Automatic segmentation of whole-body bone scintigrams as a preprocessing step for computer assisted diagnostics

Tracking number: C0114

Abstract. Bone scintigraphy or whole-body bone scan is one of the most common diagnostic procedures in nuclear medicine used in the last 25 years. Pathological conditions, technically poor quality images and artifacts necessitate that algorithms use sufficient background knowledge of anatomy and spatial relations of bones in order to work satisfactorily. We present a robust knowledge based methodology for detecting reference points of the main skeletal regions that simultaneously works on both anterior and posterior whole-body bone scintigrams. Expert knowledge is represented as a set of parameterized rules which are used to support standard image processing algorithms. Our study includes 467 consecutive, non-selected scintigrams, which is the biggest number of images ever used in such studies to our knowledge. Automatic analysis of whole-body bone scans using our knowledge based segmentation algorithm gives more accurate and reliable results than previous studies. Obtained reference points are used for automatic segmentation of the skeleton, which is used for automatic (machine learning) or manual (expert physicians) diagnostics. Preliminary experiments show that an expert system based on machine learning closely mimics the results of expert physicians.

Keywords: whole-body bone scintigraphy, reference point detection, automatic segmentation, image processing, machine learning

1 Introduction

Whole-body scan or bone scintigraphy is a well known clinical routine investigation and one of the most frequent diagnostic procedures in nuclear medicine. Indications for bone scintigraphy include benign and malignant diseases, infections, degenerative changes ... [2]). Bone scintigraphy has high sensitivity and the changes of the bone metabolism are seen earlier than changes in bone structure detected on radiograms [15].

The investigator's role is to evaluate the image, which is of poor resolution due to the physical limitations of gamma camera. There are approximately 158 bones visible on both anterior and posterior whole-body scans [10]. Poor quality and the number of bones to inspect makes it difficult and often tedious work. Some research on automating the process of counting the bone lesions has been done, but only few studies attempted to automatically segment individual bones prior to the computerized evaluation of bone scans [6; 7; 1].

1.1 Related work

First attempts to automate scintigraphy in diagnostics for thyroid structure and function were made back in 1973 [11]. Most of the research on automatic localization of bones has been done at the former Institute of medical information science at the University of Hildesheim in Germany from 1994 to 1996. The main contribution was made by the authors Berning [7] and Bernauer [6] who developed semantic representation of the skeleton and evaluation of the images. Benneke [1] has realized their ideas in 1996.

Yin and Chiu [13] tryied to find lesions using a fuzzy system. Their preprocessing of scintigrams includes rough segmentation of six parts with fixed ratios of the whole skeleton. Those parts are rigid and not specific enough to localize a specific bone. Their approach for locating abnormalities in bone scintigraphy is limited to point-like lesions with high uptake.

When dealing with lesion detection other authors like Noguchi [10] have been using merely intensity thresholding and manual lesion counting or manual bone ROI (region of interest) labeling. Those procedures are only sufficient for more obvious pathologies whereas new emerging pathological regions are overlooked.

2 Aim and our approach

The aim of our study was to develop a robust method for segmenting whole-body bone scans to allow further development of automatic algorithms for bone scan diagnostics of individual bones.

We have developed the algorithm for detecting extreme edges of images (peaks). Here, respective skeletal regions are processed in the following order: shoulders, head, pelvis, thorax and extremities. The experience with automatic processing is presented. Several image processing algorithms are used such as binarization, skeletonization, Hough's transform, Gaussian filtering [4], least square method and ellipse fitting in combination with background knowledge of anatomy and scintigraphy specialities.

In everyday practice, when a bone is identified, it is diagnosed by the expert physician according to several possible pathologies (lesions, malignom, metastasis, degenerative changes, inflammation, other pathologies, no pathologies). This process can be supported by using some machine learning classifier [9] which produces independent diagnoses. As an input it is given a suitably parameterized bone image, obtained from detected reference points. As an output it assigns the bone to one of the above pathologies. It can therefore be used as a tool to give physician an additional insight in the problem.

