Multiscale Salient Point–Based Retrieval of Fracture Cases

Xin Zhou[a], Richard Stern[b], Henning Müller[ac]

[a]Service of Medical Informatics, Geneva University Hospitals and University of Geneva, Rue Gabrielle–Perret–Gentil 4, 1211 Geneva 14, Switzerland;
[b]Division of Orthopaedics and Trauma Surgery, Geneva University Hospitals, Rue Gabrielle–Perret–Gentil 4, 1211 Geneva 14, Switzerland;
[d]University of Applied Sciences Western Switzerland (HES SO), TechnoArk 3, 3960 Sierre, Switzerland

ABSTRACT

Fractures are common injuries, and some more complicated fractures require a surgical intervention. When such an operation is planned it can be beneficial to have access to similar past cases including follow ups to compare, which method might be the most adapted one in a particular situation. In the surgery department of the University hospitals of Geneva a database of past cases including pre– and post–operative images and case descriptions has been created over the past years with the goal to support clinical decision making.

Images play an important role in the decision making and the judgment of a fracture, but image content is currently not directly accessible for search. At the moment, search is mainly via a classification system of the fractures or in the patient record itself only by patient ID. In this paper we propose a solution that combines visual information from the images in a case to calculate similarity between cases and allow thus an access to visually similar cases. Such a system can complement the text– or classification–based search that has been used so far.

In a preliminary study, we used pixel–grid–based salient–point features to build a first prototype of case–based visual retrieval of fracture cases. Cases belonging to different fracture classes were beforehand often confused due to the similar bone structures in the various images. In this article, a multi–scale approach is used in order to perform similarity measures at both large and small scales. When compared to the first prototype, the introduction of scale and spatial information allowed improving the performance of the system. Cases containing similar bone structures but with dissimilar fractures are generally ranked lower whereas more relevant cases are returned. The system can thus be expected to perform sufficiently well for use in clinical practice and particularly for teaching.

Keywords: content–based image retrieval, fracture database, medical imaging, decision support system

1. INTRODUCTION

Fractures are a common medical condition, and for most fractures no surgical intervention is necessary. More complicated fractures sometimes require surgical interventions where images play an important role in planning the intervention. For classifying fractures the AO/OTA Classification classification was developed.[1] This allows classifying fractures by the bone, the exact place of the fracture in the bone and the complexity of the fracture.

Several databases such as teaching files exist that include fractures. For example, the University hospitals of Geneva have stored cases of fractures requiring an operation in a teaching file called Casimage* since several years.[2] In general, cases are stored with images before the operation, immediately after an operation and then whenever the patient comes for a follow up or control visit. Subsets of these data have been made available for example in the form of teaching CDs and books.[3] Access to these cases is usually via a structure of the fracture classification. In the patient record, access to cases can also be by the patient identification.

In terms of image processing regarding fracture images several applications have been proposed in the past. Automatic detection of bone fractures, for example, has been proposed several times.[4,5] Leow et al.[4,6,7] worked on automatic femur fracture detection in X–ray images and Donnelley et al.[8–10] on long bone fracture detection.[4,6,7] A common approach is to perform a contour–based bone segmentation and then to use a gradient–map–based texture analysis to localize the fracture. The performance of this approach relies mainly on the bone contour modeling. Ideally, the modeling should be flexible enough to cover fractures of all types of bones, though.

Image retrieval on fracture images has been proposed in the past as well but to our knowledge only quite rarely.[11] In general, there are different types of image retrieval systems, mainly those using textual information and those using directly the visual information of the images.[12] Systems using the visual information of images directly are also called content–based image retrieval (CBIR) systems. Such CBIR systems have been proposed for general images many times and hundreds of articles have been published on this with a large variety of approaches.[13,14] In the medical domain content–based image retrieval has also been proposed several times[15–17] and some concrete systems exist.[18] Typical application domains are the retrieval of articles form the medical literature or the analysis of the texture in lung CTs.[19,20]

To evaluate image retrieval algorithms for visual and textual retrieval systems the ImageCLEF[†] benchmark (part of CLEF[‡], the Cross–Language Evaluation Forum) was started in 2003. Since 2004, a medical retrieval task was added, first working with teaching files and then later with images from the medical literature.[21] This benchmark has been run every year since and has led to a variety of interesting observations regarding the techniques employed. Tasks for images classification as well as for retrieval were evaluated in separate tasks in the competition.[21,22] Most often, for the purely visual tasks multi–scale salient point–based approaches were used[23,24] and outperformed the other visual techniques.[21,25,26] The combination of SIFT (Scale–Invariant Feature Transform)[27] features and the BoF (Bags of Features) image representation[28] is commonly used in the benchmark. As this representation does not take into account spatial distribution Nowak et al.[29] proposed to keep the SIFT descriptor but replace the multi–scale salient point detector by random sampling, which in some experiments gave equal or better results.[23,24]

Goal of the visual retrieval of fractures is to supply to the clinician cases of the past of which the images are visually similar to cases stored in the database. The approach is similar to the well know case–based reasoning technique.[30] Such cases of the past can help to decide on the potential techniques for an intervention taking into account the choices of other clinicians and also based on the long–term outcome on particular patients.

