2D-3D Registration Based on Shape Matching

Christopher M. Cyr, Ahmed F. Kamal, Thomas B. Sebastian, Benjamin B. Kimia
Division of Engineering
Brown University
Providence, RI 02912

Abstract

This paper describes a method for determining an object's pose given its 3D model and a 2D view. This 2D-3D registration problem arises in a number of medical applications, e.g. image guided spine procedures. Previous approaches often rely on a good initial estimate of the pose parameters and an optimization procedure to refine this initial pose estimate, e.g. the iterative closest point (ICP). However, such algorithms can identify local minima as global minima, leading to registration errors, if the initial pose is not carefully chosen. The specification of the appropriate initial conditions, however requires user interaction and is time consuming. We propose an approach where sample 2D views are generated from the 3D model and matched against the given view (2D-3D registration). Additional views are then generated in the vicinity of the best view and the procedure is repeated until convergence. Results of estimating the coordinates of a vertebrae spine bone from its 3D model, obtained from volumetric (CT or MR) data, and a 2D view, as might be obtained from fluoroscopic data, demonstrates that the pose can be reliably obtained without requiring extensive user interface.

1 2D-3D Shape Registration

Medical applications often require the alignment of coordinates between a 3D shape and a 2D image of it. This process of aligning spatial data to projective data is referred to as 2D-3D registration [8]. For example, in image guided spine procedures, pre-operative volumetric datasets, such as MR or CT of a patient, need to be matched against intra-operative two dimensional images, such as X-rays or fluoroscopic images [2]. While there is extensive work on 3D-3D and 2D-2D registration, 2D-3D registration has received relatively less attention. Problems associated with extrinsic registration methods, such as the invasive nature of the procedure, or poor accuracy if less invasive methods are used, have encouraged the development of intrinsic methods. Shape-based registration is a particularly attractive intrinsic method especially when registering
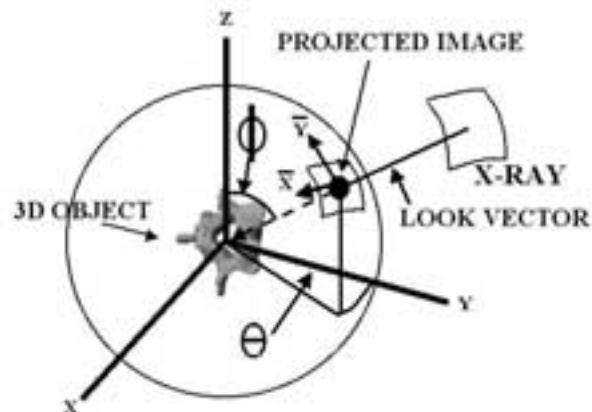

Figure 1: This figure illustrates the process of pose estimation by shape matching. The 3D object is shown at the center of a viewing sphere, depicted by the larger circle with an associated object-centered coordinate system. We assume the 3D model of the object is described in this coordinate system. A point on the sphere represents a projection (e.g., an X-ray direction) or a pose. This point is completely specified by the two angles, θ and ϕ as shown. The projection gives rise to an image described in an image-centered coordinate system. The goal of this 2D-3D registration is to recover the projection direction which produces the given 2D image from projection given the 3D model.

rigid structures such as bone. Such methods rely on geometric or anatomical features such as corners [8] or curvature [4] to determine registration parameters.

A popular registration method is the iterative closest point (ICP) algorithm [1]. In this approach, the shape to be registered is first placed in close proximity, in location and orientation, to the target shape. Second, for each point on one shape, the closest point on the target shape is found and the distance between the two shapes is computed as a least square sum of their distances. Third, the pose parameters of the shape are adjusted incrementally in the direction which minimizes the sum squared distances between the two shapes. The process of calculating and minimizing the distance is iterated
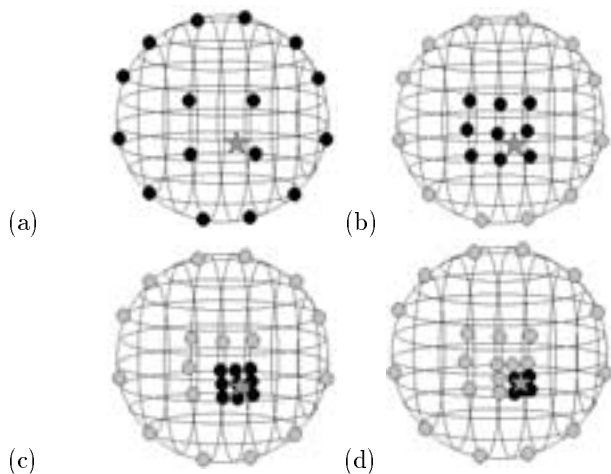
(a)

