Content–based image retrieval from a database of fracture images

Henning Müller[a], Phuong Anh Do Hoang[a], Adrien Depeursinge[a], Pierre Hoffmeyer[b], Richard Stern[b], Christian Lovis[a], Antoine Geissbuhler[a]

[a]University Hospitals of Geneva, Service of Medical Informatics, 24, Rue Micheli-du-Crest, CH-1211 Geneva 14, Switzerland;
[b]University Hospitals of Geneva, Department of Surgery, 24, Rue Micheli-du-Crest, CH-1211 Geneva 14, Switzerland;

ABSTRACT

This article describes the use of a medical image retrieval system on a database of 16'000 fractures, selected from surgical routine over several years. Image retrieval has been a very active domain of research for several years. It was frequently proposed for the medical domain, but only few running systems were ever tested in clinical routine. For the planning of surgical interventions after fractures, x–ray images play an important role. The fractures are classified according to exact fracture location, plus whether and to which degree the fracture is damaging articulations to see how complicated a reparation will be. Several classification systems for fractures exist and the classification plus the experience of the surgeon lead in the end to the choice of surgical technique (screw, metal plate, ...). This choice is strongly influenced by the experience and knowledge of the surgeons with respect to a certain technique. Goal of this article is to describe a prototype that supplies similar cases to an example to help treatment planning and find the most appropriate technique for a surgical intervention.

Our database contains over 16'000 fracture images before and after a surgical intervention. We use an image retrieval system (GNU Image Finding Tool, GIFT*) to find cases/images similar to an example case currently under observation. Problems encountered are varying illumination of images as well as strong anatomic differences between patients. Regions of interest are usually small and the retrieval system needs to focus on this region. Results show that GIFT is capable of supplying similar cases, particularly when using relevance feedback, on such a large database. Usual image retrieval is based on a single image as search target but for this application we have to select images by case as similar cases need to be found and not images. A few false positive cases often remain in the results but they can be sorted out quickly by the surgeons.

Image retrieval can well be used for the planning of operations by supplying similar cases. A variety of challenges has been identified and partly solved (varying luminosity, small region of interested, case–based instead of image–based). This article mainly presents a case study to identify potential benefits and problems. Several steps for improving the system have been identified as well and will be described at the end of the paper.

1. INTRODUCTION

Image retrieval has been a very active domain of research for almost 20 years[1–3] to allow for a better management of the increasingly produced digital visual information. Many techniques have shown their utility such as relevance feedback and interacting with a user in the loop to optimise results.[4, 5] The medical domain was often mentioned as a prime example for content–based image retrieval,[6–8] but only few running systems were ever tested in clinical routine such as the *ASSERT* system[9] that showed a significant improvement in the diagnostic quality with using the system particularly for less experienced radiologists.

In general, there are two large application groups using image retrieval techniques:

- information retrieval

Further author information: (Correspondence to Henning Müller) henning.mueller@sim.hcuge.ch, tel. ++41 22 372 61 75, fax ++41 22 372 8680

*`http://www.gnu.org/software/gift/`

- and image classification.

Systems such as the *IRMA*[†] (Image Retrieval in Medical Applications)[10] follow the classification approach that requires training data sets and labels new incoming images to one of the pre–existing classes of the database. Applications of this approach are the automatic correction of DICOM headers that have shown to contain errors[11] or a semi–automatic annotation of images. Other applications classify images into one disease class.[9] The information retrieval approach is mainly applied to large collections without or with only a limited ground truth and no clear class labels. The *medGIFT*[‡] project[12] follows this information retrieval approach and uses rather general features supplying similar images instead of a hard class membership for a query. This approach is often used to navigate in large information repositories such as many medical teaching files (i.e. Pathopic[§], Casimage[¶]). In general, systems are evaluated based on responses to example images/cases based on what a user of the system would judge to be relevant for a certain task. Both classification and information retrieval based approaches use fairly similar techniques such as visual features and distance measures, only the learning approaches are often different.