Challenges include difference in the basic data recorded for each case as the exact views and absolute grey–scale of the x–ray images can vary quite strongly. Some cases might also have more images than others so incomplete information needs to be taken into account. The retrieval also needs to be on a case basis and not on an image basis, so similarity values of image in a case need to be combined for a similarity score of cases. As x–rays are often taken under standardized conditions, images of the same bone but at difference places often contain very similar views and only differ in very small parts of the image. Differences in anatomy between the persons thus are often more important than the difference due to a small fracture in a slightly different region. As some cases contain many CT images whereas others do not, there also needs to be a normalization at this level to avoid unwanted influences.

The prototype of bone fracture retrieval described in this paper is based on a pixel–grid sampling using a SIFT descriptor followed by BoF representation. As SIFT has been combined successfully with image representations such as spatial pyramids,[31] sparse flexible models[32] and cluster co–occurrence matrices,[33] in this article we implement a multi–scale SIFT detector and compare several case representation models.

This article is structured as follows: Section 2 describes the material and methods of the article, namely the database used and the basic techniques reused. Chapter 3 then describes the performance results obtained before the article analyzes the obtained results critically. The article then finishes with a conclusion.
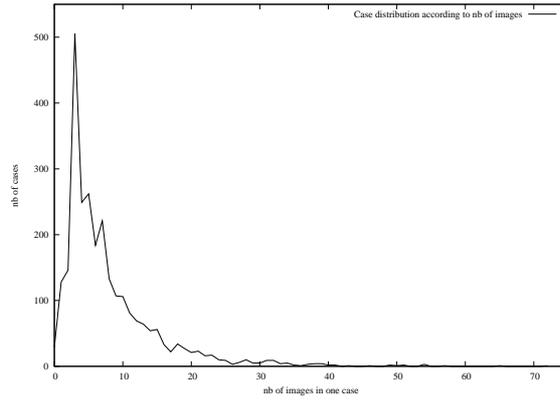
2. METHODS

This section describes the basic methods used for the work described

Figure 1. Number of cases with the respective number of images per case.

2.1. Data set used

At the surgery department of the University Hospitals of Geneva, cases of fractures including images and textual descriptions have been collected over the past more than ten years. In this article, a sub set of 23'970 images was extracted for the demonstration system. Each case in the database has images before and after the operation and wherever possible follow up visits even several years after the operation are added to the cases to see the evolution of the operation The 23'970 images used in this study are part of 2'690 fracture cases associated with 43 different fracture types based on AO/OTA classification. For malleolar (ankle) fractures, Danis–Weber classification[34, 35] is used. Each case contains from 1 to 73 images of various modalities, mainly x–ray images. In Figure 1 the distribution of the cases based on the number of images is given.

Table 1 shows the number of cases, the total number of images and their average for each. Fracture classes are available as textual labels that correspond to particular classes in the AO classification.

Around 90% of the images are x–ray images. Other image modalities include CT (Computed Tomography), MRI (Magnetic Resonance Imaging), Angiography, Scintigraphy, 3D reconstruction images, hand–drawn surgery plans and photos of injuries. X–ray images can be taken from different views and have a varying size. Normally, at least 2 x–rays are contained in a case (one antero–posterior view and one lateral view). Sometimes a third view, oblique or external rotation is given. Several cases also included a magnification of the region of interest.

2.2. Techniques

CBIR usually includes an offline image processing part to pre–process the data and then online query processing that includes calculating the similarity of a query case with past cases. In our case, offline image processing is composed of pre–processing, visual feature extraction, feature clustering to create a visual vocabulary and finally image/case modeling. The online query processing step is composed of feature extraction if a new case is treated, similarity calculation for each image in the query and then a case–based fusion to calculate similarity values on a per case and not per image basis.

Contrast normalization is performed as a pre–processing step based on histogram equalization. Multi–scale image analysis is performed by using SIFT.[27] SIFT features were extracted from 9 levels (3 octaves with 3 layers for each). The clustering algorithm is a griddified version of hierarchical KMeans quantizer, which was presented in.[36] Image representation is based on the BOF.[28]

Two distance functions were used for similarity calculation:

- the SSD(Sum of Squared Difference) distance (also known as Euclidean distance or 2nd order Minkowski distance);
- the HI(Histogram Intersection).