(b)

(c)

(d)

Figure 2: Hierarchical Iterative Registration. The viewing sphere is first coarsely sampled. For each point the projected image is compared to the target image and the best matching views are selected. These then represent a "focus" are, which is resampled at a higher resolution and the process is repeated to convergence. The global search is efficient and avoids the local minimum problem associated with gradient descent.

until convergence. This algorithm is effective, easy to implement, and robust, if the initial pose is close enough to the true pose of the target shape. However, with a distant initial pose, the process can converge to local minima, thereby requiring further user interaction or unnecessary delays.

Our 2D-3D registration approach avoids the local minima problem inherent in many gradient descent type approaches by hierarchically searching the space of possible poses using a notion of similarity between the projected shape and the 2D target shape, in a fashion reminiscent of the aspect graph approach [6, 7]. Specifically, a coarse sampling of the viewing sphere leads to a number of projected shapes which are each matched against the given view and a similarity score is obtained for each. The area in the vicinity of the best resulting view is then sampled in a finer manner and the best views are again selected. The procedure is then repeated to convergence, Figure 2.

This paper is organized as follows. Section 2 describes the process of generating views in a hierarchical, iterative fashion. Section 3 describes the shape representation and matching strategy used to select the most similar views. Section 4 presents results for spine vertebrae and carpal bones. Section 5 concludes the paper.
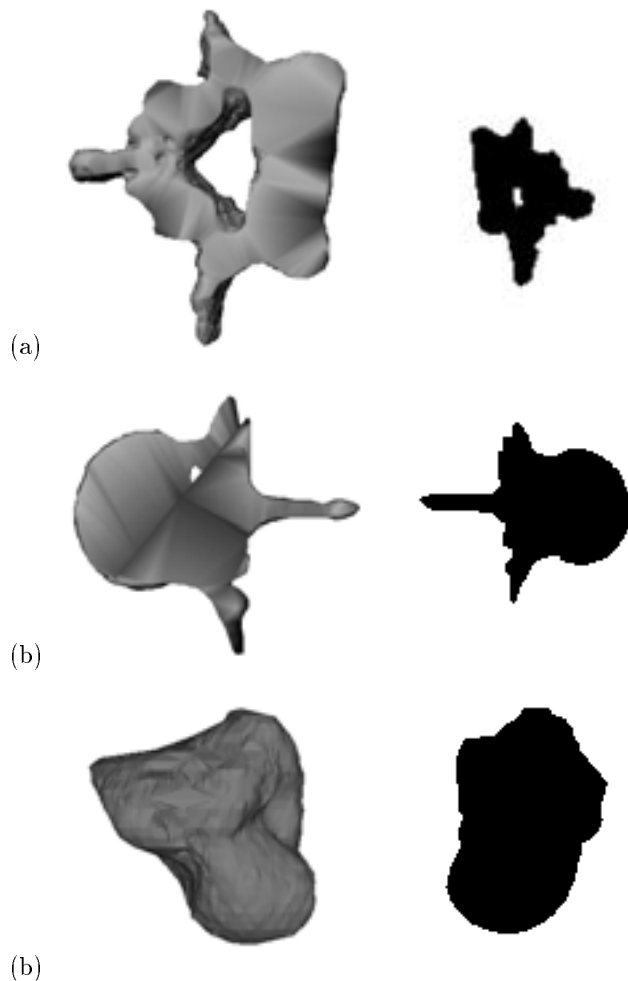

(a)

(b)

(b)

Figure 3: 3D Models and examples of projections.

