

Tina Memo No. 2014-006
Internal, IMI preliminary work

Comparison of Local-rigid and Non-rigid Registration of Diffusion Weighted MRI for Improved Estimation of ADC.

Hossein Ragheb, Neil A. Thacker, Jean-Marie Guyader, Stefan Klein and Alan Jackson

Last updated
20 / 10 / 2014



Centre for Imaging Sciences,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT;
and Biomedical Imaging Group Rotterdam,
Departments of Medical Informatics and Radiology,
Erasmus MC - University Medical Centre Rotterdam.

Comparison of Local-rigid and Non-rigid Registration of Diffusion Weighted MRI for Improved Estimation of ADC.

Abstract

Our aim is to develop methodologies for the determination of ADC, suitable for use in patient management and drug trials. In this paper we demonstrate improvements in ADC measurement which at the same time maintain the biological structures essential for analysis of tissue heterogeneity. Two motion correction methods are compared, one based upon non-rigid registration using freely available open source algorithms and the other a purpose designed local rigid registration. Performance of these methods is evaluated using metrics computed from regional ADC histograms. While the non-rigid registration method has the advantages of being applicable on the whole volume and in a fully automatic fashion, the local rigid registration method is much faster and also provides advantages with regard to data smoothness by avoiding interpolation and sub-sampling. Our study also shows that the averaging commonly applied to DW-MR images as part of the acquisition protocol should be avoided if at all possible.

1 Introduction

The IMI QuIC-ConCePT project is a multi-centre proof of concept study to investigate and develop methodologies that support measurement of clinical bio-markers [9]. Diffusion weighted (DW) magnetic resonance (MR) imaging [19] has recently become a focus of attention for facilitating the diagnosis and treatment of tumours [1, 20]. Specifically, to monitor the effect of new drugs administered to treat cancerous organs, DW-MR imaging is an attractive non-invasive method compared to other alternatives including histology or open surgery. There is potential to use this technique either in drug trials or for oncology decision support. However, there are many difficulties associated with the analysis of DW-MR images. Recommendations have been made for tackling these problems, in order to ensure the validity of any corresponding conclusions [2]. In particular, diffusion weighted MR data are acquired over tens of minutes, during which respiratory motion causes organs such as the liver to move. One requirement is that motion is prevented during the acquisition, either using breath holding, triggering or gating [22, 21, 23, 24, 25, 26] or after scanning [17, 27, 28, 29, 30, 31].

The success of this work must ultimately be determined by the ability to make quantitative measurements of ADC for patient management and drug trials. Our current assessment of this requirement has led us to conclude that to monitor the effects of drugs in tumour tissue we need to reliably detect changes in ADC at the level of 10%. This level of performance can be obtained in several ways, including averaging of observed change over many subjects for drug trials. However, for patient management we need to observe these effects in one regional sample. Working back to the original data we conclude that summary values of ADC for a region of interest (ROI) need to have an accuracy of approximately 2 to 3% (for a reliable 2.5 standard deviation difference). Our recent multi-site reproducibility studies for normal liver tissue [3, 7] suggest a repeatability in measurements of individual regions of only 6%. This is a best case scenario, as normal tissue is homogeneous and ADC estimates are relatively unaffected by motion. In more recent work [8] focused on liver tumour regions, we observe that repeated measurement of ADC could vary by more than 10% in 4 out of 19 subjects. We attribute this largely to the image blurring generated by respiratory motion. We therefore need to both stabilise image quality and also improve the accuracy of summary statistics by a factor of 2 to make ADC based patient specific therapy viable. We are looking at several aspects of this problem, including reliability of ROI selection. However, the motion generated in the abdomen offers a specific challenge to meeting this level of reliability.

In this paper we concentrate on the improvements in ADC measurement gained by aligning data via registration. One common way to assess the accuracy of registration algorithms is the use of the Dice overlap. Such a measure evaluates the level of regional agreement between labelled regions in data, but not the accuracy of registration within homogeneous labelled regions. However, we are interested not in the accuracy of labels but instead individual ADC estimates at each voxel location. Ideally we require an assessment of ADC accuracy. One problem here is the absence of ground truth values with which to perform a comparison. Instead, we estimate the apparent reduction in the degree of randomness in ADC estimates by computing summary distribution variables in regions of real data. The apparent improvement in performance can be expressed as the equivalent amount of noise (on average) which has been removed from individual ADC estimates.

Whole volume non rigid alignment (NRA) has been applied in the context of ADC studies in the abdomen and breast in several studies [17, 27]. However, only a small subset of the total volume of data (that immediately

surrounding a ROI) is often required for subsequent analysis. For small amounts of non-rigid movement, a local rigid alignment may be sufficient. In this paper, we therefore study to what extent whole volume image registration compares with local rigid alignment (LRA) of the data in the required ROI's. Compared to NRA, such LRA techniques could reduce execution time and avoid the need to interpolate data at off-lattice locations. This paper also compares these two methods (for off-scanner alignment followed by averaging) with a commonly applied protocol which applies on-scanner averaging (without alignment) of DW-MR images. Details of these protocols and methods are as follows.

2 Materials and Methods

In this section we describe our methods and the data with which we experiment and perform our evaluations and comparisons.

2.1 Data

Data sets of DW-MR images were acquired on a 1.5 T Siemens scanner. Five healthy volunteers were scanned twice within two weeks to test reproducibility of ADC in the liver based on free breathing. Images were acquired using: three b-values 100, 500 and 900 s/mm^2 , field of view $380 \times 332mm$, acquired pixel resolution $3 \times 3mm$, slice thickness $5mm$, number of slices per volume 40, and reconstructed matrix of 256×224 [32]. Two data protocols are available: protocol A and protocol B [32]. For each of the b-values, protocol B provides 12 sets of 3D images (40 image slices per image) which correspond to 4 repeated acquisitions with 3 separate diffusion gradient directions. On the other hand, protocol A data consists of only one set of 3D images per b-value: the image slices are averaged over the 4 acquisitions and 3 gradient directions and produced by the scanner itself, which is a commonly applied protocol [32]. The data is acquired using an interleaving acquisition protocol. This means that each 3D image is reconstructed from 40 slices that are not acquired contiguously. In our data the consequences are consistent with odd slices being acquired first, followed by the even slices (see Figure 2). For a given b-value, simple averaging of the 4 repeats and 3 gradient directions of the acquired protocol B images is expected to generate data equivalent to protocol A. In Fig. 1, we show sample abdominal image slices from the same acquisition corresponding to 3 different gradient directions.

2.2 Comparison of Image Registration Methods

Two image registration methods are developed which might be considered the extremes of mathematical complexity and therefore processing requirement. The intention here is to determine the need for high levels of sophistication and the potential for use of fast methods in a real-time user interface. Both were applied to protocol B images and subsequently compared. The first one is a whole volume registration based on 3D non-rigid deformations, referred to as non-rigid alignment (NRA) technique. The second method aims at aligning a 2D slice of a reference image to those of other images using a local rigid alignment (LRA) technique. The results obtained by these two methods are also compared with those obtained with protocol A data, on which no registration is performed.

2.3 Non-Rigid Alignment (NRA)

The NRA method is a fully automatic image registration pipeline which uses 3D non-rigid deformations. The deformation model is based on B-spline transformations [15] with control points spaced every 64 mm. The first step of the method consists of correcting for motion in each acquired DW-MR image. In the second step, all the DW-MR images of a given scanning session are brought to a common image space. The pipeline is based on elastix [14], a publicly available open source image registration software. The main advantage of the NRA method is that it automatically operates a global registration of the considered 3D volumes. NRA does not require any ROI to be selected for the purpose of image registration. The deformable transformation model used by NRA aims to take into account possible non-rigid misalignments caused by patient and respiratory during image acquisition. The technique consists of 2 steps [16, 17].