3 Materials and methods

3.1 Patients and images

Retrospective review of 467 consecutive, non-selected scintigraphic images from 461 different patients who visited University Medical Centre in Ljubljana from

October 2003 to June 2004 was performed. Images were not preselected, so the study included standard distribution of patients coming to examination in 9 months. 19% of the images were diagnosed as normal, which means no pathology was detected on the image. 57% of the images were diagnosed with slight pathology, 20% with strong pathology and 2% were classified as super-scans.

Images also contained some artifacts and non-osseous uptake such as urine contamination and medical accessories (i.e. urinary catheters) [5]. Segmentation was also complicated by the radiopharmaceutical site of injection. Partial scans (missing a part of the head or upper/lower extremities in the picture) were the case in 18% of the images. There were also adolescents with growth zones (5% of the images), manifested as increased osteoblastic activity in well delineated areas with very high tracer uptake.

3.2 Bone scintigraphy

All patients were scanned with gamma camera model Siemens MultiSPECT with two heads with LEHR (Low Energy High resolution) collimators. Scan speed was 8cm per minute with no pixel zooming. 99m-Tc-DPD (Techneos^R) was used. Bone scintigraphy was obtained about 3h after intravenous injection of 750 MBq of radiopharmaceutical agent. The whole body field was used to record anterior and posterior views digitally with resolution of 1024 x 256 pixels. Images represent the counts of detected gamma rays in each spatial unit with 16-bit grayscale depth.

3.3 Detection of reference points

Bone scans are very different (Figure 3) one from another even though the structure and position of bones is more or less the same. In practice many scans are only partial because only a determined part of the body is observed or due to the scanning time limitations. In our study we have observed that only on two images out of 467 the shoulders were not visible. Many other characteristic parts could have been missing in images more often (i.e. head, arms, one or both legs). We have chosen shoulders as the main reference points to start with, which means they are supposed to be visible in the images. Second and the last assumption is the upward orientation of the image. This assumption is not limiting since all scintigraphies are made with same orientation.

In order to make the detection of reference points faster and more reliable we have tried to automatically detect intuitive peaks which would represent edges and would cover roughly also the reference points. With normal Canny edge filter too many peaks were obtained. Our approach is based on orthogonal two-way Gaussian filtering [16].

Low image intensities (count level) acquired in typical studies are due to the limited level of radioactive dosage required to ensure patient's safety. They make bone scans look distorted. Bone edges are more expressive after we filter images with some averaging algorithm (i.e. wavelet based, median filter, Gaussian filter) [4]. We have used Gaussian filter so that the detection of peaks was more reliable.

Both images, anterior and posterior, are simultaneously processed in the same detection order and in each step the detected reference points visible on both images are compared and corrected adequately.

Detected points from the anterior image are mirrored to the posterior and vice versa. Some bones are better visible on anterior and some on posterior images due to the varying distances from both collimators. This improves the calculation of circles, lines and ellipses with least square method (LSM).

The order in which the reference points were detected was determined by using knowledge of human anatomy as well as physicians' recommendations. They are represented as a list of parameterized rules. Rule parameters (e.g. thresholds, spatial and intensity ratios, ...) were initially set by physicians and further refined on a separate tuning set. More details can be found in [16].

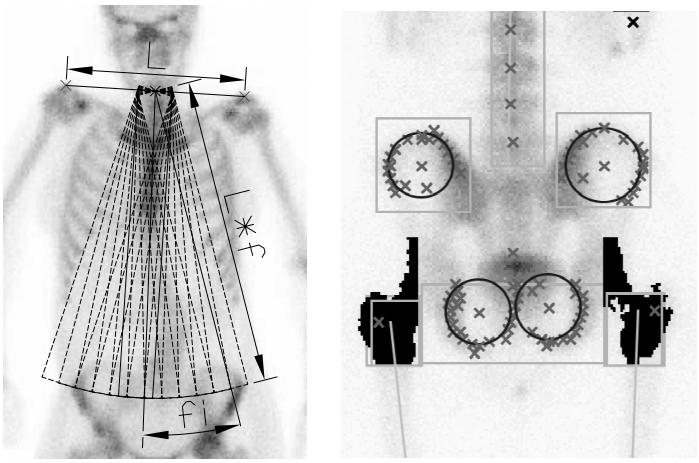
Shoulders. They are the only part of the body that is assumed to be present in every image in the upper part on both sides. The algorithm just searches for the highest detected peak on both sides of the image. Next step is to locally shift the candidate points with local maximum intensity tracing to the outermost location. Only in 5 images out of 467 the shoulders were not found correctly due to the tilted head position.

