Rigid registration of CT, MR and Cryosection images using a GLCM framework

Morten Bro-Nielsen[123]

[1]HT Medical, Inc., 6001 Montrose Road, Suite 902, Rockville, MD 20852, USA
[2]3D-Lab, School of Dentistry, Univ. of Copenhagen, Denmark
[3]Dept. of Mathematical Modelling, Technical University of Denmark
e-mail: bro@ht.com WWW http://www.imm.dtu.dk/~bro

Abstract. The majority of the available rigid registration measures are based on a 2-dimensional histogram of corresponding grey-values in the registered images. This paper shows that these features are similar to a family of texture measures based on Grey Level Cooccurrence Matrices (GLCM). Features from the GLCM literature are compared to the current range of measures using images from the visible human data set. The voxel-based rigid registration of Cryosection and CT images have not been reported before. The tests show that mutual information is the best general measure, but some GLCM features are better for specific modality combinations.

This paper discusses existing and some new voxel similarity measures for image registration. Elaborate tests are used to evaluate the different measures and compare them. Finally, a registration algorithm based on voxel similarity measures is described and some results are presented.

1 Image data

The algorithms developed in this paper have been applied to registration of images from the Visible Human data set [20]. From this data set, images of the head from the following modalitites have been used:

- MR, T1 weighted (T1)
- CT, windowed for bone (CT)
- Red channel of the cryosection colour image (R)

These images were taken from the Research Systems' Visible Human CD.

Using a combination of manual and automatic tools the images were registered to each other to get an initial ground truth. This registration was performed carefully using visual inspection for validation of the results. Unfortunately, during this registration process, the voxel size of the MR images turned out to be inconsistent with the size of the other images. By measuring the distance between anatomical landmarks the voxel size were estimated to 1.05 x 1.05 x 5 mm instead of 1.016 x 1.016 x 5 mm as given in the documentation for the MR images. The following combinations of modalities are explored in this paper: CT/T1, and CT/R. Voxel-based registration of CT and cryosection images has not been documented before.

2 Voxel similarity measures

For registration of uni-modal images, correlation has been used extensively in both remote sensing, medical imaging and other application areas. Simple correlation of grey-values assumes that a *linear* relationship between the grey-values exists [2]. This is seldom the case, and grey-level correlation has, therefore, not provided convincing results for multi-modality registration of images.

In recent years, though, renewed interest in voxel-based multi-modality registration has been revived by the successful work on PET/PET and PET/MR registration by Woods et al. [23, 24]. The basic assumption of this work is the same as for correlation, ie. that a linear mapping exists between grey-values g_1 and g_2 of the two images. As mentioned above, this assumption is seldom valid for multi-modality images. But Woods et al. circumvent this problem by looking instead at the variance of the coefficient $R = g_1/g_2$, where g_1 is the PET image grey-value. They argue that this *coefficient of variation* is minimized when the images are in register, and have achieved good results for PET/PET registration [23] using this measure. For PET/MR registration they have proposed a modified version of the initial measure [24], where the variance is calculated independently for each MR grey-value and subsequent summed weighted by the probability estimate of the MR grey-values. To achieve successful registration, only the intracranial structures are used in the registration process, and this algorithm, therefore, needs some manual segmentation to work. But, the coefficient of variation is today probably the best measure for registration of PET/PET and PET/MR [22].

Inspired by this work, Hill et al. proposed a modified algorithm for registration of CT/MR in [12, 13]. In this algorithm CT is used as the denominator g_2, and only certain ranges of CT intensities are used in the calculation of the resulting coefficient.

In [12] Hill also proposed an alternative measure based on the *third order moment* of the 2D histogram created from the images. This was inspired by intensive studies of the development of the 2D histograms for changing registration parameters. A general observation was that intensity concentrations in the histograms seemed to disperse when the registration deviated from an optimal registration.