Whereas most early systems were either purely visual[5,10] or purely textual,[13] it becomes increasingly clear that the images out of their clinical context loose most of their information. On the other hand, there is much information stored in the images implicitly that is hard and expensive to extract in textual form. A benchmarking event for evaluating medical image retrieval using both visual and textual information is based at *CLEF*[‖] (Cross Language Evaluation Forum). It is called *ImageCLEF* and has a biomedical retrieval benchmark called *ImageCLEFmed*[**]. Each year a large dataset (2006: 51'000 images) and 25–30 query topics are made available to participating research groups that can send in retrieval results that are finally evaluated and compared at a common workshop.[14,15] Other current techniques for visual retrieval include the use of local descriptors for visual retrieval to allow not only for a little variation in the images but also to find sub parts. One of the techniques also applied to medical images is using salient features and relations between these features.[16,17]

This article will first motivate the development of the described prototype before describing the base components (methods) that were used. Then, the first results will be presented before working on important issues that need to be implemented in a fully functional prototype at the next stage.

2. MOTIVATION

For the planning of surgical interventions after fractures, traditional x–ray images play the most important role. The fractures are usually classified according to the place of the fracture and whether the fracture is damaging articulations and to which degree it does so. A common schema for this is the A0 classification of fractures.[18,19] Axes in this schema are:

- the bone,
- the segment of the bone,
- the type of fracture (simple, wedge, complex),
- the group and subgroup within the type.

An example of this system is: **32-B2.1** for *femur, diaphysis, wedge fracture, bending wedge, subtrochanteric.* Several other such classification systems exist and the classification leads generally to the choice of a technique for the surgical intervention (screw, metal plate, ?). This choice is often influenced by the experience of the surgeons with respect to the technique of choice. Goal of the project described in this article is to develop a

[†] http://www.irma-project.org/
[‡] http://www.sim.hcuge.ch/medgift/
[§] http://alf3.urz.unibas.ch/pathopic/intro.htm
[¶] http://pubimage.hcuge.ch/
[‖] http://www.clef-campaign.org/
[**] http://ir.ohsu.edu/image/

Figure 1. Examples images of the fracture database used for retrieval.

technique that supplies similar cases to an example case to assist the treatment planning and find the most appropriate technique requiring less experience for the medical doctors, or rather to make the experience of several surgeons available for less experienced colleagues.

To our knowledge there is currently no image retrieval system supplying similar fracture images for intervention planning. There is also no extensive study on how the choice of the surgical technique influences the long term quality after the surgical intervention. According to the surgeons the current classification systems for fracture do not completely cover the needed information for choosing the most appropriate technique. A manual classification system together with visual retrieval has the potential to improve the information made available to the surgeons for choosing an appropriate technique.

3. METHODS

This section describes the database created at the University Hospitals of Geneva over several years that is used for our current prototype. It also describes the retrieval system that our prototype is based upon.

3.1. Database used for retrieval

At the surgery department of the University Hospitals of Geneva a database with almost 16'000 fractures has been created over the years containing images before and after an operation.[20] An institutional teaching file called *Casimage*[21] is being used for the data acquisition, so the acquisition can be completed by several surgeons and data can easily be shared and reused for teaching as well as for research projects. Periodically, teaching CDs are published with a subset of the data.[20] We received in early 2006 a subset of 15'327 images from a total of 1'637 cases. This means that every case contains on average 9.4 images with the smallest cases containing only 3 images and the largest cases 54 images. Each case also contains a textual annotation for every image and the case but neither the submission of images nor the text is controlled for any quality. In general the images show the situation before the operation in detail and then the result directly after the operation when the patient is still in the hospital. Figure 1 gives an idea of the variety of images that can occur in the data set.

After first tests on the entire data set it became increasingly clear that we would need to limit ourselves to a single bone for such a first prototype to do a slightly more detailed analysis. As there is a large number of femur images in the dataset we were working on this subpart of the database for some of the tests, containing a total of 421 cases. Figure 2 shows an example case for a femur fracture with a total of seven images in the case.

3.2. Retrieval system used

Based on this database we used a general image retrieval system that is available free of charge as open source (*GIFT*, GNU Image Finding Tool,[22]) that is also the basis of the medical retrieval engine *medGIFT* with only slight modifications to find images similar to an example case. This system is subsequently modified to optimise the retrieval outcome. The system has the following features:

Figure 2. Example images of one case of a femur fracture before and after an intervention.

