documents or 100 words and the title, had an average precision and R-precision ranging between approximately 0 and 0.1, depending on the software used to convert the PDF files to text.

The results of these experiments suggest that text-based retrieval systems are badly affected by noise of the sort commonly encountered in text conversions of scientific documents. Typically, tables are reduced to meaningless sequences of numbers, graphs to just the captions and labels, and equations to arbitrary collections of symbols and characters; many of these types of content are discarded by information retrieval systems. Multiple column layouts are commonly treated as single columns, interspersed with the text left over from figures, captions and other non-text elements of the document, making any term proximity or sequence measures ineffective. Effective retrieval of scientific documents converted from PDF (or other document formats containing a lot of layout and representation information) requires a better representation of the text structure of the document than is obtained with most widely available convertors.

6 Image experiments

The algorithm we used for our experiments is based on nearest-neighbour (NN) distances. NN methods are sensitive to feature quantization [Boiman et al., 2008],

Table 2: Image experiments: average precision (rank measure)

NN	NF25	NF100	NF250	NF300	NF350	NF500
100	0.2356	0.2431	0.2818	0.2877	0.2695	0.3413
250	0.1939	0.259	0.2957	0.3291	0.3279	0.3872
500	0.1787	0.2598	0.2948	0.3298	0.3304	0.3706
1000	0.1661	0.2794	0.2965	0.3285	0.3263	0.3789
5000	0.1545	0.2738	0.3357	0.3134	0.3399	0.3983

so for these experiments we have used the raw SIFT features; further work is needed to determine the performance characteristics of SIFT features using different local region parameters, histograms and image sizes. We use cosine similarity as our measure of similarity. Given two vectors A and B,

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \, \|B\|}$$

The similarity of each page to all other pages in the collection was calculated as follows:

1. Calculate the cosine similarities of the SIFT features (i.e. comparing each feature in the corpus with each feature from the 10 documents in the query set),
2. Select the n nearest neighbours of each feature in the query set,
3. Rank the nearest neighbours in cosine similarity order,
4. Accumulate the scores for image-to-image (i.e. page-to-page) reciprocals of ranks and cosine similarities,
5. Calculate the scores for document-to-document similarities and normalise them for the number of features.

In general, the average precision of the rank measure increases with the number of features per image (NF) selected and with the number of nearest neighbours (NN); this is shown in Figure 3. There is an anomaly in that the average precision for 100-150 features/page is only slightly less than the maxima obtained with 500 features/page; further work is needed to determine whether this is a genuine performance peak, or an artefact related to the corpus.

We also experimented with varying the minimum number of features required for a match, using minima of 2, 3, 5, and 10 features for a match between images, and discovered that the precision is unaffected by raising the number of features required for a match above three.

The results of experiments using less than 100 features/image are uniformly poor and it is evident that they do not provide a useful description of the page

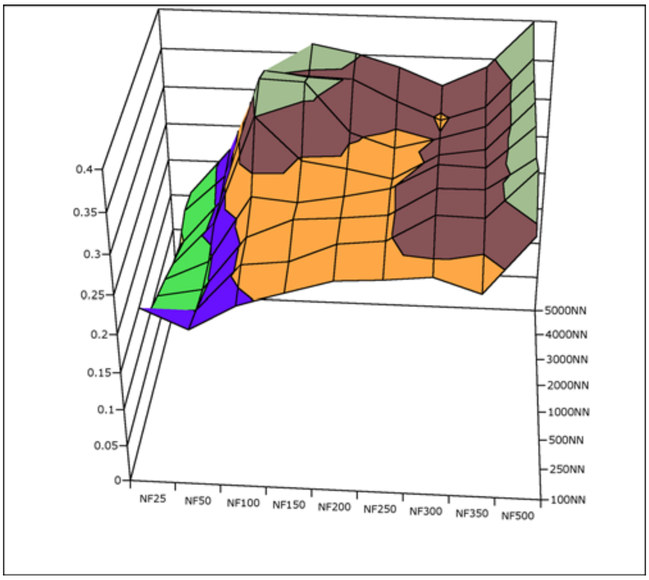

Figure 3: Precision for ranked results. The x-axis is the number of features per image, the z-axis is the number of nearest neighbours and the y-axis is the precision.

image. For the other experiments the average precision increases with the number of features per page used. However, the number of documents returned decreases, so recall decreases with higher numbers of features per page. The explanation for this appears to be that with higher numbers of features per page, density of nearest neighbours per document increases rapidly for the most similar documents and slightly less similar documents are not returned, as they have insufficient numbers of matching features.