Van den Elsen has proposed a modified correlation approach for CT/MR registration [7, 8, 9], where the images are pre-processed to extract similar structures in both modalities, typically bones. In [7, 9] these structures were extracted using complex differential operators in scale-space. Similar results were later obtained using simple ramp intensity remapping in [8].

At this point all the measures proposed for multi-modality registration had been based on heuristics. Several groups independently realized that the intrinsic problem of registering two independent image modalities, could be cast in an information theoretic framework. Collignon et al. [3] and Studholme et al. [17] both suggested using the *joint entropy* of the combined images as a registration potential, and Collignon et al. [4], and Wells and Viola [21] finally suggested the *relative entropy* or *mutual information* as a registration measure. Mutual information is more robust to truncation of images than joint entropy, and has been applied to other registration tasks than medical imaging. It is a very general measure of correspondence between two images, and in a recent evaluation of a range of different multi-modality registration methods [22], mutual information was quite succesful.

2.1 GLCM matrices

Except for the work of Van den Elsen [7, 8, 9] all the voxel similarity measures introduced above can be formulated based on the 2D histogram or joint probability distribution of the two images.

A similar family of measures is found in the texture analysis literature on *Grey Level Cooccurrence Matrices* (GLCM) [5, 6, 10, 11]. The GLCM is determined as the 2D plot of grey-values of voxels in an image with a fixed displacement between them.

Let $g(\boldsymbol{x})$ be the grey-value of the pixel at position $\boldsymbol{x} = [x_1, x_2, x_3]^T$ in the image, and let $\boldsymbol{u} = [u_1, u_2, u_3]^T$ be the displacement vector between corresponding voxels. The GLCM is generated by accumulating the grey-value pairs $[g(\boldsymbol{x}), g(\boldsymbol{x}+\boldsymbol{u})]$ in a 2D histogram for all image positions $\boldsymbol{x}$. The normalized GLCM can be seen as an estimate of the joint probability distribution of voxels $g(\boldsymbol{x})$ and $g(\boldsymbol{x}+\boldsymbol{u})$.

By changing the definition of the displacement vector $\boldsymbol{u}$ to be, not between different voxels in one image, but between the same voxels in *different* images, the GLCM turns out to be the 2D histogram of voxel intensities used by Hill et al. [12, 13, 15], and the normalized GLCM becomes an estimate of the joint probability distribution of voxels in the two images.

In the GLCM texture analysis literature a range of different measures exists. On the following pages we evaluate these measures as voxel similarity measures for multi-modality image registration, and compare them to the existing voxel similarity measures.

2.2 GLCM features

Most of the GLCM features are derived by weighting the entries of the GLCM with a weighting function and summing the result. The features fall in three classes based on the character of the weighting function.

Using the notation $P(i, j)$ for elements of the normalized GLCM, the general form of the GLCM features is:

$$F = \sum_{i,j} w(i, j) P(i, j) \tag{1}$$

where the weighting function w depends either on the normalized GLCM value ($P(i,j)$), the spatial position in the GLCM ((i,j)), or both.

Notation As above $P(i,j)$ is the value of the normalized (n_i, n_j) GLCM at position (i,j).

$$N = n_i n_j \qquad P_i(i) = \sum_j P(i,j) \qquad P_j(j) = \sum_i P(i,j)$$

$$\mu_i = \sum_i i P_i(i) \qquad \mu_j = \sum_j j P_j(j)$$

$$\sigma_i^2 = \sum_i (i-\mu_i)^2 P_i(i) \qquad \sigma_j^2 = \sum_j (j-\mu_j)^2 P_j(j)$$