Table 1. Number of cases and images of each fracture class.

fracture class	nb of cases	nb of images	average nb of images per case
Acetabulum	31	433	13.97
Ankle	8	60	7.50
Ankle Weber A	44	270	6.14
Ankle Weber B	244	1784	7.31
Ankle Weber C	156	1222	7.83
Calcaneus	33	340	10.30
Carpal–Other	1	5	5.00
Clavicle	40	232	5.80
Elbow	4	21	5.25
Femur–Subtrochanteric	132	1342	10.17
Femur Diaphysis	169	1960	11.60
Femur Distal–Extraarti	55	614	11.16
Femur Distal–Intraarti	50	673	13.46
Femur Proximal–Head	2	18	9.00
Femur Proximal–Intertr	54	406	7.52
Femur Proximal–Neck	72	499	6.93
Femur Proximal–Pertroc	419	2132	5.09
Foot	3	36	12.00
Hip	3	25	8.33
Humerus Diaphysis	119	1146	9.63
Humerus Distal–Extraar	30	325	10.83
Humerus Distal–Intraar	61	590	9.67
Humerus Proximal	172	1522	8.85
Knee	7	34	4.86
Metacarpal-Phalanx hand	10	30	3.00
Metatarsal-Phalanx foot	62	476	7.68
Patella	34	198	5.82
Pelvic Ring Fracture	43	438	10.19
Radius/Ulna Diaphysis	46	262	5.70
Radius/Ulna Distal	14	68	4.86
Radius/Ulna Proximal	58	357	6.16
Scapula	10	131	13.10
Shoulder	16	73	4.56
Spine Cervical	1	1	1.00
Spine Lumbar	2	8	4.00
Spine Thoracic	1	3	3.00
Talus	32	497	15.53
Tarsal–Other	17	244	14.35
Tibia/Fibula Diaphysis	205	2160	10.54
Tibia/Fibula Distal–Ex	52	737	14.17
Tibia/Fibula Distal–In	57	599	10.51
Tibia/Fibula Proximal–Ex	23	310	13.48
Tibia/Fibula Proximal–In	98	1689	17.23
total	2690	23970	8.94

Distances $\mathcal{D}$ are calculated between a pair of models. Let $\mathcal{C}$ be the set of cases and C_i represents a single case i. To calculate the similarity, C_i is represented by a set of models, named $\mathcal{M}(i)$. Using $\mathcal{M}(i)$ as a query, each model $M(i)_k$ of $\mathcal{M}(i)$ will generate a list of models $\mathcal{M}_{similar}(k)$ ranked by the distance, among which several models can refer to the same case C_j. Fusion of first level is performed to convert the list of similar models $\mathcal{M}_{similar}(k)$ into a list of similar cases $\mathcal{C}_{similar}(k)$. As $\mathcal{M}(i)$ can contain more than one model, more than one list of similar cases are produced. Fusion of second level will combine all lists into one.

Fusion strategies are based on classical fusion strategies combMAX, combMIN, combSUM, combMNZ proposed in[37] and and combSUM(n)MAX proposed in.[38]

The equations are listed to make the various combination rules explicit:

$$D_{\texttt{combMAX}}(m) = \arg \max_{l=1:N} D(m)_l, \tag{1}$$

$$D_{\texttt{combMIN}}(m) = \arg \min_{l=1:N} D(m)_l, \tag{2}$$

$$D_{\texttt{combSUM}}(m) = \sum_{l=1}^{N} D(m)_l, \tag{3}$$

$$D_{\texttt{combMNZ}}(m) = D_{\texttt{combSUM}}(m) * F(m), \tag{4}$$

$$D_{\texttt{combANZ}}(m) = D_{\texttt{combSUM}}(m)/F(m), \tag{5}$$

$$D_{\texttt{combDANZ}}(m) = D_{\texttt{combANZ}}(m)/F(m). \tag{6}$$

where m is one returned model, $D(m)$ is its distance to query model, l is one list of returned model and $F(m)$ is the number of times that m appeared in all lists. $\texttt{combSUM(}n\texttt{)MAX}$ is proposed have a trade–off between *combMAX* and *combSUM*.

$$V_{\texttt{combSUM(}n\texttt{)MAX}}(i) = \sum_{l=1}^{N} \arg \max_{k \in \mathcal{E}_N \setminus \mathcal{E}_l} (\overline{D(m)_l}), \tag{7}$$

Here n is a parameter. When $n = 1$, only 1 maximum is taken, $\texttt{combSUM(}n\texttt{)MAX}$ is equivalent to $\texttt{combMAX}$. When $n = N$, $\texttt{combSUM(}n\texttt{)MAX}$ is equivalent to $\texttt{combMAX}$. In this article $n = 2$ is chosen as proposed in.[38]

Similarity is represented by small SSD value or large HI value. En consequence, combMIN + combDANZ are used to combine SSD distances, and combMAX + combMNZ for HI distances.

