Robust Multimodal Image Registration Using Local Frequency Representations

Baba C. Vemuri[1], Jundong Liu[1], and José L. Marroquin[2]

[1] Department of CISE, University of Florida
Gainesville, Fl. 32611
vemuri|jliu@cise.ufl.edu
[2] CIMAT, Guanajuato 36000, Mexico
jlm@cimat.mx

Abstract. Fusing of multi-modal data involves automatically estimating the coordinate transformation required to align the data sets. Most existing methods in literature are not robust and fast enough for practical use. We propose a robust algorithm, based on matching local-frequency image representations, which naturally allow for processing the data at different scales/resolutions, a very desirable property from a computational efficiency view point. This algorithm involves minimizing – over all affine transformations – the integral of the squared error (ISE or L_2E) between a Gaussian model of the residual and its true density function. The residual here refers to the difference between the local frequency representations of the transformed (by an unknown transformation) source and target data. The primary advantage of our algorithm is its ability to cope with large non-overlapping fields of view of the two data sets being registered, a common occurrence in practise. We present implementation results for misalignments between CT and MR brain scans.

1 Introduction

Image registration is one of the most widely encountered problems in a variety of fields including but not limited to medical image analysis, remote sensing, satellite imaging, etc. Broadly speaking, image registration methods can be classified into two classes [10] namely, feature-based and direct methods. In the former, prominent features from the two images to be registered are matched to estimate the transformation between the two data sets. In the latter, this transformation is determined directly from the image data or a derived "image-like" representation of the same.

Several feature-based schemes exist in literature. We will not describe feature-based schemes here but simply refer the reader to the survey [6]. Feature-based approaches have one commonality, i.e., they need to detect landmark features in the images and hence the accuracy of registration is dictated by the accuracy of the feature detector.

Amongst the direct approaches, one straightforward direct approach is the optical flow formulation [10] which assumes that the brightness at corresponding

M.F. Insana, R.M. Leahy (Eds.): IPMI 2001, LNCS 2082, pp. 176–182, 2001.

points in the two data to be registered is the same, an assumption that is severely violated in multi-modal images. The most popular direct approach is based on the concept of maximizing mutual information (MI) reported in [12,3]. Reported registration experiments in these works are quite impressive for the case of rigid and affine motions. Most MI-based algorithms in literature have been formulated for global parameterized motion with the exception of the work reported in Meyer et al., [8] wherein affine transformations as well as thin-plate spline warps are handled. The reported CPU execution times are quite high – of the order of several hours for estimating thin-plate warps. The problem of being able to handle local deformations in a mutual information framework is a very active area of research and some recent papers reporting results on this problem are [5,1]. For an exposition on other direct methods, we refer the reader to the survey [6] and some recent work on level-set methods in [11]. Most direct methods are known to be sensitive to outliers in data and this motivates us to seek a statistically robust scheme. Situations involving outliers arise when the field of view (FOV) of the data sets does not have a significant overlap.

In this paper, we develop a multi-modal registration technique which is based on a *local frequency* representation of the image data obtained by Gabor filtering the image tuned to a certain frequency and then computing the gradient of the phase of the filtered image. This representation is relatively invariant to changes in contrast/brightness. Because, multi-modal images of the same underlying object have very different intensities, a local frequency based representation seems apt to capture the salient commonalities between the two modalities of data. After computing the local frequency representations of the input images, a registration that matches these representations best is determined. To achieve this, we minimize a robust match measure which is based on the integral of the squared error (ISE or L_2E) between a Gaussian model of the residual and its true density function [9]. The residual here refers to the difference between the local frequency representations of the transformed source and target data. This robust estimation framework affords the advantage of not having to deal with additional tuning parameters that would be required when using M-estimators to achieve robustness. *One of the key strengths of our method is that – due to the formulation being set in a robust framework – it can handle data sets that do not have significant overlap as is the case in most practical situations.*

2 Local Frequency Representation

Tomes of research has been reported on retrieving information that is shared by the multi-modal data sets while reducing or eliminating the (imaging) sensor dependent background information. In this paper, we use a *local frequency* image representation obtained by Gabor filtering the input images and then computing the gradient of the phase. This local frequency representation has all the advantages of an "energy" representation and additionally can be tuned to any desired frequency/orientation thereby facilitating control of the alignment process. In this paper, we automatically select a bank of frequency and orientation