DETECTION OF POSE CHANGES FOR SPATIAL OBJECTS FROM PROJECTIVE IMAGES

Boris Peter Selby [a], Georgios Sakas [b], Stefan Walter [a], W. -D. Groch [c], Uwe Stilla [d]

[a] MedCom GmbH, Medical Imaging, Rundeturmstr. 12, 64283 Darmstadt, Germany
[b] Cognitive Computing and Medical Imaging, Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany
[c] Fachbereich Informatik, University of Applied Sciences, Haardtring 100, 64295 Darmstadt, Germany
[d] Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, Arcisstr. 21, 80333 Muenchen, Germany

KEY WORDS: Pose estimation, Change detection, Projection, Image registration, Mutual Information, Landmark segmentation

ABSTRACT:

Numerous fields of application effort the detection of pose changes for 3 dimensional objects in six degrees of freedom (6 DoF). Automatic procedures that exploit 2D images for the detection of pose changes can be used for example for tracking object movements, for quality control or for the verification of the alignment of patients in radiation treatment devices. In this contribution we present two different solutions for the detection of pose changes that base on the comparison of two 2D images resulting from the projection of an object in the new pose and a 3D volume of the same object in a known reference alignment. Whereas for the first solution we use an object where we can clearly extract landmarks useable as reference positions for the determination of the object's alignment, we provide a second solution for objects where these landmarks cannot be extracted, which is involved automatically if necessary. In this case grey value based pose estimation is conducted by registering the computationally projected reference 3D volume to the 2D images. As reference data for the object with known alignment, CT slices will be used, as they are provided for the alignment of patients in radiation treatment devices. Two X-ray images of the same object in an unknown pose can then be compared to the reference data to determine the respective pose change, which may consist of 3 rotations and 3 translations. Using both approaches to determine patient misalignments in treatment devices shows, that both methods result in highly accurate pose detections and that the second method, despite being less accurate and more time consuming, is an appropriate solution in cases where landmark detection fails.

1. INTRODUCTION

1.1 Motivation

Modern particle beam based radiation treatment techniques for tumours allow accurate application of the treatment dose onto carcinogen tissue with an accuracy much better than 1.0 mm and therefore require an accurate alignment of the patient in the treatment facility (Verhey et al., 1982). Common strategies like tracking of external markers or fixation of the patient's body do not suffice the requirement of high set-up precision and are not feasible whenever internal tumours are to be irradiated, because of possible movements of the treatment target relative to the outer body shape.

Today it is common practice in image guided radiotherapy to align patients manually in the treatment device according to visual evaluation of reference images as X-rays and CT volumes that allow an estimation of the patient's misalignment (Thilmann et al., 2005).

During this time consuming procedure, the alignment of the respective body region may change, which leads to unknown set-up errors and degrades the results of the treatment. Besides that, a manual alignment correction can hardly be done for 6 DoF, because rotational misalignments can hardly be detected visually in the imaged objects.

To overcome these problems an approach for the automatic determination of alignment errors is used, which is based on the comparison of the position of internal landmarks whenever fiducial markers are available.

The use of fiducial markers can be advantageous in many cases, because a marker based procedure allows determining a pose correction without being influenced by surrounding tissue, for example for pose correction of the eyeball.

However, fiducial markers are not always present or detectable. Because of the invasive character of the maker application, marker attachment is not possible for many anatomical regions. In these cases, the respective alignment of the target region has to be estimated by comparison of either natural landmarks or other image properties. Because detection of natural landmarks, especially for soft shaped objects, is not very reliable, we use an approach that estimates the pose by comparison of the grey value distribution in the projected images, which are in the case of radiation treatment X-ray images and the reference data volume, here a CT dataset of the involved body region. To be able to achieve best results for all cases, the second approach is involved automatically if the landmark-based procedure is not able to deliver acceptable results.

1.2 Aims

Our aim is to provide fast and stable algorithms that allow detecting changes of the spatial alignment of an object in respect of a reference position, using 2D projections of the object. These known changes can be used to realign the object, to achieve a correct placement. We do not intend to restrict the

procedure to any known object geometry, but limit it to rigid transformations, because of the inability of correcting object deformations.