2.3.1 Step 1: intra-image registration

Because of respiratory motion, DW-MR images acquired with an odd-even scheme include motion artefacts. These artefacts translate into spatial shifts visible between two neighbouring slices. This step re-establishes the 3D spatial

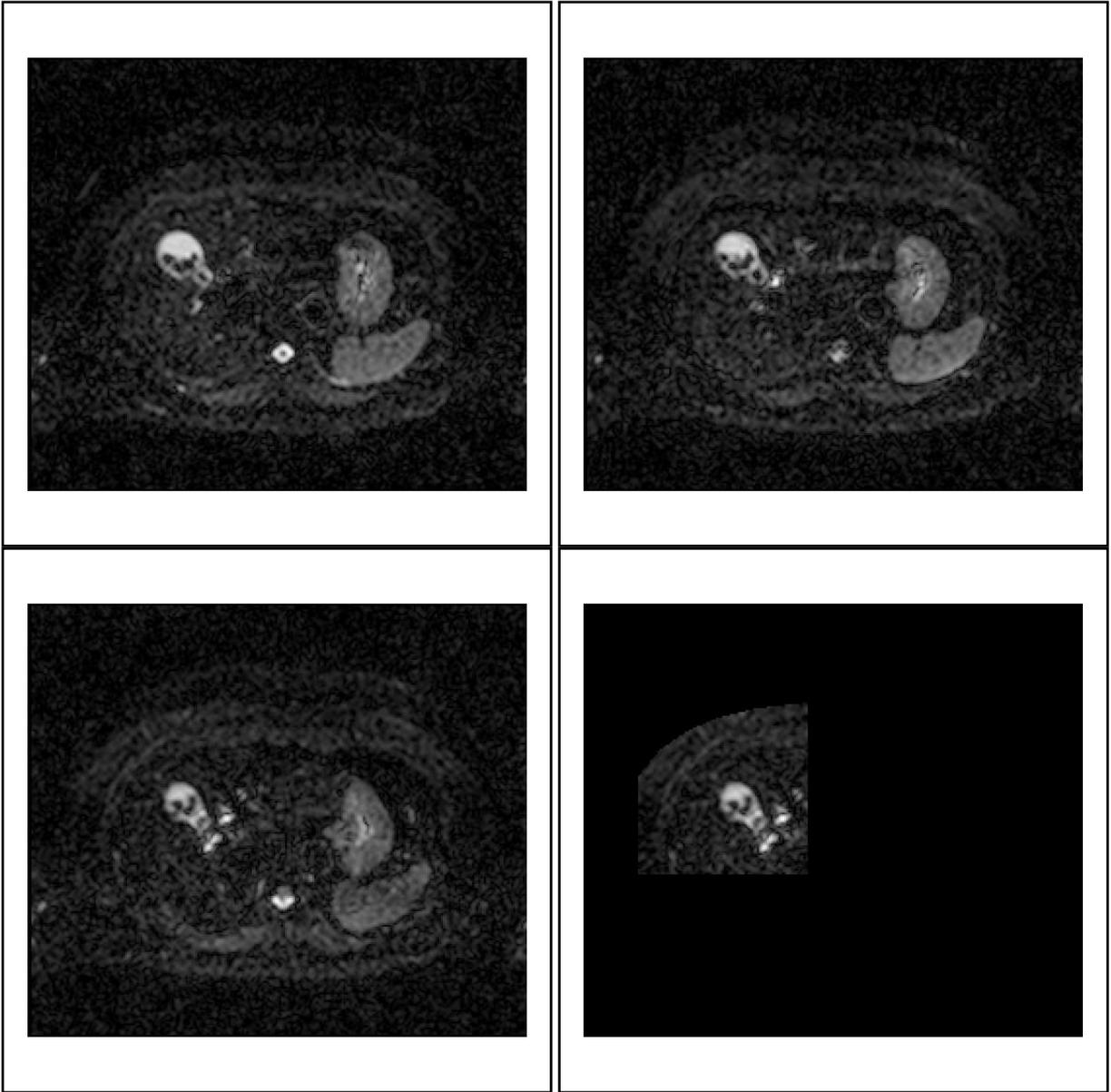


Figure 1: Healthy liver DW-MR image data (v3-20120528-slice24); b-value= $100s/mm^2$; selected image slice from one of the repeat data-sets (out of 4) and diffusion gradients in y (top-left) and z (top-right) directions; diffusion gradient in $-x$ direction which has been used as the reference image slice for local-rigid alignment (bottom-left) and the reference image shown in the selected ROI (bottom-right); this ROI is used both in local-rigid alignment and ADC histogram analysis; gamma adjusted.

alignment within each of the DW-MR images. To that purpose a sub-volume is built from the odd slices of the acquired DW-MRI, centred at their original positions but with a doubled slice thickness. Another sub-volume is extracted from the even slices in a similar fashion. These two sub-volumes are subsequently registered to one another using a group-wise scheme [18] and recombined into a single motion-corrected image.

2.3.2 Step 2: inter-image registration

For given b-value and diffusion gradient direction, the four repeated scans are first registered and brought to a common mid-point space using a group-wise registration. By means of pairwise registration, a spatial correspondence is then found between the mid-point spaces of the 9 couples b-value/diffusion gradient direction. In order to reduce the number of re-sampling steps, the 36 images generated after intra-image registration are transformed to a common image space using a transformation composing the group-wise and pairwise registrations carried out in the inter-image registration step.

