

ABSTRACT

Title of Document: PROSPECTIVE HEAD MOVEMENT
CORRECTION FOR HIGH-RESOLUTION
MRI USING AN IN-BORE OPTICAL
TRACKING SYSTEM.

Lei Qin, Doctor of Philosophy, 2009

Directed By: Professor Yang Tao, Fischell Department of
Bioengineering

In MRI of the human brain, subject motion is a major cause of magnetic resonance image quality degradation. To compensate the effects of head motion during data acquisition, an in-bore optical motion tracking system is proposed. The system comprises one or two MR compatible infrared cameras that are fixed on a holder right above and in front of the head coil. The resulting close proximity of the cameras to the object allows precise tracking of its movement. During image acquisition, the MRI scanner uses this tracking information to prospectively compensate for head motion by adjusting gradient field direction and RF phase and frequency. Experiments performed on subjects demonstrate the system's robustness, exhibiting an accuracy of better than 0.1mm and 0.15°.

PROSPECTIVE HEAD MOVEMENT CORRECTION FOR HIGH-RESOLUTION
MRI USING AN IN-BORE OPTICAL TRACKING SYSTEM.

By

Lei Qin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Yang Tao, Chair
Professor Rama Chellappa
Dr. Jeff H Duyn
Professor Elliot R McVeigh
Professor Hubert Montas

© Copyright by
Lei Qin
2009

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Jeff H Duyn, who has been an invaluable mentor for the past three and half years at NIH. His insightful guidance and advice have made substantial differences in my work, as well as my life. His confidence in my capability has enabled much of my creative work. His encouragement and understanding have made lots of difficulties seem easier.

I am also deeply grateful to my supervisor, Dr. Yang Tao, for seeing my research potential and pushing me to complete my PhD degree. His wide knowledge and his logical way of thinking have been of great value for me. His understanding and guidance have provided a good basis for the present thesis. His support and encouragement have been invaluable for the past four years.

I would also like to thank Dr. Rama Chellappa, Dr. Elliot R McVeigh and Dr. Hubert Montas for serving on my committee and for their time, support, and helpful advices during my research and academic development.

I also owe my most sincere gratitude to Dr. Alan Koretsky, for his guidance, encouragement, understanding and generous help in both professional and personal matters. He made me believe in myself when I was in doubt, showed me the right path when I was lost. I cannot thank him enough for all he did for me.

I am also grateful to Dr. John Andrew Derbyshire from NHLBI for his help on my understanding of MR physics and sequence programming. I will never forget the long hours he spent with me on the scanner. He gave me tremendous inspiration in my work and gave me help whenever I needed it.

I enjoyed every single day I worked in the Advance MRI lab. Every one in the lab is nice, generous and helpful to me. I want to thank Dr. Peter van Gelderen, Dr. Jacco de Zwart, Dr. Jonghoo Lee, Dr. Karin Shmueli, Dr. Bing Yao, Dr. Masaki Fukunaga, Dr. Marta Bianciardi, Dr. Shuming Wang, Dr. Joe Murphy-Boesch , Dr. Tieqiang Li and Susan Fulton. Special thanks go to Peter and Tieqiang for their guidance during my first steps into MR pulse sequence programming; and to Jacco for his help on my numerous computer problems. Only because of the help from all of you can I make this thesis possible.

I would also like to express my appreciation to all present and former members of the Bio-imaging and Machine Vision Laboratory at University of Maryland for their constant suggestions and friendship, including Bin Zhu, Lu Jiang, Xin Chen, Hansong Jing, Xuemei Chen, Abby Vogel, Angela Vargas, and Gary Seibel.

I would like to extend my deepest gratitude to my parents, who have always believed in me and knew I could accomplish the goal of being the first PhD in my family. And to my husband Dr. Fenghua Jin, thanks for the enduring love, encouragement, inspiration and support through all these years. His presence with me has made the whole work more enjoyable.

Last, but far from least, I want to express my thanks to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

Table of Contents

Acknowledgements.....	ii
Table of Contents	iv
List of Figures	vi
Chapter 1 Introduction	1
Chapter 2 Review of Literature.....	2
2.1 Principles of Magnetic Resonance Imaging	2
2.1.1 Nuclear Magnetic Resonance Signal	2
2.1.2 Spatial Localization for Imaging	5
2.1.3 Fast Imaging – Echo Planar Imaging.....	10
2.2 Motion Artifacts.....	11
2.2.1 Reason of motion artifacts	11
2.2.2 Correction for Motion Artifacts	13
2.3 Stereo Tracking System	17
2.3.1 Camera Calibration	17
2.3.2 3D Reconstruction	19
2.3.3 Object Tracking	20
2.3.4 3D Motion Estimation.....	21
Chapter 3 Objective	23
Chapter 4 Prospective Motion Correction for MRI using a Stereo Tracking System	24
4.1 System Setup.....	24
4.2 System Calibration.....	26
4.2.1 CCD Camera Calibration.....	27
4.2.2 Stereo Calibration	36
4.2.3 3D Reconstruction from Stereo	39
4.2.4 Integrated Optical and MRI System Calibration	42
4.3 Real-time Motion Calculation.....	46
4.3.1 Kanade-Lucas-Tomasi Tracker.....	46
4.3.2 Stereo Matching.....	48
4.3.3 Rigid Motion Estimation.....	51
4.4 Motion Calculation in MRI coordinates	54
4.4.1 Motion in MRI physical coordinates	54
4.4.2 Motion in MRI logical coordinates.....	55
4.5 Motion Correction for MRI	57
4.5.1 Rotation Correction.....	57
4.5.2 Slice Selection Displacement Correction.....	58
4.5.3 Frequency Encoding Displacement Correction	66
4.5.4 Phase Encoding Displacement Correction.....	69
4.5.5 Summary of Position Compensation	72
4.5.6 K-space Line Re-acquisition.....	73
4.6 Software Flowchart for the Motion Correction System.....	73
Chapter 5 Results and Discussion with the Stereo Motion Correction System for MRI	75
.....	75

5.1 Camera Calibration Test	75
5.2 Stereo Reconstruction Validation	76
5.3 Tracker Noise Evaluation	77
5.4 Calibration Accuracy Between Cameras and MRI Coordinate System	79
5.5 Pulse Sequence Test with a Phantom.....	80
5.6 System Performance Test	83
5.7 Volunteer Study	84
5.7.1 Breathing Related Motion.....	85
5.7.2 Minimal Motion.....	86
5.7.3 Large Stepwise Motion.....	87
5.7.4 Slow Motion.....	88
5.8 Discussion	90
Chapter 6 Prospective Motion Correction for MRI using a Single Camera	94
6.1 System Setup.....	94
6.2 Method	95
6.2.1 Motion Computation.....	97
6.2.2 Best Match Search	98
6.3 Result	100
6.3.1 Accuracy of the Motion Estimation.....	100
6.3.2 Volunteer Study	103
6.4 Discussion	104
Chapter 7 Conclusion.....	107
Chapter 8 Future Work	109
Chapter 9 Appendices	111
9.1 Derivation of motion in logical coordinate system.....	111
9.2 MRI Coordinates from DICOM Image.....	112
9.3 Gauss-Newton Algorithm	114
9.4 Stereo System QA.....	117
Chapter 10 Bibliography.....	119

List of Figures

Figure 2-1	Slice selection. Activating a given magnetic-field gradient creates a linear correspondence between the magnetic-field strength and the position along the gradient. An RF pulse with center frequency ω_c excites spins located at position z_c ; the bandwidth of $\Delta\omega$ and the gradient strength determine the slice thickness Δz	6
Figure 2-2	Sequence diagram for a 2D gradient echo imaging sequence. RF pulse is a sinc function. G_z is the slice selection z-gradient. G_y is the phase encoding y-gradient, pictured as a series of horizontal lines to denote that it is being stepped regularly through increasing values during different repetition periods. G_x is the frequency encoding x-gradient and S is the echo.	7
Figure 2-3	k-space trajectory for 2D gradient echo imaging sequence. The signal starts from $k_x=k_y=0$ after each slice selection rephasing lobe. Phase encoding gradient moves the k-space trajectory vertically to a k_y depending on the current G_y (red vertical line). The dephasing lobe of G_x transverses the trajectory horizontally to $k_{x,min}$ (red horizontal line). The reversed read gradient then brings the spins to $k_{x,max}$ through a complete set of k_x values (red arrows). The experiment is repeated by changing G_y to obtain a complete set of data (black arrows).	9
Figure 2-4	(a) An EPI sequence with blipped phase encoding gradient. (b) K-space trajectory for the corresponding EPI sequence.	11
Figure 2-5	An axial brain image exhibits ghosting due to head motion. Since the movement is random, the ghosts are smeared out along the phase encoding direction (left/right).	12
Figure 2-6	A high resolution T2* weighted GRE brain image. The resolution is $200 \times 200 \times 1000$ micrometer, and the scan time is about 5 minutes.	13
Figure 2-7	Navigator sequence and its application for motion correction. (a) X-NAV sequence for x-axis displacement measurement. The NAV echo is interleaved into the imaging sequence and is similar to an image echo, except that no phase encoding is applied. (b) A transverse section of the legs of a volunteer. Left leg moved from side to side while the right leg stayed stationary. The artifact caused by the left leg motion was marked with the arrow. (c) The X-NAV data obtained during the scan. The motion of the left leg and the static position of the right leg were clearly shown. (d) Motion corrected images using the NAV data from (c). The motion artifact of the left leg was eliminated. (Figures from Ehman and Felmlee, 1989)	15
Figure 2-8	k-space trajectory. (a) projection reconstruction; (b) spiral; (c) PROPELLER.	16
Figure 4-1	Experimental setup. The top pane is an illustration of the position of the cameras and head coil in the MRI scanner bore. The bottom images show the real instruments. Cameras are highlighted with the red ellipse.	25
Figure 4-2	Calibration model. (a) Schematics of the phantom; (b) View of the actual phantom surface.	27

Figure 4-3	Pinhole camera geometry. O_C is the camera center and also the origin of the camera coordinate system, O_p is the principle point, M is a point in the space which has different coordinates in the world and camera coordinate systems, and m_d and m_f are the image of M in the image plane with and without lens distortion.....	28
Figure 4-4	Camera calibration using a square target. (a) Projection of a square ABCD to the image plane; (b) The image and the four vanishing points $\{\tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4\}$, and the horizon line $\tilde{\lambda}_H$	31
Figure 4-5	Two views of the calibration target. The red crosses mark the extracted corners used for calibration, which have the number of 100 for each view.....	36
Figure 4-6	The epipolar geometry. M is a point in space, O_c and O_c' are the two camera centers. The camera baseline intersects each image plane at the epipoles e and e' . The camera centers, 3D space point M and its images m and m' lie in a common plane, called epipolar plane. The two lines me and $m'e'$ are epipolar lines.....	37
Figure 4-7	Calibration between the optical system and MRI. The calibration phantom was put in the field of view of both the cameras and the MRI scanner.	43
Figure 4-8	Two images from the 3D scan of the calibration phantom. The red crossings mark the corners. The blue arrow marks an air bubble inside the phantom.	44
Figure 4-9	Phantom corner position estimation in MRI coordinate system. (a) shows one side of a prism. Blue dots are the observed corner position along that side; green line is the estimated side and the red circle is the estimated corner position on the phantom surface. (b) shows the alignment between the extracted internal structure (green) of the phantom and an ideal model (red). The green lines are the side of the prisms and the red circles are the estimated corners position in MRI.	45
Figure 4-10	Epipolar lines estimated after stereo calibration. (a) shows image from one of the camera and two randomly chosen points. (b) shows the image from the other camera and the estimated epipolar lines corresponding to the two points in (a) respectively.....	49
Figure 4-11	MRI physical and logical coordinate system. $\{O_M, X_M, Y_M, Z_M\}$ is MRI physical coordinate system, which doesn't change after the MRI system is set up. $\{O_L, X_L, Y_L, Z_L\}$ is MRI logical coordinate system, which changes with different prescribed scan plane. X_L is the frequency encoding direction, Y_L is phase encoding direction and Z_L is slice selection direction.	56
Figure 4-12	The change of RF frequency to select the same slice when motion occurs. Under the same slice-selective gradient, to follow the motion, the only way is to change the RF frequency.....	59
Figure 4-13	Slice selection sequences (RF pulse and z-gradient) and the three paths that generate the RF pulse, i.e. center frequency path, delta frequency path and phase path. $t=0$ is the start of TR and only one TR is shown here.	60
Figure 4-14	Hybrid magnitude and phase image of the cylindrical phantom. Phase encoding gradient was zero. Frequency encoding was left-right direction.	62
Figure 4-15	The hybrid space image after applying a linear RF frequency. Phase encoding gradient is set to zero. For each RF excitation, its frequency changes	

linearly with a step $\Delta\omega$. Perform FFT on each measured echo in k-space can we get a hybrid image. The magnitude of each line in the frequency encoding direction (shown with a rectangle shape) represents the object profile. The red line is to select an arbitrary point in the hybrid image, and the phases of those points also change linearly with a step $\Delta\phi$. Complex means to combine the magnitude and phase of the point in that line. 63

Figure 4-16 The hybrid phase image of the phantom when phase encoding gradient is set to zero. RF transmit frequency varies linearly along the phase encoding direction. (a) Phase correction is calculated with $\Delta t=264 \mu s$; (b) Phase correction is calculated with $\Delta t=180 \mu s$; (c) Phase correction is calculated with $\Delta t=230 \mu s$; (d) Phase correction is calculated with $\Delta t=480 \mu s$ 65

Figure 4-17 Frequency encoding gradient and the acquisition window associated with it. Also the three paths that generate the receive frequency and phase. 67

Figure 4-18 Hybrid phase image used for measurement of $\Delta t'$. Phase encoding gradient is set to zero. RF receive frequency varies linearly along the phase encoding direction with a step of 100Hz. (a) Phase correction is calculated with $\Delta t'=144 \mu s$; (b) Phase correction is calculated with $\Delta t'=100 \mu s$; (c) Phase correction is calculated with $\Delta t'=200 \mu s$. (d) Phase correction is calculated with $\Delta t'=300 \mu s$ 69

Figure 4-19 Software flowchart for the prospective head movement correction system for MRI. Calibration is all done offline and programmed with MATLAB. Real-time correction was programmed with C++ running in Linux. N is the number of total phase encoding steps for a scan. 74

Figure 5-1 Distortion and correction for the camera. (a) Distortion model for the camera, containing radial and tangential distortion. The arrows and number on the curve mark the distortion direction and 76

Figure 5-2 Noise of the tracker. The maximum x translation error as a function of the number of tracked feature points. 78

Figure 5-3 Rotation and Translation parameters estimated from MRI images (solid lines) and from camera images transformed into MRI coordinates (dotted lines). The two sets of lines matching very well means the calibration between camera and MRI is accurate. 80

Figure 5-5 The MR hybrid image without phase encoding gradient. Small breathing induced head movement was observed during the scan of the volunteer. (a) without motion correction; (b) with motion correction. 84

Figure 5-6 Prospective breathing induced motion correction for the 2D gradient echo acquisition. Images (a,b) show the evolution of motion parameters reported by the tracker during the acquisition of images (c) and (d) respectively. Image (c) was scanned without motion correction, and (d) was with motion correction; (e) and (f) were zoomed in on an area in the frontal brain indicated by the red square in images (c) and (d) respectively. 86

Figure 5-7 Prospective motion correction data obtained for a volunteer that attempted to minimize head motion. Images (a,b) show the motion parameters reported by the tracker during the acquisition of images (c) and (d) respectively. Image (c) was scanned without motion correction, and (d) was with motion

	correction; (e) and (f) magnify an area in the front of the brain (the area marked by the red square in (c) and (d) respectively).	87
Figure 5-8	Example of the performance of the prospective motion correction method in the presence of large stepwise motion during the acquisition of multi-shot 2D gradient echo images on a normal volunteer. Images (a-c) show the motion parameters reported by the tracker during the acquisition of images (d-f) respectively. Image (d) was scanned without intentional motion during the scan, whereas (e,f) were scanned in the presence of motion; (e) was acquired without correction but (f) was acquired using the proposed method. (g-i) show additional detail in an area in the front of the brain that was indicated by the red square in images (d-f).....	88
Figure 5-9	Prospective motion correction in the presence of slow motion. The motion parameters reported by the tracker during the scan of images (d-f) are displayed in (a-c), respectively. Image (d) was acquired in the absence of intentional head motion, (e,f) during slow head motion, where, unlike (e), (f) was acquired with the proposed correction method. A magnification of the area in the red box in (d-f) is shown in (g-i).	89
Figure 6-1	System setup. The camera was fixed right on top of the head coil.	95
Figure 6-2	Method of the motion estimation using a single camera. N training EPI and camera images were acquired. From EPI images, 6DOF motion parameters were calculated from registration for each training data. During the real-time scan, the most similar camera image in the training set was found for the newly acquired camera image. The corresponding motion parameters from EPI were sent to MRI pulse sequence for prospective motion correction.	96
Figure 6-3	Motion parameters from the training data set. The total number of training images is 145. The motion is mainly rotation around z axis and translation in x axis.....	101
Figure 6-4	Correlation of two randomly chosen images with all training images. Red crossings mark the chosen image, (a) the correlation between image No. 47 with others; (b) the correlation between image No. 100 with others.....	101
Figure 6-5	Prospective correction of large motion. (a) was acquired without intentional motion during the scan. (b c) were scanned in the presence of motion; (b) was acquired using the proposed method but (c) was without correction. (d,e) show the evolution of motion parameters during the acquisition of image (b,c), respectively.	104
Figure 9-1	3D reconstruction error if camera parameters change. The solid lines projection is correct and the dotted projection is the result from different camera parameters with a small focal length and principle point difference. It's obvious that the 3D estimation magnify the camera parameters distinction.....	117
Figure 9-2	3D position estimation for a checker board pattern. Every circle represents a corner of the checker board. Red color shows position estimation from the left camera; blue color shows position estimation from the right camera; and green color shows those from the stereo. (a) when correct calibration parameters were used. (b) when focal length was changed with only 0.1mm. .	118

Chapter 1 Introduction

Magnetic resonance imaging (MRI) has been widely used as a noninvasive clinical and research modality for the study of human anatomy. However, subject motion during scanning remains a severe problem and may degrade image quality below levels acceptable for clinical diagnosis. Recent improvements that yield high spatial resolution MRI of the brain (~ 0.2 mm) make this problem more acute, since for this application the tolerance to motion is reduced, and scan time is increased. The latter makes it more difficult for the subject to maintain the same position throughout the scan, which is especially problematic for children and patients.

To compensate the effects of head motion during data acquisition, an in-bore optical motion tracking system is proposed. The system comprises one or two MR compatible infrared cameras that are fixed on a holder right above and in front of the head coil. The resulting close proximity of the camera(s) to the object allows precise tracking of its movement. During image acquisition, the MRI scanner uses this tracking information to prospectively compensate for head motion by adjusting gradient field direction and RF phase and frequency. Experiments performed on subjects demonstrate the system was able to improve MR image quality if motion occurred during acquisition.

Chapter 2 Review of Literature

Magnetic resonance imaging (MRI) has been widely used as a noninvasive clinical and research modality for the study of human anatomy. It exploits the phenomenon of nuclear magnetic resonance (NMR) in an external magnetic field, whereby nuclei absorb and re-emit certain radio frequency (RF) waves during change in nuclear spin state. Image formation using NMR signals was developed by Lauterbur and Mansfield (Lauterbur, 1973; Mansfield and Grannell, 1973) in 1973 based on spatial encoding principles, which won them the 2003 Nobel Prize in Physiology and Medicine. Since then, MRI has undergone dramatic improvements in all the features that define image quality, such as resolution, signal-to-noise ratio (SNR), contrast enhancement and speed. Specific structures, such as arteries, lesions, white matter fiber tracts, can also be visualized by manipulating RF fields and local magnetic field.

2.1 Principles of Magnetic Resonance Imaging

2.1.1 Nuclear Magnetic Resonance Signal

The nucleus is composed of protons and neutrons, each of which intrinsically possesses angular momentum \vec{J} , often called spin (Liang et al., 2000; Gerlach and Stern, 1924). Nuclei with an odd number of protons and/or an odd number of neutrons have a net spin. Because the nucleus is positively charged and possesses angular momentum, it possesses a magnetic moment $\vec{\mu}$ also (Rabi et al., 1938). The relationship between the spin angular momentum and magnetic moment is

$$\vec{\mu} = \gamma \vec{J} \quad [2.1]$$

Where γ is the gyromagnetic ratio. A related constant \mathcal{Z} is also widely used, which is defined as

$$\mathcal{Z} = \frac{\gamma}{2\pi} \quad [2.2]$$

The value of γ or \mathcal{Z} is nucleus-dependent. For hydrogen (proton), γ is 2.675×10^8 rad/s/T, \mathcal{Z} is 42.58MHz/T. Because of its high concentration in the body in water and fat molecules, and their relatively high gyromagnetic ratio, the hydrogen proton is the source of signal for most clinical MRI exams.

In the absence of an externally applied magnetic field, the hydrogen spins in biological materials have random orientations, therefore the vector sum of the magnetic moments, often referred to as net magnetization, denoted by \vec{M}_0 , is zero. However, if an external magnetic field \vec{B}_0 is applied, the spins will take only a discrete number of orientations, which equals to twice the nuclear-spin value plus one. For protons with a spin of $\frac{1}{2}$, $\vec{\mu}$ has two possible orientations with respect to \vec{B}_0 direction.

Since spins in different orientations have different energy of interaction with the external magnetic field \vec{B}_0 , two energy states can be measured for protons, a high- and a low-energy state. The high-energy state is when spin is aligned opposing \vec{B}_0 , or spin-down E_{\downarrow} , while the low-energy state is when spin is aligned with \vec{B}_0 , or spin-up E_{\uparrow} . At equilibrium, the relative populations of the up and down states are given by the Boltzmann distribution as:

$$\frac{N_{\downarrow}}{N_{\uparrow}} = \exp\left(-\frac{\gamma\hbar B_0}{kT}\right) \approx 1 - \frac{\gamma\hbar B_0}{kT} \quad [2.3]$$

Where \hbar is Planck's constant with the value $\hbar=1.0546\times 10^{-34}$ J.s; k is Boltzmann's constant, with the value $k=1.381\times 10^{-23}$ J/K; T is the absolute temperature of the sample in Kelvin.

Since the net magnetization \vec{M}_0 is the vector sum of the individual magnetic moments, the excess of spins in the spin-up state results in an \vec{M}_0 aligned with the external field B_0 . To measure MRI signal by the RF receiver coil, \vec{M}_0 needs to be tipped away from the longitudinal direction (B_0 direction) to the transverse plane (orthogonal to B_0). This can be achieved by applying a radiofrequency (RF) pulse at the Larmor frequency,

$$\omega_0 = \gamma B_0 \quad [2.4]$$

The RF pulse is normally turned on for a few microseconds or milliseconds. The created B_1 magnetic field makes the spins to precess, As a result, the net magnetization is tipped away from B_0 direction, creating a measurable transverse component \vec{M}_{xy} .

At equilibrium, the fractional population difference $\frac{\gamma\hbar B_0}{kT}$ is very small, which is about 1 in 100,000 nuclei at 1.5T. Because the population difference is so small, the equilibrium net magnetization \vec{M}_0 is also small, making the measured MR signal \vec{M}_{xy} small. Therefore, to get a better signal-to-noise ratio (SNR) image, a stronger B_0 is desirable.

2.1.2 Spatial Localization for Imaging

A basic example of spatial localization is selective excitation and Fourier spatial encoding. To form an image, the first step is to localize the signals by exciting the magnetization only within a specified slice or slab of body (Bernstein et al., 2004; Lauterbur, 1973; Mansfield and Grannell, 1973). This slice selection is achieved by applying the RF pulse in the presence of a magnetic field gradient G_z . As given by the Larmor equation, the precession frequency is proportional to the field strength. The gradient therefore creates a variation in the Larmor frequency along the direction of the gradient. So, as shown in Figure 2-1, the RF pulse with a specific frequency ω_c excites only spins at a specific position z_c because only those spins possess a resonant frequency that equals to the frequency of the applied RF pulse. The selection of another slice can be accomplished by changing the RF frequency, for example in Figure 2-1, increasing the RF frequency will excite a slice more superior while decreasing the RF frequency will select a slice more inferior. The thickness of the slice is determined by the bandwidth of the RF pulse and the amplitude of the gradient, as:

$$\Delta z = \frac{\Delta \omega}{\gamma G_z} \quad [2.5]$$

Where Δz is the slice thickness, G_z is the applied gradient field strength and $\Delta \omega$ is the bandwidth of the RF pulse.

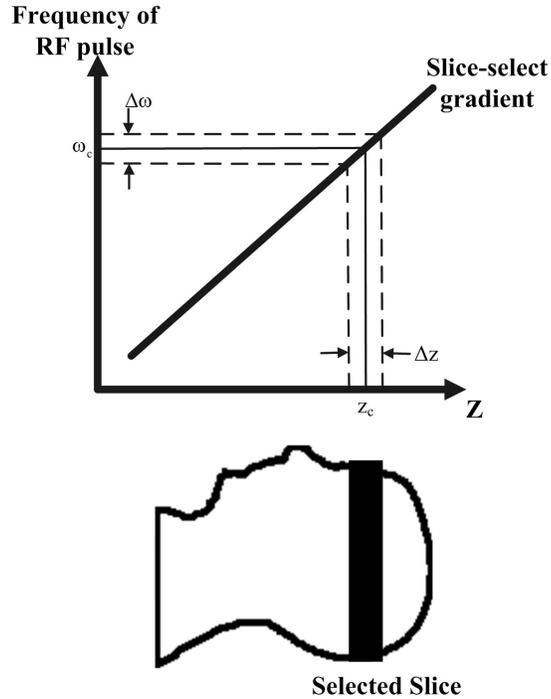


Figure 2-1 Slice selection. Activating a given magnetic-field gradient creates a linear correspondence between the magnetic-field strength and the position along the gradient. An RF pulse with center frequency ω_c excites spins located at position z_c ; the bandwidth of $\Delta\omega$ and the gradient strength determine the slice thickness Δz .

After the slice selection gradient is turned off, a phase encoding gradient (G_y) can be implemented. The magnetization accumulates a y-dependent phase when the y-gradient is kept on for a time τ_y :

$$\phi_y = \gamma G_y \tau_y y \quad [2.6]$$

The encoded phase can be written in terms of the k-space spatial frequency (Twieg, 1983) in y direction:

$$\phi_y = 2\pi k_y y \quad \text{with} \quad k_y = \gamma G_y \tau_y \quad [2.7]$$

The gradient G_y varies each time after excitation in a step-like fashion, to gather enough information for localizing the spin density in y direction.

The frequency encoding gradient G_x is applied after G_y . A negative dephasing lobe is followed by a positive lobe, during which an echo of the magnetization is formed and the signal is measured. The accumulated x-dependent phase is:

$$\phi_x = \gamma G_x t_x x = 2\pi k_x x \quad \text{with} \quad k_x = \gamma G_x t_x \quad [2.8]$$

The gradient G_x is a constant, and the time t_x is a variable along the measurement period.

These spatial encoding events can be depicted in a graphical format, called a pulse sequence. Figure 2-2 shows the basic gradient echo (GRE) sequence described above.

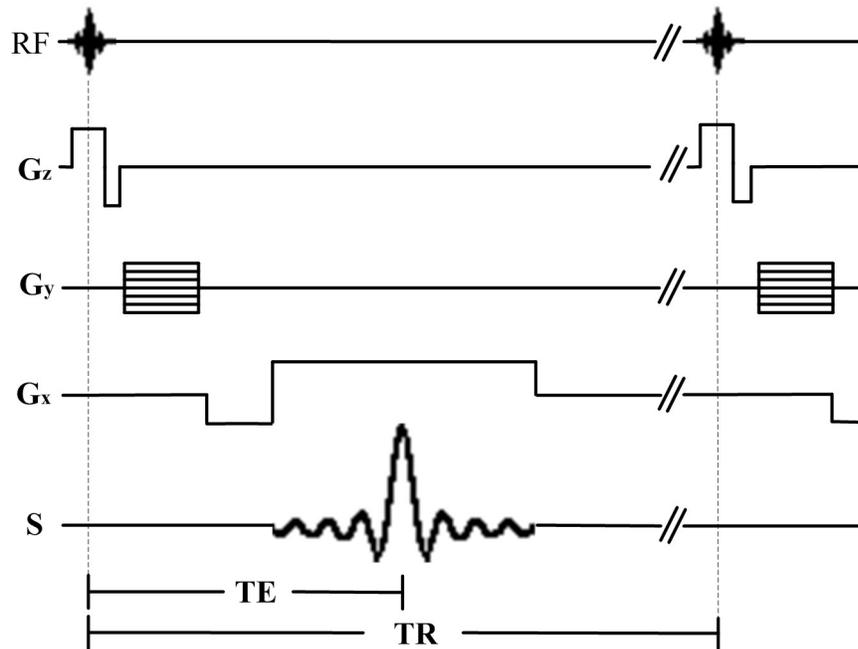


Figure 2-2 Sequence diagram for a 2D gradient echo imaging sequence. RF pulse is a sinc function. G_z is the slice selection z-gradient. G_y is the phase encoding y-gradient, pictured as a series of horizontal lines to denote that it is being stepped regularly through increasing values during different repetition periods. G_x is the frequency encoding x-gradient and S is the measured echo signal.

The spin density $\rho(x, y, z_0)$ and the measured signal $s(k_x, k_y)$ satisfy the following equation:

$$s(k_x, k_y) = \iint_{fov} \rho(x, y, z_0) \exp(-i2\pi(k_x x + k_y y)) dx dy \quad [2.9]$$

Therefore, sufficient k-space coverage is needed to reconstruct the image of the spin density with inverse 2D Fourier transform,

$$\rho(x, y, z_0) = \iint_{fov} s(k_x, k_y) \exp(i2\pi(k_x x + k_y y)) dk_x dk_y \quad [2.10]$$

The k-space trajectory of the GRE imaging sequence is illustrated in Figure 2-3. The initial phase is $k_x = k_y = 0$ after each slice selection, corresponding to the center of the k-space. A phase encoding gradient moves the k-space trajectory vertically to a k_y depending on the current G_y (red vertical line). The dephasing lobe of the frequency encoding gradient transverses the trajectory horizontally to $k_{x,min}$ (red horizontal line). The positive read gradient then brings the spins to $k_{x,max}$, with signal being measured (red arrows). This gives one line in k-space, so the experiment is repeated by stepped changing G_y to obtain a complete set of data (black arrows).

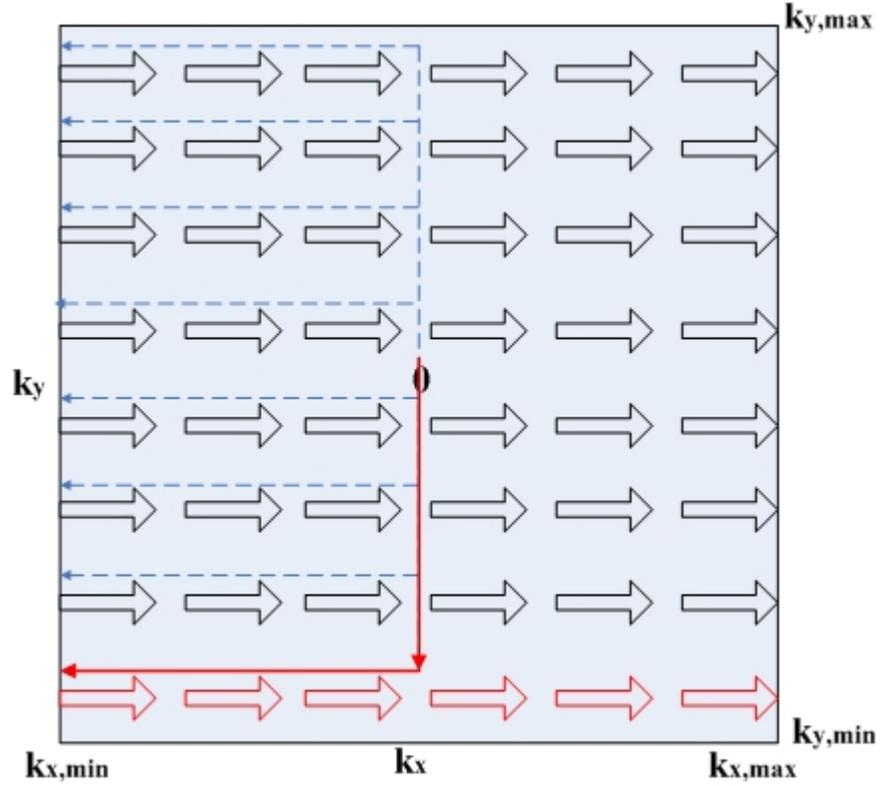


Figure 2-3 k-space trajectory for 2D gradient echo imaging sequence. The signal starts from $k_x=k_y=0$ after each slice selection rephasing lobe. Phase encoding gradient moves the k-space trajectory vertically to a k_y depending on the current G_y (red vertical line). The dephasing lobe of G_x transverses the trajectory horizontally to $k_{x,min}$ (red horizontal line). The reversed read gradient then brings the spins to $k_{x,max}$ through a complete set of k_x values (red arrows). The experiment is repeated by changing G_y to obtain a complete set of data (black arrows).

The MR signal is collected by uniformly sampling points in the k-space rather than by continuous monitoring, leading to discrete k-space coverage.

$$s(k_x, k_y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \rho(m\Delta x, n\Delta y) \exp(-i2\pi(m\Delta x k_x + n\Delta y k_y)) \quad [2.11]$$

Where M and N are the number of sampled k-space data points in x and y direction; Δx and Δy are the pixel size. Therefore, the image of the spin densities can be reconstructed from the discrete inverse Fourier transform of the k-space data.

The total scan time required for 2D GRE acquisition is described by the following equation:

$$T_{acq} = TR \times M \times NEX \quad [2.12]$$

Where TR equals the repetition time and NEX equals the number of repetitions used (signal averaging). For a scan with TR 1s, phase encoding 256, and NEX 2, the total scan time is about 5 minutes. A higher resolution scan using much more phase encoding can prolong the scan time even more. Fast imaging strategies are therefore developed.

2.1.3 Fast Imaging – Echo Planar Imaging

The first ultra high-speed imaging technique was proposed by Mansfield in 1977 (Mansfield, 1977), named Echo Planar Imaging (EPI). Since then, various EPI techniques have been proposed.

EPI encodes complete image position information in a single acquisition line rather than phase encoding a single line of k-space with every TR interval. Therefore, with a single excitation, an image can be formed. A GRE-EPI sequence is shown in Figure 2-4(a). Phase encoding of individual gradient echoes is through the use of a series of blipped G_y pulses. Figure 2-4(b) illustrates the k-space trajectory of the EPI sequence.

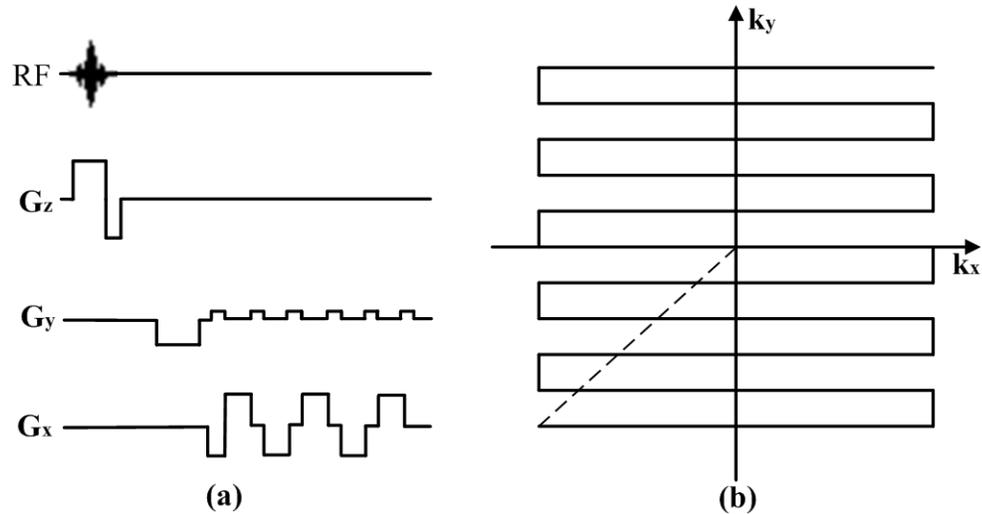


Figure 2-4 (a) An EPI sequence with blipped phase encoding gradient. (b) K-space trajectory for the corresponding EPI sequence.