2 Registration Based on Shape Matching

The 2D-3D registration method proposed here is based on a sampling of the viewing sphere and matching the given 2D view with the series of projected shapes to determine the best view. It is related to the "aspect graph" approach [6, 7], which parcellates the viewing sphere into equivalent views bounded by the singularities of the projection. The method presented here is similar in that pose estimation uses a parcellation of the viewing sphere, but one which is based on a uniform and hierarchical sampling. Specifically, the method iterates on the following steps. We assume the object is centered at the origin of the coordinate system, Figure 1.

First, in the initialization stage, the viewing sphere is sampled coarsely and uniformly into N views, Figure 2(a). The 3D object model is projected on the

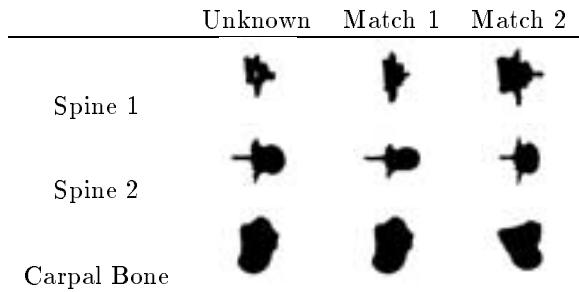|	Unknown	Match 1	Match 2
Spine 1			
Spine 2			
Carpal Bone			

Figure 4: The result of the first stage of matching is shown: the two best views of spinal vertebra (first and second row) and a carpal bone (third row) are shown.

target plane for each view. This generates N binary images of the projected shape (Figures 5 and 6). Each view is described by ϕ and θ, where ϕ is the angle between the viewing vector $\vec{v}$ and the $\vec{Z}$ axis, while θ is the angle between the projection of $\vec{v}$ and the X-Y plane and the $\vec{Y}$ axis. The two parameters θ and ϕ determine the unit viewing vector $\vec{v}$ uniquely as $\vec{v} = (sin\theta sin\phi, cos\theta sin\phi, cos\theta)$. We assume that the viewing vector, $\vec{v}$, is aligned with $\bar{Z}$, where $(\bar{X}, \bar{Y}, \bar{Z})$ describes the image coordinate axis. Having determined (ϕ, θ), this leaves one degree of freedom in relating $(\bar{X}, \bar{Y}, \bar{Z})$ to (X, Y, Z), namely the rotation of $(\bar{X}, \bar{Y}, \bar{Z})$ around $\bar{Z}$. This is specified by $\delta < (X, \bar{X})$. These three parameters, ϕ, θ, and δ, completely describe the relationship between the object-centered coordinate system (X, Y, Z) and the image-based coordinate system $(\bar{X}, \bar{Y}, \bar{Z})$. Out approach is to determine ϕ and θ via a shape matching procedure, and determine δ from the rotation angle that brings the projected shape into alignment with the viewed shape.

Second, the projected shapes are each matched against the viewed shape and ranked according to the extent of similarity. The basic assumption here is that projected views which are closer to the object pose generate more similar views. This is strictly not true for highly symmetric objects such as spheres, but holds for objects with any degree of asymmetry which allow various views to be distinguished. We will show that while for shapes with various protrusions the algorithm performs well, it also performs well with blobby shapes, such as carpal bones. As an example, consider Figure 4, where the unknown views of a spine vertebra and a carpal bone are matched against N samples from the viewing sphere. The matching strategy is discussed in Section 3. The assumption is that this matching procedure returns a similarity measure which can be used to rank order projected views according to extent of similarity to the given view.

