Content-based queries on the casimage database with the IRMA framework: A field report

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

Department of Medical Informatics, Aachen University of Technology

cthies@mi.rwth-aachen.de

Abstract

Recent research has suggested that there is no general similarity measure which can be applied parameter free on arbitrary databases. In contrast 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 dataset could be transferred to another image repository without extensive re-design of all relevant components. Transfer of data corpora is vital if image retrieval is integrated into complex environments such as picture archiving and communication systems (PACS). Image retrieval in medical applications (IRMA) is a framework that strictly separates data administration and application logic. This permits an efficient transfer of the data abstraction from one database to another without re-designing the software. It supports the loop of estimating a combination of distance measures, parameter adaption and result visualization, which is characteristic if an image retrieval application is used for varying data corpora. In this work the casimage dataset has been added as a data corpus to the IRMA system. Thereon the query performance has been evaluated without optimization of the currently applied feature combination. It consists of scaled representations of the images, global texture, aspect ratio, and an evaluation of deformation between pixels of different images.

1 Introduction

Classical architectures of image retrieval systems consist of an image repository, along with visualization tools and query functionality. The principle of data storage and visualization does not vary significantly for 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].

The first class of approaches associates and stores the secondary annotated content descriptions with each image. In general this means textual information, which makes the retrieval task a text search in the descriptions. Since the information is added by a human observer it is likely to be correct and considers the semantical context of the image content. However inter- and intra-individual variance in perception, knowledge and capability of expression leads to different descriptions for a single issue, as do the homonyms and synonyms in medical terminology. 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 mean 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, i.e. image content. 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 content based image retrieval depends on an appropriate selection of the similarity measure which again depends

on the considered image features. In general, selection of features and similarity measures cannot be done by a physician in clinical routine. Thus an abstraction layer from the feature handling is required.

Besides the data entry cost and the content representation a third problem is the state of the image database: In a clinical environment the set of available images continuously grows. This must be considered since it is one aim of content based image retrieval to make clinical routine data available as a source of knowledge for education and diagnosis. Consequently the feature selection and similarity computation must be as flexible as possible to adopt to the current state of the database.

The image retrieval in medical applications (IRMA) project integrates these three aspects of query design into a single framework [2]. In this paper the application of the IRMA framework to the previously unknown casimage database of the University Hospitals of Geneva is described. Therefore the general query principle of the IRMA System is introduced (Sec. 2), then the specific features and similarity computations for query execution are described (Sec. 3). Based on these methods a suite of experiments was conducted (Sec. 4). The results that were obtained are summarized (Sec. 5) and discussed (Sec. 6). Finally a conclusion is drawn (Sec. 7).

2 Designing a query

As stated in the introduction a content based query consists of an automated feature extraction combined with a similarity measure.

2.1 Feature Computation

Existing literature describes numerous features, which can be extracted from an image. They are roughly categorized in shape descriptions such as edges [3], and color information, for instance texture features [4, 5]. 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. On behalf of 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, either for a query or for a database entry, is automatically transformed into all available feature representations. This ensures short update cycles when entering new images as well as implementing new features since only the new results must be integrated. For this purpose IRMA provides an automated storage concept that applies the implemented image to feature mappings at data entry time. A new feature computation is integrated into an image processing chain by providing the transformation code. The actual database handling is hidden to this implementation by an interface which provides a view on the image exclusively.

2.2 Feature Comparison

In this concept a query corresponds to a nearest-neighbor classification of the query image to all images in the currently considered image corpus. Consequently it is designed by defining a similarity measure as a metric in the images' feature space. Result of this computation is a representation of the corpus as a sorted list, where the most similar images with the smallest distance form the head. It must be considered how many of the first images in the list are relevant images matching the query. This question is task specific and cannot be determined in advance. Furthermore it 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 take 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 over the corpus respectively. Those processing sequences are combined on a binary execution level by abstract methods. Consequently the data handling is also hidden from the application by special data flow interfaces.

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 knowledge on the classes of images and their sizes. Thus finding the appropriate features and similarity measures is similar to an optimization task, without knowing the target function. However, brute force learning approaches for the 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 changed, 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.

