Image Registration: Features and Applications

By

Jie Wang

A Thesis Submitted

For the Degree of Doctor of Philosophy

at

Department of Computer Science

School of Computing

National University of Singapore

August, 2011

Acknowledgement

I would like to express my deep and sincere gratitude to my advisor, Professor Chew Lim Tan in School of Computing, National University of Singapore, for his invaluable guidance and constant support throughout this research work. His wide knowledge and constructive advice have inspired me with various ideas to tackle the difficulties and attempt new directions. In particular, his understanding and help in every aspect have supported me through the chaos and confusion in those difficult days. This thesis would not have been possible without his generous contributions in one way or another.

I wish to express my warm and sincere thanks to Dr. Shi Jian Lu, who gave me important guidance during my first steps into this research area. I sincerely appreciate his ingenious ideas on document image restoration and detailed suggestions and efforts throughout the writing of our paper on document skew detection. I also want to thank Dr. Shi Miao Li, for her insightful advice and comprehensive comments on the work about CT scan normalization. Her expertise in computer

vision and image registration has enlightened me several evaluation strategies to better demonstrate the effectiveness of the proposed method.

I wish to express my deep appreciation to Associate Professor Michael S. Brown currently with School of Computing, National University of Singapore, for his generous sharing of their existing work on historical document restoration and document data with us, and for his constructive suggestions and efforts in improving our paper on historical document restoration.

I owe my sincere gratitude to Dr. Kok Lim Low and Associate Professor Ee-Chien Chang in School of Computing, National University of Singapore, for their detailed reviews, constructive comments and suggestions to my graduate research paper and thesis proposal during the whole research program.

I wish to extend my warmest thanks to all those colleagues and friends who have helped me and encouraged me in one way or another during my research study in the Center of Information Mining and Extraction (CHIME) of School of Computing, National University of Singapore.

Last but not least, I wish to express my special gratitude to my parents and my husband Shuai Hao, for their continuous support and understanding throughout my study for all these years.

Contents

Contents

Abstract

Nowadays images provide more and more information about this world. Often multiple images share the same scene observed from different angles, at different times or with different devices. Image registration is a method of aligning two or more images of the same scene into the same coordinate system so that the aligned images can be directly compared and combined. It is a fundamental step in many image analysis tasks in which the final knowledge has to be gained from the combination of multiple data sources. Identifying the correspondence between two images is simple for human visual system but challenging for computer algorithms. In general, four components are important for a typical image registration framework: image feature extraction, similarity metric, transformation model and optimization strategy. Due to the variety of image types and application domains, it is impossible to design a universal method for all image registration tasks.

In this thesis, we have developed several contributions to the field of image registration. These contributions stand on their own as valuable components within

their particular application domains, but are linked under the common theme of image registration. First, we have developed a method which is capable of estimating the skew distortion and orientation of printed document images. It registers a skewed document image with an imaginary image that would be captured if the document was posed in exactly upright position during the scanning procedure. Within this method, we have presented a novel image feature called interline white run to perform this registration task. Interline white run can be accurately derived from white run histograms which are obtained through one-time fast scanning of the document. Although the new feature seems simple, our experiments on real-world documents have demonstrated its efficiency in estimating the skew angle of printed document images.

We have also developed a framework to register the two sides of a double-sided historical document. As historical document images are usually degraded by various noises and distortions, we have designed an algorithm to extract salient control points from historical images for the purpose of registration. For documents with slight geometric distortions, a representative block is selected and used to estimate a rigid transformation model. When severe local deformation is present, mainly warping effects and local uneven surfaces, a fine registration procedure which combines salient points extraction, free-form transformation model and residual complexity similarity measure is additionally applied. Our experiments have shown that this registration framework significantly improves the performances of subsequent bleed-through correction methods.

Finally, we have proposed a groupwise image registration framework to build a brain CT atlas with the CT scans of multiple patients. The groupwise registration method is built upon a non-rigid pairwise image registration method which shares the same transformation model with the method we have proposed for historical

document images. CT slices which are from normal study cases and labeled with the same level number are first clustered into different groups. Among each group, all slices are registered to the center of the group and an intermediate average slice is computed for the group. The final average slice for a particular level is the combination of the average slices of all groups on this level. With the built atlas, we can efficiently estimate the level of an input CT slice in the axial direction of brain, which will significantly speed up subsequent content based retrieval systems. In addition, by comparing the input slice which are affected by traumatic brain injury against the atlas, we can identify the abnormal regions on the input slice.

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Image Registration

Image registration refers to the process of overlaying two or more images of the same scene (or similar scenes) that are taken at different times, from different perspectives and by different sensors for the purpose of comparison or fusion [ZF03]. With two images to be registered, one of them is usually called the reference image and kept untouched, and the other image is called the target image and transformed to the coordinate system where the reference image is. When multiple images need to be registered, they are often uniformly called the subject images. Image registration is a crucial step in many image analysis tasks and has been studied in various research areas, such as remotely sensed data processing, medical image analysis, computer vision and pattern recognition. Within different applications, image registration can also be called image alignment, matching, stabilization, fusion or stitching. In general, the applications of image registration could be

divided into four main groups, according to the manner of the image acquisition:

- **Different viewpoints:** Images of the same scene are captured from different viewpoints. Registering such kind of images is usually to gain a larger or a higher dimensional representation of the scene. Representative applications include image mosaicing in remote sensing and 3D shape recovery in computer vision.

- **Different times:** Images of the same scene are acquired at different times and probably under different conditions. One of the purposes of registering such images is to detect changes in the consecutively acquired images. Examples of applications include detecting scene changes for security purpose in compute vision and monitoring the healing therapy or the evolution of tumors in medical imaging.

- **Different sensors:** Images of the same scene are obtained with different types of sensors. These images are registered so that more complex or detailed scene representation can be achieved by integrating all the information from different sources. One example of such applications is registering computer tomography (CT) scans to magnetic resonance image (MRI) scans to get detailed information on anatomical structures.

- **Different scenes:** Images to be registered are captured from different scenes. One typical situation is to register multiple medical scans, e.g. MRIs from different patients. The aim is to construct an atlas which describes the anatomical variations of populations. The other situation of registering images from different scenes is to register the image of a scene and a model of the scene. The model can be a computer representation of the scene, such as a CT atlas, the imaginary image of a skewed document posed in precisely upright

position (as in Chapter 3). The aim of registering these images is to localize the target image in the scene or model for comparison.

In spite of large amount of work and relative success, the problem of image registration is far from being solved [CHH04, Hol08]. Many challenges remain in the definition of the correspondences between image elements, of similarity measures and of the transformation models between the two images to be registered. For instance, in medical imaging domain, the correspondences between anatomical components are difficult to be formulated mathematically. Meanwhile, real-world images are often corrupted by noise, illumination changes and spatially varying bias fields. Furthermore, if the two images to be registered are from different modalities, they can have completely different intensities. An image can also be corrupted by outliers, e.g. contrast agent, growing tumor or moving cells, which might have no matching counterparts in the other image. All these circumstances make the definition of a metric measuring the similarity between the two images to be registered complicated and challenging. What's most difficult is that the true underlying deformation between the two images to be registered is often unknown. Currently, researchers tend to assume a certain transformation model that produces physically realizable adequate approximation for a particular application, e.g. smooth or locally rigid deformation. Comprehensive surveys on image registration and its applications can be found in [Bro92, MV98, HBHH01, ZF03, CHH04, Sze06].

1.2 Contributions

Most of the contributions of this thesis have been successfully completed and reported during the course of the research. In summary, the following concrete and substantial contributions to the study of image registration techniques and their

applications have been made:

Interline White Runs for Skewed Document Registration: We have proposed a novel image feature, called Interline White Runs for the skew correction of degraded document images. With this feature, we register a skewed document to an imaginary image of the document posed in precisely upright position to achieve the purpose of skew correction. This feature accurately captures the spatial relationship between the two images to be registered, and it can be efficiently and accurately extracted from document images. In addition, this image feature is capable of detecting the orientation of document images. We have developed a skew correction system using interline white runs and compared its performance with other skew correction methods. Experiments on real-world documents have shown that our system is much faster and estimate more accurate skew angles.

Non-rigid Pairwise Registration for Historical Document Restoration: We have filled in the gap between document capturing and historical document restoration by providing fully automated techniques for the registration of the two sides of a document. First we have developed an algorithm to automatically extract and match control point pairs from the two images of a historical document. The algorithm takes into account the image characteristics of the document images and the forming mechanism of the bleed-through distortions on these images. Then with the detected control point pairs, we have designed a non-rigid image registration framework which combines the advantages of Residual Complexity and Free-form transformation model. We have integrated the whole registration algorithm with a wavelet based bleed-through correction method and evaluated the overall performance of document restoration on real-world historical documents.

Groupwise Registration for Brain CT Atlas Construction: We have

developed a cluster based groupwise image registration approach to construct a brain CT atlas with the medical scans of different patients. The groupwise registration method has been built upon a non-rigid pairwise registration method and a hierarchical cluster structure. Free-form transformation model and normalized mutual information are employed in the pairwise registration method. The built atlas has been used to estimate the position of an input slice on the axial direction of the brain. This procedure is referred to as slice indexing which significantly accelerates content based image retrieval systems or computer-assisted diagnosis systems. We have also demonstrated that by registering an input slice that is affected by traumatic brain injury to the atlas, the abnormal regions on the slice can be identified and located.

A Unified Framework for Historical Document Restoration: We have developed a useful image processing tool to restore historical document images. It incorporates multiple preprocessing functions, the proposed coarse and fine registration methods, several bleed-through correction methods and some postprocessing routines. It is convenient for the users to try different processing methods or the combination of them on real-world historical documents. If large amount of documents need to be processed for experiments or practical use, the system can also conduct batch processing without interrupting the users.

1.3 Thesis Outline

Chapter 2 gives an overview of the general image registration framework which consists of four major components: feature detection, feature matching, mapping function estimation and re-sampling. The essential ideas and existing techniques for each component is discussed. The idea of groupwise image registration is also

introduced in this chapter.

In Chapter 3, we introduce a new image feature, called Interline White Run. We present the method to extract this feature from document images and the method of using the detected features to estimate documents' skew angles. Then we evaluate the proposed skew estimation method with real-world skewed document images and compare its performance with other skew correction methods.

In Chapter 4, we present a framework to register the two side images of a historical document. The registration framework consists of a coarse rigid registration procedure and a fine non-rigid registration procedure. For the coarse registration procedure, we extract a pair of sub-images from the two images and use them to estimate an Euclidean transformation model. The fine registration method incorporates a control point selection method, a spline-based free-form transformation model and a similarity measure based on residual complexity. To evaluate the performance of the proposed registration approaches, we build a unified document restoration framework which incorporates image preprocessing routines, the proposed registration methods, several bleed-through correction methods and some post-processing methods. With this restoration framework, we quantitatively show that the proposed image registration method significantly improves the bleed-through correction results.

Chapter 5 describes a cluster-based groupwise registration method which is capable of constructing a brain CT atlas by registering multiple CT scans from different patients. As the groupwise registration method is built upon pairwise registration techniques, the underlying pairwise image registration method is introduced first. Later in this chapter, we demonstrate that the built atlas can be used to determine the position of an input slice on the axial direction of the brain

and also to identify the abnormal regions on the slices that are affected by traumatic brain injury.

Chapter 6 summarizes the contributions of this thesis and suggests some possible extensions to this work and several future directions.

CHAPTER 2

Background

As described in Chapter 1, image registration has been well studied in various research areas because of its importance in image analysis tasks and its complicated nature. According to the database of the Institute of Scientific Information (ISI), in the last 10 years more than 1000 papers were published on this topic [ZF03]. In early days, image registration was mainly approached by correlation based methods. These methods are mostly reviewed in the first survey paper on image registration presented by Ghaffary et al. [GS83]. Later, Brown provides a much more comprehensive survey of the general-purpose image registration methods [Bro92]. In particular, registration techniques applied in medical imaging are summarized in [EPV93, MF93, MV98]. Zitova et al. provide probably the latest survey paper which covers the majority of the recently emerged as well as some classic methods to image registration [ZF03].

2.1 General Framework

As mentioned before, designing a proper image registration framework to a particular application should take into account the assumed type of geometric deformation between images to be registered, the radiometric deformation and application-dependent data characteristics. Therefore, it is impossible to develop a universal approach which is applicable to all image registration tasks. Nevertheless, most image registration techniques share the same framework which consists of four components as follows:

- **Feature detection:** Depending on the source of information used, the approaches to image registration fall into two categories: feature-based and intensity-based ones [CHH04]. Feature-based image registration methods need to extract a set of geometrical features from the two images to be registered. These features are usually distinctive objects such as corners, edges or anatomical tissues and described with their point representatives (centers of gravity, line endings, distinctive points), also known as control points (CPs) in the literature. Section 2.2 summarizes different types of image features which have been used for image registration and the strategies of detecting these features. The key advantage of feature-based methods is their dimensionality reduction property, which significantly reduces the computational cost and load. Whereas, the major problem with these methods is that they heavily rely on the precise extraction and matching of the image features. Automatic feature extraction and correspondence estimation themselves however are large research areas in computer vision.

  In contrast, intensity-based methods directly register images with their intensities and need no feature extraction procedures. These methods are popular

as dense intensity information is not only readily available at each pixel but also more accurate in estimating local deformations. One of the major disadvantages of intensity-based methods is the extremely high computational cost and computer memory consumed especially when tremendous number of images or 3D volumes are involved. Another challenge with intensity-based methods is the definition of similarity measures, as in many image registration tasks, especially multi-modal registration applications, the intensities of the subject images are significantly different.

- **Feature matching:** With feature-based image registration methods, the feature sets detected in the previous component should be matched so that the correspondences between them can be used to estimate the transformation between the images. To establish the correspondences, a similarity measure and an optimization strategy are required. A similarity measure is usually an objective function which achieves its optimum when two objects (features or images) verify a certain relationship. In Section 2.3, we discuss some commonly used similarity measures such as L_2 norm, sum-of-squared-differences, correlation coefficients and mutual information. The optimization method is an algorithm to find a set of parameters which optimize a given similarity measure with the observed data. Popular optimization methods include Gradient Descent, Quasi-Newton, Conjugate Gradient, Levenberg-Marquardt, BFGS and Stochastic Gradient Descent methods [KSP07].

- **Mapping function estimation:** In order to align two or more images, a transformation model which consists of a transformation or a set of transformations needs to be defined and estimated. We discuss several typical transformation models in Section 2.4. Transformation models can be subdivided into rigid and non-rigid ones. Rigid transformations include only rota-

tions, translations, or their combination (sometimes called roto-translations). The simplest non-rigid transformation is affine which also allows anisotropic scaling. In real-world applications, non-rigid transformation is more often used and challenging. For instance, medical images are usually related with non-rigid transformations due to the physical properties of body organs and tissues. With simple transformation types, the parameters may be directly computed with the detected features. In most cases, search strategies and optimization methods are required to find the optimal value for these parameters. Therefore, appropriate search strategy and optimization function are needed to be carefully chosen.

Another vital mechanism in mapping function estimation is regularization which constrains the estimated transformations to be smooth or invertible. As the existence and uniqueness of the demanded transformation are not guaranteed, regularization is essential. In some registration methods, regularization even defines the key properties and behavior of the transformation model.

- **Image re-sampling and interpolation:** This component transforms the target images using the mapping functions (transformation) estimated in the above component. As transformed coordinates may be fractional, interpolation methods are necessary to obtain the final registered images. Section 2.5 reviews the key strategies for image re-sampling and interpolation.

Figure 2.1 demonstrates the above four components and in the following sections of this chapter, we discuss them in more details. As any solutions in all other applications, a complete image registration framework should include proper techniques and measures to verify the system. Therefore, in Section 2.6 we discuss the

(3) transform estimation

(1) Feature detection   (2) feature matching   (4) re-sample
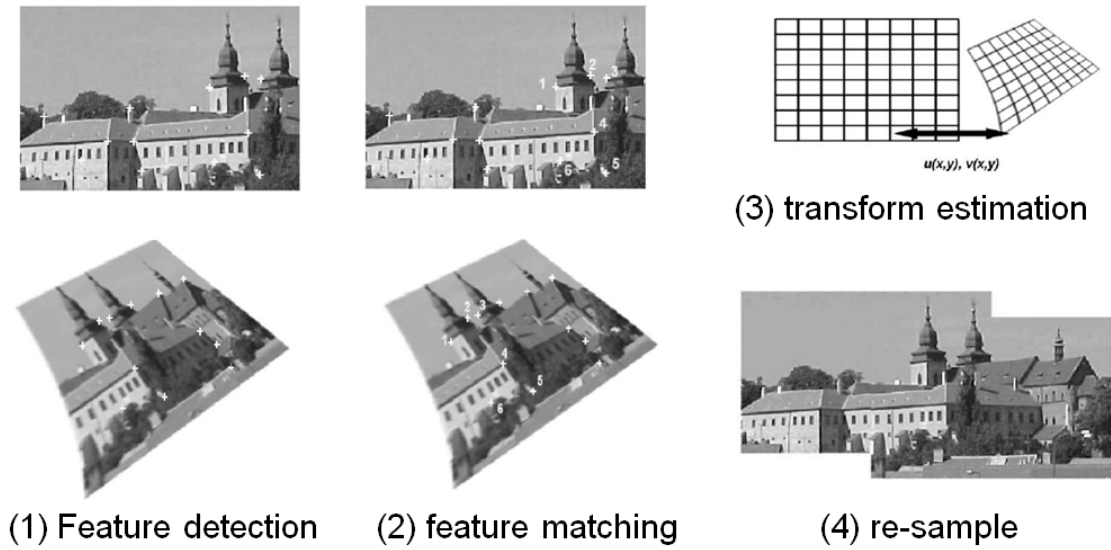
Figure 2.1: Illustration of the four components in a general image registration framework. In image (2), the matched features are labeled with the same numbers.

components of registration errors and review some most commonly used techniques and measures to evaluate the accuracy of the proposed registration approaches.

Finally, in many applications, it is not a pair but a set of images to be transformed to a common coordinate system. One typical application in medical imaging is to construct 3D atlas by registering multiple 2D scans from different study cases. The technique that solves this problem is called groupwise image registration which is usually built upon conventional pairwise image registration. In Section 2.7, we briefly review some groupwise image registration approaches.

2.2 Feature Selection and Detection

As we have discussed, the first step of feature-based image registration method is to extract proper image feature sets from the images to be registered. Features refer

to salient structures or objects, which capture the spatial relationship between the images to be registered. In term of pure image concepts, they can be classified into three categories.

Region Features: A classical region feature is the result of the projections of general high contrast, closed boundary regions of an appropriate size [GS85, GSP86]. Region features are usually detected by means of segmentation methods [PP93] and represented with their centroids or centers of gravity.

Line Features: Commonly used line features include line segments [HMP92, MH97, WC97], object contours [LMM95, DK97, GSC98]. In particular, Lu et al. [LT03] detect line segments connecting the centroids of the nearest connected components to estimate the skew distortion on document images. In this thesis, we use the line segments exactly lying between the baseline and xline of adjacent text lines to register a skewed document with its correctly-posed imaginary image.

Point Features: Traditional point features are line intersections [VZB98], centroids of connected components [LTW94, Bai87], corners [WSYR83, BS97].

Choosing proper feature sets to use in a particular image registration application depends on the characteristics of images to be registered. In general, if typical images in the application contain a lot of details, for instance, in remote sensing domain, distinctive objects such as lakes, roads, rivers are usually selected as matching features for the purpose of image registration. While in medical imaging domain, since most images are dominated by homogeneous areas and are not rich in details, regions with prominent illumination changes are often employed.

However, some criteria should be commonly satisfied by all features used for image registration. Firstly, since they are used to estimate the mapping functions between images, the chosen features should be invariant to the deformation

assumed in the application. Table 2.1 summarizes most commonly used transformation models and their corresponding invariants. Secondly, features should be distinctive enough so that the correspondence between them can be precisely established. This also helps to accurately locate these features on the target image and reference image. Thirdly, the chosen features should spread all over the reference image as well as the target image so that sufficient number of common elements can be identified. As the images to be registered are usually dissimilar, missing of matching candidates is always a serious problem of image registration. Take the registration of historical documents for example, the registration is actually between the foreground strokes and their corresponding blurred seeped ones. In many cases, the ink do not seep to the reverse side, so for many strokes, there are no corresponding points. We have to make sure there are enough number of common elements. On the other hand, the number of features should not be too large. Otherwise, too much computation will be involved and the probability of mismatches also increases. Fourthly, the chosen features should be easily detected from both images to be registered. Moreover, the accuracy of feature location can significantly influence the resulting registration. Usually features are independently pre-detected and remain constant in the whole registration procedure. Goshtasby et al. [GSP86] proposes a refinement approach where feature detection is iteratively conducted together with the registration. It is claimed that subpixel accuracy of registration could be achieved with this method.

2.3   Feature Matching

As discussed in Section 2.1, the feature sets detected in the first step of feature-based image registration methods need to be matched before they can be used

to estimate the transformation of images. The aim of feature matching is to find pairwise correspondences between the detected features. To achieve this, usually a matching metric such as similarity measure, dissimilarity or cost function is predefined and certain searching strategies are adopted to optimize the metric. Apart from feature matching, the subsequent transformation estimation procedure in these methods also need a proper similarity measure. Intensity-based image registration methods don't detect and match features but still need a similarity measure for transformation estimation. Therefore, in this section, we review and discuss some commonly used similarity measures. The reviewed works are organized based on the core ideas they use.

2.3.1 Feature-based Similarity Measures

When advanced image features other than pixel intensity are used for registration, similarity measures are usually defined directly based on the geometric features extracted. One of the simplest similarity measure is the L_2 norm between the corresponding pairs of landmarks.

$$E_{sim} = \sum_{k=1}^{K} \|\tau(c_k^j) - c_k^i\|^2 \tag{2.1}$$

where c_k^j and c_k^i are the locations of feature points extracted from the image J and I. One of the major advantages of L_2 norm is that the optimal transformation model can be found in closed form for both rigid and some non-rigid parameterizations [BM92]. Robust distance measures including L_1 norm can also be used [Kar01]. These measures assume that the two images to be registered have the same number of feature points with known correspondences. In real-world however, a feature extraction engine often detects different number of points from

the two images and the correspondences between the two sets of points are usually unknown. Therefore, point set matching methods are required. A simple way is to assign the correspondences based on the nearest distance criterion. Alternatively, the correspondences can be represented with the probabilities of all possible combinations of points [MS10]. Accordingly, the similarity measure is generalized as:

$$E_{sim} = \sum_{k=1}^{K} \sum_{l=1}^{l} P_{kl} \|\tau(c_k^j) - c_k^i\|^2 \tag{2.2}$$

where P_{kl} is the probability of point c_k^j corresponding to point c_k^i. Some point set matching methods only estimate the correspondences between the extracted feature points, whereas advanced methods simultaneously estimate the correspondences and the transformation model. Usually, the methods based on simultaneous estimation alteratively update the estimated correspondences and optimize the similarity measure with the new correspondences.

2.3.2 Sum-of-squared-differences

As we have mentioned in Section 2.1, intensity-based image registration methods register images directly with the dense intensities of the images. Accordingly, these methods use dense pixel-wise (voxel-wise for 3D registration) similarity measures which are suitable to estimate local dense deformation regions. One of the simplest intensity-based similarity measures is the sum-of-squared-differences (SSD):

$$E_{SSD}(\tau) = \sum_{n=1}^{N} (I_n - \tau(J_n))^2 \tag{2.3}$$

where I_n and J_n are the intensities of the nth pixel on the two images to be registered and τ is the transformation model to be estimated.

SSD is widely used in image registration methods for it simplicity in terms of understanding and implementation. SSD is a good choice for image registration methods of which the input images only differ by Gaussian noise. A downside of this similarity measure is its sensitivity to outliers and image artifacts due to the squaring of each term which actually weights large errors more heavily than small ones. To reduce this bad effect, researchers have proposed sum-of-absolute-differences (SAD):

$$E_{SAD}(\tau) = \sum_{n=1}^{N} \|(I_n - \tau(J_n))\| \tag{2.4}$$

Like SSD, SAD also achieves the best performance when the images to be registered have identical dense intensities if they are perfectly aligned. In the real-world however, especially in multi-modal image registration tasks, the images are often inherently different in term of dense intensities, which limits the usage of this type of mean squared measures including SSD and SAD.

2.3.3 Correlation Coefficient

Cross correlation is a classical similarity metric for template-based image registration. Its application is first motivated by squared Euclidean distance and the most commonly used version of this measure, which is the normalized cross correlation (NCC) can be represented as [Lew95]:

$$\gamma(u, v) = \frac{\sum_{x,y}(f(x,y) - \bar{f}_{u,v})(t(x-u, y-v) - \bar{t})}{\sqrt{\sum_{x,y}(f(x,y) - \bar{f}_{u,v})^2}\sqrt{\sum_{x,y}(t(x-u, y-v) - \bar{t})^2}}. \tag{2.5}$$

where f is the image and the sum is over (x, y) under the window containing the feature t positioned at (u, v). $\bar{t}$ is the mean of the feature and $\bar{f}_{u,v}$ is the mean of $f(x, y)$ in the region under the feature.

As shown in Equation 2.5, this metric measures the similarity between a pair of windows, of which one is on the target image and the other is on the reference image. In order to register two images, this measure is computed for each possible pair of windows and the window pairs with the maximum are set as the corresponding ones.