2.3. Evaluation methodology

Results presented previously using similar techniques in[39] were produced by a random selection of only 10 cases from each class as query examples. Only early precision P10 and P30 were presented, and cases with less than 10 images were automatically discarded. In this article, every case was taken to query the database. MAP (Mean Average Precision) was calculated retrieving the most similar 1000 cases. Both, MAP and a class–based geometric MAP are presented. Early precision P10 and P30 are also shown, as early precision might be what most end users are actually interested in.

3. RESULTS

This section details the main results of the article and a performance evaluation.
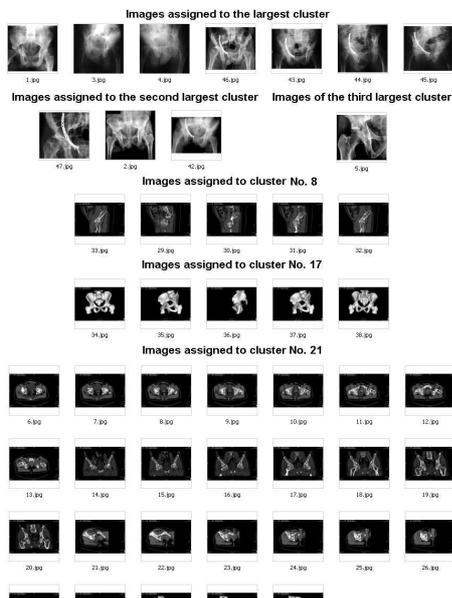
Figure 2. Images clustering in one large example case: Case Acetabulum #5

3.1. Novel techniques developed

Fracture classification is based upon the fracture location and morphology. Not all the images in one case represent location and morphology well. Images of antero–posterior view or lateral view usually show very well the bone architecture as well as the fracture location. Other images such as CT present a intersectional view, which is less relevant for fracture classification. Less relevant views can provide useful information, but in case that one case contains more irrelevant views than relevent ones, the system will return low quality results.

We tried to improve the performance by emphasizing the representation models of relevant images should be highly weighted whereas the number of less relevant views should be reduced. One simple way is to separate images by thumbnails, as the thumbnails of images taken from same view are usually very similar. To reduce the computation time, images are converted into 16x16 thumbnails. Separation is performed by KMean clustering algorithm is used, number of clusters k varies from 10 to 100. In Figure 2, one example case containing various types of images is shown. $k = 70$ is used in this example. The same color meaning that images are of the same cluster. Almost all the images of antero–posterior view are associated to red, which is the largest cluster of the database. As this phenomena is widespread in Acetabulum, Femur, Humerus and Ankle fracture cases, which

Table 2. Minimum, maximum and average numbers of images assigned to clusters

k	min	average	median	max
10	406	2402.3	1883.5	7747
20	175	1201.2	1068.5	6995
30	1	800.8	640.5	6760
40	68	600.6	455	6092
50	17	480.5	385	6211
60	48	400.4	290.5	6182
70	34	343.2	264.5	5583
80	37	300.3	239	5428
90	34	266.9	204.5	5485
100	23	240.2	192	5001

are 80% of the whole database. Emphasizing antero–posterior view images can be approximated to emphasizing the images associated to the largest cluster. One way is to reduce the numbers of representation vectors for other type of images by aggregating their BoF vectors into one. The reason of choosing this approach is also due to the fact that BoF model has better performance when the number of features is larger.

Let $\mathcal{I}(i)$ be the set of images of case C_i, and each image $I(i)_j$ associated to cluster s ($1 <= s <= k$) is noted $I(i)-> s$. By using BoF+SIFT feature extraction, $I(i)_j$ is represented by a vector V_j. The dimension of vector V depends on the number of keywords pre–defined. In[23] 1000 keywords was proved to be a stable choice and is used in our experiments. For images $I(i)-> \overline{s}$ where $\overline{s}$ is not the largest cluster, BoF vectors are aggregated into one vector $V-> s$ by calculate the average value for each bin. Therefore, case C_i is represented by $N_{s(i)} + K_{\overline{s(i)}}$ representation vectors, where $N_s(i)$ is the number of images associated to s in C_i, and $K_{\overline{s}}$ is the number of clusters appeared in C_i. This case model is named BoF(cluster) in the rest of the article.