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

3 Applied features and queries

The methods applied here have been taken without optimization for the casimage image set from recent applications on the IRMA database. This database consists of 10,000 images from clinical routine which have been categorized by medical experts and is used to train parameters and verify query quality [6]. From the medical point of view it was assumed, that the distribution of image classes from clinical routine in two comparable university hospitals does not vary significantly. Thus the unmodified transfer of the methods is reasonable. There is one difference since the IRMA system processes only gray-scale versions of the images. For color images that are contained in a new database the gray-scale conversion is done by using the standard color weighting

$$Y = \frac{6969 \cdot R + 23434 \cdot G + 2365 \cdot B}{32768}.$$

Furthermore recent experiments have indicated that spatial and intensity features must be considered coevally to obtain reasonable results [7, 8]. Those were the global texture features by TAMURA et al. and the image distortion model [9, 11]. In the following descriptions Q means the query image and R denotes a reference image from the database.

3.1 Texture features by TAMURA

TAMURA et al. use coarseness, contrast and directionality to capture an image's texture properties [9]. Those features are computed per pixel and reflect the texture affiliation. The value ranges for coarseness, contrast and directionality in the current image are quantized into 6, 8, 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 therfore incomparable histogram counts. To obtain comparable features, each image is scaled to a size of 256×256 pixels. This, on the other hand, ignores the aspect ratio.

To compare the TAMURA histograms of two images $TH(Q)$ and $TH(R)$ with $M = 384$ bins each, the Jensen-Shannon divergence is used [10]:

$$\text{JST}(Q,R) = \frac{1}{2} \sum_{m=1}^{M} \left[TH_m(Q) \log \frac{2TH_m(Q)}{TH_m(Q) + TH_m(R)} + TH_m(R) \log \frac{2TH_m(R)}{TH_m(Q) + TH_m(R)} \right]$$

3.2 Aspect ratio

Comparing the aspect ratio of images is an unspecific measure yet it is useful if the dimension of images has to be considered. Since normalization for some texture features requires the deformation of the image dimensions to a square shape the aspect ratio of an original image is a mean of image comparison. Furthermore the aspect ratio is characteristic for different classes of medical images. MRI slices have identical edge dimensions while radiographs of limbs are rectangular elongated in direction of the principal bone. The aspect ratio is compared by:

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

Where $X(I)$ and $Y(I)$ are the size of the x- and y-dimension of an image I.

3.3 Image distortion model

While histogram-based methods provide invariance against some transformations, e.g. translation, scaled representations of the original images can preserve spatial properties, which are especially important to recognize medical images [11]. A drastic reduction in size also reduces noise and small image defects. The image distortion model (IDM) expands the naïve 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' denote the offsets within the search window for a corresponding pixel, the distance is computed by

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

Where $I(x, y), I \in \{R, Q\}$ indicates the pixel value of an image R, Q at position (x, y). 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 height of 32 pixels keeping their original aspect ratio.

3.4 Classifier combination

Classifier combinations can be grouped into three main categories [12]: parallel, serial (like a sieve), and hierarchical (comparable to a tree). In the current application a parallel classifier combination is used, since it is a straightforward method to integrate results that were obtained from the single classifiers. Normally this is a linear combination of the single classifier results such as the weighted sum. However, the result depends of the value ranges of the single addends. If there is a single large addend that dominates the sum, it will eventually shadow discriminative results with low value ranges. In order to avoid such value domination the results of each single classifier for each image are transformed to a common scale first. 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 measures in this paper determine the metric:

$$\rho(Q, R) = \alpha \cdot JST(Q, R) + \beta \cdot \text{IDM}(Q, R) + \gamma \cdot \text{AR}(Q, R).$$

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 Relevance Determination

Relevance of an image with respect to the query image Q is computed by sorting the database DB with the similarity metric into a sequence:

$$\text{REL}(Q, DB) = (R_1, ..., R_n) | \rho(Q, R_1) \leq \rho(Q, R_2) \leq ... \leq \rho(Q, R_n), R_1..R_n \in DB, n = |DB|.$$

DB denotes the entire image repository that is to be searched and $n = |DB|$ is 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. In this work the sieve is applied to the IDM classifier by the following steps. First, compute a neighbor list 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. On the other hand, the IDM can only reorder the results. The sieve is defined by:

1. $DS(I, S)$, Compute a $S \times S$ representation for I
2. $\delta(A, B) = ||DS(A, 16), DS(B, 16)||_2$
3. $DB'(Q, DB, k) = (R_1, ..., R_k) | \delta(Q, R_1) \leq \delta(Q, R_2) \leq ... \leq \delta(Q, R_n), R_1..R_n \in DB, n = |DB|$

Based on the sieve $DB'(Q, DB, k)$, the most relevant images are now selected by application of

$$\text{REL}(Q, DB'(Q, DB, k))$$

4 Experiments

Aim of the experiments is the verification whether the query design that yielded good results on the IRMA database could be transferred to another database with minimal changes to the parameterization.