Because of the longer readout, EPI suffers more than GRE from various types of artifacts. In regions of high field inhomogeneities or at interfaces between tissues of different susceptibilities, image distortions or signal drop-out can be generated. The odd and even echoes shown in Figure 2-4(b) may have slight timing mismatches caused by imperfections in the gradients, which translate into phase errors and manifest themselves as duplicate images of the main image along the phase direction.

2.2 Motion Artifacts

2.2.1 Origin of motion artifacts

For a slow high resolution structural scan, the head motion of an uncooperative volunteer causes motion artifacts primarily in the phase-encoding direction. The frequency encoding direction is less affected since the k-space data corresponding to the readout direction are collected over a relatively short period of a few milliseconds or less, and can be regarded as instantaneous on the time scale of most physiologic

motion. However, in the phase encoding direction, the positions are encoded by a step-wise changing magnetic field gradient prior to the acquisition of each line of k-space. Therefore, the time interval between neighboring two points in phase encoding direction is TR, which ranges from tens of milliseconds or several seconds and cannot be regarded as instantaneous. The motion will cause additional phase shifts to the spins from line to line, which will corrupt the spatial encoding and give rise to ghosting.

Head motion is usually random motion, which makes the artifacts smeared out along the phase encoding direction and can obscure pathology (Figure 2-5).

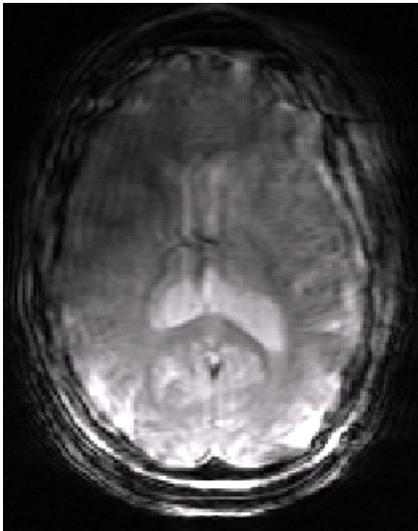


Figure 2-5 An axial brain image exhibits ghosting due to head motion. Since the movement is random, the ghosts are smeared out along the phase encoding direction (left/right).

Recent improvements that yield high spatial resolution MRI of the brain (~0.2 mm, Figure 2-6) make this problem more acute, since for this application the tolerance to motion is reduced, and scan time is increased (Duyn et al., 2007; van Gelderen et al., 2007). The latter makes it more difficult for the subject to maintain

the same position throughout the scan, which is especially problematic for children and patients.

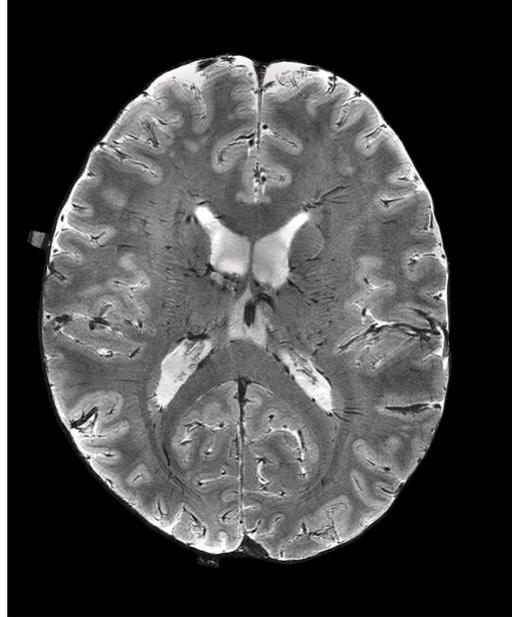


Figure 2-6 A high resolution T2* weighted GRE brain image. The resolution is $200 \times 200 \times 1000$ micrometer, and the scan time is about 5 minutes.

2.2.2 Correction for Motion Artifacts

Several methods have been proposed to solve the head movement problem in MRI. All model head movement as rigid body motion with six degrees of freedom (DOF), namely three rotations and three translations along the MRI coordinate system. These parameters are then used to either retrospectively or prospectively compensate for the effects of motion on the image data.

Retrospective motion correction addresses motion artifacts after the acquisition of a complete set of raw image data. While this might work well for in-plane motion, it is generally inadequate for through-plane motion, primarily because it cannot correct for the effects of this motion on the local magnetization history (i.e.

changes in saturation level of longitudinal magnetization due to motion-induced changes in the image-slice location). To avoid this problem prospective motion correction techniques have been proposed recently (Lee et al., 1998; Lee et al., 1996; Thesen et al., 2000; van der Kouwe et al., 2006; Ward et al., 2000). They track the head motion and rectify the acquisition planes correspondingly, by adjusting gradient direction and RF phases and frequencies. This avoids problems related to the effects of motion on spin history.

The object motion parameters used by prospective and retrospective correction methods can be derived from MRI data using image or navigator based methods. Image based methods use image registration algorithms to detect the motion parameters (Friston et al., 1995; Friston et al., 1996; Jenkinson et al., 2002; Thesen et al., 2000). These techniques, while straightforward, can only calculate motion after acquiring a volume and are therefore always lagging the motion. Also, any ghosting and blurring artifacts that motion may cause may affect the accuracy of the registration algorithms.

Navigator based motion correction methods acquire a motion-sensitive reference signal with the image (Ehman and Felmlee, 1989; Fu et al., 1995; Lee et al., 1998; Lee et al., 1996; van der Kouwe et al., 2006; Barnwell et al., 2007; Kadah et al., 2004). The earliest navigator method was capable of detecting only one-dimensional (1D) translation (Ehman and Felmlee, 1989) by employing a frequency encoding gradient but no phase encoding gradient before image acquisition. Soon after, orbital navigators (Fu et al., 1995; Ward et al., 2000), spherical navigators (Welch et al., 2002; Wyatt et al., 2005; Petrie et al., 2005), cloverleaf navigators (van

der Kouwe et al., 2006) and other navigators (Spuentrup et al., 2002; Thiel et al., 2002) (Wang et al., 2005; Pipe et al., 2002) were proposed to allow full tracking of 3D motion. However, the extra time required for measuring the navigator echoes led to an increased scan time.

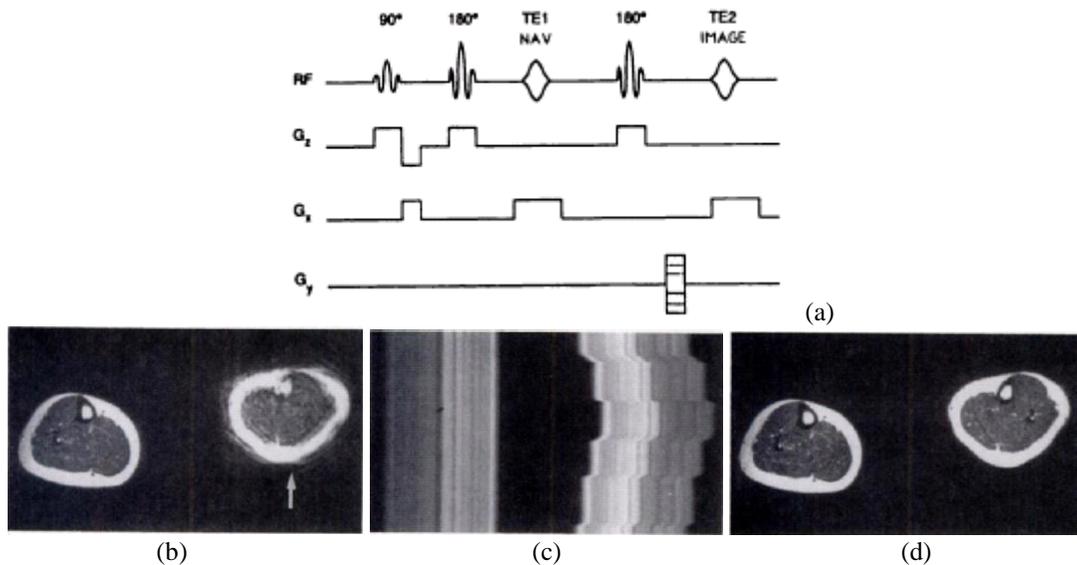


Figure 2-7 Navigator sequence and its application for motion correction. (a) X-NAV sequence for x-axis displacement measurement. The NAV echo is interleaved into the imaging sequence and is similar to an image echo, except that no phase encoding is applied. (b) A transverse section of the legs of a volunteer. Left leg moved from side to side while the right leg stayed stationary. The artifact caused by the left leg motion was marked with the arrow. (c) The X-NAV data obtained during the scan. The motion of the left leg and the static position of the right leg were clearly shown. (d) Motion corrected images using the NAV data from (c). The motion artifact of the left leg was eliminated. (source: Ehman and Felmlee, 1989)

Self-navigating methods can also be used to retrospectively correct motion when combined with particular image acquisition techniques including projection acquisition (Glover and Noll, 1993; Glover and Pauly, 1992; Kim et al., 2008), spiral acquisition (Glover and Lee, 1995; Liu et al., 2004; Meyer et al., 1992; Hu and Glover, 2007), and PROPELLER (Pipe, 1999; Pipe and Zwart, 2006; Wang et al., 2005; Pipe et al., 2002). These all involve the collection of redundant MRI data,

usually including the central region of the k-space. The over-sampled central region of k-space provides intrinsic averaging of image features that reduces motion artifacts, and can be further used to correct spatial inconsistencies in position, rotation, and phase between acquisitions. One drawback of these methods is the increased scan time related to redundant k-space sampling.

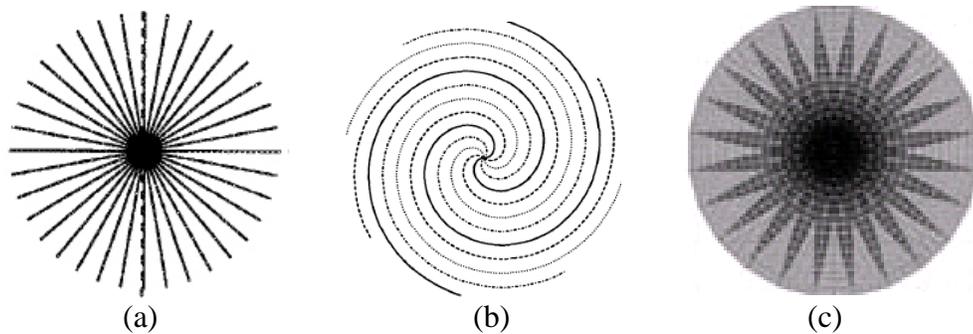


Figure 2-8 k-space trajectory. (a) projection; (b) spiral; (c) PROPELLER.

Alternatively, external motion tracking systems can be used for MRI motion correction. These tracking systems rely on additional hardware, such as miniature coils (Derbyshire et al., 1998; Nevo et al., 2002), fiducial samples (Eviatar et al., 1997), optical reflectors (Eviatar et al., 1999) or stereo cameras (Dold et al., 2006; Tremblay et al., 2005; Zaitsev et al., 2006). Among them, stereo optical systems have been shown to give good motion correction results with reasonable accuracy retrospectively (Tremblay et al., 2005) as well as prospectively (Dold et al., 2006; Zaitsev et al., 2006). The stereo tracking system works in parallel with the scanner thus needs no extra scan time for motion detection in the imaging sequence. However, there are limitations to the current tracking systems, the most important being that the tracking target required for monitoring object position is either uncomfortable for the subjects (e.g., a mouthpiece used in Zaitsev et al., 2006), or

hard to combine with close-fitting receiver arrays (e.g., a cap used in Tremblay et al., 2005). To overcome these limitations, we developed an in-bore video tracking system for prospective motion correction that monitors the position of features on the subject's face. The system was designed to allow high resolution ($\sim 0.2\text{mm}$) MRI at 7T in the presence of substantial head motion.

2.3 Stereo Tracking System

Stereo tracking system includes two cameras, acquiring images of the same object simultaneously from slightly different view points. To determine the 3D motion of an object, the system needs to be accurately calibrated beforehand. The object's motion parameters can be estimated by analyzing the image pairs captured by the stereo system. The tracking feature points are first detected in each image pair and the 3D position of these feature points are calculated through stereo triangulation. The 6-DOF motion parameters (3-DOF translations plus 3-DOF rotations) can be further estimated based on the 3D coordinates of these feature points.

2.3.1 Camera Calibration

Camera calibration is a crucial issue in computer vision. It serves to determine the relationship between a 2D image perceived by a camera (or images perceived by multiple cameras) and the 3D information of the real object. Precise camera calibration is especially important for applications that involve metric measurements, such as dimensional measurements (Tomasi and Kanade, 1992), object motion

(Broggi, 1998; Charbonnier and Fournier, 1995), visual inspection (Newman and Jain, 1995), etc.

Camera models usually involve of two sets of parameters: extrinsic and intrinsic. Extrinsic parameters, including a rotation and translation vector, model the geometrical relationship between the camera and the scene. Intrinsic parameters, including focal length, principal points, and distortions, model the internal geometry of a camera. They determine how to derive an image point position given the spatial position of the point with respect to the camera. The position and orientation of a stereo setup is also intrinsic since it is independent of scene structure. Intrinsic parameters remain constant after cameras are set up while extrinsic parameters may change with different scenes. The estimation of these two sets of parameters is called camera calibration, which allows one to relate the image measurements to the spatial structure of the observed scene.

The existing camera calibration approaches can be classified into three categories:

a) Linear closed-form methods. These methods construct a linear transformation to relate 3D points with their 2D projections (Abdel-Aziz and Karara, 1971; Faugeras and Toscani, 1986; Hall et al., 1982). Then least squares methods are used to get the closed-form solution. The advantage of these methods is the simplicity of the model and fast computation. However, they have the following disadvantages. First, the linear models don't include lens distortion, which make these methods less accurate. Second, camera parameters are not explicit in these models, and constraints (i.e. rotation matrix should be orthonormal) are not considered.

b) Non-linear optimization methods. When lens distortion is considered in the camera model, the relationship between the 3D scene points and their 2D projections becomes non-linear. The non-linear error function is usually defined as the distance between 2D image points and the projection of the 3D points. Camera parameters can then be optimized iteratively by minimizing the defined error function (Faig, 1975; Slama et al., 1980). The advantage of these methods is that any distortion can be considered in the model and good accuracy can be achieved by reaching correct convergence. However, a good initial guess is required to guarantee correct convergence.

c) Two-step methods. These methods first find closed-form solutions for some parameters, then do iterative optimization for the other or all of the parameters (Tsai, 1987; Weng et al., 1992b; Zhang, 2000). These methods take the advantages of the previous two methods and can always converge quickly to a correct solution due to the closed-form guess in the first step. Explicit intrinsic and extrinsic camera parameters and major distortions (namely, radial, tangential, decentering) can all be estimated with high accuracy in a short time.

2.3.2 3D Reconstruction

3D reconstruction computes the position of a point in 3D given its image in two views and the camera intrinsic parameters. The linear triangulation method (Bouguet, 1999; Hartley and Zisserman, 2003) back projects rays from the measured image points and find the intersection of the two rays for the 3D point. Since errors are inevitable in the measured points, the epipolar constraint (explained in Section 4.2.2

in detail) may not be satisfied and the intersection of the back-projected rays may not exist. Sampson error correction is applied to these measured points to correct first-order errors (Torr and Zisserman, 1997). This approximation is valid only when the error of the measured points is small. However, the epipolar constraint is still not satisfied exactly. A more costly triangulation method (Hartley and Sturm, 1997) can find the optimal 3D points by solving a sixth-degree polynomial.

2.3.3 Object Tracking

Visual tracking is a very important component for systems with applications in robot control (Milella and Siegwart, 2006), surveillance (Coifman et al., 1998), human-computer interface (Gorodnichy and Roth, 2004), and visual reconstruction (Tomasi and Kanade, 1992). Various tracking methods have been proposed, which can be classified in four categories: model based (Kakadiaris and Metaxas, 2000; Zhou et al., 2004), region based (Hager and Belhumeur, 1998), contour based (Blake et al., 1993) and feature based (Shi and Tomasi, 1994; Yao and Chellappa, 1995; Zheng and Chellappa, 1993) methods. In this project, feature based tracking is used, because of its subpixel precision, speed, as well as robustness.

The feature tracking process can be divided into two major subtasks: feature selection and feature tracking. Good features can be identified as edges or corners (Harris and Stephens, 1988) or zero crossings of the Laplacian of the image intensity (Marr et al., 1979). However, these feature selection criteria are often defined independently of the tracking algorithm. With the Kanade-Lucas-Tomasi (KLT) tracker, instead, a good feature is defined as the one that can be tracked well or that

can make the tracker work best (Lucas and Kanade, 1981; Tomasi and Kanatani, 1991). The tracker starts from Lucas and Kanade's fast image registration technique that makes use of the spatial intensity gradient of the images to find a good match (Lucas and Kanade, 1981). After that initial work, Tomasi and Kanade developed a feature tracker based on sum of squared intensity differences (SSD) and assumed translational inter-frame motion (Tomasi and Kanatani, 1991). Then, Shi and Tomasi proposed an affine model to make the region matching better over longer time spans (Shi and Tomasi, 1994). Subsequently, Tommasini proposed an extension of this tracker to automatically reject spurious features, which made the tracker even more robust (Tommasini et al., 1998). Now, the KLT tracker has been widely used in many systems for high accuracy tracking, such as facial feature tracking, robot control, and flow velocity measurement.

2.3.4 3D Motion Estimation

Computing a 3D rigid motion that maps a set of 3D points to another set is the basis of the absolute orientation problem (Horn, 1987; Horn et al., 1988). It is an important computer vision task and has numerous applications, such as robotics, pose estimation, and motion analysis. Solutions to this problem include closed-form and iterative methods, between which closed-form solution is preferred because of its robustness and efficiency. Iterative methods, however, may be trapped in local minima of the error function and always need a good initial guess. Because of these reasons, we adopted closed-form methods for our motion estimation.

Four major closed-form solutions are used for 3D rigid body transformations. The first one was proposed by Arun (Arun et al., 1987). In their derivation, rotation is represented as a standard orthonormal matrix and the solution is based on the singular value decomposition (SVD) of a covariance matrix of the data. However, this method will fail if the data is seriously noisy or becomes planar. Umeyama (1991) and Kanatani (1994) corrected this with different ways of derivation. Horn (1987) represented rotation using quaternions and the solution is the eigenvector corresponding to the largest positive eigenvalue of a symmetric matrix. Horn (1988) later also used an orthonormal matrix to represent rotation, and derived an alternative method that used manipulation of matrices and their eigenvalue-eigenvector decomposition. Walker et al (1991) presented a solution by representing rotation and translation together as a dual quaternion. These four algorithms were compared with respect to accuracy, stability and efficiency by Eggert et al (1997). The accuracy shows no discernible differences, while the SVD method (Arun et al., 1987; Umeyama, 1991) gives the best stability. Even though it is somewhat slower than Horn's (1988) method, the superior accuracy and stability made us choose this method in our research.

Chapter 3 Objective

The overall objective of this project is to design a new tracking system to prospectively correct head motion for MRI. This optical system should be compatible with MRI. It should be fast, robust, and accurate for tracking of head motion and should also be able to communicate with MR scanner before the MRI data are acquired. With the accomplishment of this project, sensitivity of brain MRI to head motion should be substantially reduced. Specifically, the objectives are:

- 1) To setup the optical tracking system. One or two MR compatible infrared cameras should be used for tracking with illumination of infrared LEDs. The tracking system needs to be able to determine the subject's head position in real time (three rotations plus three translations).
- 2) To program MR pulse sequence to receive motion parameters from the optical tracking system and apply motion correction in real-time. Before each excitation, the MRI scan computer should be able to read the motion parameters to adjust the gradient field direction and RF phase and frequency in the imaging sequence to compensate the changes in position.
- 3) To evaluate the system accuracy and performance. Specifically, we need to evaluate the accuracy of the optical tracking system, the calibration accuracy of the combined optical and MRI system, and the acquired MRI image quality with and without motion compensation.

Chapter 4 Prospective Motion Correction for MRI using a Stereo Tracking System

4.1 System Setup

MRI experiments were performed on a GE (Milwaukee, WI, USA) 7T MRI system, equipped with a detunable volume transmit coil and a 32-channel receive-only head coil (Nova Medical, Wilmington, MA, USA). The optical tracking system included two MR compatible infrared cameras (MRC Systems GmbH, Germany) as part of a stereo-vision system to measure the subject's head motion in 3D. The size of the cameras was $28 \times 18 \times 30\text{mm}^3$. They were fixed on a holder right above and in front of the head coil (Figure 4-1). The distance and the angle between them were 12 cm and 28° respectively. Surrounding each camera lens, six infrared light emitting diodes were used to illuminate the field of view. The distance from the cameras to the face was around 8cm. The high spatial resolution of this setup (640×480 pixels, about $13 \times 10\text{cm}^2$ field of view) allowed detection of very small movements.

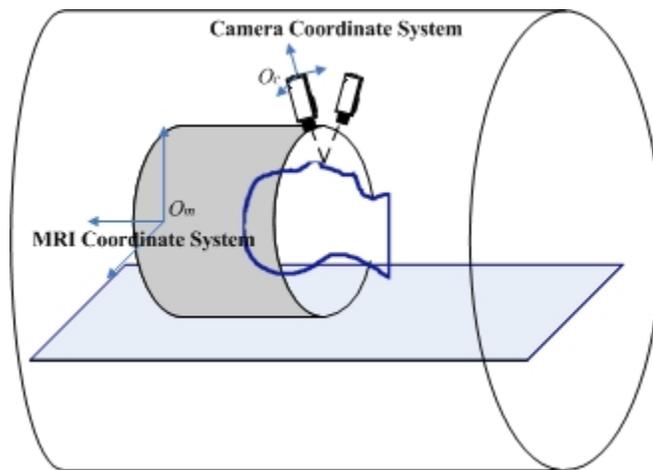




Figure 4-1 Experimental setup. The top pane is an illustration of the position of the cameras and head coil in the MRI scanner bore. The bottom images show the real instruments. Cameras are highlighted with the red ellipse.

The two cameras were connected to a filter box (MRC Systems GmbH, Germany) via the camera connector cable. This low pass filter, which has a cutoff frequency of 1MHz, was used to prevent interferences caused by the high frequency signals of the MR scanner, as well as possible infiltration of MR-frequency noise into the magnet room. The video output signals were then captured by a Matrox Morphis frame grabber (Matrox Electronic Systems Ltd. Quebec, Canada), using a standard BNC cable connection, by a tracking computer running a Linux operating system. Real-time tracking and motion estimation software was developed using C++. The tracking speed was 10Hz. The tracking computer communicated with the MRI scanner through a TCP/IP connection (Santos et al., 2004) to provide the console (scan computer) with the real-time motion parameters. This allowed the image acquisition pulse sequence to alter gradient direction and RF offset phases and frequencies to compensate for the changes in object position.

4.2 System Calibration

A precise calibration is necessary to achieve the goal of motion correction for high-resolution MRI. A three step calibration procedure was developed, which included the calibration for each single camera, the calibration for the stereo, and the calibration between the stereo cameras and the MRI scanner.

To achieve this, a calibration phantom was designed, the internal of which consisted of a grid of rectangular prisms (Figure 4-2a) carved out of a plastic form. The prisms were all of equal size, $10 \times 10 \times 18\text{mm}^3$ (length \times width \times height), and were $10 \times 10\text{mm}^2$ apart, except those on the four corners, which were cut into triangular prisms to provide differentiation in the height direction. The open spaces in the phantom were filled in with saline to make them visible to MRI. The cover of the phantom was 2.4mm thick. On the surface of the phantom, a piece of paper with a black-and-white checkerboard representation of the internal grid was attached (Figure 4-2b). This design provided many corners for calibration, and the clear sharp corners improved calibration accuracy.

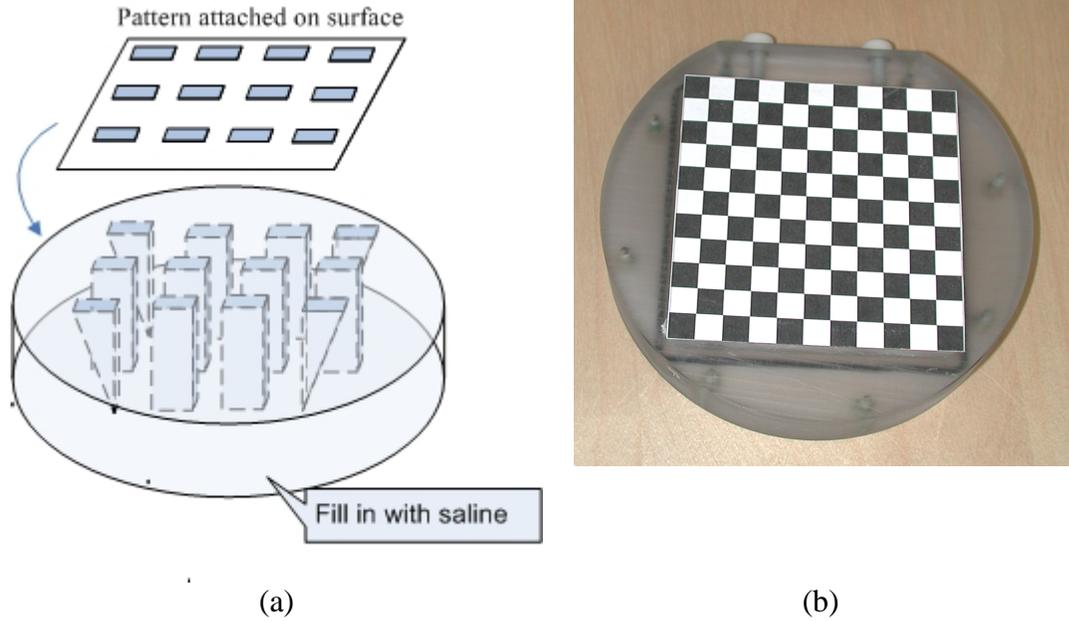


Figure 4-2 Calibration model. (a) Schematics of the phantom; (b) View of the actual phantom surface.

4.2.1 CCD Camera Calibration

Camera calibration is required to build a model that can map between a 3D world and a 2D image. The pinhole model is a widely used one and its geometric illustration is presented in Figure 4-3:

In this camera model, a 3D point $\mathbf{M}_w = [X_w, Y_w, Z_w]^T$, is transformed from the world coordinate system to the camera coordinate system, $\mathbf{M}_c = [X_c, Y_c, Z_c]^T$. Through central projection, the ideal (undistorted) image coordinate of the object point is $\mathbf{m}_u = [x_u, y_u]^T$. However, common cameras are often characterized by a somewhat distorted projection behavior which causes the observed (distorted) image point shifting to $\mathbf{m}_d = [x_d, y_d]^T$.

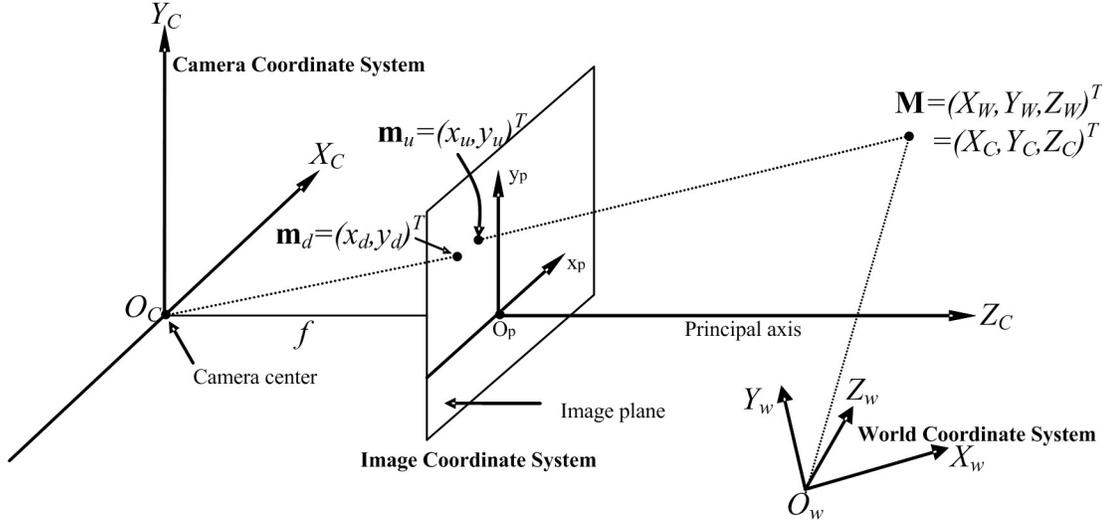


Figure 4-3 Pinhole camera geometry. O_C is the camera center and also the origin of the camera coordinate system, O_p is the principle point, M is a point in the space which has different coordinates in the world and camera coordinate systems, and \mathbf{m}_d and \mathbf{m}_u are the image of M in the image plane with and without lens distortion.

In practice, the image in the image plane will be further sampled by a frame grabber card and stored in an image memory buffer. The relationship between a point \mathbf{M}_w in the world coordinate system and its image \mathbf{m}_p in image pixel coordinate can be represented as

$$\mathbf{M}_c = \mathbf{R}(\mathbf{M}_w - \mathbf{C}) \quad [4.1]$$

$$\mathbf{m}_u = \begin{bmatrix} x_u \\ y_u \end{bmatrix} = f \begin{bmatrix} X_c / Z_c \\ Y_c / Z_c \end{bmatrix} = f \mathbf{m}_n = f \begin{bmatrix} x_n \\ y_n \end{bmatrix} \quad [4.2]$$

$$\mathbf{m}_d = \begin{bmatrix} x_d \\ y_d \end{bmatrix} = \begin{bmatrix} x_u (1 + k_1 r^2 + k_2 r^4) + 2p_1 x_u y_u + p_2 (r^2 + 2x_u^2) \\ y_u (1 + k_1 r^2 + k_2 r^4) + 2p_2 x_u y_u + p_1 (r^2 + 2y_u^2) \end{bmatrix} \quad [4.3]$$

Where $r^2 = x_u^2 + y_u^2$

$$\mathbf{m}_p = \begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} D_x x_d + x_0 \\ D_y y_d + y_0 \end{bmatrix} \quad [4.4]$$

\mathbf{R} and \mathbf{C} are rotation matrix and translation vector which characterize the transformation between the world coordinate system and the camera coordinate system; f is the focal length; \mathbf{m}_n is the normalized coordinate; k_1 and k_2 are the radial lens distortion coefficients; p_1 and p_2 are the tangential lens distortion coefficients; D_x, D_y are the number of pixels per metric unit distance (For most imaging sensors currently manufactured, pixels may be assumed perfectly square, implying $D_x=D_y$. In the general case, they may be different); (x_0, y_0) are the coordinates of the optical center (principal point) in the image coordinate system. The objective of the camera calibration procedure is to determine optimal values for these parameters based on image observations of a known calibration target. Since any calibration point \mathbf{M}_w and its corresponding image in the CCD sensor \mathbf{m}_p should satisfy Equation 4.1-4.4, the unknown parameters can be determined when the computed projection of a known structure (i.e. a checkerboard pattern) matches the best with the observed projection on the image. If enough (larger than the number of unknown camera parameters) calibration point pairs are given, the camera parameters can be estimated.

A closed-form solution for the unknown parameters without considering lens distortion gives a good initial guess. In absence of lens distortion, the projection of \mathbf{M}_w to \mathbf{m}_p can be simplified as:

$$\mathbf{m}_p = \begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} D_x f \frac{X_c}{Z_c} + x_0 \\ D_y f \frac{Y_c}{Z_c} + y_0 \end{bmatrix} = \begin{bmatrix} f_x x_n + x_0 \\ f_y y_n + y_0 \end{bmatrix} \quad [4.5]$$

Where $f_x = fD_x$ and $f_y = fD_y$ represent the focal length of the camera in terms of pixel dimensions in x and y direction.

Denoting the homogeneous coordinates of a vector $\mathbf{x} = [x, y, \Lambda]^T$ by $\tilde{\mathbf{x}}$, i.e.

$\tilde{\mathbf{x}} = [x, y, \Lambda, 1]^T$, Equation 4.1 and 4.5 can be written as:

$$\begin{aligned} s\tilde{\mathbf{m}}_p &= \begin{bmatrix} f_x & x_0 & 0 \\ & f_y & y_0 & 0 \\ & & 1 & 0 \end{bmatrix} \tilde{\mathbf{M}}_c \\ &= \mathbf{K} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R} & -\mathbf{RC} \\ 0 & 1 \end{bmatrix} \tilde{\mathbf{M}}_w = \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \tilde{\mathbf{M}}_w = \mathbf{P} \tilde{\mathbf{M}}_w \end{aligned} \quad [4.6]$$

Where s is an arbitrary scale; \mathbf{K} is a 3×3 matrix, mapping the normalized image coordinates to the image pixel coordinates; $\mathbf{T} = -\mathbf{RC}$; and \mathbf{P} is a 3×4 matrix, called the perspective projection matrix. In matrix \mathbf{P} , there are a total of 10 unknowns: 4 for \mathbf{K} , 3 for \mathbf{R} and 3 for \mathbf{C} . The parameters contained in \mathbf{K} are called the intrinsic camera parameters (distortion parameters are intrinsic parameters too, but are not estimated in this closed-form solution), and those in \mathbf{R} and \mathbf{C} are extrinsic parameters.

Vanishing points can be used for calibration for its simplicity and ease of computation (Bouquet, 1999; Daniilidis and Ernst, 1996; Wang and Tsai, 1991). Figure 4-4 shows a perspective image of a square ABCD in space. \mathbf{m}_{n1} , \mathbf{m}_{n2} , \mathbf{m}_{n3} and \mathbf{m}_{n4} are the four corners of the square observed in the image plane after normalization, which means:

$$\tilde{\mathbf{m}}_{ni} = s \begin{bmatrix} 1/f_x & & -x_0 \\ & 1/f_y & -y_0 \\ & & 1 \end{bmatrix} \tilde{\mathbf{m}}_{pi} \quad [4.7]$$

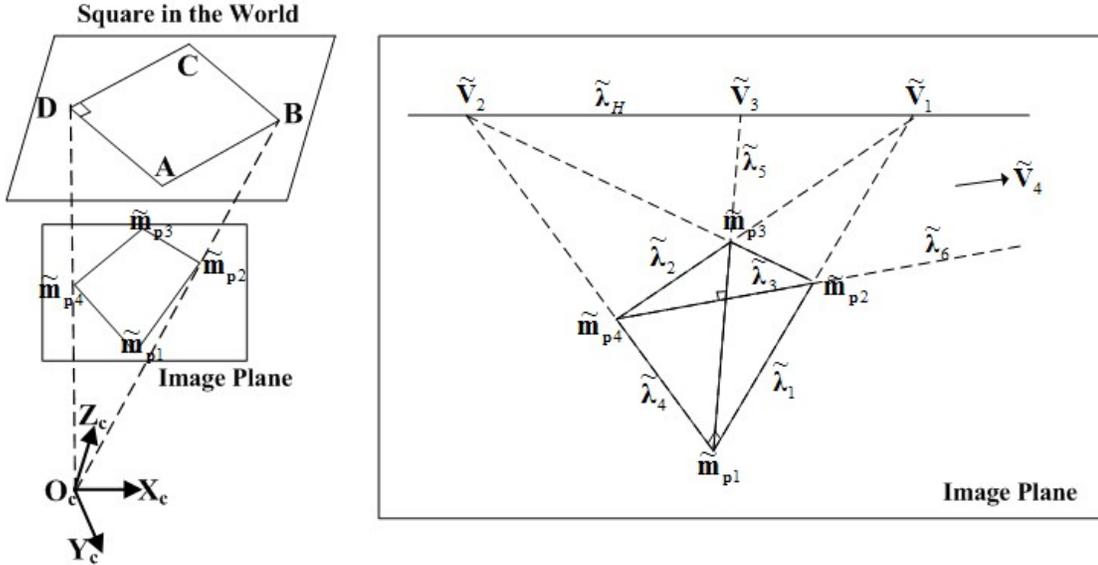


Figure 4-4 Camera calibration using a square target. (a) Projection of a square ABCD to the image plane; (b) The image and the four vanishing points $\{ \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4 \}$, and the horizon line $\tilde{\lambda}_H$.