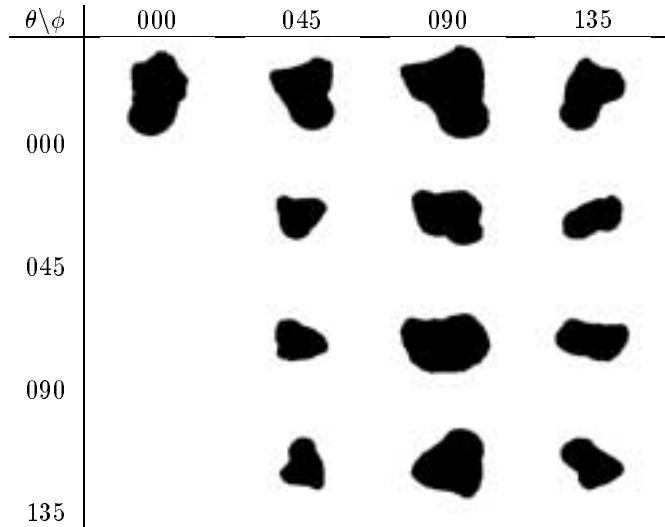| $\theta \backslash \phi$	000	045	090	135
000				
045				
090				
135				

Figure 5: Results of generating projections of carpal bone model at 45 degree intervals.

Third, having obtained an initial estimate of the object pose, the procedure is repeated but with novel projected samples closely centered around the best N views serving as foci of sampling. Figure 2 illustrates the iterative process. The black dots represent the views being used to generate the images compared in the shape matching. The gray dots represent previously used views. The process uses the closest matches in each iteration to determine a focus point for the set of views in the next iteration. The gray star represents the position of the camera when the 2D projection was generated. This is the unknown pose of the object which our algorithm is attempting to recover. The matching procedure is repeated to convergence as determined by the desired accuracy, and the pose of the object is recovered.

3 Matching Projected and Target Views

The method described in Section 2 relies on a matching strategy to indicate the proximity of two views of the same 3D object. Several methods have been proposed to measure the similarity of the shape based on silhouettes [3] or based on their medial axis [9, 10]. We selected the shock graph matching strategy known as the "graduated assignment algorithm" [5] used in [9]. This approach represents the shape by its medial axis with associated velocities, and matches two shapes by matching the resulting graphs (known as shock graphs). Two shock graphs are matched by a pair-wise assignment of nodes represented by a match matrix and a goodness of match functional which is the
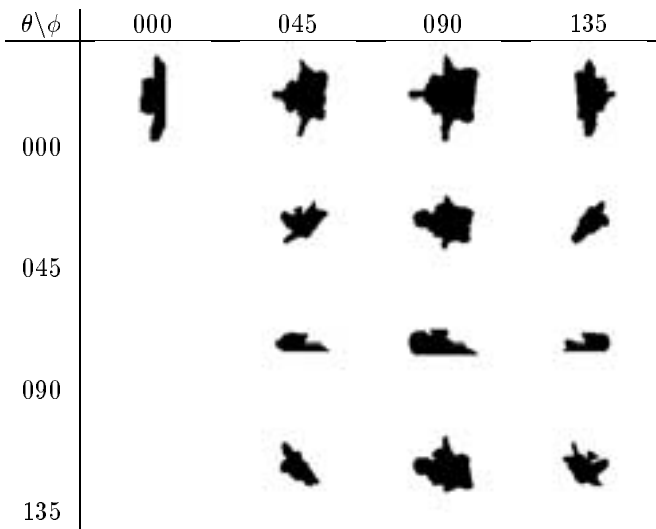
$\theta\backslash\phi$	000	045	090	135
000				
045				
090				
135				

Figure 6: Results of generating projections of first spine vertebra model at 45 degree intervals.

$\theta\backslash\phi$	000	045	090	135
000				
045				
090				
135				

Figure 7: Results of shock generation on projections of first spine vertebra model at 45 degree intervals.

sum over the goodness of match or any pair of links connecting any two pairs of nodes. The best match is the one that maximizes this functional which is found by a simulated annealing type algorithm. The energy of the best matches gives the extent of the similarity between the two shapes. In [9] this approach was used for indexing into image databases, whereas we use the matching strategy for a comparison of views of an object. Figure 7 illustrates the shock graphs for each of the views in Figure 6.