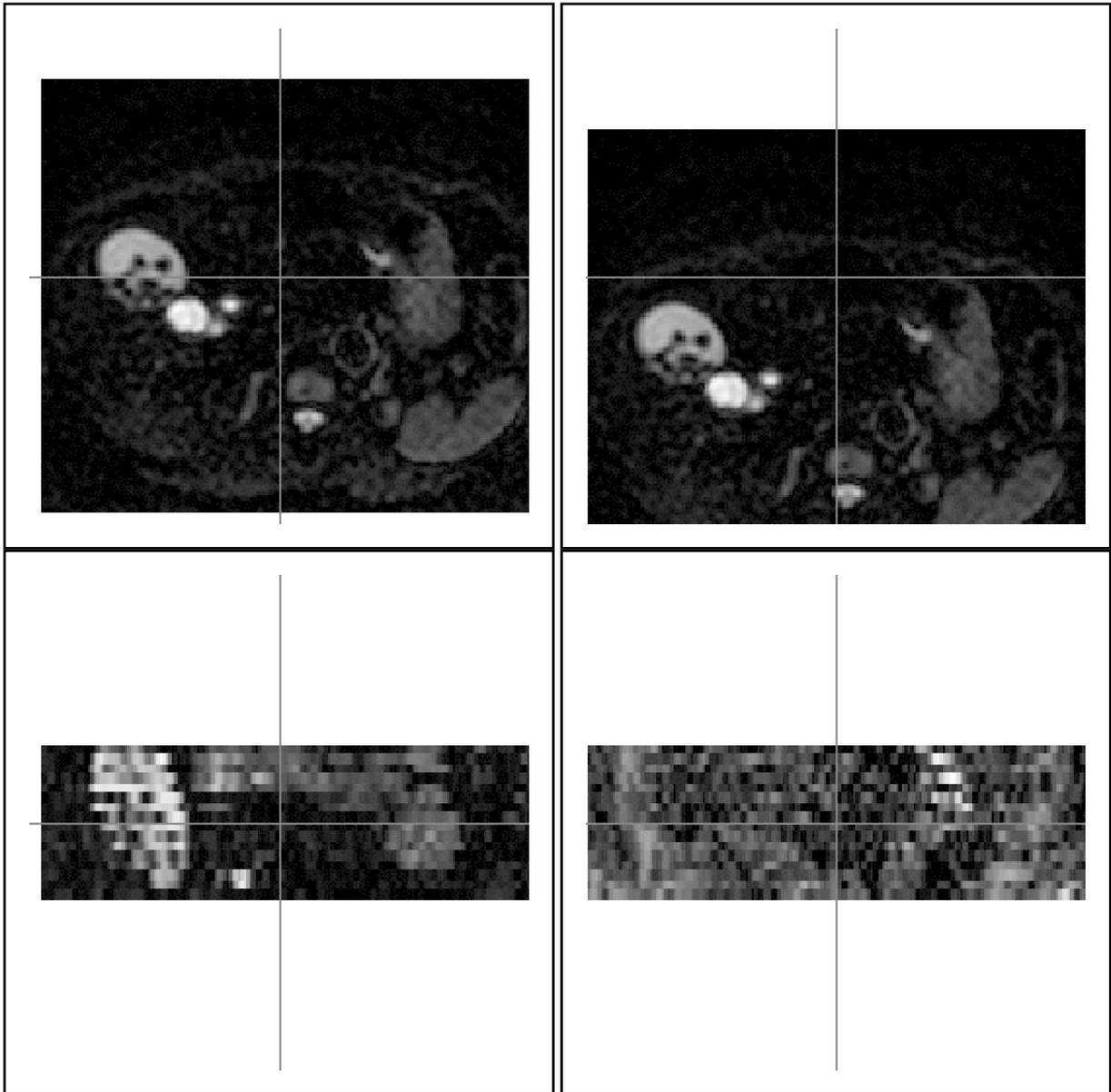


Figure 2: Healthy liver DW-MR image data (v3-20120528) showing the odd-even effect (in the coronal view) as a result of an interleaving acquisition protocol; $b\text{-value}=100s/mm^2$; image volume selected is one of the repeat data-sets (out of 4) with diffusion gradient in $-x$ direction; the coronal view (bottom row) is the view seen from the cross section of the horizontal line drawn on the transverse (slice) view (top row); the two columns show different cross sections on the same volume.

2.3.3 Need for comparison with a local method

One of the main disadvantages of such a global technique is its computational cost. On a single computer, the typical processing time for one data-set is currently 5.5 hours. The most suitable application for NRA is therefore offline. Another drawback of the NRA technique is that 3D global image registration implies re-sampling the registered images. This re-sampling involves interpolating the images, which is done twice in the case of NRA: the first for intra-image registration, the second for inter-image registration. This explains our interest to study to what extent the results obtained with NRA compare with those obtained using a local alignment method, both in terms of ADC values and execution time.

2.4 Local Rigid Alignment (LRA)

When developing our local rigid alignment (LRA) method we are interested in studying ADC metrics corresponding to a single region in a specific slice of the image. Our LRA method aligns the data corresponding to one slice location, but can potentially also be used for 3D ROIs by aligning adjacent slices separately.

Our method involves alignment of a single reference slice to all nearby repeat acquisitions of slice data. The reference slice is taken from the lowest b-value image (in our case, $b = 100s/mm^2$), which has the best signal to noise characteristics. Otherwise, it can be taken from any of the 3 gradient directions and any of the 4 repeats. The region on which matching is performed is a region covering the tissue on the left-hand side of the abdomen (including the liver) and above the level of spinal cord where motion occurs. This area contains various boundaries which provide the information needed for matching. We observe that data away from the spine and around the edge of the body generally appears to move in one (outward) direction. This area is manually identified for the reference slice but as this needs only be done approximately we believe the process could be automated if necessary. Alignment then works as follows.

Using observations of respiratory motion in typical data-sets we were able to set limits on the amount of motion present in the region of the liver. We expect no more than 3 pixels movements ($\approx 1cm$) in vertical and horizontal directions on each image slice and 2 slices along the z axis. There is little or no observable rotation (less than a pixel at the edge of a ROI), so that movement can be modelled on the basis of shifts of the origin. For each gradient direction there are 4 repeat ‘volume’ acquisitions. We allow 5 adjacent slices per repeat data as candidates for alignment, the central slice and two slices from either side. Hence for each gradient direction there are 20 candidate slices from which 4 slices are selected with best matching scores as follows. Specifically, for each of the 5 slices, the best matching score is found out of the $7 \times 7 = 49$ possible cases of shifted slices. Out of these 5 adjacent slices (and in accordance with the observed odd-even acquisition described in the introduction), we keep one shifted slice from the odd slices and one from the even slices with the best matching scores. This gives 8 candidate shifted slices for each gradient direction. We then keep only 4 of these which correspond to the best matching scores. Once 4 slices are selected from each of the 3 gradient directions, these 12 slices are combined to give the average aligned slice for a specific b-value. This process is repeated for other b-values using the same reference image slice as previously (in this case from the $b = 100s/mm^2$ image acquisition). At the end of this process 3 b-value slices are generated which can be analysed to extract voxel based ADC estimates in the same way as for protocol A.

2.4.1 Template matching

The standard literature suggests several metrics for use in image alignment. In our experience the use of small image regions mitigates against the use of cost functions such as mutual information [10, 12, 13, 11]. For the current data roughly one third of the rectangular area around the body (after exclusion of air) would suffice (see Figure 1 for example). Instead we develop our own matching algorithm for a reference slice against a target slice based upon conventional statistics. Applying the variational principle to the problem of matching two scaled noisy image patches I and J (with N pixels and standard deviation of image noise σ), one can define the optimisation function [4]

$$\chi^2 \propto \sum_n^N (\alpha I_n - \beta J_n)^2 \quad s.t. \quad \alpha^2 + \beta^2 = 1 \quad (1)$$

where χ^2 is proportional (and not equal) to the term shown on the right-hand side (which theoretically suffice for optimisation). However, we can explicitly derive formulations and rewrite the cost function in form of detailed equations as follows. Specifically, rather than using two scale factors α and β , one may use a single scale factor $\gamma = \alpha/\beta$. For similar patches it can be shown that $\gamma = \sqrt{B/A}$ where

$$A = \sum_n^N I_n^2 \quad , \quad B = \sum_n^N J_n^2 \quad \text{and} \quad C = \sum_n^N I_n J_n \quad (2)$$

To avoid lengthy execution times, one may expand the patch similarity measure and write

$$\chi^2 = \frac{\sum_n^N (\gamma I_n - J_n)^2}{\sigma^2(1 + \gamma^2)} = \frac{\gamma^2 \sum_n^N I_n + \sum_n^N J_n^2 - 2\gamma \sum_n^N I_n J_n}{\sigma^2(1 + \gamma^2)} \quad (3)$$

$$\chi^2 = \frac{\gamma^2 A + B - 2\gamma C}{\sigma^2(1 + \gamma^2)} = \frac{2(B - \gamma C)}{\sigma^2(1 + \gamma^2)} \quad (4)$$

When choosing the reference image patch to be J , it follows that B is the constant term while A is varying as the target image patch is translated. Also σ has a fixed value during optimisation. Hence minimising χ^2 in Eq. (4) is equivalent to minimising

$$\chi^2 = \frac{-2\gamma C}{\sigma^2(1 + \gamma^2)} \quad (5)$$

The $\gamma/(1 + \gamma^2)$ term can be eliminated during optimisation in accordance with the general proof of convergence of the Expectation Maximisation (EM) algorithm [33], so that optimising the term C is equivalent to optimising the χ^2 . The term C works well when used in a conventional optimiser for alignment of similar images. However, here rather than using grey-level image patches and least-square differences directly, their gradients (based on differential kernels) in the horizontal and vertical directions are used. Previous work [35, 34] in medical image analysis and machine vision has demonstrated that this reduces the dependency upon absolute scaling of the data and is more suitable for matching with MRI and CT datasets, though at the expense of reducing the capture range of the cost-function. In this case, C can be interpreted as the summation of the ‘dot products’ of the two-component gradient vectors originated from the reference and target image patches.