Pelvic region (ilium bone, pubis bone, great trochanter of femur). The most identifiable bone in pelvic region is ilium bone which has higher uptake values than it's neighboring soft tissue. Ilium bone has circular shape in the upper part and therefore it is convenient for circle detection with LSM method. This bone is well described with already detected peaks by as shown in Figure 1(b). Ilium position is roughly estimated with regions of interest (ROIs) which are found on the basis of skeleton's anticipated ratios and reference points found up to this step of detection.

The pelvis is located at the end of the spine and has approximately the same width as shoulders. In order to find the pelvis, the calculation of the spine position is required. This is done with a beam search (Figure 1(a)). The anticipated spine length is determined from the distance between shoulders. Beam starting point is the middle point of the shoulders and it's orientation is perpendicular to the shoulder line. The angle at which the beam covers most peaks, is a rough estimation of spine direction since there is most of the uptake in the vertebrae and hence peaks are dense in that region.

Pubis bone is detected by estimating the pubis ROI using detected ilium location, distance between detected ilium circles and their inclination. The experimentally determined ROI's size is narrowed and additional vertical peaks are added and circles detected as shown in Figure 1(b).

Head and neck. When at least image orientation and the location of the shoulders are known, some part of the neck or even head is visible since they are between the shoulders. Finding the head is not difficult but its orientation is, especially in cases where a part of the head in scan is not visible. The most reliable method for determining head orientation and position is ellipse fitting of the head contour determined by thresholding. Neck is found by local vertical



(a) Beam search sketch (b) Detection of bones in the pelvic region

Fig. 1. Beam search and detection in pelvic region

shifting of a stripe determined by the ellipse's semiminor axis (position and orientation).

Thoracic part (vertebrae, ribs). Vertebrae have more or less constant spatial relations, the only problem is that on a bone scintigraphy only a planar projection of the spine is visible. Since the spine is longitudinally curved, the spatial relations vary due to different longitudinal orientation of the patients. Average vertebrae relations have been experimentally determined from normal skeletons.

Ribs are the most difficult skeleton region to detect since they are quite unexpressive on bone scans, their formation can vary considerably and their contours [14] can be disconnected in the case of stronger pathology (Figure 2).

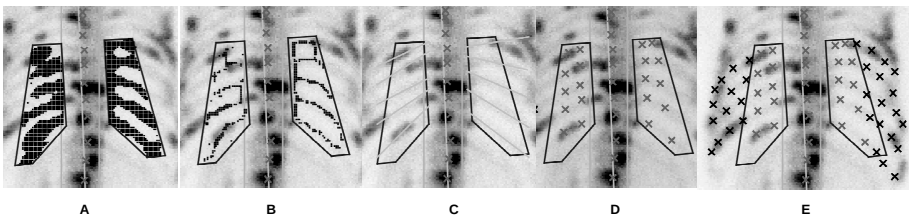


Fig. 2. Rib detection steps example on a skeleton with strong pathology. Rib ROI is binarized (A), binarized image is skeletonized (B), Hough transform of linear equation is calculated on skeleton points (C), reference points are estimated using results of the Hough transform (D), rib contours are individually followed by the contour following algorithm (E).

