

Content-Based Queries on the CasImage Database Within the IRMA Framework

A Field Report

Christian Thies, Mark Oliver Güld, Benedikt Fischer,
and Thomas M. Lehmann

Department of Medical Informatics, University of Technology Aachen,
Pauwelsstrae 30, 52057 Aachen, Germany
`cthies@mi.rwth-aachen.de`

Abstract. Recent research has suggested that there is no general similarity measure, which can be applied on arbitrary databases without any parameterization. Hence, the optimal combination of similarity measures and parameters must be identified for each new image repository. This optimization loop is time consuming and depends on the experience of the designer as well as the knowledge of the medical expert. It would be useful if results that have been obtained for one data set can be transferred to another without extensive re-design. This transfer is vital if content-based image retrieval is integrated into complex environments such as picture archiving and communication systems. The image retrieval in medical applications (IRMA) project defines a framework that strictly separates data administration and application logic. This permits an efficient transfer of the data abstraction of one database on another without re-designing the software. In the ImageCLEF competition, the query performance was evaluated on the CasImage data set without optimization of the feature combination successfully applied to the IRMA corpus. IRMA only makes use of basic features obtained from grey-value representations of the images without additional textual annotations. The results indicate that transfer of parameterization is possible without time consuming parameter adaption and significant loss of retrieval quality.

1 Introduction

Classical architectures of content-based image retrieval (CBIR) systems consist of an image repository, along with visualization tools and query functionality. The principle of data storage and visualization does not vary notably among different databases whereas classes of retrieval approaches are differentiated by the query principle. In medical applications, this becomes a fundamental question since several requirements for data entry, retrieval time, and content representation must be considered [1, 2].

The first class of approaches associates and stores the secondary annotated content descriptions with each image. When textual information is used, the

retrieval task becomes a text search in the descriptions. Since the information is added by a human observer, it can be subjective based on the annotator and reflects his semantical view on the image content. However, inter- and intra-individual variances in perception, knowledge and capability of expression lead to different descriptions for a single issue, as do homonyms and synonyms in standardized medical language. Due to the fact that each image has to be categorized manually, the effort at data entry time is high, which is infeasible in clinical applications. Furthermore, there is no means of objective verification for the added data.

The second class of retrieval approaches overcomes the data entry problem by making use of information that is exclusively contained in the image. Here, the retrieval task is the detection of the nearest neighbors to the query image in the image database. This is based on the similarity of abstract representations of images in a feature space. Consequently, CBIR depends on an appropriate selection of the similarity measure, which again depends on the considered image features. In general, explicit selection of feature computations and similarity measures cannot be done by a physician in clinical routine. Thus, an abstraction from the low-level feature handling is required [3].

Besides the data entry cost and the content representation a third problem arises from the state of the image database: In a clinical environment the set of available images is continuously growing. This must be considered since one aims at CBIR to make clinical routine data available as a source of knowledge for education and diagnostics. Consequently, the feature selection and similarity computation must be as flexible as possible [4].

The image retrieval in medical applications (IRMA) project integrates these three aspects of query design into a single framework [5]. In this paper, the application of the IRMA framework to the previously unknown CasImage database of the University Hospitals of Geneva [6] is described with respect to the Image-CLEF competition. This work has two main goals. It is verified if it is possible to transfer the IRMA query approach to another domain without significant loss of retrieval quality and if it is sufficient to focus on basic image features for content description in contrast to text or text/feature combinations.

2 Designing a Query in the IRMA Framework

2.1 Feature Computation

Numerous features are described in the literature. They are roughly categorized into shape [7], color [8] and texture [9]. Those features are extracted from an image and form a size-reduced representation of the content. The first task in query design is the definition of relevant features. With respect to the large variety it is useful to provide as many features as possible and to select an appropriate subset for a distinct task. In the IRMA system, each newly presented image is automatically transformed into all available feature representations. This causes computational time for initial database processing but ensures short update and querying cycles when presenting unknown single images as well as

implementing new features since only the new information must be calculated. For this purpose, IRMA provides an automated storage concept that applies the implemented image-to-feature mappings at data entry time [10]. A new feature computation is integrated into an image processing chain by providing the transformation code. The actual database handling is hidden from this implementation by an interface providing an exclusive view on the image.