As we do not need to consider sub-pixel shifting, we exhaustively compute the cost function for the limited number of possible variations and find the values which give the best matching score to avoid local optima. Among our candidate patches there will be sub-pixel shifted versions of the region (due to the anatomical motion). As we select only the best matching images from those available we therefore recover some amount of sub-pixel precision whilst avoiding interpolation smoothing.

2.4.2 Constrained movement space

As mentioned above, the movement space defined for our LRA method for the test data-sets studied covers a 3D space with $7 \times 7 \times 5 = 245$ possible movement combinations. When inspecting the shift patterns used to align different b-value images, we observed that b-value images with good image quality (high signal-to-noise ratio) exhibit similar movement patterns, consistent with repeated underlying respiratory motion. This movement pattern could be seen to closely approximate a line in 3D. However, this correlation in estimated movement became less pronounced in poor SNR data. As all datasets are acquired with the same underlying cyclic respiratory motion we can use the correlation seen in good SNR data to improve the estimation of motion in lower SNR data.

For the reference b-value, LRA works as previously, searching the whole movement space for best shift combinations in order to match the data. The shift combinations (as 3D points) for the 12 selected image slices are used to obtain a robust estimate of the line corresponding to subject specific motion. For the non-reference b-values, we then only allow a subset of the original movement space lying close to this line to be searched. For our data we allow all immediate neighbouring pixels of the line to be included. Constraining movement not only results in speeding up the alignment process but also enforces consistent motion between b-value image data.

2.5 Fits of ADC Parameters

For an image slice location given, at each image pixel we fit the signal values from the corresponding b-value image slices to an exponential curve. The decay parameter of the resulting fit is the ADC for that pixel [2]. Specifically in clinical data, as the noise distribution is skewed, we find that a first order Rician correction factor improves the quality of fit [5].

We estimate ADC D through a likelihood-based parameter optimisation $\log P(I|D, S_0)$ (the probability of the image data given the assumed parameters).

$$\log P(I|D, S_0) = -\frac{1}{2\sigma^2} \sum_{b_k} [I(b_k) - f(b_k, D, S_0)]^2 + \text{const} \quad (6)$$

where $f(b_k, D, S_0)$ is the theoretical value of the Rician corrected exponential function and $I(b_k)$ is the signal value from the b-value image pixel. f is a function of b-value b_k and the current estimate of ADC D and no-diffusion signal S_0 (at $b_k = 0$), and is computed using

$$f^2(b_k, D, S_0) = S_0^2 \exp(-2b_k D) + \alpha \sigma^2 \quad ; \quad k \in [1, 2, 3] \quad (7)$$

where $k \in [1, 2, 3]$ refers to the three b-values b_k used, e.g. 100, 500 and 900 s/mm^2 . The signal value for no diffusion S_0 ($b = 0s/mm^2$) is the second parameter which is estimated. α is a fixed value defining the amount of Rician correction applied and it may be adjusted depending on the amount of image smoothing (for our data α

was set to the theoretical value of one). An estimate of the standard deviation of noise in the image σ is computed from the distribution of second derivatives (for x and y) around zero [6], in a central rectangular region on the tissue. Our ADC measurement software has been implemented on the QuIC-ConCePT platform and has become available to all sites within the project to use.

2.6 Assessment of Registration Performance

Our first approach to assessing the quality of ADC data is to observe joint distributions of the ADC parameters arising from fits. Scatter plots of the estimated parameters S_0 and D should show groups of data which correlate with specific biological structures. Plots which allow identification of regions of characteristic behaviour, i.e. clusters, are expected to be indicative of good quality data. This is a qualitative assessment only.

Although several quantitative evaluation methods were considered (including average goodness of fit, measurement reproducibility and entropy based assessments of information), the most sensitive and reliable approach was found to be based upon measurement of the standard deviation of ADC estimates. For this approach, histograms of computed ADC values from within the re-aligned regions are used directly to estimate the amount of random noise removed from data. We make the assumption that the effects of noise, genuine biological variation and motion are all independent. Then the observed standard deviation σ_1 of ADC values can be assumed to be describable by the quadrature addition of such terms. Adjusting one of these (by for example aligning the data) and measuring a new σ_2 while taking steps to control the other factors, should allow the change in this contribution $\Delta\sigma$ to be estimated via quadrature subtraction.

$$\Delta\sigma = \frac{\sigma_1^2 - \sigma_2^2}{|\sigma_1^2 - \sigma_2^2|} \sqrt{|\sigma_1^2 - \sigma_2^2|} \quad (8)$$

For equivalent region of data the biological variation is fixed, whilst controlling for noise involves applying levels of blurring which result in equivalent levels of image smoothness (high spatial frequency image noise).

2.7 Comparison of Fit Results Obtained with Protocol A and B

We first compare the results obtained from protocol B data using NRA and LRA. It has already been stated that NRA is expected to generate smoother aligned images compared to LRA because of the interpolations performed during image re-sampling. To reach a similar degree of blurring as in NRA, the DW-MR aligned images generated using LRA are smoothed via Gaussian blurring (standard deviation 1 pixel). This case is referred to in the tables as LRA-blur.

We also investigate how the results obtained without registration with protocol A (PA) compare with the case in which protocol B images are averaged on the repeated acquisitions and diffusion gradient directions, for a given b -value. This case is referred to as AVG. To compare with the NRA results, the AVG images are also smoothed via Gaussian blurring. This is referred to as AVG-blur.

3 Results

All results (figures and tables) reported in this paper correspond to single slice region of interests (2D ROI's) selected for individual volunteer scans. These 2D ROI's are used here both for template matching using LRA and for generating each ADC histogram using individual measurements. Mean and median ADC metrics and the width of ADC distribution¹ are then extracted from these histograms. In Fig. 3, we show, for a selected data-set and image slice location, the protocol A image slice (top row, left) and the average image slice obtained when applying alternative methods on the corresponding protocol B data. These methods are: non-rigid alignment NRA (top row, right), simple averaging AVG and AVG-blur (middle row), and finally, LRA and LRA-blur (bottom row). LRA and LRA-blur seem the best in terms of recovering the genuine fine structures between the stomach (and partially pancreas) on one side and liver and gallbladder on the other. There is also fat contained between boundaries of these organs. This is obviously more important if one requires to extract ADC metrics from small ROI's in the proximity of these fine structures.

In Fig. 4, we show scatter plots of S_0 signal against ADC obtained by fitting a mono-exponential to the 3 b -value images obtained using different methods.

In Table 1, we tabulate the width of ADC histogram σ_D for each data-set. Here we also tabulate the percentage difference in the width of ADC histogram (shown in brackets) $\Delta\sigma_D = \sqrt{|\sigma_{D_1}^2 - \sigma_{D_2}^2|}$ (including the sign as in

¹Width of ADC histogram is equivalent to the standard deviation of all fitted ADC values in the corresponding ROI.