By assuming the optical center is at the center of the image, the intrinsic parameters to be estimated are therefore f_x and f_y only, and Equation 4.7 can be simplified as:

$$\tilde{\mathbf{m}}_{ni} = s \begin{bmatrix} 1/f_x & 0 \\ 1/f_y & 0 \\ 1 & 1 \end{bmatrix} \tilde{\mathbf{p}}_i = s \mathbf{K}_p^{-1} \tilde{\mathbf{p}}_i \quad [4.8]$$

Where \mathbf{K}_p is the intrinsic camera matrix without principle points, and $\tilde{\mathbf{p}}_i (i = 1, \Lambda 4)$ is the homogeneous pixel coordinates of the corners after subtraction of the optical center.

From the geometry in Figure 4-4, the two vanishing points \tilde{V}_1 and \tilde{V}_2 can be determined by:

$$\left. \begin{array}{l} \tilde{\lambda}_1 \simeq \tilde{\mathbf{m}}_{n1} \times \tilde{\mathbf{m}}_{n2} \\ \tilde{\lambda}_2 \simeq \tilde{\mathbf{m}}_{n3} \times \tilde{\mathbf{m}}_{n4} \\ \tilde{\lambda}_3 \simeq \tilde{\mathbf{m}}_{n2} \times \tilde{\mathbf{m}}_{n3} \\ \tilde{\lambda}_4 \simeq \tilde{\mathbf{m}}_{n4} \times \tilde{\mathbf{m}}_{n1} \end{array} \right\} \Rightarrow \left. \begin{array}{l} \tilde{\mathbf{V}}_1 \simeq \tilde{\lambda}_1 \times \tilde{\lambda}_2 \\ \tilde{\mathbf{V}}_2 \simeq \tilde{\lambda}_3 \times \tilde{\lambda}_4 \end{array} \right\} \Rightarrow \tilde{\lambda}_H \simeq \tilde{\mathbf{V}}_1 \times \tilde{\mathbf{V}}_2 \quad [4.9]$$

Where \times is the standard cross product. All the equalities in Equation 4.9 are defined up to scale. However, the normalized coordinates of the corners are not available since the focal lengths are unknown. But the pixel coordinates of the four corners $\tilde{\mathbf{p}}_i (i = 1, \Lambda 4)$ are available, so Equation 4.9 should be written in pixel unit:

$$\left. \begin{array}{l} \tilde{\lambda}_1 \simeq (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_1) \times (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_2) \simeq \mathbf{K}_p^{-1*} \tilde{\lambda}_1^p \\ \tilde{\lambda}_2 \simeq (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_3) \times (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_4) \simeq \mathbf{K}_p^{-1*} \tilde{\lambda}_2^p \\ \tilde{\lambda}_3 \simeq (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_2) \times (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_3) \simeq \mathbf{K}_p^{-1*} \tilde{\lambda}_3^p \\ \tilde{\lambda}_4 \simeq (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_4) \times (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_1) \simeq \mathbf{K}_p^{-1*} \tilde{\lambda}_4^p \end{array} \right\} \Rightarrow \left. \begin{array}{l} \tilde{\mathbf{V}}_1 \simeq (\mathbf{K}_p^{-1*} \tilde{\lambda}_1^p) \times (\mathbf{K}_p^{-1*} \tilde{\lambda}_2^p) \simeq \mathbf{K}_p^{-1} \tilde{\mathbf{V}}_1^p \\ \tilde{\mathbf{V}}_2 \simeq (\mathbf{K}_p^{-1*} \tilde{\lambda}_3^p) \times (\mathbf{K}_p^{-1*} \tilde{\lambda}_4^p) \simeq \mathbf{K}_p^{-1} \tilde{\mathbf{V}}_2^p \end{array} \right\} \quad [4.10]$$

$$\Rightarrow \tilde{\lambda}_H \simeq \mathbf{K}_p^{-1*} \tilde{\lambda}_H^p$$

Where \mathbf{K}_p^* is the adjoint of \mathbf{K}_p . Since \mathbf{K} is a 3×3 matrix, and \mathbf{u}, \mathbf{v} are 3-vectors, we have $(\mathbf{K}\mathbf{u}) \times (\mathbf{K}\mathbf{v}) = \mathbf{K}^*(\mathbf{u} \times \mathbf{v})$ (Faugeras and Robert, 1996). Because \mathbf{K} is invertible, $\mathbf{K}^* = |\mathbf{K}|(\mathbf{K}^T)^{-1}$, therefore $\mathbf{K}^{**} \simeq \mathbf{K}$. And also explicitly,

$$\left. \begin{array}{l} \tilde{\lambda}_1^p \simeq \tilde{\mathbf{p}}_1 \times \tilde{\mathbf{p}}_2 \\ \tilde{\lambda}_2^p \simeq \tilde{\mathbf{p}}_3 \times \tilde{\mathbf{p}}_4 \\ \tilde{\lambda}_3^p \simeq \tilde{\mathbf{p}}_2 \times \tilde{\mathbf{p}}_3 \\ \tilde{\lambda}_4^p \simeq \tilde{\mathbf{p}}_4 \times \tilde{\mathbf{p}}_1 \end{array} \right\} \Rightarrow \left. \begin{array}{l} \tilde{\mathbf{V}}_1^p \simeq \tilde{\lambda}_1^p \times \tilde{\lambda}_2^p \\ \tilde{\mathbf{V}}_2^p \simeq \tilde{\lambda}_3^p \times \tilde{\lambda}_4^p \end{array} \right\} \Rightarrow \tilde{\lambda}_H \simeq \tilde{\mathbf{V}}_1^p \times \tilde{\mathbf{V}}_2^p \quad [4.11]$$

Because the two vanishing points $\tilde{\mathbf{V}}_1$ and $\tilde{\mathbf{V}}_2$ are mutually orthogonal:

$$\begin{aligned} \tilde{\mathbf{V}}_1 \perp \tilde{\mathbf{V}}_2 &\Leftrightarrow (\mathbf{K}_p^{-1} \tilde{\mathbf{V}}_1^p) \perp (\mathbf{K}_p^{-1} \tilde{\mathbf{V}}_2^p) \\ &\Leftrightarrow (\tilde{\mathbf{V}}_1^p)^T (\mathbf{K}_p^{-1T} \mathbf{K}_p^{-1}) \tilde{\mathbf{V}}_2^p = 0 \end{aligned} \quad [4.12]$$

In Equation 4.12, $\tilde{\mathbf{V}}_1^p$ and $\tilde{\mathbf{V}}_2^p$ can be calculated from Equation 4.11, and are denoted here as $\tilde{\mathbf{V}}_1^p \simeq [a_1, b_1, c_1]^T$ and $\tilde{\mathbf{V}}_2^p \simeq [a_2, b_2, c_2]^T$, so Equation 4.12 provides one constraint in the focal lengths:

$$\frac{a_1 a_2}{f_x^2} + \frac{b_1 b_2}{f_y^2} + c_1 c_2 = 0 \quad [4.13]$$

Another constraint for the focal lengths comes from the fact that the diagonals of the square ABCD are orthogonal too. Therefore, similarly, we get:

$$\left. \begin{aligned} \tilde{\boldsymbol{\lambda}}_5 &\simeq (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_1) \times (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_3) \simeq \mathbf{K}_p^{-1*} \tilde{\boldsymbol{\lambda}}_5^p \\ \tilde{\boldsymbol{\lambda}}_6 &\simeq (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_2) \times (\mathbf{K}_p^{-1} \tilde{\mathbf{p}}_4) \simeq \mathbf{K}_p^{-1*} \tilde{\boldsymbol{\lambda}}_6^p \end{aligned} \right\} \Rightarrow \begin{cases} \tilde{\mathbf{V}}_3 \simeq (\mathbf{K}_p^{-1*} \tilde{\boldsymbol{\lambda}}_5^p) \times (\mathbf{K}_p^{-1*} \tilde{\boldsymbol{\lambda}}_6^p) \simeq \mathbf{K}_p^{-1} \tilde{\mathbf{V}}_3^p \\ \tilde{\mathbf{V}}_4 \simeq (\mathbf{K}_p^{-1*} \tilde{\boldsymbol{\lambda}}_6^p) \times (\mathbf{K}_p^{-1*} \tilde{\boldsymbol{\lambda}}_5^p) \simeq \mathbf{K}_p^{-1} \tilde{\mathbf{V}}_4^p \end{cases} \quad [4.14]$$

Then, since

$$\begin{aligned} \tilde{\mathbf{V}}_3 \perp \tilde{\mathbf{V}}_4 &\Leftrightarrow (\mathbf{K}_p^{-1} \tilde{\mathbf{V}}_3^p) \perp (\mathbf{K}_p^{-1} \tilde{\mathbf{V}}_4^p) \\ &\Leftrightarrow (\tilde{\mathbf{V}}_3^p)^T (\mathbf{K}_p^{-1T} \mathbf{K}_p^{-1}) \tilde{\mathbf{V}}_4^p = 0 \end{aligned} \quad [4.15]$$

We get another constraint for focal lengths with denoting $\tilde{\mathbf{V}}_3^p \simeq [a_3, b_3, c_3]^T$, and

$$\tilde{\mathbf{V}}_4^p \simeq [a_4, b_4, c_4]^T :$$

$$\frac{a_3 a_4}{f_x^2} + \frac{b_3 b_4}{f_y^2} + c_3 c_4 = 0 \quad [4.16]$$

So, f_x and f_y are the solution of Equation 4.13 and 4.16.

Next, the extrinsic parameters can be initialized through homography (Zhang, 2000). Without loss of generality, the calibration pattern plane is assumed to be on $Z_w = 0$ plane. Denote the i^{th} column of the rotation matrix \mathbf{R} by \mathbf{r}_i . From Equation 4.6, we have

$$s \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{T} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \quad [4.17]$$

Therefore, a calibration point $\tilde{\mathbf{M}}_w$ (since Z_w is 0, $\tilde{\mathbf{M}}_w = [X_w, Y_w, 1]^T$) and its image $\tilde{\mathbf{m}}_p$ is related by a homography \mathbf{H} :

$$s\tilde{\mathbf{m}}_p = \mathbf{H}\tilde{\mathbf{M}}_w \quad \text{with} \quad \mathbf{H} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} \quad [4.18]$$

To calculate the homography, we first define a 9×1 vector \mathbf{H}_a as $[\bar{\mathbf{h}}_1^T, \bar{\mathbf{h}}_2^T, \bar{\mathbf{h}}_3^T]^T$, where $\bar{\mathbf{h}}_i^T$ ($i=1,2,3$) is the i^{th} row of \mathbf{H} . Then Equation 4.18 can be written as

$$\begin{aligned} s \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} &= \begin{bmatrix} \bar{\mathbf{h}}_1^T \\ \bar{\mathbf{h}}_2^T \\ \bar{\mathbf{h}}_3^T \end{bmatrix} \tilde{\mathbf{M}}_w \Rightarrow \begin{cases} sx_p - \bar{\mathbf{h}}_1^T \tilde{\mathbf{M}}_w = 0 \\ sy_p - \bar{\mathbf{h}}_2^T \tilde{\mathbf{M}}_w = 0 \\ s = \bar{\mathbf{h}}_3^T \tilde{\mathbf{M}}_w \end{cases} \\ &\Rightarrow \begin{bmatrix} \tilde{\mathbf{M}}_w^T & \mathbf{0}^T & -x_p \tilde{\mathbf{M}}_w^T \\ \mathbf{0}^T & \tilde{\mathbf{M}}_w^T & -y_p \tilde{\mathbf{M}}_w^T \end{bmatrix} \mathbf{H}_a = 0 \end{aligned} \quad [4.19]$$

When all the corners in our calibration checkerboard pattern were used (Figure 4-2b), which was 100, we had 100 of the above equations. These can be written as $\mathbf{L}\mathbf{H}_a = \mathbf{0}$, where \mathbf{L} is a 200×9 matrix. Therefore, the solution of \mathbf{H}_a is the right singular vector of \mathbf{L} associated with the smallest singular value. Note that \mathbf{H} is defined up to a scale factor. Denoting \mathbf{H} by its column vector $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3]$, from Equation 4.18, we get

$$\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} = \lambda \mathbf{K}^{-1} \mathbf{H} \Rightarrow \begin{cases} \mathbf{r}_1 = \lambda \mathbf{K}^{-1} \mathbf{h}_1 \\ \mathbf{r}_2 = \lambda \mathbf{K}^{-1} \mathbf{h}_2 \\ \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{T} = \lambda \mathbf{K}^{-1} \mathbf{h}_3 \end{cases} \quad [4.20]$$

To assure \mathbf{r}_1 and \mathbf{r}_2 are unit vectors, $\lambda = \left(\frac{1}{\|\mathbf{K}^{-1}\mathbf{h}_1\|} + \frac{1}{\|\mathbf{K}^{-1}\mathbf{h}_2\|} \right) / 2$ (or choose either

$\lambda = \frac{1}{\|\mathbf{K}^{-1}\mathbf{h}_1\|}$ or $\lambda = \frac{1}{\|\mathbf{K}^{-1}\mathbf{h}_2\|}$ since they are very close). However, because of noise,

the computed $\mathbf{R} = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]$ does not in general satisfy the properties of a rotation matrix, therefore we need to do the following:

$$\begin{aligned}
 \mathbf{r}_1 &\leftarrow \mathbf{r}_1 / \|\mathbf{r}_1\| \\
 \mathbf{r}_2 &\leftarrow \mathbf{r}_2 - \langle \mathbf{r}_1, \mathbf{r}_2 \rangle \mathbf{r}_1 \\
 \mathbf{r}_2 &\leftarrow \mathbf{r}_2 / \|\mathbf{r}_2\| \\
 \mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2
 \end{aligned} \tag{4.21}$$

Where “ \leftarrow ” means “to be replaced by”, and \langle, \rangle is the standard inner product.

The solution of the intrinsic parameters from Equation 4.9 – 4.16, and the extrinsic parameters from Equation 4.18 – 4.21 assumes no lens distortion of the cameras. However, the MR compatible cameras used in this project exhibited significant lens distortion, especially radial distortion. So to estimate all the camera parameters, we need to use the already calculated parameters as an initial guess for the minimization of the nonlinear cost function:

$$\sum_{i=1}^M \sum_{j=1}^N \left\| \mathbf{m}_{p_{ij}} - \mathbf{m}_p(f_x, f_y, x_0, y_0, k_1, k_2, p_1, p_2, \mathbf{R}_i, \mathbf{T}_i, \mathbf{M}_{w_j}) \right\|^2 \tag{4.22}$$

Where $\mathbf{m}_p(f_x, f_y, x_0, y_0, k_1, k_2, p_1, p_2, \mathbf{R}_i, \mathbf{T}_i, \mathbf{M}_{w_j})$ is the projection of point \mathbf{M}_{w_j} from the world coordinate system into the i^{th} calibration image according to Equation 4.1 – 4.4. The 3×3 rotation matrix \mathbf{R}_i has 3 DOF only and is converted to a 3×1 vector through Rodrigues formula for computation efficiency (Faugeras, 1993). $\mathbf{m}_{p_{ij}}$ is the measured pixel coordinate. M is the total number of views used for calibration

and N is the number of corners on the calibration target in each view. It has been shown that calibration errors decrease when more images are used (Zhang, 2000), so 20 ~ 30 calibration images were used in our project. Figure 4-5 shows two calibration images, with red “+” marking the $N=100$ corners (Harris and Stephens, 1988). The nonlinear minimization is solved with sparse Gauss-Newton algorithm (Hartley and Zisserman, 2003, Section 9.3).

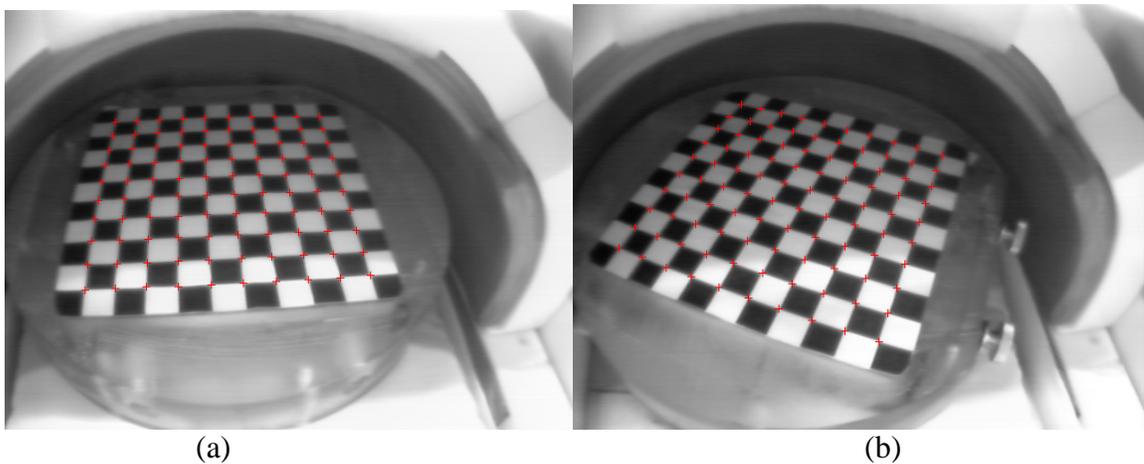


Figure 4-5 Two views of the calibration target. The 100 red crosses on each image mark the extracted corners used for calibration.

4.2.2 Stereo Calibration

After calibration of both cameras, the relationship between the two cameras can be estimated. Let \mathbf{O}_c and \mathbf{O}_c' be the camera centers of the left and right cameras and \mathbf{I} , \mathbf{I}' the corresponding image planes respectively (Figure 4-6). Given a point \mathbf{M} in the world, its projection on the left image is \mathbf{m} , and on the right image is \mathbf{m}' . For every point on the line $\mathbf{O}_c\mathbf{M}$, their projections in the left image are all \mathbf{m} , but their projections in the right image are on a line \mathbf{I}'_m , which is called the epipolar line of \mathbf{m} . The line \mathbf{I}'_m is the intersection of the plane \mathbf{I}' with the plane X , defined by \mathbf{m} , \mathbf{O}_c and

O_c' , which is called the epipolar plane. Because the epipolar geometry is symmetric, point m' in the right image corresponds to an epipolar line l_m in the left image, which is the intersection of the same plane X with the left image plane I . Furthermore, all epipolar planes (with different M) contain the line $O_c O_c'$. They all intersect I at a common point e and intersect I' also at a common point e' , which are called epipoles.

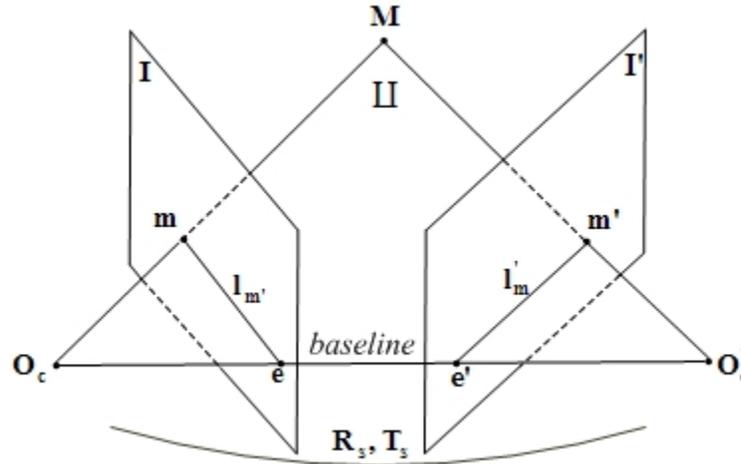


Figure 4-6 The epipolar geometry. M is a point in space, O_c and O_c' are the two camera centers. The camera baseline $O_c O_c'$ intersects each image plane at the epipoles e and e' . O_c , O_c' , M and its images m and m' all lie in a common plane, called epipolar plane. The two lines me and $m'e$ are epipolar lines.

The epipolar geometry provides a computational significance in stereo vision that for a point in one image, its correspondence in the other image must lie on the epipolar line, and therefore the search space for a correspondence is reduced from 2D to 1D. But to get that, the rotation R_s and translation T_s between the two camera coordinate systems need to be determined.

Given a set of points in space M_{ij} ($i=1, \dots, M$ represents the number of images used for calibration, $j=1, \dots, N$ represents the number of corners in each calibration image), and let $M_{cij} = [X_{cij}, Y_{cij}, Z_{cij}]^T$, $M'_{cij} = [X'_{cij}, Y'_{cij}, Z'_{cij}]^T$ be the Euclidean coordinates of M_{ij} in the two camera coordinate systems, then

$$\mathbf{M}'_{cij} = \mathbf{R}_s \mathbf{M}_{cij} + \mathbf{T}_s \quad [4.23]$$

From each camera's calibration step, we get the extrinsic parameters. Also, since the two cameras get images of the same object simultaneously, they have the relation:

$$\left. \begin{aligned} \mathbf{M}_{cij} &= \mathbf{R}_i \mathbf{M}_{wj} + \mathbf{T}_i \\ \mathbf{M}'_{cij} &= \mathbf{R}'_i \mathbf{M}_{wj} + \mathbf{T}'_i \end{aligned} \right\} \Rightarrow \mathbf{M}'_{cij} = \mathbf{R}'_i \mathbf{R}_i^{-1} \mathbf{M}_{cij} + \mathbf{T}'_i - \mathbf{R}'_i \mathbf{R}_i^{-1} \mathbf{T}_i \quad [4.24]$$

Therefore, with the M calibration images, we get M rotation and translation estimations. The median of these is used as the initial guess to search for the optimal \mathbf{R}_s and \mathbf{T}_s by minimizing the following cost function:

$$\sum_{i=1}^M \sum_{j=1}^N \left\| \mathbf{m}_{p_{ij}} - \mathbf{m}_p(\mathbf{K}, k_1, k_2, p_1, p_2, \mathbf{K}', k'_1, k'_2, p'_1, p'_2, \mathbf{R}_i, \mathbf{T}_i, \mathbf{R}_s, \mathbf{T}_s, \mathbf{M}_{wj}) \right\|^2 \quad [4.25]$$

Where $\mathbf{m}_{p_{ij}} = [\mathbf{m}_{p_{ij}}^T, \mathbf{m}_{p_{ij}}^T]^T$ is the observed corner position in both the left and right images; \mathbf{M}_{wj} is the point in world coordinate system; $\mathbf{m}_p(\mathbf{K}, k_1, k_2, p_1, p_2, \mathbf{K}', k'_1, k'_2, p'_1, p'_2, \mathbf{R}_i, \mathbf{T}_i, \mathbf{R}_s, \mathbf{T}_s, \mathbf{M}_{wj})$ is the projection of a space point \mathbf{M}_{wj} to the two cameras by combining Equation 4.1 – 4.4 and 4.25; \mathbf{K} and \mathbf{K}' are the calibration matrix of the two cameras, including f_x, f_y, x_0, y_0 and f'_x, f'_y, x'_0, y'_0 ; k_1, k_2, p_1, p_2 and k'_1, k'_2, p'_1, p'_2 are distortion parameters; \mathbf{R}_i , and \mathbf{T}_i are the extrinsic parameters corresponding to the i^{th} calibration image captured by one camera. The nonlinear minimization is also solved with the sparse Gauss-Newton algorithm (Hartley and Zisserman, 2003). Note that in Equation 4.25, all the intrinsic parameters of the two cameras are to be refined even though they have been determined already after calibration of each camera. The reason to do it here is that

the epipolar constraint can provide some refinement on these parameters for better calibration accuracy.

4.2.3 3D Reconstruction from Stereo

With all the parameters of the stereo known, from any corresponding image points \mathbf{m} and \mathbf{m}' in the two camera images, we can retrieve the 3D coordinate of that point based on stereo triangulation (Hartley and Zisserman, 2003; Hartley and Sturm, 1997; Bouguet, 1999). From epipolar geometry in Figure 4-6, \mathbf{m} , \mathbf{m}' , \mathbf{O}_c and \mathbf{O}_c' must lie in the epipolar plane, therefore the back-projection of the two lines $\mathbf{O}_c\mathbf{m}$ and $\mathbf{O}_c'\mathbf{m}'$ intersect in 3D at the point \mathbf{M} . The computation of a 3D point position from its image in two 2D views is called stereo triangulation.

Let $\mathbf{m}_n = [x_n, y_n]^T$ and $\mathbf{m}'_n = [x'_n, y'_n]^T$ be the normalized coordinates for points \mathbf{m} and \mathbf{m}' respectively, calculated using Equation 4.2 – 4.4 after the two cameras being calibrated. Let $\mathbf{M}_c = [X_c, Y_c, Z_c]^T$ and $\mathbf{M}'_c = [X'_c, Y'_c, Z'_c]^T$ be the coordinates of \mathbf{M} in the two camera coordinate systems respectively, from Equation 4.2, we can get:

$$\mathbf{M}_c = Z_c \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = Z_c \tilde{\mathbf{m}}_n, \quad \mathbf{M}'_c = Z'_c \begin{pmatrix} x'_n \\ y'_n \\ 1 \end{pmatrix} = Z'_c \tilde{\mathbf{m}}'_n \quad [4.26]$$

With the known \mathbf{R}_s and \mathbf{T}_s from stereo calibration, and $\mathbf{M}'_c = \mathbf{R}_s \mathbf{M}_c + \mathbf{T}_s$ from Equation 4.23, we get:

$$Z'_c \tilde{\mathbf{m}}'_n = Z_c \mathbf{R}_s \tilde{\mathbf{m}}_n + \mathbf{T}_s \quad [4.27]$$

Since Z_c and Z'_c are the unknowns, the above equation can be rewritten as:

$$\begin{bmatrix} -\mathbf{R}_s \tilde{\mathbf{m}}_n & \tilde{\mathbf{m}}'_n \end{bmatrix} \begin{bmatrix} Z_c \\ Z'_c \\ Z_c \end{bmatrix} = \mathbf{T}_s \quad [4.28]$$

Denote $\mathbf{A} = [-\mathbf{R}_s \tilde{\mathbf{m}}_n \quad \tilde{\mathbf{m}}'_n]$ (a 3×2 matrix), the least square solution is then:

$$\begin{bmatrix} Z_c \\ Z'_c \\ Z_c \end{bmatrix} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{T}_s \quad [4.29]$$

If there is no noise in \mathbf{m} and \mathbf{m}' , Equation 4.29 returns the exact 3D position of the point \mathbf{M} , which is the intersection of the two back-projecting rays $\mathbf{O}_c \mathbf{m}$ and $\mathbf{O}_c' \mathbf{m}'$. However, if noise exists, the two rays don't intersect in general, it is thus necessary to estimate a best solution for the point in 3D world. This can be realized by Sampson correction of the measured points (Hartley and Zisserman, 2003). But before Sampson correction, the epipolar constraint needs to be reiterated here.

For Equation 4.27, by taking the cross product with \mathbf{T}_s , followed by the inner product with $\tilde{\mathbf{m}}'_n$, we obtain $\tilde{\mathbf{m}}'_n \cdot (\mathbf{T}_s \times \mathbf{R}_s \tilde{\mathbf{m}}_n) = 0$. This can also be written as:

$$\tilde{\mathbf{m}}_n^T \mathbf{E} \tilde{\mathbf{m}}_n = 0 \quad \text{where } \mathbf{E} = [\mathbf{T}_s]_{\times} \mathbf{R}_s \quad [4.30]$$

\mathbf{E} is called the essential matrix, and $[\mathbf{T}_s]_{\times}$ is the skew-symmetric matrix of \mathbf{T}_s , defined

as:

$$[\mathbf{T}_s]_{\times} = \begin{bmatrix} 0 & -T_{s3} & T_{s2} \\ T_{s3} & 0 & -T_{s1} \\ -T_{s2} & T_{s1} & 0 \end{bmatrix} \quad [4.31]$$

Equation 4.30 mathematically describes the epipolar constraint illustrated in

Figure 4-6. In particular it means that $\tilde{\mathbf{m}}'_n$ lies on the epipolar line $\mathbf{l}'_m = \mathbf{E} \tilde{\mathbf{m}}_n$.

A typical observation of a noisy point correspondence $\mathbf{m}_n \leftrightarrow \mathbf{m}'_n$ doesn't in general satisfy the epipolar constraint. They have to be corrected to $\mathbf{m}_{nc} \leftrightarrow \mathbf{m}'_{nc}$ which have minimum distance to the measured $\mathbf{m}_n \leftrightarrow \mathbf{m}'_n$ but satisfy the epipolar constraint $\tilde{\mathbf{m}}_{nc}^T \mathbf{E} \tilde{\mathbf{m}}_{nc} = 0$ exactly. To emphasize the dependency on $\mathbf{x} = [\mathbf{m}_n^T, \mathbf{m}'_n{}^T]^T$, we will write the epipolar constraint as $C_E(\mathbf{x}) = 0$. This cost function may be approximated by a Taylor expansion to first order:

$$C_E(\mathbf{x} + \boldsymbol{\delta}_x) = C_E(\mathbf{x}) + \frac{\partial C_E}{\partial \mathbf{x}} \boldsymbol{\delta}_x \quad [4.32]$$

The corrected $\mathbf{x}_c = [\mathbf{m}_{nc}^T, \mathbf{m}'_{nc}{}^T]^T$ satisfy the constraint, i.e. $C_E(\mathbf{x}_c) = 0$. Since

$\mathbf{x}_c = \mathbf{x} + \boldsymbol{\delta}_x$, we have $C_E(\mathbf{x}) + \frac{\partial C_E}{\partial \mathbf{x}} \boldsymbol{\delta}_x = 0$. Denote the partial derivative matrix

(Jacobian matrix) $\frac{\partial C_E}{\partial \mathbf{x}}$ as \mathbf{J} and the cost $C_E(\mathbf{x})$ as ε , the Sampson correction term

$\boldsymbol{\delta}_x$ is thus the solution of equation $\mathbf{J} \boldsymbol{\delta}_x = -\varepsilon$:

$$\boldsymbol{\delta}_x = -\mathbf{J}^T (\mathbf{J} \mathbf{J}^T)^{-1} \varepsilon \quad [4.33]$$

The corrected point is therefore:

$$\mathbf{x}_c = \mathbf{x} - \mathbf{J}^T (\mathbf{J} \mathbf{J}^T)^{-1} \varepsilon \quad [4.34]$$

Explicitly, the cost here is $\varepsilon = \tilde{\mathbf{m}}_n^T \mathbf{E} \tilde{\mathbf{m}}_n$, and the Jacobian matrix is:

$$\mathbf{J} = \frac{\partial \varepsilon}{\partial \mathbf{x}} = [(\mathbf{E}^T \tilde{\mathbf{m}}_n)_1, (\mathbf{E}^T \tilde{\mathbf{m}}_n)_2, (\mathbf{E} \tilde{\mathbf{m}}_n)_1, (\mathbf{E} \tilde{\mathbf{m}}_n)_2] \quad [4.35]$$

The elements of \mathbf{J} in the previous equation are computed by the chain rule as:

$$\mathbf{J}_1 = \frac{\partial \varepsilon}{\partial x_n} = \frac{\partial \tilde{\mathbf{m}}_n^T \mathbf{E} \tilde{\mathbf{m}}_n}{\partial x_n} = \frac{\partial \tilde{\mathbf{m}}_n}{\partial x_n} \frac{\partial \tilde{\mathbf{m}}_n^T \mathbf{E} \tilde{\mathbf{m}}_n}{\partial \tilde{\mathbf{m}}_n} = (1, 0, 0) \mathbf{E}^T \tilde{\mathbf{m}}_n = (\mathbf{E}^T \tilde{\mathbf{m}}_n)_1 \quad [4.36]$$

Then from Equation 4.34 the corrected point is:

$$\begin{bmatrix} x_{nc} \\ y_{nc} \\ x'_{nc} \\ y'_{nc} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \\ x'_n \\ y'_n \end{bmatrix} - \frac{\tilde{\mathbf{m}}_n^T \mathbf{E} \tilde{\mathbf{m}}_n}{(\mathbf{E}^T \tilde{\mathbf{m}}'_n)_1^2 + (\mathbf{E}^T \tilde{\mathbf{m}}'_n)_2^2 + (\mathbf{E} \tilde{\mathbf{m}}_n)_1^2 + (\mathbf{E} \tilde{\mathbf{m}}_n)_2^2} \begin{bmatrix} (\mathbf{E}^T \tilde{\mathbf{m}}'_n)_1 \\ (\mathbf{E}^T \tilde{\mathbf{m}}'_n)_2 \\ (\mathbf{E} \tilde{\mathbf{m}}_n)_1 \\ (\mathbf{E} \tilde{\mathbf{m}}_n)_2 \end{bmatrix} \quad [4.37]$$

After Sampson error correction, the 3D position of a point can be estimated using Equation 4.29. But since Sampson correction is a first-order correction, it is accurate only if the noise in each image is small, making it a suboptimal solution. The optimal solution could be found by minimizing the function:

$$C_E(\mathbf{m}_n, \mathbf{m}'_n) = d(\mathbf{m}_n, \mathbf{m}_{nc})^2 + d(\mathbf{m}'_n, \mathbf{m}'_{nc})^2 \quad \text{subject to} \quad \tilde{\mathbf{m}}_{nc}^T \mathbf{E} \tilde{\mathbf{m}}_{nc} = 0 \quad [4.38]$$

Where $d(*,*)$ is the Euclidean distance between the points. This cost function can be minimized using Newton iteration algorithm, but it is more computationally expensive. The Sampson error correction therefore gives a close solution but a much cheaper computation.

4.2.4 Integrated Optical and MRI System Calibration

The optical tracking system measured the motion parameters in the camera coordinate system. These needed to be converted to the MRI coordinate system to be used by the imaging sequence. Therefore, an accurate calibration was necessary to estimate the rotation (\mathbf{R}_{mc}) and translation (\mathbf{T}_{mc}) between the camera and MRI coordinates.

To achieve this, the phantom was put in the target field (approximately the position of a subject's face) within the field of view of the scanner and video cameras (Figure 4-7). A 3D image of the phantom was acquired by the MRI scanner (resolution $0.4 \times 0.4 \times 0.7 \text{ mm}^3$, TR=55ms, TE=6ms, FA=10°, Slice Number=40) and

the stereo measured the 3D position of the calibration corners on the surface pattern based on stereo triangulation including distortion correction for the cameras (Section 4.2.3).

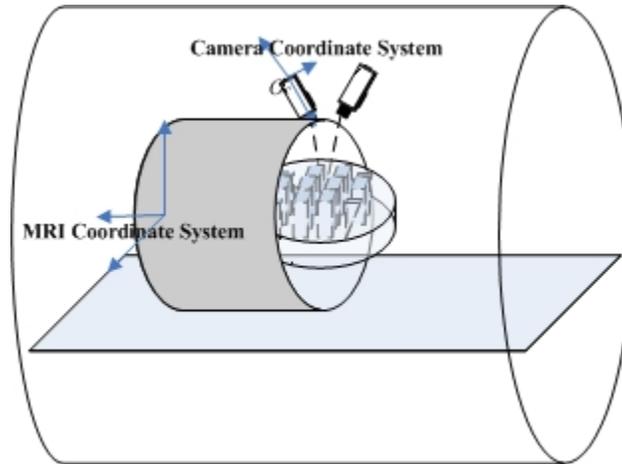


Figure 4-7 Calibration between the optical system and MRI scanner. The calibration phantom was put in the field of view of both the cameras and the MRI scanner.