The rank measure gave better average precision than the distance measure in the experiments with 100 or fewer features per page. The R-precision obtained from all the experiments was low (maximum 0.17) and declined sharply with both larger numbers of features per page and larger numbers of nearest neighbours. This is because experiments with larger numbers of features returned fewer documents and many of these experiments returned fewer documents than the number of relevant documents in the corpus for each query.

The average precision for the cosine similarity measure shows a clear peak region with 300-350 features/page and 100-1000 nearest neighbours, with a maximum of 0.43 (Figure 4); selected figures are also shown in Table 3. This level of retrieval performance clearly shows that retrieval based on image similarity can be comparable with text-based retrieval.

Table 3: Image experiments: average precision distance (cosine similarity)

	NF25	NF100	NF250	NF300	NF350	NF500
100	0.0838	0.2368	0.373	0.372	0.3954	0.2912
250	0.0953	0.2526	0.3901	0.4348	0.4329	0.3218
500	0.094	0.2537	0.3782	0.3872	0.4311	0.3483
1000	0.0877	0.2509	0.3902	0.3859	0.4129	0.3433
5000	0.1146	0.2384	0.293	0.3012	0.389	0.2836

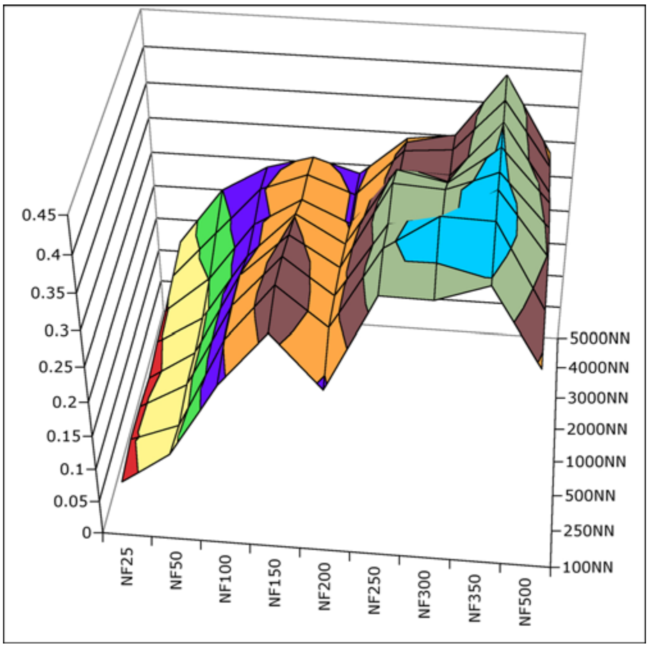

Figure 4: Precision for distance (cosine similarity) results. The X axis is the number of features per image, the Z axis is the number of nearest neighbours and the Y axis is the precision.

7 Discussion

We have looked in more detail at the individual results from the two best performing combinations: first, the cosine similarity results for 350 features per image and 500 nearest neighbours for each query feature, and second, the ranked results for 100 features per image and 5000 nearest neighbours for each query feature.

The NF350/NN500 cosine similarity experiment returned an average of 9.1 documents per query, of which 42% were relevant. The non-relevant documents in the results of this experiment almost all have similar figures and tables to the query documents. The relevant documents in the result set have fewer obvious visual similarities. A small test on these results suggests that identifying any document having similar figures and tables in the results as not relevant to the query document is generally accurate; further work is required to substantiate this. It seems likely that documents from different conferences (i.e. with a different page format) may be under-represented in the results, although a much larger heterogeneous corpus would be needed to determine whether this is the case. This analysis suggests, first, that the relevant documents are being returned because SIFT is describing similarities in the text (e.g. word shapes), second, the results would be improved if the non-text elements were treated separately from the text, and third, that the features with the largest magnitudes may be hindering the recognition of similar documents with different overall formatting. The experiments we conducted with the very largest magnitude features removed did not show any obvious improvements in the precision of the results; further experiments may determine if removing large magnitude features can consistently affect retrieval performance.