Eq. 8) between results obtained using AVG and those obtained by applying each of the alternative methods to protocol-B data. At the bottom of the table, we compute an average value and an accuracy (error on standard deviation) for $\Delta\sigma_D$ values. The largest modification to simple averaging, which is equivalent to standard deviation of the noise removed, is obtained using the LRA and its blurred version followed by NRA and blurred version of simple averaging.

data-set	PA	AVG	LRA ($\Delta\sigma_D$)	NRA	LRA-blur	AVG-blur
v1-20120503	27.25	30.62	27.54 (13.38)	22.36 (20.91)	23.78 (19.28)	26.58 (15.20)
v1-20120515	31.05	28.85	23.58 (16.62)	20.53 (20.26)	19.31 (21.43)	24.69 (14.92)
v2-20120521	54.25	55.09	44.01 (33.13)	63.29 (-31.15)	42.00 (35.64)	53.69 (12.34)
v2-20120528	62.04	56.12	48.15 (28.82)	61.49 (-25.13)	45.62 (32.68)	54.24 (14.40)
v3-20120524	56.73	67.85	65.79 (16.59)	67.59 (5.93)	62.94 (25.34)	64.72 (20.37)
v3-20120528	71.82	70.71	58.64 (39.51)	64.25 (29.52)	57.18 (41.59)	68.55 (17.34)
v4-20120611	32.07	32.92	30.54 (12.28)	25.99 (20.20)	26.79 (19.13)	28.15 (17.06)
v4-20120614	38.30	33.94	33.85 (2.47)	28.27 (18.78)	29.53 (16.73)	29.61 (16.58)
v5-20120731	32.40	32.68	33.60 (-7.80)	29.41 (14.24)	29.81 (13.39)	28.88 (15.29)
v5-20120809	32.59	31.06	34.34 (-14.64)	28.94 (11.27)	30.09 (7.70)	27.01 (15.33)
mean±accuracy			(14.03±4.17)	(8.48±4.92)	(23.29±2.55)	(15.88±0.51)

Table 1: Width of ADC histogram σ_D and difference in the width of ADC histogram (in brackets) $\Delta\sigma_D$ (Eq. 8) between results obtained using AVG and those obtained using each of the alternative methods on protocol-B data; same units as ADC $10^{-5} \text{ mm}^2/\text{s}$; from ADC fits for protocol-A and for the average slice using protocol-B with and without motion correction (computed in the tissue region used for motion correction); the bottom row gives the mean of $\Delta\sigma_D$ values for each column and its corresponding accuracy.

data-set	PA	AVG	LRA	NRA	LRA-blur	AVG-blur
v1-20120503	113.90	113.67	108.35	109.27	107.53	112.64
v1-20120515	127.66	105.36	105.51	108.91	105.45	104.18
v2-20120521	119.22	111.43	99.45	109.45	102.19	111.32
v2-20120528	129.53	105.74	102.61	105.95	99.52	105.50
v3-20120524	115.24	100.60	93.38	95.67	92.39	98.51
v3-20120528	104.82	97.54	89.86	92.35	88.96	96.31
v4-20120611	129.42	101.63	92.54	94.47	91.65	100.98
v4-20120614	129.89	99.30	95.23	92.88	93.74	97.72
v5-20120731	104.68	78.72	73.65	76.21	71.58	76.70
v5-20120809	105.81	77.57	76.57	76.65	75.25	75.77

Table 2: ADC mean values of the two scans for each volunteer for protocol-A and for the average slice using protocol-B with and without motion correction (computed in the tissue region used for motion correction); ADC units $10^{-5} \text{ mm}^2/\text{s}$.

In Table 2, we tabulate the mean ADC measurement for each data-set. Using numbers given in Table 2, we compute in Table 3 the percentage change in ADC mean measurements $\Delta D\% = \frac{2|D_1 - D_2|}{(D_1 + D_2)} \times 100$ for each volunteer. At the bottom of the table, we compute the mean and unbiased standard deviation of these measurements for each method. In terms of absolute mean value, NRA gives the best reproducibility followed by LRA and its blurred version. However, these mean values are quite close within their corresponding standard deviations.

In Table 4, we tabulate the accuracy of mean ADC measurements (σ_D/\sqrt{N}), where σ_D is the width of ADC histogram and N is the number of histogram entries. It is clear from this table that mean (or median) ADC values are quite accurate (around one percent). These errors play no major role in reproducibility errors where more important sources of variability dominate.

In Tables 5 and 6, we investigate the reproducibility of ADC median measurements extracted from different ADC histograms. The pattern found here is consistent with that found in Tables 2 and 3 where mean ADC measurements were used. Note that similar to Table 3, the unbiased standard deviations are computed in Table 6. Here, these unbiased standard deviations can be interpreted as variability figures within the population (not accuracy as in Table 1).

data-set	PA	AVG	LRA	NRA	LRA-blur	AVG-blur
v1-20120503, v1-20120515	11.39	7.58	2.65	0.33	1.95	7.80
v2-20120521, v2-20120528	8.28	5.24	3.12	3.24	2.64	5.36
v3-20120524, v3-20120528	9.47	3.08	3.84	3.53	3.78	2.25
v4-20120611, v4-20120614	0.36	2.31	2.86	1.69	2.25	3.28
v5-20120731, v5-20120809	1.07	1.47	3.88	0.57	4.99	1.21
mean [standard deviation]	6.11 [5.64]	3.93 [2.76]	3.27 [0.62]	1.87 [1.64]	3.12 [1.39]	3.98 [2.92]

Table 3: Percentage ADC change $\Delta D\% = \frac{2|D_1-D_2|}{(D_1+D_2)} \times 100$ from ADC mean values for each volunteer (from Table 2) for protocol-A and for the average slice using protocol-B with and without motion correction (computed in the tissue region used for motion correction); the bottom row gives the mean and unbiased standard deviation (variability) for each column.

data-set	PA	AVG	LRA	NRA	LRA-blur	AVG-blur
v1-20120503	0.4257	0.4800	0.4322	0.3485	0.3712	0.4146
v1-20120515	0.5196	0.4861	0.3964	0.3441	0.3240	0.4134
v2-20120521	0.6221	0.6337	0.5050	0.7241	0.4789	0.6132
v2-20120528	0.6996	0.6361	0.5449	0.6924	0.5125	0.6091
v3-20120524	0.7188	0.8825	0.8562	0.8533	0.8054	0.8273
v3-20120528	0.9050	0.9062	0.7541	0.8030	0.7225	0.8625
v4-20120611	0.4171	0.4292	0.4005	0.3379	0.3493	0.3660
v4-20120614	0.4681	0.4172	0.4169	0.3448	0.3605	0.3609
v5-20120731	0.3462	0.3568	0.3653	0.3141	0.3185	0.3108
v5-20120809	0.3825	0.3701	0.4088	0.3388	0.3537	0.3176
mean [st dev]	0.551±0.180	0.560±0.201	0.508±0.167	0.510±0.226	0.460±0.177	0.510±0.206

Table 4: Mean ADC measurement accuracy (σ_D/\sqrt{N}) from ADC fits for protocol-A and for the average slice using protocol-B with and without motion correction (computed in the tissue region used for motion correction); σ_D is the width of ADC histogram and N is the number of histogram entries; units for the corresponding mean ADC values $10^{-5} \text{ mm}^2/\text{s}$; the bottom row gives the mean and standard deviation of each column.

data-set	PA	AVG	LRA	NRA	LRA-blur	AVG-blur
v1-20120503	109.42	112.94	107.51	107.87	106.99	111.56
v1-20120515	120.64	103.81	104.01	107.62	103.71	103.20
v2-20120521	104.43	101.76	96.47	95.58	95.63	100.40
v2-20120528	115.69	95.87	91.47	93.59	91.31	94.93
v3-20120524	110.26	84.49	77.53	77.78	76.86	82.75
v3-20120528	93.19	82.02	77.91	75.34	78.93	82.49
v4-20120611	128.76	100.64	92.26	94.00	91.52	99.05
v4-20120614	127.98	99.44	94.19	91.64	93.33	97.37
v5-20120731	106.10	79.30	73.20	76.27	71.54	77.59
v5-20120809	105.85	77.34	75.45	75.94	73.70	76.23

Table 5: ADC median values of the two scans for each volunteer for protocol-A and for the average slice using protocol-B with and without motion correction (computed in the tissue region used for motion correction); ADC units $10^{-5} \text{ mm}^2/\text{s}$.