3.2.1. *MRML* – Multimedia Retrieval Markup Language

To separate the actual query engine from a user interface, the Multimedia Retrieval Markup Language (*MRML*[††]) was developed. This query language is based on direct communication of search engine and interface via sockets and eases a variety of applications such as meta–search engines and also the integration of a retrieval tool into a variety of environments and applications. The entire communication is based on the XML standard, which allows for quick development of tools. *MRML* also serves as a language to store log files of user interaction. This information can also be used to improve the query performance by long–term learning from user behaviour.[23] We used this query language to modify the user interface and filter the transmitted image information.

3.2.2. User interface

As *medGIFT* is a domain–specific search tool, the user interface has different requirements from a non–medical retrieval engine. One important part is the display of not only thumbnail images for the browsing but also the text of the diagnosis. This means that a number of images are stored together with an ordered description of a case. On the other hand, not much control is being performed on the quality of the text entered which results in records of extremely varying quality with several being empty and other containing spelling errors and non–standard abbreviations.

Figure 3 shows a web interface after a query was executed. The query results are displayed ordered by their visual similarity to the query, with a similarity score shown underneath the images. The diagnosis is also shown underneath the images. A click on the image links with the case database system and allows to access the full–size images.

3.3. Features and feature weightings

For the ease of processing all images are first converted to 256x256 pixels. Then, *GIFT* relies on four main groups of features for retrieval:

[††] `http://www.mrml.net/`

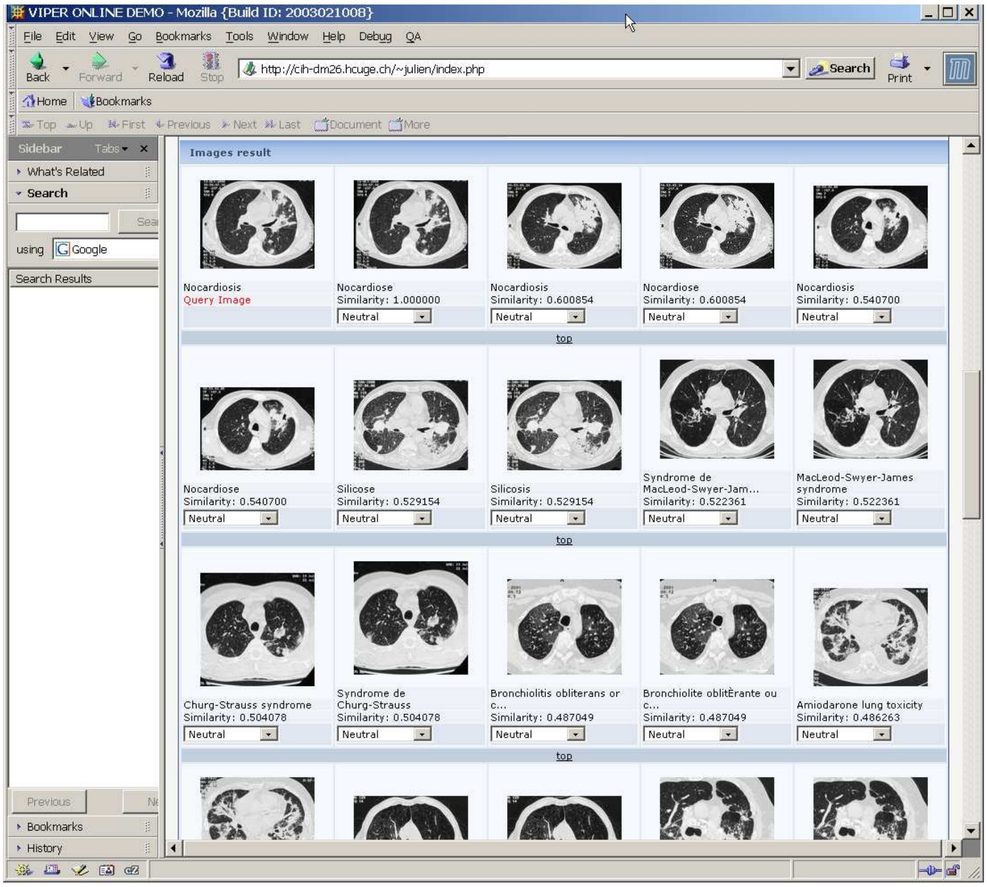

Figure 3. A screenshot of a typical web interface for medical image retrieval system allowing query by example(s) with the diagnosis underneath the image.