Features: Weighting dependent on $P(i,j)$

$$Energy = \sum_{i,j} P(i,j)^2$$

$$Variance = \sum_{i,j} (P(i,j) - \frac{1}{N})^2$$

$$Entropy = -\sum_{i,j} P(i,j) log P(i,j)$$

$$MI = -\sum_{i,j} P(i,j) log \left(\frac{P(i,j)}{P_i(i) P_j(j)} \right)$$

$$IDM = \sum_{i,j} \frac{1}{1+(i+j)^2} P(i,j)$$

$$Inertia = \sum_{i,j} (i-j)^2 P(i,j)$$

$$Dmoment = \sum_{i,j} |i-j|(i+j-\sigma_i-\sigma_j) P(i,j)$$

$$Correlation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} P(i,j)$$

$$Cshade = \sum_{i,j} (i+j-\sigma_i-\sigma_j)^3 P(i,j)$$

$$Cprominence = \sum_{i,j} (i+j-\sigma_i-\sigma_j)^4 P(i,j)$$

$$Woods = \sum_{i>0,j} \frac{s_i}{a_i} P(i,j)$$

$$a_i = \frac{1}{P_i(i)} \sum_j P(i,j) j \qquad s_i = \frac{1}{P_i(i)} \sum_j P(i,j)(j-a_i)^2$$

Note, that the Woods MR/PET registration measure is not symmetric.

Some of these features have been used before for multi-modality image registration (entropy, mutual information, correlation coefficient, and Woods MR/PET) whereas the rest are new for medical image registration. Both groups are included for comparison.

2.3 Implementation

A sample size or sample frequency must be decided upon when the GLCM is calculated. The sample has to be large enough to incorporate enough information about the registration, but at the same time small enough to allow efficient computation. We use a scheme similar to that of Collignon et al. [4]. The tests described in this section have all been performed using super-sampling with a factor of 2.

When the GLCM is calculated for two images, which overlap in such a way that voxels of one image maps to inter-voxel positions in the other image, it is necessary to estimate the grey-values using interpolation. In this work tri-linear interpolation is used.

The joint probability $P(i, j)$ is estimated from the GLCM. For 12-bit images the raw GLCM contains 4096×4096 bins which is bigger than some images. A reduction of the number of bins is therefore necessary to allow efficient computation. We use simple binning witg 256 bins, implemented with binary shifts.

2.4 Plotting GLCM features

In [14, 15] Hill et al. used socalled *similarity measure plots* to determine the quality of voxel similarity measures. These plots show curves for displacements in the different directions, and rotations around the three axes under the assumption that the other parameters are zero.

Obviously this kind of plot does not provide any information about the quality of the measures for deviations where several parameters are non-zero. In addition, these plots do not allow quantitative evaluation of the measures and objective comparison is not easy. On the other hand they do tend to give a good impression of the behaviour of the measures in terms of local minima and precise localization.

In the following the measures are evaluated using two types of plots:

The similarity measure plot that Hill et al. have used. The similarity measure is determined for a sequence of deviations with a single parameter at a time. This gives a curve for each parameter and these curves are combined in a single plot.

Distance/Feature plots. For a large number of random displacements, the length of the parameter vector is plotted against the feature. It turns out that these plots are reasonably linear for good similarity measures. We therefore choose the linearity as an objective measure of the feature quality. Linear regression is used to determine the best approximating line (using the Splus software package) and the R^2 is used as a quality measure.

Correcting for wrong scaling of rotation When the length of the parameter vector is determined, an implicit choice of scaling for the rotation parameters, compared to the translation parameters, has to be made. The obvious choice is using millimeters for translations and degrees for rotations. In the medical image registration literature this has been used widely (if not exclusively), eg. [4, 9, 15]. There is no theoretical basis for this choice and any other could just as well have been used. Work in this paper indicates that it is often poor choice.

For algorithms that use a brute-force approach to determine the minimum of the similarity function [9, 15] this has little influence. But where more advanced methods such as Newton-Raphson [23, 24] or Powell's method [4] are used, different scaling of the rotation and translation axes can influence the direction of steps or stop-requirements. For calculation of the distance/feature plots the scaling also has an effect. It is therefore necessary to estimate the correct scaling.