4 Experiments and Results

We now describe the details of our method used in a sequence of experiments to verify the validity of this approach in determining the relative pose of an object from a single view. We use three CT volumetric datasets, two from the spine and one from the wrist. The spinal vertebrae and carpal bone were segmented and are visualized in Figure 3. The algorithm initially selects N=13 viewing samples, corresponding roughly to a 45 degree separation between views. These views are the following in the (ϕ, θ) coordinate system: [(0,0), (45,0), (45,45), (45,90), (45,135), (90,0), (90,45), (90,90), (90,135), (135,0), (135,45), (135,90), (135,135)]. The 3D object model obtained from the segmentation from the CT data is projected onto these views. The shock graph of the projected view is obtained and compared to the target view, rank ordered by similarity, and the top M matches are selected.

The results generated at each step determine a range, between which the unknown position lies. This range can be found by starting at the closest match and traversing down the chart until 2 different results
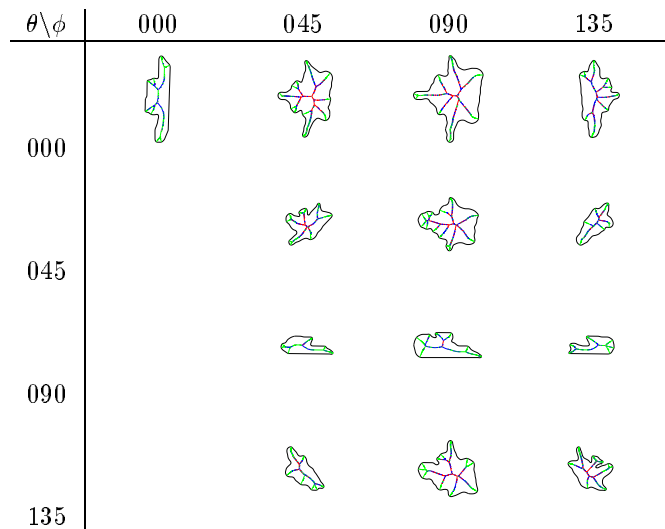
are found, in each of the ϕ and θ directions. Using Table 1 as an example, 45.0 degrees in ϕ becomes the first boundary. Moving to the next match displays 0.0 degrees as the next match. This forms the new range. The same process is used in the θ direction. The new range of projections to be used for matching becomes 0 and 45 degrees (ϕ) and 90 and 135 degrees (θ). The angle increment is decreased by half, from 45 degrees to 22.5 degrees, and new projection images are generated for this limited range. Shocks are generated, then matched and the results then decide the next range of angles for projection. This process is repeated until a resulting range is sufficiently small enough for the desired result, or the specific angle is discovered; Table 2 shows anther example. Figure 2 shows that the effect of decrementing the step and resetting the boundaries, causes a "focusing" into the unknown angles.

We now describe the experiments. The first spine vertebra, Figure 3(a), where the **3D model** was projected at a randomly generated view ($\phi = 19, \theta = 162$); this represents the unknown view. Table 3(a) represents the top three results of the rank ordering of the 13 initial views (separated by 45 degrees). The top two

Rank	ϕ	θ
1	**045.00**	**135.00**
2	**000.00**	**090.00**
3	045.00	090.00
4	090.00	135.00

Table 1: Sample results of shock matching.

Rank	ϕ	θ
1	**090.00**	**045.00**
2	090.00	**000.00**
3	**045.00**	000.00

Table 2: Results of carpal bone distance matching (unknown angles: (80,15) degrees)

	Rank	ϕ	θ
(a)	1	**000.00**	**135.00**
	2	000.00	**180.00**
	3	**045.00**	180.00
(b)	1	**022.50**	**157.50**
	2	000.00	157.50
	3	000.00	**180.00**
(c)	1	**022.50**	**157.50**
	2	**011.25**	**168.75**
	3	011.25	157.50
(d)	1	**023.00**	**157.00**
	2	**017.00**	157.00
	3	017.00	**163.00**
(e)	1	**020.00**	**160.00**
	2	**017.00**	160.00
	3	017.00	**163.00**
(f)	1	**018.00**	**160.00**
	2	019.00	162.00
	3	018.00	162.00

