

FIRE in ImageCLEF 2005: Combining Content-based Image Retrieval with Textual Information Retrieval

Thomas Deselaers, Tobias Weyand, Daniel Keysers, Wolfgang Macherey, and Hermann Ney
Lehrstuhl für Informatik VI, RWTH Aachen University, Aachen, Germany
surname@informatik.rwth-aachen.de

Abstract

In this paper we describe the methods we used in the 2005 ImageCLEF content-based image retrieval evaluation. For the medical retrieval task, we combined several low-level image features with textual information retrieval. Combining these two information sources, clear improvements over using one of these sources alone are possible.

Additionally we participated in the automatic annotation task, where we used FIRE, our content-based image retrieval system, on the one hand and a subimage based method for object classification on the other hand.

The results achieved are very good. In particular, we obtained the first and the third rank in the automatic annotation task out of 44 submissions from 12 groups.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval
I.5 [Pattern Recognition]: I.5.4 Applications

Keywords

content-based image retrieval, object recognition, textual information retrieval

1 Introduction

It is known that in content-based image retrieval (CBIR) benchmarking of systems is a major problem. ImageCLEF, as part of the Cross language evaluation forum, is one major step in the direction of creating standard benchmarking tasks and setting up competitions to compare content-based image retrieval systems. One of the main conclusions that can be drawn from the 2004 ImageCLEF image retrieval evaluation is that textual information and user feedback can strongly improve the results if available. Especially if the queries are of semantic nature it is intrinsically difficult to solve them using visual information alone.

Especially in real life applications, as e.g. in medicine, where textual information is available and pictures alone are not sufficient to describe a medical case, any information available should be used. If, for example, the query image is a microscopic photo of a bacteria culture, a standard image retrieval system will easily find other pictures of bacteria cultures, but it will hardly be able to distinguish between different kinds of bacteria. With additional, textual query information, like "E. Coli bacteria", the query, and thus the result, is more precise.

As we obtained the best score in the category "visual information only, no user interaction" in the 2004 ImageCLEF evaluation, it was an interesting challenge to extend our FIRE system¹

¹<http://www-i6.informatik.rwth-aachen.de/~deselaers/fire.html>

towards using textual information.

Other groups already proposed their approaches of combining textual information retrieval and content-based image retrieval, e.g. [22, 19, 1, 28].

In this paper, we describe the techniques we used for the 2005 ImageCLEF evaluation, in particular how we combine textual information retrieval and content-based image retrieval.

In 2005 ImageCLEF, four tasks were available: a) automatic annotation, b) medical image retrieval, c) bilingual information retrieval, and d) interactive retrieval. We participated in the automatic annotation task and the medical image retrieval task. Our approach to the medical retrieval task is described in Section 2, the two approaches to the automatic annotation task are described in Section 3.

2 Medical Retrieval Task

For the medical retrieval task in the 2005 ImageCLEF Image Retrieval Evaluation, 25 queries were given. Each query was defined by a short textual query description and one to three example images. One query contained a negative example image, all other example images were positive. A more detailed description of the task and an overview on the results can be found in [3]. In the following we describe our setup of FIRE for the medical retrieval task in the 2005 ImageCLEF Image Retrieval Evaluation.

2.1 Decision Rule

Given a set of positive example images Q^+ and a (possibly empty) set of negative example images Q^- a score $S(Q^+, Q^-, X)$ is calculated for each image X from the database:

$$S(Q^+, Q^-, X) = \sum_{q \in Q^+} S(q, X) + \sum_{q \in Q^-} (1 - S(q, X)). \quad (1)$$

where $S(q, X)$ is the score of database image X with respect to query q and is calculated as $S(q, X) = e^{-\gamma D(q, X)}$ with $\gamma = 1.0$. $D(q, X)$ is a weighted sum of distances calculated as

$$D(q, X) := \sum_{m=1}^M w_m \cdot d_m(q_m, X_m). \quad (2)$$

Here, q_m and X_m are the m th feature of the query image q and the database image X , respectively. d_m is the corresponding distance measure and w_m is a weighting coefficient. For each d_m , $\sum_{X \in \mathcal{B}} d_m(Q_m, X_m) = 1$ is enforced by re-normalization.

Given a query (Q^+, Q^-) , the images are ranked according to descending score and the K images X with highest scores $S(Q^+, Q^-, X)$ are returned by the retriever.