Lower and upper extremities (femur, knee, tibia, fibula, humerus, elbow, radius, ulna). They are often partly absent from whole-body scan because of limited gamma camera detector width. In our patients, a maximum of 61cm width is usually not enough for the entire skeleton. The regions of humerus, ulna and radius as well as femur, tibia and fibula bone are located with the use of controlled beam search. The beam lengths can be estimated from skeletal relationships (i.e. femur length is estimated as 78% of the distance between the neck and ilium bone center). The detection is designed so that a part or all of the extremities and/or the head may not be visible.

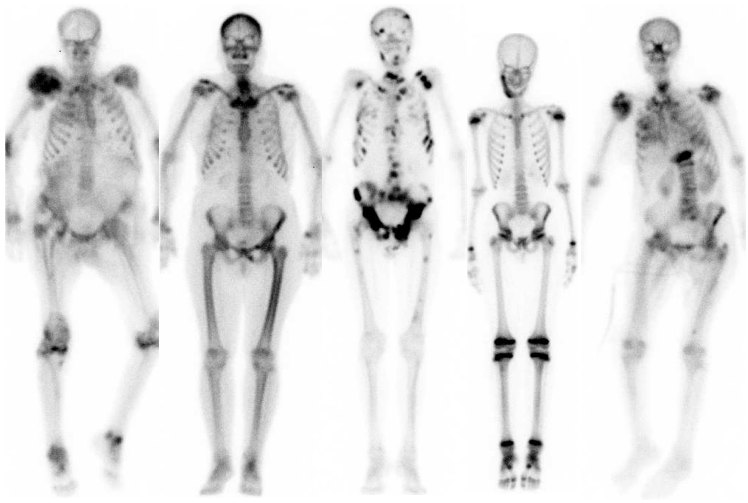


Fig. 3. Examples of body scan variety

3.4 Diagnosing pathologies with machine learning

When all reference points are obtained, every bone is assigned a portion of original scintigraphic image, according to relevant reference points. Obtained image is parametrized by using the ArTeX algorithm [8]. It uses association rules to describe images in rotation-invariant manner. Rotation invariance is very important in our case, since it accounts for different patients' positions inside the camera.

Bones were described with several hundreds of automatically generated attributes. They were used for training the SVM [12] learning algorithm. In our preliminary experiments pathologies were not discriminated, i.e. bones were labelled with only two possible diagnoses (no pathology, pathology). In 19% of patients no pathology or other artifacts were detected by expert physicians. In the remaining 81% of the patients at least one pathology or artifact was observed.

4 Results

4.1 Segmentation

Approximately half of the available images were used for tuning rule parameters to optimize the recognition of the reference points and another half to test it. All 246 patients examined from October 2003 to March 2004 were used as the tuning set and 221 patients examined from April 2004 to June 2004 were used as the test set. In the tuning set there were various non-osseous uptakes in 38.9% of the images, 47.5% images with the visible injection point and 6.8% images of adolescents with the visible growth zones. Similar distribution was found in the test set (34.5% non-osseous uptakes, 41.0% visible injection points and 2.85% adolescents). Most of the artifacts were minor radioactivity points from urine contamination in genital region or other parts (81.4% of all artifacts) whereas only few other types were observed (urinary catheters 13%, artificial hips 4% and lead accessories 1.6%). We have observed that there were no ill-detected reference points in adolescents with the visible growth zones since all the bones are homogenous, have good visibility and are clearly divided with growth zones. Results of detecting reference points on the test set are shown in the Table 1.

Table 1. False reference point detection on test set. Both frequencies and percentages are given.

Bone	*no pathology*		*slight pathology*		*strong pathology*		*super-scan*		all	
	46		*133*		*39*		*3*		*221*	
ilium	0		2	0.9%	6	2.7%	1	0.5%	9	4.1%
pubis	2	0.9%	3	1.4%	2	0.9%	0		7	3.2%
trochanter	0		1	0.5%	0		0		1	0.5%
shoulder	0		0		1	0.5%	0		1	0.5%
extremities	5	2.3%	11	5.0%	0		0		16	7.2%
spine	0		2	0.9%	1	0.5%	0		3	1.4%
ribs	11	5.0%	17	7.7%	3	1.4%	0		31	14.0%
neck	2	0.9%	4	1.8%	0		0		6	2.7%

4.2 Machine learning results

From our complete set of 467 patients, pathologies were thoroughly evaluated by physicians only for 268 patients. These 268 patients were used for evaluation of machine learning approach by using ten-fold cross validation. Results are shown in Table 2. The bones were grouped in ten relevant groups. They are quite impressive, given high numbers of different bones (158 visible for an individual adult patient). Classification accuracy was obtained for a two-class problem.