Correlation-like methods are popular in that they directly make use of image intensities and thus no feature detection is needed. Also, it can be efficiently implemented in the spatial domain and transformation domain. However it has a serious limitation and two major disadvantages. First, it can only register images with only translation distortion and possibly slight rotation distortion. Second, it is quite sensitive to the intensities differences between the target image and reference image. Third, it is highly computational complicated. Therefore, enormous generalizations are made to this metric to tackle the limitation and the two disadvantages.

The method presented in [Sim96,Ber98] mainly aim to extend correlation-based registration methods to images with more complicated geometric deformations. In order to reduce the computational cost, Pratt [Pra74] applied filters on noisy images to reduce the size of source data. Meanwhile, Wie [WS77] and Anuta [Anu70] improve the efficiency of correlation-based registration methods by applying them on edges extracted instead of the original images. Apart from these generalization, other metrics similar to correlation are also employed to improve the registration accuracy in particular application areas. Such examples include the correlation ratio metric used in multimodal registration [RMPA98] and Hausdorff distance (HD) [HKR93] for images with perturbed pixel locations.

2.3.4 Mutual Information

Mutual Information (MI) is the recently emerged similarity metric for image registration. It measures the statistical dependency between two images and is particularly suitable for the registration of medical images. The MI between two random variables X and Y is defined as [ZF03]:

$$MI(X,Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y) \qquad (2.6)$$

where $H(X) = -E_X(\log(P(X)))$ represents the entropy of the random variable X and $P(X)$ is the probability distribution of X. $H(Y|X) = -E_{Y|X}(\log(P(Y|X)))$ represents the conditional entropy and $H(X,Y) = -E_{X,Y}(\log(P(X,Y)))$ is the joint entropy. Similar to correlation, the matching pairs with maximum MI value are set as corresponding ones. The major issue with this metric is the even higher computational cost than correlation based methods. Therefore, much effort have been made to speed up the MI optimization procedure. Generally speaking, pyramidal approaches are used for this purpose, such as Marquardt-Levenberg method [TU98] and the method combining hierarchical search and simulated annealing [ROC+99].

2.3.5 Speedup Techniques

Due to the large size and number of document images to be processed, speed-up strategies are usually employed in most registration approaches in order to reduce computational cost. In general, pyramidal methods also known as coarse-to-fine hierarchial approaches are used. For instance, a sub-window is first used to find probable candidates of the corresponding window in the reference image and then

the full-size window was applied.

In general, this coarse-to-fine hierarchical strategy applies the usual registration methods, but it starts with the reference and sensed images on a coarse resolution. Then they gradually improve the estimates of the correspondence or of the mapping function parameters while going up to the finer resolutions. At every level, they considerably decrease the search space and thus save the necessary computational time. Another important advantage resides in the fact that the registration with respect to large-scale features is achieved first and then small corrections are made for finer details. On the other hand, this strategy fails if a false match is identified on a coarse level. To overcome this, a backtracking or consistency check should be incorporated into the algorithms.

Due to its inherent multi-resolution character, wavelet decomposition of the images has been recommended for the pyramidal approach. Methods can differ in the type of the applied wavelet and the set of wavelet coefficients used for finding the correspondence. Most frequently used methods decompose the image recursively into four sets of coefficients by filtering the image successively with two filters, a low-pass filter and a high-pass filter H, both working along the image rows and columns.

2.4 Mapping Function Estimation

Mapping function defines the way to deform the target image to match the reference image. In general, the type of the mapping functions should be chosen according to the priori knowledge about the image acquisition process and the expected image degradations. If no such priori information is available, the model should be flexible

and general enough to handle all possible degradations which might appear. After the correspondences between features have been established, the parameters of the assumed mapping function are estimated. This mapping function is expected to overlay the sensed image over the reference image as close as possible. In order to achieve this, the correspondences between features and the constraint conditions like the continuity are employed in the process. Therefore, the task to be solved consists of choosing the proper type of mapping functions and accurately estimating their parameters. Deciding the proper type of mapping functions for the input images should take into account the assumed geometric deformation, the method of image acquisition and the required accuracy of the registration.

According to the amount of image data they use as their support, transformation models can be categorized into global models and local models. Global models use all features to estimate one set of transformation parameters that are assumed to be valid for the entire image. On the other hand, the local models treat the image as a composition of patches and the transformation parameters depend on the location of their supporting features in the image. Global transformation models require the tessellation of the image, like a triangulation, and the defining of mapping functions for each patch separately.

Aside from the source of information used, transformation models are also subdivided into rigid ones and non-rigid ones based on the behaviors of the transformations. Historically, rigid transformation models such as Euclidean and similarity transformations are used in many applications. These models have a small set of parameters including rotation and translation parameters. For their simplicity, rigid transformation models are good candidates for global transformation models. The simplest non-rigid transformation model is affine transformation which also allows skews and shearing. Figure 2.1 illustrates some of the commonly used trans-

formation models and their behaviors and invariant values. In practice, however, affine transformation is often regarded as rigid due to its simplicity. The most popular non-rigid transformation models are piecewise affine, radial basis function (RBF) and B-splines. In the following parts of this section, we discuss some of these transformation models in more details.

Group	Matrix	Distortion	Invariant properties
Projective 8 dof	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$		Concurrency, collinearity, **order of contact**: intersection (1 pt contact); tangency (2 pt contact); inflections (3 pt contact with line); tangent discontinuities and cusps. cross ratio (ratio of ratio of lengths).
Affine 6 dof	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Parallelism, ratio of areas, ratio of lengths on collinear or parallel lines (e.g. midpoints), linear combinations of vectors (e.g. centroids). The line at infinity, $\mathbf{l}_\infty$.
Similarity 4 dof	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Ratio of lengths, angle. The circular points, $\mathbf{I}, \mathbf{J}$ (see section 2.7.3).
Euclidean 3 dof	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Length, area

Table 2.1: Geometric properties of commonly occurring planar transformations [HZ04]. The matrix $A = [a_{ij}]$ is an invertible $2 * 2$ matrix, $R = [r_{ij}]$ is a 2D rotation matrix, and (t_x, t_y) a 2D translation. The distortion column shows the typical effects of the transformations on a square. Transformations higher in the table can produce all the actions of the below ones. These range from Euclidean, where only translations and rotations occur, to projective where the square can be transformed to any arbitrary quadrilateral (provided no three points are collinear).

2.4.1 Global/Local Mapping Function

In general, the number of CPs is usually bigger than the minimum number of CPs that are required by the determination of the transformation model. The parameters of the transformation functions are then computed by means of least-quare fit, so that the polynomials minimize the sum of squared errors at the CPs. Higher order polynomials usually are not used in practical applications because

they may unnecessarily warp the sensed image in areas away from the CPs when aligning with the reference image.

In many cases, a global mapping cannot properly handle the local deformation on the images. This happens in historical documents where uneven surfaces are formed near the spine areas and in medical images where the growth of a intracerebral hemorrhage (ICH) locally affects the image. In this case, the least square technique used for global mapping function estimation actually averages out the local geometric distortion equally over the entire image, which is obviously undesired. Currently proposed local mapping functions include Goshtasby's piecewise linear mapping [GSP86] and piecewise cubic mapping [Gos87] and Akima's quintic approach [WRSS96]. These methods require the images being subdivided into rectangular or triangular blocks and apply a simple transformation (usually rigid ones) to each block. Such methods are usually fast but tend to introduce approximation errors to the truly non-rigid deformations.

2.4.2 Radial Basis Function

Radial basis functions are the representatives of globally estimated and locally sensitive mapping functions. The estimated mapping function can be represented with a linear combination of translated radially symmetric function plus a low-degree polynomial [ZF03]:

$$(u, v) = a_0 + a_1 x + a_2 y + \sum_{i=1}^{N} c_i g(x, x_i) \tag{2.7}$$

where a_i and c_i are unknown parameters and $g(x, x_i)$ is the basis function which can be Gaussian kernel, multiquadric and thin-plate spline. The most often

used representatives of the radial basis functions are the thin-plate splines (TPS), where the 2D radial terms have the form:

$$g(x, x_i) = \| x - x_i \|^2 \ln(\| x - x_i \|) \qquad (2.8)$$

The TPS can be viewed as a very thin plate, which is fixed at the positions determined by the CPs in the reference image in the heights that are given by the x or y coordinates of the corresponding CPs in the reference image [Gos88]. These methods are first used in mechanic and engineering for the interpolation of irregular surfaces and now are widely used for medical image registration [Kar01]. One of the reasons that TPS is popular is that it can be decomposed into linear and non-linear parts. However, as each CPs has a global influence on the transformation, TPS transformation suffers a high computational cost when large number of CPs are used for mapping function estimation. Therefore, TPS transformation is more suitable for feature-based image registration methods where a set of feature points are extracted and matched.

2.4.3 Regularization

As we have discussed in Section 2.1, regularization plays an important role in image registration. It enforces certain properties of the estimated transformation model such as smoothness, rigidity and continuity. In particular, when we try to estimate a transformation model with a set of CPs or feature points, the existence and uniqueness of the transformation model are not guarantied. In other words, there may be infinite number of transformations that can match the CPs or feature points but have different behaviors in other pixels of the images. Therefore, by constraining on the behavior of the transformation model through regulariza-

tion procedures, we attempt to find precise and unique transformation model that exactly describes the deformation between the two images.

Actually, regularization is a key components in many research areas such as machine learning and computer vision. Many problems in these research areas are ill-posed [CH02]. By ill-posed, we mean the solutions to these problems don't satisfy all the three conditions: existence, uniqueness and continuity. The theory of regularization was first proposed by Tikhonov [Tik77]. A traditional regularization method is to add a regularization term to the optimization procedure. In this way, the cost function to be optimized becomes:

$$C(f) = S(f) + \alpha R(f) \tag{2.9}$$

where $S(f)$ is the original objective function to be optimized and $R(f)$ is the regularization term. α represents the trade-off between the two terms. A popular operator for $R(f)$ is the first or second order derivative operator.

2.5   Image Re-sampling and Interpolation

Once the mapping functions between images are estimated, they are used to transform the target image to obtain the registered new image. The transformation can be realized in a forward or backward manner. Forward transformation is straightforward in theory but complicated to implement. With this strategy, the coordinates of each pixel in the target image are mapped to compute the coordinates of corresponding point on the registered image. As the transformed coordinates are not always integers, discretization and rounding are inevitable to happen. Therefore, holes will be formed at places where no transformed coordinates are discredited

to. Meanwhile, overlaps occur at points which multiple transformed coordinates are discredited to. The simple way to address this problem is to detect these holes and overlaps and then interpolate the gray value at these positions using the gray values of nearby non-hole/overlap points on the registered image.

No holes or overlaps will be formed on the registered image if background re-sampling strategy is used. With this method, the registered image is in the same coordinate systems with the reference image. For each point on the registered image, the coordinates of its counterpart on the target image is computed by applying the inverse of the estimated mapping function to its coordinates. Similarly, the transformed coordinates may be not integers, therefore, its gray value is interpolated from other points on the target image.

Interpolation is usually realized via convolution of the image with an interpolation kernel. An ideal interpolation kernel such as sinc function is difficult to implement in practice because it spatially is unlimited. Therefore, truncated and windowed sinc interpolators are investigated as reported in the literature. Most commonly used interpolation methods include nearest neighbor function, the bilinear and bicubic functions, quadratic splines [BB95, Dod97], cubic B-splines [UAE91], higher-order B-splines [CP04], Catmull-Rom cardinal splines [RU98], Gaussian [App96]. As interpolation methods are essential for medical image processing, proposed interpolation methods are usually compared and evaluated with experiments on medical images [LCS99]. Generally speaking, nearest neighbor function should be avoided when medical images are registered. Bilinear interpolation is most commonly used for it's probably the best trade-off between accuracy and computational complexity. Cubic interpolation is recommended when the geometric transformation involves a significant enlargement of the sensed image. Nearest neighbor interpolation should be considered only when the number of intensities is

low.

2.6   Evaluation of Registration Accuracy

Regardless of the types of images to be processed, the registration approaches adopted and the application areas, an evaluation method is needed to illustrate how accurate the proposed registration methods are. This problem is non-trivial as each component of the image registration framework could introduce errors to the final registration results.

Firstly, localization error referring to the inaccurate detection of features will result from imperfect feature detectors. It is impossible to directly measure this error with given images. Even if we manage to adopt well-implemented feature detectors, this localization error is inevitable as there is always a trade-off between the number of detected features and the quality of detected features. For some applications, large number of low-quality features may produce better registration results than small number of high-quality features.

Secondly, matching errors which is due to mismatches could be introduced in the feature matching procedure. In certain cases, this error can be measured with the number of mismatches. As this error significantly affects the registration accuracy and even leads to failure of the registration process, they should always be avoided. To correct this error, consistency check or cross-validation approaches could be used to identify mismatches. If multiple matching methods for given images are available, they all are applied to the given images and only the feature pairs agreed by most or all matching methods are considered as valid corresponding ones. If no other reliable matching methods are suitable for the particular applica-

tion, each time one pair of features are excluded from the calculation of mapping parameters and are used to check the mapping function. If the displacement of this excluded feature pairs is below a given threshold, they are accepted as a valid corresponding pair.

Thirdly, modeling error is introduced by the component estimating the mapping functions between images. It refers to the difference between the estimated mapping function and the actual geometric deformation between images. This error consists of two parts. First, the chosen mapping function type may not precisely represent the actual geometric deformation. Second, the parameters of chosen mapping function may not be accurately estimated. The first error is due to the lack of priori knowledge about the geometric distortion and the later one may be caused by the small number of detected features or the low quality of detected features. Traditionally, modeling error is measured with mean square error at control points (CP), known as control point errors (CPE) or test point error (TPE). These metrics however actually quantifies how the features can be fitted by the chosen mapping function instead of how the mapping function reflects the actual deformation. Moreover, the localization error of these test points may negatively affect this measure.

Fourthly, with rounding and interpolation approaches employed, additional error will be inevitably added to the total registration error. As no ground truths about the deformation between images are available in most image registration tasks, this error cannot be quantified.

Finally, although the error of image registration approaches is complicated in nature, there are some ways to measure it. First of all, we can always visually assess and compare the performance of different image registration approaches.

Then similar to the evaluation of matching errors, consistency check and cross-validation are also adoptable. The only difference is now they are conducted between multiple entire image registration frameworks instead of different feature matching approaches. Different variations of consistency check approaches are available in [BCT+98, HHD+00].

To sum up this section, estimation of accuracy of registration algorithm is a substantial part of registration process. Without quantitative evaluation, no registration method can be accepted for practical utilization. A lot of work has been done on validation of rigid-body registration while validation of non-linear, local and elastic registration methods is still at its infancy [ZF03].

2.7   Groupwise Image Registration

As indicated by their names, groupwise image registration methods register more than two images to a common coordinate system. It is not a single transformation but a group of transformations that need to be determined [ZLMGW05, TCM+06]. A typical application of groupwise image registration is to register multiple medical scans, e.g. MRIs from different patients to construct an atlas which describes the anatomical variations of populations.

There are several approaches to groupwise image registration. The simplest way is selecting a certain image (the first or random one) as the reference image and then registering the remaining images to the reference image using pairwise image registration. In other words, such approaches organize the images to be registered in a centralized and flat structure. On the one hand, the centralized structure makes these approaches lose the variations of the target images, which is a crucial problem

with atlas construction applications. More importantly, these approaches are very sensitive to the choice of the reference image. If an outlier image is chosen as the reference image, the built atlas could be completely wrong. On the other hand, as all target images are equal in the flat structure, these groupwise registration methods neglect the possible relationship between the images to be registered. For instance, if the images to be registered are acquired at consecutive time instance, two consecutive images should be close to each other in terms of the underlying transformation.

As an improvement, sequential image registration is then proposed in research areas such as remote sensing or time lapse video imaging. The idea is to sequentially register next image to the previous one and track the deformation field over time. This approach is supported by the fact that two neighboring images usually have a large overlap, whereas some distant frames have small or zero overlap due to the progressive object or camera shift. Sequential image registration not only significantly improves the processing speed, but also benefits from a large area overlap between the consecutive frames in contrast to the fixed-reference-image approach. To further improve the robustness of this approach, one can register the next image to the average of previously aligned images. The main benefit of averaging is to remove noise and artifacts, while the disadvantage is the reduction of the texture statistics.

To address the centralization problem, so-called 'true' groupwise image registration methods register all images simultaneously [JDJG04, ZLMGW05, TCM$^+$06, MTT06]. These approaches fundamentally reduce the bias of choosing any particular reference image. Twining et al. [TCM$^+$06] proposed to use a Minimum Description Length (MDL) to align the target images. The MDL principle states that the transmission of a model of the data, together with the parameters of the

model should be as short as possible. Zollei et al. [ZLMGW05] presented the congealing approach for groupwise image registration. They used the total element-wise entropy of the input image sequence as the objective function to optimize. The entropies are computed at each coordinate location and then added together. Vedaldi et al. [VS07] improved the congealing approach by minimizing the total complexity of the data set together with the amount of image distortions. Joshi et al. [JDJG04] computed the sum of the SSDs between all images and the reference image simultaneously with the registration procedure. By choosing the SSD similarity measure the reference image can be found analytically as an arithmetical average of all the images.

2.8   Summary

In this chapter, we mainly reviewed the four essential components in the general framework of image registration. They are feature selection, feature matching, mapping function estimation and image re-sampling. For each of these components, we briefly described its workflow and then discussed its commonly used methods, measures and models. To complete this general introduction to image registration, we also included some discussion on analyzing registration errors and evaluating registration methods. Finally, we introduced the idea of groupwise image registration and gave a brief overview on the major methods used in this problem.

CHAPTER 3

Single Registration of Printed Documents

As mentioned in Section 1.1, some image registration methods register the images of a scene to a model of the same scene. A typical application of this kind in medical imaging is to register an input scan to an atlas, such as brain CT or lung MRI. The model can also be a computer representation of the scene, such as an imaginary image that would be captured if the skewed document was posed in precisely upright position. By registering the skewed document to this imaginary image of itself, we actually correct the skew distortion on the skewed document image. Based on this idea, in this chapter, we present a new image feature called inter white run and use it to estimate the skew angle of a skewed document image. In this application, the transformation model between the two images to be registered is rigid and has only one parameter for rotation, which significantly simplifies the mapping function estimation procedure. In real implementation, we can even directly compute the transformation model with the feature we extract.

As in this chapter and the next one, we are both dealing with document images,

we give a brief overview to document imaging and document image restoration in Section 3.1. In section 3.2, we review the four groups of methods that have been widely used for document skew correction. The new feature is presented in Section 3.3.1 and we estimate the skew angle of a degraded document images with it through the method described in Section 3.3.2. We evaluate the proposed feature and skew correction method with real-world data and provide some discussion on the experimental results in Section 3.4. Finally, we conclude this chapter with a summary of the contributions and some potential future research directions related to this work.

3.1 Document Imaging

In this digital age, more and more information is captured and stored in electronic forms, which can be conveniently accessed and modified on-line through digital libraries or other web services. However, there are still a large collection of documents that are printed or handwritten in physical media and therefore are difficult to locate and access [Bai03]. This is particularly true for many historical documents that are of great value but either out of print, deteriorated, or sealed in archives for preservation. A useful digital library is expected to integrate these resources into a database so that they are searchable, readable or even modifiable by a worldwide large population. As a result, document digitization is playing an important role in the advancement of current digital libraries.

In general, a complete digitization cycle consists of three phases: imaging phase, recognition phase and content recovery phase [DWPL04]. First, electronic images or microfilms of physical resources are acquired using scanners or cameras. These images are usually produced in batches and the digitizing devices are hardly

aware of the content of the material, which makes indexing a difficult task. In the second phase, the content of these resources are extracted. For this purpose, image processing techniques are firstly applied to remove noises and to correct various distortions that may affect the subsequent document image analysis tasks. Then text regions or images can be detected and extracted from these images using some layout analysis methods. The extracted text regions or blocks can be fed to Optical Character Recognition (OCR) engines for recognition. Some image enhancement procedures may be applied prior to the recognition step. Upon finishing, plain text is available for reading and editing purposes. As structural information is usually important for human perception and useful for information preservation and dissemination, the third phase is employed to convert the plain text to logical or physical formats. In this way, the original physical document is completely converted to its electronic form, which can be easily accessed from the large on-line database by a wide range of users. Nevertheless, a link between the reconstructed electronic document and its original image form may still be maintained for further verification purposes.

The traditional way of converting a physical document to its electronic form is through flatbed scanners, which usually consist of a glass pane, an underneath bright light across the pane and a moving optical array being either a charge-coupled device (CCD) or contact image sensor (CIS). Images to be scanned are placed face down to the scanning plane and the sensor array and light source move through the pane to capture the entire area. In practice, this procedure usually introduce various geometric and photometric distortions to the captured images. On the one hand, photometric distortions are caused when non-planar materials, such as rolled scripts and folded papers, could not be flattened on the scanning plane. This is the case for thick bounded documents which can hardly be pressed

down near the spine regions. It is especially true for delicate historical materials which are so fragile that they could be easily damaged under external forces thus resulting in a permanent loss of information. On the other hand, imprecise positioning of documents or different settings in the scanning process often lead to errors, such as translation, rotation, and scaling.

These distortions must be removed to facilitate human perception and machine recognition. On one hand, clean and well-posed images are necessary to enable human readability. On the other hand, most document analysis tasks are extremely sensitive to these distortions. A large-scale test of current commercial OCR systems [G. 99] has demonstrated that the accuracy of current OCR devices falls abruptly when image defects exist, such as heavy and light print, stray marks, curved baselines, and shaded background. Improved image processing techniques are thus needed to alleviate the effects of such distortions in the pre-processing step prior to the subsequent document analysis tasks.

3.2   Document Skew Correction

Document skew, as shown in Figure 3.4 is a kind of rotation-induced geometric distortion that has been introduced during the digitizing process. As discussed in the previous section, the existence of skew distortions degrades and complicates subsequent document analysis tasks, therefore they must be detected and corrected before the ensuing document processing tasks such as document layout analysis, OCR and document image retrieval. Theoretically, correcting the skew distortion on a degraded document image can be solved by registering the document to an imaginary image that would be captured if the document was posed in exactly upright position. In this image registration task, the usually used image features

include the line segments connecting the centroids of the nearest connected components [LT03], the centroids of the connected components [Bai87] and the points of the last black runs [LTW94]. As only global skew distortion is considered, the mapping function between the two images to be registered is rigid and can be described with an one-parameter Euclidean transformation. The single parameter corresponds to the skew angle by which the document was deformed.

(a)   (b)
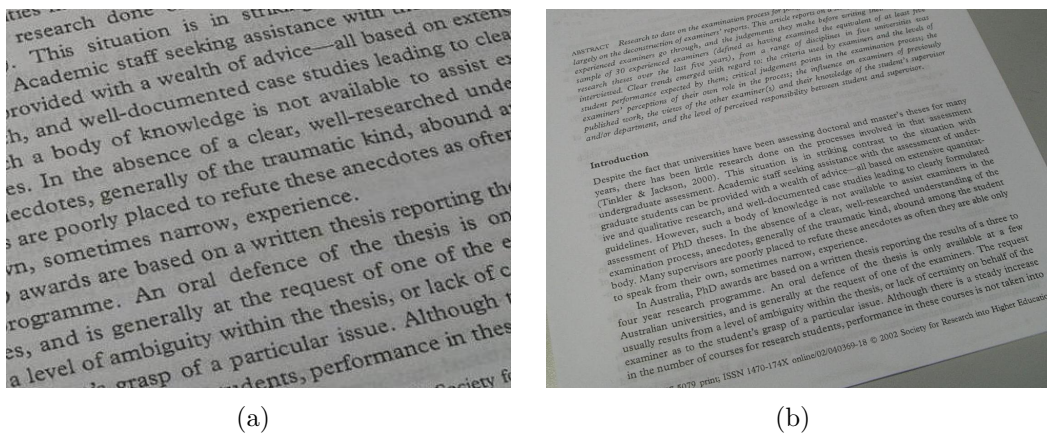
Figure 3.1: Two sample images that are degraded by skew distortions. The left image was cropped from a larger image.

In the literature, three groups of registration methods have been proposed to estimate the skew angle of degraded document images. The first group of methods which are based on projection profile analysis [Bai87] search for the projection angle by which the resultant projection profile forms the most distinguishable peaks and troughs. These methods commonly assume that the documents' skew angles lie within a small range. When the skew angle is large, the computational load increases dramatically and the registration accuracy decreases significantly. The second group of skew correction methods are based on hough transform analysis [LTW94]. Such methods calculate the orientations of the normals from the origin to all possible lines crossing a point on the text line and inspect for local maxima, which represent the orientations orthogonal to text lines. These methods

have no detectable angle restriction, however, they are computationally intensive even with some data reduction techniques. A common limitation shared by hough transform based methods and projection profile based methods is that they require sufficient number of text lines on the document to function well. The third group of skew correction methods are called nearest neighbor based methods [LT03]. They fit straight line segments that connect the centroids of the nearest connected components and the slant angles of the detected line segments vote for the skew angle of the document image. Such kind of methods can detect skew angles in very high precision only if the connected components on the document can be accurately detected. Furthermore, connected component analysis is highly time consuming and thus these skew correction methods are usually slow. In addition, all these three types of method assume that document images are captured in an upright orientation. As a result, few of them can deal with documents that were captured upside-down.

3.3   Registration with Interline White Runs

In this chapter, we present a new method to correct the skew distortion on degraded document images. This method registers the skewed document to an imaginary image of the document itself posed in exactly upright orientation using a new image feature called Interline White Run (IWR). IWRs are the continuous white pixels that lie exactly between the baseline and x line of adjacent text lines. In order to extract this feature, we horizontally scan the document and accumulate all the white runs to form the white run histogram. As the transformation between the two images to be registered has one single rotation parameter, we directly compute this parameter with the extracted IWRs.