To estimate the surface corner positions in MRI coordinates, the internal structure needed to be extracted. Figure 4-8 shows two slices from the 3D scan of the calibration phantom. We used corners to describe the internal structure, which when connected are the sides of prisms. Harris corner detection (Harris and Stephens, 1988) was applied on these MR images. A potential problem may be caused by air bubbles inside the phantom, i.e. the blue arrow in Figure 4-8a, which could result in a false corner. To correct this, lines were fitted on the corners and the crossings of these lines were marked as real corners, shown with red “+” in Figure 4-8. The MRI coordinates of these corners can be determined from their pixel coordinates and the slice position information from DICOM image header (Section 9.2).

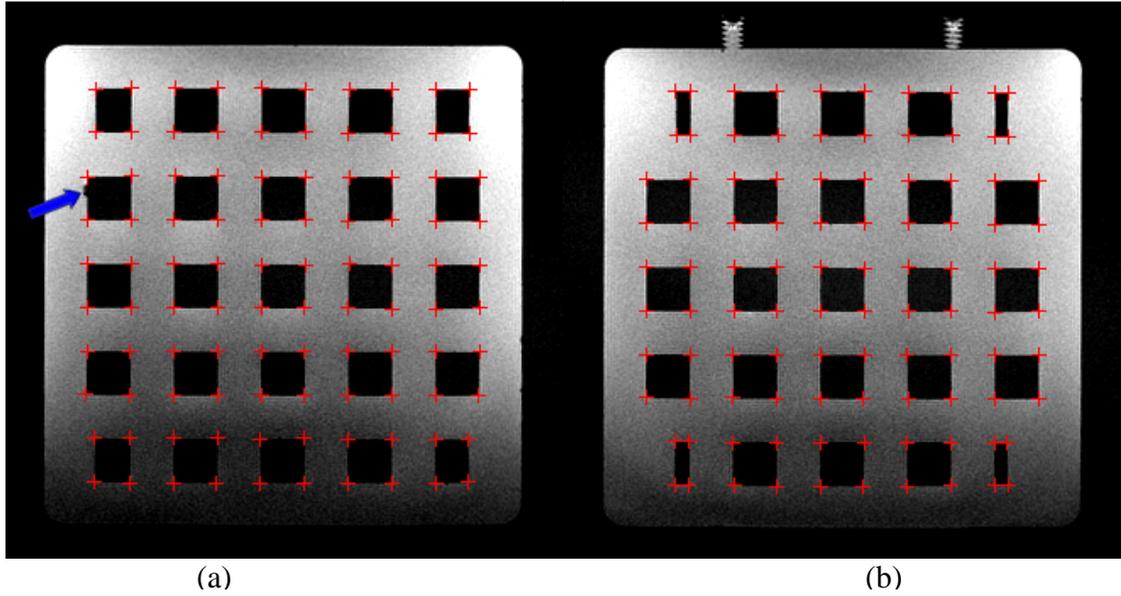


Figure 4-8 Two images from the 3D scan of the calibration phantom. The red crossings mark the corners. The blue arrow marks an air bubble inside the phantom.

For each side of the prism, we have forty positions in MRI coordinates since we scanned 40 slices. They were used to estimate the corresponding corner position on the phantom surface. Figure 4-9a shows the estimation of one side of a prism from the extracted corners. Blue dots were the observed corners (Figure 4-8), which obviously did not lie on a line because of noise. To avoid this, a line that minimizes the perpendicular distances from each of the points to the line needs to be fitted. Principal Components Analysis (PCA) is a good solution for 3D line fitting. Denote all the blue dots as $\mathbf{M}_{mi}=[X_{mi}, Y_{mi}, Z_{mi}]^T$ and $\mathbf{M}_m=[\mathbf{M}_{m1}, \mathbf{M}_{m2}, \dots, \mathbf{M}_{mk}]$ ($i=1,2,\dots,k$) where k is the total number of slices (k equals to 40 here). Principal components ($\mathbf{P}=[\mathbf{p}_1 \mathbf{p}_2 \mathbf{p}_3]$) of \mathbf{M}_m are the eigenvectors of the covariant matrix of \mathbf{M}_m :

$$\mathbf{M}_m \mathbf{M}_m^T = \mathbf{P}^T \mathbf{D} \mathbf{P} \quad [4.39]$$

Since \mathbf{M}_m is of size $3 \times k$, PCA generates three orthogonal components. The first principal component (denote as \mathbf{p}_1) gives the direction vector of the line because

it explains as much of the variance in the points as is possible with one dimension (green line in Figure 4-9a). The second and third components (\mathbf{p}_2 and \mathbf{p}_3) define directions perpendicular to the line, and are the error term in the fitting (blue lines in Figure 4-9a). The function of the fitted line is therefore:

$$\mathbf{l}_f = \text{mean}(\mathbf{M}_n) + t\mathbf{p}_1 \quad [4.40]$$

Where t is the free variable for the line function. Then the position of the corresponding corner on the surface can be estimated because the thickness of the phantom cover is known as 2.4mm (red circle in Figure 4-9a).

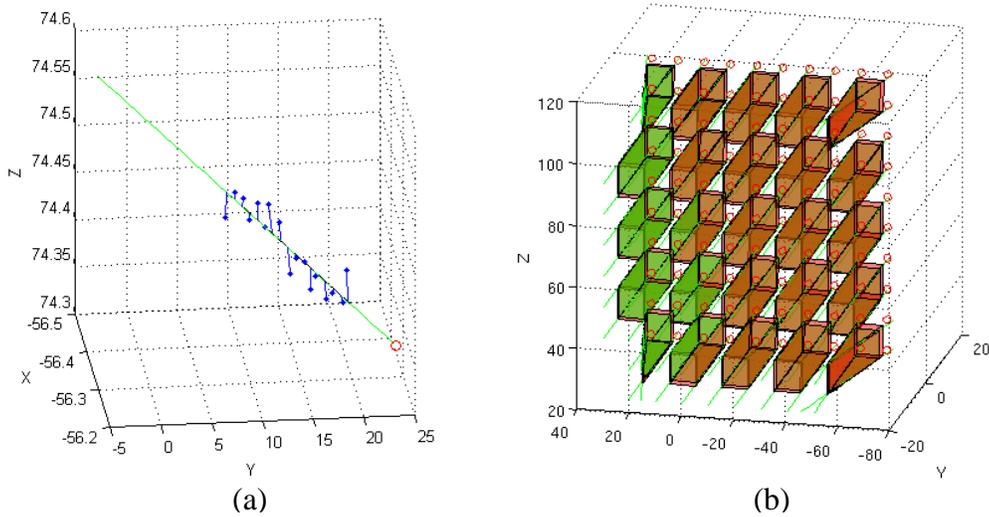


Figure 4-9 Phantom corner position estimation in MRI coordinate system. (a) shows one side of a prism. Blue dots are the observed corner position along that side; green line is the estimated side and the red circle is the estimated corner position on the phantom surface. (b) shows the alignment between the extracted internal structure (green) of the phantom and a simulated ideal model (red). The green lines are the side of the prisms and the red circles are the estimated corners position in MRI (all dimensions are in mm).

By performing the line fitting on each side of all the prisms inside the phantom, one can estimate all the corner positions on the surface. These corners do not lie in a plane exactly because of noise. So they were further registered to a simulated phantom surface (checkerboard pattern). The corners on the simulated

phantom were transformed into the MRI coordinate system and were the corrected positions used for calibration. Figure 4-9b shows the results after alignment. Red prisms represent the simulated phantom, which has $10 \times 10 \times 18 \text{ mm}^3$ rectangular prisms in the middle and triangular prisms on the four corners. Green prisms are structures extracted from the MRI 3D image. Red circles represent the calibration corner positions (on the camera-visible checkerboard grid, 2.4 mm from the liquid surface).

Therefore, based on the simultaneous measurements of the same points in both the MRI scanner and stereo coordinates, the rotation (a 3×3 orthonormal matrix \mathbf{R}_{mc}) and translation (a 3×1 vector \mathbf{T}_{mc}) between them can be calculated using a singular value decomposition (SVD) method (Section 4.3.3).

4.3 Real-time Motion Calculation

4.3.1 Kanade-Lucas-Tomasi Tracker

The Kanade-Lucas-Tomasi (KLT) tracker (Lucas and Kanade, 1981; Tomasi and Kanatani, 1991; Shi and Tomasi, 1994) is a feature tracker widely used in the computer vision community. It offers sub-pixel accuracy and is very computationally efficient, therefore is utilized for the head tracking in this project.

Denote $I(\mathbf{x}, t)$ as the image intensity, where \mathbf{x} is the position of a point, and t is time. As the object moves, the image intensity changes in a complex way, but if the camera grabbing frequency is sufficiently high, or the movement between successive images is sufficiently small, we can model the two successive images as:

$$I(\mathbf{x} + \mathbf{d}, t + \tau) = I(\mathbf{x}, t) + n(\mathbf{x}) \quad [4.41]$$

Where τ is the time difference between sampling of the two frames; \mathbf{d} is the displacement vector $(d_x, d_y)^T$; and n is noise.

The displacement vector \mathbf{d} is the one that can minimize the residual error:

$$\varepsilon(\mathbf{x}) = \sum_W [I(\mathbf{x}_i + \mathbf{d}, t + \tau) - I(\mathbf{x}_i, t)]^2 \quad [4.42]$$

Where \mathbf{x}_i are points in the Gaussian window W centered on \mathbf{x} .

When the displacement vector is small, the intensity function can be approximated by Taylor series expansion truncated to the linear term, as:

$$I(\mathbf{x}_i + \mathbf{d}, t + \tau) = I(\mathbf{x}_i, t + \tau) + \frac{\partial I}{\partial \mathbf{x}}(\mathbf{x}_i, t + \tau) \mathbf{d} \quad [4.43]$$

By dropping the time term in the intensity for simplicity and denoting

$$h(\mathbf{x}_i) = I(\mathbf{x}_i, t + \tau) - I(\mathbf{x}_i, t), \quad \mathbf{g}(\mathbf{x}_i) = \frac{\partial I}{\partial \mathbf{x}}(\mathbf{x}_i, t + \tau),$$

we can rewrite the residual error

$$\varepsilon(\mathbf{x}) = \sum_W [h(\mathbf{x}_i) + \mathbf{g}(\mathbf{x}_i) \mathbf{d}]^2 \quad [4.44]$$

To find the displacement \mathbf{d} , we let the derivative of the error to zero:

$$\begin{aligned} \frac{\partial \varepsilon(\mathbf{x})}{\partial \mathbf{d}} &= 2 \sum_W [h(\mathbf{x}_i) + \mathbf{g}(\mathbf{x}_i) \mathbf{d}] \mathbf{g}(\mathbf{x}_i) = 0 \\ \Rightarrow \sum_W \mathbf{g}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_i) \mathbf{d} &= \sum_W h(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_i) \end{aligned} \quad [4.45]$$

In other words, we must solve the equation:

$$\mathbf{Gd} = \mathbf{e} \quad [4.46]$$

Where \mathbf{G} is a symmetric 2×2 matrix $\mathbf{G} = \sum_w \mathbf{g}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_i)$ and \mathbf{e} is a 2×1 vector $\mathbf{e} = \sum_w h(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_i)$. However, because of the linearization of Equation 4.43, the solution of \mathbf{d} is only approximately correct. Iteration in Newton style is therefore needed to search for the right displacement by resampling $h(\mathbf{x}_i)$ with new \mathbf{d} in each step. Iteration is stopped when difference in \mathbf{d} is negligible, which usually takes less than 5 iterations.

It can also be seen from Equation 4.46 that to get a reliable displacement, both eigenvalues of \mathbf{G} must be larger than noise level. Explicitly, two small eigenvalues mean a roughly flat intensity within a window; a large and a small eigenvalue correspond to an edge and two large eigenvalues represent corners. So, to select good reliable features, we accept a window if:

$$\min(\lambda_1, \lambda_2) > \lambda \quad [4.47]$$

Where λ_1, λ_2 are the eigenvalues of \mathbf{G} and λ is a predefined threshold.

4.3.2 Stereo Matching

In Section 4.2.3, it was shown that the position of a 3D point can be reconstructed from two perspective images. To achieve a correct 3D reconstruction, point correspondences between two images need to be established. A large amount of work has been carried out for matching different images of a single scene (Chou and Chen, 1990; Weng et al., 1992a), but the results are not satisfactory if the only geometric constraint, i.e., the epipolar constraint, is not used. In our application, after

stereo calibration, the epipolar geometry is known, therefore robust correspondences can be established between the two images (Zhang et al., 1994).

After applying KLT on one of the camera image, we got N feature points suitable for tracking. For each point \mathbf{m}_i ($i=1,2,\dots,N$) in image 1, a corresponding epipolar line can be found as $\mathbf{l}'_m = \mathbf{E}\tilde{\mathbf{m}}_{ni}$, where \mathbf{E} is the essential matrix defined in Equation 4.30, $\tilde{\mathbf{m}}_{ni}$ is the normalized homogeneous coordinate for point \mathbf{m}_i . Figure 4-10 shows the two epipolar lines (red lines in b) calculated for two randomly picked points (red “+” in a). Both lines contain the corresponding points. Therefore the search range for a matching point is narrowed to a band centered on the epipolar line. All corners \mathbf{m}'_{ij} ($j=1,\dots,\text{number of corners}$) are detected in this band as candidates (Harris and Stephens, 1988) and correlations are computed to search for the corresponding point.

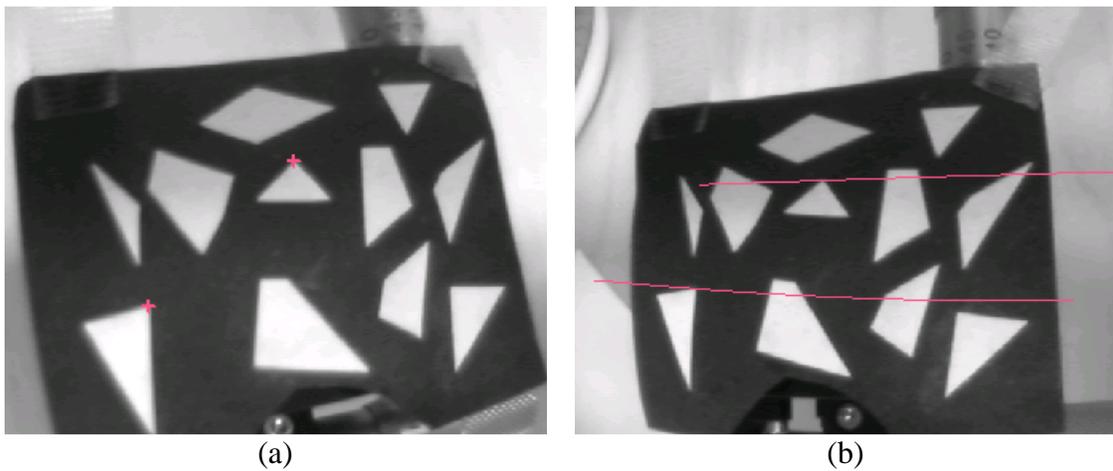


Figure 4-10 Epipolar lines estimated after stereo calibration. (a) shows an image from one of the camera and two randomly picked points. (b) shows the image from the other camera and the estimated epipolar lines corresponding to the two points in (a) respectively.

Given a point \mathbf{m}_i in image one and candidate points \mathbf{m}'_{ij} in image two, a correlation window of size $(2n+1) \times (2m+1)$ centered at each of these points was used. The correlation coefficient is defined as:

$$c(\mathbf{m}_i, \mathbf{m}'_{ij}) = \frac{\sum_{\Delta \in W} [I_1(\mathbf{m}_i + \Delta) - \bar{I}_1(\mathbf{m}_i)] [I_2(\mathbf{m}'_{ij} + \Delta) - \bar{I}_2(\mathbf{m}'_{ij})]}{\sqrt{\left(\sum_{\Delta \in W} [I_1(\mathbf{m}_i + \Delta) - \bar{I}_1(\mathbf{m}_i)] \right)^2} \sqrt{\left(\sum_{\Delta \in W} [I_2(\mathbf{m}'_{ij} + \Delta) - \bar{I}_2(\mathbf{m}'_{ij})] \right)^2}} \quad [4.48]$$

Where Δ are the points in the Gaussian window W centered on zero.

$\bar{I}_k(\mathbf{m}) = \sum_{\Delta \in W} I_k(\mathbf{m} + \Delta) / [(2n+1)(2m+1)]$ is the average at point \mathbf{m} of I_k ($k=1,2$).

Correlation measures the relationship between two windows. Possible correlations range from +1 to -1. A zero correlation indicates that there is no relationship between the windows. A correlation of -1 indicates a perfect negative correlation, meaning that as the intensities in one window go up, the other go down. A correlation of +1 indicates a perfect positive correlation, meaning that intensities in both windows have the same pattern. A best match point for \mathbf{m}_i is therefore the one in \mathbf{m}'_{ij} which has the highest correlation. However, there could be cases where no match point exist in image two for a point in image one (occlusion, outside of field-of-view, etc), so a minimum correlation threshold was set to avoid this problem. Therefore, the best match point is

$$m'_r = \max\{c(\mathbf{m}_i, \mathbf{m}'_{ij})\} > c_threshold \quad [4.49]$$

In our implementation, $n=m=7$ for the correlation window, and a correlation threshold of was set to 0.7.

4.3.3 Rigid Motion Estimation

3D positions from the corresponding points in the two views can be calculated based on stereo triangulation. When these feature points are tracked in real-time, rigid motion parameters (rotation \mathbf{R}_c , translation \mathbf{T}_c) can be calculated.

Given two sets of points $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ in 3D space at successive time, rigid motion parameters are the solution by minimizing the cost function (Arun et al., 1987; Horn, 1987; Horn et al., 1988; Kanatani, 1994; Umeyama, 1991; Haralick et al., 1989):

$$\varepsilon^2 = \sum_{i=1}^N \|\mathbf{y}_i - (\mathbf{R}_c \mathbf{x}_i + \mathbf{T}_c)\|^2 \quad [4.50]$$

The minimum should occur when the partial derivative of ε^2 with respect to \mathbf{T}_c equals zero.

$$\begin{aligned} \frac{\partial \varepsilon^2}{\partial \mathbf{T}_c} &= 2 \sum_{i=1}^N (\mathbf{y}_i - (\mathbf{R}_c \mathbf{x}_i + \mathbf{T}_c))(-1) = 0 \\ \Rightarrow \mathbf{T}_c &= \bar{\mathbf{y}} - \mathbf{R}_c \bar{\mathbf{x}} \end{aligned} \quad [4.51]$$

Where

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad \bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \quad [4.52]$$

Therefore once \mathbf{R}_c is known, \mathbf{T}_c can be calculated easily.

To minimize ε^2 with respect to \mathbf{R}_c , the orthonormal constraint of the rotation matrix needs to be used. Let $\mathbf{R}_c = [\mathbf{r}_1^T, \mathbf{r}_2^T, \mathbf{r}_3^T]^T$, where each \mathbf{r}_k^T is a row of \mathbf{R}_c . The constraints can be written as:

$$\begin{aligned} \mathbf{r}_k^T \mathbf{r}_k &= 1, \quad k = 1, 2, 3 \\ \mathbf{r}_j^T \mathbf{r}_k &= 0, \quad j \neq k \end{aligned} \quad [4.53]$$

Construct the Lagrangian function for ε^2 with the above constraint, we get

$$\begin{aligned} \varepsilon^2 = & \sum_{i=1}^N \sum_{k=1}^3 (y_{ik} - \mathbf{r}_k^T \mathbf{x}_i - T_{ck})^2 \\ & + \sum_{k=1}^3 \lambda_k (\mathbf{r}_k^T \mathbf{r}_k - 1) + 2\lambda_4 \mathbf{r}_1^T \mathbf{r}_2 + 2\lambda_5 \mathbf{r}_1^T \mathbf{r}_3 + 2\lambda_6 \mathbf{r}_2^T \mathbf{r}_3 \end{aligned} \quad [4.54]$$

Where $\mathbf{y}_i = [y_{i1} \quad y_{i2} \quad y_{i3}]^T$, $\mathbf{T}_c = [T_{c1} \quad T_{c2} \quad T_{c3}]^T$.

Substitute the constraint of \mathbf{T}_c in Equation 4.51 to Equation 4.54, we get

$$\begin{aligned} \varepsilon^2 = & \sum_{i=1}^N \sum_{k=1}^3 (y_{ik} - \bar{y}_k - \mathbf{r}_k^T (\mathbf{x}_i - \bar{\mathbf{x}}))^2 \\ & + \sum_{k=1}^3 \lambda_k (\mathbf{r}_k^T \mathbf{r}_k - 1) + 2\lambda_4 \mathbf{r}_1^T \mathbf{r}_2 + 2\lambda_5 \mathbf{r}_1^T \mathbf{r}_3 + 2\lambda_6 \mathbf{r}_2^T \mathbf{r}_3 \end{aligned} \quad [4.55]$$

Then take the partial derivative of ε^2 with respect to \mathbf{r}_k , and set them to zero, we get:

$$\begin{aligned} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{r}_1 + \lambda_1 \mathbf{r}_1 + \lambda_4 \mathbf{r}_2 + \lambda_5 \mathbf{r}_3 &= \sum_{i=1}^N (y_{i1} - \bar{y}_1)(\mathbf{x}_i - \bar{\mathbf{x}}) \\ \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{r}_2 + \lambda_4 \mathbf{r}_1 + \lambda_2 \mathbf{r}_2 + \lambda_6 \mathbf{r}_3 &= \sum_{i=1}^N (y_{i2} - \bar{y}_2)(\mathbf{x}_i - \bar{\mathbf{x}}) \\ \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{r}_3 + \lambda_5 \mathbf{r}_1 + \lambda_6 \mathbf{r}_2 + \lambda_3 \mathbf{r}_3 &= \sum_{i=1}^N (y_{i3} - \bar{y}_3)(\mathbf{x}_i - \bar{\mathbf{x}}) \end{aligned} \quad [4.56]$$

Let

$$\mathbf{A} = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & \lambda_4 & \lambda_5 \\ \lambda_4 & \lambda_2 & \lambda_6 \\ \lambda_5 & \lambda_6 & \lambda_3 \end{bmatrix}$$

And

$$\begin{aligned} \mathbf{B} &= \left[\sum_{i=1}^N (y_{i1} - \bar{y}_1)(\mathbf{x}_i - \bar{\mathbf{x}}) \quad \sum_{i=1}^N (y_{i2} - \bar{y}_2)(\mathbf{x}_i - \bar{\mathbf{x}}) \quad \sum_{i=1}^N (y_{i3} - \bar{y}_3)(\mathbf{x}_i - \bar{\mathbf{x}}) \right] \\ &= \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \end{aligned} \quad [4.57]$$

Equation 4.56 can be rewritten as

$$\mathbf{A}\mathbf{R}_c^T + \mathbf{R}_c^T\boldsymbol{\lambda} = \mathbf{B} \Rightarrow \mathbf{R}_c\mathbf{A}\mathbf{R}_c^T + \boldsymbol{\lambda} = \mathbf{R}_c\mathbf{B} \quad [4.58]$$

Since \mathbf{A} is symmetric, $\mathbf{A}^T = \mathbf{A}$, therefore $(\mathbf{R}_c\mathbf{A}\mathbf{R}_c^T)^T = \mathbf{R}_c\mathbf{A}^T\mathbf{R}_c^T = \mathbf{R}_c\mathbf{A}\mathbf{R}_c^T$ is also symmetric. Both of the two terms on the left-side are symmetric, which means the right-side must also be symmetric

$$\mathbf{R}_c\mathbf{B} = \mathbf{B}^T\mathbf{R}_c^T \quad [4.59]$$

If the singular value decomposition (SVD) of \mathbf{B} is $\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ where \mathbf{U} and \mathbf{V} are orthonormal and \mathbf{D} is diagonal, then

$$\mathbf{R}_c\mathbf{U}\mathbf{D}\mathbf{V}^T = (\mathbf{U}\mathbf{D}\mathbf{V}^T)^T\mathbf{R}_c^T = \mathbf{V}\mathbf{D}\mathbf{U}^T\mathbf{R}_c^T \quad [4.60]$$

It can be observed that a solution for \mathbf{R}_c is simply

$$\mathbf{R}_c = \mathbf{V}\mathbf{U}^T \quad [4.61]$$

But Kanatani (Kanatani, 1994) proves that this solution gives a reflection instead of rotation if $\det(\mathbf{V}\mathbf{U}^T) = -1$. So the correct solution should be adjusted to

$$\mathbf{R}_c = \mathbf{V} \begin{pmatrix} 1 & & \\ & 1 & \\ & & \det(\mathbf{V}\mathbf{U}^T) \end{pmatrix} \mathbf{U}^T \quad [4.62]$$

Which obviously still satisfies Equation 4.60.

Given 3D feature point positions at any two successive times, real time motion parameters can be calculated by first constructing the correlation matrix \mathbf{B}

using Equation 4.57, then calculating the rotation \mathbf{R}_c from the SVD of \mathbf{B} using Equation 4.62 and finally determining translation \mathbf{T}_c from Equation 4.51.

4.4 Motion Calculation in MRI coordinates

4.4.1 Motion in MRI physical coordinates

During the actual experiment, a number of feature points on the object (or face) were automatically selected in the left camera image after which the corresponding points were identified in the right camera image (Section 4.3.2). After correction for the image distortion of the camera lenses, the position of these feature points could be calculated in 3D using triangulation (Section 4.2.3). The changes in position of these feature points in the camera images were tracked over time (Section 4.3.1), allowing calculation of changes in their 3D position. The translation and rotation parameters were then estimated from this collection of 3D points (Section 4.3.3). However, these parameters were in the camera coordinate system and needed to be transformed into MRI coordinate system based on the prior calibration data.

At time t_0 , the relationship between the points MRI coordinates (\mathbf{X}_m) and camera coordinates (\mathbf{X}_c) can be described as follows, based on calibration data:

$$\mathbf{X}_m(t_0) = \mathbf{R}_{mc} \mathbf{X}_c(t_0) + \mathbf{T}_{mc} \quad [4.63]$$

where \mathbf{R}_{mc} and \mathbf{T}_{mc} are the rotation and translation parameters for the conversion between MRI and camera system (Section 4.2.4).

At time t_i , the tracker measures the movement parameters in camera coordinate system,

$$\mathbf{X}_c(t_i) = \mathbf{R}_c \mathbf{X}_c(t_0) + \mathbf{T}_c \quad [4.64]$$

Where \mathbf{R}_c and \mathbf{T}_c are the rotational and translational motion parameters. Using the above two equations, we can calculate the movement (rotation \mathbf{R}_m and translation \mathbf{T}_m) in MRI coordinates,

$$\begin{aligned} \mathbf{X}_m(t_i) &= \mathbf{R}_{mc} \mathbf{R}_c \mathbf{R}_{mc}^{-1} \mathbf{X}_m(t_0) - \mathbf{R}_{mc} \mathbf{R}_c \mathbf{R}_{mc}^{-1} \mathbf{T}_{mc} + \mathbf{R}_{mc} \mathbf{T}_c + \mathbf{T}_{mc} \\ \Rightarrow \mathbf{R}_m &= \mathbf{R}_{mc} \mathbf{R}_c \mathbf{R}_{mc}^{-1} \\ \mathbf{T}_m &= -\mathbf{R}_{mc} \mathbf{R}_c \mathbf{R}_{mc}^{-1} \mathbf{T}_{mc} + \mathbf{R}_{mc} \mathbf{T}_c + \mathbf{T}_{mc} \end{aligned} \quad [4.65]$$

4.4.2 Motion in MRI logical coordinates

Since the pulse sequence used MRI logical coordinate system instead of physical coordinate system, the motion parameters calculated from Equation 4.65 need to be transformed into the logical coordinate system.

MRI physical coordinate system refers to the standard axes of an MRI magnet system, shown in Figure 4-11. The subject is usually placed head first into the magnet and carried in by a sliding gantry on the patient table. If the person is supine (lying on his/her back), the head point along $O_M Z_M$ (direction of the main magnetic field), right shoulder along $O_M X_M$ (left-right direction), and nose along $O_M Y_M$ (up-down direction). O_M is the gradient isocenter, where the gradients produce zero magnetic field and which typically corresponds to the center of the magnet. This coordinate system does not change and is used in camera and MRI system calibration in Section 4.2.4.

The MRI logical coordinate system, however, depends on the prescribed scan plane. As in Figure 4-11, $O_L X_L$, $O_L Y_L$ and $O_L Z_L$ correspond to frequency encoding,

phase encoding and slice selection directions respectively. These change with different scan plane prescription and are used in MRI pulse sequence.

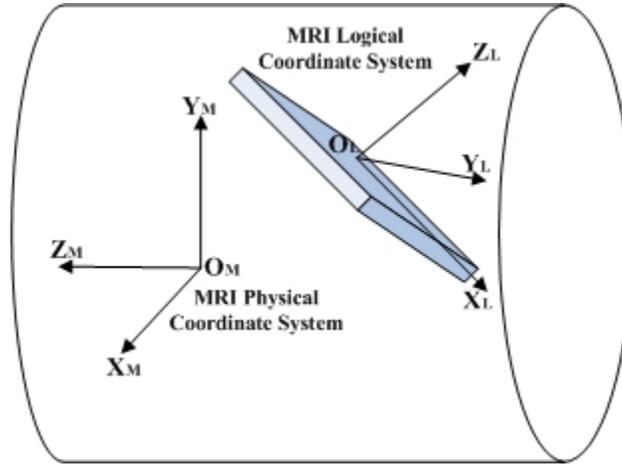


Figure 4-11 MRI physical and logical coordinate system. $\{O_M, X_M, Y_M, Z_M\}$ is MRI physical coordinate system, which doesn't change after the MRI system is set up. $\{O_L, X_L, Y_L, Z_L\}$ is MRI logical coordinate system, which changes with different prescribed scan plane. X_L is the frequency encoding direction, Y_L is phase encoding direction and Z_L is slice selection direction.

The rotation and translation (\mathbf{R}_{ml} and \mathbf{T}_{ml}) between these two coordinate systems can be read out from the pulse sequence after the scan-plane is prescribed:

$$\mathbf{X}_m(t_0) = \mathbf{R}_{ml}(\mathbf{X}_l(t_0) + \mathbf{T}_{ml}) \quad [4.66]$$

Therefore, the final motion parameters as will be used by the pulse sequence should be (see derivation in Section 9.1):

$$\mathbf{R}_{corr} = \mathbf{R}_m \mathbf{R}_{ml} \quad [4.67]$$

$$\mathbf{T}_{corr} = \mathbf{R}_{corr}^{-1} \mathbf{T}_m \quad [4.68]$$

The motion parameters (\mathbf{R}_{corr} and \mathbf{T}_{corr}) were sent from the tracking system to the MRI scanner over a TCP/IP connection (Santos et al., 2004) for real time compensation of motion.

4.5 Motion Correction for MRI

Motion correction was evaluated using a GRE pulse sequence. The sequence updated the scan parameters before each RF excitation. Correction for rotation was performed by adjusting the gradient rotation matrix; correction for translation in the slice selection and read out direction was performed by adjusting the RF transmit and receive frequencies and phases respectively; and correction for translation in the phase encoding direction was performed by adjusting the receive phase offset. Since the correction for translation $\mathbf{T}_{\text{corr}} = [d_x, d_y, d_z]^T$ is different in all three directions, we will discuss each of them separately.

4.5.1 Rotation Correction

MRI can acquire images in any plane, which is realized by varying the frequency-encoding, phase-encoding, and slice-selection gradient waveforms. The directions of these three gradients define the X_L , Y_L , and Z_L for MRI logical coordinate system (Section 4.4.2). Therefore, after the operator's prescription of a scan plan, the logical gradient waveforms that are played out should be in this coordinate system. However, a magnetic field gradient system normally consists of only three orthogonal gradient coils that generate gradient field in the physical X_M , Y_M and Z_M axis. In order to create the three logical gradient waveforms, three physical waveforms should be digitally mixed. The mixing ratios are determined by angles of the prescribed scan plane and can be described with an orthonormal rotation matrix.

$$\begin{bmatrix} G_{mX}(t) \\ G_{mY}(t) \\ G_{mZ}(t) \end{bmatrix} = \mathbf{R}_{ml}(t) \begin{bmatrix} G_{lX}(t) \\ G_{lY}(t) \\ G_{lZ}(t) \end{bmatrix} \quad [4.69]$$

Where G_{mk} and G_{lk} ($k=X, Y, Z$) are the physical and logical gradient waveforms respectively, $\mathbf{R}_{ml}(0)$ is the same matrix in Equation 4.66. When motion occurs during the acquisition, the orientation of the gradient fields needs to be adjusted in real time to follow the motion. The rotation matrix \mathbf{R}_{corr} from Equation 4.67 is thus used in the pulse sequence to correct the combination of physical gradient waveforms to generate the desired logical gradient waveforms.

4.5.2 Slice Selection Displacement Correction

It has been shown in Figure 2-1 that a slice can be excited by applying a slice-selective RF pulse. Since RF pulse is frequency-selective, only spins resonating at the same frequency can be excited. To make an RF pulse spatially selective, it is necessary to make the spin resonance frequency position-dependent, which is accomplished by applying a linear z-gradient field during the excitation. In the presence of this gradient, the Larmor frequency at position z is given by

$$\omega(z) = \gamma(B_0 + G_z z) = \omega_0 + \gamma G_z z \quad [4.70]$$

Therefore a slice at position z can be excited by an RF pulse with frequency $\omega(z)$. If, however, the object moves d_z in this direction, frequency of the RF pulse needs to be adjusted to select the same slice for each excitation (Figure 4-12),

$$\omega'(z, d_z) = \gamma(B_0 + G_z(z + d_z)) = \omega_0 + \gamma G_z(z + d_z) \quad [4.71]$$

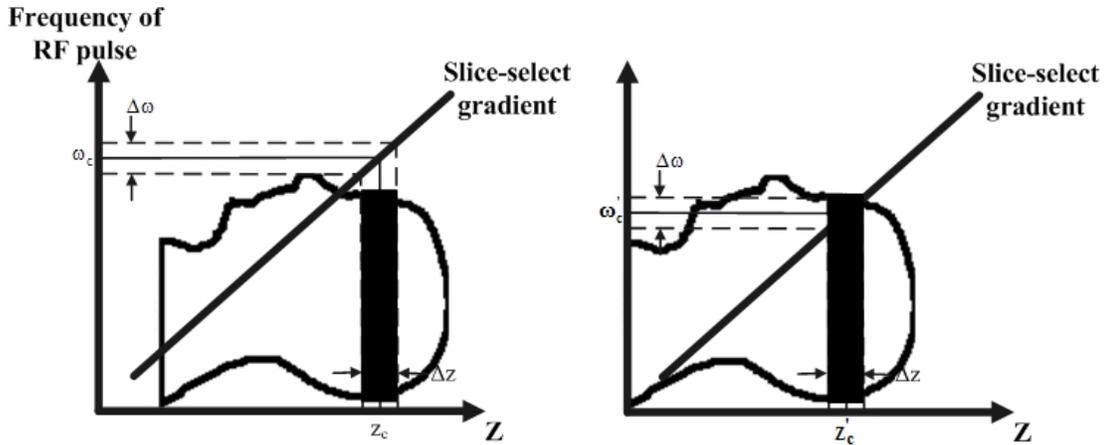


Figure 4-12 The change of RF frequency to select the same slice when motion occurs. Under the same slice-selective gradient, to follow the motion, the only way is to change the RF frequency.