- global color features in the form of a color histogram in HSV space (Hue=18, Saturation=3, Value=3, Gray=4);
- local color features in the form of the mode color of blocks in various sizes and various regions by successively dividing the image into four equally–sized regions;
- global texture features in the form of a Gabor filter histogram using four directions, and three scales. The filter responses are quantised into 10 bins;
- local Gabor filter responses.

Gabor filter responses have often shown their good performance for texture characterisation.[24] Equally the HSV color space has proven to be closer to human perception that spaces such as RGB (Red, Green, Blue) and it still is easy to calculate.[25] Surprisingly small numbers of grey levels (8–16) lead generally to best retrieval results, although x–ray images do not contain any color. The small number actually allows for a certain variety in the illumination of the images.

Two weightings are used for the four feature groups. The two global histogram features are weighted according to a simple histogram intersection.[26] The two block feature groups that represent around 80% of the features are weighted according to a simple tf/idf weighting:

Figure 4. Results when using the basic medGIFT system for fracture retrieval.

$$feature\ weight_j = \frac{1}{N}\sum_{i=1}^{N}\left(tf_{ij}\cdot R_i\right)\cdot log^2\left(\frac{1}{cf_j}\right), \tag{1}$$

where tf is the *term frequency* of a feature, cf the *collection frequency* of a feature, j a *feature number*, q corresponds to a query with $i = 1..N$ input images, and R_i is the *relevance* of an input image i within the range $[-1; 1]$.

Then a *score* is assigned to a possible result image k with query q containing features $1..j$:

$$score_{kq} = \sum_j \left(feature\ weight_j\right), \tag{2}$$

Scores are calculated for all four feature groups separately and then added in a normalised way, which leads to better results than a simple addition.[27]

4. RESULTS

This section present the results of our pilot project. Due to the limited time available for the student performing the work, mainly a qualitative evaluation was attempted showing the feasibility of our proposed approach but no quantitative evaluation.

4.1. Using the basic *medGIFT* for retrieval

In a first step we used the basic *medGIFT* system for retrieval without any modification. Results for the radiologists were satisfying as for the performed queries the majority of the results was of the exactly correct region and sort of fracture, although a few false positives remain. Figure 4 shows an example query with the system. Only a single image does not correspond to the exact anatomic region. Several images are images after an operation although the query image was before the intervention.

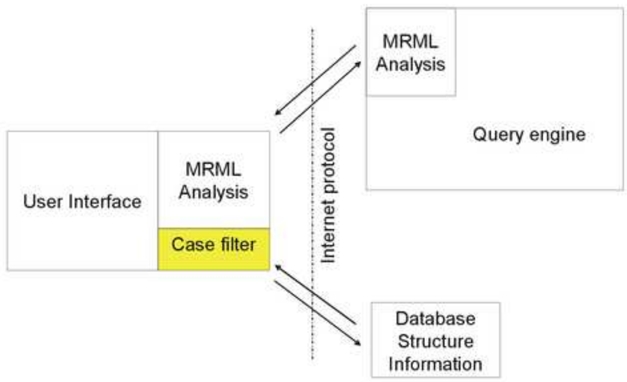

Figure 5. The schema of a system that filters the query response of *medGIFT* to allow case–based retrieval.

Figure 6. Screenshot of a retrieval results after filtering out double images of the same case, and a screenshot showing the possibility to extent a single image to all other images of the same case.

4.2. Performing case–based retrieval with medGIFT

One straight forward step in improving the current retrieval system is to modify the user interface to request a larger number of images than to be shown on screen and filter out those images, where the case is already represented by another image of the same case. Figure 5 shows the information flow in this system.

After the filtering steps, the results are often only slightly different, but for some cases much more information is available after filtering out double images of the same case. Although these images might be visually similar, they do not contain any new information. Figure 6 shows the query for the same image as beforehand but after the filtering step. The quality is slightly superior as this is the case for the majority of queries. The figure also contains a screenshot extending an image to all images of the same case. This allows to get a global picture of the comparison looking at several views of a case, for example.