The NF100/NN5000 rank similarity experimental results had an average of 18.9 documents per query, of which 39% were relevant. There is no significant rank correlation between the documents retrieved in this experiment and those in the NF350/NN500 cosine similarity experiment . The larger number of documents retrieved results in a modest improvement in recall, and the non-relevant documents in the result set appear to share the same characteristics as those in the NF350/NN500 cosine similarity experiment – although to a less obvious degree.

8 Conclusions

We have shown that it is possible to perform useful retrieval - and hence classification - of scientific papers using visual features alone. The initial experiments have focused on retrieving whole documents, working within a conventional information retrieval evaluation framework. These experiments have shown that an average precision of over 0.4 is obtained in several experimental configurations. These are somewhat surprising results – text retrieval has virtually noise free data from which to work. Image data is corrupted by digitisation, differing choice of font, the appearance of graphics and differing formatting – a simple choice, hyphenation for example, can completely alter the appearance of a word. Yet the visual retrieval is successful probably because it can generate a large number of key points and ignore the ones that are non-informational.

There are three main elements of our future work. First, we will validate our results and replicate these experiments using one or more different corpora (depending on the availability of suitable collections with relevance judgements).

Second,to improve the retrieval and computational performance we will improve the efficiency of our code, and experiment with reduced feature sets, and look at alternatives to the vector space model for retrieval. Third, we will isolate the non-text items (images, graphs, equations, etc.) as the basis for discovering similarities in methods and mathematical underpinnings across a wider range of subject areas.

Acknowledgments

The authors would like to thank Sarah Hilder, Alberto Pastrana and Jake Newman for their help with this work.

References

[Ablameyko and Uchida, 2007] Ablameyko, S. and Uchida, S. (2007). Recognition of engineering drawing entities: Review of approaches. *Int. J. Image Graphics*, 7(4):709–733.

[Bay et al., 2008] Bay H., Ess A., Tuytelaars T., and Van Gool L. (2008) Speeded-Up Robust Features (SURF) *Computer Vision and Image Understanding*, 110(3), 346–359

[Boiman et al., 2008] Boiman, O., Shechtman, E., and Irani, M. (2008). In defense of nearest-neighbor based image classification. In *CVPR*. IEEE Computer Society.

[Brouwer et al., 2008] Brouwer, W., Kataria, S., Das, S., Mitra, P., and Giles, C. L. (2008). Automatic identification and data extraction from 2-dimensional plots in digital documents. *CoRR*, abs/0809.1802.

[Buckley and Voorhees, 2005] Buckley, C. and Voorhees, E. M. (2005). *Retrieval System Evaluation*, chapter 3, pages 53–75. MIT Press.

[Deselaers et al., 2008] Deselaers, T., Keysers, D., and Ney, H. (2008). Features for image retrieval: an experimental comparison. *Inf. Retr.*, 11(2):77–107.

[Deserno et al., 2007] Deserno, T. M., Antani, S., and Long, L. R. (2007). Exploring access to scientific literature using content-based image retrieval. *Proceedings of the International Society for Optical Engineering (SPIE)*, 6516:OL1–OL8.

[Fan et al., 2006] Fan, Q., Barnard, K., Amir, A., Efrat, A., and Lin, M. (2006). Matching slides to presentation videos using sift and scene background matching. In Wang, J. Z., Boujemaa, N., and Chen, Y., editors, *Multimedia Information Retrieval*, pages 239–248. ACM.

[Jansen and Spink, 2006] Jansen, B. J. and Spink, A. (2006). How are we searching the world wide web? a comparison of nine search engine transaction logs. *Inf. Process. Manage.*, 42(1):248–263.