4.1 Reference Data

The casimage database consists of 8,723 images and represents a mixture of images from clinical routine and drawings from medical education. Furthermore there are images with secondary added contents such as pseudo-colorings of segmented ultrasound images or manually placed marker arrows for operation planning, which do not represent original data.

From this dataset 26 samples have been arbitrarily selected as query images. The experimental task was to extract similar images to each of the 26 samples and provide a list of query results for manual evaluation. Thus the ground truth for each query was a-priori unknown and could not be optimized. This setting corresponds to free database exploration without any knowledge about its structure and not to mention the relevance of the results. On behalf of that there is no specific quantization and threshold computation to cut off the list from $\text{REL}(Q, DB)$ with respect to each Q. Since the actual amount of relevant images to a query Q that are contained in the database is unknown, a fixed set of possible results is to be presented. In the experiments the cutoff is set to 100 images since this number is manually manageable in a browser view. These are lifelike values for free database exploration as used to understand the structure of a database. For comparable quantitative evaluation, a ground truth is mandatory. It was provided by three medical experts form the Geneva University Hospital for the casimage dataset for analysis after deadline for result submission.

4.2 Quality of results

The capacity of the concept is best verified by examination of the effect of different query designs on the results. Since the function $\rho(Q, R)$ provides the central query functionality it is manipulated

	α	β	γ
AR	0	0	1
JST	1	0	0
IDM	0	1	0
Comb 1	0.25	0.75	0.0
Comb 2	0.225	0.675	0.1
Comb 1'	0.25	0.75	0.0
Comb 2'	0.225	0.675	0.1

Table 1: The weights for the classifiers as used for the experiments

to optimize the results. As explained in section 3.4 this means the adaptation of the parameters α, β and γ. Table 1 lists the settings for the combined classifier weights, that have been used for the experiments.

For each of the 26 query images Q the first 100 images from the sequence $\text{REL}(Q, DB)$ have been considered. They have been compared to the ground truth provided for the casimage data by the usual measures of precision and recall. Precision quantifies the percentage of relevant images that were among the 100 images returned. Recall denotes the ratio of relevant images to all images that should have been returned. Due to the restriction to 100 replies, recall will never reach 100 percent for queries with more than 100 relevant results in the database. Also precision will be low for query images which have significantly less than 100 images among the dataset.

4.3 Runtime behavior

Besides the quality of results the runtime of the algorithms must be considered. Especially if parameters must be adjusted in a feedback cycle. Furthermore the retrieval system must be capable of handling several queries at a time with the respective response times. And finally the setup of the parameters has to be efficient for efficient verification cycles. For this purpose the extraction of features at the image entry time (Sec. 2.1) must be considered separately from the feature comparison (Sec. 2.2). Furthermore runtime is reduced by preliminary application of the sieve $DB'(Q, DB, k)$ to reduce the number of necessary comparisons. In the conducted experiments the cutoff value k was set to 500. For quality comparison the the combinations Comb 1 and Comb 2 are also applied on $DB'(Q, DB, k)$ which extends the experiment set by Comb 1' and Comb 2'.

5 Results

5.1 Quality of results

The results for precision and recall for each of the classifier combinations are listed in Table 2 and Table 4. For the combined classifiers Comb 1 and Comb 2 the best precision was obtained for image 24 and the worst precision for image 14. While best recall for the combined measures was also for image 14 the worst recall for Comb 1 and Comb 2 was for image 23. Overall Comb 2 yielded the highest average precision. The results for Comb 1' and Comb 2' on the reduced datasets are only slightly inferior in the average precision. Several single results are even better such as for query 5. Figure 1 illustrates excerpts from the results for three queries. For four query images, precision is perfect or near perfect, whereas several query images yielded unsatisfactory results, especially queries 4 (43 relevant images), 11 (9), 14 (11), 17 (31), and 23 (74).

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

Table 2: Precision results for the experiments. Comb 1 combines JST and IDM Comb 2 combines AR, JST, and IDM, while Comb 1$'$ and Comb 2$'$ use the sieve $DB'(Q, DB, k)$. The best result was obtained for query 24, and the least precise result for query 14 respectively (Fig. 1). Both were obtained with Comb 2.