The second novelty is to develop a tree–based image representation to replace BoF. BoF is often critised due to the lack of spatial information of local features, and it is argued that the most important spatial information for one feature comes from its nearest–neighbors.[31, 32] We propose to structure each SIFT feature as a descendant of the nearest–neighbor feature in neighbor scale, which forms a tree structure that takes into account the spatial information as well as the multiscale information. Let $\mathcal{F}$ be the set of SIFT feature extracted from one image. Let $\{$ be a single feature. $\mathcal{F}_\sigma$ is the subset of features extracted at scale σ. σ_0 is the bottom level and σ_8 is the top level. As feature extraction is always extracted from bottom to top, when $\{_{\sigma_i}$ is extracted, all features of $\mathcal{F}_{\sigma_j}(j < i)$ are already known. The location information of $\mathcal{F}_{\sigma_j}$ can be used to find out the lower scale nearest–neighbors. Level by level a tree of SIFT features is established, where low scale level features are always linked to the nearest higher scale level features.

As KMeans clustering is performed both on SIFT descriptor vectors ($k = 1000$) and thumbnail vectors ($k = 70$), each node of the tree is then replace by the cluster index. The classical TED (Tree Edit Distance) function[40] is used as distance function. Two trees are considered as similar if only small nubmer of nodes need to be moved to two trees identical. This approach is named MultiScaleTree(image) in the rest of the article.

3.2. Performance evaluation

In thumbnails clustering step, k is the parameter to be defined. From 10 and 100 clusters were tested. Table 2 shows per cluster how many images on average are contained in each cluster as well as the maximum and minimum size in the clusters. The largest cluster contains 20–30% of the images of the database depending on the number of clusters. In Figure 3 early precision as well as MAP obtained by BoF(cluster) with two distance methods (SSD and HI) and various numbers of clusters k is given. MAP is little influenced by the value of k, especially using SSD distance function. HI gave better results then SSD distance fucntions. The best results are obtained for $k = 30$ for HI distance, and $k = 70$ for SSD distance.
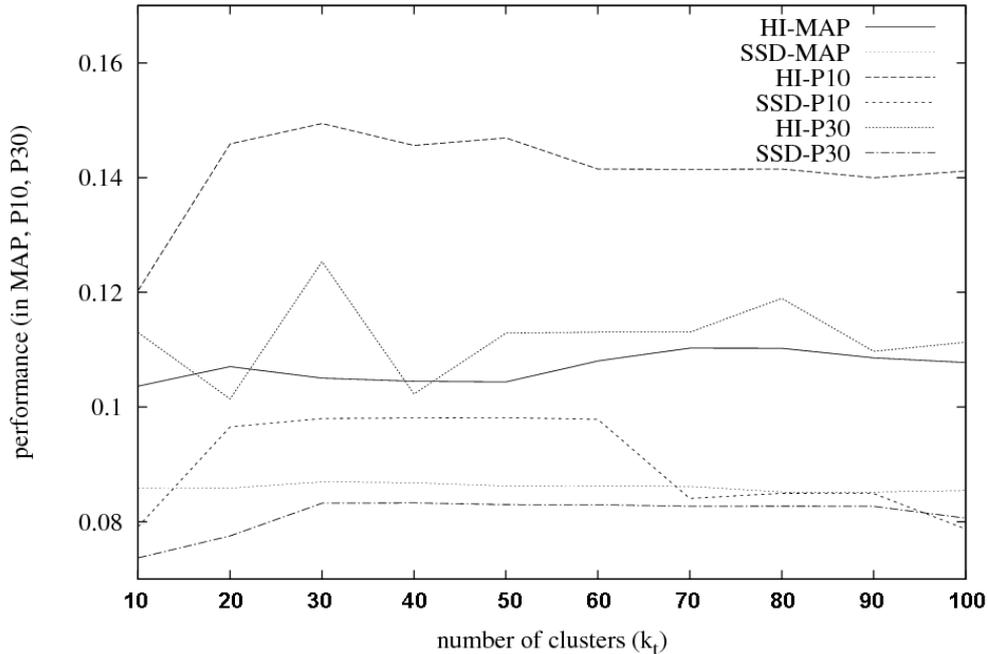
Figure 3. Performance obtained by SIFT+BoF(cluster) with various number of k.

Table 3. Performance comparison using various sampling strategies, representation models, distance measure and fusion approaches are shown (HI = Histogram Intersection; SSD is Sum of Squared Difference; TED is Tree Edit Distance.)