3.3.1   White Run Histogram

At the very beginning, the input document images need to be preprocessed to re-
move certain noises before any ensuing analysis. In particular, small-sized noises
such as salt and pepper noise should be suppressed with a global mean filter. Then
we binarize the input images in order to detect white runs. Plenty of document im-
age binarization methods [TT95,WJC02] have been reported to server this purpose
in the literature. In this work, we adopt Niblack's text binarization algorithm [W.
86], as experiments have shown that Niblack's thresholding technique generally
outperforms most of other global and local thresholding techniques in terms of
accuracy and efficiency.

(a)   (b)
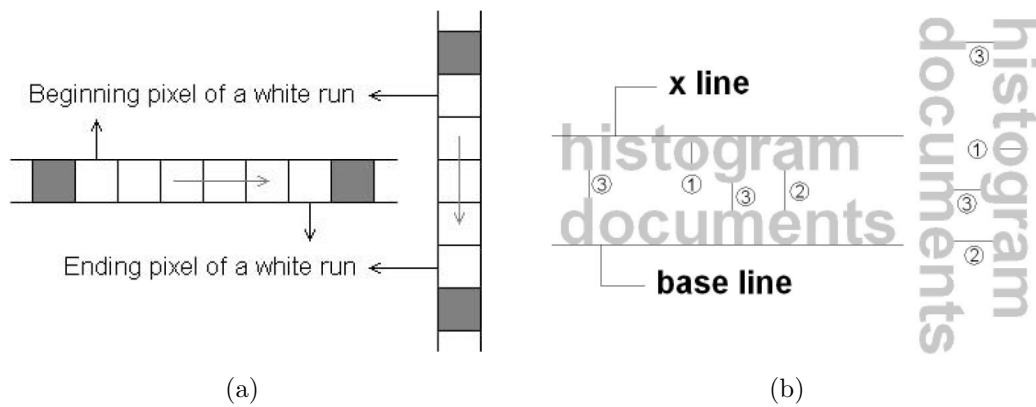
Figure 3.2: Illustration of the three types of white runs. The ones labeled with
② are called interline white runs which produce the second peak in the white run
histograms and are used to estimate the skew angles of degraded document images.
The ones labeled with ③ are used to detect the orientations of document images.

A white run refers to a continuous sequence of white pixels along the direction
by which a document image is scanned. It can be identified by a beginning pixel
(black-to-white pixel) and an ending pixel (white-to-black pixel) as shown in Figure
3.2(a). The length of a white run can be determined with the Euclidean distance
between the beginning pixel and the ending pixel. Figure 3.2(b) illustrates three
types of white runs and the definitions of base line and x line (also known as

top line) of text lines. Base line and x line can be seen as the lower and upper boundary lines of character 'x' in the direction that is perpendicular to that of the straight lines (as the straight line in character 'm') of the text. [LCK05] has proven that these two lines connect the eigen-points of half number of characters and the direction of them is the most close to the direction of text lines. Therefore, by estimating the direction of base lines or x lines, we can determine the direction of text lines and compute the skew angle with respect to horizontal direction.

As shown in Figure 3.2(b), the white runs labeled with ② span exactly between the base line of a text line and the x line of the adjacent text line. We define this type of white runs as interline white run (IWR). In this work, we use this image feature to fit base line and x line for the purpose of skew angle estimation. In order to extract IWR, we first scan the input image horizontally and vertically to generate two sets of white runs with various lengths. Then we collect all occurring white run lengths in two accumulator arrays with which we construct the white run histograms on horizontal direction and vertical direction. Figure 3.3 shows two examples of the obtained histograms for a document image skewed by 20° and 80° respectively. In the figure, the X-axis represents all possible white run lengths and Y-axis represents the corresponding frequency of occurring.

As shown in Figure 3.3, two evident peaks are present in either the horizontal histogram or the vertical histogram of a distorted document image. To be specific, if the document's skew angle is large, horizontal histogram will show two peaks (as shown in Figure 3.3(a)) and vertical histogram will do so if the skew angle is small (as shown in Figure 3.3(c)). The first peak corresponding to small run length is caused by the white runs occurring within text lines, as labeled with ① in Figure 3.2(b). The second peak corresponds to the white runs that exactly span interline spacings (ie. IWR) because text documents normally hold a large amount
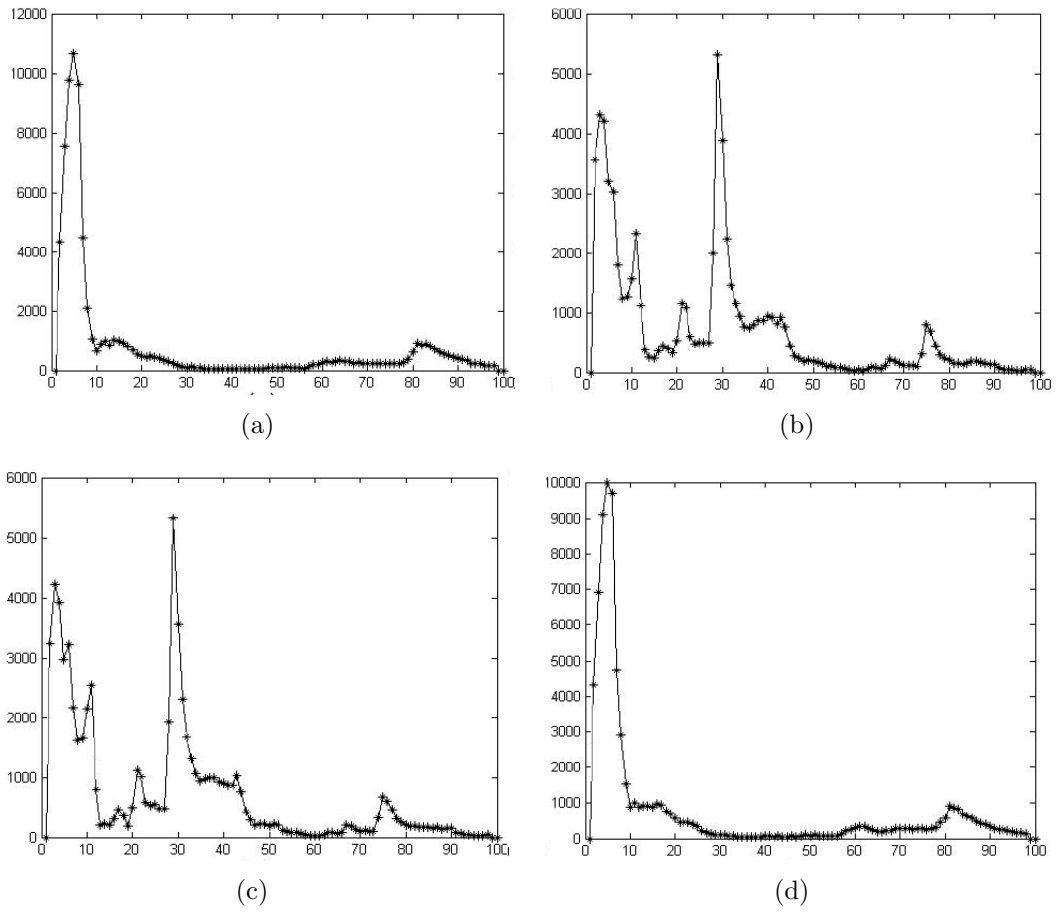
Figure 3.3: Horizontal and vertical white run histograms for the documents shown in Figure 3.4. Images (a-b) are for the document in Figure 3.4(a); Images (c-d) are for the document shown in Figure 3.4(c).

of equidistant interline spacings. Furthermore, most white runs that lie between the two peak bins normally cross interline spacings and span over character ascenders or descenders as labeled by ③ in Figure 3.2(b).

(a)

(b)

(c)

(d)
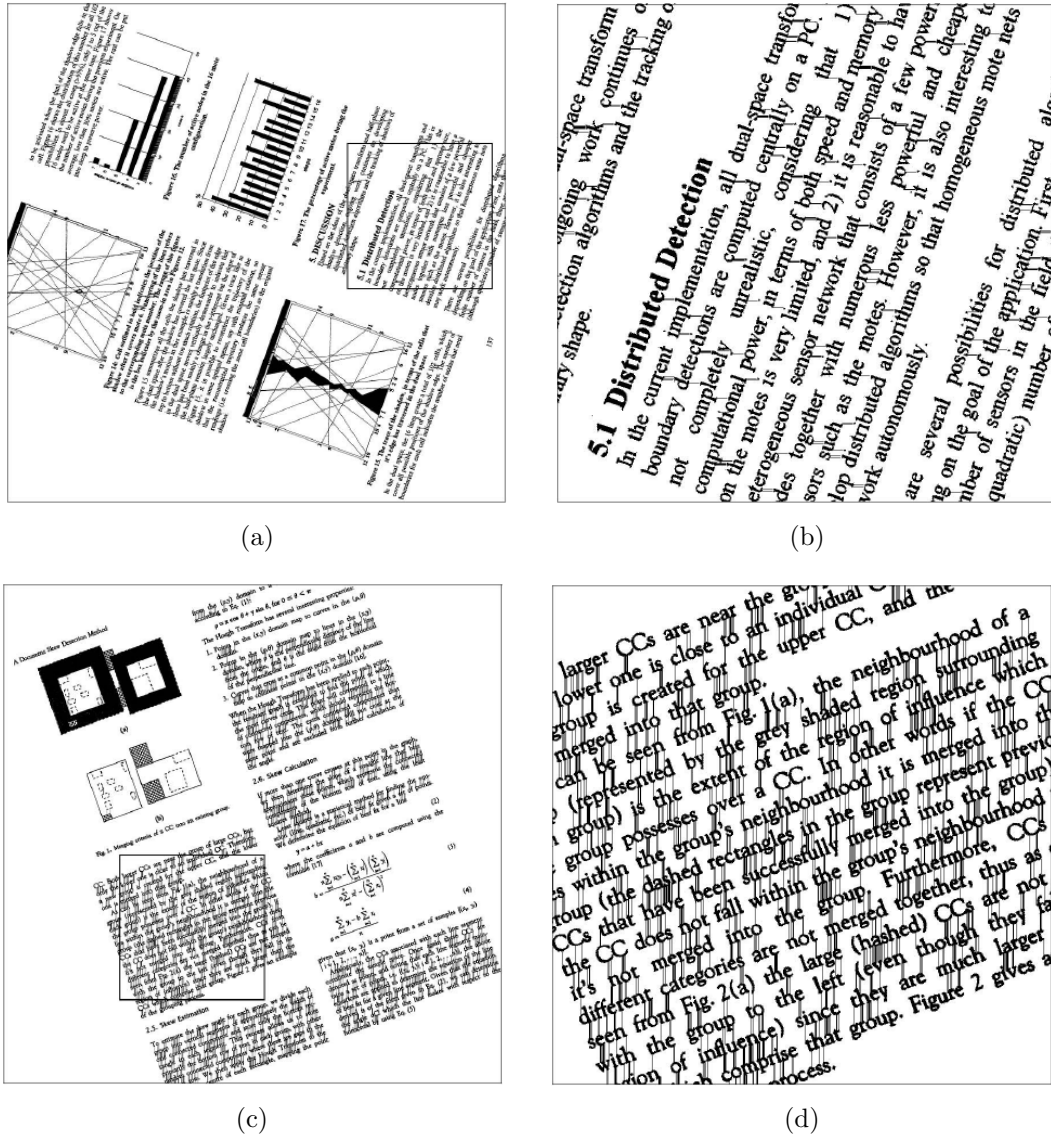
Figure 3.4: Two skewed document images and the interline white runs that were identified from their horizontal or vertical white run histograms

3.3.2   Skew Angle Estimation

As discussed in Section 3.3.1, we estimate document skew by fitting base lines or x lines with the starting points or the ending points of the detected IWR. Also we know from the previous section that IWR corresponds to the second peak on the horizontal or vertical white run histogram. Therefore, in order to detect the white runs exactly spanning the interline spacing, we need to locate the second peak on the white run histograms first. For this purpose, we search for a thresholding white run length k^{th} which maximizes the inter-class variance $\sigma_s(k)$ of the constructed white run histogram:

$$\sigma_s(k) = \frac{(\varphi(k) \sum_{i=1}^{S}(i \cdot p(i)) - \sum_{i=1}^{k} i \cdot p(i))^2}{\varphi(k)(1 - \varphi(k))} \tag{3.1}$$

where S refers to the maximum length of the detected white runs. $p(i)$ gives the normalized density of the white run histogram and $\varphi(k)$ gives the zeroth-order cumulative moment of the histogram as follows:

$$p(i) = n(i)/N_t; \varphi(k) = \sum_{i=1}^{k} p(i) \tag{3.2}$$

where N_t denotes the number of the white runs detected and $n(i)$ gives the number of the white runs with length i. We search for the global maxima (Y-axis value) among the histogram bins with white run length (X-axis value) larger than the threshold determined in Equation 3.1. The X-axis value corresponding to this maxima is the length of IWR.

As we have mentioned, the beginning or ending pixels of the peak runs that exactly span a specific interline spacing normally lie along the x line or the base line of the text lines adjacent to the interline spacing. At the same time, the orientations

of those x lines or base lines exactly represent the document skew angle. Therefore, the document skew can be estimated by fitting straight lines to the beginning or ending pixels of the peak runs that belong to the same interline spacing.

Firstly, we built a local cluster for each peak run with a distance threshold set exactly equal to the length of the peak runs. In this way, the peak runs within the adjacent interline spacings will not be included. Therefore, the peak runs satisfying the following constraint are included into the local cluster:

$$LC = R(x, y) : |x - x_0| < L \bigwedge |y - y_0| < L \tag{3.3}$$

where $R(x, y)$ refers to all peak runs and P denotes the length of the peak runs. (x_0, y_0) is the coordinate of the beginning pixel of a peak run. After the clustering is specified, a local cluster is constructed for each peak run and then sorted based on their size. Those first $C\%$ (G is set at 10 in our system) largest local clusters are chosen for document skew estimation. The orientation of the straight line is estimated as follows by using the least square algorithm:

$$k = \frac{\sum_{i=1}^{N} x_i y_i - n\bar{x}y}{\sum_{i=1}^{N} -n(\bar{x})^2} \tag{3.4}$$

where $(x_i; y_i)$ refers to the coordinate of the beginning pixel of peak runs within the local cluster under study. x and y refer to the mean of the x and y coordinate of the beginning pixel of those peak runs. Document skew can therefore be estimated using the median of those straight line orientations as follows:

$$\theta = \arctan(median(K)) \tag{3.5}$$

where K, $[k_1, k_2, \cdots, k_n]$ refers to the straight line orientations estimated by using

the first $G\%$ largest local clusters.

3.3.3 Orientation Estimation

As mentioned in Section 3.2, most skew estimation methods can not deal with document images that were captured upside-down. They usually assume the skew angles of all degraded document images are within small ranges such as $0°$ to $15°$. However, in real-world, capturing a document upside-down is common. Therefore, in this section, we present a document orientation detection method which also uses the white runs that can be extracted from the histograms. Our method is based on the observation that the number of character ascenders is statistically much larger than that of character descenders in English. In particular, the number of character ascenders and descenders can be estimated by the number of white runs crossing interline spacings, but lying over character ascender and descenders, which normally lie between the two histogram peaks. Such kind of white runs (labeled as 3 in Figure 3.2(b)) normally satisfy three constraints. Firstly, they are usually longer than the white runs lying within text lines (labeled as 1 in Figure 3.2(b)) but shorter than those exactly spanning interline spacings (labeled as 2 in Figure 3.2(b)). Therefore, they normally lie between the two histogram peaks. Secondly, their midpoints lie within the interline spacing since they cross interline spacings. Thirdly, their beginning or the ending pixel normally lie a bit below or above the base line and x line of two adjacent text lines.

The detection of the white runs lying over character ascenders and descenders can be summarized as follows. For each white runs lying between the two histogram peaks, the nearest peak run is first located. The x line xl and base line bl of the two adjacent text lines are then estimated by using the beginning and ending

pixel of the nearest peak run and the skew angle θ estimated in Equation 3.5. Thus, the white run under study is detected to lie over character ascenders if its midpoint lies between xl and bl and its ending pixel lies a bit higher than the xl. On the other hand, it is detected to lie over character descenders if its midpoint lies between xl and bl and its beginning pixel lies a bit lower than the bl. Figure 3.4(c) shows the white runs lying over character ascenders and descenders and Figure 3.4(d) gives the close-up view of the rectangle labeled in Figure 3.4(c). The numbers of character ascenders and character descenders can thus be roughly estimated as the numbers of white runs lying over character ascenders and character descenders, respectively. Based on the character ascender and descender statistics, the orientation of document images can be determined as follow:

$$ DO = \begin{cases} upright & \text{if } AS_{wr} > DS_{wr} \\ upside - down & \text{if } AS_{wr} < DS_{wr} \end{cases} \tag{3.6} $$

where DO refers to the orientation of the document image under study. AS_{wr} and DS_{wr} instead denote the number of white runs lying over character ascenders and descenders and crossing interline spacings, respectively. Combined with the skew angle estimated in Equation 3.5 in the last subsection, the skew angle can be finally estimated as:

$$ \theta_d = \begin{cases} \theta & \text{if upright} \\ 180 + \theta & \text{if upside - down} \end{cases} \tag{3.7} $$

3.4 Experiments and Discussion

To evaluate the proposed skew correction method, we have created a small data set consisting of 38 document images that were downloaded from our university digital library and 14 downloaded from the Internet. All these images contain at

least 15 text lines as well as some non-text components such as tables and graphics. Such an image is shown in Figure 3.4. For generality, the texts on these documents are in different fonts, styles, and sizes. For quantitative evaluation, we deliberately rotated these documents with known angles using Adobe Photoshop to form the 'ground truth'. The rotation angle ranged from 0 degrees to 360 degrees and 21 test documents are positioned upsidedown. We have implemented the proposed method in C++ and tested it on an AMD Opteron CPU 2GHz machine with 2GB RAM. We also implemented other typical skew estimation methods for comparison. They were Lu's connected component based method, Baird's projection profile based method and Le's Hough transform based method.

Experimental results have shown that our skew correction method is significantly faster than the three other methods. Averaging on the 52 test images, our method took approximately 0.038 second to correct the skew distortion of a document image, while the other three methods took 5.42, 11.28, and 18.37 seconds, respectively. The high speed of the proposed method should be mainly ascribed to the efficient extraction of image features for registration. Apparently, detecting IWR from images is much faster than Hough transform, connected component labeling and projection on multiple angles.

Although the image feature used seemed simple, our skew estimation technique were proven by the experiments to be quite accurate. Table 3.1 shows the actual and estimated skew angles of a few typical test documents. For comparison, the estimated results of the other three methods are also listed. As it shows, the proposed method outperformed other estimation methods by providing estimated skew angles which are most closed to the actual skew angles in most cases. In particular, Baird's projection profile based method is less accurate in most cases.

Actual Angle	Le's	Lu's	Baird's	Ours
45	44.9714	45.3004	44.0951	44.9766
35	34.9429	34.9429	35.4239	34.9641
25	24.9286	24.8053	24.5243	24.9588
15	14.9429	14.8145	14.6435	14.9825
5	4.9857	5.0737	5.3492	4.9828
2	1.9714	1.8757	2.5402	1.9874
-2	-1.9687	-2.1582	-2.6089	-1.9718
-5	-4.9870	-4.8944	-5.4783	-4.9774
-15	-14.9379	-14.8278	-14.6531	-14.9612
-25	-24.9306	-24.6069	-25.4390	-24.9832
-35	-34.9348	-34.8252	-34.3365	-34.9729
-45	-44.9489	-44.6196	-44.1889	-44.9681

Table 3.1: Experimental results of the proposed method to document skew estimation

The proposed method is further examined on the orientation detection of degraded document images. In our experiment, the orientations of 47 of the 52 test documents were correctly detected. When examining the five images which failed our proposed method, we found they all contained too little texts. In this case, statistics of character ascenders and descenders may not be captured properly, therefore, the estimation of orientation may fail. As a comparison, most reported skew estimation methods assume that the skew angle lies within a reasonable range, therefore, few of them are able to deal with documents that are captured upside-down.

In addition, our proposed method can be easily extended to deal with documents that are printed in languages other than English. We have tested the method with 12 documents that were printed in Chinese and achieved promising results. The orientation of these documents, however, is almost impossible to be detected as no character ascenders and descenders exist in this language.

3.5 Conclusion

In this chapter, we have defined a new image feature, called interline white run, to estimate the skew angle of a degraded document image by registering it with an imaginary image of the document posed in exactly upright orientation. The key advantage of this new feature is its simplicity in terms of both computational complexity ($O(N)$) and implementation. On the one hand, interline white runs can be easily detected from white run histograms which are obtained by scanning the document image once. On the other hand, this new feature has shown robust and accurate performance on estimating the skew angles of real-world document images. In particular, interline white runs can also be used to detect the document's orientation.

The proposed method can be extended to deal with document images with multiple skew angles. To achieve this, we need to explore more sophisticated clustering methods to extract multiple interline white runs from the white run histograms. Alteratively, we need a segmentation procedure to separate text blocks with different skew angles before constructing the white run histograms. As we have discussed, the proposed method can also be applied on documents printed in other languages like Chinese or Korean. To make the method work better with documents in these scripts, a possible way is to blur the text lines on these documents before scanning them.

CHAPTER 4

Pairwise Registration of Historical Documents

In the previous chapter, we presented a method to register a skewed document image to an imaginary image of itself. Although the registration method is a very special case because of the extremely simple transformation model and the direct computing of the transformation parameter, strictly speaking, the method is still feature-based. As we have discussed in Section 2.1, feature-based image registration methods rely on the precise extraction of image features. So a key factor in the success of the method is the easy and accurate detection of the new feature we proposed.

In this chapter, we are also working on document images. The aim is to register the two side images of a double-sided historical document. As shown in Section 3.1, historical documents are valuable and important but usually degraded by various noises and distortions. Especially due to the long-term preservation and aging, the ink on one side of the document seeped through the paper and mixed with the content on the other side. So in this work, we actually attempt to

match the foreground text on one side of a document with the interfering strokes caused by the foreground text itself. Compared with the image registration task of aligning two images (one is imaginary image) with exactly the same scene but being degraded by a global rotation distortion in Chapter 3, the registration task in this work is different in three ways. First, the two images to be registered in this work have significantly different intensities. Second, it is difficult to extract sophisticated image features from the two input images (historical document images). Third, the transformation model between the two sides of a historical document is much more complicated and non-rigid local transformation is required. We elaborate on these three points in Section 4.2. For these reasons, we present a two-parts intensity-based method to register the two sides of a historical document. The method combines rigid coarse registration (in Section 4.4) and non-rigid fine registration (in Section 4.5). The rigid registration method extracts a pair of representative sub-images from the two images and estimates an Euclidean transformation model with them. In contrast, the non-rigid registration method extracts salient points from the two images, matches them and estimates a transformation which combines affine and spline-based transformations.

4.1 Bleed-through Distortion

As discussed in Section 3.1, document images are usually subject to various photometric distortions and geometric distortions including background noises, translation/rotation displacement and warping effects. In particular, for a double-sided historical document, the foreground contents on one side are additionally impaired by the showing of the contents on the other side of the document. This type of image distortion is usually called bleed-through distortion or show-through distor-

tion, as shown in Figure 4.1. Furthermore, if the document is printed, the term of show-through distortion is used. Otherwise, it is called bleed-through distortion if the document is handwritten. Show-through distortion and bleed-through distortion are mainly caused by the transparency and thinness of the carrying medium and the seeping of the ink that was used to print or write. Moist environment and long-term preservation also contribute to the form of these distortions.

These distortions annoy human users by making it difficult to understand these documents. Under extreme situations as in Figure 4.1(c), they even make the document completely unreadable. More importantly, bleed-through distortions severely impair subsequent document analysis tasks like text extraction or content recognition. Take Figure 4.1(a) for example, the foreground texts are even lighter than the bleed-through interferences that were caused by the texts on the other side. This phenomenon could probably fail the ensuing binarization procedure which may mistakenly treat bleed-through interferences as foreground texts. Therefore, as discussed before, bleed-through distortions should be corrected with all means that are available.