5.2   Runtime behavior

Computation of all requires feature representations takes approximately 7.5 hours while the query computation for the combined measures for a single image takes about 5 minutes (Tab. 3) on a standard Pentium PC running at 2.4 GHz. The sieve based computation of the combined measures Comb 1$'$ and Comb 2$'$ yields a significantly faster runtime of 18.7 seconds for a single query.

6   Discussion

The application of image features and their respective distance measures, including a parameter set, which was optimized on the IRMA database, yields useful retrieval results on the previously unknown casimage image data set. However, the casimage data set also includes color images, which demand additional features to capture their specific properties.

The variability of the precision computed for all queries is high, since the actual number of relevant images for all queries varies. For 16 query images there are less than 100 relevant images in the database. Consequently the number of false positive results is at least the difference between the relevant images and the preset threshold. On the other hand the recall is biased, for the ten images for which more than 100 relevant results exist. Thus methods for an automated threshold estimation are needed to improve query results.

	feature extraction	Query 26 Images	Single Query
AR	0.5 h	< 1 sec	<< 1 sec
JST	4 h	13 sec	< 1 sec
IDM	3 h	150 min	5 min
Comb 1	7 h	150 min	5 min
Comb 2	7.5 h	150 min	5 min
Comb 1′	7 h	9.3 min	18.7 sec
Comb 2′	7.5 h	9.3 min	18.7 sec

Table 3: Integral running times of the feature extraction for all 8728 images, of the feature comparison for all 26 query images and of a single query on a standard PC running at 2.4GHz.

By application of the query related sieve on the database the number of costly IDM comparisons is significantly reduced. On the other hand the query results have only slightly lost in average precision and recall. This encourages the use of more sophisticated classifiers for online retrieval applications such as PACS integration.

7 Conclusion

The presented approach considers the entire application cycle of an image retrieval system. By separating the application logic from the storage concept the transfer of new features classifiers and image data in the existing system requires no changes of existing implementations. Yet the quality of the results is not as good as it could be. The transfer of existing retrieval solutions onto new data repositories still requires some adaptive efforts. On the other hand the results can be compared to middle-rate results of manually optimized approaches. This is encouraging since optimization is always time consuming with respect to verification times of good classifiers such as the IDM. It is not necessary to start from scratch if a retrieval concept is transferred to another database. In contrast if the system is well defined the actual optimization is efficiently prepared.

However this conclusion has to be verified on various image repositories. Furthermore there is still the gap between fast computable query designs and good retrieval results. Powerful classifiers as required by medical applications still require long computation times. This remains a field of ongoing research where the IRMA system provides a supporting framework for efficient verification and also application.
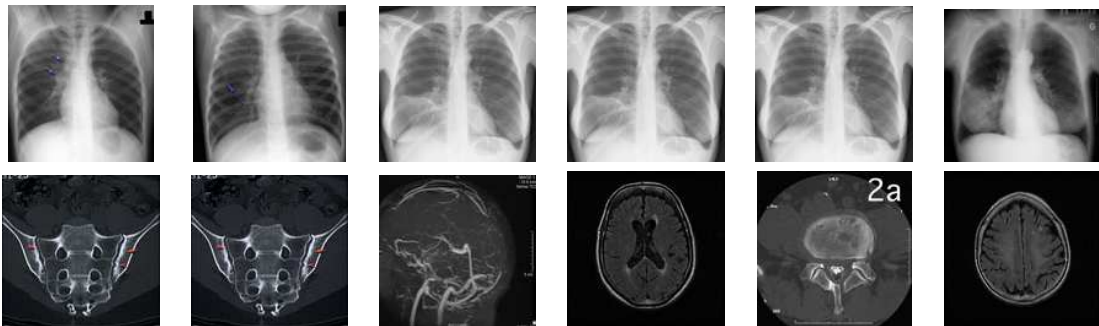

Figure 1: Results for Medical Query images 24 (top row), 14 (bottom row), left column: query image, remaining columns: reference images, ascending distance. For query image 24, a precision of 100% was achieved for 100 returned images, whereas Query images 14 yielded only 1% precision among 100 returned images.