Two distance/feature plots are created, where one uses only rotation and the other only translation in the parameter vector. Using linear regression, approximating lines are determined for these two plots. Assuming that the estimates of the slopes of the lines are α_{rot} and α_{tr} for rotation and translation, respectively, a correction factor is determined as $\gamma_{rot2tr} = \alpha_{rot}/\alpha_{rtr}$ This correction factor is pre-multiplied all rotation parameters before the length of the parameter vector is determined. Using corrected rotation parameters, a final distance/feature plot is calculated where all parameters take random values.

2.5 Results

The actual set of plots is not shown here for space reasons (Refer to [1]). Instead derived information from the plots is described.

The similarity measure plots call for a subjective evaluation and we have performed this evaluation using the following scale:

1. Useless,
2. Poor localization with serious local minima,
3. Reasonable localization of optimum, some small local minima,
4. Reasonable localization of optimum, smooth curve without local minima,
5. Perfect localization of optimum, smooth curve without local minima.

The results of the classification and corrected linear regression are shown in tables 1, 2 and 3. They show that the information theoretic measures entropy and mutual information perform consistently well. This is in line with the image registration literature [3, 4, 17, 18, 19] which also indicates that mutual information is better than entropy for truncated images [18], ie. where parts of one image is not present in the other.

The results of the other measures are mixed, but it is interesting to note that the measures with weights based on the position (i, j) (and $P(i, j)$) in the normalized GLCM do quite well in the CT/R experiments. Indeed the Diagonal Moment perform better than the entropy and mutual information measures. Without jumping to any conclusions, this could indicate that position weighted measures can do well if the weighting matches the problem.

Table 1. *Pd-T1: Similarity measure plot of quality result and R^2 compared with uncorrected R^2. 500 samples are used.*

	Quality	R^2	Uncorrected R^2
Energy	2	0.9384729	0.9088758
Variance	2	0.9384698	0.9088723
Entropy	5	0.9760593	0.9708134
MI	5	0.9260448	0.9214459
IDM	5	0.9685928	0.9599456
Inertia	4	0.9173130	0.9061781
Dmoment	4	0.9187858	0.9159433
Correlation	4	0.9057705	0.8883932
Cshade	3	0.8380936	0.8114711
Cprominence	3	0.8265533	0.8120783
Woods (X:Pd)	3	0.7361181	0.7412743
Woods (X:T1)	3	0.5480094	0.5482602

Table 2. *CT/T1: Similarity measure plot quality results and R^2 compared with uncorrected R^2. 500 samples are used.*

	Quality	R^2	Uncorrected R^2
Energy	4	0.9505296	0.9493070
Variance	4	0.9505302	0.9493080
Entropy	5	0.9666757	0.9638251
MI	4	0.8077108	0.7595025
IDM	2	0.9416978	0.9384330
Inertia	1	0.6917992	0.6948765
Dmoment	1	0.2931991	0.2798332
Correlation	1	0.5524907	0.5277128
Cshade	1	0.0773453	0.0773202
Cprominence	1	0.3667449	0.3952463
Woods (X:CT)	4	0.6376055	0.6018653
Woods (X:T1)	4	0.8147630	0.8139735

3 Image registration using voxel similarity measures

A registration algorithm similar to that of Collignon et al. [4] has been implemented. The method optimizes the registration using Powell's algorithm for optimization without derivatives [16]. Multi-resolution is used to speedup the algorithm.

A Quasi-Newton algorithm was tested, but problems calculating stable estimates of the first derivatives caused the results to be somewhat poor.

All the voxel similarity measures may be used for the registration. But in practice we have preferred the mutual information most of the time, since it provides consistent results for different modalities.