2.2 Feature Comparison

A query is modeled as a nearest neighbor classification. Consequently, it is designed by defining a similarity measure as a metric in the feature space. The corpus is represented as a sorted list, where the most similar images with the smallest distance to the query image form the head. The selection of actually correct matches must be made from this list. This is task specific and, therefore, a runtime parameter of the system that cannot be determined in advance. Furthermore, the selection depends on the actual number of relevant images in the database. Thus, a sufficient set of results must be presented to the user who has to make the final decision. For this purpose, the IRMA system offers a set of database processing sequences, which enable the sequential or parallel access to the stored image features via iteration or fan-in/fan-out processing over the corpus. Those sequences are combined on a binary execution level by abstract methods. Consequently, the data handling is also hidden from the application by specialized data flow interfaces [10].

2.3 Integration of New Image Data

The component-based software architecture provides a platform where new data is integrated without re-implementing the available features and distance measures. New images typically require the adoption of existing feature extractions and similarity computations. Yet when introducing a new image corpus there is no a-priori knowledge on the classes of images and their sizes. Therefore, finding the appropriate features and similarity measures equals the optimization task to find an unknown target function. However, brute force learning approaches for optimal query parameters can only be performed if the ground truth is known. Alternatively, a manual optimization of query methods to a distinct database is inapplicable for clinical routine solutions, since there is simply no time to supervise the learning process. Thus, the designer of an application combines a set of features and similarity methods in advance and then hands it over to the medical expert who has to verify the results. Once such a retrieval engine is integrated, for instance, into a picture archiving and communication system (PACS), it can hardly be modified or optimized, since the database continuously evolves. In contrast, the IRMA framework allows a hot swap of the feature extraction, similarity computation and database without affecting each other [10].

In case of the CasImage data set there was no ground truth given, so the results could only be generated by transfer of successfully applied query settings from other applications. This was performed for the ImageCLEF task. Main objective was the transfer of already implemented code and associated experience from recent experiments onto a new domain without parameter adaption.

3 Applied Features and Queries

The methods were taken from recent applications on the IRMA database consisting of 10,000 images from clinical routine, which were categorized by medical experts and used to train parameters [5]. Thus, the unmodified transfer of the methods is reasonable. Since the IRMA system processes only gray-scale images, RGB color conversion was done by using the standard color weighting [11]:

$$Y = \frac{6969 \cdot R + 23434 \cdot G + 2365 \cdot B}{32768} \quad (1)$$

Recent experiments indicated that spatial and intensity features must be considered equally to obtain reasonable results [12, 13].

3.1 Texture Features by TAMURA

TAMURA et al. use coarseness, contrast and directionality to capture an image's texture properties [14]. Those features are computed per pixel and reflect the texture affiliation. The value ranges for coarseness, contrast and directionality are quantized into 6, 8 and 8 equidistant intervals, respectively. They form the $6 \times 8 \times 8 = 384$ bins of a three-dimensional histogram, which serves as the global texture description. However, different image sizes result in different and therefore incomparable histogram counts. To obtain comparable features, each image is scaled to a size of 256×256 pixels, ignoring the aspect ratio.

