

# MULTI-MODAL 3D IMAGE REGISTRATION USING THE KULLBACK-LEIBLER DISTANCE

Gjenna Stippel  
University of Cape Town  
MRC/UCT Medical Imaging Research Unit  
Department of Human Biology  
<http://www.gjenna.com>  
email: [gstippel@cormack.uct.ac.za](mailto:gstippel@cormack.uct.ac.za)

## ABSTRACT

In this paper we address the problem of multi-modal co-registration of medical 3D images. Several techniques for the rigid registration of multi-modal images have been developed; in one of those the Kullback-Leibler distance is used to align 2D-3D angiographic images [1]. In this paper we refine this technique and investigate its performance on the registration of pairs of 3D CT/US images. We test the method with regard to accuracy of registration, robustness with regard to the starting point, and speed of convergence, and obtain promising results.

## KEY WORDS

multi-modal registration, image guided surgery, ultrasound, Kullback-Leibler distance, SPSA algorithm, computed tomography

## 1 Introduction

Registering intraoperative ultrasound (US) to preoperative images like computed tomography (CT) or magnetic resonance (MR) remains a challenging problem in the field of image guided surgery. For many surgical applications, ultrasound provides adequate clinical information to carry out the required intervention. There are some applications, however, where being able to interpret the ultrasound images in the context of the higher quality preoperative imaging has been shown to be helpful [2].

In this paper, we introduce a new rigid registration technique that aligns US and CT images by applying linear transformations to the coordinate system in order to minimize the Kullback-Leibler distance (KLD) between the observed joint intensity distribution, and a reference distribution representing the prior knowledge of the expected joint intensity histogram when the images are aligned properly [1].

We calibrate our KLD registration by taking an ultrasound scan, and manually aligning it with a CT scan to establish the fixed reference distribution. Afterwards, as more 3D ultrasound acquisitions are acquired over the course of the surgery, this calibration does not need to be repeated. Hence, the cost of the calibration is amortized over all the follow up scans during the procedure. If there

are many scans, the calibration cost is a small percentage of the total effort.

## 2 Description of the technique

The proposed technique is an intensity based registration algorithm. As explained in the Introduction, we align one image pair consisting of an image of the same organ from each modality manually. From this initial manual alignment a joint intensity histogram  $P(g_1, g_2)$  with  $0 \leq g_1, g_2 \leq 255$  is constructed as follows:

Let  $I_1$  and  $I_2$  be the intensity values of the two images with spatial domain  $X$  and  $Y$  respectively. Now the joint intensity histogram  $P$  is defined as:  $P(g_1, g_2) := |\{x \in X \cap Y | i_1(x) = g_1 \wedge i_2(x) = g_2\}|$  ( $0 \leq g_1, g_2 \leq 255$ ), where  $|S|$  denotes the cardinality of the set  $S$ .

This histogram serves as a reference distribution in all further alignments; in this way, the registration is steered by *a priori* knowledge. To register the new image automatically, one of the two images (the “floating image”) is translated and rotated iteratively while the other image is kept fixed (the “fixed image”); the aim is to determine the rigid transformation for which the resulting joint intensity distribution corresponds most with our reference distribution. As a measure to quantify the similarity of the two histograms, the so-called Kullback-Leibler distance is used.

### 2.1 Kullback-Leibler distance

Let the reference distribution be denoted by  $P_{\text{ref}}$  and the observed joint intensity distribution after the transformation  $T$  has been applied to the floating image by  $P_T$ . Now the Kullback-Leibler distance between the two distributions is defined as follows:

$$D(P_T || P_{\text{ref}}) := \sum_{i_1, i_2=0}^{255} P_T(i_1, i_2) \log \frac{P_T(i_1, i_2)}{P_{\text{ref}}(i_1, i_2)}.$$

The Kullback-Leibler distance is not a distance in the mathematical sense, since in general it is not symmetric:  $D(P_{\text{ref}} || P_T) \neq D(P_T || P_{\text{ref}})$  and thus does not satisfy one of the three criteria any distance has to satisfy. Two other

distance properties are met and are the most important for our purpose [3, 4]:

1.  $D(P_T||P_{\text{ref}}) \geq 0$ ,
2.  $D(P_T||P_{\text{ref}}) = 0$  if and only if  $P_T = P_{\text{ref}}$

These two properties make the Kullback-Leibler distance a useful measure to determine how far a particular alignment of the two images is from our “reference alignment” obtained from the manual calibration. For an image pair that is identical to the image pair used in the manual calibration, we observe that when these two images are perfectly aligned, the joint intensity histogram coincides exactly with the reference histogram, so in that case the KLD equals zero; otherwise the KLD is strictly positive.