For a GE scanner, the RF frequency generation path can be broken up into 3 components: center frequency, slice selection delta frequency, and phase. The center frequency is ω_0 , which is only B_0 dependent, e.g. 298MHz for a 7T scanner. The delta-frequency component creates an output which is added to the output of the center frequency path for slice selection ($\gamma G_z z$ in Equation 4.70). The phase component determines the phase of the RF transmit pulse.

Without loss of generality, we assume $z=0$ from now on, because it is a constant after scan plane prescription and only d_z is the variable changing with motion. Figure 4-13 shows the sequences for slice selection and the three components that determine RF frequency and phase. For the GRE sequence used in our application, a SINC RF excitation pulse in conjunction with a slice-selection gradient is used. The center frequency is a constant, which is 298MHz as mentioned above. After prescription by an operator, delta frequency d_ω is set, which is zero here because $z=0$ (the black straight line of d_ω). Also, the phase d_ϕ is zero.

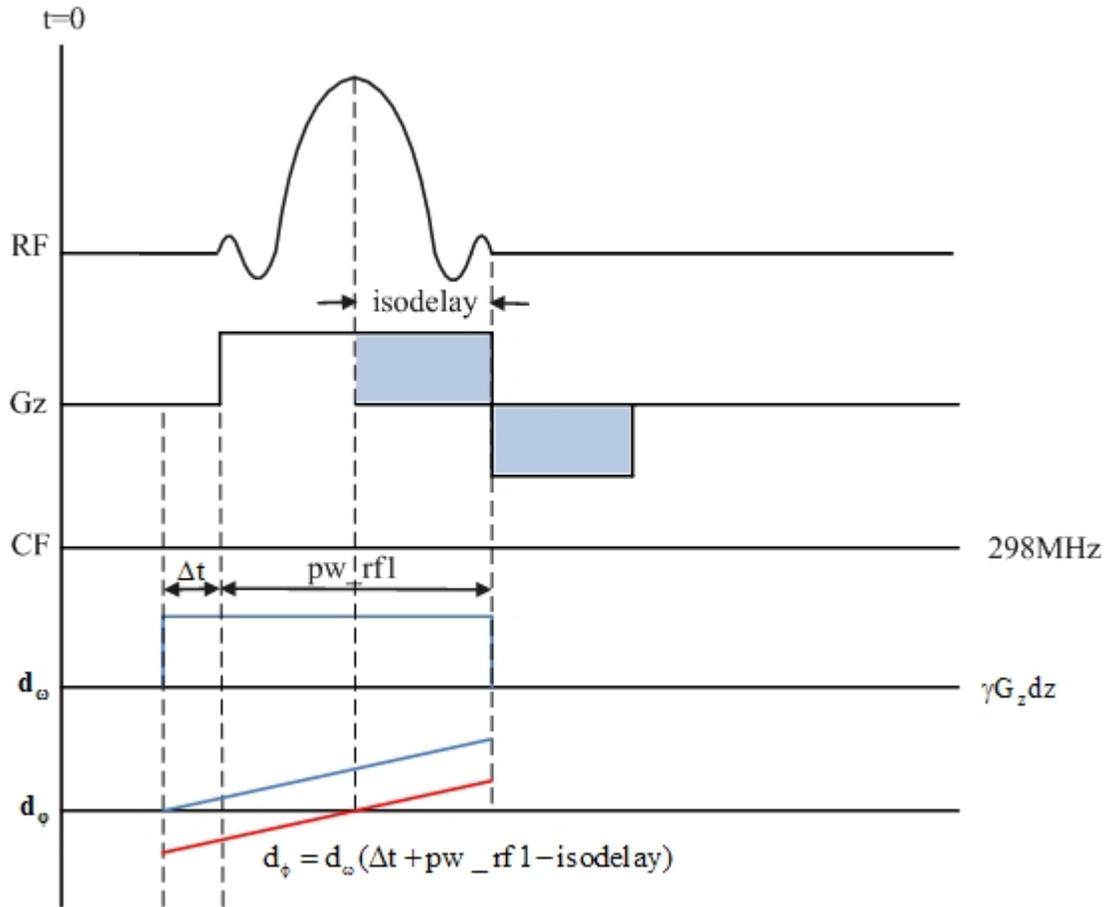


Figure 4-13 Slice selection sequences (RF pulse and z-gradient) and the three components that generate the RF pulse, i.e. center frequency, delta frequency and phase. $t=0$ is the start of TR and only one TR is shown here.

Figure 4-13 also shows a parameter isodelay, which is a very important parameter for an RF pulse. After the RF pulse is played out, because of precession caused by the slice selection gradient, the transverse magnetization will generally not be coherent across the slice. Instead there will be phase dispersion ($\phi = \gamma G_z z t$). A rephasing lobe with an opposite polarity to the slice selection gradient is therefore needed to compensate for the phase dispersion. The isodelay is the parameter used to calculate the optimal area of rephasing gradient. It is defined as the effective precession or dephasing time that results in the phase dispersion across the slice. As

shown in the figure, the rephasing gradient lobe (negative blue lobe) should have the same area as the slice selection gradient under isodelay:

$$G_z \cdot isodelay = A_R \quad [4.72]$$

Where A_R is the area of the slice rephasing lobe. The shape of the rephasing lobe does not affect the phase, only the area matters. For the SINC RF pulse used in GRE, the isodelay is nearly equal to one-half the RF pulse width.

After the rephasing gradient lobe, phase dispersion of the transverse magnetization is eliminated for each RF excitation pulse. Since multiple RF excitations have to be applied for imaging (Section 2.1.2), the phase at the start point of isodelay should therefore also be consistent for each excitation.

If no motion occurred, $d_\omega = d_\phi = 0$, meaning no phase difference adjustment is made between RF excitations. However, if there is displacement in the z direction, the delta frequency component needs to be changed with an offset $d_\omega = \gamma G_z d_z$, a phase therefore is accumulated. At the start of isodelay, the phase is:

$$d_\phi = \gamma G_z d_z (\Delta t + pw_rf1 - isodelay) \quad [4.73]$$

This phase obviously is not zero anymore (the blue line of d_ϕ in Figure 4-13), which will not cause phase dispersion across slice, but rather phase difference between every RF excitations. This phase needs to be subtracted in the RF phase generation path to recover net zero phase at isodelay (the red line of d_ϕ in Figure 4-13). All parameters in Equation 4.73 are known, except Δt , which is the period of turning on the delta frequency before the RF pulse plays out. Therefore, to calculate the correct phase, the constant Δt has to be measured.

To measure Δt , a cylindrical shaped phantom was used. The advantage of this phantom is that when scanned axially (and the phantom axis is aligned with B_0), every slice is exactly the same. We set phase encoding gradient to zero, and reconstructed the data with FFT transformation only in the frequency encoding direction. The resultant is therefore a hybrid space image, with every line in the frequency encoding direction corresponding to the profile of the phantom. Figure 4-14 shows the hybrid magnitude and phase image of the phantom (256×128 matrix, $TR=100\text{ms}$, $TE=6\text{ms}$, $FA=10^\circ$, Slice Thickness=1mm). Every line is exactly the same due to identical phase encoding.

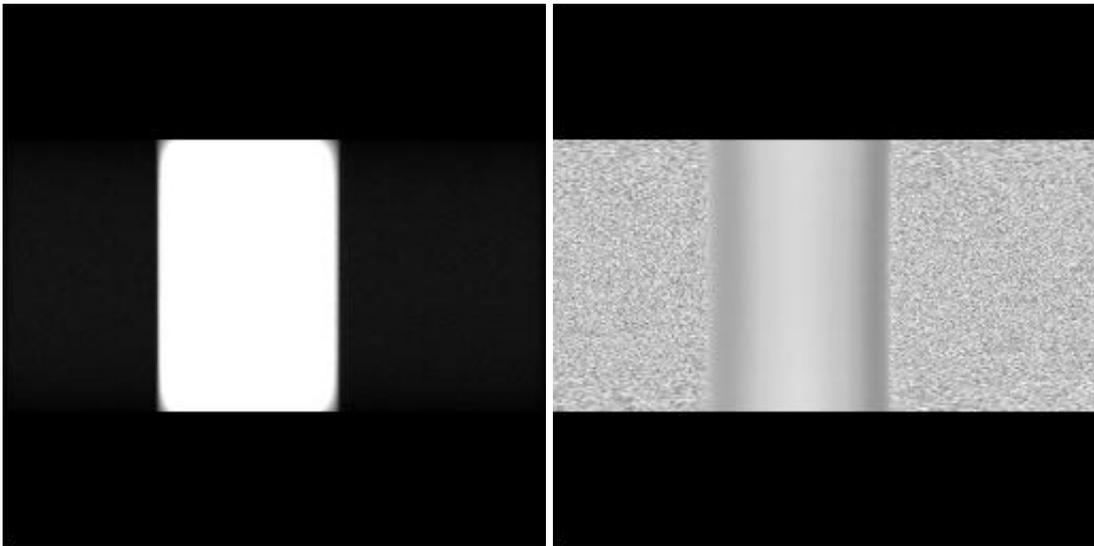


Figure 4-14 Hybrid magnitude and phase image of the cylindrical phantom. Phase encoding gradient is zero. Frequency encoding is left-right direction.

If a linear frequency is added to each RF excitation frequency, the selected slices will vary correspondingly. However, because the phantom is cylindrical, every slice along the slice-selection direction is the same. Therefore the measured signal could be regarded as from the same slice. When the phase encoding gradient is set to

zero, the hybrid magnitude image will not change, but the phase along the phase encoding direction will change linearly, as shown in Figure 4-15.

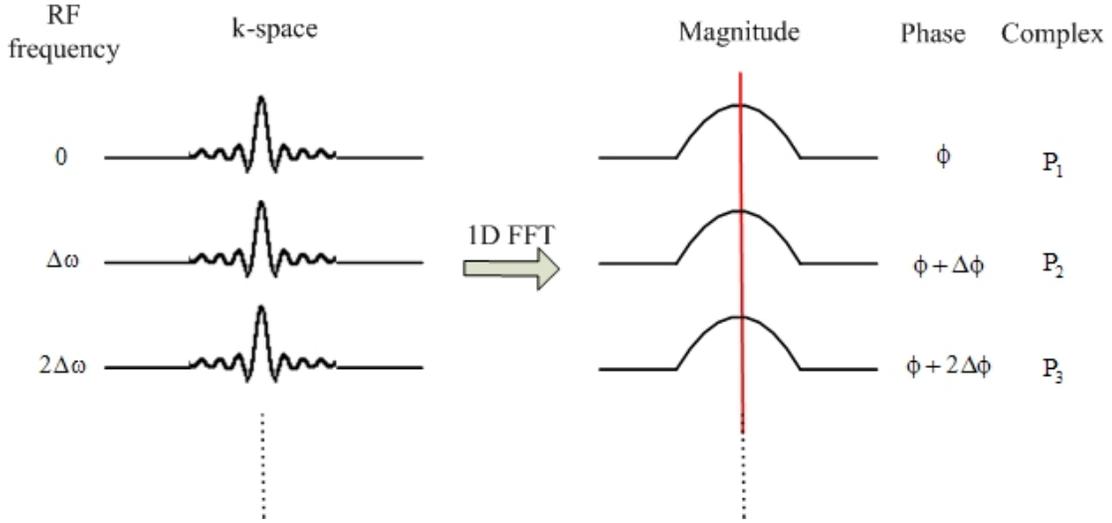


Figure 4-15 The hybrid space image after applying a linear RF frequency. Phase encoding gradient is set to zero. For each RF excitation, its frequency changes linearly with a step $\Delta\omega$. Perform FFT on each measured echo in k-space we can get a hybrid image. The magnitude of each line in the frequency encoding direction (shown with a curvature shape) represents the object profile. The red line is to select an arbitrary point in the hybrid image, and the phases of those points also change linearly with a step $\Delta\phi$. Complex means to combine the magnitude and phase of the point in that line.

Only three lines are shown for illustration. The RF frequency changed with a step $\Delta\omega$. The acquired k-space signals were reconstructed by FFT along only the frequency encoding direction. The magnitude profiles all have the same shape because the profile of the phantom is independent of slice position. However, the phase of any arbitrary point along the phase encoding direction (red line) should vary linearly corresponding to $\Delta\omega$, shown as ϕ , $\phi + \Delta\phi$, $\phi + 2\Delta\phi$. Even though $\Delta\phi$ can be calculated by using only two lines, improved precision is achieved when using all the information from an acquisition

$$\Delta\phi = \arg\left(\sum_{i=1}^{view-1} P_i^* P_{i+1}\right) \quad [4.74]$$

Where $view$ is the total number of phase encoding, P_i is the reconstructed complex value of an arbitrary point (must be on the scanned object) on the i^{th} phase encoding line.

Then because

$$\Delta\phi = \Delta\omega(\Delta t + pw_rf1/2 - isodelay) \quad [4.75]$$

We can find the only unknown parameter Δt , which equals to $264 \mu s$ for the scanner used.

If we calculate $\Delta\phi$ as the negative value of Equation 4.75 and change the phase of the RF pulse with this $\Delta\phi$, the linear phase change along the phase encoding direction caused by $\Delta\omega$ should disappear (keep the phase encoding gradient still at zero), meaning the phase accumulation at the start of isodelay is compensation to zero. Since this compensation depends on a correct measurement of Δt , the effect of using various Δt for RF phase compensation is illustrated in Figure 4-16, in which Δf ($\Delta f = 2\pi\Delta\omega$) was set to 100Hz; (a) uses the correct $\Delta t=264 \mu s$ and the resultant phase image is very homogeneous showing the fully compensated phase, while (b-d) uses $\Delta t=180 \mu s$, $230 \mu s$ and $480 \mu s$ and the resultant images show obvious uncorrected residual phases.

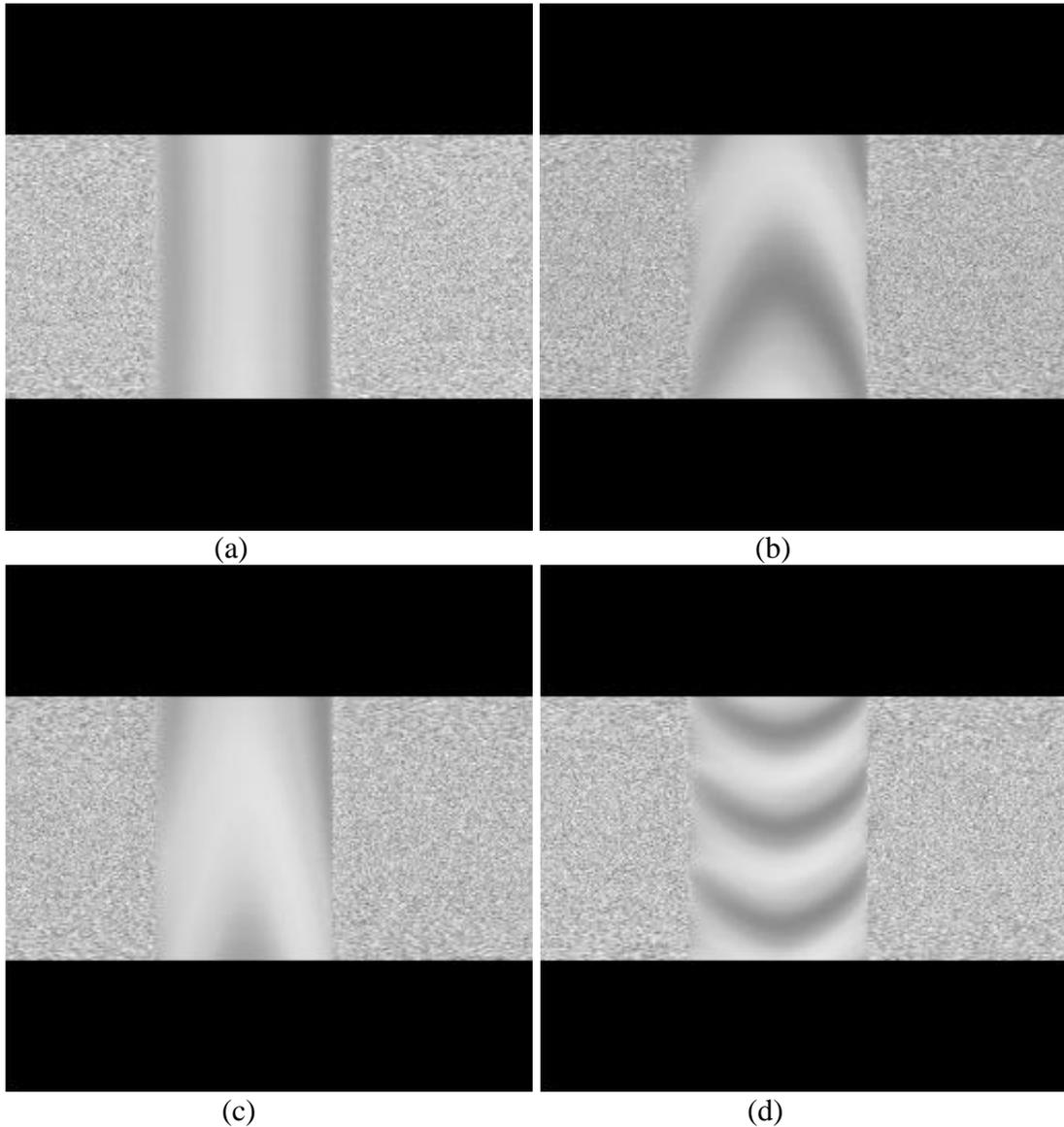


Figure 4-16 The hybrid phase image of the phantom when phase encoding gradient is set to zero. RF transmit frequency varies linearly along the phase encoding direction. (a) Phase correction is calculated with $\Delta t = 264 \mu s$; (b) Phase correction is calculated with $\Delta t = 180 \mu s$; (c) Phase correction is calculated with $\Delta t = 230 \mu s$; (d) Phase correction is calculated with $\Delta t = 480 \mu s$.

It can also be seen from Figure 4-16 that the phase image is very sensitive to the change of Δt (if Δf is set big, e.g. 100Hz), therefore the image itself can be used for the measurement of Δt . After setting a Δf for RF pulse, start changing its phase

using various Δt . The Δt that results in the most homogeneous hybrid phase image is the correct Δt for the scanner.

4.5.3 Frequency Encoding Displacement Correction

A frequency encoding gradient waveform consists of two portions, a dephasing gradient lobe and a readout gradient lobe. The polarity of the dephasing gradient lobe is opposite to the readout gradient lobe, which is used to prepare the transverse magnetization so that an echo signal can be created at a later time. The time of the echo peak is determined by the area under the dephasing gradient lobe. Because of the similar reason as in the previous section, the echo reaches its maximum (all spins are in phase, without considering B_0 inhomogeneity, susceptibility variations, etc) when the area under the readout lobe equals to the area of the dephasing lobe. To describe this quantitatively, denote the dephasing gradient as $-G_x$, the spins along the frequency encoding direction precess at different frequencies:

$$\omega(x) = \gamma(B_0 - G_x x) \quad [4.76]$$

The phase accumulation due to the dephasing gradient (in a rotating frame with frequency of $\omega_0 = \gamma B_0$) is

$$\phi_d(x) = -\gamma G_x x T \quad [4.77]$$

Where T is the duration of dephasing gradient.

When the readout gradient G_x is applied, an additional phase dispersion is introduced. The accumulated phase will be:

$$\phi(x) = \gamma G_x x (t - T) \quad [4.78]$$

At a specific time $t_{echo}=T$, the phase in the above equation equals zero, meaning all the spins are in phase and form an echo. If we want the echo to be at the center of the k-space, the readout gradient must be on for a time $T_{acq}=2T$ (Figure 4-17). So during the acquisition period from 0 to T_{acq} , the sampled data produces a symmetric k-space line spanning $k_{x,min}$ to $k_{x,max}$ (Figure 2-3).

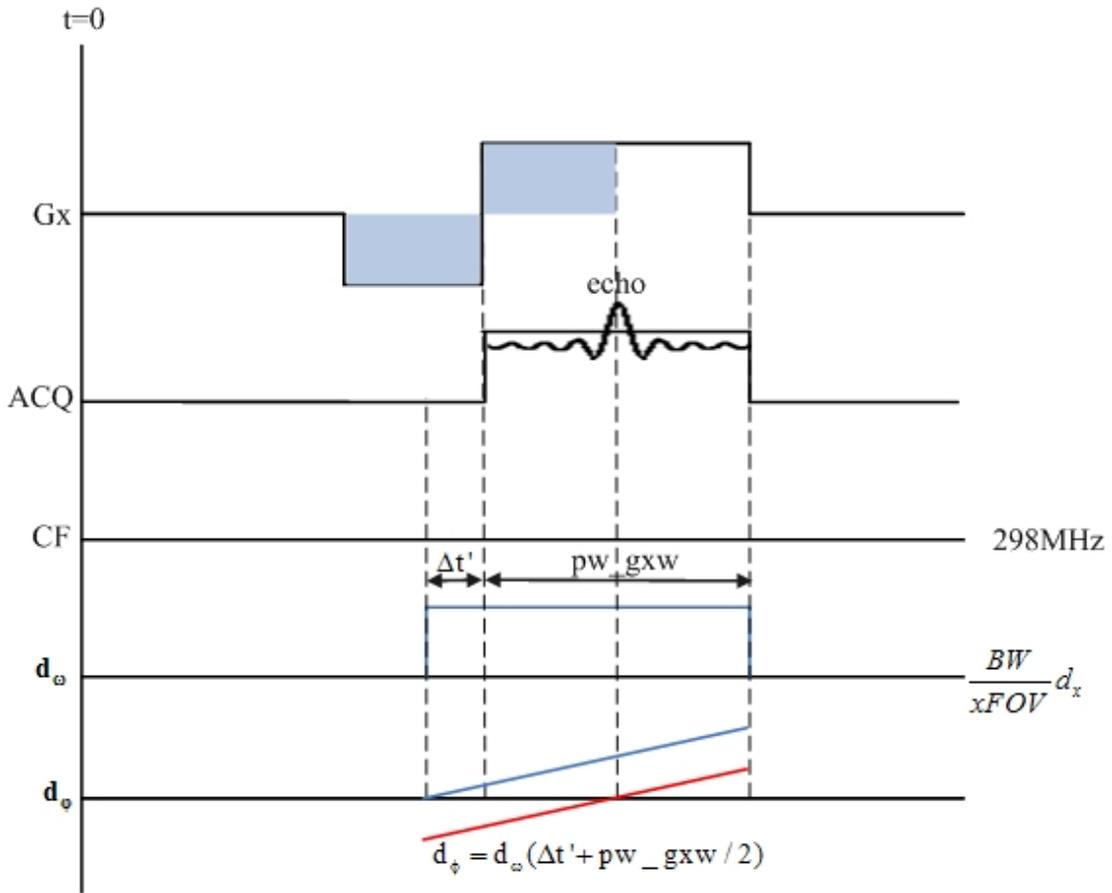


Figure 4-17 Frequency encoding gradient, the acquisition window associated with it, and the three components that generate the receive frequency and phase.

When motion d_x occurs at this direction, the RF receive frequency needs to be offset by

$$d_{\omega} = \gamma G_x d_x \quad [4.79]$$

Given a receive bandwidth $\pm BW$ and the target $xFOV$ (x direction FOV), the readout gradient should be

$$G_x = \frac{BW}{\gamma \cdot xFOV} \quad [4.80]$$

By substituting Equation 4.80 to 4.79, we can get the offset receive frequency

$$d_\omega = \frac{BW}{xFOV} d_x \quad [4.81]$$

Similarly as in Figure 4-13, the offset frequency change results in an extra phase on the echo signal. It has to be subtracted from the receive phase for each acquisition to get rid of phase difference caused by different x displacement (Figure 4-17).

$$d_\phi = d_\omega (\Delta t' + pw_gxw/2) \quad [4.82]$$

Where $pw_gxw/2$ is the pulse width (duration) of the readout gradient, $\Delta t'$ is the time interval between start of the delta frequency and data acquisition.

To measure $\Delta t'$, we again set phase encoding gradient to zero and reconstructed the hybrid phase image. Change the receive frequency linearly with a step Δf (e.g. 100Hz) for each k-space line acquisition, and compensate receive phase using Equation 4.82 with various $\Delta t'$. The $\Delta t'$ corresponding to the most homogeneous phase image was therefore the correct $\Delta t'$ of the scanner. Figure 4-18 shows four hybrid phase images corresponding to $\Delta t' = 144 \mu s$, $100 \mu s$, $200 \mu s$ and $300 \mu s$. The homogeneous image (a) means the receive phase caused by frequency offset was all compensated and therefore $\Delta t' = 144 \mu s$. The diagonal shape was due to the applied linear change of the receive frequency.

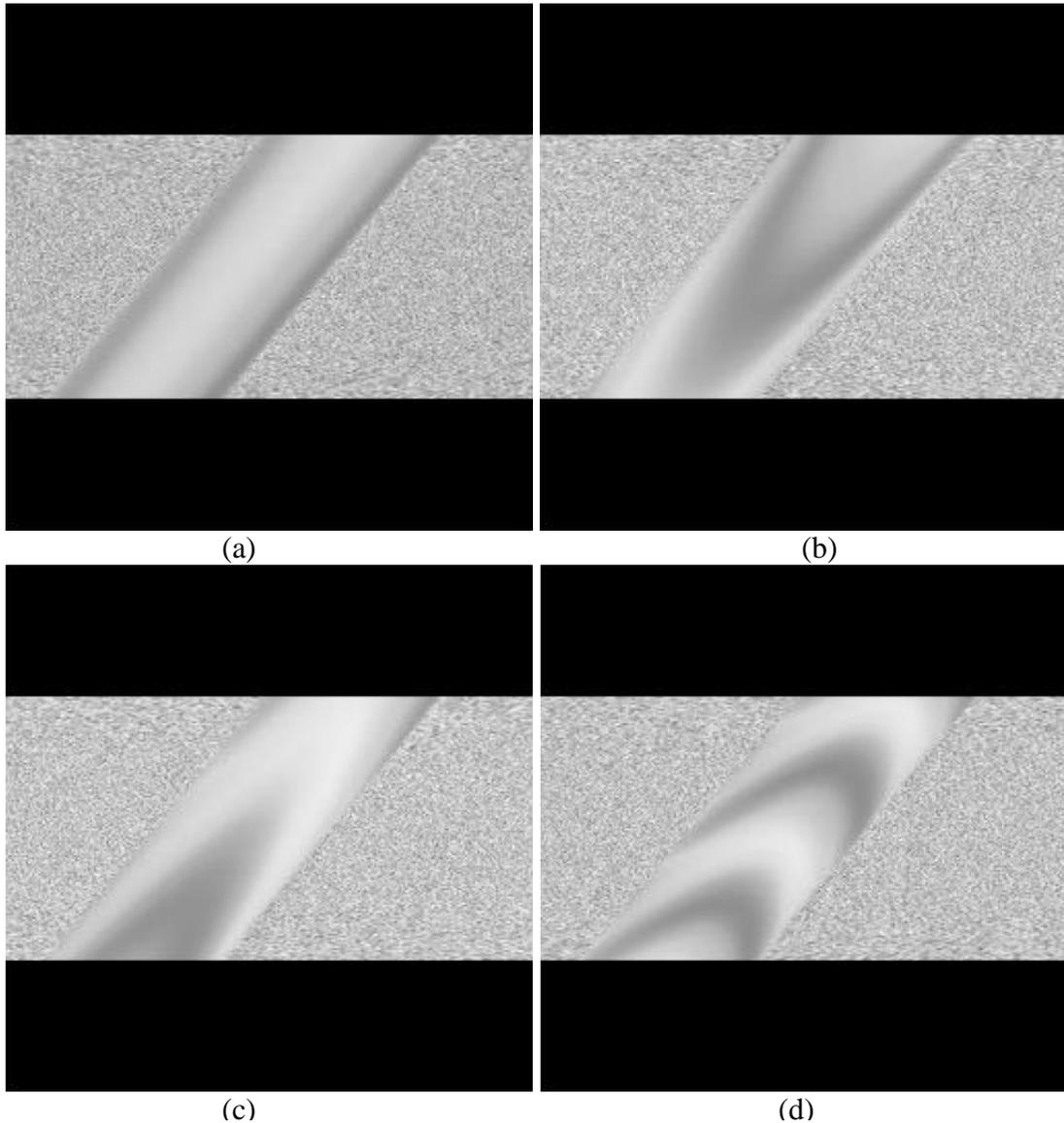


Figure 4-18 Hybrid phase image used for measurement of $\Delta t'$. Phase encoding gradient is set to zero. RF receive frequency varies linearly along the phase encoding direction with a step of 100Hz. (a) Phase correction is calculated with $\Delta t'=144 \mu s$; (b) Phase correction is calculated with $\Delta t'=100 \mu s$; (c) Phase correction is calculated with $\Delta t'=200 \mu s$. (d) Phase correction is calculated with $\Delta t'=300 \mu s$.

4.5.4 Phase Encoding Displacement Correction

Phase encoding gradient is applied while the magnetization is in the transverse plane but before the readout (Figure 2-2). By varying the phase encoding linearly, a linear

phase is introduced to the magnetization in the phase encoding direction (Figure 2-2). FFT is used to reconstruct the spatial information of the object.

Since the frequency and phase encoding can be analyzed independently, Equation 2.11 can be simplified to one dimension:

$$s(k_y) = \sum_{n=0}^{N-1} \rho(n\Delta y) \exp(-i2\pi k_y n\Delta y) \quad [4.83]$$

Where $y = n\Delta y$, Δy and N are the pixel size and number of pixels in phase encoding direction respectively. For each phase encoding step, the gradient G_y changes with a constant step creating a constant Δk_y ($\Delta k_y = \gamma \Delta G_y \tau_y$). The N steps phase-encoding starting at the bottom edge of k-space (Figure 2-3) can therefore be written as:

$$k_y(m) = (m - \frac{N}{2})\Delta k_y \quad m = 0, 1, \dots, N-1 \quad [4.84]$$

The signal is:

$$s(m) = \sum_{n=0}^{N-1} \rho(n\Delta y) \exp(-i2\pi n\Delta y(m - N/2)\Delta k_y) \quad [4.85]$$

To satisfy the Nyquist criterion, the phase encoding step size Δk_y must be chosen so that:

$$\Delta k_y = \frac{1}{yFOV} = \frac{1}{N\Delta y} \quad [4.86]$$

Substitute Equation 4.86 to 4.85, we have

$$\begin{aligned} s(m) &= \sum_{n=0}^{N-1} \rho(n\Delta y) \exp(-i2\pi n\Delta y(m - N/2)\frac{1}{N\Delta y}) \\ &= \sum_{n=0}^{N-1} \rho(n\Delta y) \exp(i\pi n) \exp(-i2\pi nm / N) \end{aligned} \quad [4.87]$$

The image can be reconstructed from the discrete inverse Fourier transform:

$$\rho(n\Delta y)(-1)^n = \sum_{m=0}^{N-1} s(m) \exp(i2\pi m / N) \quad [4.88]$$

The factor $(-1)^n$ means to change the sign of the alternate elements of the output image.

If there is d_y displacement occurred in this direction, the phase of the receiver needs to be changed:

$$\begin{aligned} \rho(n\Delta y + d_y)(-1)^{n+d_y/\Delta y} &= \sum_{m=0}^{N-1} s(m) \exp\left(\frac{i2\pi m}{N}(n + d_y / \Delta y)\right) \\ \Rightarrow \rho(n\Delta y + d_y) &= \sum_{m=0}^{N-1} \left[s(m)(-1)^{d_y/\Delta y} \exp\left(\frac{i2\pi m}{N} \frac{d_y}{\Delta y}\right) \right] (-1)^n \exp(i2\pi mn / N) \end{aligned} \quad [4.89]$$

Note also that

$$k_y(0) = -\frac{N}{2} \Delta k_y = \cancel{\pi} G_y(0) \tau_y \quad [4.90]$$

Where $G_y(0)$ is the gradient for the first step phase encoding.

$$(-1)^{d_y/\Delta y} = \exp\left(\frac{-i\pi}{\Delta y} d_y\right) = \exp\left(\frac{-i\pi N}{N\Delta y} d_y\right) = \exp(-i\pi N \Delta k_y d_y) \quad [4.91]$$

Substitute Equation 4.90 to 4.91, we get

$$(-1)^{d_y/\Delta y} = \exp(i2\pi \cancel{\pi} G_y(0) \tau_y d_y) \quad [4.92]$$

Therefore, the phase offset needed to follow d_y displacement is:

$$d\phi = \left(2\pi \cancel{\pi} G_y(0) \tau_y + \frac{2\pi m}{yFOV} \right) d_y \quad [4.93]$$

The first phase offset term in Equation 4.93 is the phase accumulated in the first encoding step; the second term is a linear phase ramp, meaning the phase increases $2\pi d_y / yFOV$ per phase encoding step.

Note that most of the time, when off-center FOV is required, the shift theorem for Fourier transforms was simply used, which says to shift the image along a direction, a linear phase ramp is needed in the frequency space (the second term in Equation.4.93). But in our case, this is not exactly true because phase encoding started from 0 instead of $-N/2$. When d_y is a constant, the linear phase ramp does produce an off-center FOV. The constant phase (first term in Equation 4.93) does not affect the image contrast because magnitude image does not change at all and phase image has only a global shift. However, when d_y varies with motion, the first term multiplied with d_y is not constant anymore. Thus if not compensated, artifacts will appear.

4.5.5 Summary of Position Compensation

The pulse sequence updated the scan parameters before each RF excitation. The correction of motion for the GRE sequence can be summarized as follows.

- a. Correction for rotation was performed by adjusting the gradient rotation matrix with \mathbf{R}_{corr} (Equation 4.67).
- b. Correction for slice selection translation was performed by adjusting the RF transmit frequency offset with $\gamma G_z d_z$, and adjusting the RF phase with $d_{\phi_x} = -\gamma G_z d_z (\Delta t + pw_rf1 - isodelay)$ (Equation 4.73).
- c. Correction for read out direction translation was performed by adjusting the RF receive frequency with $d_{\omega} = \frac{BW}{xFOV} d_x$ (Equation 4.81), and adjusting the receiver phase with $d_{\phi_x} = -d_{\omega} (\Delta t' + pw_gxw/2)$ (Equation 4.82).

d. Correction for translation in the phase encoding direction was performed by

adjusting the receive phase offset with $d_{\phi_y} = \left(2\pi\gamma G_y(0)\tau_y + \frac{2\pi m}{yFOV} \right) d_y$

(Equation 4.93). (The total receiver phase offset is therefore the summation of phase shift in c and d, $d_{\phi} = d_{\phi_x} + d_{\phi_y}$.)

4.5.6 K-space Line Re-acquisition

An additional feature of the pulse sequence was that it allowed for a re-acquisition of motion corrupted k-space lines. This feature was added to reduce the effect of motion that was too fast to be compensated by gradient and RF frequency and phase adjustment. Re-acquisition was performed when the motion that occurred during the acquisition of a given k-space line exceeded a preset threshold. The displacement threshold for the reacquisition of a k-space line was set to 0.2mm (because image resolution was chosen to be about 0.2mm). A rotation threshold was not used because head rotation was always perceived as a displacement as well. This is because rotation is defined relative to the center of the image, whereas the head generally pivots around the back of the head, where it is supported by the receive coil array, opposite of the part of the head viewed by the cameras.

4.6 Software Flowchart for the Motion Correction System

Figure 4-19 shows the software flowchart of the proposed motion correction system. Calibration was done offline and was programmed with MATLAB. Real time

tracking software was running on Linux with C++. The tracking computer communicated with the MRI scanner through a TCP/IP connection to provide the console (scan computer) with the real-time motion parameters.