[Ke and Sukthankar, 2004] Ke, Y. and Sukthankar, R. (2004). PCA-SFT: A more distinctive representation for local image descriptors. In *CVPR (2)*, pages 506–513.

[Ke et al., 2004] Ke, Y., Sukthankar, R., and Huston, L. (2004). An efficient parts-based near-duplicate and sub-image retrieval system. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 869–876, New York, NY, USA. ACM.

[Ladoux et al., 2009] Ladoux, P.-O., Rosenberger, C., and Dorizzi, B. (2009). Palm vein verification system based on SIFT matching. In *ICB 2009*, LNCS 5558, pages 1290–1298.

[Lank et al., 2001] Lank, E., Thorley, J. S., Chen, S., and Blostein, D. (2001). On-line recognition of UML diagrams. In *ICDAR*, pages 356–360. IEEE Computer Society.

[Li and Ma, 2009] Li, C. and Ma, L. (2009). A new framework for feature descriptor based on SIFT. *Pattern Recogn. Lett.*, 30(5):544–557.

[Li and Gray, 2000] Li, J. and Gray, R. M. (2000). Context-based multiscale classification of document images using wavelet coefficient distributions. *IEEE Transactions on Image Processing*, 9(9):1604–1616.

[Liu et al., 2007] Liu, Y., Bai, K., Mitra, P., and Giles, C. L. (2007). Tableseer: automatic table metadata extraction and searching in digital libraries. In Rasmussen, E. M., Larson, R. R., Toms, E., and Sugimoto, S., editors, *JCDL*, pages 91–100. ACM.

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

[Lu et al., 2007] Lu, X., Wang, J. Z., Mitra, P., and Giles, C. L. (2007). Automatic extraction of data from 2-d plots in documents. In *ICDAR*, pages 188–192. IEEE Computer Society.

[Mikolajczyk and Schmid, 2003] Mikolajczyk, K. and Schmid, C. (2003). A performance evaluation of local descriptors. In *CVPR (2)*, pages 257–263. IEEE Computer Society.

[Mitra et al., 2007] Mitra, P., Giles, C. L., Sun, B., and Liu, Y. (2007). Chemxseer: a digital library and data repository for chemical kinetics. In Mitra, P., Giles, C. L., and Carr, L., editors, *CIMS*, pages 7–10. ACM.

[Moreno et al., 2009] Moreno, P., Bernardino, A., and Santos-Victor, J. (2009). Improving the SIFT descriptor with smooth derivative filters. *Pattern Recognition Letters*, 30(1):18–26.

[Névéol et al., 2009] Névéol, A., Deserno, T. M., Darmoni, S. J., Güld, M. O., and Aronson, A. R. (2009). Natural language processing versus content-based image analysis for medical document retrieval. *JASIST*, 60(1):123–134.

[Ounis et al., 2007] Ounis, I., Lioma, C., Macdonald, C., and Plachouras, V. (2007). Research directions in terrier. *Novatica/UPGRADE Special Issue on Web Information Access, Ricardo Baeza-Yates et al. (Eds), Invited Paper.*

[Schütze, 1995] Schütze, H. (1995). Distributional part-of-speech tagging. In *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*, pages 141–148, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Valle and Cord, 2006] Valle, E. and Cord, M. (2006). CBIR in cultural databases for identification of images: A local-descriptors approach.

[Vedaldi and Fulkerson, 2008] Vedaldi, A. and Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms.

[Wangming et al., 2008] Wangming, X., Jin, W., Xinhai, L., Lei, Z., and Gang, S. (2008). Application of image SIFT features to the context of CBIR. In *CSSE '08: Proceedings of the 2008 International Conference on Computer Science and Software Engineering*, pages 552–555, Washington, DC, USA. IEEE Computer Society.

[Yi and Newsam, 2008] Yi, Y. and Newsam, S. (2008). Comparing SIFT descriptors and Gabor texture features for classification of remote sensed imagery. In *15th IEEE International Conference on Image Processing*, pages 1852–1855.