4.2 Historical Document Restoration

In the literature, many approaches have been proposed to remove bleed-through distortions from historical documents. At first, thresholding-like methods [FK01, Don01, LVPG02] were presented to remove all background noise including bleed-through strokes on the document images. This type of methods are quite effective when the gray intensities of the bleed-through strokes (caused by the texts on the other side) are significantly different from those of the foreground texts. Especially, when the gray intensities of bleed-through interferences are lower than those of
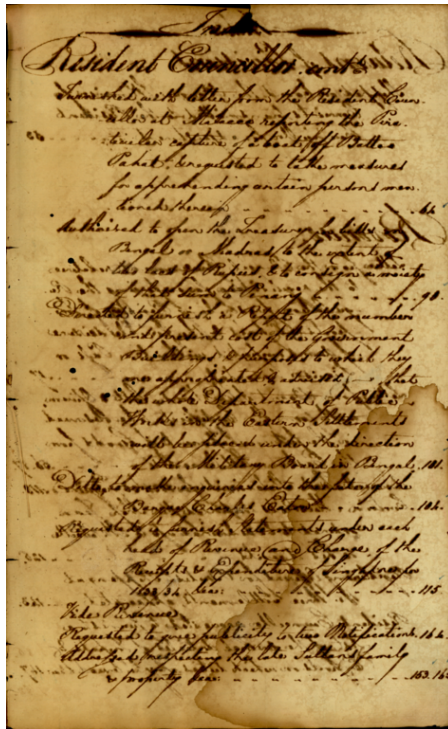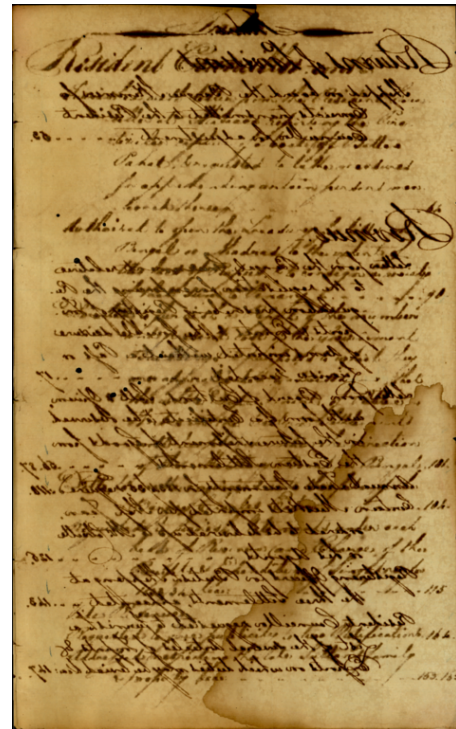
(a)   (b)
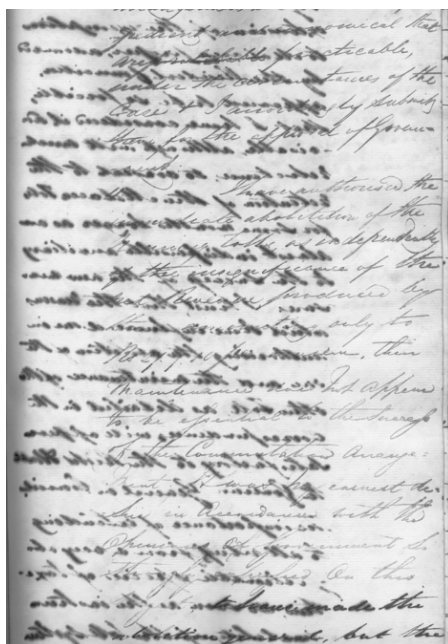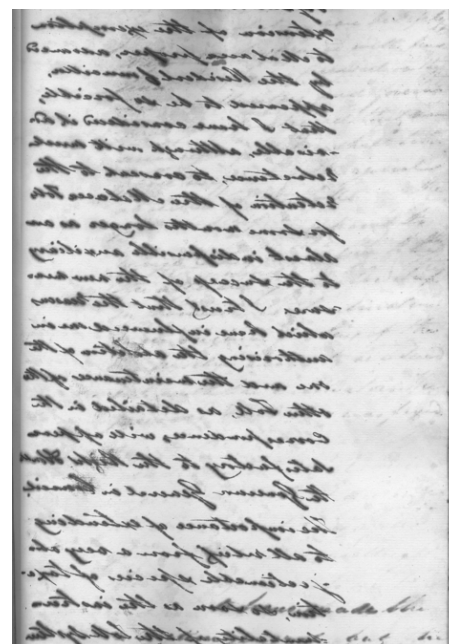
(c)   (d)

Figure 4.1: Sample document images that are impaired by bleed-through distortions. Image (a) is the recto side of document 1; Image (b) is the flipped verso side of document 1; Image (c) is the recto side of document 2; Image (d) is the flipped verso side of document 2

the foreground texts, these method can achieve their best performance. On the contrary however, when the intensities of bleed-through strokes and foreground texts are similar, these methods tend to either remove many foreground texts or leave too many bleed-through strokes.

Considering the fact that bleed-through strokes on one side were caused by the foreground texts on the reverse side of a document, sophisticated separation methods like directional wavelet decomposition [TCS02], ICA (Independent Component Analysis) [TBS04, TSB07] were then proposed to segment foreground texts from bleed-through strokes. Since more information is employed in the classification procedure, this class of methods are believed to be much more accurate. Most importantly, they are more robust to work on both slightly degraded documents and severely degraded documents. The bottleneck of these methods however is that the two sides of a document need to be perfectly aligned before any classification based restoration procedure.

Perfectly registering the two sides of a document is difficult for five reasons. From the very beginning, we should be clearly aware that the matching components in this image registration application is the foreground text on one side and the corresponding bleed-through shown on the other side. Then we elaborate on these five reasons as follows. First, unlike other image registration applications such as panoramic image creation, image features used for image registration are difficult to be accurately extracted from historical documents. In general, both the foreground texts which cause bleed-through and the caused bleed-through interferences are blur, which makes it difficult to precisely locate image features such as corners, boundaries. Second, the gray intensities of matching components are generically highly different, which also raises the requirement for image similarity.

Third, the high frequency of missing matchable counterparts can even fail the image registration procedure. Depending on the practical condition of each document, the total number of and the severity of the bleed-through interferences are various. Fourth, the transformation model between the two images to be registered is unknown and complicated. On the one hand, due to long-term preservation, especially in moist environment, these documents variously stretched/shrank and formed uneven surfaces, which resulted non-linear local deformation on captured images. On the other hand, most historical documents are too fragile to be pressed flattened on scanning planes, which causes geometrical distortions like warping effects, translation/rotation displacements. Fifth, although pixel-level registration accuracy is required, there are no proper ways to evaluate the performance of the proposed registration method. Given a distorted document, there is no such ground truth as the deformation pattern and parameters of the document. Nor could we determine which registered document is the best.

For the reasons discussed above, currently the registration task for historical document restoration is mostly done manually. In the image processing tool box of Matlab, there is a function routine called 'cpselect' which provides a GUI for users to select control points from a pair of images. Then by providing the type of a transformation model, another routine can estimate the parameters of the selected transformation model. The first problem with this semi-automatic strategy is that in most cases, the type of transformation model is unknown. What's more important is that there are only two rigid transformation models available right now.

To fill in the gap between historical document capturing and historical document restoration, a few studies on automating this registration procedure have been made [WT01]. However, a common problem remains as that the accura-

cies of these methods are unsatisfactory due to the rigid transformation models they used. Therefore, in this chapter, we present a unified framework which combines rigid coarse registration and non-rigid fine registration methods for historical manuscripts to be effectively restored without human intervention.

4.3   Framework Overview

At the beginning, we manually pair the two side images of a document and roughly align them. Based on the layout of the document, we vertically or horizontally flip one of the images to get its mirror image. Without losing generality, we refer the flipped image as target image (verso image) and the intact side one as reference image (recto image). The two images first go through a rigid coarse registration procedure as presented in Section 4.4. In this procedure, a similarity transformation is estimated so that the global translation and rotation deformations between the two images are corrected. In many cases, this coarse registration is sufficient for subsequent bleed-through correction tasks. However, with documents that are severely distorted by local deformations such as uneven surfaces or warping effects, a fine registration as presented in Section 4.5 is required.

For the fine registration of these documents, we first extract certain number of salient control points from the two images and establish the correspondences between them by minimizing a combined distance measure (as shown in Section 4.5.1). These matched point pairs are then used to estimate a free-form transformation model that is based on B-splines. During this procedure, we use residual complexity (RC) as similarity measure and regular the estimated transform with smoothness constrain. Once the two images of a historical document have been precisely registered, we apply a wavelet-based method (as in Section 4.6) to cor-
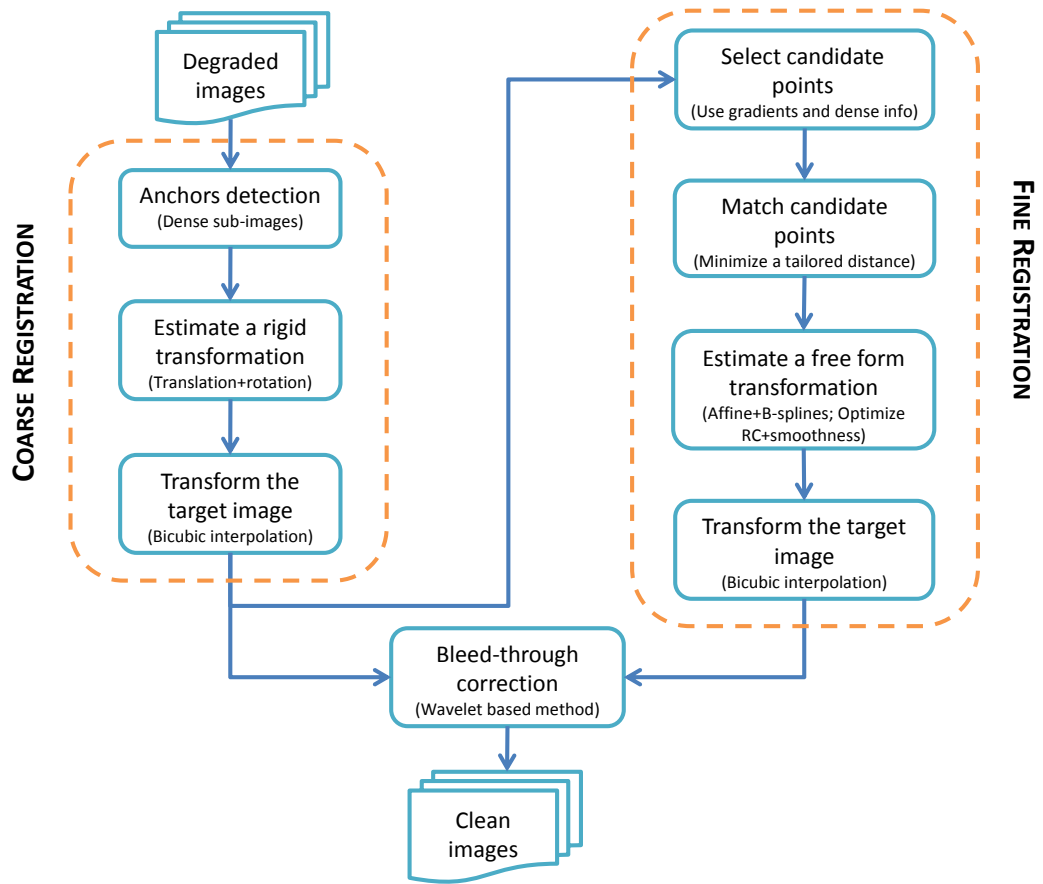
Figure 4.2: The built framework for historical document image restoration.

rect the bleed-through distortions on the two images. Figure 4.2 shows the general structure of the proposed historical document restoration framework.

4.4   Rigid Coarse Registration

As we have discussed in the previous section, one major problem that arises when aligning the two sides of an old handwritten document is the varying intensities of the interfering strokes relative to the corresponding original strokes. In certain cases, the interfering strokes have very similar intensity to the original strokes. While in other cases, interfering strokes are very faint or can hardly be found because their original strokes slightly seeped or did not seep through the page. As the alignment is actually between the interfering strokes and the corresponding original strokes, weak or lacking of interfering strokes will significantly impair the alignment accuracy, especially for documents without dominant text areas.

To overcome this problem, at this coarse registration stage, we use a pair of extracted anchors instead of the two whole images to globally align the document. The anchors are defined as a pair of sub-images that are located at the same position of the recto image and the verso image respectively. These two sub-images are expected to capture all and only capture the global transformation between the two images of a document, such as translation distortion and rotation deformation. In our approach, we chose the anchors in an intensively overlapping region, where the observed "texts" on both sides of the document have the highest intensity. The high intensity is generally caused by the combination of foreground strokes and bleed-through strokes. As the interfering strokes on both sides are intensive, the extracted anchors capture more information on the correspondence between the original strokes and their bleeding strokes. Moreover, since only a pair of small

(a) recto   (b) verso
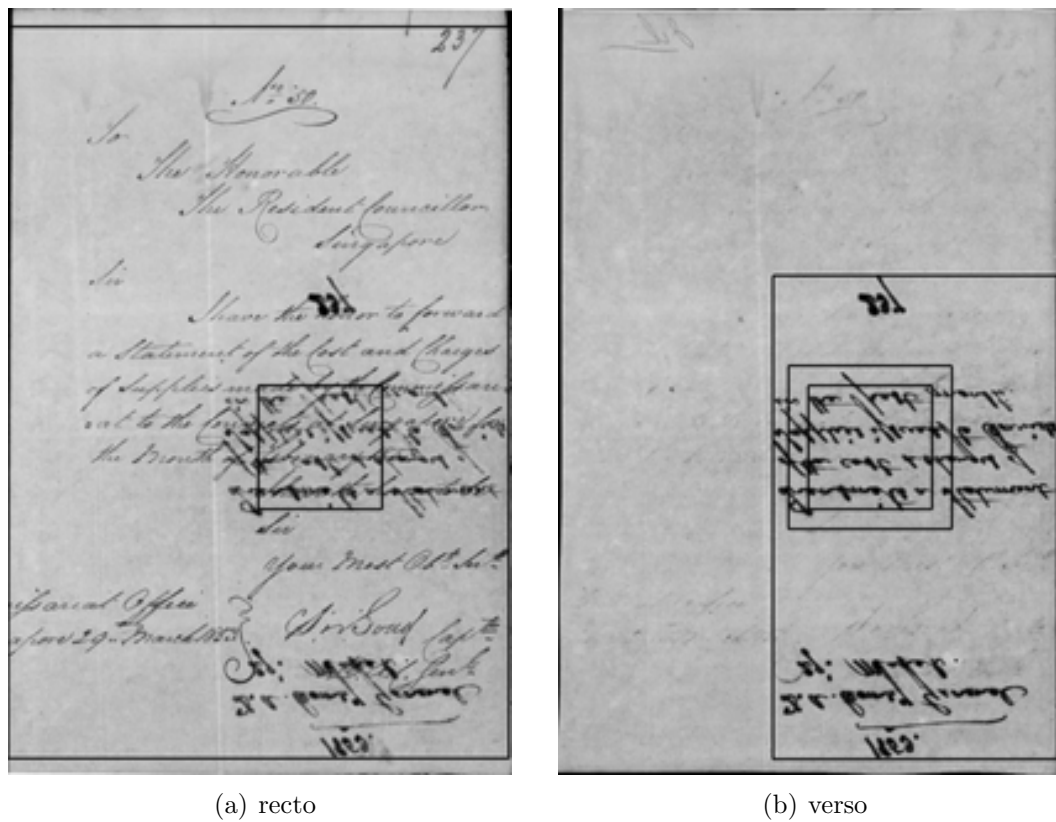
Figure 4.3: Illustration of the extracted main text areas, the intensively overlapping regions and the search window on the verso image.

sub-images are involved in computation, this strategy can significantly reduce the processing time and memory consumed by considerably large images.

To extract the anchors from the recto image and the verso image, we need to identify the intensively overlapping region. For this purpose, the main text areas on the two images are detected using a global thresholding technique. The size of this overlapping region is automatically adjusted based on the minimum size of main text areas on the two images and the average intensities of the two images, as shown in Equation 4.1:

$$S_{over} \propto \min\{S_{r_{main}}, S_{v_{main}}\} \tag{4.1}$$

with constraints $I_{r_o} > \alpha I_r$ and $I_{v_o} > \alpha I_v$, where I_{r_o} and I_{v_o} are the average identities of the overlapping region on the recto image and the verso image respectively. α is a constant and is set to 1.2 in our experiments.

By computing the average intensities of all sub-images with size equal to S_{over} and globally searching on the two images to be registered, an optimal solution that satisfies the above requirements can be achieved. This procedure, however, is extremely computationally expensive even dynamic programming techniques are applied to reduce the computational load. Therefore, in our method, we search for a local optimal solution. The detailed procedure to search for this overlapping region can be described as follows:

- Step 0: set the target images to be the two original images and the grid size to be 200*200;

- Step 1: divide the target images into grids;

- Step 2: search for the grids with the highest intensities on the two target

images;

- Step 3: merge all neighboring grids of the most intensive grid to form new target images;

- Step 4: decrease the grid size and go to step 1;

- Step 5: iterate the above steps until the current grid size reaches the specified value;

It is noticed that when selecting the overlapping region, boundary areas should be avoided, because warping distortion due to binding effect usually appear in these areas and this local warping deformation should not be interpolated to the whole image.
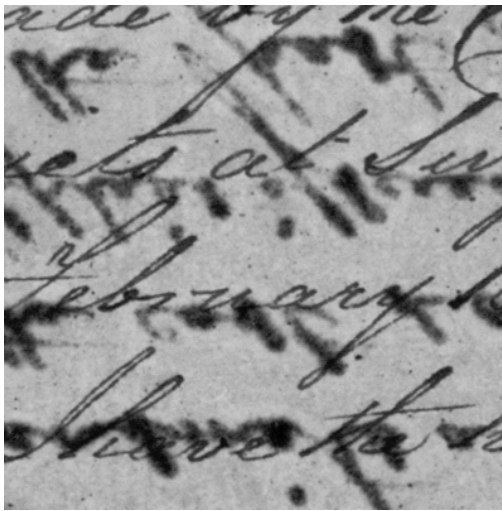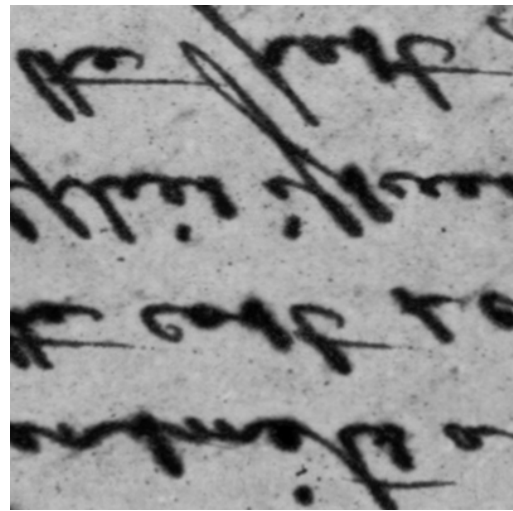
(a) recto   (b) verso

Figure 4.4: A pair of sub-images that have been extracted from the recto image and the verso image of the document shown in Figure 4.3

Once such a region is found, the sub-image that centers at this location on the recto image is chosen as one of the anchors, which is shown in Figure 4.3(a). Then we search for the corresponding sub-image on the verso image of the same

document. The searching procedure is performed within a window with a larger size than the recto anchor, as shown in Figure 4.3(b). At this coarse registration stage, we only focus on the global translation and rotation deformations between the two images, therefore, only candidate sub-images in the area as shown in Figure 4.5 are examined. The similarities between the recto anchor and the matching candidates on the verso image are measured using block matching strategy. In particular, we adopt the normalized correlation coefficients that are defined as follows for this purpose:

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{r_{ij}} - I_r)(I_{v_{ij}} - I_v)}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{r_{ij}} - I_r)^2} \cdot \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{v_{ij}} - I_v)^2}} \tag{4.2}$$

where I_r, I_v are the average intensities of the recto and verso images respectively. Figure 4.4 shows a pair of extracted anchors. The correspondence between them now can be used to align the recto and verso images.

(a) translation   (b) rotation
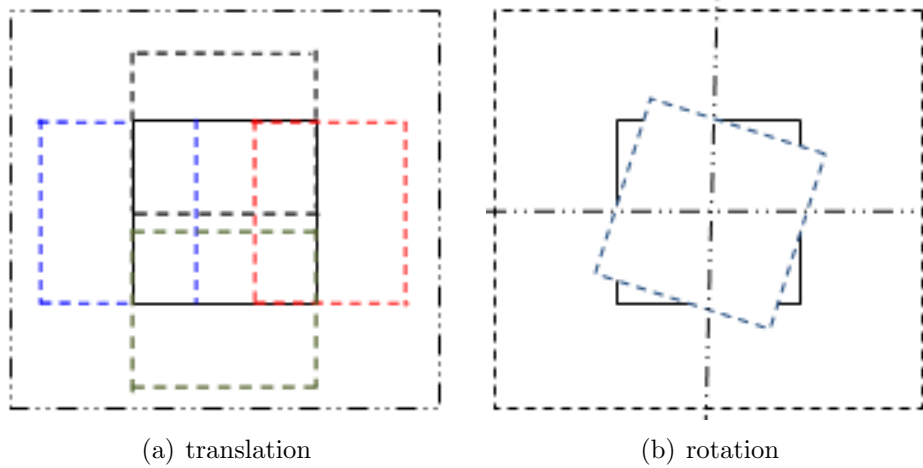
Figure 4.5: Illustration of the search strategies to correct the global translation and rotation deformations on a document image.

To show the effect of this coarse registration procedure, we apply a bleed-through correction technique (as presented in Section 4.6) to the originally un-

aligned document images and the coarsely registered images and compare the results. As we can see from Figure 4.6, this coarse registration procedure significantly improves the performance of the subsequent bleed-through correction procedure. However, as also shown in the figure, many bleed-through strokes still remain. In particular, in some areas, such as the circled regions in Figure 4.6(b), 4.6(d), originally only bleed-through strokes are present in these regions and the intensities of these interfering strokes are very similar to those of the corresponding foreground strokes originating from the reverse side of the paper. Using our bleed-through correction method, these bleed-through interferences should be fully removed, if the two sides of the document are precisely aligned. Therefore, finer registration of the recto and verso image is essential to fully remove bleed-through interferences.

4.5 Non-rigid Fine Registration

As we have discussed in Section 4.4, some historical documents that were degraded by severe local deformations need a finer registration method so that the present bleed-through distortions can be fully removed. These local deformations are usually caused by stains (water, coffee or other liquids), moist environment and binding effects. The first two factors make a historical document locally uneven through long-term preservation in the archives. Meanwhile, due to aging, these documents become too fragile or these documents were bound in too many pages so that they can not be pressed flat towards the scanning plane, which also causes local warp in the captured images. These local deformations are non-rigid and can not be estimated with the coarse registration method described in the previous section. Therefore, in this section, we present a non-rigid fine registration method that is tailored to severely degraded document images.

(a) no alignment

(b) coarse alignment

(c) no alignment

(d) coarse alignment
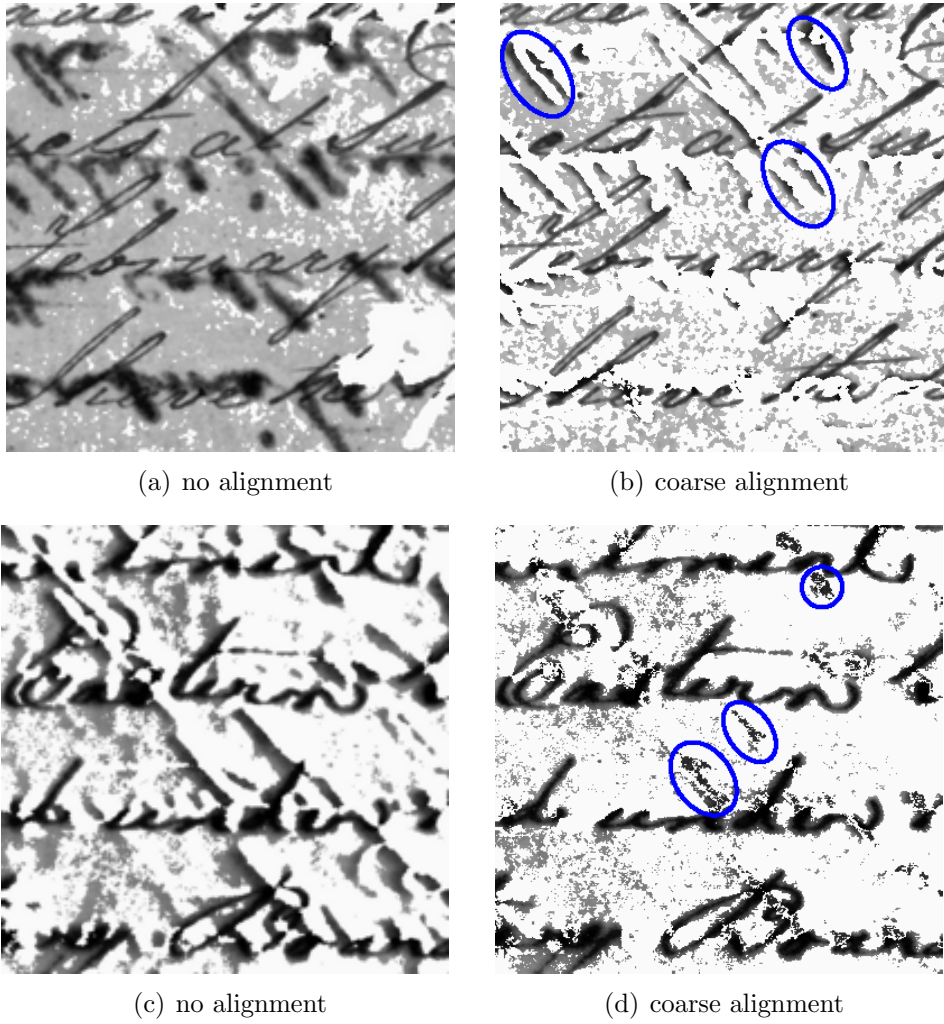
Figure 4.6: Resultant images after applying bleed-through removal technique on the originally unaligned images and the coarsely aligned images. Images (a-b) are for sample image 1; Images (c-d) are for sample image 2

4.5.1 Control Point Selection

As we have mentioned before, the components to be registered in this application are the foreground texts on one side of a document and the bleed-through interferences that are present on the other side of the document. As discussed in the previous section, depending on the conditions of the physical documents, the intensity patterns of the showing bleed-through interferences are highly different. On the one hand, both the original foreground strokes which cause bleed-through interferences on the other side and the corresponding bleed-through strokes are blur. On the other hand, as these documents are handwritten, characters, words or even text lines are often connected, which makes automatic component detection or object labeling almost impossible. Therefore sophisticated image features that can be used for image registration such as corners, object boundaries are difficult to be precisely extracted from these historical document images. Meanwhile, as we have said historical documents are often affected by unevenness effect which is caused by moist environment during long-term preservation. This unevenness effect usually results local deformations between the two side images of a document. In order to address these local deformations, pixel-level image registration accuracy is required in this application.