The idea behind the registration technique is thus, to find a transformation  $T_0$ , acting on the floating image, that minimizes the KLD between the joint intensity distribution  $P_{T_0}$  and the reference distribution  $P_{\text{ref}}$ . Or, in formula:

$$T_0 = \arg \min_T D(P_T||P_{\text{ref}}) \quad (1)$$

The strategy we follow in the technique to find the optimal alignment (i.e., the one closest to the reference alignment) is as follows:

We keep the CT scan fixed during the entire procedure. The US image can be translated and rotated in any direction. Using the Simultaneous Perturbation Stochastic Approximation (SPSA) method [5–7], we search, by iterative translations and rotations, the position for the floating image at which the KLD between the corresponding joint intensity distribution and the reference distribution is minimal. According to Equation 1, this position is the optimal alignment.

## 2.2 Optimizer

In our experiments, we used the SPSA algorithm [8–10] as the optimizer. The SPSA algorithm is not as computationally complex as most gradient based methods, and it converges almost equally fast. An important thing to keep in mind when studying a system using the SPSA optimizer is that, since SPSA involves *stochastic* approximation, the convergence behaviour of such system shows a certain unpredictability; the objective function does not necessarily improve at each iteration.

## 3 Dataset

To generate our test set, we used a 3D abdominal phantom (CIRS, Norwalk, VA). The CT scan (512x512, 1mm slices) was acquired using the Somatom Plus4 (Siemens Medical Systems, Iselin, NJ). Two 3D US images of the liver (left lobe, right lower lobe) were generated using the Stradx software (Cambridge University, Cambridge, UK). The 3D US system consists of a Lynx ultrasound unit (BK Medical Systems, Wilmington, MA) and a miniBIRD tracking

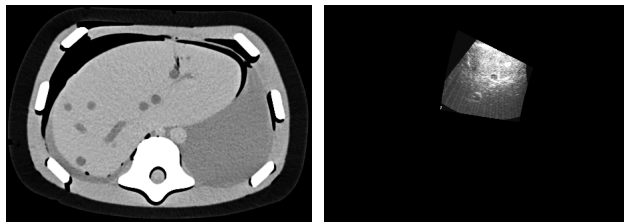


Figure 1. 2D section of volumetric CT and US scans as used in the experiments

system (Ascension Technology, Burlington, VT). For the calibration the CT and US images were manually aligned using the Slicer software (Brigham & Women’s Hospital, Boston, MA). In Figure 1 two 2D sections that are aligned, one of the CT scan and one of the US scan, are shown. The size of one 3D US image is: 460x324x18 voxels.

## 4 Experiment

In the validation of the performance of an image registration technique, the following three points are of concern:

1. Accuracy; How precise does the technique find the optimal alignment?
2. Robustness; How sensitive is the method to the choice of the initial position? At what distance of the initial position from the optimal alignment does the method still converge to the correct position?
3. Speed of convergence; How many iterations does the algorithm need to find the optimal solution?

Therefore, we designed the following experiment to address these three questions:

As explained in Section 3, we constructed our reference joint intensity histogram from the manual alignment of the US image of the left lobe of the liver and the CT scan. Furthermore, in all subsequent tests we performed to align the two images, we always kept the CT scan fixed, and moved only the US image.

We performed registrations with our newly developed technique from several initial points. As mentioned in Section 2.1, the algorithm searches for the optimal alignment iteratively: at each new position of the US image, a new joint intensity histogram is constructed, and the KLD between this joint intensity histogram and the reference histogram is calculated. Then, following the approach of the SPSA procedure, we do a stochastic translation, with step size dependent on the value of the KLD and the fixed parameters of the algorithm. In [7], the authors prove that with this SPSA procedure, the KLD converges to its minimum.

In separate arrays, we keep track of the points visited, the Manhattan distance<sup>1</sup> of each point from the “true (or ideal) alignment (0,0,0), and the corresponding KLD.