	Rank	ϕ	θ
(a)	1	**090.00**	**000.00**
	2	**045.00**	000.00
	3	045.00	**045.00**
(b)	1	**067.50**	**000.00**
	2	**090.00**	**022.50**
	3	067.50	022.50
(c)	1	**078.75**	**022.50**
	2	**090.00**	**011.25**
	3	078.75	011.25
(d)	1	**078.00**	**011.00**
	2	**084.00**	**017.00**
	3	084.00	011.00
(e)	1	**081.00**	**017.00**
	2	**078.00**	**014.00**
	3	078.00	017.00
(f)	1	**080.00**	**014.00**
	2	080.00	016.00
	3	079.00	015.00

Table 3: Results of pose recovery for spine model 1.

Table 4: Results of pose recovery for carpal bone model.

	Rank	ϕ	θ
(a)	1	**045.00**	**180.00**
	2	000.00	**135.00**
	3	000.00	045.00
(b)	1	**022.50**	**180.00**
	2	000.00	**157.50**
	3	000.00	180.00
(c)	1	**011.25**	**168.75**
	2	**022.50**	168.75
	3	011.25	**180**
(d)	1	**011.00**	**168.00**
	2	**017.00**	**174.00**
	3	023.00	168.00
(e)	1	**011.00**	**168.00**
	2	**014.00**	**171.00**
	3	011.00	171.00
(f)	1	**013.00**	**170.00**
	2	014.00	170.00
	3	014.00	171.00

	Rank	ϕ	θ
(a)	1	**045.00**	**000.00**
	2	**000.00**	000.00
	3	000.00	**045.00**
(b)	1	**022.50**	**045.00**
	2	**000.00**	**022.50**
	3	000.00	045.00
(c)	1	**011.25**	**022.50**
	2	**022.50**	**033.75**
	3	022.50	022.50
(d)	1	**023.00**	**022.00**
	2	**017.00**	**028.00**
	3	023.00	028.00
(e)	1	**020.00**	**022.00**
	2	020.00	**025.00**
	3	**023.00**	028.00
(f)	1	**021.00**	**025.00**
	2	021.00	023.00
	3	020.00	025.00

Table 5: Results of pose recovery for spine model 2.

Table 6: Results of pose recovery for spine model 2, on a second test run.

views "trap" the unknown pose between $0 < \phi < 45$ and $135 < \theta < 180$. The second iteration increases the sampling frequency to 22.5 degrees and the process is repeated in the interval selected by the first step. The results of the second step are shown in Table 3(b), where the interval $0 < \phi < 22.5$ and $157.5 < \theta < 180.0$ is correctly identified. Tables 3(c), 3(d), 3(e) and 3(f) reduce the sampling interval by half for each iteration so that after six iterations the approximate pose ($\widetilde{\phi}=18$, $\widetilde{\theta}=160$) is selected.

The second experiment used the second spine vertebrae, Figure 3(b) at the pose of ($\phi = 13, \theta = 170$). The initial round of matching gives a range of $0 < \phi < 45$ and $135 < \theta < 180$, which is increasingly refined, Tables 5(a) through 5(f). After six iterations, the resulting pose estimate is found to be ($\widetilde{\phi}=13$, $\widetilde{\theta}=170$). Since the exact pose is found, no further iterations are necessary.

The third experiment repeated the second experiment with a second unknown pose ($\phi = 20, \theta = 24$). Tables 6(a) through 6(f) show that the iterative procedure correctly zones in onto the target view for a final estimate after six iterations of ($\widetilde{\phi}=21$, $\widetilde{\theta}=25$).

The fourth experiment used a shape with no protrusions, a carpal bone, Figure 3(c), with an unknown pose of ($\phi = 80, \theta = 15$). Table 4(a) through 4(f) show that the iterations focus the projections correctly to the final increment, and select ($\widetilde{\phi}=80.0$, $\widetilde{\theta}=14.0$).