Table 2. Experimental results with machine learning on two-class problem.

Bone group	Classification accuracy
Cervical spine	75.94
Feet	83.82
Skull back	94.74
Illium bone	87.31
Lumbal spine	71.43
Femur and tibia	88.89
Pelvic region	92.16
Ribs	98.05
Scapula	91.42
Thoracic spine	81.95
Total	89.93

5 Discussion

The testing showed encouraging results since the detection of proposed reference points gave excellent results for all bone regions but the extremities, which was expected.

We have payed special attention to the images with partial skeletons since it is often the case in clinical routine (in our study 18% of the images were partial and no particular problem appeared in detecting) and a robust segmentation algorithm should not fail on such images. The detection of ribs showed to be the most difficult, yet that was expected. Results show that in 14% to 20% of images there were difficulties in detecting the ribs. This usually means one rib is missed or not followed to the very end which we intend to improve in the future. In the present system such reference points can be manually repositioned by the expert physicians.

The automatically detected reference points can be used for mapping a standard skeletal reference mask, which is to our belief the best way to find individual bones on scintigrams since bone regions are often not expressive enough to follow their contour. An example of such mask mapping is shown in Figure 4.

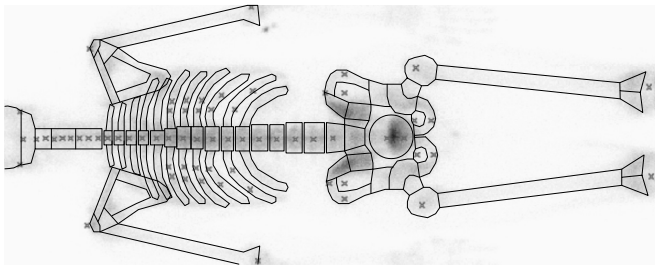


Fig. 4. Example of mapped standard skeletal mask with the detected reference points

While our experimental results with machine learning are quite good, one must bear in mind that they were obtained for a simplified (two class) problem. Simply extending a problem to a multi-class paradigm is not acceptable in our case, as the bone may be assigned several different pathologies at the same time. A proper approach, the one we are currently working on, is to rephrase a problem to the multi-label learning problem, where each bone will be labelled with a nonempty subset of all possible labels [17; 3].

6 Conclusion

The presented computer-aided system for bone scintigraphy is a step forward in automating routine medical procedures. Some standard image processing algorithms were tailored and used in combination to achieve the best reference point detection accuracy on scintigraphic images which have very low resolution. Poor quality, artifacts and pathologies necessitate that algorithms use as much background knowledge on anatomy and spatial relations of bones as possible in order to work satisfactorily. This combination gives quite good results and we expect that further studies on automatic scintigraphy diagnosing using reference points for image segmentation will give more accurate and reliable results than previous studies, negligent to the segmentation.

This approach opens a new view on automatic scintigraphy evaluation, since in addition to detection of point-like high-uptake lesions there are also:

- more accurate and reliable evaluation of bone symmetry when looking for skeletal abnormalities. Many abnormalities can be spotted only when the symmetry is observed (differences in length, girth, curvature etc.),
- detection of lesions with low-uptake or lower activity due to metallic implants,
- possibility of comparing uptake ratios among different bones,
- more complex pathology detection with combining pathologies of more bones (i.e. arthritis in joints)
- possibility of automatic reporting of bone pathologies in written language.

Machine learning approach in this problem is in a very early stage, so its usefulness in practice cannot yet be objectively evaluated. However, preliminary results are encouraging and switching to the multilabel learning framework may make them even better.