Two different solutions are provided, because usage of landmarks comes with several advantages, but is not applicable in all cases. If the procedure using landmarks does not lead to proper results, the alternative approach shall be used to accomplish the pose estimation. The decision which of the two procedures is to be used shall be automated.

The radiation therapy field of application is used for the implementation and test of the procedures, because high accuracy and reliability are of special importance in this scope. Using high-resolution images we aim to achieve accuracies better than 1.0 mm and 0.5° for the detection of pose changes. Besides that, volumetric data as well as projective images of respective objects are available in this scope and specific fixation devices, as for example stereotactic frames, can be used to validate the results in a controlled environment.

1.3 Overview

This contribution consists of two main parts. Part one introduces methods for pose estimation that base on fiducial markers in two projections of an object and a 3D volume dataset. The markers used in this work are tantalum clips attached to an eyeball. We use 4 or 5 clips, which are resided in a CT dataset of the said eyeball and 2 X-ray images of the eye in different alignment. To be able to conduct an automatic detection of the pose change, the landmark positions are extracted from the 3D and the 2D image data. Then an inverse projection of the 2D positions is performed to be able to compute the transformation between the resulting points and the reference markers from the volumetric dataset. This is done by a rigid registration of the point-sets that has to be tolerant against single inaccuracies from the landmark detection.

If the first approach fails or leads to inconsistent results, the second approach is conducted automatically.

The second approach provides a solution for cases where no proper landmarks can be found. Then the volume dataset is projected into the imager planes of the X-ray beamlines for different virtual object poses. Optimisation of grey value distribution based image comparators gives the alignment of the reference dataset. The modification of the reference volume alignment results in the change of the object's pose relative to the original pose.

1.4 References to related work

A solution for X-ray based pose estimation is given in (Bhunre et al. 2007). In this approach, a single 3D model is used to estimate the pose of a bone, visible in X-ray images. As the optimal shape for the model can vary from patient to patient, a single model does not suffice the demand for a highly flexible solution. Besides that, the model-based solution requires a segmentation of the respective object in the X-ray image. As X-ray images can be noisy and the same anatomical objects may appear different, depending on the X-ray beam's energy, it is not possible to rely on a segmentation, especially if, as in our case, the goal is to detect pose changes for any rigid object, without restricting the detection to bones.

The approach described in (Frahm et al. 2004) uses a Harris Corner Detector to extract features from several images obtained from different camera positions. Through comparison of these features, the pose of an object can be estimated. This approach is not applicable in the case of X-ray images, where detected corners may be placed anywhere inside the object, because of the nature of the X-ray imaging, not mapping the surface onto the image detector, but the integral of absorptions of the ray on its way through the object. For X-ray images it is not possible to determine whether a visible feature is located on the surface, within or at the back of an object.

In (Tang et al. 2000) a feature based method is proposed, basing on the comparison of features in one single X-ray image to known feature positions in 3D space. Accuracies of about 1 mm and 2° could be reached. Using one single X-ray image implies, that a point to line registration has to take place. The accuracy of the estimation for the distance of the 3D object from the X-ray source depends on the beam width of the X-ray beam pyramid. Because shifts in direction of the beam can only be determined by size changes of the projected 2D image, the accuracy of the results is restricted on how accurate the image-scaling factor could be determined. In a typical radiation machine device, where the source – detector distance is up to 3 m and more, much higher accuracies can be reached, using two X-ray images.

2. METHODS

2.1 Landmark based pose estimation

The first approach is based on the comparison of artificial landmarks.

2.1.1 Identification of landmarks

Detection of landmarks in 3D data
To enable very fast detection, clips are segmented in the CT data using different levels of volume resolution. The CT is resampled to a 4D pyramid containing several instances of the volume, each with a different resolution. The search for a clip starts at a low-resolution level. As soon as a potential clip voxel is found the search continues at a higher resolution. To determine if a voxel belongs to a clip, two Hounsfield thresholds are used. A voxel between those thresholds is considered to belong to a clip. As soon as a clip is found, the thresholds are adapted, assuming that upcoming clips are of the same material and will be represented by similar Hounsfield values. If no clip can be found, the thresholds are modified and the search continues (Fig. 1).