To compare the TAMURA histograms of two images $H_T(Q)$ and $H_T(R)$ with $M = 384$ bins each, the Jensen-Shannon divergence is used [15], where Q and R denote the query and verified image, respectively:

$$D_{JS}(Q, R) = \frac{1}{2} \sum_{m=1}^M \left[H_T^m(Q) \log \frac{2H_T^m(Q)}{H_T^m(Q) + H_T^m(R)} + H_T^m(R) \log \frac{2H_T^m(R)}{H_T^m(Q) + H_T^m(R)} \right] \quad (2)$$

3.2 Aspect Ratio

Comparing the aspect ratio of images is an unspecific measure. Yet it is useful to consider the dimension of images. Since normalization for some texture features requires the deformation of the image dimensions into a square shape, the aspect ratio of an original image is a means of image comparison. Furthermore, the aspect ratio is characteristic for different classes of medical images. For instance, slices from magnetic resonance imaging (MRI) have identical edge dimensions while radiographs of limbs are rectangular elongated in direction of the principal bone. The aspect ratio is compared by:

$$D_{AR}(Q, R) = \left| \frac{X(Q)}{Y(Q)} - \frac{X(R)}{Y(R)} \right| \quad (3)$$

where $X(I)$ and $Y(I)$ denote the size of an image I : $(X, Y) := x \in \{0..X-1\}, y \in \{0..Y-1\}$.

3.3 Image Distortion Model

While histogram-based methods provide invariance against some transforms such as translation, scaled representations of the original images can preserve spatial properties, which are especially important to recognize medical images [16]. A drastic reduction in size also reduces noise and small image defects. The image distortion model (IDM) expands the naive pixel-by-pixel comparison of the scaled representations. It allows local displacements for each pair of pixels compared within the distance measure. This is especially useful for medical images due to individual anatomical properties in each image. The policy is to match each pixel of the sample image to one in the reference image. This ensures that all sample information is evaluated. To prevent a completely unordered vector field of pixel mappings between two images, it is useful to include the local context into the search process for a correspondence hypothesis. Denoting the coordinate offsets by x'' and y'' , while x' and y' term the offsets within the search window for a corresponding pixel, the distance is computed by:

$$D_{\text{IDM}}(Q, R) = \sum_{x=1}^X \sum_{y=1}^Y \min_{|x'|, |y'| \leq W_1} \left\{ \sum_{|x''|, |y''| \leq W_2} \|R(x + x' + x'', y + y' + y'') - Q(x + x'', y + y'')\|_2 \right\} \quad (4)$$

The results are improved if the image gradient is used instead of the intensity values. For our experiment, we used $W_1 = 2$ (5×5 pixel-sized search window for corresponding pixels) and $W_2 = 1$ (3×3 pixels of local context). The images were scaled to a fixed maximal height or width of 32 pixels keeping their original aspect ratio.

3.4 Classifier Combination

A parallel classifier combination is used. In order to avoid value domination of a single large addend, the results of each classifier are transformed to a common scale. This is done by dividing each result for a single classifier by the sum of all distances of the respective classifier. The weighting for each addend determines the combined vote for a distinct classifier. The described similarity measure is finally obtained from:

$$\rho(Q, R) = \alpha \cdot D_{\text{JS}}(Q, R) + \beta \cdot D_{\text{IDM}}(Q, R) + \gamma \cdot D_{\text{AR}}(Q, R) \quad (5)$$

As a matter of fact α , β and γ are parameters of the function ρ . Yet for the retrieval application described in this paper they are considered as constants that were empirically determined beforehand on the IRMA medical image corpus.

3.5 Determination of Relevance

The relevance of an image with respect to the query image Q is computed by sorting the database DB into a sequence:

$$S_{\text{REL}}(Q, B) = (R_1..R_n) | \rho(Q, R_1) \leq \dots \leq \rho(Q, R_n), R_1..R_n \in B, n = |B| \quad (6)$$