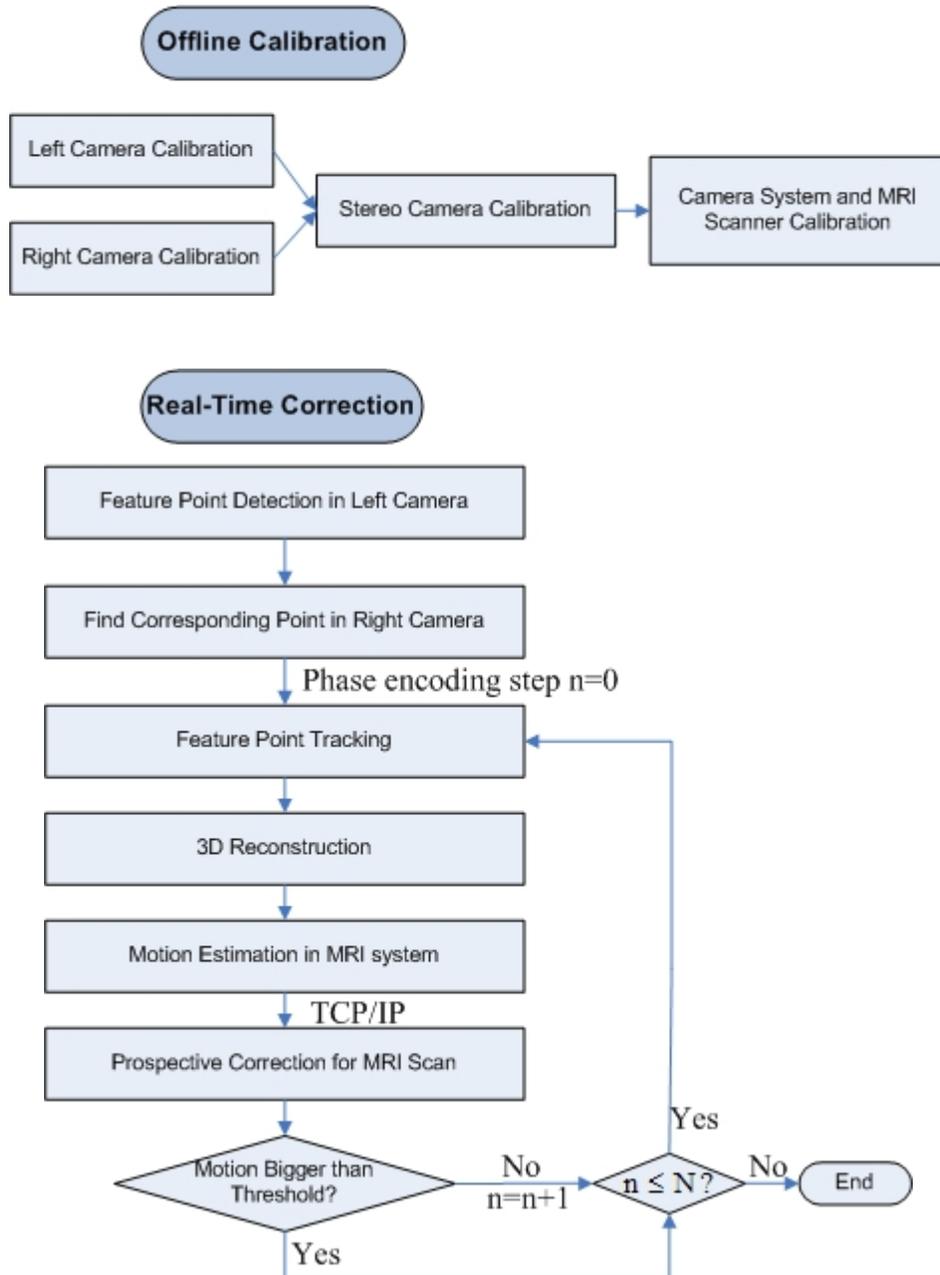
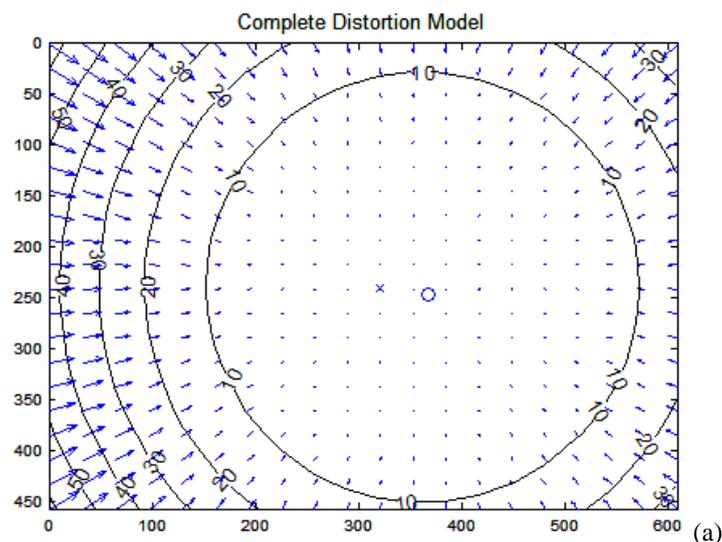


Figure 4-19 Software flowchart for the prospective head movement correction system for MRI. Calibration is all done offline and programmed with MATLAB. Real-time correction is programmed with C++ running in Linux. N is the number of total phase encoding steps for a scan.

Chapter 5 Results and Discussion with the Stereo Motion Correction System for MRI

5.1 Camera Calibration Test

The cameras used in this project have 4 mm lenses and were positioned about 8cm away from the volunteer's face. Distortion of the images is very serious. Figure 5-1a shows the distortion model on each pixel of one of the stereo cameras. The blue cross in the center marked the image center and the blue circle marked the principal point. Each arrow represented the displacement of the pixel caused by lens distortion. The displacement was bigger at the corners of the image, which reached more than 50 pixels. It meant that without a correct camera calibration, the 3D estimation from stereo would not be accurate. Image (b) shows a checkerboard image acquired with the corresponding camera. The edges of the checkerboard pattern were not straight. Instead, they were curved because of the distortions. Image (c) shows the image after distortion correction, which recovered the straight checkerboard edges.



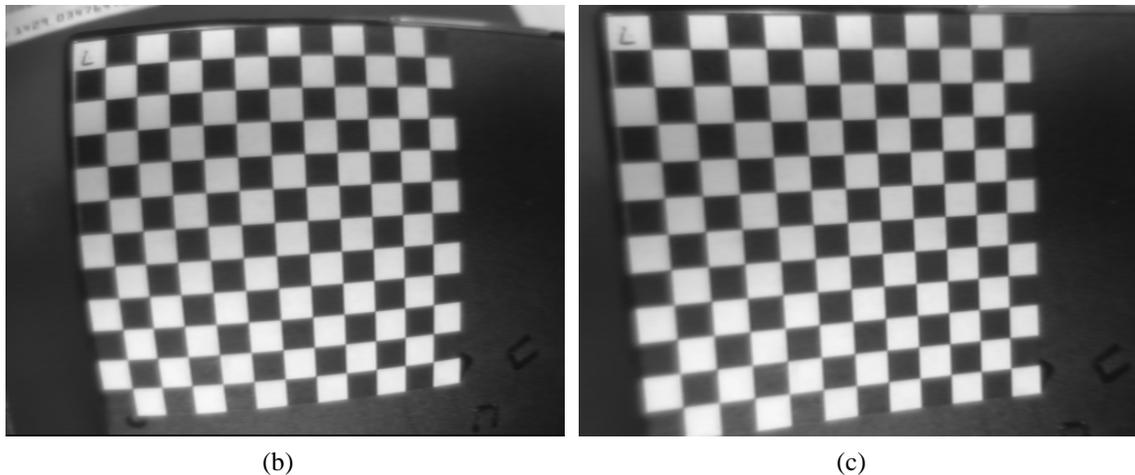


Figure 5-1 Lens distortion and correction for a camera. (a) Distortion model of the camera. The arrows and number on the curve mark the distortion direction and amplitude. (b) an image of a checkerboard pattern. The lens distortion is obvious especially close to the corners of the image. (c) lens distortion corrected image of b.

5.2 Stereo Reconstruction Validation

The checkerboard pattern was attached to a highly sensitive, 6-axis-positioning system (M-824 compact 6-axis-positioning system, Physik Instruments, GmbH & Co. KG, Germany) to test the stereo's accuracy. The hexapod was driven by six high-resolution actuators with sub-micron precision. Experiments were done outside of the scanner since the positioning system was not MR-compatible. The cameras were fixed firmly on top of the moving platform at a distance of about 8cm to simulate the position of the stereo with respect to the human face. The platform was controlled by a computer to move in the x, y, and z direction over known distances and rotate around the three axes with known degrees, where the x, y and z were defined as the two orthogonal axes of the platform, and the z axis was defined as pointing out of it. Camera images were acquired at each position to measure these movements. The discrepancy between the real and measured positions was evaluated

to determine the stereo's accuracy, which turned out to be $0.04\pm 0.03\text{mm}$, $0.03\pm 0.02\text{mm}$ and $0.02\pm 0.01\text{mm}$ in x, y and z direction, and errors of rotation around x, y and z axes were $0.03\pm 0.13^\circ$, $-0.02\pm 0.05^\circ$ and $0.01\pm 0.03^\circ$ respectively. This high accuracy is well within the highest available MRI resolution, which guarantees adequate motion correction.

5.3 Tracker Noise Evaluation

Since the motion correction was performed for each acquired k-space line, the tracking noise was also an important factor. Our real-time tracking was based on feature points, which provided sub-pixel accuracy but was also sensitive to noise in the video images. The tracking noise was evaluated by monitoring a stationary object for several minutes by tracking 3 feature points.

When the tracker was set up outside of the scanner, the noise was very small even by tracking only 3 feature points, as shown in Table 1. However, when the tracker was put inside the MRI, a significant increase of the error was observed, even without scanning. During a scan the noise level increased even further. This might be attributed to remaining interference of the MRI scanner with the video cameras. These findings suggest that there is an effect of the static magnetic field as well as of the switching gradients and/or RF. The maximum translation error exceeded 2mm, which was unacceptable for the high-resolution MRI motion correction.

Table 1 Accuracy and noise of the tracker when it was put outside or inside of the scanner.

	Tracker Outside			Tracker Inside (no scan)			Tracker Inside (during scan)		
	Mean	σ	Max	Mean	σ	Max	Mean	σ	Max
x (mm)	0.065	0.047	0.215	0.002	0.315	1.057	0.702	0.609	2.772
y (mm)	0.027	0.014	0.068	0.076	0.034	0.179	0.052	0.057	0.314
z (mm)	0.029	0.025	0.107	0.158	0.221	0.918	0.111	0.288	1.138
rx (°)	0.021	0.014	0.061	0.146	0.169	0.740	0.047	0.226	0.881
ry (°)	0.016	0.024	0.081	0.068	0.080	0.338	0.034	0.143	0.588
rz (°)	0.038	0.028	0.132	0.019	0.249	0.864	0.545	0.485	2.166

To reduce the tracking noise, multiple feature points were used. Figure 5-2 shows the maximum x-translation error as a function of the number of feature points. It shows that tracker noise decreased with increasing number of feature points. For our application, we opted to select around 20 points to keep the maximum error below 0.2 mm and 0.2 degree.

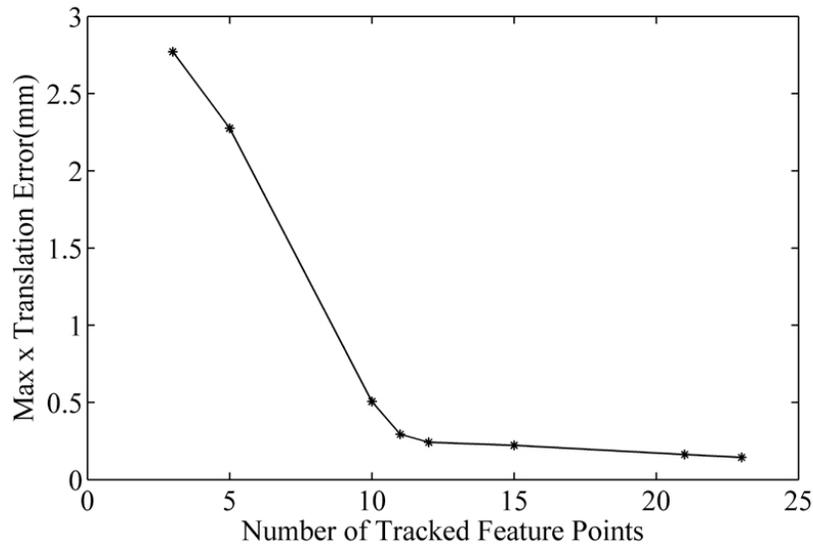


Figure 5-2 Noise of the tracker. The maximum x translation error as a function of the number of tracked feature points.

However, the number of feature points on the human face detectable with both video cameras was often below 20, in part because of the cameras' limited field of

view of about $13 \times 10\text{cm}^2$. To overcome this we decided to place a sticker on the subject's forehead that displayed a number of white triangles on a black background.

5.4 Calibration Accuracy Between Stereo and MRI Coordinate System

The calibration phantom (Figure 4-2) was used to test the calibration accuracy between the stereo system and the MRI scanner. First, the phantom was put at an arbitrary position. The stereo measured the corner positions in 3D and transformed them to MRI coordinates based on the calibration results. Then 3D MR images were acquired, from which the 3D MRI coordinates of the corners were determined. Accuracy was obtained by measuring the discrepancy between the estimated position (obtained by using the camera model and calibration parameters) and real ones (measured from MR images). This procedure was repeated at different arbitrary phantom positions. The accuracy of these points was found to be $0.06 \pm 0.05\text{mm}$, $0.10 \pm 0.08\text{mm}$ and $0.15 \pm 0.13\text{mm}$ in the x, y, and z direction in the MRI coordinate system respectively, which should be sufficient for the motion compensation task.

The results are also visualized in Figure 5-3 where rotation and translation parameters are calculated from these points at different phantom positions. Solid lines show parameters estimated using MRI images. Dotted lines show parameters estimated using camera images but further transformed into MRI coordinate system. The two sets of estimation matched very well, which means the calibration between the two coordinate systems was accurate.

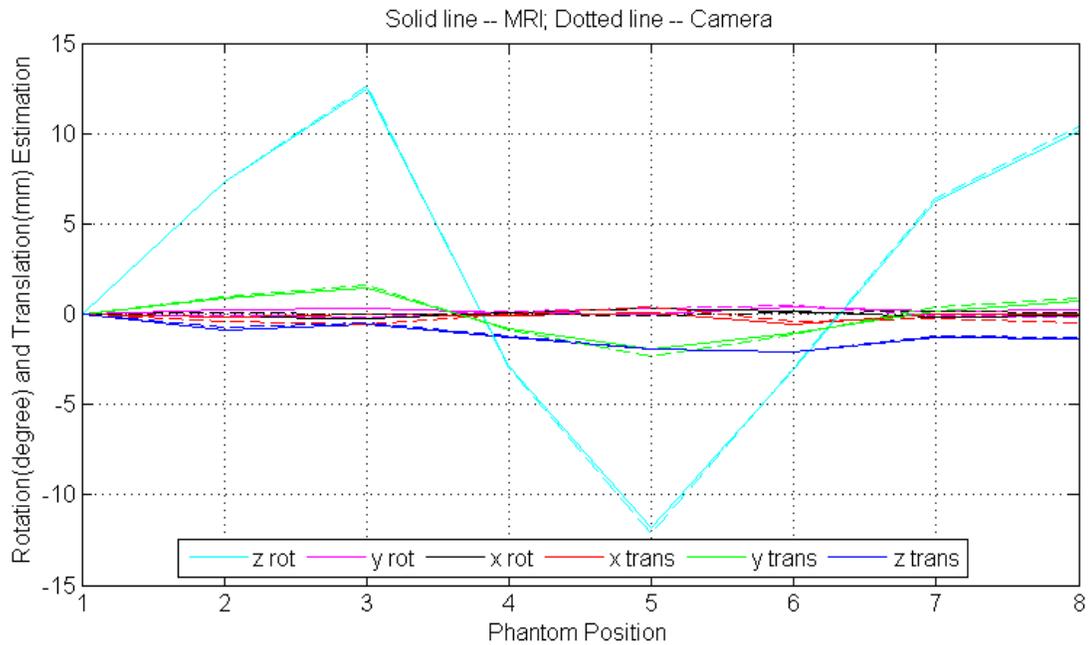


Figure 5-3 Rotation and Translation parameters estimated from MRI images (solid lines) and from camera images transformed into MRI coordinates (dotted lines). The two sets of lines matching very well means the calibration between camera and MRI is accurate.

5.5 Pulse Sequence Test with a Phantom

To test the accuracy of RF frequency and phase adjustment for translational displacement in slice-selection, phase-encoding and frequency-encoding directions, a high resolution phantom was scanned with our modified GRE sequence. (Rotation doesn't need to be tested for the sequence because it only requires rotation matrix adjustment but doesn't need any frequency or phase compensation.) Halfway through the experiment, a translation of 10mm was executed and the scan was continued until the end with the phantom staying at the new position. The particular motion pattern was used to maximize errors near the center of k_y -space, which causes the most salient artifacts.

Figure 5-4 shows the results of three sets of scans (256×128 matrix, $TR=100\text{ms}$, $TE=6.9\text{ms}$, $FA=30^\circ$). The first column represents the control scan without motion, second and third columns represent scans with displacements, without correction and with correction respectively. Displacements were executed all in the $\mathbf{O_MZ_M}$ direction, so for axial scans (a-c), displacements were in the slice-selection direction; for coronal scans (d-f) with frequency-encoding superior-inferior ($\mathbf{O_MZ_M}$), displacements were in the frequency-encoding direction; for coronal scans (g-i) with phase-encoding superior-inferior ($\mathbf{O_MZ_M}$), displacements were in the phase-encoding direction. It can be seen that for all cases, the displacements were compensated successfully, which means the measurement of scanner specific time Δt and $\Delta t'$ is accurate and the strategy of RF frequency and phase adjustments are correct.

Note that image (g-i) show an aliasing artifact because of the small FOV prescribed in the phase-encoding direction, which, however, does not affect the displacement compensation test.

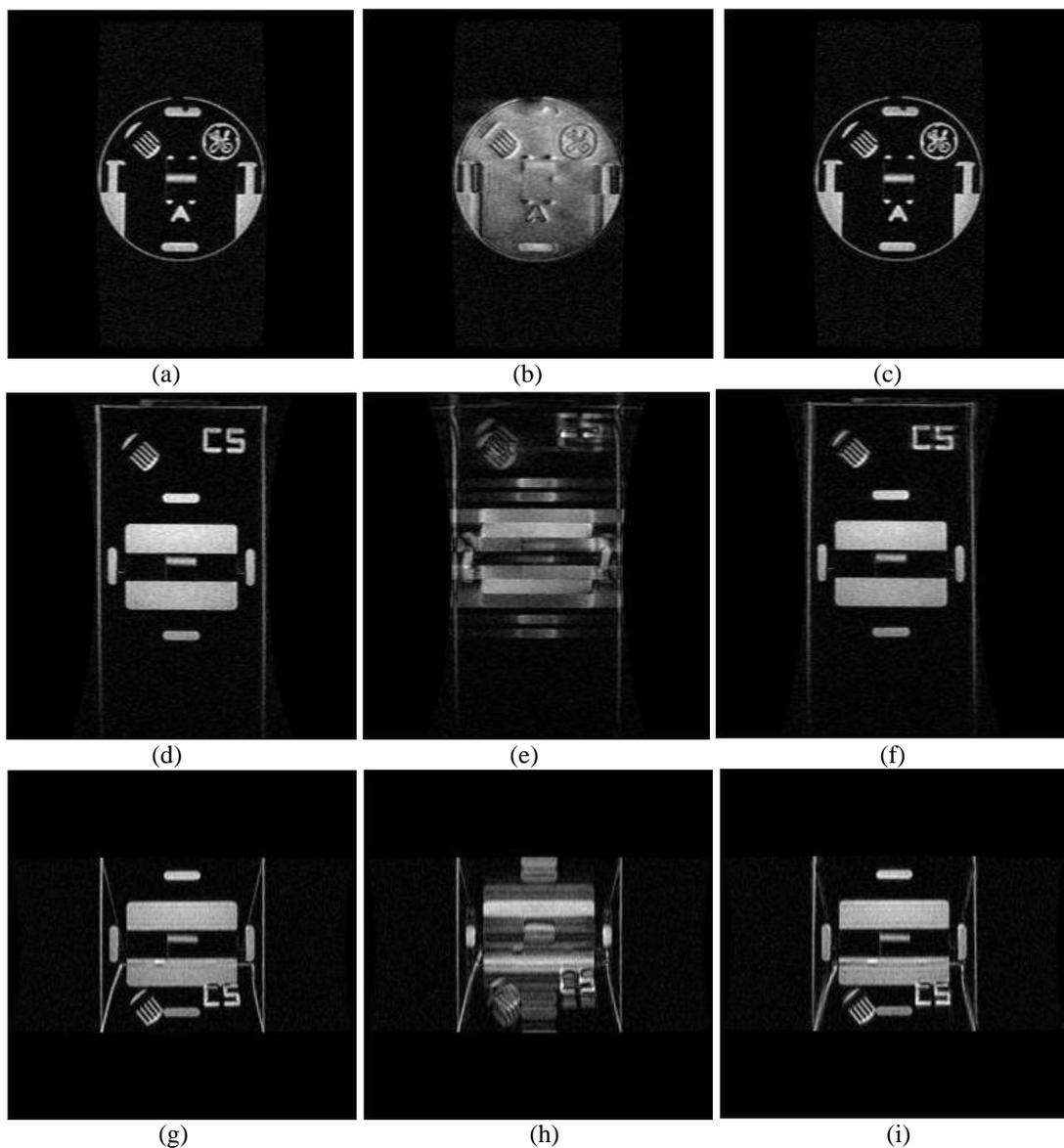


Figure 5-4 Translation compensation of a phantom. (a,d,g) were scanned statically; all others were scanned with the phantom locating in one position for the first half of the scan and the second half in another position 10mm apart. (b,e,h) were scanned without position correction, but (c,f, i) were with correction. Displacements in (a-c) were in slice-selection direction; (d-f) in frequency-encoding direction, and (g-i) in phase-encoding direction.

5.6 System Performance Test

With all the separate parts of the prospective motion correction system validated, the performance of the system as a whole needed to be tested. The optimal way is to evaluate hybrid space data of a scan with motion. Because ideally without motion, all lines in the phase-encoding direction should be exactly aligned. Imperfect alignment would indicate imperfect motion compensation. If motion can be corrected fully, then an alignment should be achieved.

We set phase encoding gradient to zero and reconstructed the hybrid data (resolution 0.23mm^2 , TR=100ms, TE=25ms, FA=10°). Figure 5-5a shows motion induced by the breathing of a volunteer recorded with the tracker during the scan. Image (b) was scanned without real-time motion correction and the lines in the phase encoding direction shifted during the scan. When prospective motion correction was applied for image (c) acquisition, those lines were all aligned.

Since the in plane resolution was only 0.2mm, the perfectly aligned lines in Figure 5-5b suggested that the error of the whole correction system should be less than a pixel or 0.2mm. Therefore, this is a quick and coarse way to validate the accuracy and reliability of the system.

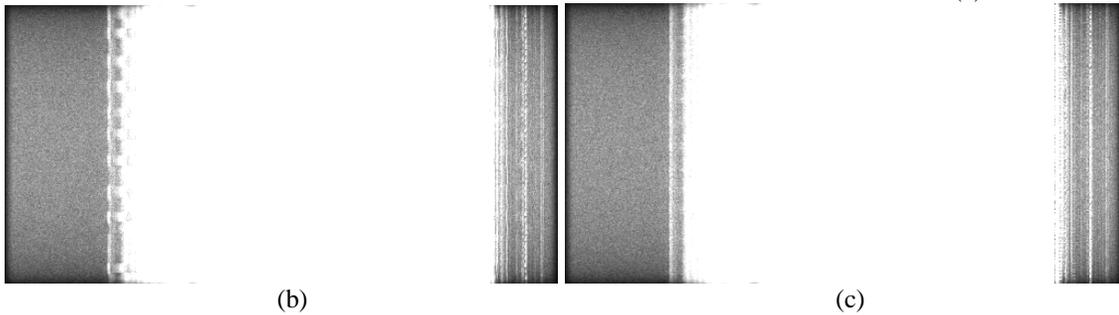
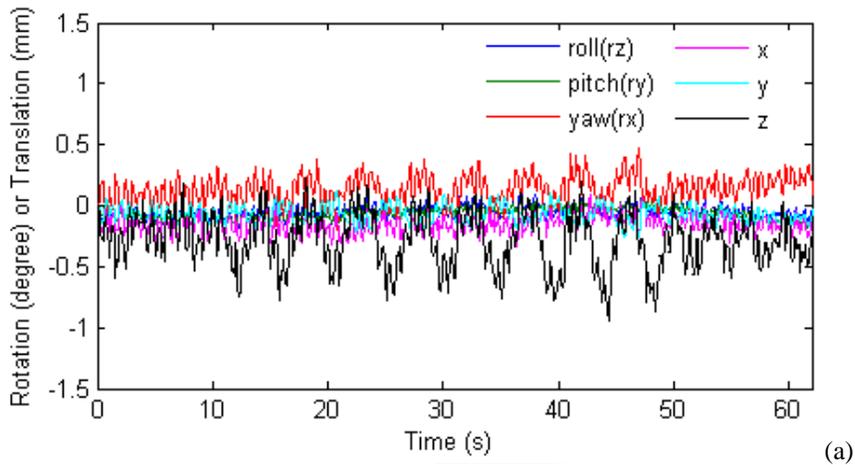


Figure 5-5 The MR hybrid image without phase encoding gradient. (a) Small breathing induced head movement was observed during the scan of the volunteer. (b) without motion correction; (c) with motion correction.

5.7 Volunteer Study

The efficacy of the tracking system was studied on a normal volunteer under an institutional review board-approved protocol. A nominal in-plane spatial resolution of $0.23 \times 0.23 \text{mm}^2$, a slice thickness of 3mm, 300ms TR, 25ms TE and 30° flip angle were used. Four motion conditions were studied: one with cyclic breathing-related head motion (Section 5.7.1), one with minimal motion (Section 5.7.2), one with large stepwise motion (Section 5.7.3), and one with slow gradual motion (Section 5.7.4).

5.7.1 Breathing Related Motion

During this experiment, the volunteer was instructed to not pay any effort to lay still. Motion parameters reported by the tracker (Figure 5-6a,b) showed cyclic head motion with a period consistent with the respiratory cycle. The observed rotation was mainly around the x-axis and the translation along the z-axis (in the MRI physical coordinate system), which suggested a nodding like motion. The scan was performed twice, once without motion correction and once by using the prospective motion correction technique proposed here. Figure 5-6c,e showed that even though the motion was only about 2mm and 1°, ghosting and blurring artifacts were clearly observed in the high-resolution images (0.23 x 0.23mm² in-plane resolution). The prospective line-by-line motion corrected image using our real-time compensation system showed much improved image quality and a virtual elimination of ghosting artifact (Figure 5-6d,f) as compared to the uncorrected images (Figure 5-6c,e). The reacquisition rate was 2.3%.

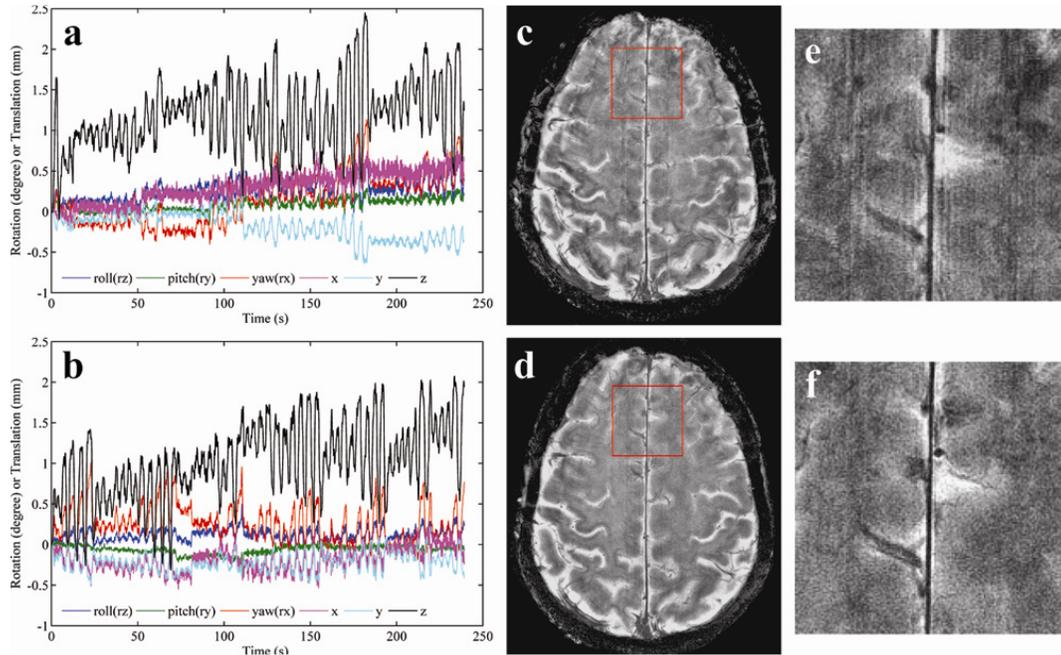


Figure 5-6 Prospective breathing induced motion correction for the 2D gradient echo acquisition. Images (a,b) show the evolution of motion parameters reported by the tracker during the acquisition of images (c) and (d) respectively. Image (c) was scanned without motion correction, and (d) was with motion correction; (e) and (f) were zoomed in on an area in the frontal brain indicated by the red square in images (c) and (d) respectively.

5.7.2 Minimal Motion

During this experiment, the volunteer was instructed to stay as still as possible. Despite the efforts of the volunteer, the motion parameters reported by the tracker (Figure 5-7a,b) showed a small gradual and jittering motion. This scan was also repeated twice, once without (a,c,e) and once with (b,d,f) motion correction. Both uncorrected (c,e) and corrected (d,f) images have similar quality and show little evidence of ghosting. The reacquisition rate of Figure 5-7d was 4.6%.

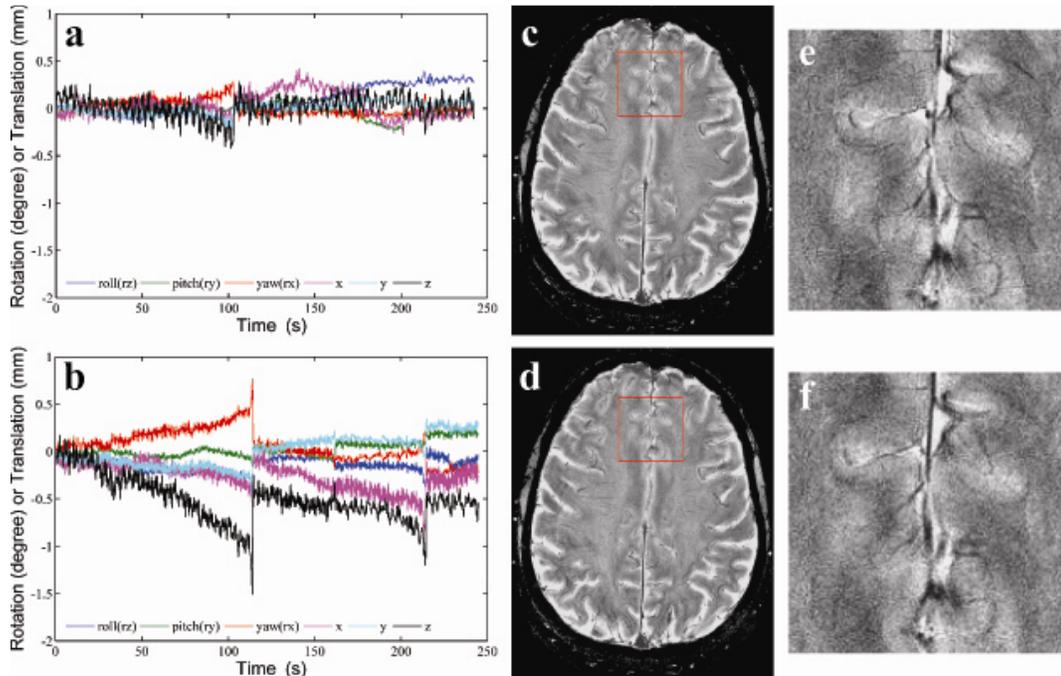


Figure 5-7 Prospective motion correction data obtained for a volunteer that attempted to minimize head motion. Images (a,b) show the motion parameters reported by the tracker during the acquisition of images (c) and (d) respectively. Image (c) was scanned without motion correction, and (d) was with motion correction; (e) and (f) magnify an area in the front of the brain (the area marked by the red square in (c) and (d) respectively).

5.7.3 Large Stepwise Motion

Experiments in the presence of strong intentional motion demonstrated the robustness of the prospective motion correction technique (Figure 5-8). The motion parameters reported by the tracker indicated a range of motion of close to 10 mm translation and 10 degree rotation. The severe ghosting and blurring artifacts observed without correction (e,h) were almost entirely avoided with prospective motion correction (f,i), and the image quality of (f,i) was very similar to (d,g), which were acquired in the absence of intentional motion and with motion correction on (motion parameters in a). The reacquisition rate for (f,i) was 10%.

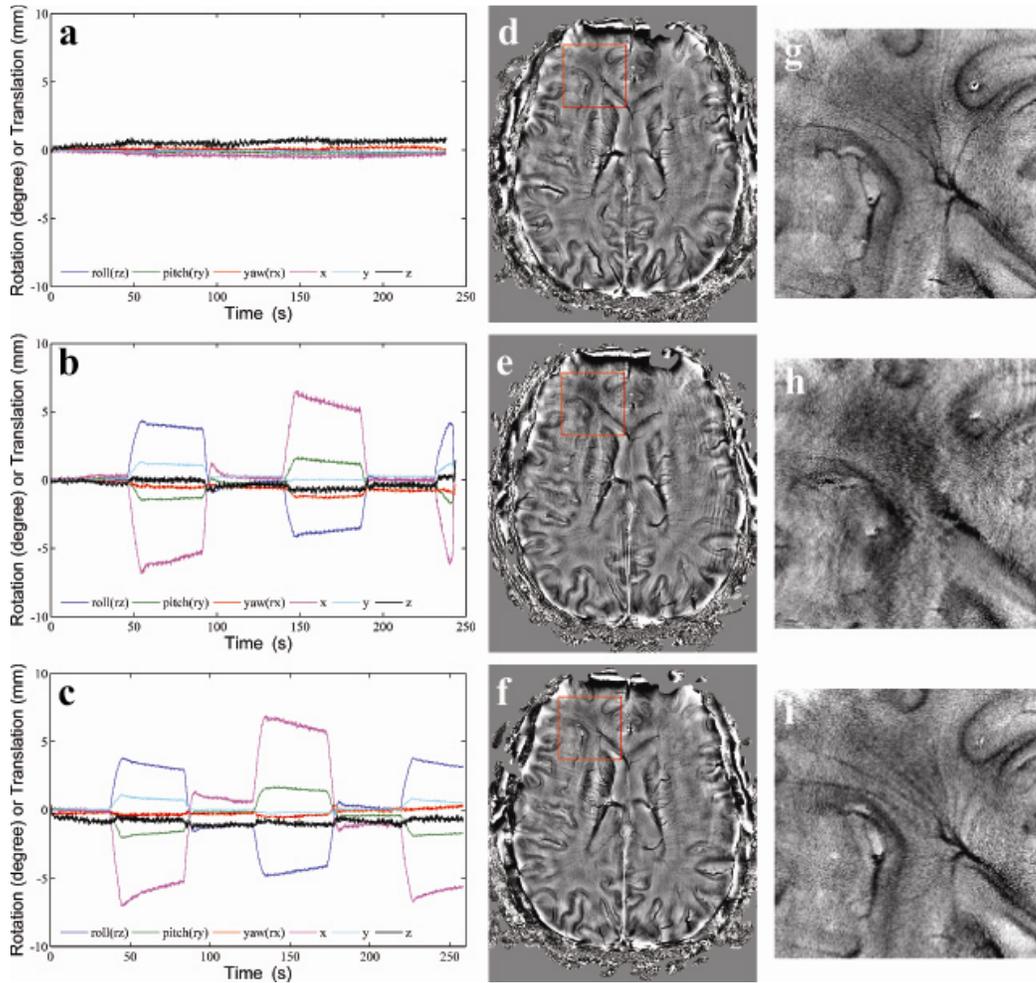


Figure 5-8 Example of the performance of the prospective motion correction method in the presence of large stepwise motion during the acquisition of multi-shot 2D gradient echo images on a normal volunteer. Images (a-c) show the motion parameters reported by the tracker during the acquisition of images (d-f) respectively. Image (d) was scanned without intentional motion during the scan, whereas (e,f) were scanned in the presence of motion; (e) was acquired without correction but (f) was acquired using the proposed method. (g-i) show additional detail in an area in the front of the brain that was indicated by the red square in images (d-f).