Figure 1. Landmark detected in consecutive slices

Detection of landmarks in 2D projections
For detecting the landmarks in the X-ray images, two different approaches are combined. Using a Harris Corner Detector (Harris et al., 1988) potential clip corners are identified. Then the convex hulls for the point-sets are determined, as the shape of the eye clips will be convex after projection onto the X-ray panel plane. The area A_{poly} of a convex polygon is calculated by equation 1:

$$A_{poly} = \sum_{i=1}^{N-2} A_{\Delta}\left(P_1, P_{i+1}, P_{i+2}\right) \quad (1)$$

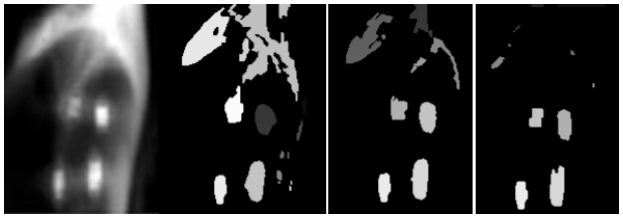
where N = number of polygon corners
P_i = ordered corners of the polygon
A = area of the triangle

In a next step, segmentation is performed inside the area of the convex polygons. A grey value threshold T is used to identify potential clip pixels. The area A of the identified pixels is determined by equation (2):

$$A = sy \sum_{i=y0}^{y1} \left(sx \sum_{j=x0}^{x1} t(i,j) \right)$$
$$t(i,j) = \begin{cases} 1 & for \quad I(i,j) \geq T \\ 0 & else \end{cases} \quad (2)$$

where $y0, y1, x0, x1$ = bounding box of the polygon
sx, sy = pixel size on image plane
$I(x, y)$ = pixel intensity

If the quotient of A and A_{poly} becomes larger than a certain threshold, the polygon is considered to belong to a fiducial marker. If the total number of resulting marker objects does not correspond to the expected number of markers, the detection is resumed, using a modified threshold T (Fig. 2).



(a) (b) (c) (d)

Figure 2. X-ray with 4 landmarks (a); Iterative refinement of areas identified as landmarks (b, c, d)

2.1.2 Back-projection

Inverse projection
To compare the markers detected in the projective images with the reference markers from the spatial dataset, an inverse projection is performed for corresponding pairs of both images. Based on a known geometric set-up of the imaging devices (Fig. 3), the inverse projection is done by calculating the intersection points of rays from the X-ray source to the centre of the segmented marker in the respective projection plane.

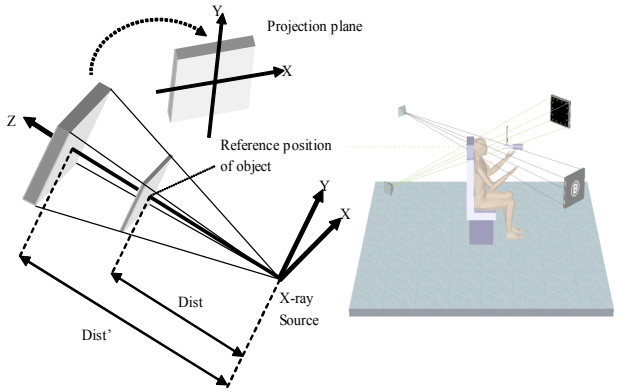


Figure 3. Imaging device (left); geometric set-up of imaging devices (right)

Handling of redundant results
Because it is not always clear which clips correspond to each other, it is possible that one clip becomes a member in several back-projection results. This is the case, if several clips are projected onto a horizontal line in the plane of one flat panel (Fig. 4).