Due to the lack of suitable training data, weights w_m were chosen heuristically based on experiences from earlier experiments with other data.

2.2 Textual Information Retrieval

To incorporate textual information in FIRE, we decided to use an existing textual information retrieval engine [20].

To incorporate textual information in FIRE, we decided to use an existing textual information retrieval engine that was developed at Lehrstuhl für Informatik VI. The text retrieval engine implements a variant of the Smart-2 retrieval metric, which is based on the well-known term frequency inverse document frequency (tf-idf) metric. First, the textual information is preprocessed by removing function words that are considered to be of no importance for the actual retrieval process (so called stopping). The stop word list used for this purpose comprises 319 of the most frequently occurring function words in the English language. Subsequent to the stopping process, the remaining words are reduced to their stems using Porter’s stemming algorithm [23]. The

stemmed words form the index terms that are used in order to index the text documents provided with the image data. In our implementation of the Smart-2 retrieval metric we use the following definition of the inverse document frequency:

$$\text{idf}(t) := \log \left\lfloor \frac{K}{n(t)} \right\rfloor \quad (3)$$

Here, t denotes an index term, and K is the number of text documents. Due to the floor operation in Eq. (3) a term weight will be zero if it occurs in more than half of the documents. According to [2], each index term t in a document \mathbf{d} is associated with a weight $g(t, \mathbf{d})$ that depends on the ratio of the logarithm of the term frequency $n(t, \mathbf{d})$ to the logarithm of the average term frequency $\bar{n}(\mathbf{d})$

$$g(t, \mathbf{d}) := \begin{cases} [1 + \log n(t, \mathbf{d})] / [1 + \log \bar{n}(\mathbf{d})] & \text{if } t \in \mathbf{d} \\ 0 & \text{if } t \notin \mathbf{d} \end{cases} \quad (4)$$

with $\log 0 := 0$ and

$$\bar{n}(\mathbf{d}) = \frac{\sum_{t \in \mathcal{T}} n(t, \mathbf{d})}{\sum_{t \in \mathcal{T}: n(t, \mathbf{d}) > 0} 1} \quad (5)$$

The logarithms in Eq. (4) prevent documents with high term frequencies from dominating those with low term frequencies. In order to obtain the final term weights, $g(t, \mathbf{d})$ is divided by a linear combination between a pivot element c and the number of singletons $n_1(\mathbf{d})$ in document \mathbf{d} :

$$\omega(t, \mathbf{d}) := \frac{g(t, \mathbf{d})}{(1 - \lambda) \cdot c + \lambda \cdot n_1(\mathbf{d})} \quad (6)$$

with $\lambda = 0.2$ and

$$c = \frac{1}{K} \sum_{k=1}^K n_1(\mathbf{d}_k) \quad \text{and} \quad n_1(\mathbf{d}) := \sum_{t \in \mathcal{T}: n(t, \mathbf{d}) = 1} 1 \quad (7)$$

Unlike tf-idf, only query terms are weighted with the inverse document frequency $\text{idf}(t)$:

$$\omega(t, \mathbf{q}) = [1 + \log n(t, \mathbf{q})] \cdot \text{idf}(t) \quad (8)$$

The SMART-2 retrieval function is then defined as the product over the document and query specific index term weights:

$$f(\mathbf{q}, \mathbf{d}) = \sum_{t \in \mathcal{T}} \omega(t, \mathbf{q}) \cdot \omega(t, \mathbf{d}) \quad (9)$$

To use the textual information for image retrieval, each image has to be attached to at least one (possibly empty) text document. These text documents are used in the above described image retrieval process. To determine the distance $d_{\text{text}}(q_m, X_m)$ between a query image q with query text q_m and a database image X with attached text X_m , first the textual information retriever is queried using the query text. Then, the textual information retriever returns the list of all relevant documents from the database. These documents are ranked by the retrieval status values (RSV) R which is high for documents similar to the query and low for dissimilar documents. The distance $d(q_m, X_m)$ is then calculated as

$$d_{\text{text}}(q_m, X_m) = \begin{cases} R_{\max} - R_X & \text{if } X \text{ is in the list of relevant documents} \\ \rho & \text{otherwise} \end{cases} \quad (10)$$

where R_{\max} is the maximum of all returned RSVs, R_X is the RSV attached to image X , q_m and X_m are the query text and the text attached to image X , respectively, and ρ is a constant chosen sufficiently large to make sure that images whose texts do not appear in the list of relevant objects have high distances. Note, the case where $\rho = R_{\max}$ corresponds to assigning RSV of 0 to all non-relevant texts.