Sampling	Models	Distance	Fusion	MAP	P10	P30
PixelGrid	BoF(image)	HI	combMAX+combMNZ	0.094	0.117	0.110
SIFT detector	BoF(cluster)	HI	combMAX+combMNZ	**0.110**	0.141	**0.113**
PixelGrid	BoF(image)	HI	combSUM(2)MAX+combMNZ	0.099	0.105	0.106
SIFT detector	BoF(cluster)	HI	combSUM(2)MAX+combMNZ	0.102	0.133	0.096
PixelGrid	BoF(image)	SSD	combMIN+combDANZ	0.076	0.099	0.053
SIFT detector	BoF(cluster)	SSD	combMIN+combDANZ	0.087	0.097	0.083
SIFT detector	MultiScaleTree(image)	TED	combMIN+combDANZ	0.106	**0.165**	0.097

Evaluation of BoF(cluster) model and MultiScaleTree(image) model is performed and classical BoF image representation is used as a baseline, noted BoF(image). For MultiScaleTree(image) model, the number of thumbnail clusters does not have important impact as it only influence one node. In Table 3, early precision and MAP obtained by three representation models with associated distance functions and fusion strategies are shown. The best overall MAP is obtained by using BoF(cluster) with histogram intersection and combMAX+combMNZ for fusion. For all approaches, BoF(cluster) outperforms the classical BoF model. In terms of distance functions and fusion strategies, histogram intersection produce always better results than SSD distance. It can be that the combMIN+combDANZ fusion strategies are not effective for SSD distance. Using the combSUM(2)MAX operator for score fusion reduced the early precision and sometimes also MAP. CombMAX together with combMNZ remains the most stable choice for the fusion of histogram intersection. Using a nearest–neighbor tree structure for SIFT produced the best early precision (P10).

Table 4. Performance comparison per class

fracture class	BoF (image)	BoF (cluster)	MultiScaleTree (image)
Acetabulum	0.012	0.048	0.116
Ankle	0.023	0.172	0.052
Ankle Weber A	0.068	0.040	0.044
Ankle Weber B	0.184	0.116	0.098
Ankle Weber C	0.072	0.100	0.094
Calcaneus	0.052	0.056	0.067
Clavicle	0.056	0.149	0.077
Elbow	0.010	0.005	0.021
Femur Diaphysis	0.140	0.147	0.159
Femur Distal–Extraarti	0.036	0.034	0.063
Femur Distal–Intraarti	0.056	0.019	0.022
Femur Proximal–Head	0.028	0.022	0.016
Femur Proximal–Intertr	0.016	0.020	0.026
Femur Proximal–Neck	0.040	0.076	0.103
Femur Proximal–Pertroc	0.236	0.251	0.217
Femur Subtrochanteric	0.062	0.101	0.060
Foot	0.134	0.219	0.175
Hip	0.120	0.435	0.151
Humerus Diaphysis	0.072	0.035	0.111
Humerus Distal–Extraar	0.024	0.033	0.038
Humerus Distal–Intraar	0.028	0.133	0.032
Humerus Proximal	0.152	0.200	0.146
Knee	0.012	0.186	0.033
Metacarpal-Phalanx hand	0.099	0.261	0.212
Metatarsal-Phalanx foot	0.108	0.114	0.104
Patella	0.052	0.008	0.004
Pelvic Ring Fracture	0.056	0.056	0.033
Radius/Ulna Diaphysis	0.052	0.038	0.032
Radius/Ulna Distal	0.020	0.010	0.021
Radius/Ulna Proximal	0.044	0.019	0.016
Scapula	0.011	0.028	0.021
Shoulder	0.008	0.014	0.004
Talus	0.052	0.030	0.061
Tarsal–Other	0.020	0.025	0.031
Tibia/Fibula Diaphysis	0.176	0.179	0.162
Tibia/Fibula Distal–Ex	0.072	0.047	0.110
Tibia/Fibula Distal–In	0.064	0.030	0.083
Tibia/Fibula Proximal–Ex	0.115	0.051	0.126
Tibia/Fibula Proximal–In	0.132	0.166	0.172
Geometric average	0.072	0.092	0.0798

In Table 4, an additional analysis is given for each fracture class. The 6 largest classes contain half the images of the database. The majority classes usually work well in classification problems. As a consequence, the geometric mean is lower than the overall MAP. Using BoF (cluster) obtained the best results also in this case. Using the tree structure obtained the second place, but the MAP for each class is more uniform than for the other two methods, which shows that a multi–scale tree structure is more resistant to varying size of classes.