Therefore, as complex image features are difficult to be located in historical document images and as pixel-level registration accuracy is required, in this work, we use corresponding control points to perform the registration task. According to our experience on manually selecting control points from historical document images, points which contain pure foreground text or pure bleed-through interference are much easier to be accurately detected. Such points are also easier to be perfectly matched due to their simple local structures. Meanwhile, due to people's writing pattern in English, foreground strokes that slant at particular angles are
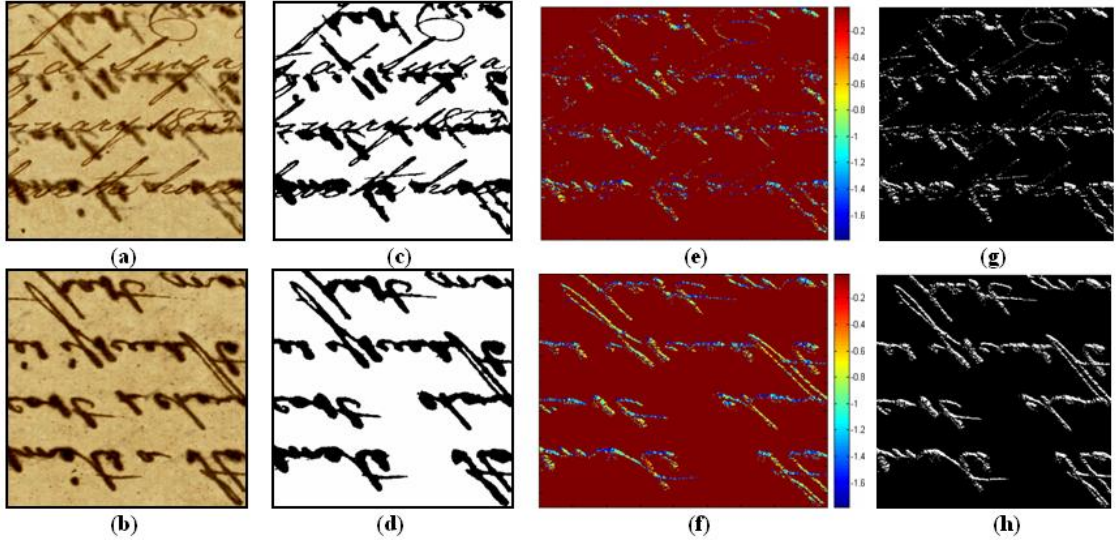
Figure 4.7: Illustration of the procedure to detect control points from the two images of a document. Images (a-b) are the two side images of a document; Images (c-d) are the binary versions of the two side images; Images (e-f) are the gradient direction maps of the two images; Images (g-h) show the candidate control points that have been identified from the two images.

usually much stronger than other foreground strokes and thus have more chances to cause bleed-through distortions on the other side of the page. Therefore, points that satisfy these two conditions are in high probability of having a matching component and are easier to be accurately located. Based on these heuristic knowledge, we select corresponding control points from the two images to be registered with the algorithm described as follows:

- Identify candidate control points that satisfy the following conditions from the two images:

$$
\begin{cases}
\theta_{xy} \otimes (\theta_{xy} > \theta_l) \otimes (\theta_{xy} < \theta_h) \otimes B_{xy} \neq 0 \\
M_{xy} \otimes (\theta_{xy} > \theta_l) \otimes (\theta_{xy} < \theta_h) \otimes B_{xy} \neq 0 \\
I_{xy} \otimes (\theta_{xy} > \theta_l) \otimes (\theta_{xy} < \theta_h) \otimes B_{xy} \neq 0
\end{cases}
\qquad (4.3)
$$

where θ_{xy}, M_{xy}, I_{xy} and B_{xy} represent the gradient direction, gradient magnitude, dense intensity and binarized value of point (x, y). To compute B_{xy}, we binarize the images with an adaptive binarization procedure [Ots79]. Parameters θ_l and θ_h are thresholds to filter the gradient direction maps and $\otimes$ represents Hadamard product. In our experiments, θ_l and θ_h have been set as -1.8 and -0.5 (approximately $120°$ and $150°$). According to our experience and experiments, points satisfying these conditions are most possible to cause strong bleed-through interference. Therefore, in this way, we make sure that most selected control points on the reference image have a corresponding point on the target image. Figure 4.7 illustrates this feature selection procedure and shows some detected candidate points.

- Establish the correspondences between the selected points on the reference image and the selected points on the target image by minimizing a distance measure that combines gray intensity I, gradient magnitude M, gradient direction θ and displacements $x - x'$, $y - y'$:

$$
\begin{aligned}
Dis = w_i(I_{xy} - I'_{x'y'}) + w_m(M_{xy} - M'_{x'y'}) \\
+ w_\theta(\theta_{xy} - \theta'_{x'y'}) + w_d\sqrt{(x - x')^2 + (y - y')^2}
\end{aligned}
\tag{4.4}
$$

where w_i, w_m, w_θ, w_d are the weights that specify the relative importance of gray intensity, gradient magnitude, gradient direction and displacement when determining the correspondences between points. In our experiments, we have empirically chosen [1, 3, 10, 20] corresponding to $[w_i, w_m, w_\theta, w_d]$.

- Check the displacements of each point pair block by block to correct non-collectively occurring mismatches. As the established correspondences are supposed to be locally continuous, point pairs that have a displacement larger

than the median displacement of the local block are regarded as mismatches and discarded. To achieve this, we group the detected points into 8×8 blocks and compute the median displacements of each block as:

$$\begin{cases} dx_B = \mathbf{Median}(x_i - x'_i) \\ dy_B = \mathbf{Median}(y_i - y'_i) \end{cases} \tag{4.5}$$

where $i = 1 \cdots n$. n is the number of control points in each block. The detected point pairs with displacements larger than $2dx_B$ or $2dy_B$ are removed as mismatches.

- Switch the roles of recto image and verso image to conduct consistency checking. This step is useful for documents with severe bleed-through distortions on both sides and could further refine detected control point pairs.

The number of the detected control points can be controlled by thresholding on Dis or directly set. Figure 4.8 shows some of the detected corresponding control points with small distances. It is noticed that the control points that have been extracted for now are randomly distributed. However the subsequent cost function estimation procedure is based on uniform control point meshes. Therefore, we employ a cubic interpolation method to construct control point meshes with the detected control points:

$$P(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j \tag{4.6}$$

where a_{ij} are 16 coefficients which are determined by the values and derivatives of the functions at the four corners of the square centering at point (x, y).
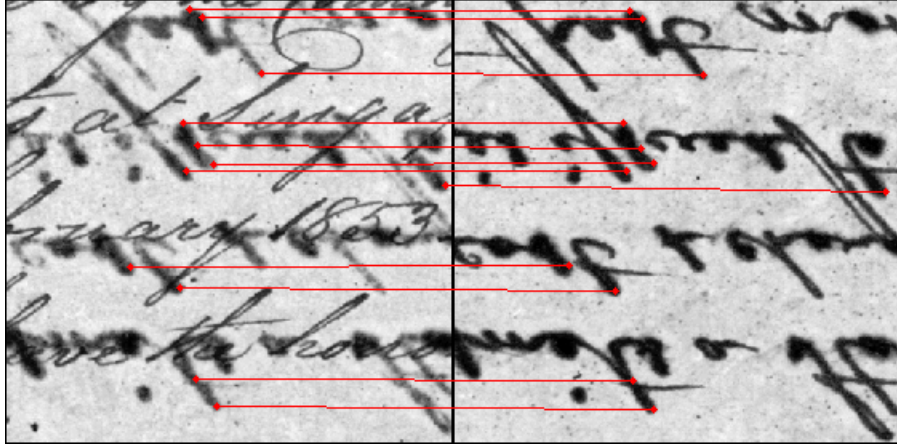
Figure 4.8: Illustration of the matched control point pairs.

4.5.2 Free-form Mapping Function

We model the spatial relationship between the two images to be registered with a free-form transformation model which is based on a 3 hierarchical levels of B-spline [RSH+99]. The transformation model consists of two components which are sequently optimized in a hierarchical structure (coarse to fine).

$$T(x, y) = T_{global}(x, y) + T_{local}(x, y) \tag{4.7}$$

T_{global} is an affine transformation (as shown in Equation 5.6) which is supposed to capture the global rotation, translation and non-isotropic scaling deformations between the two images.

$$T_{global}(x, y) = \begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{4.8}$$

T_{local} is defined as the tensor product of the cubic B-splines (as shown in Equation 5.7), which is expected to describe the local deformations between the

two images. Such a transformation model is chosen as it is flexible enough to capture the local deformations as much as possible and meanwhile it is capable of producing a locally smooth and continuous transformation. Apart from these factors, B-splines are locally controlled, which makes them computationally efficient for a large number of control points.

$$T_{local}(x, y) = \sum_{m=0}^{3} \sum_{n=0}^{3} B_n(u) B_m(v) \phi_{i+n, j+m} \tag{4.9}$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$ and $[n_x, n_y]$ is the size of the control mesh. ϕ denotes the mesh of control points and B_i refers to the ith basis function as shown below.

$$\begin{cases} B_0(u) = (1-u)^3/6 \\ B_1(u) = (3u^3 - 6u^2 + 4)/6 \\ B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) = u^3/6 \end{cases} \tag{4.10}$$

4.5.3 Cost Function Optimization

In order to estimate the parameters in the transformation model that has been defined in Section 4.5.2, we define a cost function that also consists of two components.

$$C = C_{sim}(I_{xy}, T(I_{xy})) + \lambda C_{smo}(T) \tag{4.11}$$

where C_{sim} refers to the similarity of the registered images, which ensures an accurate transformation model. C_{smo} is a regularization term to make sure the estimated transformation model is smooth. λ weights the relative importance of the two constrains and was set to 0.02 in our experiments.

The similarity measure used in this work is residual complexity which is the compression complexity of the residual images between the two registered images [MS09]. This similarity measure is used as the intensities of the matching components (foreground strokes vs bleed-through interferences as we have discussed before) in this application are significantly different, dependent and non-stationarily distorted. To be specific, residual complexity can be summarized as follows:

$$
\begin{aligned}
C_{sim} &= \sum_{n=1}^{N} log((\mathbf{q}_n^T \mathbf{r})^2/\alpha + 1) \\
\mathbf{q}_n &= dctn(\mathbf{r}) \\
\mathbf{r} &= I_{ref} - T(I_{tar})
\end{aligned}
\tag{4.12}
$$

where $\mathbf{r}$ refers to the residual image and $dctn$ is the multidimensional **DCTs**. α is a trade-off parameter and is set to 0.05 in our experiments.

For implementation simplicity and time efficiency, C_{smo} is currently chosen as the space integral of the square of the second order derivatives:

$$
C_{smo} = \frac{1}{s} \int \int [(\frac{\partial^2 T}{\partial x^2})^2 + (\frac{\partial^2 T}{\partial y^2})^2 + 2(\frac{\partial^2 T}{\partial xy})^2] dx dy
\tag{4.13}
$$

where s denotes the area of the image domain and T represents the estimated transformation. For the purpose of function optimization, gradient descent method is adopted and the estimated transformation model is iteratively refined with increasing size of control points. In our implementation, the optimization procedure terminates when the relative function difference is larger than a threshold or the maximum number of iterations has been reached. We avoid finding some local minima by reducing the optimization step size and slightly increasing the value of the objective function. Figure 5.8 shows an example of the fine registration result.
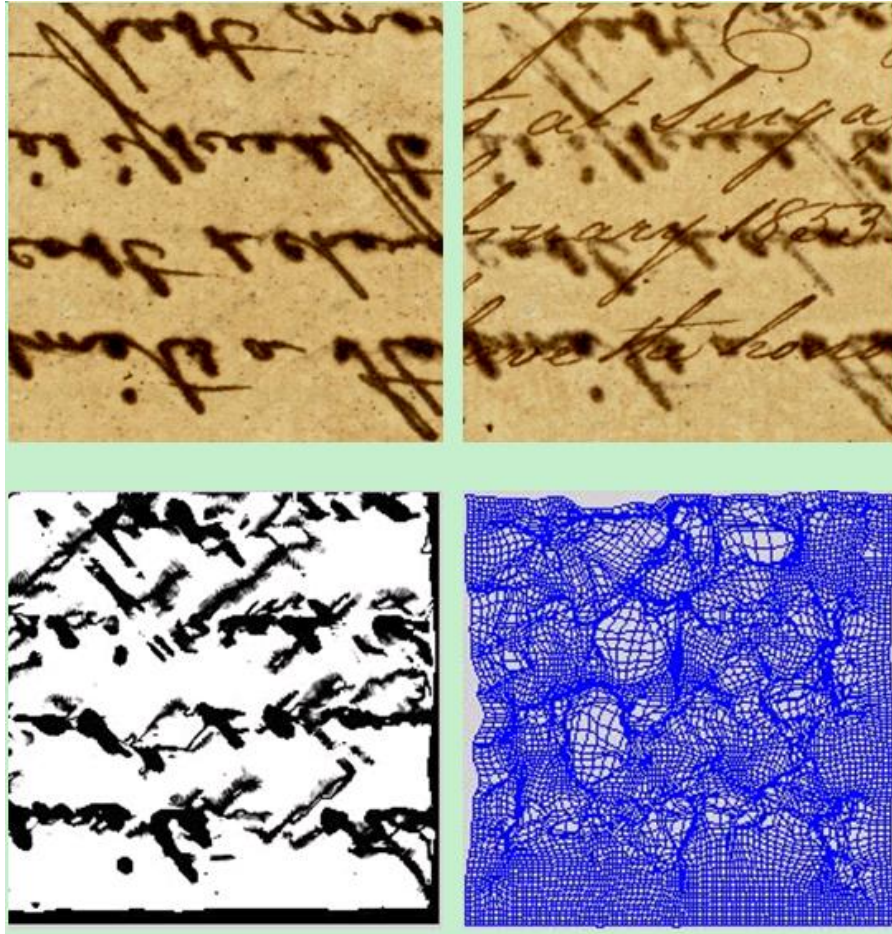
Figure 4.9: Illustration of the fine registration procedure. The images from top to bottom and left to right are: the reference image (the one to be registered to), the target image (the one to be registered), the registered target image and the estimated transformation map

4.6   Ink Bleed-through Correction

Once the two images of a document have been registered, we employ the wavelet-based method as proposed in [TCS02] to correct the bleed-through distortions on the document. In general, this method suppresses bleed-through distortions and enhances foreground texts by modifying the wavelet detail coefficients of the document images. The detailed steps for this approach is summarized as follows:

- Compute the foreground overlay image $a(x, y)$ for the recto image $f(x, y)$ of

a document as:

$$a(x, y) = flip(invert(b(x, y))) + f(x, y) \tag{4.14}$$

where $b(x, y)$ refers to the registered verso image of the same document. This step actually conducts a subtraction between the two images, which coarsely cancel some bleed-through interferences.

- Further weaken the suspected bleed-through interferences on the foreground overlay image $a(x, y)$ by scaling it with a nonlinear transformation as follows:

$$curve(x) = 2^k - 1 - ((2^k - 1)^2 - x^2)^{0.5} \tag{4.15}$$

This step is supposed to increase the density differences between the foreground texts and the bleed-through distortions, which means the foreground texts get enhanced and the bleed-through distortions get suppressed.

- Detect the foreground strokes on the foreground overlay image $a(x, y)$ using a modified canny edge detector that is regularized with orientation filters and other constraints [CTWS00, TCS$^+$00] to form the "enhancement feature image", $E(x, y)$.

- Switch the roles of $f(x, y)$ and $b(x, y)$ and conduct the above steps on the registered verso image $b(x, y)$ to gain the "smearing feature image", $S(x, y)$.

- Decompose an original image ($f(x, y)$ or $b(x, y)$) into wavelet domain while retaining the size of the image as:

$$wf(x, y) = \{C_j(x, y), D_j^k(x, y), j = 0, \cdots, k = 1, 2, 3\}$$
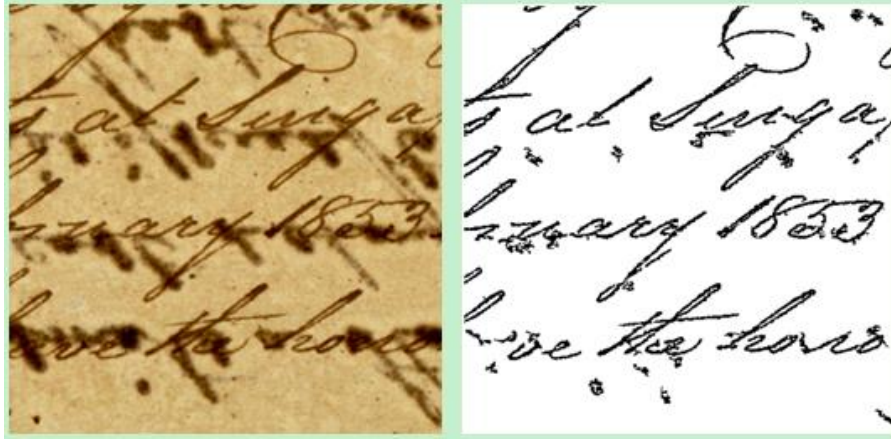
Figure 4.10: A degraded document image (cropped from a larger image) and the resultant image after fine registration and bleed-through correction.

where $C_j(x, y)$ is the wavelet approximation coefficient and $D_j^k(x, y), (k = 1, 2, 3)$ are the wavelet detail coefficients at scale j of the wavelet decomposition.

- Modify the wavelet detail coefficients by referring to $E(x, y)$ and $S(x, y)$:

$$D_j^k(x, y) = \begin{cases} e_j^k D_j^k(x, y) & \text{if } E(x, y) == 1 \\ s_j^k D_j^k(x, y) & \text{if } S(x, y) == 1 \end{cases} \tag{4.16}$$

where $j = 0, \cdots, J; k = 1, 2, 3$.

- Reconstruct the wavelet representation of the original image with the modified wavelet detail coefficients.

- Iteratively repeat the decomposition and reconstruction procedure at most 15 times to get the restored images with foreground texts enhanced and bleed-through distortion reduced.

- Apply the same edge detector to the enhanced images to obtain the final output images. Figure 4.10 shows a document image that is degraded by severe bleed-through distortions and the restored image after fine registration

and bleed-through correction. As we can see from the figure, although some foreground strokes are mistakenly broke, the restored image is much better for human perception than the original one. With certain recovering methods that can connect the broke foreground strokes, the quality of the resultant images can be further improved.

4.7 Experiments and Results

We have tested the proposed framework with 28 double-sided historical documents (56 images) from the national archives and the experimental results are encouraging. Our experiments were conducted on a Pentium 4GHZ CPU with MATLAB. Typically, it takes 5 minutes to process a pair of images scanned at 150dpi and with a size of 1800×2800. This speed is acceptable since most document restoration tasks are conducted off-line and with more powerful machines and parallel strategies, this time cost can be further reduced. To assess the proposed framework, visual assessment was first performed by the experts working at the achieves for qualitative evaluation. In particular, we compared the results from this new framework with those of other three methods [TCS+00, WBT09, WT01]. As shown in figure 4.11, this new framework produced better results for most testing images.

No. of words		47	241	134	79	83	257	209	189	217	145	97	89	Average
Precision	Q Wang [14]	87	57	63	76	82	47	58	76	62	64	71	68	62.4
(%)	J Wang [12]	92	79	84	89	88	73	81	79	78	86	84	90	83.6
	Proposed	97	88	89	95	93	81	87	90	87	84	88	89	89
Recall	Q Wang [14]	91	65	80	82	85	72	69	77	79	68	83	81	77.7
(%)	J Wang [12]	96	76	87	91	90	79	73	87	81	79	91	86	84.7
	Proposed	95	83	93	97	89	86	84	92	87	84	96	94	90

Table 4.1: Quantitative evaluation and comparison of the proposed bleed-through correction method with other methods

For quantitative assessment, we manually labeled 12 documents (4 slightly

(a) Tan [TCS02]

(b) Q Wang [WT01]

(c) J Wang [WBT09]

(d) Proposed
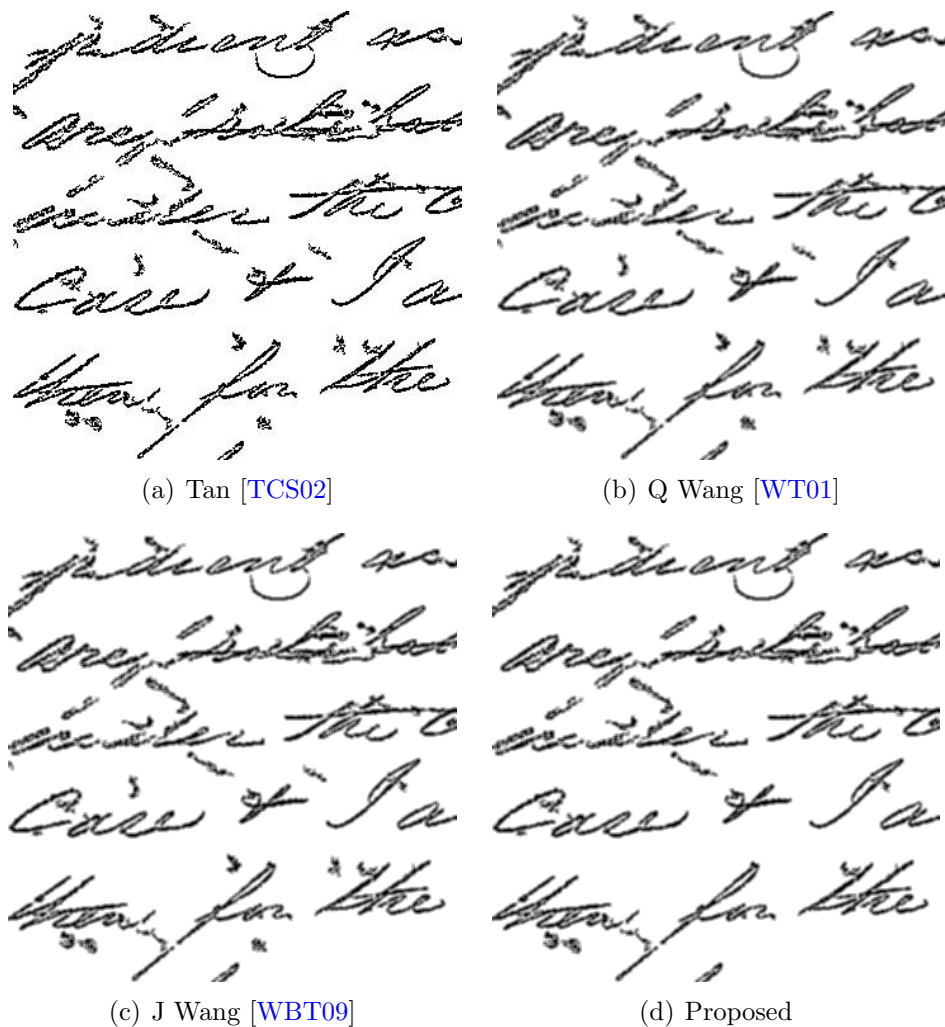
Figure 4.11: The comparison of the resultant images that have been produced by different bleed-through correction methods.

degraded, 4 moderately affected and 4 severely impaired as shown in Figure 4.1) with "ground truth". To be specific, we counted the following numbers: the total number of foreground words N_{fgd} on the image; the number of words detected after applying this restoration method N_{detect}; the number of words correctly detected after restoration $N_{correct}$. When calculating these numbers, connected words were counted separately. N_{detect} included fully detected foreground words and partially or fully detected interfering words. However, only fully detected foreground words were counted in $N_{correct}$. With these numbers, the proposed framework was evaluated with the traditional document analysis metrics as:

$$Precision = \frac{N_{correct}}{N_{detect}}, \quad Recall = \frac{N_{correct}}{N_{fgd}} \tag{4.17}$$

Table 4.1 lists the results of 12 degraded documents from the new framework and the other two methods in [WBT09,WT01]. As we can see, the average restoration precision and recall of the proposed framework are 89 and 90, respectively, which are higher than those of the other methods. Moreover, experiments show that when working on severely degraded manuscripts, the new framework is more robust.