The resulting distances $d_{\text{text}}(q_m, X_m)$ are used in the retrieval process described in the previous section.

2.3 Image Features

In the following we describe the image features we used in the evaluation, these features are extracted offline from all database images.

2.3.1 Appearance-based Image Features

The most straight-forward approach is to directly use the pixel values of the images as features. For example, the images might be scaled to a common size and compared using the Euclidean distance. In optical character recognition and for medical data improved methods based on image features usually obtain excellent results [15, 16, 17].

In this work, we used 32×32 versions of the images, these were compared using Euclidean distance. It has been observed, that for classification and retrieval of medical radiographs, this method saves as a not-top-bad baseline.

2.3.2 Color Histograms

Color histograms are widely used in image retrieval [4, 10, 24, 26]. Color histograms are one of the most basic approaches and to show performance improvements, image retrieval systems often are compared to a system using only color histograms. The color space is partitioned and for each partition the pixels with a color within its range are counted, resulting in a representation of the relative frequencies of the occurring colors. In accordance with [24], we use the Jeffrey divergence to compare histograms.

2.3.3 Tamura Features

In [27] the authors propose six texture features corresponding to human visual perception: *coarseness*, *contrast*, *directionality*, *line-likeness*, *regularity*, and *roughness*. From experiments testing the significance of these features with respect to human perception, it was concluded that the first three features are very important. Thus in our experiments we use coarseness, contrast, and directionality to create a histogram describing the texture [4] and compare these histograms using the Jeffrey divergence [24]. In the QBIC system [10] histograms of these features are used as well.

2.3.4 Global Texture Descriptor

In [4] a texture feature consisting of several parts is described: *Fractal dimension* measures the roughness or the crinkliness of a surface. In this work the fractal dimension is calculated using the reticular cell counting method [13]. *Coarseness* characterizes the grain size of an image. Here it is calculated depending on the variance of the image. *Entropy* is used as a measure of disorder or information content in an image. The *Spatial gray-level difference statistics* (SGLD) describes the brightness relationship of pixels within neighborhoods. It is also known as co-occurrence matrix analysis [14]. The *Circular Moran autocorrelation function* measures the roughness of the texture. For the calculation a set of autocorrelation functions is used [12].

2.3.5 Invariant Feature Histograms

A feature is called invariant with respect to certain transformations if it does not change when these transformations are applied to the image. The transformations considered here are translation, rotation, and scaling. In this work, invariant feature histograms as presented in [25] are used. These features are based on the idea of constructing features invariant with respect to certain transformations by integration over all considered transformations. The resulting histograms are compared using the Jeffrey divergence [24]. Previous experiments have shown that the characteristics of invariant feature histograms and color histograms are very similar and that invariant feature histograms often outperform color histograms [5]. Thus, in this work color histograms are not used.

Table 1: Error rates obtained using different features on the IRMA 10000 Validation data.

feature	distance	dev corpus	test corpus
32×32 thumbnails	Euclidean	25.3	36.8
X×32 thumbnails	IDM	13.0	12.6
tamura texture histogram	JSD	33.1	46.0

3 Automatic Annotation Task

In the automatic annotation task, the objective was to classify 1,000 images into one of 57 classes using 9,000 training images. We participated in this using two different methods. Method A is identical to the approach we have chosen for the medical retrieval task, except that here no textual information was available, and that we used appearance-based image features and Tamura Texture Features only, as we know from earlier experiments that these features perform good on medical radiographs [18].

Method B is a general object recognition method using histograms of image patches and discriminative training of log-linear models [6, 7].

The parameters of method A were optimized using 1,000 images from the 9,000 training images as development set and the remaining 8,000 images for training. The parameters of method B were chosen as they work best on the Caltech Database [11, 8, 7].

A more detailed description of the task and a detailed analysis of the results can be found in [3].

3.1 Method A: Image Distortion Model

Method A uses our CBIR system FIRE, and a subset of the above described features consisting of thumbnails of the images of the sizes 32×32 and $X \times 32$ and Tamura Texture Histograms. Error rates for using these features alone are given in Table 1.