4 Discussions

In Fig. 4, the scatter plots for fitted ADC parameters show how the local regional correlations in parameters improve following registration. In particular, LRA and its smoothed counterpart (LRA-blur) show much better clustering of the parameters from localised structures. The non-rigid alignment (NRA) technique shows an overall distribution of the fitted parameters which matches the general variation seen for LRA, while simple averaged (AVG) and smoothed results (AVG-blur) show poor clustering and a wider variability of parameters. This is interpreted as due to measurement bias introduced by volume averaging. Based on $b = 100\text{s}/\text{mm}^2$ image slices shown in Fig. 3, we can argue that the aligned image obtained using LRA is a much better image to draw ROI's

data-set	PA	AVG	LRA	NRA	LRA-blur	AVG-blur
v1-20120503, v1-20120515	9.75	8.42	3.30	0.23	3.11	7.78
v2-20120521, v2-20120528	10.23	5.96	5.32	2.10	4.62	5.60
v3-20120524, v3-20120528	16.78	2.96	0.40	3.18	2.65	0.31
v4-20120611, v4-20120614	0.60	1.19	2.07	2.54	1.95	1.71
v5-20120731, v5-20120809	0.23	2.50	3.02	0.43	2.97	1.76
mean [standard deviation]	7.51 [7.88]	4.20 [3.27]	2.82 [2.01]	1.69 [1.46]	3.06 [1.09]	3.43 [3.49]

Table 6: Percentage ADC change $\Delta D\% = \frac{2|D_1 - D_2|}{(D_1 + D_2)} \times 100$ from ADC median values for each volunteer (from Table 4) for protocol-A and for the average slice using protocol-B with and without motion correction (computed in the tissue region used for motion correction); the bottom row gives the mean and unbiased standard deviation (variability) for each column.

compared to that obtained using PA or NRA.

The results for mean ADC (Table 2) and deviations (Table 1) obtained from protocol B data following simple averaging (AVG) are observed to be consistent with those seen for protocol A (PA). Tables of regional ADC deviations allow the calculation of the equivalent amount of random noise which would need to be added to data in order to obtain measurements as poor as the simple average (bracketed numbers in Table 1). The main contribution to variance seen in this data is biological. The amounts of noise removed by registration and the difference between the alternative approaches can be compared directly in terms of numbers of standard deviations of measured change, and are all statistically significant. As the percentage ADC change ($\Delta D\%$) for individual fits in regions of good signal is often of the order of 5%, the implication of these numbers is that motion is the dominant source of instability in ADC estimation. Comparison of the LRA-blur and NRA results implies that a local rigid alignment of slice data is at least as good as the more sophisticated volume based non-rigid alignment. Nevertheless, NRA causes a significant amount of unwanted image smoothing caused by the sub-sampling required to overcome odd-even slice misalignment and registration errors. The figures for measurement repeatability (Table 3) show a slight improvement in performance for NRA over LRA. However, this may in part be due to a reduced spread in measured ADC values caused by averaging. We would argue that given the choice of these methods the LRA (non-smoothed) would be most suitable for assessment of ADC heterogeneity. The accuracies obtained for the selected regions are on average consistent with the target of 2-3% required for reliable detection of a 10% change. It should be remembered that this is for a region in a normal subject and more work is therefore required to place this in a clinical context.

With regard to computation cost, the typical processing time for the NRA method is currently 8.5 hours when applied to all data-sets corresponding to 3 b-values generating 120 (3×40) aligned slices ready for ADC measurement. This figure for the LRA method is 46 seconds for 3 b-values generating 3 aligned slices ready for ADC measurement. For both methods, source code may be slightly optimised and implementation could be parallelised to reduce execution time, (an order of magnitude improvement is possible) but as the same kinds of gains are available to both methods the ratio between these current figures would not be expected to change significantly.

The reproducibility figures against methods in Tables 3 and 6 (based on mean ADC values of Table 2 and median ADC values of Table 5) are consistent. These are percentage changes in ADC metrics from the two scans of the same volunteer. At the bottom of each table, an average reproducibility figure is computed. Also a standard deviation (in brackets) for between volunteer variability is provided. As the number of volunteers is small, this variability figure is poor for any conclusions. However, in this paper we are more interested in the average reproducibility figure and the accuracy with which each individual reproducibility figure (per volunteer) has been estimated. Here, Table 4 provides these accuracy figures for individual mean (or median) ADC measurements. Moreover, figures at the bottom of Tables 3 and 6 show that average reproducibility figures corresponding to protocol A for the mean and median ADC (6.11 and 7.51) are significantly larger than those corresponding to NRA (1.87 and 1.69) and LRA (3.27 and 2.82) methods. Hence we can not only tighten the ADC histogram using motion correction but can also achieve better reproducibility.

When it comes to comparing protocols A and B, the results show that the percentage of ADC change is at least two times larger for protocol A than for all the scenarios considered for protocol B. This proves that acquiring and storing individually DW-MR images is crucial for getting reproducible ADC measurements. Note that scan time would be comparable for protocol A and B if the same number of b-values, acquisition repeats and diffusion gradient directions are used.

This motion correction experiment illustrates some basic properties of this task. If the signal to noise in data is poor (as is often the case with clinical data) there may not be enough information in an image to support accurate

determination of a large number of parameters (such as needed for non-rigid alignment). Therefore practical application of registration methods needs an assessment of the required degree of complexity of the model. As in this case, simpler alignment methods may actually prove more suitable than generic ‘state-of the art’ techniques. Getting good results in these tasks may not require a fundamental break-through in the state of the art, but only pragmatic application of an appropriate technique.

5 Conclusions

Local-rigid alignment (LRA) is two orders of magnitude faster than non-rigid alignment (NRA), and so has the potential to be used in real-time in an ADC analysis pipeline just before the ADC measurement task. Moreover, LRA works on a single slice basis while NRA works on the whole volume. This means that LRA can not only resolve the motion correction requirements in a single slice level, it can also be used for larger scale (up to whole volume) motion correction tasks by setting up the required iterative process. Further, LRA results in a greater reduction in the noise generated by motion than NRA.

On the other hand, when it comes to heterogeneous tumours, to study single ADC measurements, region of interest ADC metrics (such as mean or median) or percentage change in ADC metrics, it is recommended to use LRA rather than NRA. This is because LRA avoids blurring image data and maintains individual voxels corresponding to specific tissue locations (collection of cells) during the alignment while NRA is based on image interpolation which results in data blurring, and so changing the definition of final registered voxels which may be difficult to interpret.

Finally, it is recommended that data should be acquired using protocol B (off-scanner averaging) and not protocol A (on-scanner averaging), so that post-acquisition motion compensation can be applied to improve the measurement of ADC change needed for clinical decisions on individual patients.

It is part of our ongoing work to investigate whether the methods discussed in this paper meet the specified reproducibility for detection of change in liver tumours.