4.8 Conclusion and Discussion

In this chapter, we have presented a unified framework which consists of rigid and non-rigid methods to register the two sides of a historical document so that the bleed-through distortion on the document can be fully corrected. The rigid registration method in the framework can handle historical document images with slight geometric distortions. It selects a sub-image with high intensities on both

<center>(a)   (b)</center>

Figure 4.12: A historical document image that has been impaired by severe bleed-through distortions and background noise and the resultant image that was produced by our restoration framework.
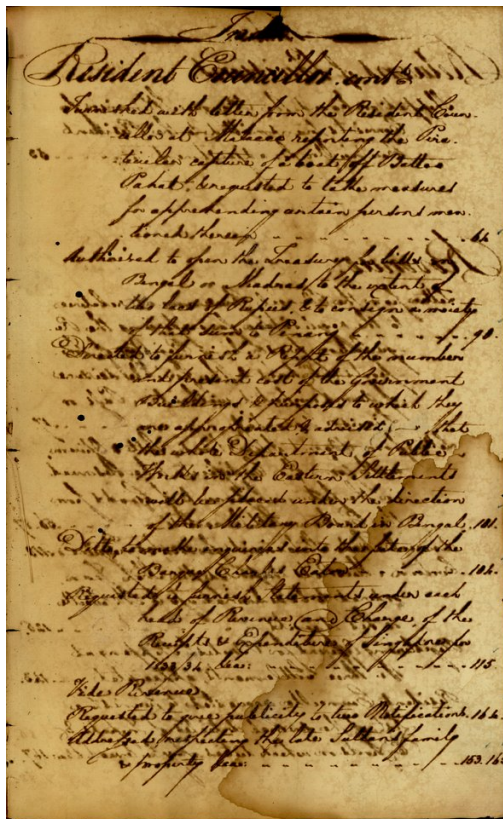
sides and uses it to estimate an affine transformation. For document images with severe geometric distortions, mainly warping effects and local uneven surfaces, the non-rigid image registration method is also applied to further fine the registration result. The non-rigid method makes use of the gradient maps of the document images and people's writing pattern to detect salient control points for registration. To capture the local deformations on severely degraded document images, we employ a free-form transformation model which combines affine and B-splines. As the two images to be registered are of different intensities, residual complexity is used for similarity measure. The estimated transformation model is regularized by the space integral of the square of the second order derivatives to ensure the smoothness of the transformation. We have evaluated the proposed method by applying a wavelet-based bleed-through correction method to the registered images and calculating restoration accuracy and recall.

In this work, we assessed the registration framework by evaluating the resultant images after bleed-through correction. On the one hand, counting those numbers (as in Section 4.7) is time consuming. On the other hand, the final accuracy is contributed by the registration procedure and bleed-through correction procedure. So technically, we didn't actually evaluate the registration framework. As shown in Figure 4.11, although most interfering strokes have been removed from the document, many foreground texts are also mistakenly broken. So we are considering adding a post-processing routine to recover broke foreground strokes after this restoration procedure so that the quality of resultant images could be further improved.

CHAPTER 5

Groupwise Registration of Brain CT Scans

As shown in Section 1.1, image registration is a fundamental task in medical imaging. Nowadays, the commonly used medical tests include Computed Tomography (CT) scans, Positron Emission Tomography (PET) scans, and Magnetic Resonance Imaging (MRI) scans. Each of these tests serves different diagnosis procedure. For instance, head CT scans are suitable for bone fracture detection and tumor identification. In contrast, brain MRI scans are more popular in evaluating and diagnosing the presence of certain diseases as they produce detailed pictures of organs, soft tissues and all other internal body structures. In general, the image registration methods used in medical domain fall into two groups: single-modal and multi-modal ones. A typical application of single-modal image registration methods is to build an atlas with the sample images. For instance, in this work, we attempt to construct a head CT atlas with the scans from different study cases. Multi-modal image registration is to transform scans that were taken by different medical tests to the same coordinate system for comparison and fusion. For instance, we often need to register CT scans to MRI scans to get more structural details.

In this chapter, we develop a groupwise image registration framework which is capable of constructing a brain CT atlas with multiple CT scans from different study cases. The groupwise registration framework is built upon a non-rigid pairwise image registration method which is presented in Section 5.4. To achieve this, we first group CT scans which are normal or have minor abnormalities into different clusters. Within each cluster, we register all the slices in the cluster to the central slice of the cluster and compute an intermediate average slice with all the registered slices. The final atlas is the combination of all these intermediate average slices as shown in Section 5.3. We have demonstrated that the built atlas can be used to index the input slice in the axial direction of the brain (as shown in Section 5.5). The atlas is also useful in detecting the abnormalities on the input slices which are affected by severe traumatic brain injury (presented in Section 5.6).

5.1 Introduction

Due to their importance in the diagnosis of traumatic brain injury (TBI), large amount of computed tomography (CT) scans are stacked in the hospital radiology department. To facilitate convenient access of these CT data, a variety of retrieval systems have been proposed. In particular, content-based image retrieval (CBIR) systems have been widely studied and have demonstrated promising results [LGW+10]. One major advantage of CBIR systems is they provide the retrieval function called 'query by example' by which CT slices that are visually or anatomically similar to the input slice are searched and retrieved. This function is useful for doctors' diagnosing procedure and the training of medical students.

In general, one axial CT scan contains 19 to 24 slices that are captured at different heights along the axial direction of the brain. Figure 5.1 shows the 18 slices

of one study case. Thus the search space of a CBIR system could be huge when extremely large scale data set is involved. Meanwhile, time efficiency is critical for all real-time CBIR systems and computer-assisted diagnosis (CAD) systems. Therefore, reducing the search spaces of CBIR systems is essential. Slice indexing refers to the procedure of determining the heights of the input slices along the axial direction of the brain. It functions as a preprocessing step in CBIR or CAD systems to reduce the search space and regularize retrieval results. Based on these knowledge, it is known that the performance of slice indexing will significantly affect the overall efficiency and precision of the whole retrieval system.
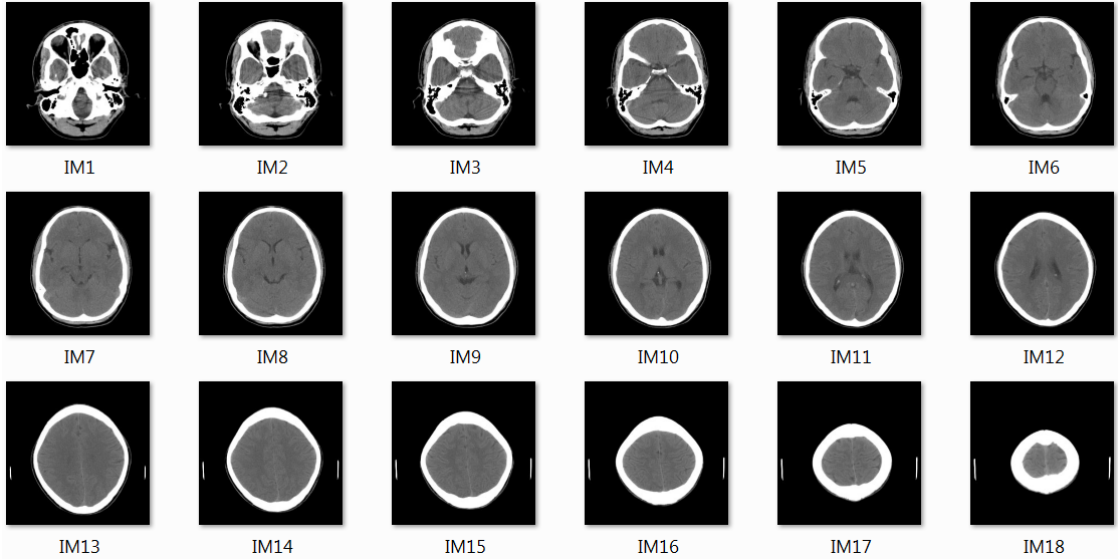
Figure 5.1: The 18 brain CT slices of a real-world study case. The numbers below the images indicate their heights in the axial direction of the brain. The number increases as the height that the slice was taken increases.

When reviewing the existing methods for slice indexing, we found two lines of approaches that can be represented by [LLP$^+$10, GKN08]. The first group of methods attempt to register 2-D slices to a 3-D atlas with planar-to-curved surface alignment method and PDE-based registration technique [GKN08]. The major problem with this approach is the time efficiency, which makes it difficult to be

applied in real-time systems. The other group of methods extract anatomical features such as basal cistern, frontal horn and cerebrospinal fluid (CSF) from the input slices. Based on these features, the input slices are grouped into predefined levels with classifiers like SVM [LLP+10]. Compared against methods in the first group, this approach is relatively faster, however, segmenting these anatomical features from the image is still time consuming. More importantly, the robustness of this approach is questionable when the input slices are significantly distorted due to severe TBI. In this scenario, distorted anatomical features either are difficult to be precisely extracted or severely impair the classification results.

In this chapter, we build a brain CT atlas by registering multiple CT scans which are normal or with minor abnormalities. As we have discussed before, registering multiple images requires a strategy for groupwise registration. In this work, we propose a cluster-structured groupwise registration method which is built on free-form pairwise image registration. First, CT slices from normal study cases are normalized with method presented in section 5.2 to correct pose difference. Normalized slices which are at the same height along axial direction are then categorized into clusters with a k-means classifier. Within each cluster, all slices (from different study cases) are registered to the center of the cluster with a free-form pairwise image registration method described in section 5.4. Following the steps in section 5.3, a brain CT atlas is built by computing an average image for each height along the axial direction. All the above work are conducted off-line once. Given an input slice with abnormalities, we can determine its height along the axial direction of the brain by registering the slice to the atlas. Furthermore, by comparing the input slice with the standard slice in the atlas, we can locate and identify the abnormal regions in the input slice.

5.2   Slice Normalization

All the CT slices used in this work are real data directly extracted from GE cen-
tricity DICOM (Digital Imaging and Communications in Medicine) system. Apart
from the intrinsic differences caused by the variety of the study cases, other defor-
mations have also been introduced in the scanning procedure such as pose differ-
ences, different starting positions in the axial direction etc. Therefore, all the input
slices need to be normalized before any subsequent analysis tasks. To achieve this,
we first fit an ellipse to the inner boundary of the skull with a numerically stable
non-iterative method [HF00]. The method is based on a least squares minimization
method [HF00] to solve the problem:

$$\arg\min_{\boldsymbol{\alpha}} \|\mathbf{D}\boldsymbol{\alpha}\|^2 \qquad \text{subject to} \qquad \boldsymbol{\alpha}^\top \mathbf{C} \boldsymbol{\alpha} = 1 \tag{5.1}$$

where $\boldsymbol{\alpha} = (a\ b\ c\ d\ e\ f)^\top$, representing the coefficients of the ellipse:

$$F(x,y) = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad \text{with constraint} \quad b^2 - 4ac < 0 \tag{5.2}$$

In Equation 5.1, $\mathbf{C}$ represents the constraint matrix (6*6):

$$\mathbf{D} = \begin{pmatrix} \mathbf{C_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{where} \quad \mathbf{C_1} = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 0 \end{pmatrix} \tag{5.3}$$

and $\mathbf{D}$, also known as *design matrix*, is a $N * 6$ matrix $(x_i^2\ x_i y_i\ y_i^2\ x_i\ y_i\ 1)$. As no
iterative steps and computational ambiguity are involved in the method, the fitting
method is very fast and stable. With the axes and centroid of the fitted ellipse, we
then estimate the rotation angle and translation displacements of the input slice.

Figure 5.2 shows the ellipse we fit and the corrected slice. After pose correction, we resize all slices to [350 350].
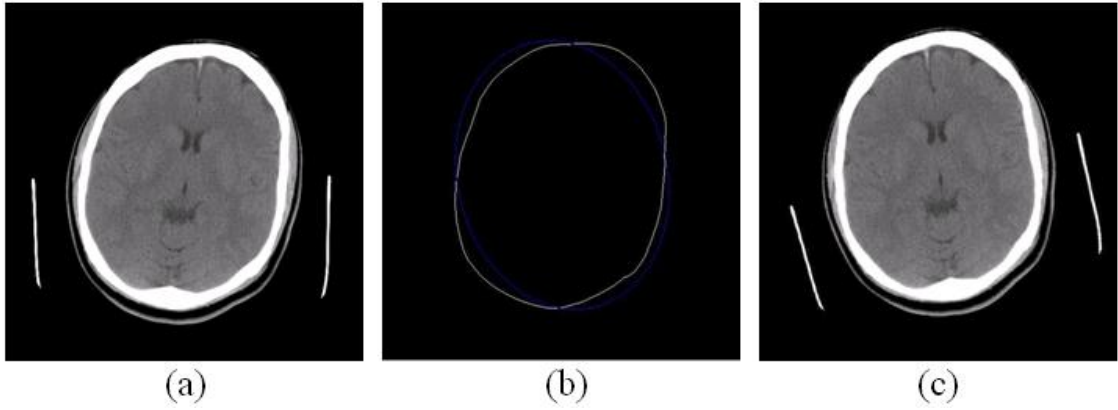

Figure 5.2: The pose correction of an input CT slice with an ellipse fitting method. Image (a) is the original slice; Image (b) shows the inner boundary of the skull and the fitted ellipse (drawn in blue); Image (c) is the slice after pose correction.
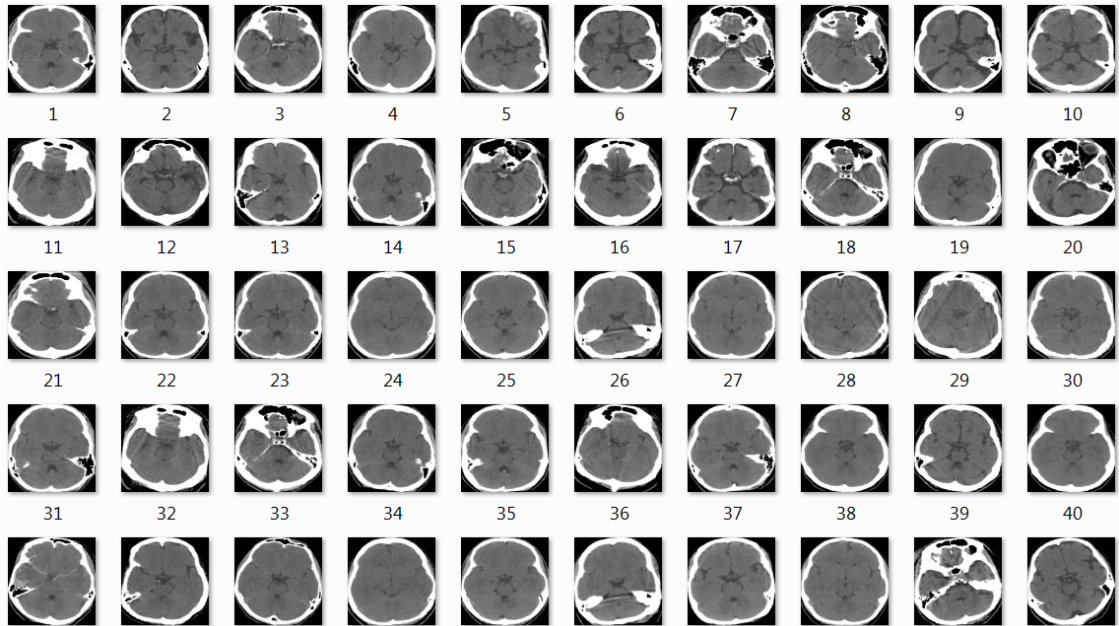
Figure 5.3: Samples of the normalized slices.

5.3   Groupwise Registration for Atlas Construction

As mentioned in Section 2.7, one of the typical applications of groupwise image registration is to construct an atlas which describes the anatomical variations of populations by registering multiple scans from different patients. In this work, we develop a cluster-based groupwise registration framework to build a brain CT atlas with the scans from different study cases. For this purpose, we have selected 63 series of CT scans which are normal or have minor abnormalities. Each of these scans consists of 19 slices that are captured at increasing heights along the axial direction of the brain. Figure 5.4 shows some of the slices that we have chosen for the $6th$ level. With the method proposed in Section 5.2, we normalize these slices. Some of the normalized slices are shown in Figure 5.6.

Unlike MRI scans, the CT scans that most hospitals are using are very sparse, which makes the inter-slice interpolation results either too blur or contain too much edge halos. Therefore, our strategy of building a brain CT atlas is to calculate an average slice for each level along axial direction of the brain instead of constructing a truly 3D model. The simplest way to compute such an average slice is to uniformly averaging on all the normal slices that are labeled with the same level number. Figure 5.5 shows part of the atlas resulting from this direct averaging method. As we can see from the figure, the resultant average slices are very blur and lose most details especially in the boundary regions and tissue areas. Therefore, registering the slices that are labeled with the same level number before computing the average slice is essential.

As discussed in Section 2.7, a simple way to register a set of slices is to register
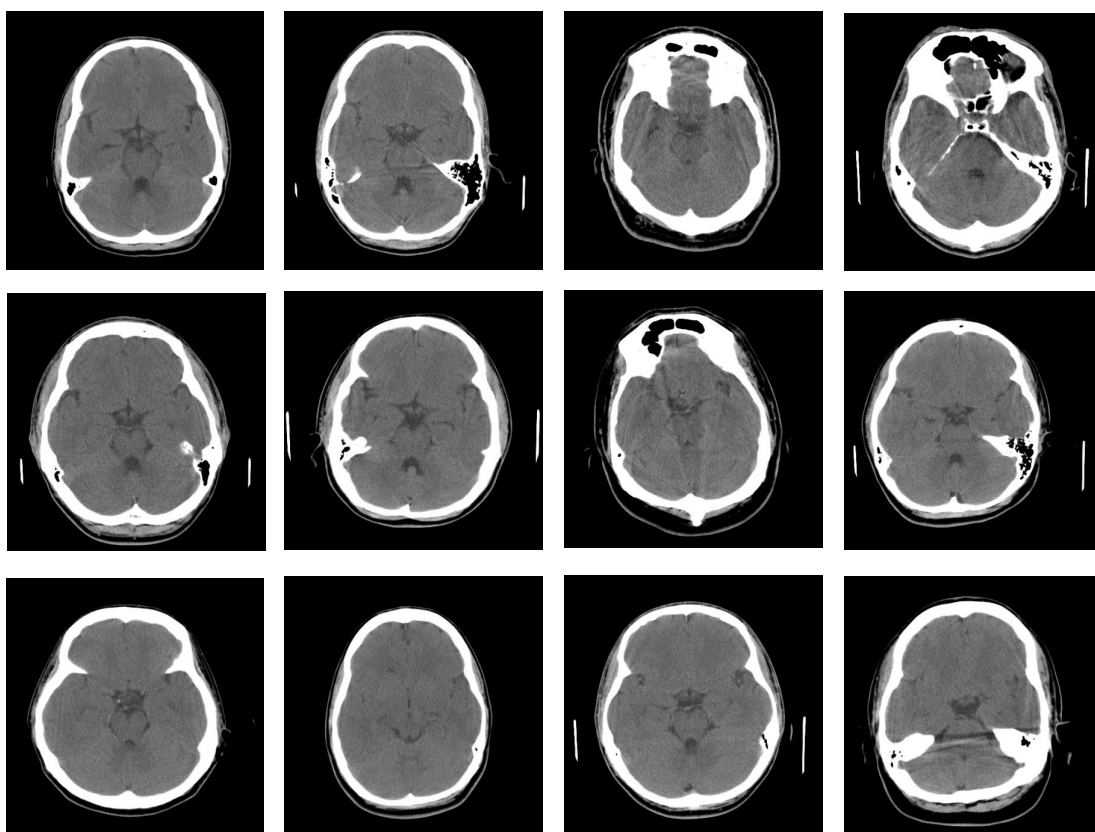
Figure 5.4: Samples of the selected normal (or with minor abnorlity) slices for level 6 (along the axial direction of the brain).
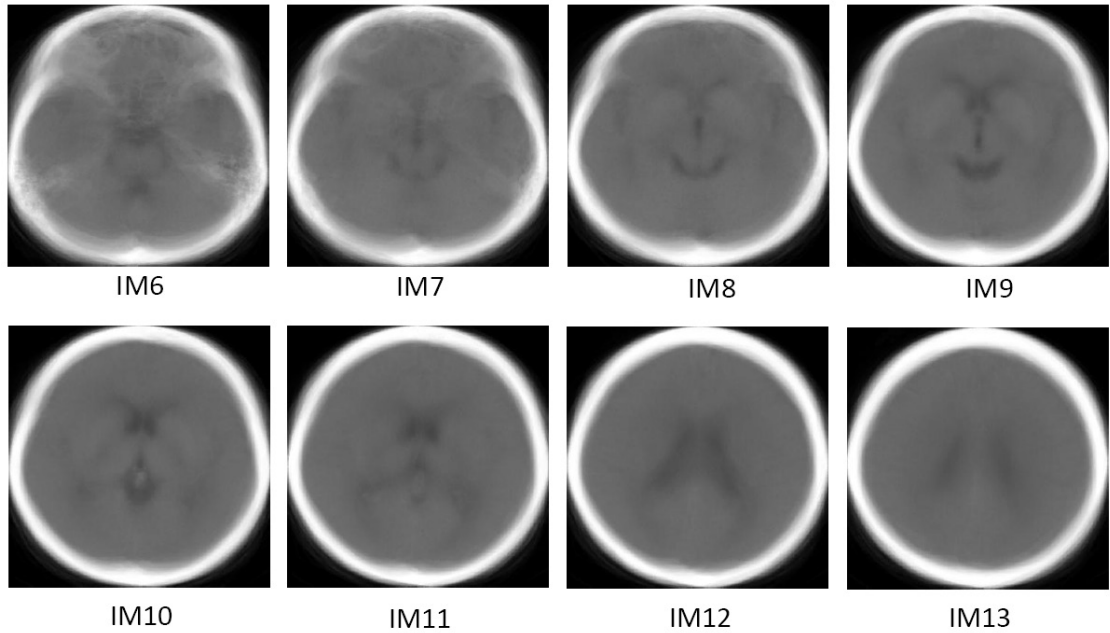
Figure 5.5: The average slices of level 6 to level 13 in the built atlas that was constructed with a direct averaging method.

all the slices (at the same level) to one particular slice. The referenced image (the one that all slices are registered to) could be the first one or a random one in the image set. For instance, an image that is the closest to the geometric mean of the image population is selected as a template and all other images are registered to it [PBIM05]. As mentioned in Section 2.7, this type of groupwise image registration methods inevitably introduce bias to the registration results due to the use of a particular template. This situation can be even worse in our registration task. Due to the variety of patients' heads, the starting positions where the first CT slice was taken can be very different, which means even the slices with the same level number (along the axial direction of the brain) can be greatly different. If the reference slice we choose is far from the majority of the slice population, the computed average slice will fail to describe the anatomical variations of populations. An extreme method to reduce the registration bias is to perform pairwise registration on all

possible image pairs and each image is deformed by all the deformations that are estimated between the image and all other images [SAM⁺04]. This method however suffers from the extensive computational cost and severe blurring effects.

Alternatively, the procedures of registering and averaging can follow certain structures like a tree and intermediate templates are used to connect individual slices [WCS09, MTW09]. For instance, [WCS09] constructs a tree of images by clustering images hierarchically into small-scale sub-groups and registration starts from the sub-groups on the leaf nodes. A more sophisticated method of this kind is to build a minimum spanning tree (MST) where each node represents an image and the edge between two nodes denotes the distance between the two images [MTW09]. The root node represents the final registration result, which can be determined by selecting a node that has the minimal edge length to all other nodes. [JWWS10] has shown that cluster-based groupwise registration methods usually outperform centralized registration methods.

In our work, as the slices labeled with the same level number have great variations, we organize the slices in a three-level tree where the leaf nodes are grouped into different clusters and the root represents the final average slice. The intermediate nodes in the second level of the tree are the average slices of each cluster. The edges between these intermediate nodes and the root denote the importance of the intermediate nodes, which are now computed as the probability of the cluster that the intermediate node belongs to. Within each cluster, for the time being, we manually assign the centroid of the cluster and all the other slices in the cluster are registered to the central slice with pairwise registration methods that will be presented in Section 5.4. Figure 5.6 shows the three clusters and some of their members for slices at level 6 in the axial direction of the brain.
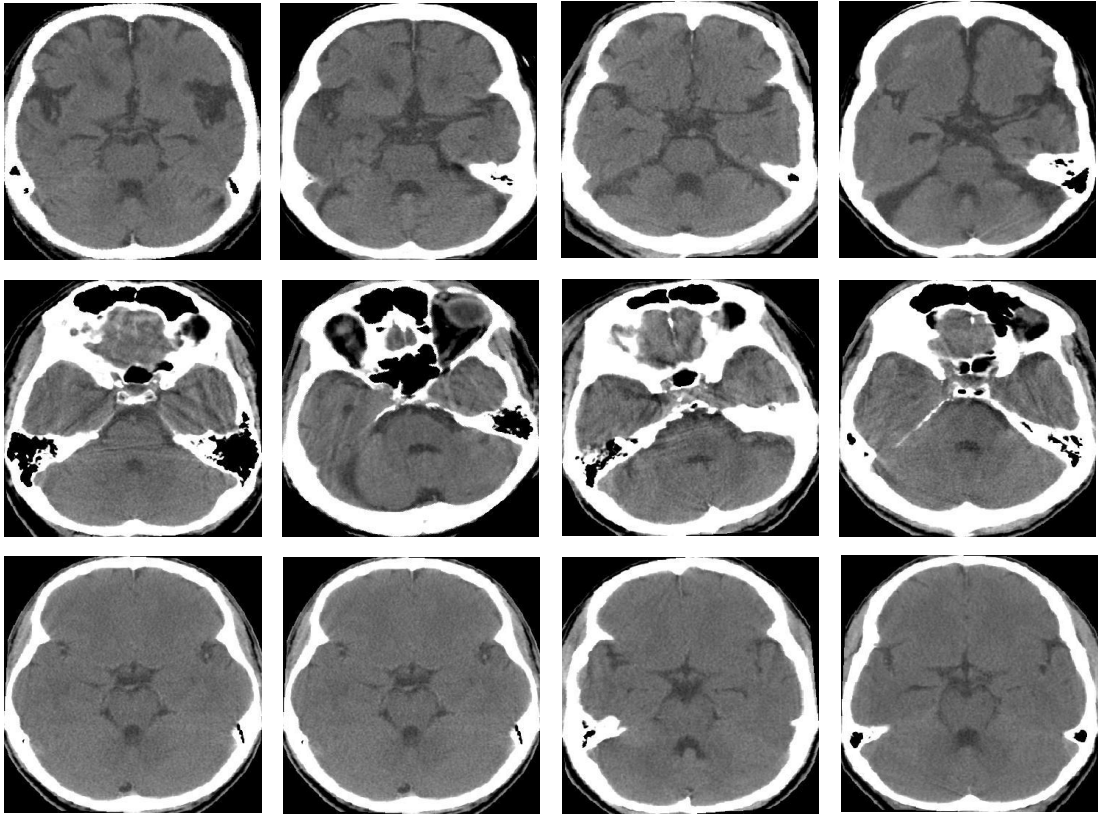
Figure 5.6: The three groups of normal slices at level 6. The slices in the same row belong to the same group and the first slice in each row represents the centroid of the group.