In general, because of its stochastic nature, the SPSSA algorithm does not necessarily converge to the minimum of the objective function monotonously, i.e. it is not necessarily true that each step yields a function value of the objective function that is closer to the minimum. This poses the problem in the practical setting, how we are actually going to *recognise* the minimum. In the, in practice non-existent, situation that the images that are being registered are the same as the ones used for the calibration, the KLD will converge to zero. However, in practice there are always noise and geometrical distortions involved. Hence, in this case the minimum will not equal 0 anymore. If the distortions are not too big, the minimum will be at the same position, though.

Since in our experimental set-up, we know that the ideal registration is the origin (0, 0, 0), we can determine for which value of the KLD the moving image is close to the ideal position. For each initial position, we searched for the minimal value of KLD with our the proposed technique. We let the algorithm run until the Manhattan distance of the position that is reached is closer than 1.0 from the origin. From that point on, we iterated 10 more times and then finished. We recorded the following six things:

1. The number of iterations after which the Manhattan distance is for the first time smaller than 1.5,
2. The KLD at this point,
3. The number of iterations after which the Manhattan distance is for the first time smaller than 1.0,
4. The KLD at this point,
5. The number of iterations after which the KLD has its global minimum for the first time,
6. The value of the KLD at its global minimum.

Several graphs deduced from the results of these experiments can be found in Section 5. By studying these, we could gain insight in the convergence behaviour of the proposed registration technique, and answer the three questions stated in the beginning of this Section.

Note, finally, that in our experiment, we only studied translations. Rotations can be included in exactly the same way; in that case the search has to be performed in a six-dimensional space instead of a three-dimensional one. Experiments including rotations are currently conducted.

## 5 Results

In Figure 2, the relation between the initial points from which we started the algorithm and the number of itera-

<sup>1</sup>The Manhattan distance between two points is the sum of the absolute values of the differences between the coordinates of the two points, i.e.,  $d_{\text{Manh}}((a_1, a_2, a_3), (b_1, b_2, b_3)) = \sum_{i=1}^3 |a_i - b_i|$ .

Table 1. Initial points

point	distance from origin
(0,0,60)	60
(0,10,-15)	25
(0,15,15)	30
(4,20,20)	44
(0,40,20)	60
(2,50,-10)	62
(4,70,-40)	114
(0,80,-80)	160
(-2,-30,60)	92
(0,-30,-30)	60
(8,-30,-10)	48
(0,-100,0)	100

tions needed to convergence for different levels of accuracy are displayed. In Table 1, we summarise the twelve initial points, and their Manhattan distances to the origin. As can be seen, the points on the abscis in Figure 2 are ordered on their second coordinate; the first eight have a positive second coordinate, and are ordered from low to high, the last four have a negative second coordinate and are ordered from least negative, to most negative.

We can conclude several things from Figure 2. First of all, we can see that once the  $\text{KLD} < 1.5$ , it takes very few iterations in order to get  $\text{KLD} < 1.0$ . This means, that at the end, the algorithm converges relatively fast. Furthermore, it takes again very few iterations (in 50% of our trials even zero iterations) to go from the point that  $\text{KLD} < 1.0$  to a global minimum. If we relate the results of Figure 2 with the distance of the initial points from the origin, as calculated in Table 1, we see that there is little correspondence between the initial distance of starting point and the number of iterations. Far more important is from which direction the ideal position is approached. There is a striking difference in convergence speed when the initial point is chosen with a negative second coordinate (last four points), or with a positive second coordinate (points five till eight). From the positive side, the algorithm converges extremely slowly. When the point is chosen near the origin (first four points), the algorithm converges relatively fast.

In Figure 3, a graph of how the value of the KLD evolves while the algorithm is running is shown. The initial point in this example is: (8,-30,-10). Note that in the SPSSA algorithm, three KLD values are calculated in each iteration. The shape of graph is typical for all the graphs we obtained in this experiment: first there is, generally, a relatively fast decrease of the function values, then they stabilize and continue to decrease very slowly. The global minimum is almost always an outlier (i.e., much lower than the adjacent KLD values). Finally, the graph gives us an idea of the normal range of the KLD values: when the algorithm is approaching the origin, the KLD value is  $\approx 0.002$ .