$n = |B|$ denotes the number of images in the repository. Relevance determination applies the classifier to all elements of the database. Consequently, time consuming ρ -functions are computed for many irrelevant comparisons. For this purpose, a sieve is applied to reduce the number of potentially relevant references. It is computed by selecting a fixed number of elements from the beginning of a list, which has been sorted with respect to the applied similarity measure. In the IRMA framework, the sieve is applied to the IDM classifier by the following steps. First, a neighbor list is computed by using Euclidian distance on 16×16 representations of the query image and the database images. Afterwards, the IDM is applied to the closest k database images. Consequently, the computation time is reduced by the factor n/k . Based on this sieve function, the most relevant images are selected by the application of $S_{REL}(Q, sieve(Q, B, k))$, where the IDM can only reorder the results.

```
images[] sieve(image QueryImage, image B[], int CutOff)
  Let image Q = scale (QueryImage, 16x16); //downscale query image
  Let int N=bound(B[]); //Size of the database B
  Define image P[N]; //Buffer for B processing
  Define double delta[N]; //for distance computation

  For (i = 0; i < N; i++) //downscale each image in
    Bs[i] = scale (B[i], 16x16); //the database and compute
    delta[i] = euclidian_distance (P[i],Q); //euclidian distance to query

  sort (P, delta); //sort database by distance
  sieve = P[0] .. P[min(CutOff,N)]; //truncate list at cutoff
```

4 Experiments

The experiments aim at verifying whether the query design that yielded good results on the IRMA database could be transferred into another image domain such as the CasImage database without parameterization. This approach was chosen since the ImageCLEF task was explicitly laid out to demonstrate the current state of CBIR research [6]. With respect to the competitive character of the workshop, it is instructive to learn to which degree sophisticated techniques for optimization on the given image domain is necessary to obtain useful results.

4.1 Reference Data

The CasImage database consists of 8,723 images and represents a mixture of diagnostic images from clinical routine and drawings for medical education. Furthermore, there are images with secondary added contents such as pseudo-colorings of segmentation or manual annotations for operation planning.

From this data set, $i = 26$ samples were arbitrarily selected as queries Q_i . The experimental task was to extract similar images to each of the samples and provide a list of retrievals for manual evaluation [6]. Thus, the ground truth for each

Table 1. The weights for the classifiers as used for the similarity functions in the experiments

Classifier	α	β	γ
D_{AR}	0.0	0.0	1.0
D_{JS}	1.0	0.0	0.0
D_{IDM}	0.0	1.0	0.0
C_1 & C'_1	0.225	0.675	0.1
C_2 & C'_2	0.25	0.75	0.0

query was a-priori unknown and optimization was only possible in an empirical and heuristic manner. In this work, no heuristic manipulation of the parameters for result optimization was performed. Hence, there is no specific quantization and threshold computation to cut off the list from $S_{REL}(Q_i, B)$ with respect to each Q_i . Since the actual amount of relevant images in the database is unknown for each Q_i , a fixed set of possible results is returned. In the IRMA concept a combined evaluation of precision, recall, and visual plausibility is used to evaluate and parameterize the system for different applications. Since automatically generated quality measures like precision and recall do not necessarily reflect the visual relevance of query results moreover it must be verified manually by an expert. This demands a trade-off between full database processing and interactively manageable results. For this purpose the cutoff value for the size of the result set is required. Its automated computation is application specific and not yet integrated in the IRMA concept. A fixed cutoff value of 100 images was found to be a suitable compromise with respect to common class sizes in the IRMA database. For comparable quantitative evaluation of the CasImage database, a ground truth is provided by three medical experts from the Geneva University Hospital [6].

4.2 Quality of Results

Similarity Function. The adjustment of $\rho(Q, R)$ as defined in (5), means the empirical adaptation of the weights α , β and γ , as explained in Section 3.4. Table 1 lists the settings that were tested for the combined classifier weights. Each of the distance measures D_{AR} , D_{JS} , and D_{IDM} is verified separately by setting the respective weight to 1 and all others to 0. Based on those experiments two weighted combinations C_1 and C_2 were acquired on the IRMA database [12], which were also applied to the CasImage database. The corresponding runs were submitted as *mi_combine1* (C_1) and *mi_combine2* (C_2). C_1 combines D_{AR} , D_{JS} , and D_{IDM} while C_2 combines D_{JS} and D_{IDM} . The sieve function $sieve(Q, B, k)$ must be evaluated separately, since the cutoff after k images eventually affects the retrieval results, causing two additional parameterizations C'_1 and C'_2 .