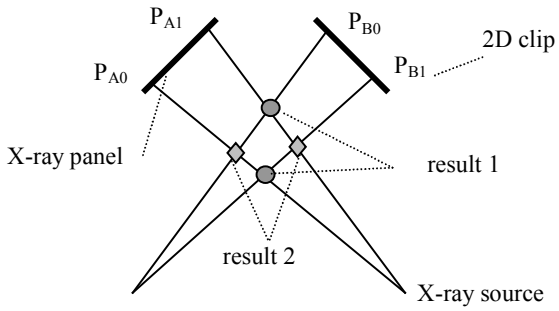


Figure 4. Back-projection with two possible results

All possible resulting point-sets B_i are kept as preliminary results. In the following registration procedure only the point-set, which can be mapped onto the marker positions A derived from the volume data is used to derive the final transformation for pose detection.

2.1.3 Point-set registration

Registration
Several registrations are performed, one for each possible back-projection result with the reference positions. In each registration, two sets of points in the 3D space are used to calculate 3 shifts and 3 rotations, which map one point-set onto the other as good as possible. The remaining mapping error can serve as an indicator for the quality of the calculated alignment deviation. We use a Downhill Simplex optimisation (Press et al., 1992) to minimize the error for the misalignment between back-projected point-set B_i and the original landmark positions A. The error metric bases on the undirected Hausdorff Distance $H(A,B_i)$ (Huttenlocher et al. 1993).

The optimisation may suffer from local minima of the error function (Fig. 5).

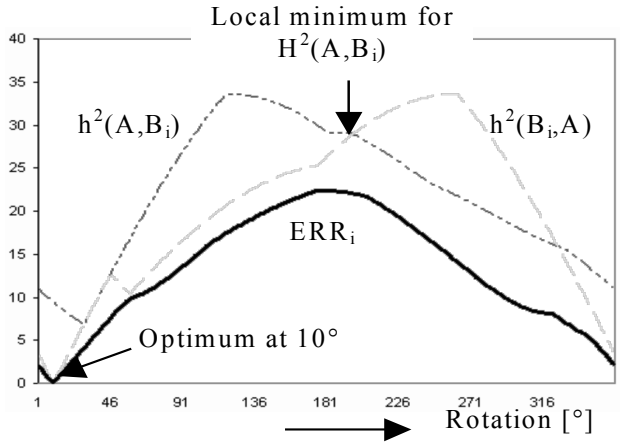


Figure 5. Comparison of error functions

The correct alignment detection for figure 5 is at 10°. The squares of the directed Hausdorff Distances $h(A,B_i)$ and $h(B_i,A)$, as well as the square of the undirected Hausdorff Distance $H(A,B_i)$, which is the maximum of the directed distances, have local minima. To avoid these, we calculate an error ERR_i for each combination of A with B_i as the square sum of a fractional undirected Hausdorff Distance of rank 1 to K, where K denotes the minimum of the number of landmarks detected in either the reference data or the projected images, as shown in equation 3:

$$
\begin{aligned}
ERR_i &= \sum_{k=1}^{K} \mathbf{max}\left(h_k(A,B_i), h_k(B_i,A)\right)^2 \\
h_k(A,B_i) &= kth_{a\in A}\ \mathbf{min}_{b\in Bi}\left(\|a-b\|\right) \\
h_k(B_i,A) &= kth_{b\in Bi}\ \mathbf{min}_{a\in A}\left(\|a-b\|\right) \\
K &= \mathbf{min}\left(N_A, N_{Bi}\right)
\end{aligned}
\qquad (3)
$$

where N = number of points
h_k = directed Hausdorff Distance of rank k

After minimization of the ERR_i values only the transformation for

$$ERR_{best} = \mathbf{min}\left(ERR_1,...,ERR_n\right) \qquad (4)$$

is kept for further calculations.

Acceptance of results
The result of the alignment detection is accepted, if the standard deviation $_{AB}$ between the closest members of the final point-sets A and B_{best} used for the optimisation lies beneath a threshold given in the program configuration as the maximal accepted error. If the result is not accepted, grey value based pose estimation is conducted.

2.2 Grey value based pose estimation

The grey value based pose estimation is conducted if the landmark based approach fails because of the lack of detectable landmarks, which lead to a good match with the 3D reference data.

