Efficient and Robust Model-to-Image Alignment using 3D Scale-Invariant Features

Matthew Toews[a], William M. Wells III[a,b]

[a] *Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA*
[b] *Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Building 32 32 Vassar Street Cambridge, MA 02139, USA*

Abstract

This paper presents feature-based alignment (FBA), a general method for efficient and robust model-to-image alignment. Volumetric images, e.g. CT scans of the human body, are modeled probabilistically as a collage of 3D scale-invariant image features within a normalized reference space. Features are incorporated as a latent random variable and marginalized out in computing a maximum a-posteriori alignment solution. The model is learned from features extracted in pre-aligned training images, then fit to features extracted from a new image to identify a globally optimal locally linear alignment solution. Novel techniques are presented for determining local feature orientation and efficiently encoding feature intensity in 3D. Experiments involving difficult magnetic resonance (MR) images of the human brain demonstrate FBA achieves alignment accuracy similar to widely-used registration methods, while requiring a fraction of the memory and computation resources and offering a more robust, globally optimal solution. Experiments on CT human body scans demonstrate FBA as an effective system for automatic human body alignment where other alignment methods break down.

Email addresses: mt@bwh.harvard.edu (Matthew Toews), sw@bwh.harvard.edu (William M. Wells III)

Corresponding Author: Matthew Toews
Phone: 617-732-6536, Fax: 617-732-7963
Email: mt@bwh.harvard.edu
Address: Surgical Planning Laboratory, Dept. of Radiology, Brigham and Women's Hospital, 75 Francis St., Boston, MA, USA, 02115

Keywords: 3D scale-invariant feature, orientation assignment, feature descriptor, probabilistic model, image alignment.

1. Introduction

Many medical imaging applications involve aligning 3D volumetric images, e.g. CT volumes of the human body or MR images of the human brain, to a standard frame of reference or model. For example, comparative studies of biomedical image data often require aligning images of different subjects to a normative atlas [46, 7], and model-to-image alignment can serve as the basis for volumetric object identification [23, 30] or computer-assisted diagnosis [42, 77]. Alignment is typically addressed via image registration or model fitting methods, where an optimal spatial transform is determined between an image and a template [20, 61, 36, 6, 41] or model [22, 23, 72], based on a combination of image similarity and geometrical constraints. Volumetric image alignment has been the focus of an important body of research for several decades, and a number of solutions now exist that are effective for analysis of well-curated data. For example, brain images of healthy subjects acquired in prospective studies can be aligned via diffeomorphic warps guaranteeing smooth, one-to-one correspondence between different subjects [6, 40].

The limitations of existing approaches become apparent when considering applications that require robust, efficient alignment of difficult data. For example, increasing numbers of publicly available human body scans [38, 19] may potentially lead to important developments in analyzing and classifying disease, possibly via large-scale data mining or content-based image retrieval methods [25, 44, 1] similar to those that have emerged in the context of 2D image data [58, 72]. Volumetric alignment remains a primary bottleneck, however, for several reasons. The human body varies greatly, due to deformations such as scaling, articulation, breathing motion, etc., but also due to the fact that the same structure may not be identifiable or present in all images, e.g. in the cases of pathology, variable organ configurations, bowel contents, partially overlapping or truncated images, etc. Accurate alignment throughout the entire image is difficult to achieve in such cases, motivating the need for robust, meaningful alignment solutions despite a lack of one-to-one correspondence. Alignment routines typically operate on entire image volumes, requiring on the order of minutes to compute a single robust linear alignment solution on an average PC. Storing, transmitting and aligning

large numbers of image volumes imposes a significant burden on memory, bandwidth and computational resources, motivating the need for improved efficiency. Algorithms often require initialization within a 'capture radius' of the correct solution, where initialization is typically provided via external labels, e.g. the body part and patient orientation as encoded in the Digital Imaging and Communications in Medicine (DICOM) protocol. DICOM labeling error rates as high as 15% have been reported [33], however, motivating the need for globally optimal alignment of image content independent of external labels.

This paper proposes a method for efficient, robust volumetric model-to-image alignment, entitled feature-based alignment (FBA). In FBA, all image data are represented as 3D scale-invariant features: distinctive, local image patterns characterized geometrically in terms of location, scale and orientation within the image, and equipped with an encoding of local image appearance. This representation is particularly useful for alignment, as features arising from the same anatomical tissues be automatically extracted and identified in different images despite global variations in image geometry and intensity, e.g. due to misalignment, sensor non-uniformity, etc. FBA models the posterior probability of a transform aligning an image to a standard reference frame or atlas, conditional on features extracted in the image. The model incorporates a latent variable enumerating a set of features characteristic of an imaging domain, e.g. features arising from ventricles or gyri in MR brain images. Model features are characterized by their occurrence probability and by conditional densities over feature appearance and geometry within atlas space, and are marginalized out in computing a maximum a-posteriori (MAP) alignment solutions. The model is trained from pre-aligned image data, automatically identifying a set of features most characteristic of anatomical structure present in training images. Efficient, globally optimal alignment is then achieved from a sparse set of highly probably feature correspondences between the model and a new image. Alignment is particularly robust to missing or unrelated image structure, and produces a set of model-to-image correspondences that can be used to initialize more detailed registration, segmentation or analysis procedures. FBA is generally applicable to a variety of imaging domains, and is efficient in both execution time and memory footprint.

The primary contribution of this article is a novel model for volumetric image alignment based on 3D scale-invariant image features. Scale-invariant features are widely used for robust, efficient matching [48, 51, 70, 86] and

appearance modeling [28, 73, 74] of 2D image data. 3D scale-invariant features extracted from volumetric data have been used for image matching [2, 16, 30] and classification [77], however leveraging 3D orientation information in determining feature geometry and efficiently encoding image intensity has proved challenging. We present novel techniques for assigning feature orientation and encoding intensity, both of which are necessary to achieve efficient, globally optimal alignment. Experiments investigate model-to-image alignment of MR brain images and CT images of the human body. In the case of brain images, FBA achieves alignment accuracy similar to established registration algorithms a fraction of the cost in terms of computation and memory, while offering a more robust, globally optimal solution. FBA is demonstrated for robust, global alignment of human body CT image volumes, in difficult scenarios including arbitrary image truncation and inter-subject variability where other alignment routines break down.

1.1. Prior Work

This article presents a general model for fast and robust volumetric image alignment, based on 3D scale-invariant image features. Here we review prior work on volumetric model-to-image alignment and the use of invariant image features.

1.1.1. Model-to-Image Alignment

Model-to-image alignment seeks identify a mapping between a model defined in particular imaging domain, e.g. brain scans, and a new image. Models used for alignment generally combine appearance and geometrical information within a standard geometrical frame of reference, e.g. the Talairach stereotaxic brain reference space [71]. Domain-specific image features and alignment strategies can be used, e.g. specialized detectors for specific structures such as the midsagittal plane of the brain [12, 59], bones [7], branch points in lung airways [46], retinal vasculature structures [70], etc. Such strategies are difficult to generalize to new imaging domains, however, as optimal features for alignment may not be known a-priori and are prone to inter-rater variability. Furthermore, alignment tends to break down if structures of interest are not present in new images, e.g. due to image truncation.