4.3. Using salient features for retrieval

The last step in our prototype was to profit from local features for retrieval and use SIFT (Scale Invariant Feature Transform) features described in.[28] We hoped that such a system would allow for more variation in illumination

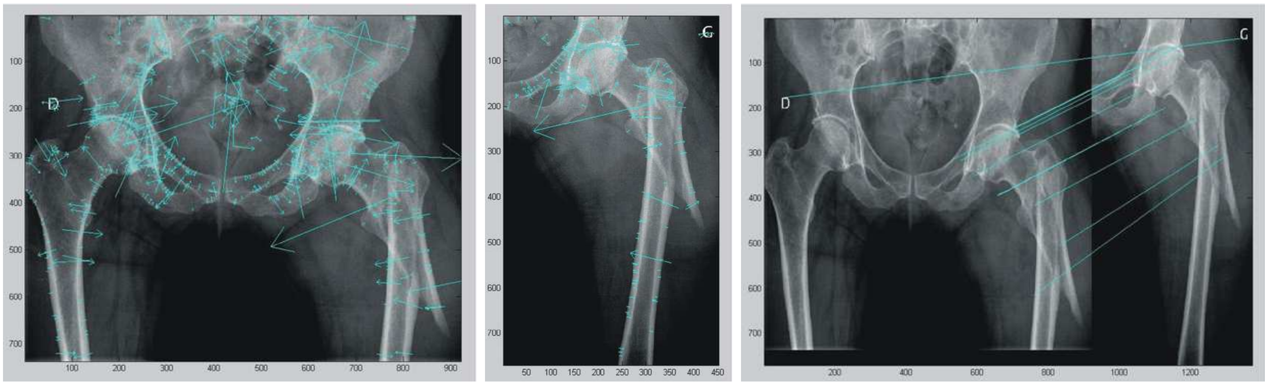

Figure 7. Correspondence of salient points between an example image and a copied subpart of this image..

and small anatomic changes between patients and still find similar fractures. A simple system for SIFT features was made available by the authors on the Internet[‡‡] and we indexed an image and a subpart of the image with the system to extract salient points. Figure 7 shows the results of the image and a subpart of the same image, plus the correspondences found between the two images. The first image contains 702 salient points and the second one 211. Surprisingly, only 13 correspondences are found between an image and a subpart of it. This is extremely surprising result and needs further clarification. We also tried to compare several images of the same anatomic region but the correspondence of the salient points was always very low.

To properly apply the SIFT features it seems necessary to well adapt the algorithm for extraction and matching to the medical x–ray images that we are using. Unfortunately in the course of this project it was not possible to perform such an in–depth work but this is foreseen in the continuation of the project. Currently, this is a very active research so many new techniques have been developed over the last few years.

4.4. Further problems identified for better retrieval

A variety of problems became clear in the course of this project, particularly for using an image retrieval system with global features and little invariances such as *medGIFT*:

- Illumination of the images can change within a large range and it is very hard to normalise.
- Images are taken using varying parameters and a varying set of background, and particularly in a teaching file cropped images can occur requiring for shift and scale invariance to be implemented.
- The region of interest is usually very small and the image retrieval system needs to be concentrated onto this region for optimal performance.
- Retrieval needs to be case–based and not image–based also for the retrieval step and not only for showing results on screen.
- Clinical parameters need to be taken into account for comparing cases, such as age and weight of the patient.
- Retrieval among only pre–operative images would be an advantage as compared to retrieval among the entire data set.
- When using a large database there need to be a two-step process, first identifying the anatomic region of the image and then searching in this sub part of the database for really similar cases.

Based on these identified problems a new retrieval prototype is foreseen.

[‡‡] `http://user.cs.tu-berlin.de/~nowozin/libsift/`

5. NEXT STEPS FOR FRACTURE RETRIEVAL

It become clear that in a 6–month project as the one described in this paper, only a feasibility analysis is possible and not the construction of a system running in practice. Still, many important parameters have been found that are required for a practical system to be efficient.