The result of the 3D registration of the MR T1 weighted image to the CT bone windowed image using mutual information, is shown in figure 1. The 3D

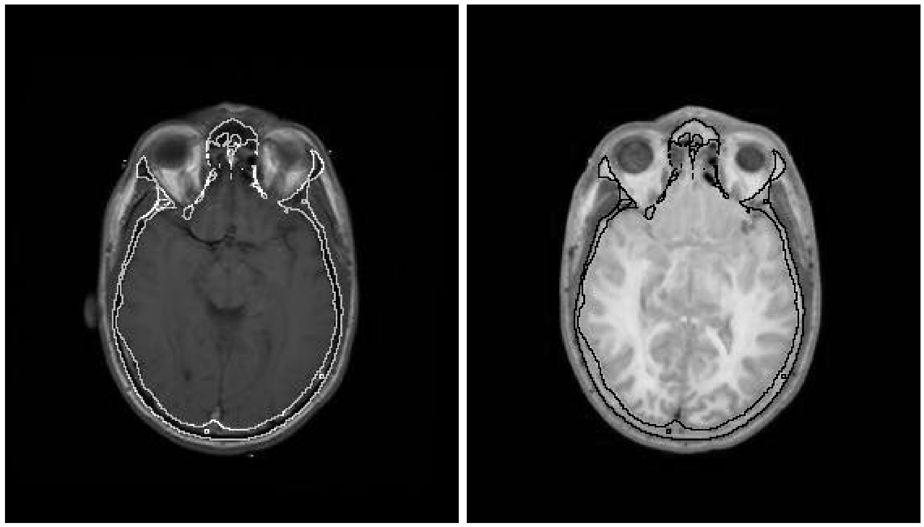

Fig. 1. *Left: Result of 3D registration using mutual information of the CT bone windowed image to the MR T1 weighted image. Right: Result of 3D registration using mutual information of the CT bone windowed image to the Red channel of the cryosection image. The outline of the thresholded CT image has been overlayed on both images.*

registration of the Red channel of the cryosection image to the CT bone windowed image is shown in figure 1.

Results of the registration could only be validated by visual inspection and exhaustive test were therefore not performed. But the visual inspection of the results showed that the registration was quite precise.

Table 3. *CT/R: Similarity measure plot quality results and* R^2 *compared with uncorrected* R^2*. 500 samples are used.*

	Quality	R^2	Uncorrected R^2
Energy	1	0.4335973	0.3803990
Variance	1	0.4335405	0.3803409
Entropy	5	0.9801963	0.9790247
MI	4	0.9016652	0.8648146
IDM	3	0.9476440	0.9375267
Inertia	5	0.9721051	0.9725199
Dmoment	5	0.9896430	0.9720567
Correlation	4	0.9176131	0.8917624
Cshade	4	0.8500580	0.7844821
Cprominence	4	0.8491666	0.7947552
Woods (X:R)	3	0.7792640	0.7764421
Woods (X:CT)	2	0.9013030	0.8753324

4 Summary

In this section voxel similarity measures for registration of the Visible Human data set have been explored.

The 2D histogram of joint voxel intensities, used in the literature as a basis for definition of many voxel similarity measures, was shown to be similar to the GLCM matrices used in texture analysis of images.

A range of features from texture analysis were compared to the state-of-the-art features. This comparison showed that the state-of-the-art features entropy and mutual information were best for general registration, since they performed consistently well for both registration of MR-T1 to CT bone, and red cryosection to CT bone. For each of the other combinations, some of the texture measures were at least as good as the information theoretic measures. But, these results were not consistent from one modality combination to the next.

Together with the information from the literature, this leads to the conclusion that mutual information is the best generally applicable voxel similarity measure.

Since most of the texture measures were dependent on the position in the GLCM, in contrast to the information theoretic measures, it should be explored whether position dependent weights adapted to the registration problem (modality combination) could improve registration results. Preliminar work in this direction did not yield positive results.