With the registered slices in one cluster, an intermediate average slice is computed for that cluster by averaging all the registered slices. The final average slice for a particular level is calculated as:

$$I = \sum_{i=1}^{N} \frac{N_i}{N} I_i \qquad \text{where} \qquad N = sum_{i=1}^{N} N_i \qquad (5.4)$$

where I_i is the intermediate average slice of the ith cluster and N_i is the number of slices in this particular cluster. Perform the above steps with all the 12 levels, we can build an atlas as partially shown in Figure 5.7.

As we can see, compared with the atlas shown in Figure 5.5, this new atlas constructed with groupwise registration procedure is much clearer and contain more details on the brain tissues. In Section 5.5, we will show how this atlas can be used to determine the axial level of an input slice. Moreover, a fine atlas like this can also contribute to the detection and location of the abnormal regions on an input slice that is affected by TBI. We will briefly show this in Section 5.6.

5.4 Pairwise Registration of Brain CT Scans

As discussed in the previous sections of this chapter, we build the groupwise image registration framework upon a non-rigid pairwise registration method. In this section, we present this pairwise registration method in details.

As we can see from the above images, brain CT scans contain a lot of homogeneous areas, which makes it difficult to extract image features or anatomical structures. Therefore, in this work, we conduct pairwise registration on pixel level and use the dense intensity information. As we will discuss in Section 5.7, this step can be improved by using some salient control points instead of randomly selected
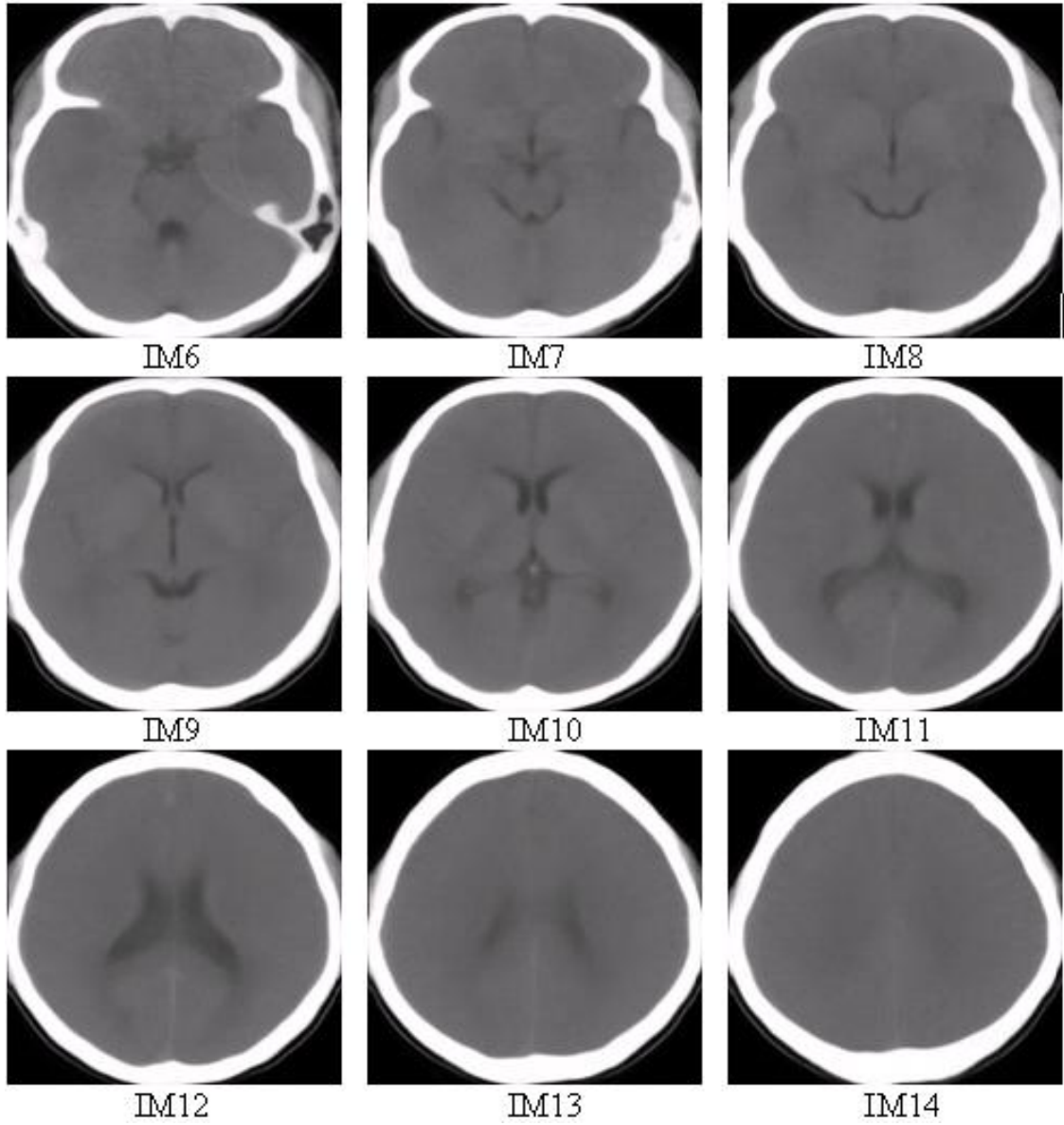
Figure 5.7: The average slices for level 6 to level 14 in the atlas that has been constructed with our groupwise image registration method.

points.

5.4.1 Transformation Model

As in other applications, in order to register two CT slices, we need to define a transformation model which can describe the deformations between the two images. Since the two CT slices to be registered were taken from different study cases ie. different patients, the transformation between them is unpredictable and nonuniform and rigid transformation models are inadequate to capture the deformation between them. Therefore, in this work, we employ the free-form transformation model as the one we used in Chapter 4:

$$T(x,y) = T_{global}(x,y) + T_{local}(x,y) \tag{5.5}$$

As we have explained in the previous chapter, the final transformation model consists of the affine-based global transformation (as shown in Equation 5.6) and the B-splines based local deformation (as shown in Equation 5.7). These two transformations together are expected to capture all the deformations between slices.

$$T_{global}(x,y) = \begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{5.6}$$

In Section 4.5.2, we have explained that this transformation model was used as it usually produces smooth and continuous transformation and B-splines are locally controlled. Apart from these reasons, this transformation model is chosen for this registration task for splines are more suitable to describe the soft-tissue

type of deformations between slices from different patients. However, compared to the work in Chapter 4, the transformation model estimated in this work needs to be smooth and continuous in larger size of sub-images or even the whole image.

$$T_{local}(x, y) = \sum_{m=0}^{3} \sum_{n=0}^{3} B_n(u) B_m(v) \phi_{i+n, j+m} \tag{5.7}$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$ and $[n_x, n_y]$ is the size of the control mesh. ϕ denotes the mesh of control points and B_i refers to i^{th} basis function as:

$$\begin{cases} B_0(x) = (1-x)^3/6 \\ B_1(x) = (3x^3 - 6x^2 + 4)/6 \\ B_2(x) = (-3x^3 + 3x^2 + 3x + 1)/6 \\ B_3(x) = x^3/6 \end{cases} \tag{5.8}$$

5.4.2 Cost function

Like the transformation model presented in the previous section, we choose a cost function that is similar to the one we used for historical document registration. As shown in Equation 5.9, the cost function consists of the similarity part for accurate matching and the smoothness part for transformation regularization.

$$C = C_{sim}(R_{xy}, T(I_{xy})) + \lambda C_{smo}(T) \tag{5.9}$$

where R_{xy} and I_{xy} stand for reference slice and target slice. λ represents the relative weight of the two constrains and has been tuned as 0.1 in our experiments. Only, for historical document registration, we used residual complexity (RC) for the similarity part. In this work, however, since the two images to be registered

were actually taken from different subjects (patients) and to reserve the variation of the populations as much as possible, we use normalized mutual information as the similarity measure. This is shown in Equation 5.10. Experiments also show that mutual information is more accurate and robust for multi-modal image registration and with normalization it is also independent on the overlap of the images.

$$C_{sim}(A, B) = (H(A) + H(B))/H(A, B) \tag{5.10}$$

where $H(A)$, $H(B)$ denote the marginal entropies of slice A, B and $H(A, B)$ is their joint entropy, which is calculated from the joint histogram. For the smoothness part, we still use the space integral of the square of the second order derivatives:

$$C_{smo} = \frac{1}{s} \int \int [(\frac{\partial^2 T}{\partial x^2})^2 + (\frac{\partial^2 T}{\partial y^2})^2 + 2(\frac{\partial^2 T}{\partial xy})^2]dxdy \tag{5.11}$$

where s denotes the area of the image domain and T represents the estimated transformation. In this work, we increase the relative weight of the smoothness part in Equation 5.9 so that the estimated transformation could be more smooth and more tissue variations could be preserved in the final average slice. The same with the work in the previous chapter, we only maximizing C_{sim} when estimating T_{global} to reduce computational cost and cascading registration errors. With the estimated transformation model, the target slice (the one to be registered) is re-sampled to obtain the registered slice. Figure 5.8 shows a sample CT slice and its registered slice.
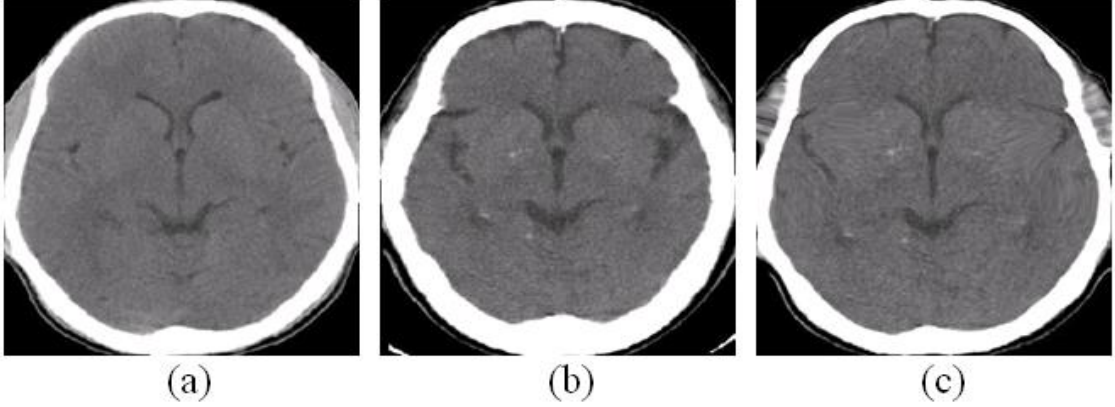
Figure 5.8: Sample results from the pairwise registration between slices. Image (a) is the reference slice (the centroid of each cluster); Image (b) is the target slice; Image (c) is the registered target slice.

5.5 Slice indexing

In this section, we demonstrate how to determine the input slice's position along the axial direction of the brain with the built atlas. As we have discussed in Section 5.1, this indexing/locating procedure is used as a pre-processing step in CBIR or CAD systems and needs to be very fast. In this work, we develop a simple method to achieve this high speed. To be specific, we compute the density differences between the input slice and all the average slices in the atlas. The input slice is labeled with the same level number as the average slice which is the most similar to the input slice. For this purpose, we first normalize the input slice with the method presented in Section 5.2 and use correlation coefficients as shown in Equation 5.12 to measure the difference between the two images.

$$Corr(I_t, I_r) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{r_{ij}} - \overline{I_r})(I_{t_{ij}} - \overline{I_t})}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{r_{ij}} - \overline{I_r})^2} \cdot \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{t_{ij}} - \overline{I_t})^2}} \qquad (5.12)$$

where I_t is the input slice and I_r is an average slice in the atlas. $\overline{I_t}$ and $\overline{I_r}$ are the expectation of slice I_t and I_r. During the experiments, we noticed that when we normalized both the input slice and the average slice in the atlas, useful information like the size of the skull is thrown away. As we can see from Figure 5.1, the size of the skull and the ratio of the fitted ellipse's axes in the beginning slices and in the ending slices is quite different. Therefore, we use these information to guide the indexing procedure by determining the coarse position (starting, middle or ending) of the input slice. These features are the by-product of the normalization step in Section 5.2, therefore no extra processing time is needed to extract them.

Level	IM6	IM7	IM8	IM9	IM10	IM11	IM12	IM13	IM14	IM15	IM16	IM17
Precision (%)	55.92	58.18	56.75	62.49	66.44	72.93	72.16	73.78	71.54	79.76	68.17	53.82
Recall (%)	96.79	74.16	72.77	78.50	84.38	85.63	84.38	86.92	73.93	75.12	67.54	88.42

Table 5.1: Quantitative evaluation of the proposed slice indexing method.

We have tested the proposed slice indexing method with 70 study cases which consist of 1520 CT slices. Among these study cases, 60 cases have severe TBI. These cases are the most frequent scenarios that are confronted and concerned by radiology doctors. To quantitatively evaluate the proposed method, we computed its indexing precision and recall for each axial level. Table 5.1 shows the results of level 6 to level 14, which are most important in TBI diagnosis. As seen from the table, the indexing method is more accurate on slices located at level 11 to 15. It's probably because slices within this interval are more distinguishable. This is consistent with the fact that CT slices after level 10 were taken with interval of 7mm instead of 5mm as for the previous slices. The precisions for slices after level 16 are relatively low, which is likely caused by the significant difference in the size of the brain matter between different patients. So far, the overall precision and recall over all 1520 slices are 67.86% and 79.71%. It worths mentioning that several test cases actually contain more than 19 slices of which we only used 19.
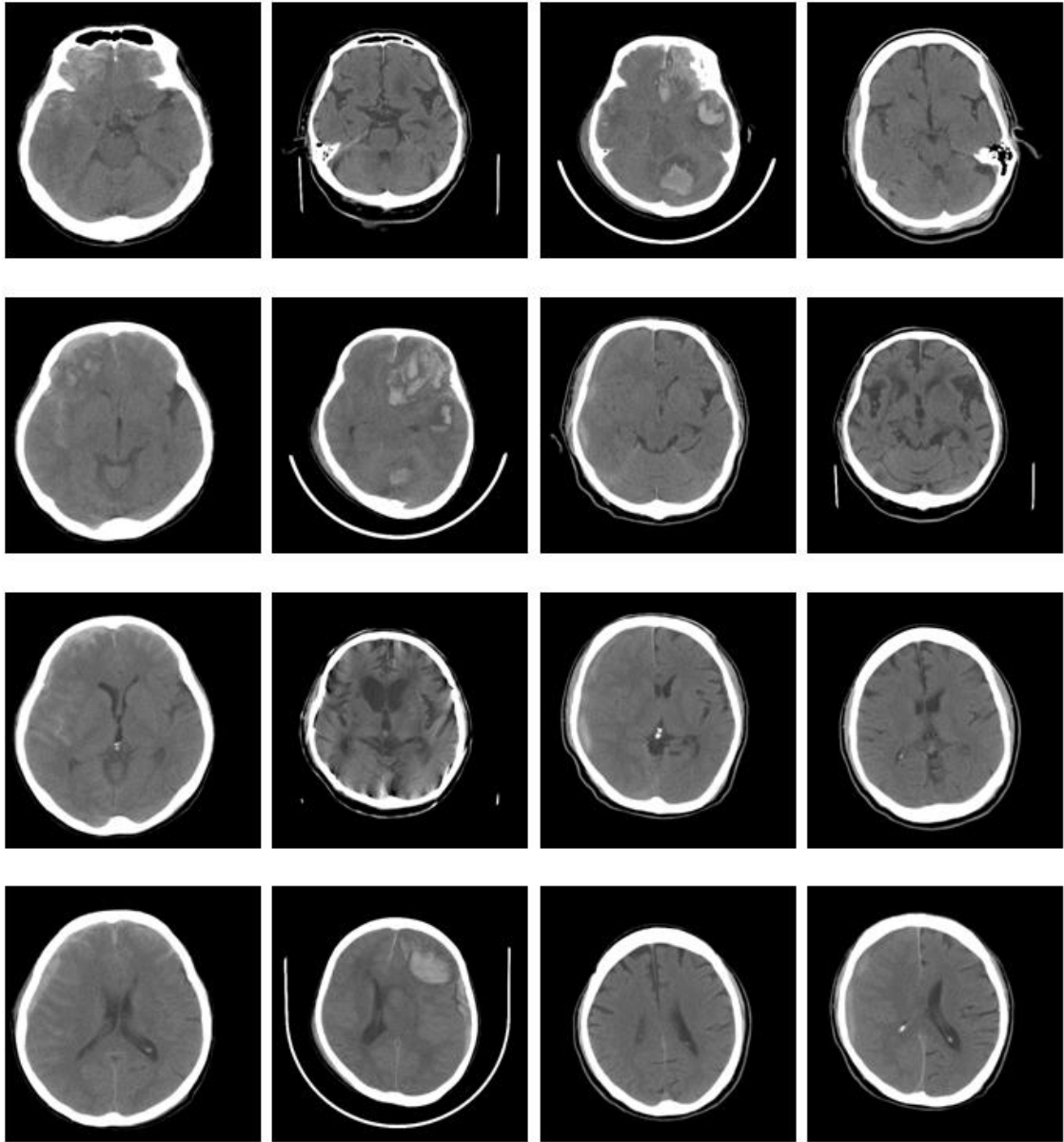
Figure 5.9: Sample results for slice indexing. The images in the same row belong to the same height. From top to bottom, the images was determined to belong to these levels: IM6, IM8, IM10, IM12

This partially impaired the overall indexing precision and recall. More importantly, the precision and recall we computed above are for exact indexing which means the exact level number of the input slice is determined. If just an interval which the level number of the input slice lies in as in [LLP+10] is needed, the precision and recall of our method are around 95% and 98%, which are much higher than the results in [LLP+10]. In addition, as most time-consuming tasks such as image registration and atlas building are conducted only once and off-line, the proposed slice indexing procedure is quite fast. Typically, less than 0.03s is needed for indexing one input CT slice, which is ten times faster than the method proposed in [LLP+10].

5.6   Abnormality Detection

With the built atlas, we can also detect and locate the abnormal regions on the CT slices that are affected by severe TBI. Given an input CT slice, we first determine its height on the axial direction of the brain with the method described in Section 5.5. Then we compare the input slice with the average slice at the same height in the atlas. For instance, with an input slice X, we determine its level is 6, which means it is probably the $6th$ slice in the scan it belongs to. So we compare it with the "standard slice" of level 6 in the atlas. In this experiment, we compute the absolute intensity differences between each pixel of the input slice and the average slice in the atlas. The regions where the intensity differences are larger than a threshold are regarded as abnormal areas. Figure 5.10 shows some sample CT slices that were distorted by TBI and the detected abnormal areas on them. With a simple classifier which takes into account the size, location and shape of the abnormal region, we can determine the type (EDH or SDH) of the abnormality.
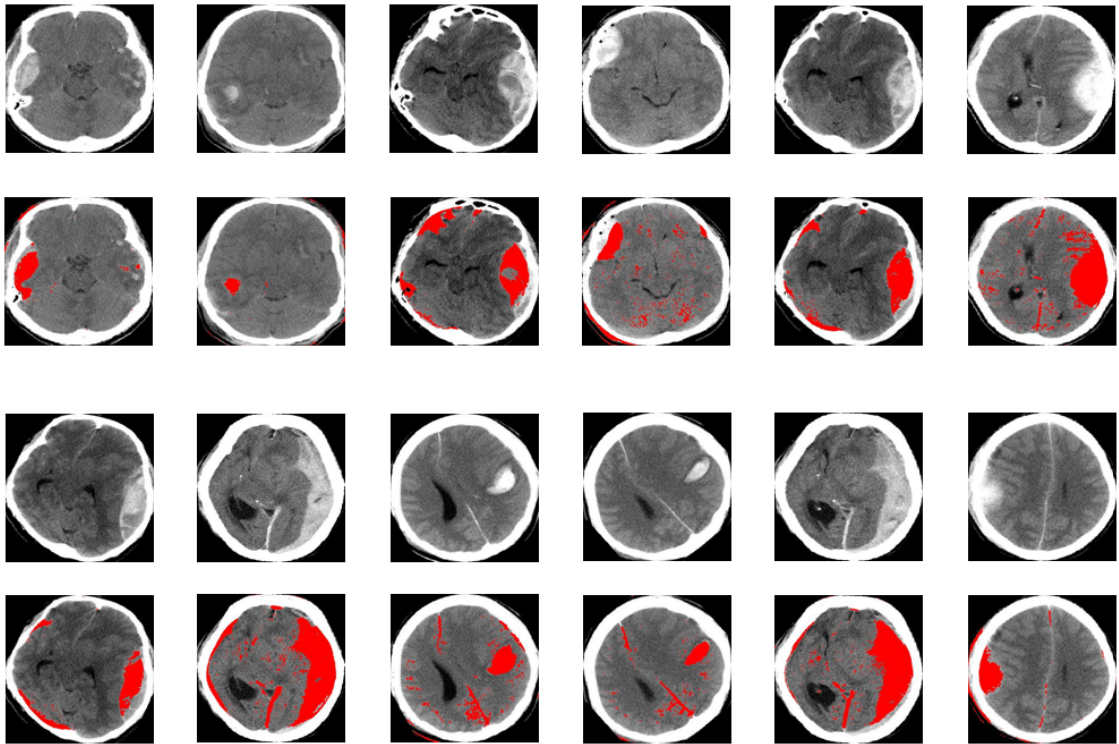
Figure 5.10: Sample results for abnormality detection. The images in the odd rows show the original CT slices and the red regions shown in the images in the even rows demonstrate the detected abnormal areas.

5.7 Conclusion and Discussion

In this chapter, we have presented a groupwise registration framework to construct a brain CT atlas which can be used for slice indexing and abnormality detection. We built the groupwise registration framework upon a non-rigid pairwise registration method. In particular, we selected 63 series of CT scans which are normal or have minor abnormalities and labeled all the slices with the level number indicating their position in the axial direction of the brain. We organized the slices with the same level number in a cluster structure. All slices in a cluster were registered to the center of the cluster and an intermediate average slice was computed with the registered slices. Combining all the intermediate average slices for a particular level, we gained the final average slice for the level. We have demonstrated that the built atlas could be used to index an input slice in the axial direction of the brain. Experiments on real brain CT scans have shown that this new method is efficient and accurate enough to be conveniently adopted by CBIR or CAD systems. We also showed that by registering an input slice which was affected by severe TBI to the atlas, we can detect and locate the abnormal regions in the slice.

CHAPTER 6

Conclusion and Future Directions

6.1   Summary

In this information age, more information is kept in digital form and the final knowledge is often gained from the combination of various data sources. Therefore, for the purpose of comparison and fusion, the ability of relating the images of the same scene (or similar scenes) is essential. Image registration refers to the procedure which transforms two or more images captured from different perspectives, at different time and with different devices into a common coordinate system. It is a crucial step in many image analysis tasks and has been deeply studied in many domains including medical imaging, remote sensing and computer vision.

In this thesis, we aim to investigate proper image registration techniques that are adaptive to document image restoration and brain CT atlas building. To achieve this goal, we first reviewed the four major components that constitute a typical image registration framework. For each component, we discussed its general

work flow and reviewed the involved key techniques.

Then we presented a new image feature which can be used to register a skewed document image with an imaginary image captured by posing the document in exactly upright direction. The feature is called interline white run which is the continuous white pixels lying exactly between two neighboring text lines. We located this feature on the histograms that were acquired by horizontally scanning the document image. As the transformation between the two images to be registered is very simple and has only one parameter for rotation distortion, we directly computed this parameter with the detected white runs. With this image feature, we can also identify the orientation of the input document. We have tested the proposed feature and registration method with real-world degraded document images. The experimental results have demonstrated that the new feature outperforms other features in estimating the skew angles of the degraded document images in terms of accuracy and efficiency.

The second approach we have developed precisely registers the recto image and verso image of a double-sided handwritten document so that bleed-through distortions on the document images can be fully corrected. This registration framework consists of a rigid coarse registration procedure and a non-rigid fine registration method. For the coarse registration method, a pair of representative sub-images were first extracted from the two images. Using the two sub-images and correlation coefficients as similarity measure, we estimated an Affine transformation for the global deformation between the two images to be registered. This coarse registration method mainly deal with document images that were degraded by slight rigid global distortions like translation and rotation. For severely degraded document images which were impaired by non-rigid local distortions like warping effects, uneven surfaces, rigid transformation and block-based registration is in-

sufficient. Therefore, we have developed the fine registration method which uses residual complexity as similarity measure and a transformation model combining affine and spline-based local deformation. In particular, we have developed an accurate method to extract and match salient point sets from the two images. Once the two images of a document have been precisely registered, we corrected the bleed-through distortions with a wavelet-based restoration method. We indirectly evaluated the proposed registration framework by visually and quantitatively assessing the final document restoration results.