Image registration can be used as a general alignment solution, where the model is represented as an average template [20, 6] or template collection [11]. Registration is effective in aligning images related by a smooth, one-to-one mapping, e.g. inter-subject alignment of healthy brain images [40]

or intra-subject lung image registration [54]. It is difficult to generalize registration to scenarios where a smooth one-to-one mapping may not apply, for example in the case of abnormality, outliers or missing data. Outlier modeling strategies have been proposed [63], however coping with significant truncation or multiple modes of anatomical morphology remains challenging. Robust image-to-image matching techniques can be used to overcome sensitivity to outliers and missing data. The FLIRT (Functional Linear Image Registration Tool) algorithm [36] involves a coarse-to-fine strategy using the robust mutual information similarity measure. Block matching methods [61, 18] achieve robust alignment by matching subsets of image data, e.g. the robust block matching (RBM) algorithm [61] computes alignment from a subset of blocks with a best image matching score, via a multi-resolution matching strategy with pre-defined block sizes and resolutions. Registration formulations focusing on distinctive local features [3, 86] or emphasizing salient image regions [60] can improve robustness to outliers. A potential weakness of registration and matching techniques is that they do not explicitly learn and exploit domain-specific image structure, and thus may be prone to aligning unrelated image content.

Model-based alignment involves learning a domain-specific model from training examples that can then be reliably fit to new images. Global models of image appearance [8, 79] or combined appearance and geometry [22] capture data variation efficiently using principal component analysis (PCA). As with image registration, robust modeling strategies have been proposed [9], however outliers, localized variation or missing data remain challenging. Modeling the image as a collection of local parts or features provides a mechanism for coping with localized variations and occlusions. Parts-based models can be described in terms of the modeling of inter-feature geometrical relationships. Naive Bayes models consider features as conditionally independent given a single feature or reference frame [83, 45, 73, 74], and lead to efficient, occlusion-resistant algorithms for recovering low-parameter image deformations, e.g. linear transforms. Markov models impose constraints between neighboring features and can represent deformable or articulated transforms [29, 27, 28, 64, 87], however computational efficiency and occlusion resistance may be more difficult to achieve.

The Haar wavelet-based strategy popularized in the context of face detection [81] is an important local feature modeling technique. Relatively uninformative Haar wavelet features are computed efficiently via the integral image representation, and feature selection techniques such as AdaBoost [31]

are used to combine features into a strong classifier. These models are typically used to reproduce manually assigned labels or landmarks, e.g. fetal anatomical measurements in ultrasound images [14], brain structure segmentation [84] or annotations in radiographs [72]. There two primary disadvantages of this approach. First, while translation and scaling can be recovered efficiently, recovering rotation requires an explicit search over orientation and increases the computational complexity and error rate of fitting. Effectively coping with a single orientation parameter is challenging in 2D image data [14, 72], difficulties are more pronounced in higher dimensions [23]. Secondly, model training requires a high degree of supervision, and anatomical landmarks or structures of interest need to be specified manually, however it may not necessarily be obvious which structures are most useful for alignment.

1.1.2. Modeling Local Invariant Features

Local image features offer a mechanism for representing image appearance in a manner robust to local variations, occlusions and missing structure, and are widely used in the computer vision [32, 83, 81, 28, 74] and medical imaging communities [3, 66, 62, 70, 73, 77, 86]. Early work proposed local interest operators for identifying salient points for image matching in the context of binocular stereo vision [50] and robotic mapping [53]. Subsequent work proposed detectors for corners [34] and landmarks [3, 66] based on spatial derivatives. Local feature geometry is useful in characterizing image patterns in a manner independent of image intensity, e.g. via landmarks [62] or point sets [37]. Image intensity information associated with features can be encoded for efficient image correspondence [78, 67, 80]. Scale-space theory extends interest operators to consider the size or scale of image patterns in addition to location [47, 68], as the notion of an image pattern is intimately linked to the scale at which it is observed. Local invariant image features have been widely adopted for image matching in the computer vision community [48, 51], e.g. the scale-invariant feature transform (SIFT) method [48], and are an attractive basis for modeling. They can be repeatably extracted and used to compute correspondence in the presence of global geometrical similarity or affine deformations [52] and intensity changes.

Scale-invariant feature extraction typically involves a search for image regions that maximize a criterion of saliency, for example the magnitude of Gaussian derivatives in scale [48] and/or space [51], image phase [15] or information-theoric measures such as entropy [39] or mutual information [76].

Once regions have been identified, they can be assigned an orientation in order to achieve invariance to image rotation. In 2D images, this is typically done by identifying dominant gradient orientations within the image region. The method of Lowe involves generating a 1D histogram of image gradient orientations, smoothing the histogram to reduce noise and finally identifying histogram maxima [48]. With location, scale and orientation identified, features can be spatially normalized and encoded for efficient image-to-image matching. Due to normalization, matching can be accomplished without requiring an explicit search over geometrical transformations under which features are invariant. In conventional 2D images, the image gradient orientation histogram (GoH) representation, popularized by the SIFT descriptor [48], has been shown to be among the most effective encoding strategies for image-to-image matching [51]. Gradient orientation samples are set into coarse histogram bins over space and orientation, providing a representation that is informative regarding yet robust to geometrical image deformations. Rank-ordering of the GoH descriptor further improves the encoding performance, providing invariance to arbitrary monotonic deformations of image gradients [75].

In the context of volumetric images, 3D scale-invariant feature methods have been used by several authors in the context of intra-subject image matching [2, 56, 57, 16, 30]. They have not been widely adopted in more difficult contexts, for example inter-subject or atlas-to-subject alignment, as popular 2D methods for orientation assignment and intensity encoding do not generalize trivially to 3D. The primary difficulty is that 3D orientation is characterized by 3 parameters, e.g. azimuth, elevation and tilt angles, instead of a single parameter. When images do not exhibit significant orientation changes, orientation assignment can be neglected [77, 43] or partially characterized [56, 57, 16]. Orientation assignment is required for efficient alignment in the presence of orientation changes, however. Most authors have followed the approach of [69, 2, 30], whereby maxima are first identified in a 2D histogram of local gradient orientations parameterized by azimuth and elevation angles, after which an orthogonal tilt angle is determined. The azimuth/elevation parameterization is subject to sampling bias, as bins are non-uniformly sized. This can be partially mitigated by normalizing histogram bins according to solid angle area [69], however bias remains a difficulty for histogram smoothing and interpolation operations. Furthermore, the popular GoH intensity encoding is difficult to generalize to 3D due to the curse of dimensionality. For example, the 2D SIFT descriptor [48] ac-

cumulates gradient orientations over 4^2 spatial bins and 8 orientation bins, resulting in a 128-element vector. Approaches extending this quantization to 3D have resulted in large descriptors with 2000+ elements [2, 57, 30], which introduce a significant burden in memory and computational resources.

2. Material and Methods

Feature-based alignment aims to identify a globally optimal spatial mapping between a volumetric image and an atlas or model, based on a sparse set of scale-invariant feature correspondences. This section describes FBA, including local invariant feature extraction, a probabilistic model that can be learned from training images, and a technique for robust model-to-image alignment.

2.1. Invariant Feature Extraction

The goal of invariant feature extraction is to identify and characterize informative image patterns in a manner independent of global variations in image geometry and appearance, e.g. due to misalignment or intensity changes. A scale-invariant feature in 3D is defined geometrically by a scaled local coordinate system S within image I. Let $S = \{X, \sigma, \Theta\}$, where $X = \{x, y, z\}$ is 3-parameter location specifying the origin, σ is a 1-parameter scale and $\Theta = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\}$ is a set of three orthonormal unit vectors $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ specifying the orientations of the coordinate axes.