4. CONCLUSIONS

In this article two new SIFT feature–based representation models are proposed. The goal is to improve the performance of visual information retrieval for fracture cases. Classical BoF model was used as a baseline for evaluation. The BoF vector aggregation based on thumbnail clustering improved the performance of all approaches. Retrieval of femur fractures remains challenging as several classes contain the exact bone structures and differences between these images and cases are very small. The performance of retrieval using BoF model is dependant to the size of the class. The multi–scale representation, however, increased the performance and allowed a better separation of the cases and more particularly of the images of the cases, thus obtain more uniform performance for all the classes.

5. ACKNOWLEDGMENTS

REFERENCES

1. J. L. Marsh, T. F. Slongo, J. Agel, J. S. Broderick, W. Creevey, T. A. DeCoster, L. Prokuski, M. S. Sirkin, B. Ziran, B. Henley, and L. Audigé, "Fracture and dislocation classification compendium - 2007: Orthopaedic trauma association classification, database and outcomes committee," *J Orthop Trauma.* **21(10 Suppl)**, pp. S1–133, 2007.
2. A. Rosset, H. Müller, M. Martins, N. Dfouni, J.-P. Vallée, and O. Ratib, "Casimage project — a digital teaching files authoring environment," *JTI* **19**(2), pp. 1–6, 2004.
3. R. Stern, P. Hoffmeyer, A. Rosset, and J. Garcia, *Fractures*, University of Geneva, Geneva, Switzerland, 2003.
4. S. E. Lim, Y. Xing, Y. Chen, W. K. Leow, T. S. Howe, and M. A. Png, "Detection of femur and radius fractures in x–ray images," in *In: Proc. 2nd Int. Conf. on Advances in Medical Signal and Information Processing*, pp. 249–256, (Sliema, Malta), September 2004.
5. M. Donnelley and G. Knowles, "Automated bone fracture detection," in *Society of Photo–Optical Instrumentation Engineers (SPIE) Conference Series*, J. M. Fitzpatrick and J. M. Reinhardt, eds., *Society of Photo–Optical Instrumentation Engineers (SPIE) Conference Series* **5747**, pp. 955–966, (San Diego, CA, USA), April 2005.
6. V. L. F. Lum, W. K. Leow, Y. Chen, T. S. Howe, and M. A. Png, "Combining classifiers for bone fracture detection in x-ray images," in *IEEE International Conference on Image Processing (ICIP'2005)*, **1**, pp. 1149–1152, (Genoa, Italy), September 2005.
7. J. C. He, W. K. Leow, and T. S. Howe, "Hierarchical classifiers for detection of fractures in x–ray images," in *Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns (CAIP 2007), Lecture Notes in Computer Science* **4673**, pp. 962–969, Springer, (Vienna, Austria), August 2007.
8. M. Donnelley, *Computer Aided Long–bone Segmentation and Fracture Detection.* PhD thesis, Flinders University of South Australia, Adelaide, South Australia, January 2008.
9. M. Donnelley and G. Knowles, "Computer aided long bone fracture detection," in *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications (ISSPA 2005)*, **1**, pp. 175–178, (Sydney, AUSTRALIA), August 2005.
10. M. Donnelley, G. Knowles, and T. Hearn, "A cad system for long–bone segmentation and fracture detection," in *Proceedings of the 3rd International Conference on Image and Signal Processing (ICISP 2008), Lecture Notes in Computer Science* **5099**, pp. 153–162, Springer, July 2008.
11. H. Müller, P. A. Do Huang, A. Depeursinge, P. Hoffmeyer, R. Stern, C. Lovis, and A. Geissbuhler, "Content–based image retrieval from a database of fracture images," in *Medical Imaging 2007: PACS and Imaging Informatics*, S. C. Horii and K. P. Andriole, eds., *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference* **6516**, p. 65160H, (San Diego, CA, USA), March 2007.
12. P. G. B. Enser, "Pictorial information retrieval," *Journal of Documentation* **51**(2), pp. 126–170, 1995.

13. 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**, pp. 1349–1380, December 2000.

14. R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys* **40**, pp. 1–60, April 2008.

15. H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medicine–clinical benefits and future directions," *Internation Journal of Medical Informatics* **73**, pp. 1–23, February 2004.

16. 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.

17. 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.

18. 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.

19. C. E. Kahn Jr. and C. Thao, "Goldminer: A radiology image search engine," *American Journal of Roentgenology* **188**, pp. 1475–1478, 2008.

20. A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "3D case–based retrieval for interstitial lung diseases," in *MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support*, B. Caputo, H. Müller, T. Syeda-Mahmood, J. Kalpathy-Cramer, and J. Duncan, eds., *Lecture Notes in Computer Science (LNCS)*, pp. 39–48, Springer, February 2010.