First, a pre–treatment of the data set and particularly of the images seems necessary to manually separate pre– and post–operative cases and be able to limit retrieval to a subset of the data. The retrieval has to be performed on the basis of a case and not a single image, not only for showing results but also for the retrieval. Normalisations need to be developed that take into account that similar cases might not have the same number of images and might also have slightly varying views. It needs to be made clear how to perform separate queries for each view of a cases and how to combine the results. Besides all the images of the cases, other meta data need to be identified such as the age and weight of the patient that can play an important role in the choice of a surgical technique. We are currently looking at automatic text analysis to see how much of these data is currently available in the free text and whether this can be extracted in an automatic fashion at least for those cases where it plays an important role. Of course it would be much better to also have images two or three years after the intervention to show the long–term quality of the operation but this will be hard to realise in the current project.

With respect to visual features it becomes clear that global features are not well adapted for fracture retrieval although the results are sufficient for general browsing. Salient feature points can be a solution once they can be adapted to x–ray images as their extraction might be slightly different from general stock photography. There comparison also needs to be optimised as a significant amount of variation will remain. A pre–treatment or normalisation of the grey levels can also help with such an application as illumination differences among similar images are enormous and currently there are little normalised models as they exist, for example, for mammography. Ultimate goal to focus retrieval would be a *fracture detector* that automatically find the region of interest in the images. For complete fractures this seems possible based on the high frequencies that occur in these regions but for partly fracture this is even different to see for a human eye.

6. CONCLUSIONS

Image retrieval can well be used for the planning of surgical operations for supplying similar cases, even in its current imperfect state. It becomes clear that when using the system a surgeon can quickly sort out false positive cases and concentrate on the few relevant cases supplied by a retrieval system. A variety of problems for similarity retrieval have been identified in the pilot study described in this paper (different luminosity of images, small region of interest, case–based retrieval instead of image–based retrieval, ...). The large database of fracture images serves as an important knowledge base for this purpose and its continuous growth will help us to improve the system. Longer term analyses containing images one or two years after the intervention would even improve the quality of the database but this is not always easy to realise as patients move between hospitals and their data is only rarely combined in one single place. Using a large archive of cases published in surgical journals would even be a stronger knowledge base and would allow for a better use of such a system. First initiatives to share images are under way[29] and a first system to retrieve images form cases described in journals already exists in *Goldminer* of the American Roentgen Ray Society. Scientists need to pressure the large publishing companies to make the knowledge available that is stored in the scientific research literature to advance the retrieval of visual information.

7. ACKNOWLEDGEMENTS

This work was partially supported by the Swiss National Science Foundation (Grants 632–066041 and 205321–109304/1).

REFERENCES

1. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content–based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22 No 12**, pp. 1349–1380, 2000.