Invariant feature extraction begins by identifying a set of location/scale pairs $\{(X_i, \sigma_i)\}$ in an image. This is done by detecting spherical image regions centered on location X_i with radius proportional to scale σ_i that locally maximize a function $f(X, \sigma)$ of image saliency. For example, SIFT feature extraction detects local maxima of the difference-of-Gaussian (DoG) function [48]:

$$\{(X_i, \sigma_i)\} = \operatorname*{local\,argmax}_{X,\sigma} \left| f(X, \kappa\sigma) - f(X, \sigma) \right|, \tag{1}$$

where $f(X, \sigma)$ is the convolution of the image I with a Gaussian kernel of variance σ^2, κ is a multiplicative scale sampling rate, and the expression 'local $\operatorname*{argmax}_X\{f(X)\}$' denotes a set of values of the argument X that locally maximize $f(X)$. DoG region detection generalizes trivially from 2D to higher dimensions and can be efficiently implemented using Gaussian scale-space

pyramids [48]. Following detection, each region is assigned an orientation Θ, after which image intensity is encoded. Orientation assignment and intensity encoding in 3D are non-trivial extensions of 2D techniques, the following sections describe the methods adopted here.

2.1.1. Orientation Assignment

Orientation assignment involves determining the local coordinate system axes Θ for each location/scale pair (X, σ) identified via feature detection. This can be done by maximizing histograms of image gradient orientation computed within the image region surrounding X. Histograms based on angular parameterizations such as azimuth/elevation [69] are affected by sampling bias due to non-uniform bin geometry. Bias is avoided by here by adopting a non-parametric 3D histogram of gradient orientation, as follows. For each pair (X, σ), a cubical image patch centered on X with side length proportional to σ is cropped and rescaled to a fixed size. Image gradient samples are then computed at each voxel within the inscribed sphere of the patch, and used to populate bins on the surface of a unit sphere in a 3D gradient orientation histogram. For each gradient sample ∇I, the histogram bin location is determined by a unit vector of gradient orientation $\hat{\nabla} I = \frac{\nabla I}{||\nabla I||}$, and bins are incremented by the gradient magnitude $||\nabla I||$. Note that the orientation histogram is sparsely populated, as only bins intersected by the unit sphere surface contain non-zero counts. Trilinear interpolation is used to spread increments over neighboring bins, and the histogram is smoothed via a standard 3D Gaussian kernel in order to reduce noise, approximating smoothing on a unit sphere [17]. Note that while operations such as interpolation, incrementing and smoothing operations may be affected by minor discretization effects due to volumetric histogram binning, they are unaffected by angular parameterization bias.

A three-step orientation assignment procedure is used to identify Θ from 3D orientation histograms, as illustrated in Figure 1 a). First, a primary orientation vector $\hat{\theta}_1$ is identified as the maximum of a 3D histogram $H_1(\hat{\theta})$ generated from gradient samples ∇I:

$$\hat{\theta}_1 = \underset{\hat{\theta}}{\operatorname{argmax}} \left\{ H_1(\hat{\theta}) \right\}. \tag{2}$$

Next, a secondary orientation vector $\hat{\theta}_2$ is determined in the great circle orthogonal to $\hat{\theta}_1$. This is done by maximizing a second histogram $H_2(\hat{\theta})$

generated from the components of gradient samples orthogonal to $\hat{\theta}_1$, i.e. defined by the operation $\hat{\theta}_1 \times (\nabla I \times \hat{\theta}_1)$:

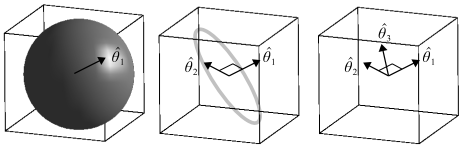
$$\hat{\theta}_2 = \underset{\hat{\theta}}{\operatorname{argmax}} \left\{ H_2(\hat{\theta}) \right\}. \tag{3}$$

Finally, the remaining orientation vector $\hat{\theta}_3$ is defined as the cross product $\hat{\theta}_3 = \hat{\theta}_1 \times \hat{\theta}_2$. Note that quadratic interpolation is used to refine histogram maxima locations to sub-bin precision.

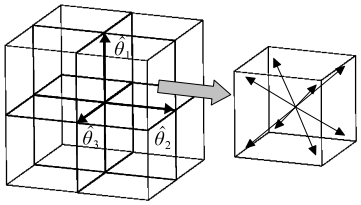
In general, orientation histograms for a single region (X, σ) may contain multiple informative maxima. Alternatively, orientation may be ambiguous for either $\hat{\theta}_1$ and/or $\hat{\theta}_2$, e.g. for spherical image patterns where unique orientations cannot be reliably assigned. As in 2D orientation assignment [48], these cases are handled by generating multiple features for each region from multiple primary and secondary orientation maxima, while limiting the number of features due to ambiguous maxima. For each region, a new feature is generated for each primary histogram peak within $0.8\times$ the maximum primary peak, and for each secondary peak with $0.8\times$ the maximum secondary peak. This results in an average of approximately 3 features per region in MR and CT volumes used later in experiments, which allows the representation of multi-modal local orientation structure while limiting multiple peaks due to orientation ambiguity.

2.1.2. Image Intensity Encoding

Once feature geometry is determined, a volumetric image patch is cropped, rescaled and reoriented according to the local coordinate system defined by $\{X, \sigma, \Theta\}$, after which intensities are encoded for subsequent matching. The goal of encoding is to convert volumetric intensity data into a small and distinctive representation for computing correspondence efficiently. We adopt a gradient orientation histogram (GoH) encoding that quantizes space and orientation uniformly into octants, resulting in a 64-bin histogram indexed by $2^3 = 8$ spatial bins and $2^3 = 8$ orientation bins, as illustrated in Figure 1 b). Gradient samples computed within the image patch are used to populate the histogram bins. For each gradient sample, the spatial histogram bin coordinate is determined by the sample location, the orientation bin coordinate is determined by the gradient vector orientation, and the bin is incremented by the gradient magnitude. Finally, the histogram is normalized by rank-ordering [75], resulting in a 64-element descriptor consisting of a permutation of the integers $\{1, \ldots, 64\}$.



a) Orientation assignment



b) Intensity encoding

Figure 1: Diagram a) illustrates the bin structure of the 3D volumetric histogram during orientation assignment, shaded regions indicate non-zero histogram bins. From left to right, the primary orientation vector $\hat{\theta}_1$ is first determined, followed by a perpendicular secondary orientation $\hat{\theta}_2$ and finally $\hat{\theta}_3$. Diagram b) illustrates the GoH image intensity encoding in 3D that quantizes gradient samples uniformly in 8 spatial bins and 8 orientations bins.

The encoding proposed here is attractive for several reasons. Spatial histogram weighting schemes used in 2D encoding are unnecessary [48], as spatial bins are distributed symmetrically about the origin. The octant sampling scheme generalizes to arbitrary spatial dimensions by considering orthants, where vectors in N-dimensions are of length $2^N \times 2^N = 2^{2N}$. Dissimilarity measures such as the L^2 distance can be used to compare rank-ordered vectors in a manner invariant to monotonic deformations in gradient magnitude [75]. Furthermore, rank-ordered vectors are small, requiring 48 bytes, and lead to efficient storage and matching.