21. H. Müller, J. Kalpathy-Cramer, C. E. Kahn Jr., W. Hatt, S. Bedrick, and W. Hersh, "Overview of the ImageCLEFmed 2008 medical image retrieval task," in *Evaluating Systems for Multilingual and Multimodal Information Access – 9th Workshop of the Cross-Language Evaluation Forum*, C. Peters, D. Giampiccolo, N. Ferro, V. Petras, J. Gonzalo, A. Peñas, T. Deselaers, T. Mandl, G. Jones, and M. Kurimo, eds., *Lecture Notes in Computer Science* **5706**, pp. 500–510, (Aarhus, Denmark), September 2009.

22. T. Deselaers, T. M. Deserno, and H. Müller, "Automatic medical image annotation in ImageCLEF 2007: Overview, results, and discussion," *Pattern Recognition Letters* **29**(15), pp. 1988–1995, 2008.

23. T. Tommasi, F. Orabona, and B. Caputo, "CLEF2007 image annotation task: an SVM–based cue integration approach," in *Working Notes of the 2007 CLEF Workshop*, (Budapest, Hungary), Sep. 2007.

24. U. Avni, J. Goldberger, and H. Greenspan, "TAU MIPLAB at ImageClef 2008," in *Working Notes of the 2008 CLEF Workshop*, (Aarhus, Denmark), Sep. 2008.

25. H. Müller, T. Deselaers, T. Lehmann, P. Clough, E. Kim, and W. Hersh, "Overview of the ImageCLEFmed 2006 medical retrieval and medical annotation tasks," in *CLEF 2006 Proceedings*, *Lecture Notes in Computer Science (LNCS)* **4730**, pp. 595–608, Springer, (Alicante, Spain), 2007.

26. H. Müller, T. Deselaers, E. Kim, J. Kalpathy-Cramer, T. M. Deserno, P. Clough, and W. Hersh, "Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks," in *CLEF 2007 Proceedings*, *Lecture Notes in Computer Science (LNCS)* **5152**, pp. 473–491, Springer, (Budapest, Hungary), 2008.

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

28. G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *In Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.

29. E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag–of–features image classification," in *Computer Vision – ECCV 2006, 9th European Conference on Computer Vision Proceedings, Part IV*, A. Leonardis, H. Bischof, and A. Pinz, eds., *Lecture Notes in Computer Science* **3954**, pp. 490–503, Springer, (Graz, Austria), 7-13 2006.

30. A. Aamodt and E. Plaza, "Case–based reasoning: Foundational issues, methodological variations, and systems approaches," *AIC* **7**(1), pp. 39–59, 1994.

31. S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pp. 2169–2178, IEEE Computer Society, (Washington, DC, USA), 2006.

32. G. Carneiro and D. Lowe, "Sparse flexible models of local features," in *In ECCV (9th European Conference on Computer Vision)*, pp. 29–43, SOS Printing, San Diego, CA, USA., (Graz, Austria), May 2006.

33. L. Setia, A. Teynor, A. Halawani, and H. Burkhardt, "Image classification using cluster cooccurrence matrices of local relational features," in *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pp. 173–182, ACM Press, (Santa Barbara, CA, USA), 2006.

34. R. Danis, *Théorie et Pratique de l'Ostéosynthése*, Masson and Cie, Paris, France, 1949.

35. B. Weber, *Die verletzungen des oberen sprungge-lenkes*, Aktuelle Probleme in der Chirurgie., Stuttgart: Huber, 1966.

36. X. Zhou, H. Krabbenhöft, M. Niinimäki, A. Depeursinge, S. Möller, and H. Müller, "An easy setup for parallel medical image processing: Using Taverna and ARC," in *Proceedings of HealthGrid 2009*, (Berlin, Germany), 2009.

37. E. A. Fox and J. A. Shaw, "Combination of multiple searches," in *Text REtrieval Conference*, pp. 243–252, 1993.

38. X. Zhou, A. Depeursinge, and H. Müller, "Information fusion for combining visual and textual image retrieval," in *Pattern Recognition, International Conference on*, IEEE Computer Society, (Los Alamitos, CA, USA), 2010.

39. X. Zhou, A. Depeursinge, R. Stern, C. Lovis, and H. Müller, "Case–based visual retrieval of fractures," in *International Journal of Computer Assisted Radiology and Surgery*, **5, Supplement 1**, pp. S162–S163, Springer, (Geneva, Switzerland), June 2010.

40. P. Bille, "A survey on tree edit distance and related problems," *Theor. Comput. Sci.* **337**(1–3), pp. 217–239, 2005.