2.2.1 The algorithm

The procedure for the grey value based pose estimation projects the reference volume A computationally onto the detector planes of the imaging devices, using a pose correction C. The results are two projections of the volume that depend on a current pose correction. For each we calculate a value Q for the quality of the match with the respective X-ray image B, acquired with the object in the current, unknown pose. The quality values are combined. Minimizing the negative combined matching quality by modification of the pose correction C gives the final pose change (Fig. 6).

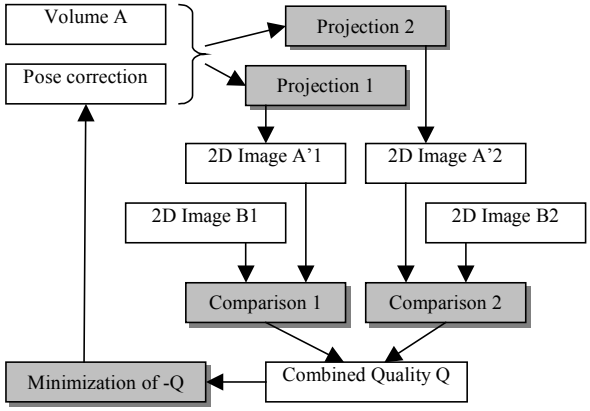


Figure 6. Algorithm for grey value based six degrees of freedom registration

2.2.2 Projection of the volume

The volume is projected onto an imager plane by a ray tracing technique, where rays to the X-ray source are computed for a sub-set of pixels, covering ¼ of the receptor area (Fig. 7).

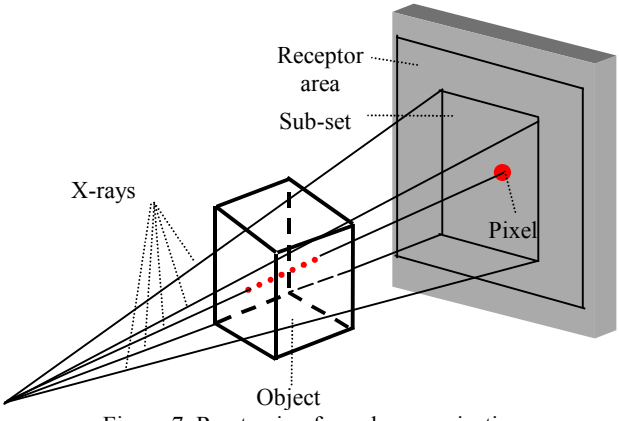


Figure 7. Ray tracing for volume projection

To speed-up the process, we stop ray tracing as soon as summing up the density values of the volume leads to a saturation of the respective pixel intensity. The pixel value is calculated by summation of the voxels, intersected by the ray. The voxels are weighted by their absorption coefficient, which can directly be obtained from the CT data. To improve the quality of the resulting image, trilinear interpolation can be applied as the ray passes through the volume. However, we used

nearest neighbourhood interpolation to improve the performance and because we did not intend to produce visual results of high quality.

2.2.3 Image comparison and optimisation

There exists a wide range of grey value based image comparators in the scope of registration. As methods like cross-correlation or usage of difference images are not applicable for images that differ in much more aspects than contrast and intensity, we decided to use mutual information as image correlation measure (PLUIM et al., 2003).

Figure 8 shows five different joint histograms, where each axis stands for the grey values of one of the images.

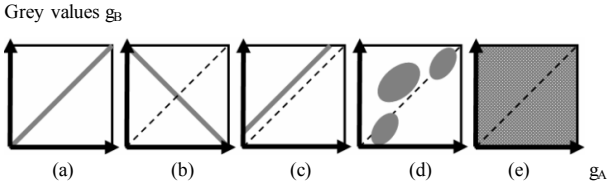


Figure 8. Joined histograms for images A and B: a) identical images; b) inverse images; c) A darker than B; d) partly different images; e) images without correspondences

A joint histogram is built up by reading the grey values of both images at the position of two overlaid pixels and incrementation of the histogram cell by one at the respective coordinates, defined by the two grey values. The mutual information value *MI* is then calculated by equation 5:

$$MI(A,B) = H(A) + H(B) - H(A,B)$$
$$H = -\sum_{g=0}^{G} p_g \ln p_g \qquad (5)$$

where G = largest grey value
p = probability for occurrence of the grey value in the image, based on the distribution in the histogram
H = entropy of either one of the image histograms or the joint histogram

Optimisation of the rigid transformation *T* for the current object pose in 6 DoF is done by minimization of a negative quality value -*Q* given in equation 6:

$$-Q = -\sum_{i=1}^{N} MI_i^{\,2} \qquad (6)$$

where MI_i = Mutual Information value for the image pair *i*
N = Number of images (here 2)

We minimize -*Q* by the downhill simplex method for 6 pose parameters (3 shifts and 3 rotations).