2.2. Feature-based Model for Alignment

Let I_j represent the intensity encoding associated with feature geometry S_j, and let $\mathcal{I} = \{(I_j, S_j)\}$ represent a set of feature appearance/geometry pairs extracted in a single image I. Furthermore, let T represent a global linear transform mapping image I to an atlas or standard reference space, about which non-linear deformations can be identified and quantified locally. FBA aims to estimate the maximum a-posteriori (MAP) transform $T_{MAP} = \underset{T}{\text{argmax}} \{p(T|\mathcal{I})\}$ maximizing the posterior probability of T conditional on

$\mathcal{I}$. Assuming conditionally independent feature observations (I_j, S_j), Bayes rule results in:

$$p(T|\mathcal{I}) = \frac{p(T)\prod_j p(I_j, S_j|T)}{p(\mathcal{I})}. \tag{4}$$

In Equation (4), $p(\mathcal{I})$ is constant as $\mathcal{I}$ are fixed data during alignment and $p(T)$ is a prior density over transform parameters. Factor $p(I_j, S_j|T)$ is a density over the appearance and geometry of an observed feature conditional on T, and is constructed by marginalizing over feature data extracted from training images as follows. Images are naturally described in terms of distinctive local structure shared across similar images, for instance the human brain can be described by the corpus callosum, ventricles, etc. Distinctive structure is incorporated here by defining density $p(I_j, S_j|T)$ via the marginalization of a latent discrete random variable f taking on values $\{f_0, \ldots, f_N\}$:

$$p(I_j, S_j|T) = \sum_i^N p(I_j, S_j, f_i|T) = \sum_i^N p(I_j, S_j|f_i, T)p(f_i). \tag{5}$$

In Equation (5), f_i can be viewed as an indicator for a specific image pattern or model feature i that occurs with stable appearance and geometry across a population. $p(f_i)$ quantifies the probability of observing model feature f_i; we assume statistical independence of f and T. $p(I_j, S_j|f_i, T)$ is a joint density over observed feature data (I_j, S_j) conditional on model feature f_i and global transform T. This density can be thought of intuitively as quantifying the probability of feature (I_j, S_j) when brought into *correspondence* with f_i via transform T. For different distinctive model features, these densities are generally concentrated and non-overlapping, meaning that for a single fixed observed feature, $p(I_j, S_j|f_i, T) \approx 0$ for all but at most one f_i. A special case is reserved for f_0, which represents a generic, spurious background feature. The density $p(I_j, S_j|f_0, T)$ associated with f_0 by definition exhibits high variance in appearance and geometry, and is thus approximated as uniform or constant, i.e. $p(I_j, S_j|f_0, T) \approx \alpha$ for some positive constant α. As a result, Equation (5) is locally approximated as:

$$p(I_j, S_j|T) \approx p(I_j, S_j|f_i, T)p(f_i) + \alpha p(f_0), \tag{6}$$

In Equation (6), the value of $p(I_j, S_j|T)$ for a single fixed observed feature (I_j, S_j) is approximated by contributions from the background f_0 and at

most one distinctive model feature f_i. The approximation in Equation (6) is key to model learning and alignment procedures described later. Model feature-conditional densities $p(I_j, S_j|f_i, T)$ are defined as:

$$p(I_j, S_j|f_i, T) = p(I_j|S_j, f_i, T)p(S_j|f_i, T), \quad (7)$$
$$= p(I_j|f_i)p(S_j|f_i, T), \quad (8)$$

where Equation (7) follows from Bayes rule and Equation (8) from the assumption of conditional independence of I_j and (S_j, T) given f_i. $p(I_j|f_i)$ is taken to be an isotropic Gaussian density over intensity descriptor elements. Density $p(S_j|f_i, T)$ represents the local variation of feature geometrical variation associated with model feature f_i, given global transform T. This accounts for local deformations about T due to inter-subject variability, and is factored into conditional densities over feature location, scale and orientation:

$$p(S_j|f_i, T) = p(X_j|f_i, T)p(\sigma_j|f_i, T)p(\Theta_j|f_i, T). \quad (9)$$

In Equation (9), factor $p(X_j|f_i, T)$ is an isotropic Gaussian density over feature location conditioned on (σ_j, f_i, T). $p(\sigma_j|f_i, T)$ is a Gaussian density over log feature scale $\ln \sigma_j$ conditioned on (f_i, T). $p(\Theta_j|f_i, T)$ is a von Mises density [26] over independent angular deviations of coordinate axes, here approximated as an isotropic Gaussian density over Θ for simplicity under a small angle assumption. Note that all densities used here take the computational form of the isotropic Gaussian density, which for an n-dimensional random vector Y is parameterized by an n-dimensional mean vector μ and a scalar variance λ:

$$Y \sim \mathcal{N}(\mu, \lambda) = \frac{\exp(-\frac{1}{2\lambda}(Y-\mu)^T(Y-\mu))}{(2\pi\lambda)^{\frac{n}{2}}}. \quad (10)$$

Thus for densities over feature appearance and geometry in Equations (8) and (9), variables (Y, n) in Equation (10) take on the values of $(Y = I_j, n = 64)$, $(Y = X_j, n = 3)$, $(Y = \ln \sigma_j, n = 1)$ and $(Y = \Theta_j, n = 9)$.

2.2.1. Learning

Learning aims to identify a set of model features $\{f_i\}$ and to estimate associated distribution parameters from training images. Training images are first pre-aligned into a standard reference space and features are extracted as in Section 2.1. Pre-alignment fixes T as a constant in Equation (4), allowing

estimation of parameters for factors $p(S_j, I_j|f_i, T)$ and $p(f_i)$ in Equation (5) that takes the form of a Gaussian mixture model. Parameter estimation could potentially be achieved via clustering approaches such as expectation-maximization [24] or Dirichlet process modeling [65]. The number of model features N is generally large and unknown, however, making these approaches difficult to apply. A process similar to the mean-shift algorithm [21] is thus adopted to identify concentrated clusters of features extracted across subjects that are similar both in geometry and appearance, where features in a cluster represent instances of the same underlying anatomical structure or model feature f_i.

First, for each extracted feature index i, set G_i of geometrically similar features is identified such that:

$$G_i = \left\{ j : \begin{array}{c} \frac{||X_i - X_j||}{\sigma_i} \leq \varepsilon_x, \\ \left|\ln \frac{\sigma_j}{\sigma_i}\right| \leq \varepsilon_{\ln \sigma}, \\ \Theta_j \cdot \Theta_i \geq \varepsilon_{\cos \theta} \end{array} \right\}. \tag{11}$$

In Equation (11), ε_x and $\varepsilon_{\ln \sigma}$ are thresholds on the maximum acceptable difference in location and scale. $\varepsilon_{\cos \theta}$ is a threshold on the minimum acceptable angular difference cosine, where $\Theta_j \cdot \Theta_i$ represents the maximum dot product between corresponding orientation unit vectors. These three thresholds define a binary measure of inter-feature geometrical similarity, and can be set to fixed values that generally apply to all features.

Next, a set A_i of features similar in appearance to i is identified such that the L^2 distance $||I_i - I_j||$ is less than a threshold ε_{I_i}:

$$A_i(\varepsilon_{I_i}) = \{ j : ||I_i - I_j|| \leq \varepsilon_{I_i} \}. \tag{12}$$

Threshold ε_{I_i} is feature-specific, as the amount of appearance information varies from one feature to the next, and is maximized for each i as follows:

$$\varepsilon_{I_i} = \sup \left\{ \varepsilon_I \in [0, \infty) : \gamma \leq \frac{|G_i \cap A_i(\varepsilon_{I_i})|}{|\overline{G_i} \cap A_i(\varepsilon_{I_i})|} \right\}. \tag{13}$$

Features in intersection $C_i = G_i \cap A_i(\varepsilon_{I_i})$ are considered to be samples of f_i for which $p(I_j, S_j|f_i, T)p(f_i) \gg 0$ in Equation (6). In Equation (13), ε_{I_i} is thus maximized such that the relative probability of distinctive vs. background features $\frac{|G_i \cap A_i(\varepsilon_{I_i})|}{|\overline{G_i} \cap A_i(\varepsilon_{I_i})|} \approx \frac{p(f_i)}{p(f_0)}$ within $A_i(\varepsilon_{I_i})$ is greater than a positive

constant γ. Here, $\gamma = 1$ is empirically used, reflecting equal probabilities of distinctive vs. background features. Much smaller γ result in large ε_{I_i} and inefficient alignment, much larger γ result in small ε_{I_i} and inaccurate estimation of $p(I_j, S_j|f_i, T)p(f_i)$. Note the set in Equation (13) contains at least one element, i.e. feature i corresponding to $\varepsilon_I = 0$, and ε_{I_i} is thus always defined.