Result Evaluation. For each query image Q_i , the first 100 images from the sequence $S_{REL}(Q_i, B)$ were compared to the ground truth by the usual measures of precision and recall:

$$\text{precision} := \frac{\# \text{ of relevant images}}{100} \quad (7)$$

$$\text{recall} := \frac{\# \text{ of relevant images}}{\# \text{ of relevant images in } B} \quad (8)$$

Ground truth was the `pisec_total` data set, which was provided by the medical experts from the Geneva University Hospital [6]. Due to the restriction to 100 replies, the recall will never reach 100% for queries with more than 100 relevant results in the database. Precision will as well be low for query images, which have significantly less than 100 images among the data set. This bias is accepted with respect to the compromise between visual verification and automated quality measurement.

4.3 Runtime Behavior

Finally, the setup of the parameters has to be efficient for fast verification cycles. For this purpose, the IRMA framework supports the separate consideration of the feature extraction at the image entry time (Sec. 2.1) from the actual feature comparison (Sec. 2.2). Furthermore, runtime is optimized by preliminary application of the sieve function to reduce the number of necessary similarity computations. In the conducted experiments, the cutoff value was set to $k = 500$. For quality comparison the combinations C_1 and C_2 are applied to the result of `sieve(Q, B, 500)`, which extends the set of experiments by C'_1 and C'_2 (Tab. 1).

5 Results

5.1 Quality of Results

The precision for each of the classifier combinations is listed in Table 2. For the combined classifiers C_1 and C_2 , the best precision was obtained for Q_{24} and the worst precision for image Q_{14} . While best recall for the combined measures was also for image Q_{14} the worst recall for C_1 and C_2 was for image Q_{23} . Overall, C_1 yielded the highest average precision. For query Q_7 only, D_{AR} returns no relevant image while the precision constantly increases with D_{JS} , and D_{IDM} and finally obtains the highest value of 0.36 with C_1 . Only for Q_{11} , no useful result could be retrieved. The results for C'_1 and C'_2 on the reduced datasets are only slightly inferior with respect to average precision. Several single results are even better such as for query Q_5 . For query images Q_1 , Q_6 , Q_{15} , Q_{24} , precision is perfect or near perfect, whereas several query images yielded unsatisfactory results. Especially, queries Q_4 , Q_{11} , Q_{14} , Q_{17} and Q_{23} returned only 43, 9, 11, 31, and 74 relevant images, respectively.

5.2 Runtime Behavior

The computation of all required feature representations takes approximately 7.5 hours while the query computation for the combined measures for a single image requires about 5 minutes on a standard Pentium PC running at 2.4 GHz (Tab. 3). The sieve-based computation of the combined measures C'_1 and C'_2 yields a significantly faster runtime of 18.7 seconds for a single query.

Table 2. Precision for the experiments. The boxes indicate the best and least precise result