Acknowledgement

This work was supported by the Slovenian Ministry of Education, Science and Sport through the research programme P2-0209. Special thanks to nuclear medicine specialist Jure Fettich at the University Medical Centre in Ljubljana for his help and support.

Bibliography

[1] Benneke A. Konzeption und realisierung eines semi-automatischen befundungssystems in java und anbindung an ein formalisiertes begriffssystem am beispiel der skelett-szintigraphie. Diplom arbeit, Institut für Medizinische Informatik, Universität Hildesheim, mentor Prof. Dr. D.P. Pretschner, 1997.

[2] Hendler A. and Hershkop M. When to use bone scintigraphy. it can reveal things other studies cannot. *Postgraduate Medicine*, 104(5):54–66, 11 1998.

[3] McCallum A. Multi-label text classification with a mixture model trained by em. In *Proc. AAAI'99 Workshop on Text Learning*, 1999.

[4] Jammal G. and Bijaoui A. Dequant: a flexible multiresolution restoration framework. *Signal Processing*, 84(7):1049–1069, 7 2004.

[5] Weiner M. G., Jenicke L., Mller V., and Bohuslavizki H. K. Artifacts and non-osseous uptake in bone scintigraphy. imaging reports of 20 cases. *Radiol Oncol*, 35(3):185–91, 2001.

[6] Bernauer J. Zur semantischen rekonstruktion medizinischer begriffssysteme. Habilitationsschrift, Institut für Medizinische Informatik, Univ. Hildesheim, 1995.

[7] Berning K.-C. *Zur automatischen Befundung und Interpretation von Ganzkörper-Skelettszintigrammen.* PhD thesis, Institut für Medizinische Informatik, Universität Hildesheim, 1996.

[8] Bevk M. and Kononenko I. Towards symbolic mining of images with association rules: preliminary results on textures. In Brito P. and Noirhomme-Fraiture M., editors, *ECML/PKDD 2004: proc. of the workshop W2 on symbolic and spatial data analysis: mining complex data structures*, pages 43–53, 2004.

[9] Kukar M., Kononenko I., Grošelj C., Kralj K., and Fettich J. Analysing and improving the diagnosis of ischaemic heart disease with machine learning. *Artificial Intelligence in Medicine*, 16:25–50, 1999.

[10] Noguchi M., Kikuchi H., Ishibashi M., and Noda S. Percentage of the positive area of bone metastasis is an independent predictor of disease death in advanced prostate cancer. *British Journal of Cancer*, (88):195–201, 2003.

[11] Maisey M.N., Natarajan T.K., Hurley P.J., and Wagner H.N. Jr. of a rapid computerized method of measuring 99mtc pertechnetate uptake for routine assessment of thyroid structure and function. *J Clin Endocrinol Metab*, 36:317–322, 1973.

[12] Cristianini N. and Shawe-Taylor J. *An introduction to support vector machines (and other kernel-based learning methods).* Cambridge University Press, 2000.

[13] Yin T.K. and Chiu N.T. A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach. *IEEE Transactions on Medical Imaging*, 23(5):639–654, 5 2004.

[14] Kindratenko V. *Development and Application of Image Analysis Techniques for Identification and Classification of Microscopic Particles.* PhD thesis, Universitaire Instelling Antwerpen, Departement Scheikunde, 1997.

[15] Müller V., Steinhagen J., de Wit M., and Bohuslavizki H. K. Bone scintigraphy in clinical routine. *Radiol Oncol*, 35(1):21–30, 2001.

[16] Šajn L., Kononenko I., Fettich J., and Milčinski M. Automatic segmentation of whole-body bone scintigrams. *Nuclear Medicine Comm.*, 2005. To appear.

[17] Shen X., Boutell M., Luo J., and Brown C. Multi-label machine learning and its application to semantic scene classification. In *Proceedings of the 2004 International Symposium on Electronic Imaging (EI 2004)*, San Jose, California, 2004.