5.7.4 Slow Motion

The final motion type that was investigated was slow motion resulting in large displacements and/or substantial rotation angles. Even without any reacquisition this type of motion was handled well by the prospective motion correction system. The motion parameters reported by the tracker (Figure 5-9b,c) showed that the range of

the motion was close to 10 mm of translation and 10 degrees of rotation. The reacquisition rate for (f,i) was 0% since the reacquisition mode was turned off. Image quality of (f,i), when motion correction was used, was notably better than (e,h), acquired in the absence of motion correction. Furthermore, image quality of (f,i) was similar to (d,g), acquired in the absence of intentional motion (motion parameters in a).

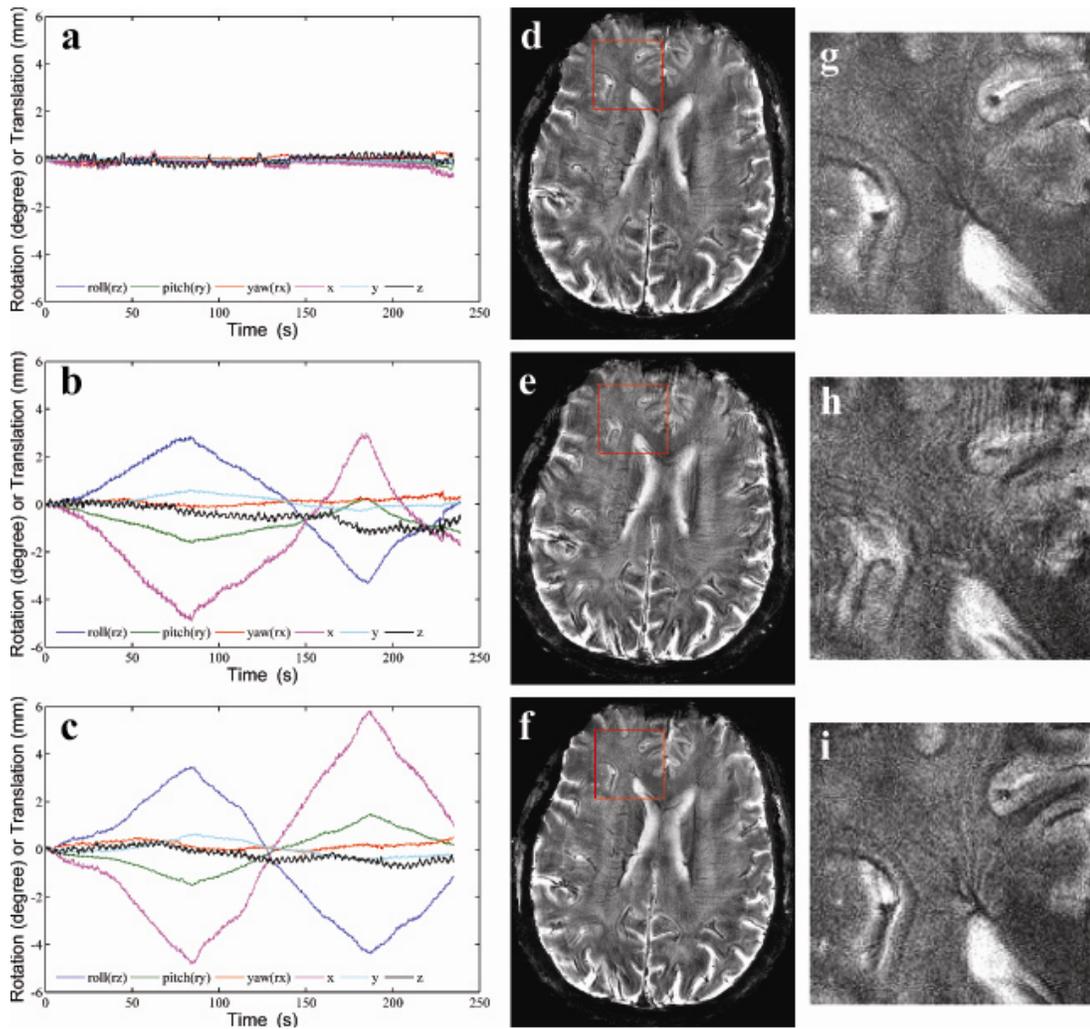


Figure 5-9 Prospective motion correction in the presence of slow motion. The motion parameters reported by the tracker during the scan of images (d-f) are displayed in (a-c), respectively. Image (d) was acquired in the absence of intentional head motion, (e,f) during slow head motion, where, unlike (e), (f) was acquired with the proposed correction method. A magnification of the area in the red box in (d-f) is shown in (g-i).

5.8 Discussion

In this study, a method for the prospective correction of 3D motion for high-resolution MRI was developed and evaluated. Preliminary data obtained on a human volunteer show substantial image quality improvement for high resolution gradient echo imaging at 7T, in particular under conditions of periodic head motion related to breathing and intentional head rotation.

To obtain adequate tracking accuracy, the cameras were put right above the MRI receiver coil, monitoring the human face from a very small distance. Since the stereo 3D reconstruction accuracy depends on the angle and relative distance between the two cameras, the best setup of the two cameras should be explored further for better accuracy.

The accuracy of the calibration between the tracker and the MRI scanner is also essential for system performance. Since all the motion parameters were estimated in the camera coordinate frame, they had to be transformed precisely to the MRI coordinate frame to be used in the sequence. However, calibration was difficult in this system because the tracker can only see the surface of the calibration object, whereas the MRI can only see its internal structure. A previous publication (25) dealt with this issue in an iterative way. They used three gel-filled reflective spheres for estimating the initial coordinate transformation, and then iteratively refined this transformation based on the internal structures of the phantom. This method, though accurate, took more than 30 minutes to accomplish. This limitation was addressed in our project by designing a phantom with high resolution, well ordered structures. Since every parameter of the phantom was known, the relationship between the external

appearance (black-and-white checkerboard) and internal MR-visible structure was easily established, which simplified the calibration and shortened the calibration time to several minutes.

The noise of the tracker was another important factor essential for reliable correction of MRI data. It was demonstrated here that at least 20 feature points needed to be tracked for adequate precision. Because of the setup of our cameras, only a part of a face, lacking a large number of feature points, could be seen from the cameras. This was the reason for using a “tracking target” mounted to the forehead. If the cameras could be put further away from the face, direct tracking of the face curvatures could be used instead of these feature points. Fixing the cameras to the inside of the magnet bore, instead of to the transmit coil as was done here may be a solution. Another solution to this issue of limited field of view would be the use of cameras with a fish-eye lens, which would give a larger view for a given camera distance from the face. However, the camera distortion correction may be more complicated. It has also been demonstrated that the noise comes from the interference between the cameras and the MRI scanner, so a better shielding or hardware filtering could be a solution too.

Due to the tracker’s relative low frame rate some fast motion could not be adequately corrected. For this reason, a reacquisition function was added to the sequence in this project. The number of re-acquired lines (reacquisition rate) was dependent on the type of head motion and quite variable, e.g. 2.3% for the conditions presented in Figure 5-7d and 10% for those in Figure 5-8f. Considering the resulting improvement in image quality and the small increase in total imaging time, this

sacrifice appears worthwhile. Nevertheless, improved tracking speed might reduce or eliminate the need for re-acquisition. This was confirmed in the “slow motion” experiment, during which the re-acquisition was turned off. Under this condition, the tracking speed exceeded the head motion speed, resulting in excellent image quality without any reacquisition, as was seen in Figure 5-9(f,i). With the continuing increase of computational power, and improved motion tracking algorithms, it is therefore expected that the need for re-acquisition will be reduced. One possible area for improved tracking is the use of predictive estimates for object motion, for example by using Kalman filtering.

Although the motion correction scheme presented here was effective in reducing the effects of motion on image quality (e.g. compare Figure 5-6d and Figure 5-8f with correction versus the not corrected versions in Figure 5-6c and Figure 5-8e), there remain possibilities for improvement. For example, the motion-corrected images in Figure 5-8f and Figure 5-9f are slightly inferior to the corresponding images acquired during the minimal head motion (Figure 5-8d and Figure 5-9d). This apparently imperfect motion compensation could be due to a number of factors, for example phase changes due to small changes in local B_0 amplitude, or subtle changes in B_1 amplitude and phase. The contribution of these effects and their potential mitigation remains to be investigated.

In this current method, quantitative validation of image quality improvement on volunteer study has not been fulfilled. One of the reasons is that since image resolution is only 0.2mm, the reference image (no motion) can hardly be exactly the same slice as in the with motion case, which makes quantitative comparison very

difficult. To get a better validation, we need to test this technique on more volunteers and ask neurologists to give objective judgments on the image quality.

Chapter 6 Prospective Motion Correction for MRI using a Single Camera

A substantial amount of clinical and research MRI images are rendered unusable because of subject motion. To solve this problem, prospective motion correction based on external tracking systems has been proposed. Stereo optical tracking system has been shown to give good real-time motion correction with reasonable accuracy. However, the calibration of this system is very complicated and time consuming, as it requires a camera system calibration as well as a calibration between cameras and MRI system using dedicated phantoms. We therefore propose an alternative motion correction method for MRI that does not require calibration and can work with just a single video camera.

6.1 System Setup

All experiments were performed on a GE (Milwaukee, WI, USA) 7T MRI system. An MR compatible infrared camera (MRC Systems GmbH, Germany) was fixed on a holder right above and in front of the head coil (Figure 6-1). Six infrared emitting diodes were used to illuminate the field of view. The distance from the camera to volunteer's face was around 8cm. The high spatial resolution of this setup (320 x 240 pixels, about 13 x 10cm² field of view focusing mostly on the nose of the volunteer) allowed the detection of sub-millimeter movements.

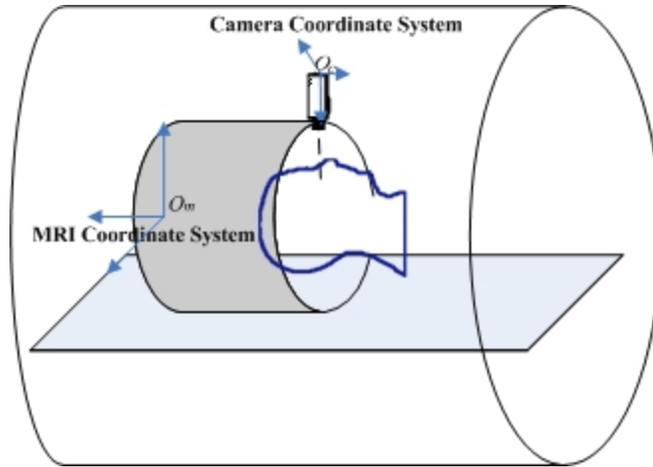


Figure 6-1 System setup. The camera was fixed right on top of the head coil.

6.2 Method

Rather than using phantom calibration scans, this method uses training scans to estimate head position and link the camera image measurement to the MRI scanner coordinate system.

During the training scan, the subject was asked to slowly move his head in the directions that were least restrained, including rotations in the axial and sagittal planes. N EPI (Section 2.1.3; parameters: 128×96 voxels over $24 \times 18 \text{ cm}^2$, 11 3mm slices with 0.5mm gap, TE 50ms, TR 1.2s) and camera images were acquired simultaneously (Figure 6-2). From the EPI images, motion parameters in the MRI coordinate system (3 rotations and 3 translations) between each volume were calculated with image registration method. Since camera images were acquired simultaneously with EPI, each camera image in the training data thus corresponded to a motion vector. The training data were then sorted based on the motion parameters to help speed up the real-time search. An ROI region (the nose) was also selected based on these training camera images.

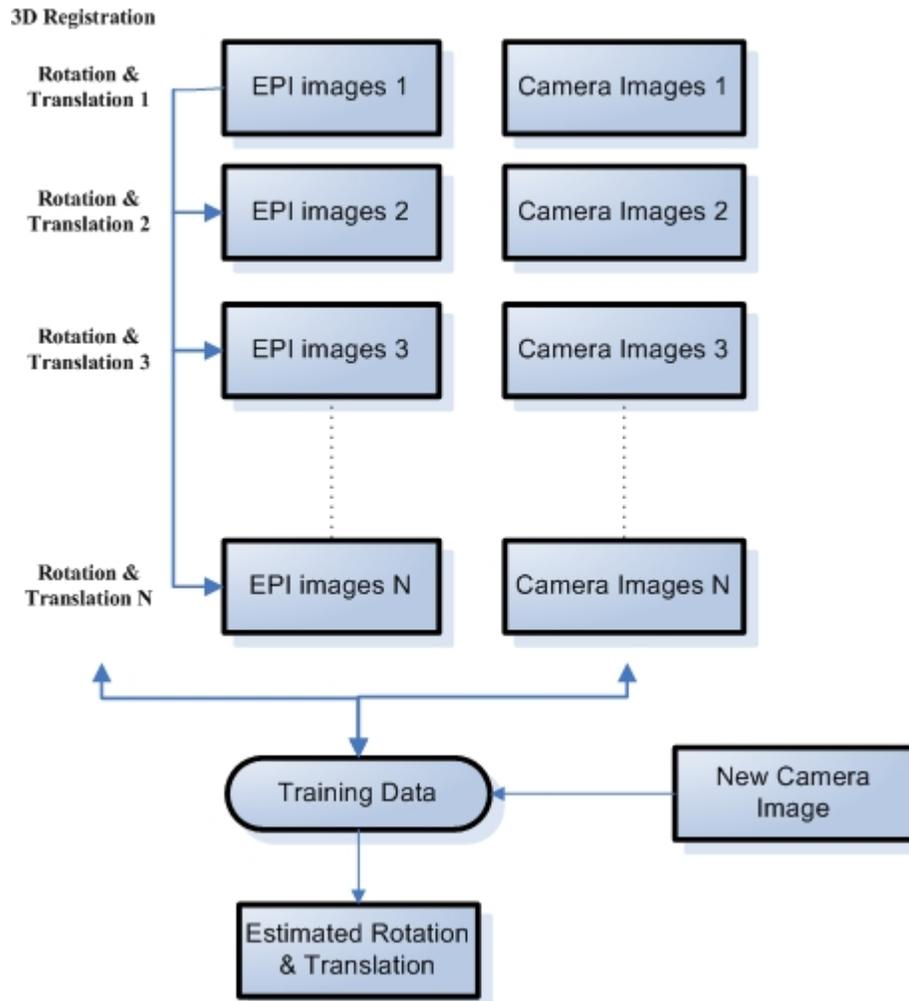


Figure 6-2 Method of the motion estimation using a single camera. N training EPI and camera images were acquired. From EPI images, 6DOF motion parameters were calculated from registration for each training data. During the real-time scan, the most similar camera image in the training data was found for the newly acquired camera image. The corresponding motion parameters from EPI were sent to MRI pulse sequence for prospective motion correction.

During the real-time scan, the correlation between the ROI of the newly captured camera image and the ones in the training camera images was calculated to find the most similar one. The corresponding motion parameters of that training image served as the estimation of the current head position. These motion parameters were then sent to the MRI scan computer at 10Hz speed. A gradient echo sequence

was modified to incorporate motion correction by adjusting gradient rotation and RF frequencies and phases before each RF excitation (Section 4.5).

6.2.1 Motion Computation

The motion parameters can be calculated from the N EPI images in the training data through image registration. Image registration involves the mapping between a pair of images, so that one image after transforming with the mapping parameters can be matched to the other one.

$$\mathbf{Y}_i = \mathbf{R}_t \mathbf{X}_i + \mathbf{T}_t \quad [6.1]$$

Where $\mathbf{X}_i = [x_{i1}, x_{i2}, x_{i3}]^T$ and $\mathbf{Y}_i = [y_{i1}, y_{i2}, y_{i3}]^T$ are the corresponding points in the two images in MRI physical coordinate system (after transformation as in Section 9.2) and $\mathbf{R}_t, \mathbf{T}_t$ are the rigid body transformation between them.

The registration parameters can therefore be determined by minimizing the sum of squared intensity differences between images (Friston et al., 1995; Friston, 2007):

$$\varepsilon^2 = \sum_i \|\mathbf{Y}_i - \mathbf{R}_t \mathbf{X}_i - \mathbf{T}_t\|^2 = \sum_i \|\mathbf{Y}_i - f(\mathbf{X}_i; \mathbf{p}_t)\|^2 \quad [6.2]$$

Where f is the transformation function, and $\mathbf{p}_t = [p_{t1}, p_{t2}, \dots, p_{t6}]^T$ is the vector containing three rotation and three translation parameters.

Given n corresponding points in the two images, Equation 6.2 can be solved using Gauss-Newton algorithm (Section 9.3). For the m^{th} iteration, the vector \mathbf{p}_t is updated as:

$$\mathbf{p}_t^{(m+1)} = \mathbf{p}_t^{(m)} + (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \mathbf{b} \quad [6.3]$$

$$\text{Where } \mathbf{J} = \begin{bmatrix} \frac{\partial f(\mathbf{X}_1; \mathbf{p}_t^{(m)})}{\partial p_{t1}} & \frac{\partial f(\mathbf{X}_1; \mathbf{p}_t^{(m)})}{\partial p_{t2}} & \Lambda & \frac{\partial f(\mathbf{X}_1; \mathbf{p}_t^{(m)})}{\partial p_{t6}} \\ \frac{\partial f(\mathbf{X}_2; \mathbf{p}_t^{(m)})}{\partial p_{t1}} & \frac{\partial f(\mathbf{X}_2; \mathbf{p}_t^{(m)})}{\partial p_{t2}} & \Lambda & \frac{\partial f(\mathbf{X}_2; \mathbf{p}_t^{(m)})}{\partial p_{t6}} \\ \mathbf{M} & \mathbf{M} & \Lambda & \mathbf{M} \\ \frac{\partial f(\mathbf{X}_n; \mathbf{p}_t^{(m)})}{\partial p_{t1}} & \frac{\partial f(\mathbf{X}_n; \mathbf{p}_t^{(m)})}{\partial p_{t2}} & \Lambda & \frac{\partial f(\mathbf{X}_n; \mathbf{p}_t^{(m)})}{\partial p_{t6}} \end{bmatrix} \quad (3n \times 6 \text{ matrix}) \text{ and}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \mathbf{Y}_1 - f(\mathbf{X}_1; \mathbf{p}_t^{(m)}) \\ \mathbf{Y}_2 - f(\mathbf{X}_2; \mathbf{p}_t^{(m)}) \\ \mathbf{M} \\ \mathbf{Y}_n - f(\mathbf{X}_n; \mathbf{p}_t^{(m)}) \end{bmatrix} \quad (3n \times 1 \text{ vector}).$$

For each iteration step, \mathbf{Y}_i are re-sampled using the new transformation matrix. The iterative process is repeated until $\boldsymbol{\varepsilon}^2$ can not be decreased or for a fixed number of iterations. This approach is based on the assumption that the starting estimate is close to the correct solution and the image is smooth, which is the case in our application because the motion between successive images acquired in the training data is very small and the relative low resolution EPI images are smooth. Therefore the possibility that a solution stuck in a local minimum instead of a global minimum is small

6.2.2 Best Match Search

The search for the camera image in the training data that most closely matched the newly acquired image was realized through a similarity measure. The correlation coefficient (CC) is optimal in similarity measure as defined in Equation 4.48, reiterated here for a correlation between images A and B within a window W :

$$CC = \frac{\sum_{\mathbf{x} \in W} (A(\mathbf{x}) - \bar{A})(B(\mathbf{x}) - \bar{B})}{\sqrt{\left(\sum_{\mathbf{x} \in W} (A(\mathbf{x}) - \bar{A})\right)^2} \sqrt{\left(\sum_{\mathbf{x} \in W} (B(\mathbf{x}) - \bar{B})\right)^2}} \quad [6.4]$$

Where \bar{A} and \bar{B} are the mean pixel values of image A and B within the window W respectively. To speed up the computation, the correlation in image space can be calculated in frequency space using FFT.

Let $\hat{A}(\mathbf{x}) = A(\mathbf{x}) - \bar{A}$, $\hat{B}(\mathbf{x}) = B(\mathbf{x}) - \bar{B}$, then the numerator in Equation 6.4 is

$$CC_n = \sum_{\mathbf{x} \in W} \hat{A}(\mathbf{x}) \hat{B}(\mathbf{x}) = \sum_{n \in 0}^{N-1} \hat{A}(n) \hat{B}(n) \quad [6.5]$$

Where N is the size of the window. It can be calculated through $\hat{F}_A(k)$, Fourier transform of $\hat{A}(n)$; and $\hat{F}_B(k)$, Fourier transform of $\hat{B}(n)$.

Since

$$\begin{aligned} IFFT\left(\hat{F}_A(k) \cdot \hat{F}_B^*(k)\right) &= \frac{1}{N} \sum_{k=0}^{N-1} \hat{F}_A(k) \cdot \hat{F}_B^*(k) e^{i2\pi nk / N} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} \hat{A}(n) e^{-i2\pi nk / N} \cdot \sum_{n'=0}^{N-1} \hat{B}(n') e^{i2\pi n'k / N} e^{i2\pi nk / N} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \hat{A}(n) \hat{B}(n') \left(\sum_{k=0}^{N-1} e^{-i2\pi(n-n'-m)k / N} \right) \end{aligned}$$

The quantity in parentheses is 0 for all values of n' except those of the form $n' = n - m + pN$, where p is any integer. At those values, it is 1. Therefore,

$$IFFT\left(\hat{F}_A(k) \cdot \hat{F}_B^*(k)\right) = \sum_{n=0}^{N-1} \hat{A}(n) \hat{B}(n - m + pN) \quad [6.6]$$

By comparing Equation 6.6 with 6.5, the correlation coefficient can be calculated as follows: convert the two images to 1D vectors, FFT the two vectors, multiply one resulted transform by the complex conjugate of the other, inverse FFT

transform the product and the first component ($m=0$) is the correlation coefficient. All other components of $IFFT(\hat{F}_A(k) \cdot \hat{F}_B^*(k))$ are the values of the correlation at different lags.

6.3 Result

6.3.1 Accuracy of the Motion Estimation

To estimate the accuracy of the method, several minutes of data were acquired, consisting of 145 EPI volumes and the corresponding 145 camera images. During this time, the subject slowly moved his head in the directions that were least restrained, mainly rotations in the axial plane. All the EPI images in the training set were registered to a reference (the first volume in the training set can be used as a reference) to get the motion parameters. The range of rotation around Z_M axis was about 17 degrees and translation in X_M direction was 18mm (Figure 6-3).

Figure 6-4 shows the correlation of two randomly chosen images in the training data set with all images. The curve suggests that the correlation is higher when positions are close, which means CC is a good measurement for most closely match image search.

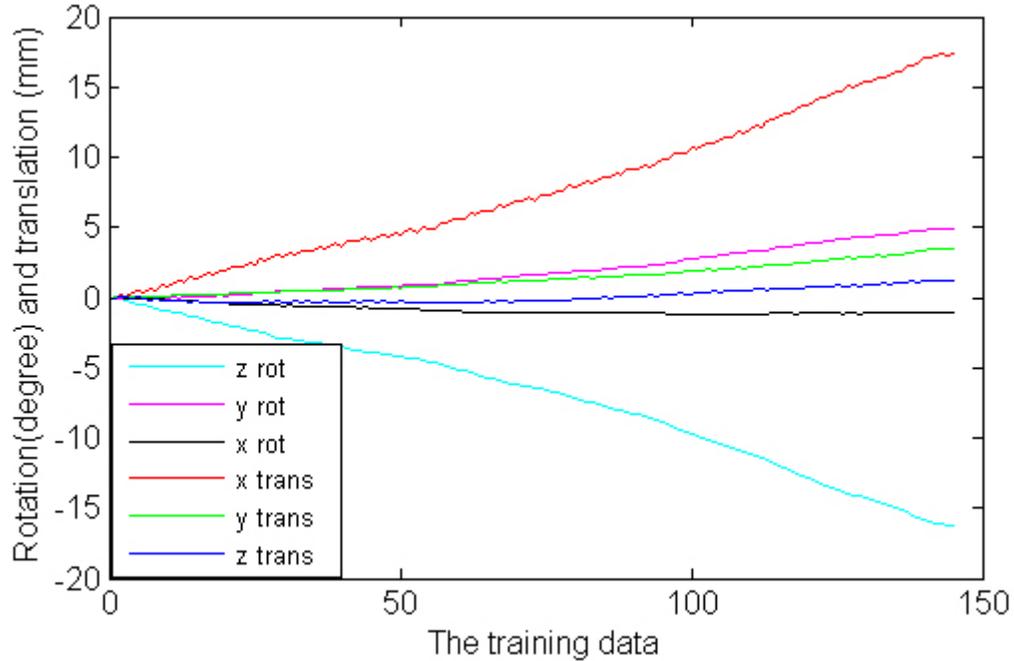


Figure 6-3 Motion parameters from the training data set. The total number of training images is 145. The motion is mainly rotation around Z_M axis and translation in X_M axis.

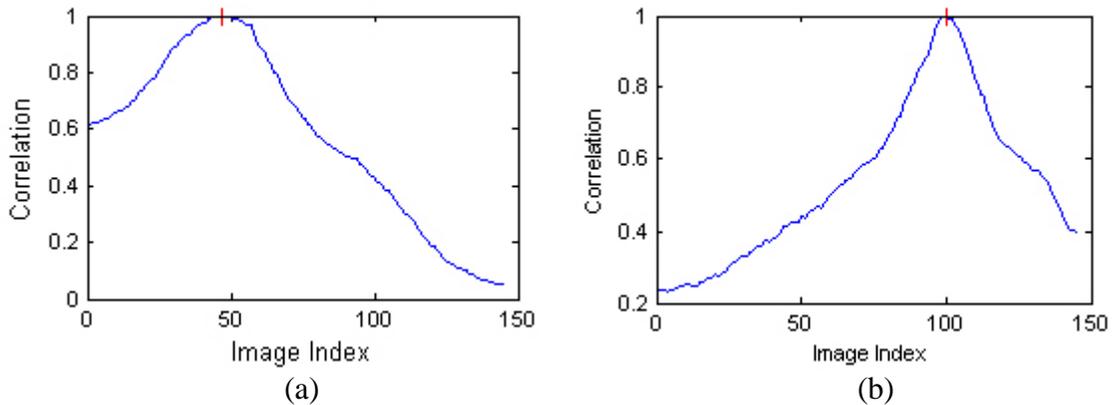


Figure 6-4 Correlation of two randomly chosen images with all training images. Red crossings mark the chosen image, (a) the correlation between image No. 47 with others; (b) the correlation between image No. 100 with others.

To test the accuracy of motion estimation, the data were divided into two sets, set A and set B. For every camera image in set B, the most similar image in set A was determined based on correlation. The motion parameters of this image in set A served as motion estimate for that image in set B, whereas the true motion was derived from

registration of the EPI image in set B to the reference. Difference between camera derived position and EPI registration parameters was used to validate motion estimation accuracy.

A k -fold cross-validation method (Duda et al., 2001) was used, where k was chosen to be 10. Training images were divided into k folds. For each trial, one fold images were used as set B, while the remaining $k-1$ fold images formed set A. After repeating the validation k times by using all different sets of B, the average error across the k trials was computed. Translation error in X_M , Y_M and Z_M direction is $0.021 \pm 0.249\text{mm}$, $0.004 \pm 0.052\text{mm}$ and $0.001 \pm 0.066\text{mm}$ and rotation error around X_M , Y_M and Z_M axis is $0.001 \pm 0.026^\circ$, $-0.001 \pm 0.050^\circ$, and $-0.006 \pm 0.129^\circ$ respectively.

However, note that this error estimation was only true if the sampled motion space covered the whole space the volunteer could move inside the head coil. For instance, if the sampled space during training stage only covered rotations about Z_M , while during the real-time scan, the volunteer moved around X_M axis, even though a most closely matched image in the training data could be determined based on the highest correlation coefficient, the estimated motion parameters were not correct. Therefore the key issue for this method is to sample as complete as possible the motion space (the volunteer moves as much as he can) during the training stage.

6.3.2 Volunteer Study

A GRE sequence modified to incorporate motion correction by adjusting gradient rotation and RF frequencies and phases before each RF excitation was used on a normal volunteer under an institutional review board-approved protocol (parameters: 512×384 voxels over 24×18 cm², 3mm slice thickness, TR=100ms, TE=30ms). During the real-time scan, the most similar camera image in the training data was found for the newly captured camera image based on the correlation coefficient inside the ROIs (nose area chosen from the training data). The corresponding motion parameters of that training image were sent to the MRI scan computer at 10Hz speed for prospective motion correction.

Experiments were performed with and without motion correction each with similar motion. Results are shown in Figure 6-5. The image quality of (b) is similar to (a), and much better than (c). The results suggest that the new system has good precision to be used in high resolution MRI. Motion parameters shown in (d,e) were not smooth because of the not dense enough sampling of the training data acquisition.

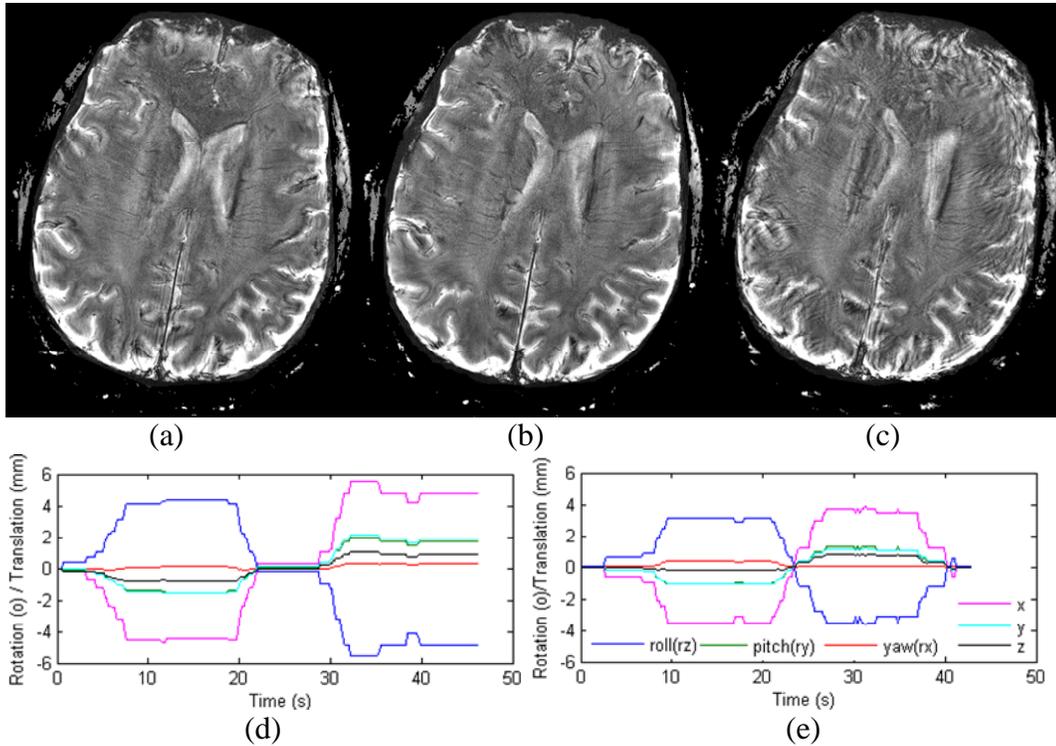


Figure 6-5 Prospective correction of large motion. (a) was acquired without intentional motion during the scan. (b c) were scanned in the presence of motion; (b) was acquired using the proposed method but (c) was without correction. (d,e) show the evolution of motion parameters during the acquisition of image (b,c), respectively.

6.4 Discussion

We propose a new motion correction method for MRI using single camera without any calibration. A short training scan is required to allow object motion estimation from the camera images. It does not need any tracking target, therefore is more user friendly. Initial testing results show that this method is very promising.

The major factor that affects the motion estimation accuracy is the registration error from EPI images. Even though registration can get sub-voxel accuracy, if motion occurs between the acquisition of the first and last slice of a volume, rigid body registration is not able to correct. This could be a problem in our training data

acquisition, because the TR of EPI acquisition was set relatively long (1s) to get a good coverage of the brain, and the volunteer was asked to continuously move his head during that period of time. In addition, ghost artifacts in a single slice could make the registration even harder. A solution to this is that during the training period of time, the volunteer should move very slowly to introduce negligible artifacts coming from these two effects.

Another issue that could affect the accuracy is the EPI image quality. The tradeoff for EPI's fast imaging speed is that it suffers image distortions, especially in a high magnetic field like in our case, which may in turn affect the registration performance. Field inhomogeneity can cause image distortions and even signal loss. Even though a high order shim can ameliorate this problem to some extent, the inhomogeneities remain, particularly close to tissue-air and tissue-bone interfaces. And even the field can be shimmed to almost homogeneous in the brain for acquisition of the first image in the training, the field changes when volunteer moves for about 20 degrees and 20mm, which could result in different distortion in the images and cannot be recovered by solely registration. Furthermore, gradient non-linearity, Nyquist ghost, etc, could also distort EPI acquisitions (Cohen and Weisskoff, 1991; Farzaneh et al., 1990; Jezzard and Clare, 1999).

The correlation coefficient was proved to be optimal in searching for the most similar image in the training data for a new image. It was not only accurate, but also insensitive to noise because global information in an image has been used. However, the accuracy for motion estimation does not just rely on the best image search. It depends on the completeness of the sampled motion space during training stage. If the

sample space is very limited, then the best match image is not able to provide correct motion parameters.

In conclusion, to get the best performance with this method, several conditions have to be satisfied: EPI image quality should be good and the volunteer should move slowly and to the full extent (not large because of the tight receive coil used) during the training data acquisition.

Chapter 7 Conclusion

Head motion during scanning can greatly affect MRI image quality and is a major reason for rendering data useless in both clinical and research application. High-resolution brain structural imaging is especially compromised by head movement because high-resolution scans have longer scan time. Two new motion correction methods were proposed here using either stereo or a single camera to measure and correct rigid body motion in real time. A gradient echo sequence was modified to accept motion estimates before each TR to adjust the gradient orientations and RF frequencies and phases for each excitation. Images were reconstructed on the scanner computer for immediate access. Experiments done with both of the two methods demonstrated a consistent improvement in image quality if motion occurred during the acquisition.

The advantage of using an optical system was its totally parallel computation to the pulse sequence, therefore no extra scanning time was needed for motion detection. The camera set up right on top of the head coil provided very high sensitivity for motion estimation.

For the motion correction method using stereo cameras, the position estimation accuracy was better than 0.1mm and 0.15°, which has been shown to be good enough for high-resolution imaging. However, it did have several drawbacks that a good calibration took some extra preparation time before experiments, and the “tracking target” used for motion detection was not totally user friendly. The first problem was not a big issue since stereo calibration only needed to be done offline

once as long as the cameras were fixed solidly on the holder (Section 9.4), and the calibration between the cameras and the MRI scanner was simplified to just several minutes. For the second problem, the tracking target was used only to remove noise caused by the scanner. A better shielding of the cameras should be able to solve the problem.