Q_i	D_{AR}	D_{JS}	D_{IDM}	C_1	C_2	C'_1	C'_2
1	0.10	0.63	0.97	0.97	0.97	0.97	0.97
2	0.01	0.70	0.66	0.81	0.82	0.71	0.72
3	0.08	0.23	0.25	0.35	0.29	0.27	0.27
4	0.09	0.02	0.02	0.04	0.03	0.02	0.02
5	0.03	0.03	0.36	0.40	0.39	0.43	0.42
6	0.20	0.81	0.94	0.97	0.99	0.96	0.95
7	0.00	0.24	0.24	0.36	0.31	0.35	0.33
8	0.09	0.06	0.11	0.23	0.20	0.11	0.11
9	0.01	0.16	0.15	0.25	0.29	0.26	0.26
10	0.04	0.17	0.41	0.42	0.38	0.37	0.42
11	0.00	0.01	0.00	0.03	0.03	0.00	0.00
12	0.23	0.47	0.69	0.67	0.72	0.72	0.71
13	0.02	0.10	0.36	0.42	0.38	0.37	0.35
14	0.01	0.03	0.01	0.01	0.04	0.02	0.01
15	0.14	0.96	0.87	0.97	0.98	0.88	0.89
16	0.02	0.57	0.34	0.58	0.51	0.34	0.34
17	0.00	0.04	0.07	0.10	0.11	0.11	0.11
18	0.20	0.10	0.43	0.36	0.38	0.38	0.39
19	0.01	0.81	0.50	0.73	0.78	0.67	0.68
20	0.06	0.06	0.10	0.10	0.09	0.09	0.08
21	0.02	0.11	0.52	0.40	0.39	0.35	0.33
22	0.10	0.36	0.68	0.59	0.64	0.59	0.60
23	0.03	0.06	0.10	0.15	0.08	0.15	0.09
24	0.15	0.80	1.00	1.00	1.00	0.99	1.00
25	0.38	0.41	0.36	0.46	0.41	0.42	0.40
26	0.13	0.21	0.02	0.30	0.20	0.32	0.21
avg	0.08	0.31	0.39	0.45	0.44	0.42	0.41

Table 3. Integral running times of the feature extraction for all 8,728 images, of the feature comparison for all 26 query images and of a single query, on a standard PC running at 2.4 GHz

Classifier	feature extraction	Query 26 Images	Single Query
D_{AR}	0.5 h	< 1 s	<< 1 s
D_{JS}	4 h	13 s	< 1 s
D_{IDM}	3 h	0.25 h	300 s
C_1	7.5 h	0.25 h	300 s
C_2	7 h	0.25 h	300 s
C'_1	7.5 h	0.15 h	18.7 s
C'_2	7 h	0.15 h	18.7 s

5.3 ImageCLEF Ranking

Within the ranking of all ImageCLEF submissions the mean average precision (MAP) was chosen as a measure for the quality of a result [6]. According to the overall evaluation the IRMA approach yielded a MAP of 0.2980 with C_1 and a MAP of 0.2809 for C_2 , which corresponded to rank 24 and 31 from 44 runs submitted [6]. The MAPs were ranging from the best value of 0.48 to the worst of 0.1, where the mean of all MAPs was 0.29 with a standard deviation of 0.11. There were 19 submissions where no query expansion and additional textual information was used. In this more comparable ranking, the IRMA approach achieves the 9th and 11th position respectively. Here, the mean of all MAPs is 0.26 while the standard deviation remains 0.11.

6 Discussion

Since the MAPs lie within the standard deviation of the MAPs for all submitted runs in the ImageCLEF task, the outcome is encouraging. Note that there was no effort taken in task specific optimization of the parameters. In the ImageCLEF task, the ranked retrieval results are a mixture of text-based, content-based, and hybrid approaches with eventual query expansion. However, the IRMA framework neither takes advantage of

- multichannel information such as color nor of
- textual annotations,

but still, the results are in the center field of the ranking.

By application of the query related sieve on the database, the number of costly IDM comparisons is significantly reduced with only slight loss of average precision and recall. This also encourages the use of sophisticatedly implemented classifiers for online retrieval applications such as differential diagnosis support via queries to a PACS. In such routine applications the physician needs immediate response to compare a given image to possibly related cases with known findings.

7 Conclusion

The application of a parameterization and evaluation concept that was optimized for the IRMA domain yields useful retrieval results on the previously unknown CasImage domain. It was one of the two main goals of this work to show that it is possible to obtain good results with the IRMA system without parameter adoption. Even if the results are not optimal time consuming training cycles are avoided. This is important since optimization of powerful classifiers such as D_{IDM} is infeasible in online systems such as PACS due to running times of several hours. The second goal was to verify the need for sophisticated features and similarity measures. It can be stated that complex integration of multichannel and textual information yields better results in comparison to the basic IRMA approach.