2.3 Results

2.3.1 Landmark based pose estimation

For all tests a standard PC has been used. Tests have been performed using CT datasets (0.2 mm slice distance, 250 slices) and X-rays of a pig's eye attached with 4 and 5 tantalum clips of 2.5 mm in diameter (Fig. 9).

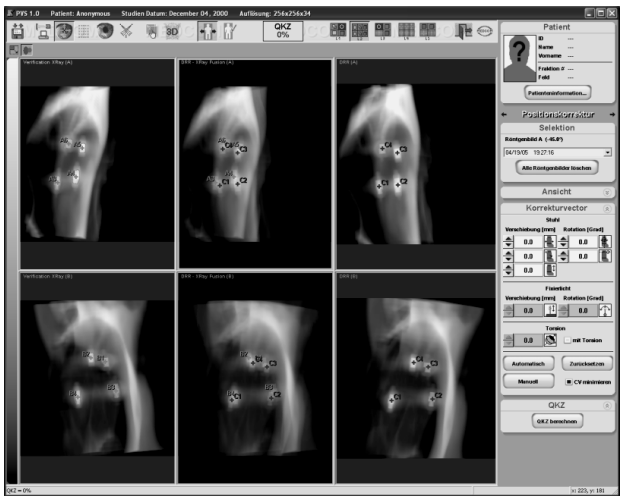


Figure 9. Images of pig's eye with 4 clips: X-rays (left); Fused (centre); DRRs (right)

The X-ray equipment has been calibrated with geometric accuracy of about 0.25 mm. In all cases it was possible to detect all landmarks and to perform a correct mapping of the back-projected marker positions to the reference data (Tab. 1).

Number of Landmarks	Pose change		Calculation error		Time
	shift	rotation	shift	rotation	
4	2.0 mm	2.0°	0.2 mm	0.1°	1sec
4	5.0 mm	10°	0.3 mm	0.1°	1sec
4	10 mm	20°	0.2 mm	0.2°	2sec
5	2.0 mm	2.0°	0.2 mm	0.1°	2sec
5	5.0 mm	10°	0.1 mm	0.1°	3sec
5	10 mm	20°	0.2 mm	0.2°	3sec

Table 1. Pose estimation errors for landmark based approach

2.3.2 Grey value based pose estimation

In this case, tests have been performed using a CT dataset with 295 slices of 0.8 mm slice distance and X-ray images of a human scull. No landmarks have been attached to the scull (Fig. 10).

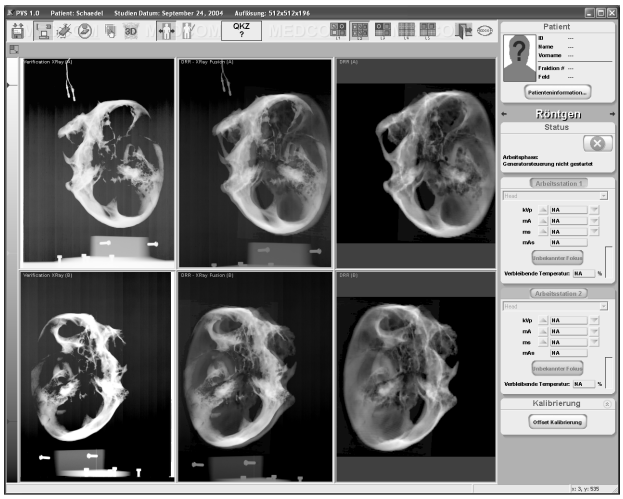


Figure 10. Images of a human scull: X-rays (left); Fused (centre); DRRs (right)

Starting with the landmark based approach the procedure recognized that it was not possible to find a consistent set of landmarks. The automatic procedure for the grey value based pose estimation was started.
For all used X-ray images the grey value based method was able to detect the respective pose change. In table 2 the results are shown for a number of given pose deviations (for better comparability, we chose the same initial poses as for the landmark based procedure).