Some experiments with different weightings of Tamura features and thumbnails on our development corpus have shown that using the image distortion model alone outperforms the combinations. In particular the combination of image distortion model (weighted 5) and Tamura texture features (weighted 2) is interesting, as this performed best in previous experiments on smaller versions of the IRMA database [18]. In our experiments, this combination yielded an error rate of 13.5% on the development corpus, and it achieves a very good error rate of 13.2% on the test data. Based on these results we decided to use the image distortion model for our submission.

3.2 Method B: Object Recognition with Subimages and Discriminative Training

For method B we used the object recognition and classification approach using histograms of image patches and maximum Entropy training to classify the 1000 test images as described in [6, 7].

To reduce the time and memory requirements for the clustering process, we used only 4000 images for estimating the Gaussian mixture model and created the histograms for the remaining 5000 training images using this mixture model. For the discriminative training of the log-linear model, we used all training histograms.

The model submitted used multi-scale features where the first PCA component was discarded to account for brightness changes and 4096-dimensional histograms. This combination was reported to work best on the Caltech database [11] and in the PASCAL Visual Object Classes Challenge [9]. The model achieved an error rate of 13.9% and thus is slightly better than the model by Raphaël Marée who follows a similar approach [21].

4 Experiments and Results

In the following we describe the exact setup of the submitted runs to the automatic annotation task and the medical retrieval task and discuss the results. Furthermore, we discuss our methods, points to errors we made, and present results with errors corrected.

4.1 Automatic Annotation Task

Our submission using model A obtained the first rank in the automatic annotation task. The submission following the object recognition approach obtained the third rank. In total, 44 runs were submitted by 12 groups. The second rank was obtained by the IRMA group² using a similar approach to our model A and the fourth rank was obtained by University of Liège, Belgium using an approach with image patches and boosted decision trees. A clear improvement over the baseline result of 36.8% error rate can be observed. This baseline results is obtained by a nearest neighbor classifier using 32x32 thumbnails of the images and Euclidean distance.

4.2 Medical Retrieval Task

For the medical retrieval task, we used the features described in Section 2.3 with different weightings in combination with text features. In total, we submitted 10 runs which are shortly described here.

Runs using textual information only: We submitted two full-automatic runs, where only textual information was used. These runs were labelled **En** and **EnDeFr**. In **En** only the English texts were used, for **EnDeFr** the English, the German, and the French texts were used and combined with equal weights.

Runs using visual information only: We submitted three full-automatic runs, where only visual information was used. The runs 5000215, 0010003, and 1010111 only differ in the weighting of the image features. The exact weightings can be seen in Table 2. The run labelled 5000215 uses exactly the same setting as our submission to the 2004 ImageCLEF evaluation which had the best score from all 23 submissions in the category “visual features only, no user interaction”. From the bad score of 0.06, it can be seen that this years tasks differ strongly from last year’s task.

Runs using visual and textual information: We submitted three full-automatic runs and two runs with relevance feedback where textual and visual information was used. For the run **i6-3010210111**, the features were combined in exactly the way described above. For the runs **i6-3(1010111-min(111))** and **i6-3(3030333)-min(111)** before combining the text- and the visual features, first the minimum distance of all three text distances was taken for each image, to better account for images that have texts in one language only.

The runs **i6-vistex-rfb1** and **i6-vistex-rfb2** used relevance feedback from the first 20 results of the automatic run **i6-3(1010111-min(111))** and differ only in the user feedback. In both cases the feedback was given by a computer scientist familiar to the FIRE system with little background in medicine. Furthermore, the textual information was not available for the user feedback, thus the feedback is based on visual information only.

Table 2 gives an overview of all runs we submitted to the medical retrieval runs. Unfortunately, in advance to the competition we were unable to test our combination of textual- and visual information retrieval, which led to a very unlucky choice of ρ in Eq. 10, such that any combination with textual information retrieval was strongly disturbed. Furthermore Table 2 gives results of experiments we performed after the evaluation where ρ was chosen properly and it can clearly be seen that the results are much better. In particular, using English textual information retrieval only we could reach a MAP of 0.25 which would have been a third rank in the 2005 ImageCLEF evaluation in the category “textual and visual information, no relevance feedback”.