Note that learning produces many similar, redundant clusters, as each extracted image feature results in a cluster. Redundant clusters are discarded to obtain a reduced set of model features. Here, all clusters C_j associated with features j that are themselves elements of a larger cluster $j \in C_i, |C_i| > |C_j|$ are discarded. Finally, parameters of distributions in Equation (8) are estimated from remaining clusters. For densities over appearance and geometry associated with cluster C_i, mean parameters are taken to be the values of I_i and S_i, respectively. Density means, variances and discrete distribution parameters for $p(f)$ are estimated via maximum likelihood (ML).

2.2.2. Alignment

Alignment seeks to identify T_{MAP} maximizing $p(T|\mathcal{I})$ given feature data $\mathcal{I}$ extracted in a new image. A two-step algorithm similar to the Interpretation Tree approach [32] is adopted, where candidate image-model correspondences are first identified and used to generate transform candidates, after which candidate transforms are evaluated under $p(T|\mathcal{I})$. Candidate model-image correspondences (i, j) are determined by comparing the intensity encoding I_j of each extracted image feature to densities $p(I_j|f_i)$ associated with model features. A candidate correspondence (i, j) exists if $||I_i - I_j|| \leq \varepsilon_{I_i}$. This procedure can be efficiently performed in $O(MlogN)$ time complexity via approximate nearest neighbor techniques [10], where N and M are the numbers of model and image features respectively, and by considering a small subset of the most probable model features as determined by $p(f)$.

For each correspondence candidate (i, j), a model-image transform candidate T_{ij} is estimated via ML, i.e. $T_{ij} = \underset{T}{\text{argmax}} \{p(S_j|f_i, T)\}$. Transform candidates are then evaluated under $p(T|\mathcal{I})$ in order to determine T_{MAP}, as follows. For each transform candidate T_{ij}, all image features $\{S_j\}$ with model correspondences are transformed to normalized model space via T_{ij}. Image features whose transformed geometry falls within thresholds ε_x, $\varepsilon_{\ln\sigma}$ and $\varepsilon_{\cos\theta}$ of their corresponding model features are considered to be *inliers* of $T_{i,j}$, as $p(I_j, S_j|f_i, T)p(f_i) \gg 0$, and used to compute $p(T|\mathcal{I})$ via

Equations (4) and (5). Integral computation in Equation (5) is efficient, as $p(I_j, S_j|f_i, T)p(f_i) \approx 0$ for non-inlier pairs (i, j) and these terms are discarded. The candidate transform maximizing $p(T|\mathcal{I})$ is taken as T_{MAP}.

Note that T_{MAP} represents a global linear model-to-image transform, while inlier model-image correspondences within geometrical thresholds of T_{MAP} account for local deformations. Correspondences could be used in order to estimate deformable alignment solution throughout the image, for example as an interpolated deformation field [5]. This is beyond the scope of this article and we leave this for future work, here least squares estimation (LSE) is used to derive a more accurate model-to-image similarity transform from correspondences associated with T_{MAP}. LSE is based on corresponding feature image location X, and thus requires a minimum of 3 non co-linear correspondences. Feature scale and orientation parameters could also be incorporated, however they tend to be more reflective of local image-specific variations, e.g. enlarged/reduced ventricle sizes in brain images, and less informative regarding the global model-to-image transform.

3. Results

Experiments here have two goals, the first is to demonstrate that FBA can be used to compute model-to-image alignment solutions in difficult cases with accuracy on par with state-of-the-art alignment approaches, at a fraction of the cost in terms of memory and computational time. The second is to demonstrate the ability of FBA to recover robust, globally optimal alignment solutions in scenarios where alignment approaches break down. Two distinct contexts are considered: T1-weighted MR images of the human brain in Section 3.2 and CT images of the human body in Section 3.3.

3.1. Experimental Details

Features are extracted from all images used in experiments as described in Section 2.1, our implementation is available online [3]. DoG extrema in Equation (1) are identified using a 3D equivalent of the Gaussian pyramid method of Lowe [48]. Briefly, the range of X spans the image volume, scale is initialized at $\sigma = 1.6$ voxels and increased at a multiplicative rate of three samples per scale doubling or octave, i.e. $\kappa = 2^{1/3}$. The image is

[3] 3D SIFT Implementation: www.spl.harvard.edu/publications/item/view/2267

sub-sampled by a factor of two at each octave for efficiency, and feature extraction terminates once 2σ exceeds the minimum image dimension. DoG extrema within 2σ of the image border are discarded, as they are subject to border effects. Extrema arising from degenerate image structure whose 3D location is unstable are also identified and discarded, e.g. planar or tubular structures. This is done using the method of Rohr [66], based on the 3×3 structure tensor or second moment matrix M generated from image gradient samples within the region (X, σ). Briefly, the ratio $\beta = \frac{Det(M)}{(\frac{1}{3}Tr(M))^3}$ calculated from the determinant $Det(M)$ and the trace $Tr(M)$ of M lies on the range $[0, 1]$, where low β indicate degenerate image structure. Here, all extrema for which $\beta < 0.2$ are discarded, lower thresholds result in similar alignment solutions but increased processing requirements, higher thresholds begin to negatively impact alignment.

Cubical image patches of side length 4σ are rescaled to 11x11x11 voxels for orientation assignment and intensity encoding. A side length of 4σ allows inclusion of informative peripheral image structure into the encoding of intensity. The size of the volumetric histogram used in orientation assignment is determined such that the number of gradient orientation samples is sufficient to populate histogram bins. The image patch contains an inscribed sphere of integer radius $\frac{11}{5} = 5$ voxels bearing $\frac{4}{3}\pi 5^3 \approx 523$ gradient samples, and a volumetric histogram of size 11x11x11 bins is used. This results in greater than one vote per bin on average, as votes are accumulated on a spherical shell of radius 5 bins with a surface area of $4\pi 5^2 \approx 314$ bins. The number of features extracted in an image is dependent on the content, here for example, a brain volume at 1mm voxel resolution produces approximately 2500 features.

Learning is applied as in Section 2.2.1 with geometrical thresholds empirically set to $\varepsilon_{\ln \sigma} = \ln 1.5$, $\varepsilon_x = 0.5$ and $\varepsilon_{\cos \theta} = 0.8$. These thresholds can be thought of as determining the maximum permissible geometrical variability of model features. Looser thresholds result in a smaller set of model features with higher geometrical variability, potentially grouping features arising from different underlying anatomical structures and leading to reduced alignment accuracy. Tighter thresholds result in a larger set of model features, with instances of the same anatomical structure potentially split into different clusters, resulting in reduced alignment efficiency. Here, geometrical thresholds are set to allow a high range of geometrical variability, as potential ambiguities in geometrical clustering are later resolved by appearance

clustering. Fitting is performed as in Section 2.2.2, where the output is the optimal global 7-parameter similarity transform T_{MAP}, in addition to a set of probable model-image correspondences representing localized deformations. Alignment is relatively insensitive to the background feature constant α and similar alignment solutions are achieved for a range of positive values, here $\alpha = 1$ is used.