However the trade-off between time input for parameter adjustment and flexible domain adoption must be considered.

By separating the application logic from the storage concept, the software architecture also supports the transfer of new features and classifiers as well as images without changes in the existing implementation. Furthermore, there is still the commonly observed gap between fast computable query designs and good retrieval results. Powerful classifiers as required for medical applications still need computation times, which are not applicable in fast reacting retrieval environments. This remains a field of ongoing research, where the IRMA system provides a supporting framework for efficient verification and also application.

References

1. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22** (2000) 1349–1380
2. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications. clinical benefits and future directions. *International Journal of Medical Informatics* **73** (2004) 1–23
3. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., Equitz, W.: Efficient and effective querying by image content. *Journal of Intelligent Information Systems* **3** (1994) 231–262
4. Tagare, H.D., Jaffe, C.C., Duncan, J.: Medical image databases: A content-based retrieval approach. *Journal of the American Medical Informatics Association* **4** (1997) 184–198
5. Lehmann, T.M., Güld, M.O., Thies, C., Fischer, B., Spitzer, K., Keyzers, D., Ney, H., Kohnen, M., Schubert, H., Wein, B.: Content-based image retrieval in medical applications. *Methods of Information in Medicine* **43** (2004) 354–361
6. Clough, P., Müller, H., Sanderson, M.: The CLEF Cross-Language Image Retrieval Track (ImageCLEF)2004. In Peters, C., Clough, P., Gonzalo, J., Jones, G., Kluck, M., Magnini, B., eds.: Fifth Workshop of the Cross-Language Evaluation Forum (CLEF 2004), *Lecture Notes in Computer Science (LNCS)*, Springer, Heidelberg, Germany (in this volume) (2005)
7. Zhou, X.S., Huang, T.S.: Edge-based structural features for content-based image retrieval. *Pattern Recognition Letters* **22** (2001) 457–468
8. Swain, M.J., Ballard, D.H.: Color indexing. *International Journal of Computer Vision* **7** (1991) 11–32
9. Castelli, V., Bergman, L.D., Kontoyiannis, I., Li, C.S., Robinson, J.T., Turek, J.J.: Progressive search and retrieval in large image archives. *IBM Journal of Research and Development* **42** (1998) 253–268
10. Güld, M.O., Thies, C., Fischer, B., Keyzers, D., Wein, B.B., Lehmann, T.M.: A platform for distributed image processing and image retrieval. In: *Procs SPIE. Volume 5150*. (2003) 1109–1120
11. ITU: Basic parameter values for the hdtv standard for the studio and for international programme exchange. ITU-R Recommendation BT.709, [formerly CCIR Rec. 709] ITU, 1211 Geneva:Switzerland (1990)
12. Güld, M.O., Keyzers, D., Deselaers, T., Leisten, M., Schubert, H., Ney, H., Lehmann, T.M.: Comparison of global features for categorization of medical images. In: *Procs SPIE. Volume 5371*. (2004) 211–222

13. Lehmann, T.M., Güld, M.O., Keysers, D., Schubert, H., Kohnen, M., Wein, B.B.: Determining the view position of chest radiographs. *Journal of Digital Imaging* **16** (2003) 280–291
14. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-8** (1978) 460–472
15. Puzicha, J., Rubner, Y., Tomasi, C., Buhmann, J.: Empirical evaluation of dissimilarity measures for color and texture. In: *Procs International Conference on Computer Vision*. Volume 2. (1999) 1165–1173
16. Deselaers, T., Keysers, D., Ney, H.: Features for image retrieval: A quantitative comparison. In: *DAGM 2004 Pattern Recognition, 26th DAGM Symposium Tübingen, Germany*. Volume 3175., *Lecture Notes in Computer Science (LNCS)*, Springer, Heidelberg, Germany (2005) 228–236