²<http://www.irma-project.org>

Table 2: Runs submitted to the medical retrieval task together with feature weightings and achieved MAP with wrongly chosen ρ and with properly chosen ρ . * means that the minimum among all lines marked with * in this column was taken and weighted by 1.

run	textual information only		visual information only			visual and textual information				+relevance feedback	
	En	EnDeFr-min	1010111	5000215	0010003	3010210111	3(3030333)-min(111)	3(1010111)-min(111)	-	vistex-rfb1	vistex-rfb2
X×32 image features	-	-	1	5	3	3	9	3	1	1	1
32×32 image features	-	-	1	0	0	0	0	3	1	1	1
color histograms	-	-	1	0	1	1	9	3	1	1	1
tamura features	-	-	1	2	0	2	9	3	1	1	1
invariant feat. histo.	-	-	1	1	0	1	9	3	1	1	1
English text	1	*	-	-	-	1	*	*	2	*	*
German text	0	*	-	-	-	1	*	*	0	*	*
French text	0	*	-	-	-	1	*	*	0	*	*
relevance feedback	-	-	-	-	-	-	-	-	-	+	+
score w/ wrong ρ	0.21	0.05	0.07	0.06	0.05	0.07	0.07	0.06	-	0.09	0.08
score w/ properly chosen ρ	0.21	0.15				0.22		0.20	0.25		

5 Conclusion and Outlook

In this paper we presented the methods we used in the 2005 ImageCLEF CBIR evaluation. We participated in the automatic annotation task, where we obtained the first and the third rank, and we participated in the medical image retrieval task, where our results were quite bad due to wrong settings, results with correct settings are presented in this work, and it can be seen that the method of combining textual information retrieval and content-based image retrieval performs very well. In particular, the result obtained would have been ranked 3rd in the medical retrieval task in the category “full automatic runs using textual and visual information”.

References

- [1] C. Alvarez, A. I. Oumohmed, M. Mignotte, and J.-Y. Nie. Toward Cross-Language and Cross-Media Image Retrieval. In *Multilingual Information Access for Text, Speech and Images. Proceedings of the 5th Workshop of the Cross-Language Evaluation Forum. CLEF 2004*, volume 3491 of *LNCS*, Springer, Bath, UK, pages 676–687, September 2004.
- [2] J. Choi, D. Hindle, J. Hirschberg, I. Magrin-Changnonleau, C. Nakatani, F. Pereira, A. Singhal, and S. Whittaker. An Overview of the AT&T Spoken Document Retrieval. In *Proc. 1998 DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, Va, USA, pages 182–188, February 1998.
- [3] P. Clough, T. Deselaers, M. Grubinger, W. Hersh, T. Lehmann, and H. Mueller. The CLEF 2005 Cross-Language Image Retrieval Track. In *CLEF 2005 Workshop Working notes*, Vienna, Austria, September 2005.
- [4] T. Deselaers. Features for Image Retrieval. Diploma thesis, Lehrstuhl für Informatik VI, RWTH Aachen University, Aachen, Germany, December 2003.
- [5] T. Deselaers, D. Keysers, and H. Ney. Features for Image Retrieval – A Quantitative Comparison. In *DAGM 2004, Pattern Recognition, 26th DAGM Symposium*, number 3175 in *LNCS*, Tübingen, Germany, pages 228–236, September 2004.