References

1. M. Bro-Nielsen, *Medical image registration and surgery simulation*, Ph.D. thesis in submission, 1996
2. L.G. Brown, *A survey of image registration techniques*, ACM Computing Surveys, 24(4):325-376, 1992
3. A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal, *3d multi-modality image registration using features space clustering*, Proc. Computer Vision, Virtual Reality, and Robotics in Medicine (CVRMed'95), 195-204, 1995
4. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, *Automated multi-modality image registration based on information theory*, Proc. Information Processing in Medical Imaging, 263-274, 1995
5. R.W. Conners and C.A. Harlow, *A theoretical comparison of texture algorithms*, IEEE Trans. Pattern Analysis and Machine Intelligence, 2(3):204-222, 1980
6. R.W. Conners, M.M. Trivedi, and C.A. Harlow, *Segmentation of a high-resolution urban scene using texture operators*, Computer Vision, Graphics, and Image Processing, 25:273-310, 1984
7. P.A. van den Elsen, *Retrospective fusion of CT and MR brain images using mathematical operators*, AAAI Spring Symposium Series: Applications of Computer Vision in Medical Image Processing, 30-33, 1994
8. P.A. van den Elsen, T.S. Sumanaweera, P.F. Hemler, S. Napel, and J. Adler, *Grey value correlation techniques used for automatic matching of CT and MR brain and spine images*, Proc. SPIE Vol. 2359: Visualization in Biomedical Computing, 227-237, 1994

9. P.A. van den Elsen, J.B.A. Maintz, E-J.D. Pol, and M.A. Viergever, *Automatic registration of CT and MR brain images using correlation of geometrical features*, IEEE Trans. on Medical Imaging, 14(2):384-396, 1995
10. R.M. Haralick, K. Shanmugan, and I. Dinstein, *Textural features for image classification*, IEEE Trans. on Systems, Man, and Cybernetics, 3:610-621, 1973
11. R.M. Haralick, *Statistical and structural approaches to texture*, Proc. of the IEEE, 67(5):786-804, 1979
12. D.L.G. Hill, *Combination of 3D medical images from multiple modalities*, Ph.D. thesis, University of London, 1993
13. D.L.G. Hill, D.J. Hawkes, N. Harrison, and C.F. Ruff, *A strategy for automated multimodality registration incorporating anatomical knowledge and imager characteristics*, Proc. Information Processing in Medical Imaging, Lecture Notes in Computer Science 687, Springer-Verlag, Berlin, 182-196, 1993
14. D.L.G. Hill and D.J. Hawkes, *Medical image registration using voxel similarity measures*, AAAI Spring Symposium Series: Applications of Computer Vision in Medical Image Processing, 34-37, 1994
15. D.L.G. Hill, C. Studholme, and D.J. Hawkes, *Voxel similarity measures for automated image registration*, Proc. Visualization in Biomedical Computing, 205-216, 1994
16. M.J.D. Powell, *An efficient method for finding the minimum of a function of several variables without calculating derivatives*, Comput. J., 7:155-163, 1964
17. C. Studholme, D.L.G. Hill and D.J. Hawkes, *Multiresolution voxel similarity measures for MR-PET registration*, Proc. Information Processing in Medical Imaging, 287-298, 1995
18. C. Studholme, D.L.G. Hill and D.J. Hawkes, *Automated 3D registration of truncated MR and CT images of the head*, Proc. British Machine Vision Conference (BMVC'96), pp. 27-37, 1996
19. C. Studholme, D.L.G. Hill and D.J. Hawkes, *Automated 3D registration of MR and CT images of the head*, Medical Image Analysis, 1(2), 1996
20. *The Visible Human Project*, WWW http://www.nlm.nih.gov/research/visible/visible_human.html, 1996
21. W.M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, *Multimodal volume registration by maximization of mutual information*, Medical Image Analysis, 1996
22. J. West, J.M. Fitzpatrick, et al., *Comparison and evaluation of retrospective intermodality image registration techniques*, Proc. SPIE Vol. 2710: Medical Imaging - Image Processing, 1996
23. R.P. Woods, S.R. Cherry, and J.C. Mazziotta, *Rapid automated algorithm for aligning and reslicing PET images*, Journal of Computer Assisted Tomography, 16(4):620-633, 1992
24. R.P. Woods, J.C. Mazziotta, and S.R. Cherry, *MRI-PET registration with automated algorithm*, Journal of Computer Assisted Tomography, 17(4):536-546, 1993

This article was processed using the LaTeX macro package with LLNCS style