3.2. MR Brain Images

The experiments here demonstrate FBA in the context of model-to-subject alignment of T1-weighted MR brain images. A brain model, or atlas, is constructed from a set of 100 healthy training subject images from the OASIS data set which have been aligned into a standard reference space at resolution $176 \times 208 \times 176$ (isotropic, 1.0mm voxel size) via a 12-parameter affine transform [49]. Note that learning is performed once off-line on a set of typical images of healthy subjects. A set of 12 difficult testing images (isotropic, voxel sizes 0.8-1.0mm) are used to demonstrate atlas-to-subject alignment, these are described in Figure 2.

FBA alignment is generally successful, identifying reasonable global alignment solutions despite arbitrary initialization and significant abnormality. On average 40 atlas-to-subject correspondences are identified per alignment trial, Figure 3 illustrates examples identified in test subject (10) despite an unknown global orientation difference. These correspondences represent a locally linear alignment solution between distinctive, healthy anatomical structure shared by the atlas and subject. The probabilities of corresponding structures are quantified locally throughout the image from distribution $p(f)$ estimated during learning, see the graph in Figure 3. Note that the tumor region produces no atlas-to-subject correspondences, as it bears no likeness to structure modeled in healthy training subjects, and thus does not impact alignment.

Alignment trials are also performed using three common robust image alignment algorithms: RBM [61] [4], FLIRT [36] [5] and ELASTIX [41] [6] based on the open source ITK [7] project. RBM, FLIRT and ELASTIX are image-based techniques, that operate by determining a linear transform between

[4]Implementation: NiftyReg, M. Modat, www.cs.ucl.ac.uk/staff/m.modat
[5]Implementation: 3D Slicer, C. Goodlet, www.slicer.org
[6]Implementation: ELASTIX, elastix.isi.uu.nl
[7]Insight Segmentation and Registration Toolkit, www.itk.org

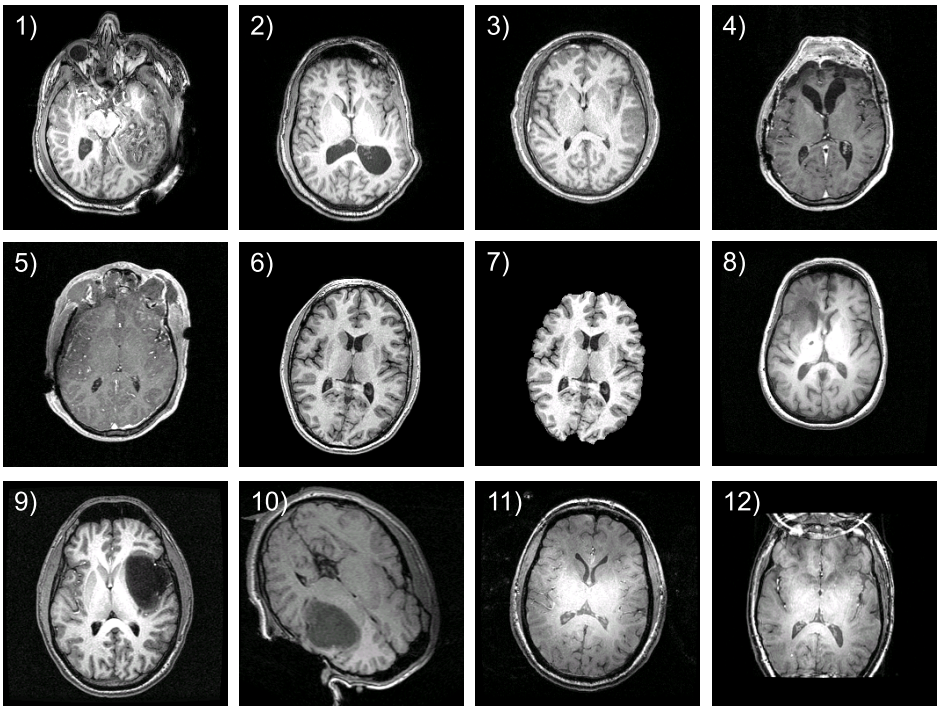


Figure 2: Axial slices of 12 test subject images. (1-5) exhibit significant anatomical abnormalities due to traumatic brain injury [55]. Subjects (6,7) represent a healthy subject before and after skull-stripping. (8,9) exhibit large tumors [4]. (10) is an intra-operative scan of subject (9) acquired at an unspecified oblique angle due to the surgical protocol [4]. (11,12) exhibit non-uniformity and wrap-around artifacts.

test subjects and an average atlas template image constructed from the 100 training subjects. RBM, FLIRT and ELASTIX recover a 12-parameter affine transform commonly used in atlas-to-subject brain analysis [41], RBM and ELASTIX also recover a 6-parameter rigid transform, which may be more robust for the difficult test images used here. FBA alignment is the most efficient of the techniques, both in terms of computational resources and memory. The approximate running times for one alignment trial on a 2.5GHz Intel Core 2 processor are (in seconds) FBA: 20 , ELASTIX: 33 , RBM: 140 and FLIRT: 220. In terms of the memory, FBA requires approximately 0.5MB to store both model and image features, while image registration requires approximately 70MB to store floating point images. In contrast to image registration algorithms, FBA focuses computational resources primarily on

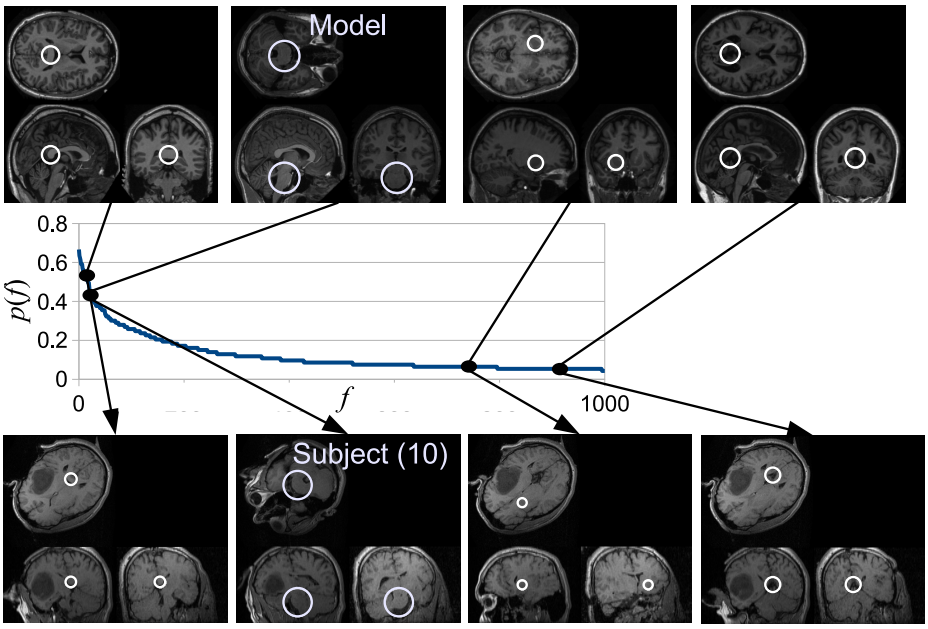


Figure 3: Examples of model-to-subject brain image correspondences across global orientation changes. The central graph plots the probability distribution $p(f)$ over model features, in descending order of probability. Images show a subset of 4 model features (upper row), their probabilities, and corresponding image features identified in subject 10 (lower row).

feature extraction in individual images prior to alignment. Virtually all FBA running time is due to Gaussian convolution during feature extraction, this represents a one-time pre-processing step that could be significantly reduced via GPU optimization. With features extracted, FBA alignment requires less than 1 second.