- [6] T. Deselaers, D. Keysers, and H. Ney. Discriminative Training for Object Recognition using Image Patches. In *CVPR 05*, volume 2, San Diego, CA, pages 157–162, jun 2005.
- [7] T. Deselaers, D. Keysers, and H. Ney. Improving a Discriminative Approach to Object Recognition using Image Patches. In *DAGM 2005*, LNCS, Vienna, Austria, page in press, August/September 2005.
- [8] P. Dreuw, D. Keysers, T. Deselaers, and H. Ney. Gesture Recognition Using Image Comparison Methods. In *GW 2005, 6th Int. Workshop on Gesture in Human-Computer Interaction and Simulation*, Vannes, France, May 2005.
- [9] M. Everingham, L. V. Gool, C. Williams, and A. Zisserman. Pascal Visual Object Classes Challenge Results. Technical report, University of Oxford, Oxford, UK, April 2005.
- [10] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and Effective Querying by Image Content. *Journal of Intelligent Information Systems*, 3(3/4):231–262, July 1994.
- [11] R. Fergus, P. Perona, and A. Zissermann. Object Class Recognition by Unsupervised Scale-Invariant Learning. In *Conference on Computer Vision and Pattern Recognition*, Blacksburg, VG, pages 264–271, June 2003.
- [12] Z. Q. Gu, C. N. Duncan, E. Renshaw, M. A. Mugglestone, C. F. N. Cowan, and P. M. Grant. Comparison of Techniques for Measuring Cloud Texture in Remotely Sensed Satellite Meteorological Image Data. *Radar and Signal Processing*, 136(5):236–248, October 1989.
- [13] P. Haberäcker. *Praxis der Digitalen Bildverarbeitung und Mustererkennung*. Carl Hanser Verlag, München, Wien, 1995.
- [14] R. M. Haralick, B. Shanmugam, and I. Dinstein. Texture Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, November 1973.
- [15] D. Keysers, C. Gollan, and H. Ney. Classification of Medical Images using Non-linear Distortion Models. In *Bildverarbeitung für die Medizin*, Berlin, Germany, pages 366–370, March 2004.
- [16] D. Keysers, C. Gollan, and H. Ney. Local Context in Non-linear Deformation Models for Handwritten Character Recognition. In *International Conference on Pattern Recognition*, volume 4, Cambridge, UK, pages 511–514, August 2004.
- [17] D. Keysers, W. Macherey, H. Ney, and J. Dahmen. Adaptation in Statistical Pattern Recognition using Tangent Vectors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):269–274, February 2004.
- [18] T. M. Lehmann, M.-O. Güld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, H. Ney, and B. Wein. Automatic Categorization of Medical Images for Content-based Retrieval and Data Mining. *Computerized Medical Imaging and Graphics*, 29:in press, 2005.
- [19] W.-C. Lin, Y.-C. Chang, and H.-H. Chen. From Text to Image: Generating Visual Query for Image Retrieval. In *Multilingual Information Access for Text, Speech and Images. Proceedings of the 5th Workshop of the Cross-Language Evaluation Forum. CLEF 2004*, volume 3491 of LNCS, Springer, Bath, UK, pages 664–675, September 2004.
- [20] W. Macherey, H.-J. Viechtbauer, and H. Ney. Probabilistic Aspects in Spoken Document Retrieval. *EURASIP Journal on Applied Signal Processing*, Special Issue on "Unstructured Information Management from Multimedia Data Sources" (2):1–12, February 2003.

- [21] R. Marée, P. Geurts, J. Piater, and L. Wehenkel. Random Subwindows for Robust Image Classification. In C. Schmid, S. Soatto, and C. Tomasi, editors, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, San Diego, CA, USA, IEEE, pages 34–40, June 2005.
- [22] H. Müller and A. Geissbühler. How to Visually Retrieve Images from the St. Andrews Collection Using GIFT. In *Multilingual Information Access for Text, Speech and Images. Proceedings of the 5th Workshop of the Cross-Language Evaluation Forum. CLEF 2004*, volume 3491 of LNCS, Springer, Bath, UK, pages 633–642, September 2004.
- [23] M. F. Porter. An algorithm for suffix stripping, July 1980. Programm.
- [24] J. Puzicha, Y. Rubner, C. Tomasi, and J. Buhmann. Empirical Evaluation of Dissimilarity Measures for Color and Texture. In *International Conference on Computer Vision*, volume 2, Corfu, Greece, pages 1165–1173, September 1999.
- [25] S. Siggelkow. *Feature Histograms for Content-Based Image Retrieval*. PhD thesis, University of Freiburg, Institute for Computer Science, Freiburg, Germany, 2002.
- [26] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval: The End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.
- [27] H. Tamura, S. Mori, and T. Yamawaki. Textural Features Corresponding to Visual Perception. *IEEE Transaction on Systems, Man, and Cybernetics*, 8(6):460–472, June 1978.
- [28] M. van Zaanen and G. de Croon. Multi-model Information Retrieval Using FINT. In *Multilingual Information Access for Text, Speech and Images. Proceedings of the 5th Workshop of the Cross-Language Evaluation Forum. CLEF 2004*, volume 3491 of LNCS, Springer, Bath, UK, pages 728–739, September 2004.