When we compare Figure 3 with the values of the

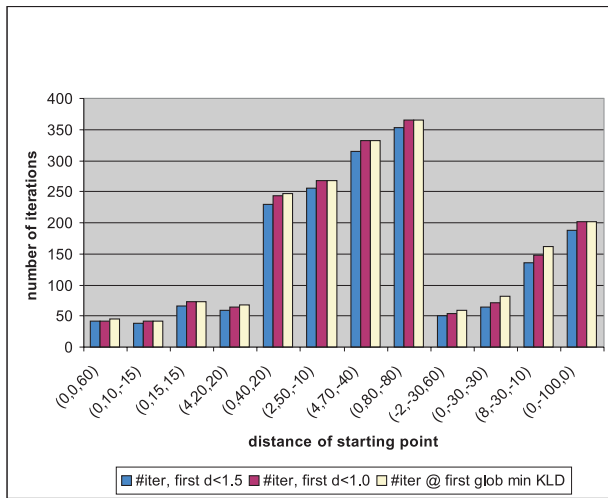


Figure 2. Relation between distance of initial position and number of iterations

Manhattan distance in Figure 4, we see that the KLD gives indeed a good indication of how close the position of the image is to the ideal alignment. The measurements of the Manhattan distance follow that of the KLD closely, and the point where the KLD reaches its global minimum, at the 161<sup>st</sup> iteration, is indeed exactly the same point where the Manhattan distance is smallest.

In Figure 5, we summarize the value that the KLD reaches at the three “test points” close to the origin for all the initial points we considered. First, we notice that the KLD is generally significantly lower when the Manhattan distance is decreasing below 1.0, than when it is just smaller than 1.5. This shows, once again, how sensitive the KLD is: when the image is within one pixel accuracy from the optimal alignment, the KLD falls dramatically. In 9 of the 12 cases, the point where the Manhattan distance decreases below 1.0, is also the point where the KLD equals 0, i.e., the optimal alignment is reached. Finally, in all cases the KLD at the global minimum is less than 0.0012. Therefore, in practice, we can use this value as a limit to determine that the algorithm converged: if the value drops below 0.0012, we can stop iterating.

Note, that the global minimum of the values of the KLD of an one experiment is in three cases bigger than zero. The reason for that is, that for those three images the values KLD decreases so slowly, that the value is never reached. (Remember that we let the algorithm run, until the Manhattan distance was smaller than 1.0, and that we then iterated ten more times. In these three cases those ten times were not enough to let the KLD converge to zero).

## 6 Conclusion

In this paper we introduced a novel technique for the rigid registration of volumetric multi-modal medical images. We

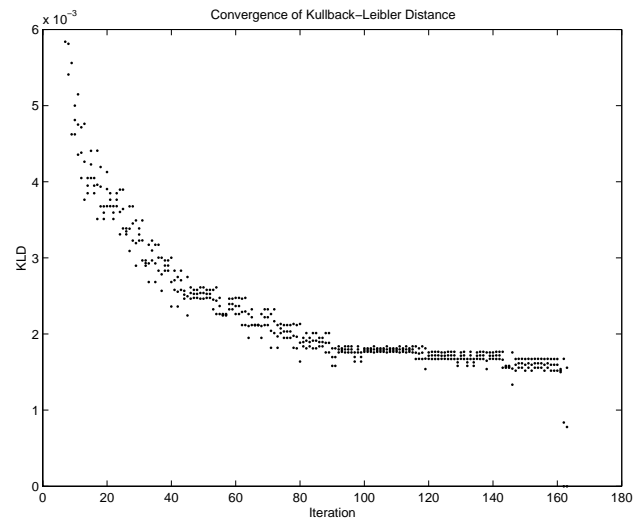


Figure 3. Value of KLD at each iteration, when converging with proposed algorithm from the initial point (8,-30,-10)