Pose change		Calculation error		Time
shift	rotation	shift	rotation	
2.0 mm	2.0°	0.9 mm	0.4°	62sec
5.0 mm	10°	1.1 mm	0.7°	98sec
10 mm	20°	2.3 mm	0.9°	113sec

Table 2. Pose estimation errors for grey value based approach

3. DISCUSSION

Both presented methods are able to provide reliable and accurate detection of pose changes. The major disadvantage of the landmark-based method is that it is only applicable if some detectable markers are present.

Regarding the results in tables 1 and 2, the advantages of the landmark based approach become apparent:

- Pose estimation is much more accurate even if we consider that the CT resolution was higher in the case of the landmark based approach;
- The calculation can be done in a few seconds on a standard PC;

Performance of the grey value based approach could be improved, if faster rendering algorithms would be used for the volume projection. However, combination of the two approaches is an ideal solution to exploit the advantages of either method, whenever it is possible.

For the radiotherapy field of application, each method provides an enormous advance in accuracy compared to manual alignment methods.

4. CONCLUSIONS

We presented two different procedures for image based pose estimation. A landmark-based approach has been tested with 2.5 mm tantalum clips, which could be detected relatively easy. Further efforts have to be done to assure a correct detection for the wide variety of applicable metal clips. Under adverse circumstances, as when clips are occluded by other clips or by bony structures of the skull, not all clips can be detected, which results, depending on the total number of clips used, in results less accurate.

If the landmark-based approach failed, grey value based pose detection is started automatically and leads to acceptable results. However, the grey value based approach was less accurate and more time consuming, but whenever no landmarks can be detected, the grey value based approach is an optimal solution to this problem.

REFERENCES

Bhunre, P. K.; Leow, W. K.; Howe, T. S. 2007. Recovery of 3D pose of bones in single 2D X-ray images. IEEE Workshop on Applications of Cumputer Vision 2007. pp. 48.

Frahm, J.-M; Köser, K.; Koch, R. 2004. Pose Estimation for Multi-Camera Systems. Proceedings of Deutsche Arbeitsgemeinschaft für Mustererkennung Vol. 26: pp. 27-35.

Harris, C.; Stephens M. 1988. A combined corner and edge detector. Procs of the 4th Alvey Vision Conference: pp. 147-151.

Huttenlocher, D. P.; Klanderman G. A.; Rucklidge W. J. 1983. Comparing images using the Hausdorff Distance. IEEE transactions on pattern analysis and machine intelligence 15 (9): pp. 850-863.

Pluim, J. P. W.; Maintz, A.; Viergever, M. A. 2003. Mutual registration based registration of medical images: a survey, IEEE Transactions on medical imaging Vol. XX.

Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery B. P. 1992. Numerical Recipes in C 2. Cambridge University Press.

Tang, T. S. Y.; Ellis, R. E.; Fichtinger, G. 2000. Fiducial registration from a single X-ray image: A new technique for fluoroscopic guidance and radiotherapy. Springer Lecture Notes in Computer Science Vol. 1935: pp. 502-511.

Thilmann, C.; Nill, S.; Tücking, T; Höss, A.; Hesse, B.; Dietrich, L.; Bendl R.; Rhein, B.; Häring, P.; Thieke, C.; Oelfke, U.; Debus,J.; Huber, P. 2005. Correction of patient positioning errors based on in-line cone beam CTs: clinical implementation and first experiences. International Journal of Radiation Oncology Biology Physics Vol. 63, Part 1, pp. 550-551.

Verhey, L. J.; Goitein, M.; McNulty, P.; Munzenrider, J. E.; Suit, H. D. 1982. Precise positioning of patients for radiation therapy. International Journal of Radiation Oncology Biology Physics. Vol. 8, Part 2, pp. 289-294.