References

- [1] B. Turkbey, O. Aras, N. Karabulut, et al, “Diffusion-Weighted MRI for Detecting and Monitoring Cancer: A Review of Current Applications in Body Imaging”, *Diagnostic Interventional Radiology, Turkish Society of Radiology*, 18(1):46-59, 2012.
- [2] A.R. Padhani, G. Liu, D. Mu-Koh, T.L. Chenevert, H.C. Thoeny, T. Takahara, A. Dzik-Jurasz, B.D. Ross, M. Van Cauteren, D. Collins, D.A. Hammoud, G.J.S. Rustin, B. Taouli, and P.L. Choyke, “Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations”, *Neoplasia*, 11(2):102-125, 2009.
- [3] H. Ragheb, N.A. Thacker, D.M. Morris, N.H.M. Douglas, and A. Jackson, “Predicting the Quantitative Accuracy of In-Vivo ADC using an Ice-Water Phantom”, *Proc. Int’l Soc’ Magnetic Resonance in Medicine*, pp. 2650, 2014.
- [4] N.A. Thacker, J.V. Manjon, and P.A. Bromiley, “A Statistical Interpretation of Non-Local Means”, *Proc. VIE 2008*, pp. 250-255, 2008.
- [5] H. Gudbjartsson and S. Patz, “The Rician Distribution of Noisy MRI Data”, *Magnetic Resonance in Medicine*, 34(6):910-4, 1995.
- [6] N.A. Thacker, “Useful Image Processing Methods”, *Tina memo 2008-010*, 2008, <http://www.tina-vision.net/docs/memos/2008-010.pdf>.
- [7] H. Ragheb, N.A. Thacker, D.M. Morris, and A. Jackson, “Interpreting Ice-Water Phantom Data for Prediction of Clinical ADC Measurement”, *Tina memo 2013-005*, 2013, <http://www.tina-vision.net/docs/memos/2013-005.pdf>.
- [8] H. Ragheb, N.A. Thacker, R. Pathak, D.M. Morris, and A. Jackson, “Multi-site Liver Tumour ADC Reproducibility at 1.5 T”, *Tina memo 2014-007*, 2014, <http://www.tina-vision.net/docs/memos/2014-007.pdf>.
- [9] J.C. Waterton and L. Pykkanen, “Qualification of Imaging Biomarkers for Oncology Drug Development”, *European Journal of Cancer*, 48:409-415, 2012.

- [10] P. Viola and W.M. Wells III, "Alignment by maximization of mutual information", *Int. J. Comput. Vis.*, 24(2):137-154, 1997.
- [11] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information", *IEEE Trans. Med. Imag.*, 16(2):187-198, Apr. 1997.
- [12] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration", *IEEE Trans. Image Process.*, 9(12):2083-2099, Dec. 2000.
- [13] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, "Mutual-information- based registration of medical images: A survey", *IEEE Trans. Med. Imag.*, 22(8):986-1004, Aug. 2003.
- [14] S. Klein, M. Staring, K. Murphy, M.A. Viergever, and J.P.W. Pluim, "Elastix: a toolbox for intensitybased medical image registration", *IEEE Trans Med Imaging*, 29:196-205, 2010.
- [15] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images", *IEEE Trans. Medical Imaging*, 18(8):712-721, 1999.
- [16] J.-M. Guyader, L. Bernardin, N.H.M. Douglas, D.H.J. Poot, W.J. Niessen, and S. Klein, "Influence of image registration on ADC images computed from free-breathing diffusion MRIs of the abdomen", *Proc. SPIE 9034*, Medical Imaging, Image Processing, pp.1-6, March 2014.
- [17] J.-M. Guyader, L. Bernardin, N.H.M. Douglas, D.H.J. Poot, W.J. Niessen, and S. Klein, "Influence of Image Registration on Apparent Diffusion Coefficient Images Computed from Free-Breathing Diffusion MR Images of the Abdomen", *submitted to JMRI*, 2014.
- [18] D. Seghers, E.D. Agostino, F. Maes, and D. Vandermeulen, "Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques", *Proc. MICCAI 2004*, Rennes, Brittany, pp. 696-703, 2004.
- [19] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanie, and M. Laval-Jeantet, "MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders", *Radiology*, 161:401-407, 1986.
- [20] R. Sinkus, B.E. Van Beers, V. Vilgrain, N. DeSouza, and J.C. Waterton, "Apparent Diffusion Coefficient from Magnetic Resonance Imaging as a Biomarker in Oncology drug development", *European Journal of Cancer*, 48:425-431, 2012.
- [21] B. Taouli and D.M. Koh, "Diffusion-weighted MR imaging of the liver", *Radiology*, 254:47-66, 2010.
- [22] B. Taouli, A. Sandberg, A. Stemmer, et al. "Diffusion-weighted imaging of the liver: comparison of navigator triggered and breathhold acquisitions", *J Magn Reson Imaging*, 30:561-568, 2009.
- [23] T.C. Kwee, T. Takahara, D. Koh, R.A.J. Nivelstein, and P.R. Luijten, "Comparison and reproducibility of ADC measurements in breathhold, respiratory triggered, and free-breathing diffusion weighted MR imaging of the liver", *J Magn Reson Imaging*, 28:1141-1148, 2008.
- [24] M. Bruegel, K. Holzapfel, J. Gaa, et al. "Characterization of focal liver lesions by ADC measurements using a respiratory triggered diffusion-weighted single-shot echo-planar MR imaging technique", *Eur J Radiol*, 18:477-485, 2008.
- [25] N.E. Larsen, S. Haack, L.P.S. Larsen, E.M. Pedersen, "Quantitative liver ADC measurements using diffusion-weighted MRI at 3 Tesla: evaluation of reproducibility and perfusion dependence using different techniques for respiratory compensation", *MAGMA*, 26:431-442, 2013.
- [26] M.K. Ivancevic, T.C. Kwee, T. Takahara, T. Ogino, H.K. Hussain, P.S. Liu, T.L. Chenevert, "Diffusion-weighted MR imaging of the liver at 3.0 Tesla using tracking only navigator echo (TRON): a feasibility study", *J Magn Reson Imaging*, 30:1027-1033, 2009.
- [27] L.R. Arlinghaus, E.B. Welch, A.B. Chakravarthy, et al. "Motion correction in diffusion-weighted MRI of the breast at 3T", *J Magn Reson Imaging*, 33:693-704, 2011.
- [28] Y. Mazaheri, R.K.G. Do, A. Shukla-Dave, et al. "Motion correction of multi-b-value diffusionweighted imaging in the liver", *Acad Radiol*, 19:1573-1580, 2012.

- [29] S.C. Partridge, R.S. Murthy, A. Ziadloo, S.W. White, K.H. Allison, C.D. Lehman, “Diffusion tensor magnetic resonance imaging of the normal breast”, *Magn Reson Imaging*, 28:320-328, 2010.
- [30] S.C. Partridge, A. Ziadloo, R. Murthy, et al. “Diffusion tensor MRI: preliminary anisotropy measures and mapping of breast tumours”, *J Magn Reson Imaging*, 31:339-347, 2010.
- [31] J. Liau, J. Lee, M.E. Schroeder, C.B. Sirlin, M. Bydder, “Cardiac motion in diffusion-weighted MRI of the liver: artifact and a method of correction”, *J Magn Reson Imaging*, 35:318-27, 2012.
- [32] D. Collins, N.H.M. Hogg, et al., “QuIC-ConCePT WP3 Task 2.3 Reproducibility Study: Healthy Volunteers in Liver”, *IMI QuIC-ConCePT Internal Document*, May 11, 2012.
- [33] A.P. Dempster, N.M. Laird, and D.B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm”, *J. Roy. Statist. Soc. B (Methodological)*, 39:1-38, 1977.
- [34] P.A. Bromiley, A.C. Schunke, H. Ragheb, N.A. Thacker, and D. Tautz, “Semi-automatic Landmark Point Annotation for Geometric Morphometrics” *Frontiers in Zoology*, 11(61), 2014.
- [35] H. Ragheb and N.A. Thacker, “Quantitative Localisation of Manually Defined Landmarks”, *Proc. Medical Image Understanding and Analysis*, London, UK, pp. 221-225, 2011.