Algorithms are compared in terms of approximate alignment error, which is quantified here by the average displacement error of 6 anatomical surfaces in aligned subject images relative to their locations in the atlas. The anatomical surfaces are shown by arrows in Figure 4 'Atlas Labels', and consist of the inferior cortical surfaces of the left and right temporal lobes and the superior cortical surface (coronal slice) and the anterior, posterior and superior surfaces of the corpus callosum (sagittal slice). These structures are used as they are present in all testing images, they are manually labeled by a single rater. Note the goal here is not to quantify alignment error precisely, which in

itself would be challenging for these images, but rather to provide an approximate error comparison and to identify alignment breakdown. The graph in Figure 4 plots the alignment error for the six methods tested here. All techniques achieve nominally low alignment error (i.e. less than 5mm) for nine of twelve subjects, three subjects (1, 7 and 10) represent difficult exceptions. Subject 1 is a difficult abnormal case that results in high error, particularly for ELASTIX. Subject 7 illustrates the tendency of the affine alignment to systematically stretch the cortex of the skull-stripped subject to match the atlas skull, highlighting how registration may be erroneously driven by overall object shape rather than distinctive image structure. Note that this problem is less pronounced for rigid RBM and ELASTIX alignment. Subject 10 illustrates the capacity of FBA to recover global orientation changes without initialization, while other registration techniques fail without correct initialization. FBA aligns all test cases here with error less than 5mm.

3.3. CT Body Images

This section demonstrates FBA in a context where existing image alignment techniques break down, inter-subject matching of human body scans. The American College of Radiology Imaging Network (ACRIN) [38] data set consists of CT scans acquired for the study of colorectal neoplasia. The National Library of Medicine (NLM) Liver CT data set [19] contains scans acquired for the analysis of liver tumors. It may be of interest to align these images for a joint study of structure common to all images, e.g. lower lung, liver or kidneys. While brain image alignment is relatively common, alignment of human body scans remains an open and difficult problem, for a number of reasons. The human body exhibits a high degree of variability from one person to the next due to myriad factors including age, sex, morphology (lean, obese), breathing state (inspiration/expiration), body position (e.g. prone/supine), content of bowels, which influence the geometry of soft tissues. Body image alignment approaches in the literature typically assume images contain similar structure [7, 35], however this is generally not the case, as images acquired in studies focusing on different aspects may be cropped differently. This is particularly relevant in the case of the CT image modality, where truncated image acquisition reduces the exposure of the subject to harmful ionizing radiation.

Here, 20 random ACRIN subject images are chosen (isotropic, 1.56mm voxel size), aligned manually to a single subject according to major organ geometry, and used to train a feature-based model with identical parameters

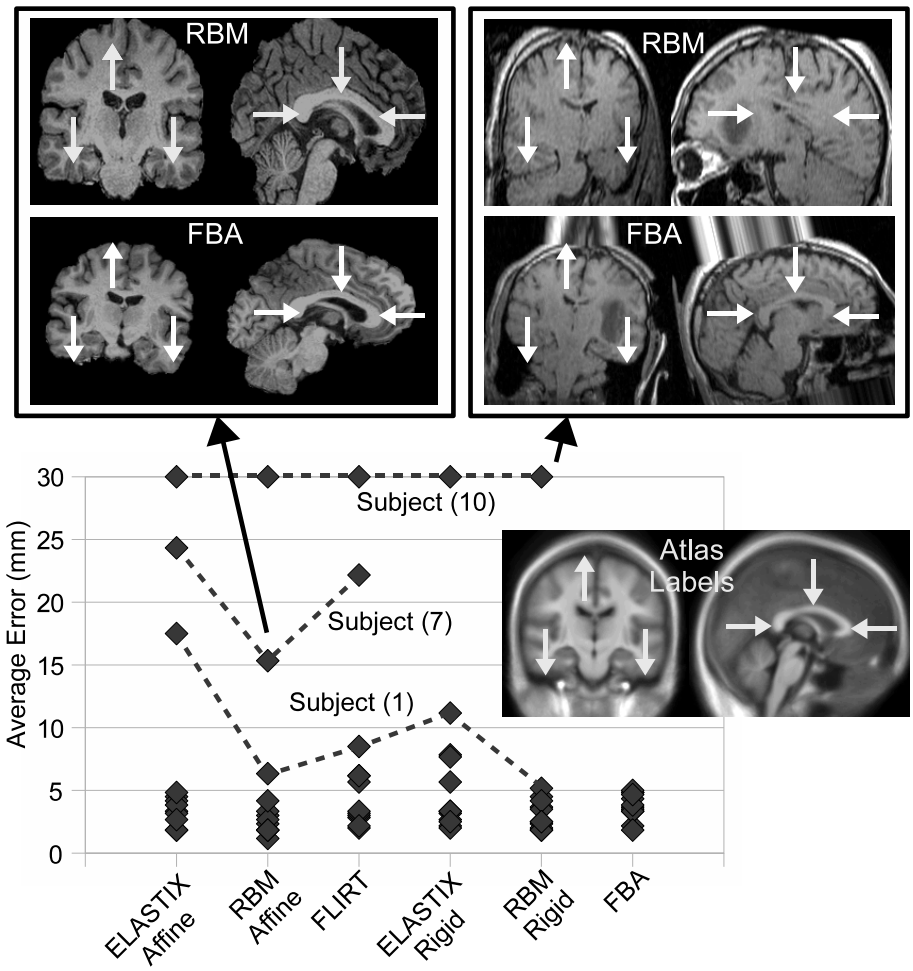


Figure 4: A graph of alignment error for ELASTIX and RBM (affine and rigid), FLIRT (affine) and FBA over 12 test subject images. The image overlaying the graph shows sagittal and coronal slices of the average atlas, arrows illustrate labels used to quantify alignment error. Dashed lines indicate error measurements for difficult subjects 1,7 and 10. The upper left and right images illustrate alignment for subjects 7 and 10 where RBM, FLIRT and ELASTIX error is noticeably higher than FBA.

to those used in brain alignment. The 4 test subjects from the NLM liver CT data set are then aligned to this model (isotropic, 1.24-1.37mm voxel size). The result of FBA alignment is shown in Figure 5, note that all sub-

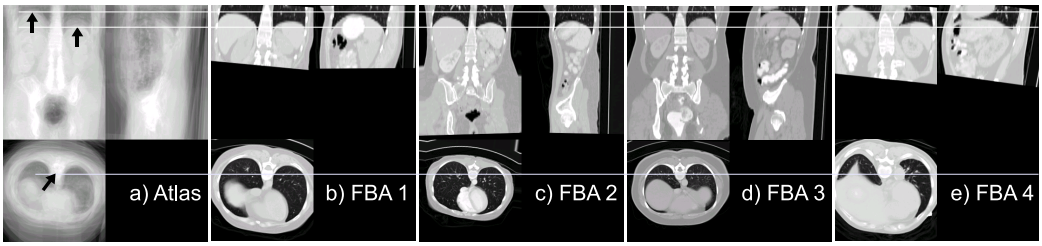


Figure 5: a) average ACRIN atlas, blurriness indicates the high degree of inter-subject variability. (b-e) NLM subjects (1-4) following successful FBA alignment, note inter-subject variability and image truncation. Arrows and lines across images indicate anatomical locations used to quantify alignment error.

jects are successfully aligned with respect to the atlas average, in particular subjects 1 and 4 with truncated abdominal regions. Approximate alignment error is quantified by the displacement of three anatomical landmarks that remain relatively stable across subjects, the superior aspect of the right and left hemidiaphram, and the location of the thoracic vertebrae, see arrows in Figure 5. An average error of 11 ± 3.2mm is obtained, which is reasonable considering the high degree of anatomical variability of the human body in the atlas.