Query	AR	JST	IDM	Comb 1	Comb 2	Comb 1$'$	Comb 2$'$
1	0.04	0.27	0.41	0.41	0.41	0.41	0.41
2	0.00	0.22	0.21	0.26	0.25	0.23	0.22
3	0.11	0.32	0.35	0.40	0.49	0.38	0.38
4	0.21	0.05	0.05	0.07	0.09	0.05	0.05
5	0.04	0.04	0.43	0.46	0.48	0.50	0.51
6	0.08	0.32	0.37	0.39	0.38	0.38	0.38
7	0.00	0.50	0.50	0.65	0.75	0.69	0.73
8	0.08	0.05	0.09	0.17	0.20	0.09	0.09
9	0.02	0.37	0.35	0.67	0.58	0.60	0.60
10	0.05	0.22	0.52	0.48	0.53	0.53	0.47
11	0.00	0.11	0.00	0.33	0.33	0.00	0.00
12	0.13	0.26	0.39	0.40	0.37	0.40	0.40
13	0.02	0.11	0.38	0.40	0.44	0.37	0.39
14	0.09	0.27	0.09	0.36	0.09	0.09	0.18
15	0.06	0.38	0.35	0.39	0.38	0.35	0.35
16	0.01	0.40	0.24	0.36	0.41	0.24	0.24
17	0.00	0.13	0.23	0.35	0.32	0.35	0.35
18	0.26	0.13	0.55	0.49	0.46	0.50	0.49
19	0.01	0.71	0.44	0.68	0.64	0.60	0.59
20	0.22	0.22	0.37	0.33	0.37	0.30	0.33
21	0.02	0.12	0.58	0.43	0.44	0.37	0.39
22	0.06	0.21	0.40	0.37	0.35	0.35	0.35
23	0.04	0.08	0.14	0.11	0.20	0.12	0.20
24	0.04	0.20	0.24	0.24	0.24	0.24	0.24
25	0.59	0.64	0.56	0.64	0.72	0.62	0.66
26	0.25	0.40	0.04	0.38	0.57	0.40	0.60
avg	0.07	0.26	0.39	0.37	0.38	0.34	0.35

Table 4: Recall results for the experiments as indicated in table 2.

References

[1] Smeulders AWM, 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(12): 1349-1380, 2000.

[2] Lehmann TM, Wein BB, Dahmen J, Bredno J, Vogelsang F, Kohnen M: Content-Based Image Retrieval in Medical Applications A Novel Multi-Step Approach. Proceedings SPIE 3972: 312-320, 2000.

[3] Zhou XS, Huang TS: Edge-based structural features for content-based image retrieval. Pattern Recognition Letters 22(5): 457-468, 2001.

[4] Haralick RM, Shanmugam, Dinstein I: Textural features for image classification. IEEE Transactions on System, Man, and Cybernetics SMC-3: 610-621, 1973.

[5] Castelli V, Bergman LD, Kontoyiannis I, Li CS, Robinson JT, Turek JJ: Progressive Search and Retrieval in Large Image Archives. IBM Journal of Research and Development 42(2): 253-268, 1998.

[6] Lehmann TM, Schubert H, Keysers D, Kohnen M, Wein BB: The IRMA code for unique classification of medical images. Proceedings SPIE 5033: 109-117, 2003.

[7] Güld MO, Keysers D, Deselaers T, Leisten M, Schubert H, Ney H, Lehmann TM: Comparison of global features for categorization of medical images. Proceedings SPIE 5371: 211-222, 2004.

[8] Lehmann TM, Güld MO, Keysers D, Schubert H, Kohnen M, Wein BB: Determining the view position of chest radiographs. Journal of Digital Imaging 16(3): 280-291, 2003.

[9] Tamura H, Mori S, Yamawaki T: Textural Features Corresponding to Visual Perception. IEEE Transactions on Systems, Man, and Cybernetics; SMC-8(6), 460-472, 1978.

[10] Puzicha J, Rubner Y, Tomasi C, Buhmann J: Empirical Evaluation of Dissimilarity Measures for Color and Texture. Proceedings International Conference on Computer Vision, Vol. 2, 1165-1173, 1999.

[11] Keysers D, Gollan C, Ney H: Classification of Medical Images using Non-linear Distortion Models. Proceedings BVM 2004, Bildverarbeitung für die Medizin 2004, Springer-Verlag, Berlin, 366-370, 2004.

[12] Jain AK, Duin RPW, Mao J: Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1): 4-36, 2000.