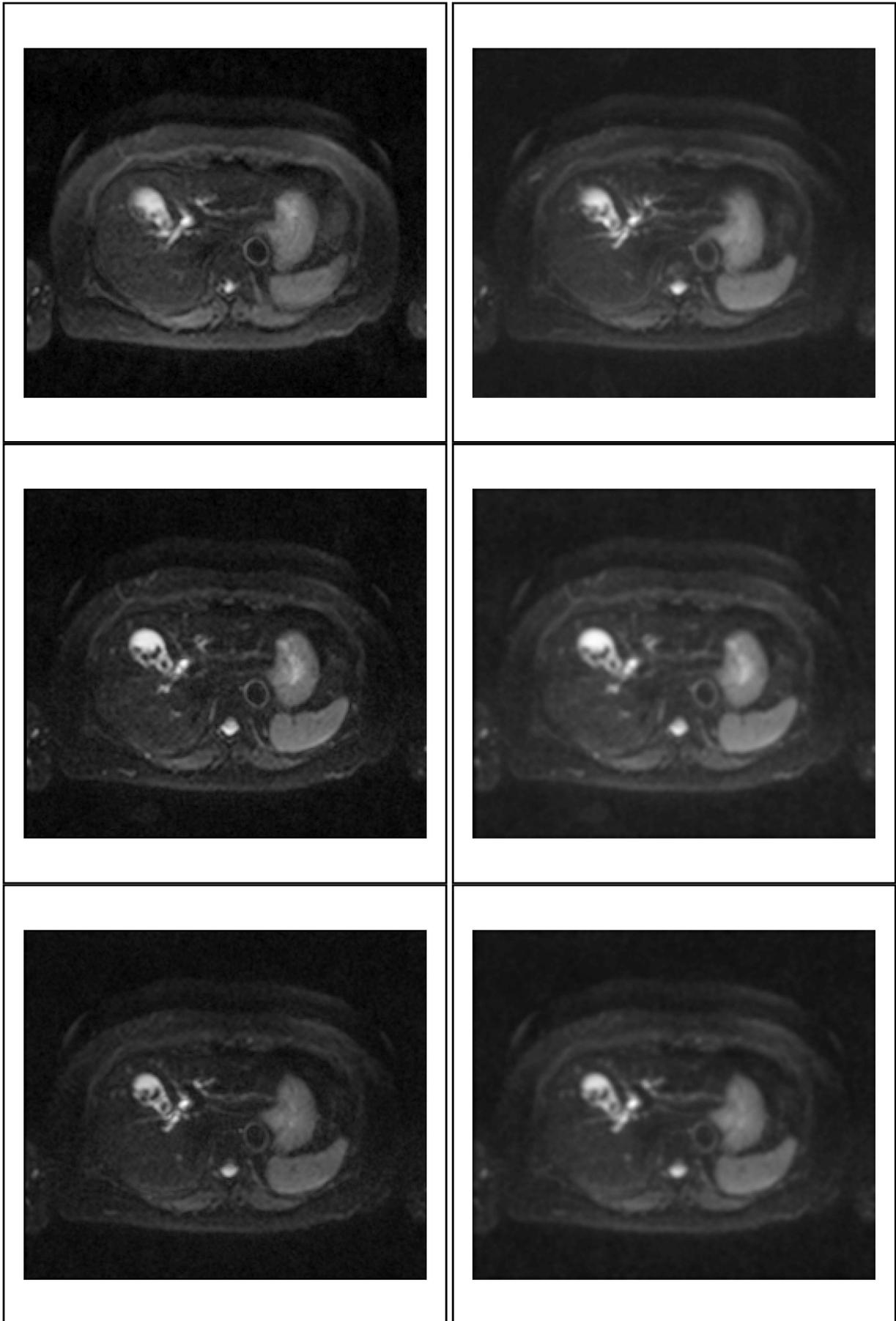


Figure 3: Healthy liver DW-MR image data (v3-20120528-slice24); selected average image slice ($b\text{-value}=100s/mm^2$) from different methods; protocol A (top-left) and NRA method (top-right); simple averaging (middle-left) and the corresponding blurred version (middle-right); LRA method (bottom-left) and the corresponding blurred version (bottom-right); for blurred images 1 pixel Gaussian blurring were applied; gamma adjusted.

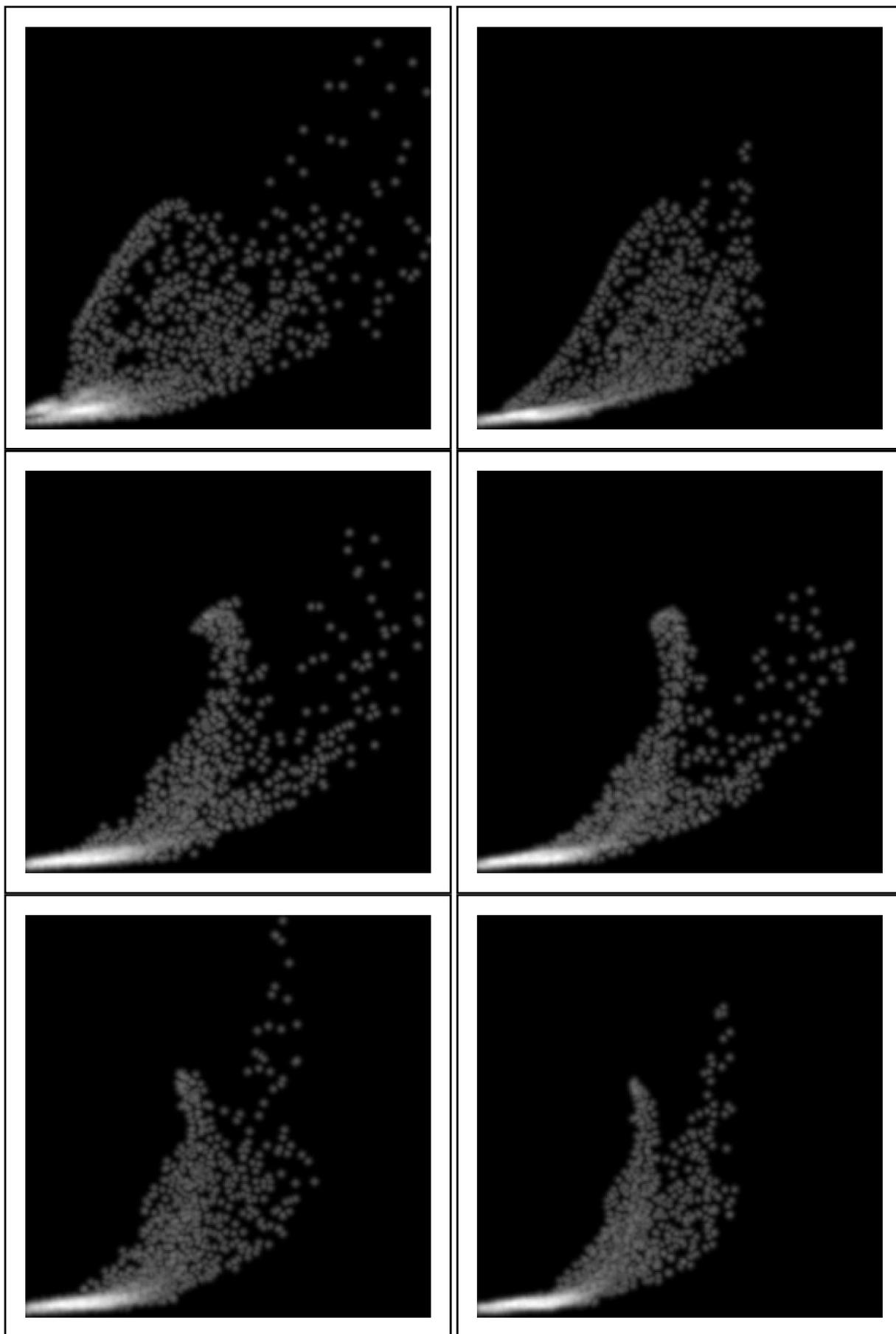


Figure 4: Healthy liver DW-MR image data (v3-20120528-slice24); scatter plot of S_0 signal against ADC for the identical selected 2D ROI (shown in Fig. 1, bottom-right); protocol A (top-left) and NRA method (top-right); simple averaging (middle-left) and the corresponding blurred version (middle-right); LRA method (bottom-left) and the corresponding blurred version (bottom-right).