For the single camera method, no calibration was needed at all, but only a short training period. The big advantage of this method was that no tracking target was needed, because face features were used directly for motion detection. It is more user friendly. However, the accuracy depended a lot on EPI image quality and sampled motion space during the training stage. For the first problem, parallel imaging (Pruessmann et al., 1999) should be able to improve both the speed and image quality. For the second problem, the volunteer should be instructed to move as much as he can inside the coil during the training data acquisition.

Chapter 8 Future Work

In this project, two systems were built for prospective motion correction for high-resolution MRI in 7T. The feasibility of using optical tracking system to help MRI image acquisition has been demonstrated. But future work can be carried out to improve the performance further.

Both methods calculated motion after acquiring video data. Therefore, there is always a time delay for correction in the sequence. A better solution is to predict motion using the Kalman filter. Another positive feature about Kalman filter is its ability to remove tracking noise, which may further lead to an unnecessary tracking target for the stereo correction method.

Currently, the motion calculation was done continuously and whenever it was ready, it was sent to the MRI scan computer. The MRI scan computer then read the motion parameters inside the registers and executed the correction. Therefore when MRI started to read the data in the registers, they may not be fully updated yet. As a result, the adjustment of gradients and RF frequencies and phases may use uncorrected parameters. A future work is to synchronize the motion computation and MRI motion correction to avoid this problem.

A better way to acquire training data for the single camera method is to synchronize camera image acquisition with navigators. The purpose of training is to link camera images with motion parameters in MRI system. The EPI takes too long to achieve this. A 3D navigator, however, will take only several milliseconds to tens of milliseconds. As a result, the sampling of complete motion space will take less than

one minute. With this method, we will be able to combine the advantage of both optical system (parallel computation with the sequence) and navigator (fast acquisition), and is believed to be more practical.

Chapter 9 Appendices

9.1 Derivation of motion in logical coordinate system

The rotation and translation (\mathbf{R}_{ml} and \mathbf{T}_{ml}) between MRI physical and logical coordinate systems can be read out from the pulse sequence after the scan-plane is prescribed as in Equation 4.66 (the time variable t_0 is dropped for brevity):

$$\mathbf{X}_m = \mathbf{R}_{ml}(\mathbf{X}_l + \mathbf{T}_{ml}) \quad [9.1]$$

New motion parameters are denoted as \mathbf{R}_m and \mathbf{T}_m in physical coordinate system. Use subscripts “o” and “n” to define the position before and after motion (short for “old” and “new”), we have

$$\mathbf{X}_m^n = \mathbf{R}_m(\mathbf{X}_m^o + \mathbf{T}_m) \quad [9.2]$$

and

$$\mathbf{X}_m^o = \mathbf{R}_{ml}(\mathbf{X}_l^o + \mathbf{T}_{ml}) \quad [9.3]$$

The new motion parameters in logical coordinate system \mathbf{R}_n and \mathbf{T}_n must also satisfy the physical and logical relation, as

$$\mathbf{X}_m^n = \mathbf{R}_n(\mathbf{X}_l^n + \mathbf{T}_n) \quad [9.4]$$

So,

$$\begin{aligned} \mathbf{X}_l^n &= \mathbf{R}_n^{-1}\mathbf{X}_m^n - \mathbf{T}_n \\ &= \mathbf{R}_n^{-1}[\mathbf{R}_m(\mathbf{X}_m^o + \mathbf{T}_m)] - \mathbf{T}_n \\ &= \mathbf{R}_n^{-1}\mathbf{R}_m\mathbf{X}_m^o + \mathbf{R}_n^{-1}\mathbf{T}_m - \mathbf{T}_n \\ &= \mathbf{R}_n^{-1}\mathbf{R}_m[\mathbf{R}_{ml}(\mathbf{X}_l^o + \mathbf{T}_{ml})] + \mathbf{R}_n^{-1}\mathbf{T}_m - \mathbf{T}_n \\ &= \mathbf{R}_n^{-1}\mathbf{R}_m\mathbf{R}_{ml}\mathbf{X}_l^o + \mathbf{R}_n^{-1}\mathbf{R}_m\mathbf{R}_{ml}\mathbf{T}_{ml} + \mathbf{R}_n^{-1}\mathbf{T}_m - \mathbf{T}_n \end{aligned} \quad [9.5]$$

Since the scan plane needs to follow the motion, it means \mathbf{X}_1^n should equal \mathbf{X}_1^o all the time, therefore the rotation term in Equation 9.5 should be the identity matrix and the translation should be zero:

$$\left. \begin{aligned} \mathbf{R}_n^{-1} \mathbf{R}_m \mathbf{R}_{ml} &= \mathbf{I} \\ \mathbf{R}_n^{-1} \mathbf{R}_m \mathbf{R}_{ml} \mathbf{T}_{ml} + \mathbf{R}_n^{-1} \mathbf{T}_m - \mathbf{T}_n &= \mathbf{0} \end{aligned} \right\} \Rightarrow \begin{cases} \mathbf{R}_n = \mathbf{R}_m \mathbf{R}_{ml} \\ \mathbf{T}_n = \mathbf{T}_{ml} + \mathbf{R}_n^{-1} \mathbf{T}_m \end{cases} \quad [9.6]$$

The motion parameters that need to be sent to the sequence for correction are the difference between the new ones and the old ones read-out from the sequence (i.e. \mathbf{R}_{ml} and \mathbf{T}_{ml}), which are:

$$\mathbf{R}_{corr} = \mathbf{R}_m \mathbf{R}_{ml} \quad [9.7]$$

$$\mathbf{T}_{corr} = \mathbf{R}_{corr}^{-1} \mathbf{T}_m \quad [9.8]$$

9.2 MRI Coordinates from DICOM Image

To get the MRI physical coordinate (millimeter) of any point from its DICOM image (index) is not very straightforward, because they use two different coordinate systems and different units.

DICOM uses patient coordinate system, meaning for axial scan

X_d increases right to left

Y_d increases anterior to posterior

Z_d increases inferior to superior

MRI physical coordinate system on the other hand is defined as:

X_m increases left to right

Y_m increases posterior to anterior

Z_m increases inferior to superior

DICOM headers have directions within the image plane, i.e. the X_d and Y_d direction,

$$\mathbf{r}_1 = \text{Header.}\mathbf{ImageOrientationPatient}[1:3] \quad [9.9]$$

$$\mathbf{r}_2 = \text{Header.}\mathbf{ImageOrientationPatient}[4:6] \quad [9.10]$$

where $\mathbf{R}_d = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]$ is the rotation matrix describing the DICOM image plane in the patient coordinate system. The direction orthogonal to the plane, i.e. the Z_d direction, is defined as

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \quad [9.11]$$

Then inter-slice distances need to be determined for unit transformation. If only one slice is scanned,

$$z = \text{Header.Slicethickness} \quad [9.12]$$

Otherwise, it can be calculated from slice positions, as

$$z = (\text{Header}(i+1).\mathbf{ImagePositionPatient} - \text{Header}(i).\mathbf{ImagePositionPatient})\mathbf{r}_3 \quad [9.13]$$

where i is the index of one slice.

Voxel size is therefore,

$$\mathbf{v} = [\text{Header.}\mathbf{PixelSpacing}(1) \ \text{Header.}\mathbf{PixelSpacing}(2) \ z] \quad [9.14]$$

where $\mathbf{PixelSpacing}$ saves in-plane pixel size in DICOM.

Position of the first voxel in DICOM's patient coordinate system is

$$\mathbf{v}_0 = \text{Header.}\mathbf{ImagePositionPatient}^T \quad [9.15]$$

To map voxels to the patient coordinates is

$$\mathbf{H}_{dicom_to_patient} = \begin{bmatrix} \mathbf{R}_d * \mathbf{diag}(\mathbf{v}) & \mathbf{v}_0 \\ \mathbf{0} & 1 \end{bmatrix} \quad [9.16]$$

Where $\mathbf{diag}(\mathbf{v})$ is a 3×3 matrix with the diagonal elements coming from \mathbf{v} .

Then map from the patient coordinate system adopted by DICOM to the MRI physical coordinate system is a rotation about the Z axis:

$$\mathbf{H}_{patient_to_MRI} = \begin{bmatrix} -1 & & & \\ & -1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \quad [9.17]$$

Combine both transformations:

$$\mathbf{H} = \mathbf{H}_{patient_to_MRI} * \mathbf{H}_{DICOM_to_patient} \quad [9.18]$$

Any voxel in DICOM image has MRI coordinate as

$$\mathbf{X}_m = \mathbf{H}\mathbf{X}_d \quad [9.19]$$

Note that if MATLAB is used, the voxel index is from 1 instead of 0, a translation

$$\text{matrix } \mathbf{mat} = \begin{bmatrix} 1 & & -1 \\ & 1 & -1 \\ & & 1 & -1 \\ & & & 1 \end{bmatrix} \text{ is needed to be multiplied with Equation 9.18 for a}$$

correct mapping.

$$\mathbf{H} = \mathbf{H}_{patient_to_MRI} * \mathbf{H}_{DICOM_to_patient} * \mathbf{mat} \quad [9.20]$$

9.3 Gauss-Newton Algorithm

Gauss-Newton iteration is a technique to locate the minimum of a multivariate function that is expressed as the sum of squares of non-linear functions (Kelley, 1999; Nocedal and Wright, 2006). In this section, we do not have the ambition to state new concepts on Newton iteration, but just to give a derivation for a better explanation in Section 4.2.1, 4.2.2 and 6.2.1.

Let f be the function that maps a measurement vector \mathbf{x} through parameter vector \mathbf{p} to another measurement \mathbf{y} , the solution \mathbf{p} is obtained when distance $\boldsymbol{\varepsilon}^2$ is minimized:

$$\boldsymbol{\varepsilon}^2 = \|\mathbf{y} - f(\mathbf{x}; \mathbf{p})\|^2 \quad [9.21]$$

By assuming $f(\mathbf{x}; \mathbf{p})$ as a linear function in the neighborhood of \mathbf{p}_0 , a Taylor series expansion will lead to the approximation

$$\hat{f}(\mathbf{x}; \mathbf{p}) = f(\mathbf{x}; \mathbf{p}_0) + \frac{\partial f(\mathbf{x}; \mathbf{p}_0)}{\partial \mathbf{p}} (\mathbf{p} - \mathbf{p}_0) \quad [9.22]$$

The derivative of the error function at \mathbf{p} is

$$\frac{\partial \hat{\boldsymbol{\varepsilon}}^2}{\partial \mathbf{p}} = -2 \frac{\partial \hat{f}(\mathbf{x}; \mathbf{p})}{\partial \mathbf{p}}^T (\mathbf{y} - \hat{f}(\mathbf{x}; \mathbf{p})) \quad [9.23]$$

And because $\hat{f}(\mathbf{x}; \mathbf{p})$ is linear,

$$\frac{\partial \hat{f}(\mathbf{x}; \mathbf{p})}{\partial \mathbf{p}} = \frac{\partial f(\mathbf{x}; \mathbf{p}_0)}{\partial \mathbf{p}} \quad [9.24]$$

Substitute Equation 9.22 and 9.24 to 9.23, we get

$$\frac{\partial \hat{\boldsymbol{\varepsilon}}^2}{\partial \mathbf{p}} = 2 \frac{\partial f(\mathbf{x}; \mathbf{p}_0)}{\partial \mathbf{p}}^T (f(\mathbf{x}; \mathbf{p}_0) + \frac{\partial f(\mathbf{x}; \mathbf{p}_0)}{\partial \mathbf{p}} (\mathbf{p} - \mathbf{p}_0) - \mathbf{y}) \quad [9.25]$$

Let the Jacobian matrix be:

$$\mathbf{J} = \frac{\partial f(\mathbf{x}; \mathbf{p}_0)}{\partial \mathbf{p}} \quad [9.26]$$

Then solve for \mathbf{p} when $\frac{\partial \hat{\boldsymbol{\varepsilon}}^2}{\partial \mathbf{p}} = 0$:

$$\mathbf{p} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \boldsymbol{\varepsilon}_0 + \mathbf{p}_0 \quad [9.27]$$

Therefore, the solution \mathbf{p} is obtained with an initial estimate \mathbf{p}_0 and refined iteratively according to:

$$\mathbf{p}_{n+1} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \boldsymbol{\varepsilon}_n + \mathbf{p}_n \quad [9.28]$$

The above algorithm is suitable for minimization with respect to a small number of parameters, but not suitable for minimizing cost functions with respect to large numbers of parameters because of the heavy burden of matrix inverse in Equation 9.28. However, in this project, when many parameters need to be estimated, e.g. $6M+8$ unknowns in Equation 4.22, the Jacobian matrix is very sparse and can be taken advantage for great time savings.

Use Equation 4.22 as an example, the parameters to be estimated are intrinsic parameters $\mathbf{a} = [f_x, f_y, x_0, y_0, k_1, k_2, p_1, p_2]^T$ and extrinsic parameters $\mathbf{R}_i, \mathbf{T}_i$ (using $\mathbf{b}_i = [r_{1i}, r_{2i}, \dots, r_{6i}]$ to denote the 6DOF parameters describing them). Then \mathbf{p} can be written as

$$\mathbf{p} = (\mathbf{a}^T, \mathbf{b}_1^T, \mathbf{b}_2^T, \dots, \mathbf{b}_M^T)^T \quad [9.29]$$

The Jacobian matrix is therefore

$$\mathbf{J} = \frac{\partial \mathbf{m}_p}{\partial \mathbf{p}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 & & & \\ \mathbf{A}_2 & & \mathbf{B}_2 & & \\ \mathbf{M} & & & \mathbf{O} & \\ \mathbf{A}_M & & & & \mathbf{B}_M \end{bmatrix} \quad [9.30]$$

Where $\mathbf{A}_i = \partial \mathbf{m}_{p_i} / \partial \mathbf{a}$, $\mathbf{B}_i = \partial \mathbf{m}_{p_i} / \partial \mathbf{b}_i$. The sparse condition is because of the fact that each projected points \mathbf{m}_{p_i} depend only on \mathbf{a} and \mathbf{b}_i but not any other $\mathbf{b}_j (j \neq i)$, so that $\partial \mathbf{m}_{p_i} / \partial \mathbf{b}_j = 0$. Then Equation 9.28 can be solved for each iteration step a lot faster based on this sparse matrix.

9.4 Stereo System QA

The stereo camera calibration takes quite some time but it does not have to be done every time before using it, as long as nothing changes. Therefore, it is important to check the stereo to make sure nothing changed, i.e. a system QA. It can be realized by comparing the 3D position estimations for a checkerboard pattern using either camera or the stereo. Because the checkerboard size is known, we are able to estimate corner positions in 3D from a single camera based on homography (Equation 4.18 – 4.21). The difference between single camera and stereo estimation can be very large even if a small change in the cameras' parameters changed (Figure 9-1)

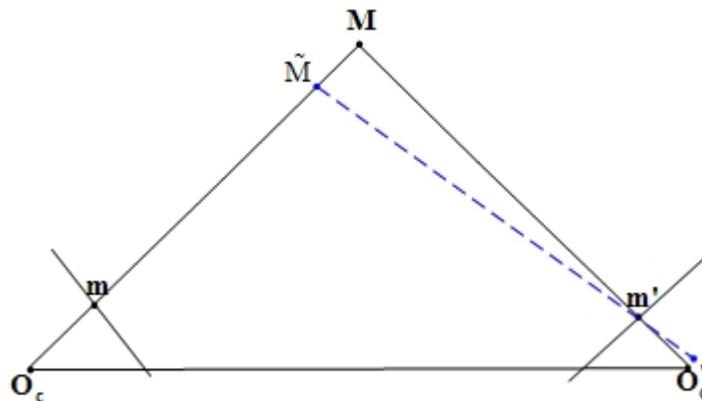


Figure 9-1 3D reconstruction error if camera parameters change. The solid lines projection is correct and the dotted projection is the result from different camera parameters with a small focal length and principle point change. It's obvious that the 3D estimation magnify the camera parameters distinction.

Figure 9-2 shows the 3D corner positions estimated using left camera (red circles), right camera (blue circles) and stereo (green circles). When correct calibration parameters were used, the three estimations match perfectly as shown in (a). But if a very small change is made on the focal length, e.g. 0.1mm, they differ quite a lot (b). Also, the difference between three estimations can be calculated. For

image (a), the maximum error between left-right camera estimation is 0.0098092; the maximum error between left-stereo camera estimation is 0.064398; and the maximum error between right-stereo camera estimation is 0.087699. But for image (b) the maximum error between left-right camera estimation is 0.48134; the maximum error between left-stereo camera estimation is 3.287; and the maximum error between right-stereo camera estimation is 3.107. Therefore, the stereo QA can be accomplished either visually or quantitatively.

In real application, the calibration parameters we use do not change, while cameras' status may change accidentally. But the error magnification should be the same as in this simulation. Therefore, each time before the stereo system is used, run the system QA to make sure if a re-calibration is necessary.

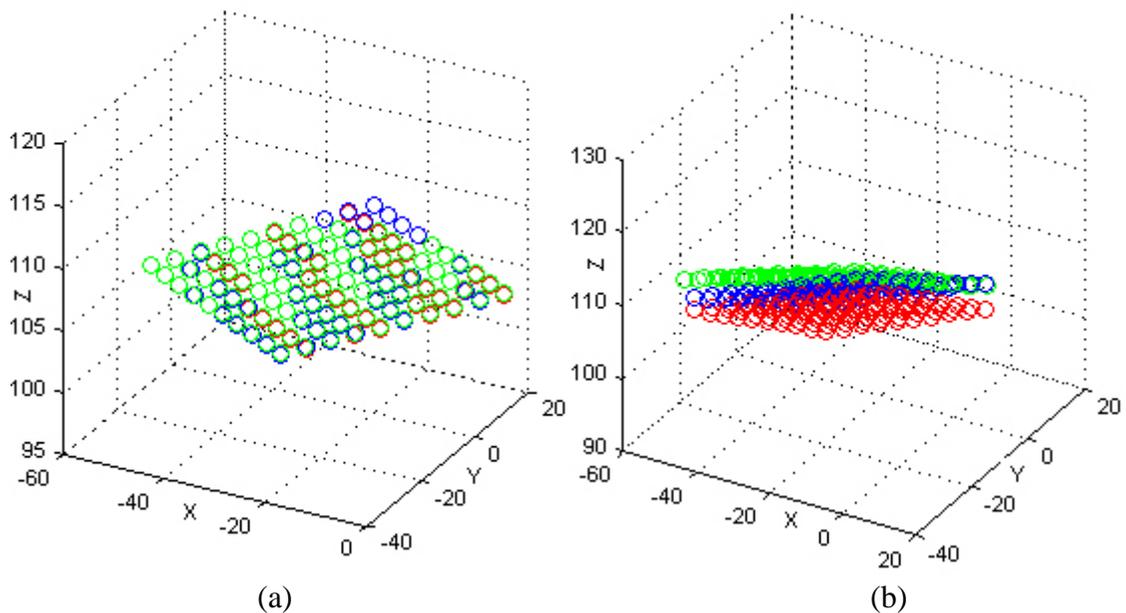


Figure 9-2 3D position estimation for a checkerboard pattern. Every circle represents a corner of the checkerboard. Red circles represent positions estimated from the left camera; blue circles represent positions estimated from the right camera; and green circles represent those estimated from the stereo. (a) correct calibration parameters were used. (b) focal length was changed with 0.1mm.

Chapter 10 Bibliography

- Abdel-Aziz, Y. I. and Karara, H. M. (1971) Direct linear transformation into object space coordinates in close-range photogrammetry. *Proc. Symposium on Close-Range Photogrammetry*, 1-18.
- Arun, K. S., Huang, T. S. and Blostein, S. D. (1987) Least-squares fitting of two 3D point sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 9, 698-700.
- Barnwell, J. D., Smith, J. K. and Castillo, M. (2007) Utility of navigator-prospective acquisition correction technique (PACE) for reducing motion in brain MR imaging studies. *AJNR Am J Neuroradiol*, 28, 790-1.
- Bernstein, M. A., King, K. F. and Zhou, Z. J. (2004) *Handbook of MRI pulse sequences*. Amsterdam ; Boston: Academic Press.
- Blake, A., Curwen, R. and Zisserman, A. (1993) A framework for spatiotemporal control in the tracking of visual contours. *International Journal of Computer Vision*, 11, 127-145.
- Bouguet, J. Y. (1999) Visual methods for three-dimensional modeling. Ph.D. Thesis: California Institute of Technology.
- Broggi, A. (1998) Vision-based diving assistance. *Intelligent Systems and their Applications, IEEE*, 13, 22-23.
- Charbonnier, L. and Fournier, A. (1995) Heading guidance and obstacles localization for an indoor mobile robot. *Advanced Robotics, 1995. ICAR '95. Proceedings., 8th International Conference on*. pp. 507-513.
- Chou, C.-H. and Chen, Y.-C. (1990) Moment-preserving pattern matching. *Pattern Recognition*, 23, 461-474.
- Cohen, M. S. and Weisskoff, R. M. (1991) Ultra-fast imaging. *Magnetic Resonance Imaging*, 9, 1-37.

- Coifman, B., Beymer, D., McLauchlan, P. and Malik, J. (1998) A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C: Emerging Technologies*, 6, 271-288.
- Daniilidis, K. and Ernst, J. (1996) Active intrinsic calibration using vanishing points. *Pattern Recognition Letters*, 17, 1179-1189.
- Derbyshire, J. A., Wright, G. A., Henkelman, R. M. and Hinks, R. S. (1998) Dynamic scan-plane tracking using MR position monitoring. *J Magn Reson Imaging*, 8, 924-32.
- Dold, C., Zaitsev, M., Speck, O., Firlé, E. A., Hennig, J. and Sakas, G. (2006) Advantages and limitations of prospective head motion compensation for MRI using an optical motion tracking device. *Acad Radiol*, 13, 1093-103.
- Duda, R. O., Hart, P. E. and Stork, D. G. (2001) *Pattern classification*. New York: Wiley.
- Duyn, J. H., van Gelderen, P., Li, T. Q., de Zwart, J. A., Koretsky, A. P. and Fukunaga, M. (2007) High-field MRI of brain cortical substructure based on signal phase. *Proc Natl Acad Sci U S A*, 104, 11796-801.
- Eggert, D. W., Lorusso, A. and Fisher, R. B. (1997) Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9, 272-290.
- Ehman, R. L. and Felmlee, J. P. (1989) Adaptive technique for high-definition MR imaging of moving structures. *Radiology*, 173, 255-63.
- Eviatar, H., Saunders, J. and Hoult, D. (1997) Motion Compensation by gradient adjustment. *Proc Intl Magn Reson Med*, 1898.
- Eviatar, H., Schattka, B., Sharp, J., Rendell, J. and Alexander, M. (1999) Real time head motion correction for functional MRI. *Proc Intl Magn Reson Med*, 269.
- Faig, W. (1975) Calibration of close-range photogrammetric systems: mathematical formulation. *Photogrammetric Engineering and Remote Sensing*, 41, 1479-1486.

- Farzaneh, F., Riederer, S. J. and Pelc, N. J. (1990) Analysis of T2 limitations and off-resonance effects on spatial resolution and artifacts in echo-planar imaging. *Magnetic Resonance in Medicine*, 14, 123-139.
- Faugeras, O. (1993) *Three-dimensional computer vision : a geometric viewpoint*. Cambridge, Mass.: MIT Press.
- Faugeras, O. and Robert, L. (1996) What can two images tell us about a third one? *International Journal of Computer Vision*, 18, 5-19.
- Faugeras, O. and Toscani, G. (1986) The calibration problem for stereo. *Computer Vision and Pattern Recognition, 1986. Proceedings., 1986 IEEE Computer Society Conference on*. pp. 15-20.
- Friston, K., Ashburner, J., Frith, C., Poline, J., Heather, J. and Frackowiak, R. (1995) Spatial registration and normalization of images. *Human Brain Mapping*, 2, 165-189.
- Friston, K. J. (2007) *Statistical parametric mapping : the analysis of functional brain images*. Amsterdam ; Boston: Elsevier/Academic Press.
- Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. and Turner, R. (1996) Movement-related effects in fMRI time-series. *Magn Reson Med*, 35, 346-55.
- Fu, Z. W., Wang, Y., Grimm, R. C., Rossman, P. J., Felmlee, J. P., Riederer, S. J. and Ehman, R. L. (1995) Orbital navigator echoes for motion measurements in magnetic resonance imaging. *Magn Reson Med*, 34, 746-53.
- Gerlach, W. and Stern, O. (1924) Ueber die Richtungsquantelung im Magnetfeld. *Ann Phys*, 74, 673.
- Glover, G. H. and Lee, A. T. (1995) Motion artifacts in fMRI: comparison of 2DFT with PR and spiral scan methods. *Magn Reson Med*, 33, 624-35.
- Glover, G. H. and Noll, D. C. (1993) Consistent projection reconstruction (CPR) techniques for MRI. *Magn Reson Med*, 29, 345-51.

- Glover, G. H. and Pauly, J. M. (1992) Projection reconstruction techniques for reduction of motion effects in MRI. *Magn Reson Med*, 28, 275-89.
- Gorodnichy, D. O. and Roth, G. (2004) Nouse []use your nose as a mouse' perceptual vision technology for hands-free games and interfaces. *Image and Vision Computing*, 22, 931-942.
- Hager, G. D. and Belhumeur, P. N. (1998) Efficient region tracking with parametric models of geometry and illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20, 1025-1039.
- Hall, E. L., Tio, J. B. K., McPherson, C. A. and Sadjadi, F. A. (1982) Measuring Curved Surfaces for Robot Vision. *Computer*, 15, 42-54.
- Haralick, R. M., Joo, H., Lee, C., Zhuang, X., Vaidya, V. G. and Kim, M. B. (1989) Pose estimation from corresponding point data. *Systems, Man and Cybernetics, IEEE Transactions on*, 19, 1426-1446.
- Harris, C. and Stephens, M. (1988) A combined corner and edge detector. *Alvey Vision Conference*. pp. 147-151.
- Hartley, R. and Zisserman, A. (2003) *Multiple view geometry in computer vision*. Cambridge, UK ; New York: Cambridge University Press.
- Hartley, R. I. and Sturm, P. (1997) Triangulation. *Computer vision and image understanding*, 68, 146-157.
- Horn, B. K. P. (1987) Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4, 629-642.
- Horn, B. K. P., Hilden, H. M. and Negahdaripour, S. (1988) Closed-form solution of absolute orientation using orthonormal matrices. *J. Opt. Soc. Am. A*, 5, 1127-1135.
- Hu, Y. and Glover, G. H. (2007) Three-dimensional spiral technique for high-resolution functional MRI. *Magn Reson Med*, 58, 947-51.

- Jenkinson, M., Bannister, P., Brady, M. and Smith, S. (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17, 825-41.
- Jezzard, P. and Clare, S. (1999) Sources of distortion in functional MRI data. *Human Brain Mapping*, 8, 80-85.
- Kadah, Y. M., Abaza, A. A., Fahmy, A. S., Youssef, A. B., Heberlein, K. and Hu, X. P. (2004) Floating navigator echo (FNAV) for in-plane 2D translational motion estimation. *Magn Reson Med*, 51, 403-7.
- Kakadiaris, L. and Metaxas, D. (2000) Model-based estimation of 3D human motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22, 1453-1459.
- Kanatani, K. (1994) Analysis of 3-D rotation fitting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16, 543-549.
- Kelley, C. T. (1999) *Iterative methods for optimization*. Philadelphia: SIAM.
- Kim, S., Dougherty, L., Rosen, M. A., Song, H. K. and Poptani, H. (2008) Automatic correction of in-plane bulk motion artifacts in self-navigated radial MRI. *Magn Reson Imaging*, 26, 367-78.
- Lauterbur, P. C. (1973) Image formation by induced local interactions. Examples employing nuclear magnetic resonance. *Nature*, 190-191.
- Lee, C. C., Grimm, R. C., Manduca, A., Felmlee, J. P., Ehman, R. L., Riederer, S. J. and Jack, C. R., Jr. (1998) A prospective approach to correct for inter-image head rotation in fMRI. *Magn Reson Med*, 39, 234-43.
- Lee, C. C., Jack, C. R., Jr., Grimm, R. C., Rossman, P. J., Felmlee, J. P., Ehman, R. L. and Riederer, S. J. (1996) Real-time adaptive motion correction in functional MRI. *Magn Reson Med*, 36, 436-44.
- Liang, Z.-P., Lauterbur, P. C. and IEEE Engineering in Medicine and Biology Society. (2000) *Principles of magnetic resonance imaging : a signal processing perspective*. Bellingham, Wash. New York: SPIE Optical Engineering Press ;

IEEE Press.

Liu, C., Bammer, R., Kim, D. H. and Moseley, M. E. (2004) Self-navigated interleaved spiral (SNAILS): application to high-resolution diffusion tensor imaging. *Magn Reson Med*, 52, 1388-96.

Lucas, B. D. and Kanade, T. (1981) An interative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence, 1981. Proceedings.,* 674-679.

Mansfield, P. (1977) Nulti-planar image formation using NMR spin-echoes. *J Phys C Solid State Phys*, 10, L55-L58.

Mansfield, P. and Grannell, P. (1973) NMR 'diffraction' in solids? *J Phys C Solid State Phys*, 6, L422.

Marr, D., Poggio, T. and Ullman, S. (1979) Bandpass channels, zero-crossings, and early visual information processing. *J. Opt. Soc. Am.*, 69, 914-916.

Meyer, C. H., Hu, B. S., Nishimura, D. G. and Macovski, A. (1992) Fast spiral coronary artery imaging. *Magn Reson Med*, 28, 202-13.

Milella, A. and Siegart, R. (2006) Stereo-Based Ego-Motion Estimation Using Pixel Tracking and Iterative Closest Point. *Computer Vision Systems, 2006 ICVS '06. IEEE International Conference on.* pp. 21-21.

Nevo, E., Roth, A. and Hushek, M. (2002) An electromagnetic 3D locator system for use in MR scanners. *Proc Intl Magn Reson Med*, 334.

Newman, T. S. and Jain, A. K. (1995) A Survey of Automated Visual Inspection. *Computer vision and image understanding*, 61, 231-262.

Nocedal, J. and Wright, S. J. (2006) *Numerical optimization*. New York: Springer.

Petrie, D. W., Costa, A. F., Takahashi, A., Yen, Y. F. and Drangova, M. (2005) Optimizing spherical navigator echoes for three-dimensional rigid-body motion detection. *Magn Reson Med*, 53, 1080-7.

- Pipe, J. G. (1999) Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging. *Magn Reson Med*, 42, 963-9.
- Pipe, J. G., Farthing, V. G. and Forbes, K. P. (2002) Multishot diffusion-weighted FSE using PROPELLER MRI. *Magn Reson Med*, 47, 42-52.
- Pipe, J. G. and Zwart, N. (2006) TurboProp: improved PROPELLER imaging. *Magn Reson Med*, 55, 380-5.
- Pruessmann, K. P., Weiger, M., Scheidegger, M. B. and Boesiger, P. (1999) SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42, 952-962.
- Rabi, I. I., Zacharias, J. R., Millman, S. and Kusch, P. (1938) A new method of measuring nuclear magnetic moments. *Phys Rev*, 53, 318.
- Santos, J. M., Wright, G. A. and Pauly, J. M. (2004) Flexible real-time magnetic resonance imaging framework. *Conf Proc IEEE Eng Med Biol Soc*, 2, 1048-51.
- Shi, J. and Tomasi, C. (1994) Good features to track. *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*. pp. 593-600.
- Slama, C. C., Theurer, C. and Henriksen, S. W. (1980) *Manual of photogrammetry*. Falls Church, Va.: American Society of Photogrammetry.
- Spuentrup, E., Stuber, M., Botnar, R. M., Kissinger, K. V. and Manning, W. J. (2002) Real-time motion correction in navigator-gated free-breathing double-oblique submillimeter 3D right coronary artery magnetic resonance angiography. *Invest Radiol*, 37, 632-6.
- Thesen, S., Heid, O., Mueller, E. and Schad, L. R. (2000) Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magn Reson Med*, 44, 457-65.
- Thiel, T., Czisch, M., Elbel, G. K. and Hennig, J. (2002) Phase coherent averaging in magnetic resonance spectroscopy using interleaved navigator scans:

compensation of motion artifacts and magnetic field instabilities. *Magn Reson Med*, 47, 1077-82.

Tomasi, C. and Kanade, T. (1992) Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9, 137-154.

Tomasi, C. and Kanatani, K. (1991) Detection and tracking of point features. Technical Report: CMU-CS-91-132.

Tommasini, T., Fusiello, A., Trucco, E. and Roberto, V. (1998) Making good features track better. *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*. pp. 178-183.

Torr, P. H. S. and Zisserman, A. (1997) Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15, 591-605.

Tremblay, M., Tam, F. and Graham, S. J. (2005) Retrospective coregistration of functional magnetic resonance imaging data using external monitoring. *Magn Reson Med*, 53, 141-9.

Tsai, R. (1987) A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *Robotics and Automation, IEEE Journal of*, 3, 323-344.

Twieg, D. B. (1983) The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Med Phys*, 10, 610-21.

Umeyama, S. (1991) Least-squares estimation of transformation parameters between two point patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13, 376-380.

van der Kouwe, A. J., Benner, T. and Dale, A. M. (2006) Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn Reson Med*, 56, 1019-32.

- van Gelderen, P., de Zwart, J. A., Starewicz, P., Hinks, R. S. and Duyn, J. H. (2007) Real-time shimming to compensate for respiration-induced B0 fluctuations. *Magn Reson Med*, 57, 362-8.
- Walker, M. W., Shao, L. and Volz, R. A. (1991) Estimating 3-D location parameters using dual number quaternions. *CVGIP: Image Understanding*, 54, 358-367.
- Wang, F. N., Huang, T. Y., Lin, F. H., Chuang, T. C., Chen, N. K., Chung, H. W., Chen, C. Y. and Kwong, K. K. (2005) PROPELLER EPI: an MRI technique suitable for diffusion tensor imaging at high field strength with reduced geometric distortions. *Magn Reson Med*, 54, 1232-40.
- Wang, L. L. and Tsai, W.-H. (1991) Camera calibration by vanishing lines for 3-D computer vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13, 370-376.
- Ward, H. A., Riederer, S. J., Grimm, R. C., Ehman, R. L., Felmlee, J. P. and Jack, C. R., Jr. (2000) Prospective multiaxial motion correction for fMRI. *Magn Reson Med*, 43, 459-69.
- Welch, E. B., Manduca, A., Grimm, R. C., Ward, H. A. and Jack, C. R., Jr. (2002) Spherical navigator echoes for full 3D rigid body motion measurement in MRI. *Magn Reson Med*, 47, 32-41.
- Weng, J., Ahuja, N. and Huang, T. S. (1992a) Matching two perspective views. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14, 806-825.
- Weng, J., Cohen, P. and Herniou, M. (1992b) Camera calibration with distortion models and accuracy evaluation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14, 965-980.
- Wyatt, C. L., Ari, N. and Kraft, R. A. (2005) Spherical navigator registration using harmonic analysis for prospective motion correction. *Inf Process Med Imaging*, 19, 738-49.
- Yao, Y.-S. and Chellappa, R. (1995) Tracking a dynamic set of feature points. *Image Processing, IEEE Transactions on*, 4, 1382-1395.

- Zaitsev, M., Dold, C., Sakas, G., Hennig, J. and Speck, O. (2006) Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system. *Neuroimage*, 31, 1038-50.
- Zhang, Z. (2000) A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22, 1330-1334.
- Zhang, Z., Deriche, R., Faugeras, O. and Luong, Q.-t. (1994) A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report: INRIA.
- Zheng, Q. and Chellappa, R. (1993) Automatic feature point extraction and tracking in image sequences for unknown camera motion. *Computer Vision, 1993. Proceedings., Fourth International Conference on*. pp. 335-339.
- Zhou, S. K., Chellappa, R. and Moghaddam, B. (2004) Visual tracking and recognition using appearance-adaptive models in particle filters. *Image Processing, IEEE Transactions on*, 13, 1491-1506.