Finally, to quantify the pose estimation error between the actual pose (ϕ, θ) and the estimated pose ($\widetilde{\phi}$, $\widetilde{\theta}$), we obtained an error measure, E

$$E = \sqrt{(\phi - \widetilde{\phi})^2 + (\theta - \widetilde{\theta})^2}$$

which is obtained for various poses of the two spine vertebrae and displayed in Table 7. It is clear that with six iterations, we can reliably arrive at a pose estimate which are approximately 2 degrees of the actual pose, with a mean error $\mu=0.991$ and a standard deviation of $\sigma=0.7$. We plan to repeat this procedure for numerous other shapes and poses to determine the nature of the error more comprehensively

5 Conclusion

We have presented a technique for 2D-3D registration of 3D volumetric images such as CT and MR, and 2D projection images such as X-ray and fluoroscopic images. The method relies on a segmentation of a known structure, such as a bone or any other organ or structure with sufficient asymmetry in both the volumetric and projection images. Once the 3D and 2D shapes are obtained, we use a hierarchical shape matching scheme where coarse samples of the viewing sphere are matched against the target view, and the process

Model	Test	ϕ	θ	Error
Spine 1	1	018.00	160.00	2.236
Spine 1	1	019.00	162.00	0.000
Spine 1	1	018.00	162.00	1.000
Spine 2	1	013.00	170.00	0.000
Spine 2	1	014.00	170.00	1.000
Spine 2	1	014.00	171.00	1.414
Spine 2	2	021.00	025.00	1.414
Spine 2	2	021.00	023.00	1.414
Spine 2	2	020.00	025.00	1.000
Carpal Bone	1	080.00	014.00	1.000
Carpal Bone	1	080.00	016.00	1.000
Carpal Bone	1	079.00	015.00	1.414

Table 7: Results

is repeated for samples close to the top matches. We have shown that for spine vertebrae and carpal bones this process robustly and reliably recovers the correct pose to about 2 degrees of accuracy within six iterations. The main advantage of this approach is that it is a global process and avoids local minima, which are problematic for gradient descent techniques. In addition the computational burden is reduced; in the experiments shown, only 58 views were generated and matched for each.

One of the limitations of the current algorithm is the dependence on the accuracy of the shape matching process. Since the matching scheme selected here focuses on the topology of the shape via the medial axis it performs well for weakly similar shapes, and as the pose is refined the matching strategy is not sensitive to small changes for highly similar shapes. We plan to use an improved silhouette matching technique for the final iterations of out procedure. This should improve the accuracy of the registration scheme to the sub-degree range while maintaining the same reliability and efficiency, without user interaction.

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3D shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):239–256, February 1992.

[2] F. Betting and J. Feldmar. 3D-2D projective registration of anatomical surfaces with their projections. In *Information Processing in medical Imaging*, Lecture Notes in Computer Science, pages 275–286. Springer, 1995.

[3] I. Cohen, N. Ayache, and P. Sulger. Tracking points on deformable objects using curvature information. In *Proceedings of the European Conference on Computer Vision*, pages 458–466, Santa Margherita, Ligure, Italy, 1992.

[4] C. Davatzikos, J. Prince, and R. Bryan. Image registration based on boundary mapping. *IEEE Trans. on Medical Imaging*, 15:212–215, February 1996.

[5] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):377–388, 1996.

[6] J. J. Koenderink and A. J. van Doorn. The singularilarities of the visual mapping. *Biol. Cyber.*, 24:51–59, 1976.

[7] D. J. Kriegman and J. Ponce. On recognizing and positioning curved 3D objects from image contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):1127–1149, October 1990.

[8] J. A. Maintz, P. A. van den Elsen, and M. A. Viergever. Comparison of edge-based and ridge-based registration of CT and MR brain images. *Medical Image Analysis*, 1(2):151–161, 1996.

[9] D. Sharvit, J. Chan, H. Tek, and B. B. Kimia. Symmetry-based indexing of image databases. *Journal of Visual Communication and Image Representation*, 9(4):366–380, December 1998.

[10] S. C. Zhu and A. L. Yuille. Forms: A flexible object recognition and modeling system. *Int'l Journal of Computer Vision*, 20(3), 1996.