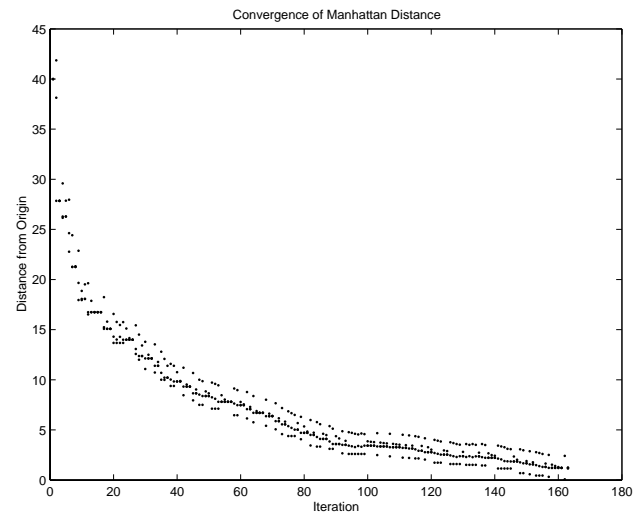


Figure 4. Manhattan distance from origin at each iteration, when converging with proposed algorithm from the initial point (8,-30,-10)

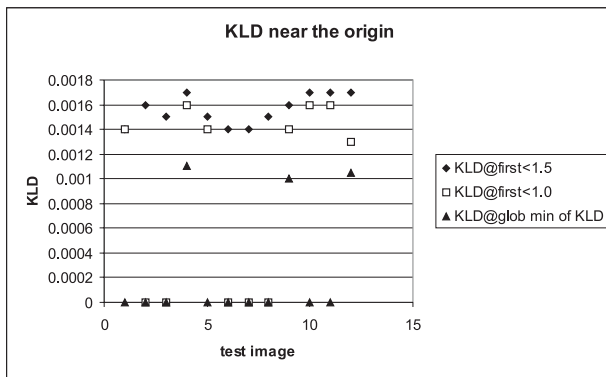


Figure 5. The KLD at the first iteration in which the Manhattan distance is  $< 1.5$ , at the first iteration in which the Manhattan distance is  $< 1.0$ , and at the global minimum that is reached in the experiment

investigated its performance on a test set of a pair of CT and volumetric US images of an abdomen phantom. The proposed technique succeeds well in aligning the US-CT pair. It turns out to be accurate and robust, in the sense that it converges accurately, regardless of the chosen start position. The speed of convergence is far more dependent on the direction of the initial point relative to the ideal registration (i.e., the contents of the image), than on the distance from the ideal position. We are currently conducting further tests on human images to assess its suitability for clinical application.

## 7 Acknowledgements

I thank Prof. dr. Simon K. Warfield, and Prof. dr. Sandy Wells for their helpful comments and for creating the possibility for me to work on this project in the Surgical Planning Lab, Brigham & Women's Hospital, Harvard Medical School, Boston, USA.

## References

[1] Albert C.S. Chung, William M. Wells III, Alexander Norbash, and W. Eric L. Grimson, "Multi-modal image registration by minimising kullback-leibler distance," in *Proceedings of MICCAI*, 2002, pp. 564–571.

[2] J. Ellsmere, J.A. Stoll, D.W. Rattner, D. Brooks, R. Kane, W.M. Wells III, R. Kikinis, and K.G. Vosburgh, "Integrating preoperative ct data with laproscopic ultrasound images facilitates interpretation," *Society of American Gastrointestinal Endoscopic Surgeons*, vol. 17, pp. S296, 2003.

[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[4] S. Kullback, *Information Theory and Statistics*, Dover Publications, Inc., 1968.

[5] James C. Spall, "An overview of the simultaneous perturbation method for efficient optimization," *John Hopkins Apl. Technical Digest*, vol. 19, no. 4, pp. 482–492, 1998.

[6] James C. Spall, "Implementation of the simultaneous perturbation algorithm for stochastic optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 3, pp. 817–823, July 1998.

[7] James C. Spall, "Adaptive stochastic approximation by the simultaneous perturbation method," in *37th IEEE Conference on Decision & Control, Tampa, Florida, USA*, December 1998.

[8] John L. Maryak and Daniel C. Chin, "Global random optimization by simultaneous perturbation stochastic approximation," in *American Control Conference, Arlington VA*, June 2001.

[9] Laszlo Gerencser, Stacy D. Hill, and Zsuzsanna Vago, "Optimization over discrete sets via spsa," in *38th Conference on Decision & Control, Phoenix, Arizona USA*, December 1999.

[10] Laszlo Gerencser, Stacy D. Hill, and Zsuzsanna Vago, "Discrete optimization via spsa," in *American Control Conference, Arlington VA*, June 2001.