2. N.-S. Chang and K.-S. Fu, "Query–by–pictorial–example," *IEEE Transactions on Software Engineering* **SE 6 No 6**, pp. 519–524, 1980.
3. V. N. Gudivada and V. V. Raghavan, "Content–based image retrieval systems," *IEEE Computer* **18**, pp. 18–22, 1995.
4. Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content–based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology* **8**, pp. 644–655, September 1998. (Special Issue on Segmentation, Description, and Retrieval of Video Content).
5. C.-R. Shyu, A. Kak, C. Brodley, and L. S. Broderick, "Testing for human perceptual categories in a physician–in–the–loop CBIR system for medical imagery," in *IEEE Workshop on Content–based Access of Image and Video Libraries (CBAIVL'99)*, pp. 102–108, (Fort Collins, Colorado, USA), June 22 1999.
6. H. D. Tagare, C. Jaffe, and J. Duncan, "Medical image databases: A content–based retrieval approach," *Journal of the American Medical Informatics Association* **4**(3), pp. 184–198, 1997.
7. H. J. Lowe, I. Antipov, W. Hersh, and C. Arnott Smith, "Towards knowledge–based retrieval of medical images. The role of semantic indexing, image content representation and knowledge–based retrieval," in *Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA)*, pp. 882–886, (Nashville, TN, USA), October 1998.
8. H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content–based image retrieval systems in medicine – clinical benefits and future directions," *International Journal of Medical Informatics* **73**, pp. 1–23, 2004.
9. A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, "Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment," *Radiology* **228**, pp. 265–270, 2003.
10. T. M. Lehmann, M. O. Güld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, and B. B. Wein, "Content–based image retrieval in medical applications," *Methods of Information in Medicine* **43**, pp. 354–361, 2004.
11. M. O. Güld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann, "Quality of DICOM header information for image categorization," in *International Symposium on Medical Imaging*, *SPIE Proceedings* **4685**, pp. 280–287, (San Diego, CA, USA), February 2002.
12. H. Müller, A. Rosste, A. Garcia, J.-P. Vallée, and A. Geissbuhler, "Benefits from content–based visual data access in radiology," *RadioGraphics* **25**, pp. 849–858, 2005.
13. C. Le Bozec, E. Zapletal, M.-C. Jaulent, D. Heudes, and P. Degoulet, "Towards content–based image retrieval in HIS–integrated PACS," in *Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA)*, pp. 477–481, (Los Angeles, CA, USA), November 2000.
14. W. Hersh, H. Müller, J. Jensen, J. Yang, P. Gorman, and P. Ruch, "Imageclefmed: A text collection to advance biomedical image retrieval," *Journal of the American Medical Informatics Association* **September/October**, pp. 488–496, 2006.
15. H. Müller, T. Deselaers, T. M. Lehmann, P. Clough, K. Eugene, and W. Hersh, "Overview of the imageclefmed 2006 medical retrieval and medical annotation tasks," in *CLEF 2006 Proceedings*, *Lecture Notes in Computer Science*, 2007 – to appear.
16. L. Setia, A. Teynor, A. Halawani, and H. Burkhardt, "Image classification using cluster–cooccurrence matrices of local relational features," in *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (MIR 2006)*, (Santa Barbara, CA, USA), Oct 2006.
17. C. Schmid and R. Mohr, "Local greyvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, pp. 530–535, May 1997.
18. M. E. Müller, *Müller AO classification of fractures — long bones*, AO Publishing, Davos, Switzerland, 2006.
19. T. P. Rüedi and W. M. Murphy, *AO principles of fracture management*, AO Publishing, Davos, Switzerland, 2001.
20. R. Stern, P. Hoffmeyer, A. Rosset, and J. Garcia, *Fractures*, University of Geneva, Geneva, Switzerland, 2003.
21. A. Rosset, H. Müller, M. Martins, N. Dfouni, J.-P. Vallée, and O. Ratib, "Casimage project – a digital teaching files authoring environment," *Journal of Thoracic Imaging* **19**(2), pp. 1–6, 2004.

22. D. M. Squire, W. Müller, H. Müller, and T. Pun, "Content–based query of image databases: inspirations from text retrieval," *Pattern Recognition Letters (Selected Papers from The 11th Scandinavian Conference on Image Analysis SCIA '99)* **21**(13-14), pp. 1193–1198, 2000. B.K. Ersboll, P. Johansen, Eds.
23. H. Müller, D. M. Squire, and T. Pun, "Learning from user behavior in image retrieval: Application of the market basket analysis," *International Journal of Computer Vision* **56(1–2)**, pp. 65–77, 2004. (Special Issue on Content–Based Image Retrieval).
24. A. Jain and G. Healey, "A multiscale representation including opponent color features for texture recognition," *IEEE Transactions on Image Processing* **7**, pp. 124–128, January 1998.
25. J.-M. Geusebroek, R. van den Boogaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(12), pp. 1338–1350, 2001.
26. M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision* **7**(1), pp. 11–32, 1991.
27. H. Müller, D. M. Squire, W. Müller, and T. Pun, "Efficient access methods for content–based image retrieval with inverted files," in *Multimedia Storage and Archiving Systems IV (VV02)*, S. Panchanathan, S.-F. Chang, and C.-C. J. Kuo, eds., *SPIE Proceedings* **3846**, pp. 461–472, (Boston, Massachusetts, USA), September 20–22 1999.
28. D. G. Lowe, "Object recognition from local scale invariant features," in *Proceedings of the International Conference of Computer Vision*, (Corfu, Greece), Sep 1999.
29. M. W. Vannier and R. M. Summers, "Sharing images," *Radiology* **228**, pp. 23–25, 2003.