Based on the pairwise image registration framework that we have developed for historical document images, we have designed a groupwise image registration framework for the purpose of constructing a brain CT atlas with multiple CT scans from different study cases. The built atlas was a set of average slices of which each describes the anatomical variations of populations at a particular position along the axial direction of the brain. To build such an atlas, we have labeled a group of normal (or have minor abnormalities) CT scans with an indexing number which denotes their position (level) on the axial direction. Then we classified the slices with the same indexing number into different clusters. Within each cluster, a central slice was selected and all the remaining slices in that cluster were then registered to the central slice using a non-rigid registration method. With the registered images, we computed an intermediate average slice for each cluster. The final average slice for that particular level was the combination of all the intermediate average slices that belonged to that level. With the constructed atlas, we are capable of locating the input slices on the axial direction of the brain. This slice indexing method will significantly accelerate content based retrieval systems especially when tremendous number of images are involved in the systems. We have also demonstrated that the built atlas could be used to detect the abnormal

regions on the input slices that were affected by traumatic brain injuries. With certain rule-based methods or simple classifiers, we can further determine the type of the abnormalities.

6.2 Future Directions

In reference to our earlier discussions in Section 3.5, Section 4.8 and Section 5.7 on the limitations of the method described in each work, further improvements and extensions can be explored in all directions as follows.

6.2.1 Future Work on Skew Correction

As discussed in Section 3.5, when the degraded document contains multiple columns, multiple fonts or multiple font sizes, our registration method based on interline white run will fail to produce satisfactory results. This is mainly due to the interline white run extraction procedure which is unable to distinguish the different types of white runs that are obtained from different text blocks. In this case, there are two directions to address this problem. We can develop more sophisticated clustering methods to extract the interline white run of each text block. This however requires more prior knowledge about the layout of the document so that the interline white runs of a text block with small font size are not counted as the inner line white runs of a text block with larger font size. Alteratively, we can add a segmentation procedure before scanning the document to collect white runs. In this way, we can segment different text blocks and conduct our registration method on each block. With sufficient text contents on the document, we can still estimate the global or even local skew angles of the document.

In Section 3.5, we showed that the proposed skew estimation method is also limited by its assumption that the skewed document contains sufficient text contents. When there are not enough text lines on the document, the histogram of white runs may be sensitive to outliers and the peak indicating interline white runs may not be easy to detect. Under this situation, we can not really do much with the white run based method. A possible solution to this problem may be using non-text objects in the document. Actually, if graphics, pictures or tables can be accurately detected from the document, their boundaries which are usually parallel to text lines can be used to estimate the skew angle of the document.

Another extension to the presented registration method is to include a script identification procedure and a supervised training procedure so that the skew estimation method can deal with multi-lingual document images. Being trained with documents printed in different scripts, the registration method can gain the distribution pattern of the white runs for each script. With an input document, the script identification procedure can be applied first and the script information will guide the registration method to choose proper domain knowledge to estimate the document's skew angle.

6.2.2 Future Work on Bleed-through Correction

In Section 4.4, we have shown that some documents that are degraded by severe local deformations need a fine registration procedure so that the bleed-through distortions on the documents can be fully corrected. For the time being, whether an input document needs the fine registration procedure or not is still manually determined by document analysis experts. We examine both the original document and the registered image resulting from the coarse registration procedure to

decide on this. An automatic algorithm to conduct this task is therefore helpful, but challenging. To achieve this goal, we need to determine the 'nature' of the deformations that are present on the document. [LBT10] has developed a classifier that can pre-characterize the noises present in a document as: bleed-through, skew and frames based on a set of human labeled training samples. Defining and determining the severity of the bleed-through distortion in a document image is however relatively easier compared to defining and determining the severity of the local deformations between the two images to be registered.

As can be seen in Section 4.5, we selected and matched the salient control points mainly based on the directions of their gradients. This method works well generally for documents that were written in English. The reason is, in a document that was handwritten in English, the strokes orientating in particular directions have dense intensities and tend to cause heavy bleed-through distortions on the other side of the page. With documents written in other scripts however, there are no such apparent writing behaviors and stroke patterns. Therefore, we need other prior knowledge to detect salient control points from these document images.

As mentioned in Section 4.8, we indirectly evaluated the proposed historical document image registration framework. The major disadvantage of this evaluation strategy is that we can not separate the registration errors with the errors that are introduced by the subsequent bleed-through correction procedure. The difficulty of quantitatively assessing the registration results is caused by the lack of ground truth about the document's real deformation. Some researchers have proposed to evaluate newly proposed registration methods with synthetic images. Ideally, by purposely distorting a document image with image processing softwares, we can gain the true deformations and then evaluate the proposed registration method. However, this strategy can actually work only if we knew the way that stains

(water or coffee), moist environment, aging procedure and image acquisition devices distorted a document.

Another direction to extend the work on historical document restoration is to investigate bleed-through correction methods that have a high tolerance for the displacements between the two sides of a document. Alteratively, we need to add more advanced post-processing methods after the bleed-through correction procedure so that the remaining bleed-through distortions can be further removed and the broken foreground strokes can be recovered. [WT01] has proposed a simple direction-based edge recovery method, which could be the starting point of this work.

6.2.3 Future Work on CT Slice Registration

As mentioned in Section 5.4, the underlying pairwise image registration method in the groupwise registration framework registered two slices with randomly picked points. Replacing these randomly picked points with salient points which are carefully extracted from the slices can further fine the built atlas and thus improve the accuracy of CT slice indexing and abnormality detection. One possible way is to segment the present tissues on the middle slices and select control points from these extracted objects. For the beginning (the first two or three slices) and ending slices (the last two or three slices), transition points on the skull and points on the middle line can be used for registration. These points are relatively stable with respect to the variety of human brains and are also easier to be extracted. However, the number of these points are probably insufficient for fine pairwise registration, therefore, accurate interpolation methods are required.

In Section 5.3, we mentioned that we manually determined the number of the

clusters for each level along the axial direction and manually assigned the centroid for each cluster. This strategy worked well since we only labeled 63 slices for one particular level. In order to build a finer atlas, we may need more normal CT slices, which makes a fully unsupervised clustering method essential.

Another direction to extend this work is to explore the registration-based methods for slice indexing and abnormality detection. As shown in Section 5.5 and Section 5.6, we currently index the input slice and detect the abnormal regions on it simply with comparison methods. We have tried to register the input slice which is affected by TBI to the atlas that we have built and use the registration errors to determine its level on the axial direction and the abnormal regions on it. The performances are better than those of the current comparison-based methods. However, the major problem was the time cost which was almost 100 times as those of the current methods.

Bibliography

[Anu70]   P. Anuta. Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques. *Journal of GeoEl*, 8:353–368, October 1970. 18

[App96]   C. Appledorn. A new approach to the interpolation of sampled data. *In Proceedings of IEEE Transactions on Medical Imaging*, 15(3):369–376, June 1996. 26

[Bai87]   H.S. Baird. The skew angle of printed documents. pages 14–21, Rochester, New York, 1987. 13, 36

[Bai03]   H. Baird. Digital libraries and document image analysis. In *In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, page 2, Washington, DC, USA, 2003. IEEE Computer Society. 33

[BB95]   S. Beauchemin and J. Barron. The computation of optical flow. *Journal of ACM Computing Surveys*, 27(3):433–466, 1995. 26

[BCT+98]   D. Becker, A. Can, J. Turner, H. Tanenbaum, and B. Roysam. Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination. *In Proceedings of IEEE Transactions on Biomedical Engineering*, 45:105–118, 1998. 29

[Ber98]   R. Berthilsson. Affine correlation. In *In proceedings of 14th International Conference on Pattern Recognition (ICPR'98)*, pages 1458–1460, 1998. 18

[BM92]   P.J. Besl and N.D. McKay. A method for registration of 3-d shapes. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'92)*, 14:239–256, 1992. 15

[Bro92]   L.G. Brown. A survey of image registration techniques. *Journal of ACM Computing Surveys*, 24:325–376, 1992. 3, 8

[BS97]   D. Bhattacharya and S. Sinha. Invariance of stereo images via the theory of complex moments. *Journal of Pattern Recognition*, 30(9):1373–1386, September 1997. 13

[CH02]   Z. Chen and S. Haykin. On different facets of regularization theory. *Journal of Neural Computation.*, 14:2791–2846, December 2002. 25

[CHH04]   W.R. Crum, T. Hartkens, and D.L.G. Hill. Non-rigid image registration: theory and practice. *The British Journal of Radiology*, pages S140–53, 2004. 3, 9

[CP04]   A. Chaniotis and D. Poulikakos. High order interpolation and differentiation using b-splines. *Journal of Computational Physics*, 197(1):253–274, 2004. 26

[CTWS00]   R. Cao, C.L Tan, Q. Wang, and P. Shen. Segmentation and analysis of double-sided handwritten archival documents. In *In Proceedings of IAPR International Workshop on Document Analysis System (DAS'00)*, pages 147–158, 2000. 72

[DK97]   X. Dai and S. Khorram. Development of a feature-based approach to automated image registration for multitemporal and multisensor remotely sensed imagery. In *In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'97)*, pages 243–245, Singapore, 1997. 13

[Dod97]   N. Dodgson. Quadratic interpolation for image resampling. *In Proceedings of IEEE Transactions on Image Processing*, pages 1322–1326, 1997. 26

[Don01]   H. Don. A noise attribute thresholding method for document image binarization. *International Journal on Document Analysis and Recognition*, 4(2):131–138, 2001. 51

[DWPL04]   X. Ding, D. Wen, L. Peng, and C. Liu. Document digitization technology and its application for digital library in china. In *In Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL'04)*, page 46, Washington, DC, USA, 2004. IEEE Computer Society. 33

[EPV93]   P.V.D. Elsen, E. Pol, and M. Viergever. Medical image matching –
           A review with classification. *In Proceedings of IEEE Engineering in
           Medicine and Biology Society (EMBC'93)*, pages 26–38, March 1993.
           8

[FK01]   K. Franke and M. Köppen. A computer-based system to support
           forensic studies on handwritten documents. *International Journal
           on Document Analysis and Recognition*, 3(4):218–231, 2001. 51

[G. 99]   G. Nagy and T.A. Nartker and S.V. Rice. *Optical Character Recog-
           nition: An Illustrated Guide to the Frontier*. Kluwer Academic Pub-
           lishers, Norwell, MA, USA, 1999. 35

[GKN08]   S. Gefen, N. Kiryati, and J. Nissanov. Atlas-based indexing of brain
           sections via 2-d to 3-d image registration. In *Journal of Biomedical
           Engineering*, volume 55, pages 147–156, 2008. 81

[Gos87]   A. Goshtasby. Piecewise cubic mapping functions for image registra-
           tion. *Journal of Pattern Recogn.*, 20(5):525–533, 1987. 23

[Gos88]   A. Goshtasby. Registration of images with geometric distortions.
           Journal of Geoscience and Remote Sensing, 26(1):60–64, 1988. 24

[GS83]   B. Ghaffary and A. Sawchuk. A survey new techniques for image
           registration and mapping. In *In Proceedings of Application of Digital
           Image Processing*, pages 222–239, 1983. 8

[GS85]   A. Goshtasby and C. Stockman. Point pattern matching using con-
           vex hull edges. *In Proceedings of IEEE International Conference on
           Systems, Man, and Cybernetics (SMC'85)*, 15:631–637, 1985. 13

[GSC98]   V. Govindu, C. Shekhar, and R. Chellappa. Using geometric properties for correspondence-less image alignment. In *In Proceedings of the 14th International Conference on Pattern Recognition (ICPR'98)*, page 37, Washington, DC, USA, 1998. IEEE Computer Society. 13

[GSP86]   A. Goshtasby, G. Stockman, and C. Page. A region-based approach to digital image registration with subpixel accuracy. *In Proceedings of IEEE Transactions on Geoscience and Remote Sensing*, 24:390–399, 1986. 13, 14, 23

[HBHH01]   D.L. Hill, P.G. Batchelor, M. Holden, and D.J. Hawkes. Medical image registration. *Journal of Physics in Medicine and Biology*, pages R1–R45, 2001. 3

[HF00]   R. Halir and J. Flusser. Numerically stable direct least squares fitting of ellipses. 2000. 83

[HHD⁺00]   M. Holden, D. Hill, E. Denton, J. Jarosz, T. Cox, T. Rohlfing, J. Goodey, and D. Hawkes. Voxel similarity measures for 3-d serial mr brain image registration. *In Proceedings of IEEE Transactions on Medical Imaging*, 19(2):94–102, February 2000. 29

[HKR93]   D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the hausdorff distance. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'93)*, 15(9):850–863, 1993. 18

[HMP92]   Y. Hsieh, D. McKeown, and F. Perlant. Performance evaluation of scene registration and stereo matching for artographic feature extrac-

tion. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'92)*, 14(2):214–238, 1992. 13

[Hol08]   M. Holden. A review of geometric transformaitons for nonrigid body registration. *In Proceedings of IEEE Transactions on Medical Imaging*, pages 111–128, 2008. 3

[HZ04]   R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. xiii, 22

[JDJG04]   S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Journal of Neuroimage*, pages 151–160, 2004. 30, 31

[JWWS10]   H. Jia, G.R Wu, Q. Wang, and D.G. Shen. Absorb: Atlas building by self-organized registration and bundling. In *In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'10)*, pages 2785–2790, 2010. 88

[Kar01]   R. Karl. *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Kluwer Academic Publishers, Norwell, MA, USA, 2001. 15, 24

[KSP07]   S. Klein, M. Staring, and J.P. Pluim. Evaluation of optimization methods for nonrigid medical image registraiton using mutual information and b-splines. *In Proceedings of IEEE Transactions on Image Processing*, 16:2879–2890, 2007. 10

[LBT10]   R.D. Lins, S. Banergee, and M. Thielo. Automatically detecting and classifying noises in document images. In *In Proceedings of the 2010*

ACM Symposium on Applied Computing, pages 33–39, New York, NY, USA, 2010. ACM. 106

[LCK05]   S.J. Lu, B.M. Chen, and C.C. Ko. Perspective rectification of document images using fuzzy set and morphological operations. *Journal of Image Vision Computing.*, 23(5):541–553, 2005. 39

[LCS99]   T. Lehmann, C. Conner, and K. Spitzer. Survey: interpolation methods in medical image processing. *In Proceedings of IEEE Transactions on Medical Imaging*, 18:1049–1075, 1999. 26

[Lew95]   J. Lewis. Fast normalized cross-correlation. In *In Proceedings of International Conference on Vision Interface (VI'95)*, pages 120–123. Canadian Image Processing and Pattern Recognition Society, 1995. 17

[LGW+10]   S. Li, T. Gong, J. Wang, R. Liu, C.L. Tan, T.Y. Leong, B.C. Pang, C.C.T. Lim, and C.K. Lee. Tbidoc: 3d content-based ct image retrieval system for traumatic brain injury. In *In Proceedings of SPIE Medical Imaging Conference*, 2010. 80

[LLP+10]   R. Liu, S. Li, C.L. Tan B.C. Pang, C.C.T. Lim, C.K. Lee, Q. Tian, and Z. Zhang. Fast traumatic brain injury ct clice indexing via anatomical feature classification. In *In Proceedings of International Conference on Image Processing (ICIP'10)*, pages 4377–4380, 2010. 81, 82, 98

[LMM95]   H. Li, B. Manjunath, and S. Mitra. A contour-based approach to multisensor image registration. *In Proceedings of IEEE Transactions on Image Processing*, 4:320–334, 1995. 13

[LT03]   Y. Lu and C.L. Tan. Improved nearest neighbor based approach to accurate document skew estimation. In *In proceedings of Seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, Edinburgh, UK, August 2003. 13, 36, 37

[LTW94]   D.S. Le, G.R. Thoma, and H. Wechsler. Automated page orientation and skew angle detection for binary document images. *Journal of Pattern Recognition*, 27(10):1325–1344, 1994. 13, 36

[LVPG02]   G. Leedham, S. Varma, A. Patankar, and V. Govindarayu. Separating text and background in degraded document images: A comparison of global threshholding techniques for multi-stage threshholding. *In Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition*, page 244, 2002. 51

[MF93]   C. Maurer and J. Fitzpatrick. A review of medical image registration. In *In Proceedings of Interactive Imageguided Neurosurgery*, pages 17–44, 1993. 8

[MH97]   S. Moss and E. Hancock. Multiple line-template matching with the em algorithm. *Journal of Pattern Recognition Letters*, 18(11-13):1283–1292, 1997. 13

[MS09]   Andriy Myronenko and Xubo B. Song. Image registration by minimization of residual complexity. In *In Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'09)*, pages 49–56, 2009. 70

[MS10]   A. Myronenko and X.B. Song. Point set registration: Coherent point drift. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'10)*, 32(12):2262–2275, 2010. 16

[MTT06]   S. Marsland, C. Twining, and C. Taylor. A minimum description length objective function for groupwise non-rigid image registration. *Journal of Image and Vision Computing*, 26:333–346, 2006. 30

[MTW09] B.C. Munsell, A. Temlyakov, and S. Wang. Fast multiple shape correspondence by pre-organizing shape instances. In *In Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'09)*, pages 840–847, 2009. 88

[MV98]   J.B. Maintz and M.A. Viergever. A survey of medical image registration. *Journal of Medical Image Analysis*, pages 1–37, 1998. 3, 8

[Ots79]   N. Otsu. A threshold selection method from gray-level histograms. *In Proceedings of IEEE Transactions on Systems, Man and Cybernetics (SMC'79)*, 9(1):62–66, 1979. 66

[PBIM05] H. Park, Peyton H. Bl, Alfred O. Hero Iii, and Charles R. Meyer. Least biased target selection in probabilistic atlas construction. In *In Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI'05)*, pages 419–426, 2005. 87

[PP93]   J. Pal and S. Pal. A review on image segmentation techniques. *Journal of Pattern Recognition*, 26:1277–1294, 1993. 13

[Pra74]   W. Pratt. Correlation techniques of image registration. *In Proceedings of Transactions on Aerospace and Electronic System*, 10:353–358, 1974. 18

[RMPA98]   A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *In Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, pages 1115–1124, London, UK, 1998. Springer-Verlag. 18

[ROC$^+$99]   N. Ritter, R. Owens, J. Cooper, R. Eikelboom, and P. Saarloos. Registration of stereo and temporal images of the retina. *In Proceedings of IEEE Transactions on Medical Imaging*, 18(5):404–418, May 1999. 19

[RSH$^+$99]   D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. In *In Proceedings of IEEE Transactions on Medical Imaging*, volume 18, pages 712–721, 1999. 68

[RU98]   U. Ruttimann and M. Unser. A pyramid approach to subpixel registration based on intensity. *In Proceedings of IEEE Transactions on Image Processing*, 7:27–41, 1998. 26

[SAM$^+$04]   D. Seghers, E.D. Agostino, F. Maes, D. Vandermeulen, and P. Suetens. Construction of a brain template from mr images using state-of-the-art registration and segmentation techniques. In *In Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI'04)*, pages 696–703, 2004. 88

[Sim96]   A. Simper. Correcting general band-to-band misregistrations. *In proceedings of International Conference on Image Processing (ICIP'96)*, B:597–600, 1996. 18

[Sze06]   R. Szeliski. Image alignment and stitching: a tutorial. *Joural of Foundations and Trends in Computer Graphics and Vision*, 2:1–104, 2006. 3

[TBS04]   A. Tonazzini, L. Bedini, and E. Salerno. Independent component analysis for document restoration. *International Journal on Document Analysis and Recognition*, 7(1):17–27, 2004. 53

[TCM⁺06]   C.J. Twining, T.F. Cootes, S. Marsland, V.S. Petrovic, R.S. Schestowitz, and C.J. Taylor. Information-theoretic unification of groupwise non-rigid registration and model building. *In Proceedings of Medical Image Understanding and Analysis*, 2:226–230, 2006. 29, 30

[TCS⁺00]   C.L. Tan, R.N. Cao, P.Y. Shen, Q. Wang, J. Chee, and J. Chang. Removal of interfering strokes in double-sided document images. In *In Proceedings of IEEE Workshop on Applications of Computer Vision (WACV'00)*, pages 16–21, California, December 2000. 72, 74

[TCS02]   C.L. Tan, R.N. Cao, and P.Y. Shen. Restoration of archival documents using a wavelet technique. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'02)*, 24(10):1399–1404, October 2002. 53, 71, 75

[Tik77]   A.N. Tikhonov. *Solutions of Ill Posed Problems (Scripta series in mathematics)*. Vh Winston, 1977. 25

[TSB07]   A. Tonazzini, E. Salerno, and L. Bedini. Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique. *International Journal on Document Analysis and Recognition*, 10(1):17–25, 2007. 53

[TT95]   O.D. Trier and T. Taxt. Evaluation of binarization method for document images. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'95)*, 17(3):312–315, 1995. 38

[TU98]   P. Thevenaz and M. Unser. An efficient mutual information optimizer for multiresolution image registration. *In Proceedings of International Conference on Image Processing (ICIP'98)*, 1:833, 1998. 19

[UAE91]   M. Unser, A. Aldroubi, and M. Eden. Fast b-spline transforms for continuous image representation and interpolation. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'91)*, 13(3):277–285, 1991. 26

[VS07]   A. Vedaldi and S. Soatto. A complexity-distortion approach to joint pattern alignment. *Journal of Advances in Neural Information Processing Systems*, 26:1425–1432, 2007. 31

[VZB98]   A. Vasileisky, B. Zhukov, and M. Berger. Automated image coregistration based on linear feature recognition. In *In Proceedings of the Second Conference on Fusion of Earch Data*, pages 59–66, 1998. 13

[W. 86]   W. Niblack. *An Introduction to Image Processing*. Prentice Hall, 1986. 38

[WBT09]   J. Wang, M. Brown, and C L Tan. A fully automatic system for restoration of historical document images. In *In Proceed-*

ings of Twenty-first Innovative Applications of Artificial Intelligence (IAAI'09), 2009. 74, 75, 76

[WC97]   W. Wang and Y. Chen. Image registration by control points pairing using the invariant properties of line segments. *Journal of Pattern Recognition Letters*, 18(3):269–281, 1997. 13

[WCS09]   Q. Wang, L. Chen, and D. Shen. Groupwise registration of large image dataset by hierarchical clustering and alignment. In *In Proceedings of SPIE Medical Imaging*, 2009. 88

[WJC02]   C. Wolf, J.M. Jolion, and F. Chassaing. Text localization, enhancement and binarization in multimedia documents. *In Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, pages 1037–1040, 2002. 38

[WRSS96]   R. Wiemker, K. Rohr, R. Sprengel, and H. Stiehl. Application of elastic registration to imagery from airborne scanners. In *In Proceedings of Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS'96)*, pages 949–954, 1996. 23

[WS77]   V. Wie and M. Stein. A landsat digital image rectification system. *Journal of GeoEl*, 15(3):130–137, July 1977. 18

[WSYR83]   C. Wang, H. Sun, S. Yada, and A. Rosenfeld. Some experiments in relaxation image matching using corner features. *Journal of Pattern Recognition*, 16(2):167–182, 1983. 13

[WT01]   Q. Wang and C.L Tan. Matching of double-sided document images to remove interference, pattern recognition. In *In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision*

and Pattern Recognition (CVPR'01), pages 11–13, Hawaii, December 2001. 54, 74, 75, 76, 107

[ZF03] B. Zitova and J. Flusser. Image registration methods: a survey. *Journal of Image and Vision Computing*, 21(11):977–1000, October 2003. 1, 3, 8, 19, 23, 29

[ZLMGW05] L. Zollei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3d data. *In Proceedings of International Conference on Computer Vision (ICCV'05)*, pages 291–301, 2005. 29, 30, 31

Author Biography

Jie Wang is a Ph.D. candidate in the Department of Computer Science, School of Computing, National University of Singapore. Her research interests include document image restoration and recognition, information retrieval, computer vision, pattern recognition and medical imaging. During her Ph.D. candidature, Jie Wang has published the following papers:

1. **J. Wang** and C L Tan. *Non-rigid Registration and Restoration of Double-sided Historical Manuscripts*. In International Conference on Document Analysis and Recognition, ICDAR 2011, 18-21 Sept 2011, Beijing, China.

2. **J. Wang** and C L Tan. *Non-rigid Image Registration for Historical Manuscript Restoration*. In 20th International Conference on Pattern Recognition, ICPR 2010, August 23-26, 2010, Istanbul, Turkey.

3. Li S, T Gong, **J Wang**, R Liu, C L Tan, T Y Leong, B C Pang, C C T

Lim, C K Lee, *TBIdoc: 3D Content-based CT Image Retrieval System for Traumatic Brain Injury*. In SPIE Medical Imaging Conference, 13-18 Feb 2010, San Diego, CA, USA.

4. **J. Wang**, M. S. Brown and C L Tan. *A Fully Automatic System for Restoration of Historical Document Images*. In Innovative Applications of Artificial Intelligence, IAAI 2009, 14-16 July 2009, Pasadena, California.

5. **J. Wang**, M. S. Brown and C L Tan. *Automatic Corresponding Control Points Selection for Historical Document Image Registration*. International Conference on Document Analysis and Recognition, ICDAR 2009, 26-29 July, 2009, Barcelona, Spain.

6. **J. Wang**, M. S. Brown and C L Tan. *Accurate Alignment of Double-sided Manuscripts for Bleed-through Removal*. In 8th IAPR International Workshop on Document Analysis Systems, DAS 2008, 16-19 Sept 2008, Nara, Japan.

7. S. Lu, **J. Wang** and C L Tan. *Fast and Accurate Detection of Document Skew and Orientation*. In International Conference on Document Analysis and Recognition, ICDAR 2007, 23-26 Sept 2007, Curitiba, Brazil.