Figure 6 illustrates example correspondences obtained in aligning truncated NLM subject 1 to the feature-based atlas, note the high degree of inter-subject variability. While it may be unrealistic to compute accurate alignment throughout the image due to inter-subject variability, truncation, etc, it may be possible within image regions associated with probable FBA correspondences. Figure 7 illustrates locally linear and deformable B-spline [8] alignment of corresponding feature regions identified in both a model and an NLM subject. The intensity difference $|I1 - I2|$ is noticeably less pronounced along tissue borders following deformable as opposed to linear alignment, indicating a more accurate alignment solution. The vertebrae of I2 are slightly stretched following deformation, however, and a more accurate alignment solution would account for bio-mechanical tissue properties such as rigid bone structure. This is a promising research direction that we leave for future work.

[8]Implementation: 3D Slicer, B. Lorensen, www.slicer.org

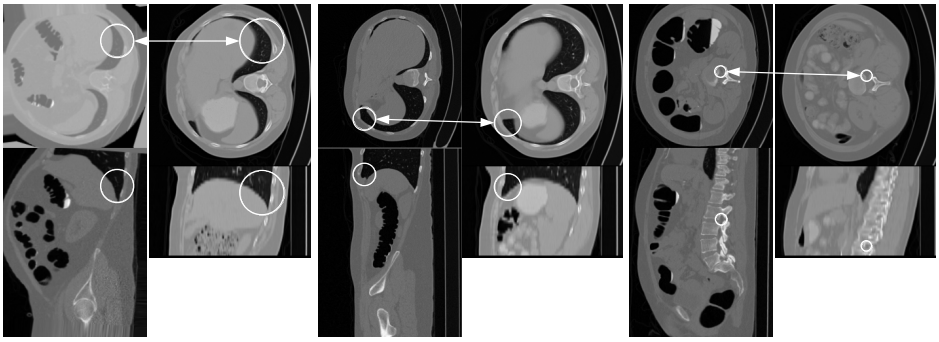


Figure 6: Examples of model-subject correspondences in CT volumes. Each image consists of axial (upper) and sagittal (lower) volume slices. White circles illustrate the location and scale of features in the ACRIN model image (left) and in the new subject image (right), arrows indicate corresponding features.

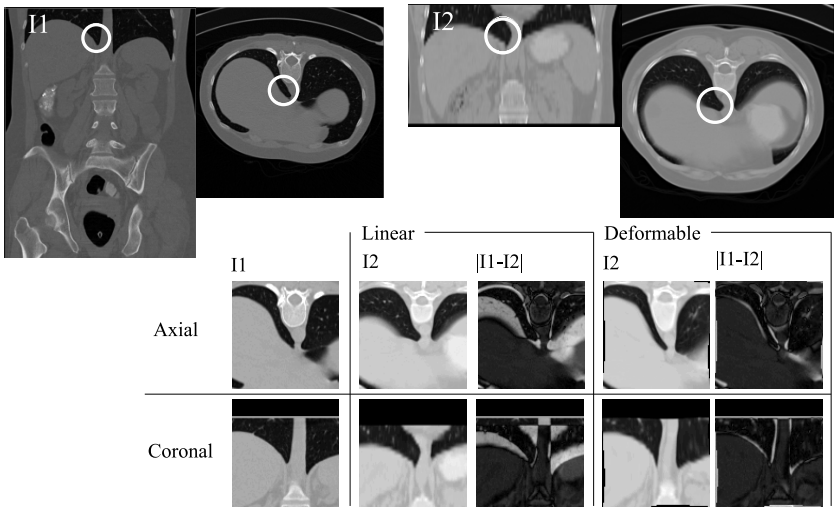


Figure 7: Deformable alignment of corresponding feature regions in CT volumes. The upper image pairs show axial and coronal slices of a feature identified in both an ACRIN model subject (left, I1) and an NLM subject (right, I2), white circles represent the location and scale of corresponding features. The lower images show fixed region I1 (left) and moving region I2 following local linear (center) and deformable (right) alignment in axial (upper row) and coronal (lower rows) views. Images $|I1 - I2|$ show the absolute difference between intensities in aligned regions, high intensity indicates alignment error.

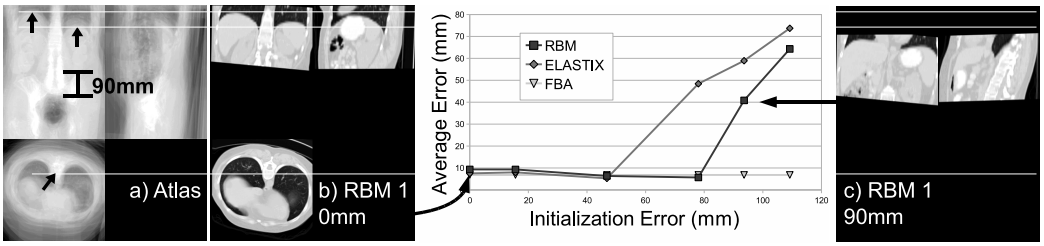


Figure 8: Alignment of NLM subject 1. The average atlas is shown in a). The graph shows average alignment error of FBA, RBM and ELASTIX associated with vertical initialization error. Images b) and c) show examples of successful and unsuccessful RBM alignment associated with vertical initialization errors of 0mm and 90mm, respectively.

To illustrate the difficulty of the task, alignment is also attempted using the FLIRT, RBM and ELASTIX algorithms. Trials are attempted using both the average ACRIN and individual ACRIN subjects as the atlas, with both subject-to-atlas and atlas-to-subject registration paradigms. Algorithm parameters are varied and the values resulting in the best alignment are considered. Alignment is generally unsuccessful, with all registration algorithms producing clearly incorrect, high error solutions. When manually initialized to the correct solution (translation, rotation, scaling), FLIRT remains unsuccessful, RBM and ELASTIX result in average errors of RBM: 13 ± 5.0mm, ELASTIX: 11 ± 3.1mm across the 4 NLM subjects. Figure 8 illustrates alignment error for FBA, RBM and ELASTIX as a function of initialization error in the form of vertical translation. ELASTIX and RBM both result in high error when initialized 80mm and 90mm from the correct solution, this poses a difficulty as protocols such as DICOM provide patient orientation but not translation within the volume. In contrast, FBA alignment is globally optimal and independent of initialization errors in location, orientation or scaling. Running times for alignment are approximately (in seconds) FBA: 28, ELASTIX: 36, RBM: 290.

4. Discussion

This article presents FBA, a method for efficient, robust and global model-to-image alignment of volumetric image data. FBA is based on a probabilistic model of 3D scale-invariant features, quantifying the appearance, geometry and occurrence probability of features in a set of training data. Novel tech-

niques for assigning feature orientation and encoding intensity are developed. MAP estimation is used to identify a globally optimal, locally linear alignment solution from correspondences between features extracted in a new image and the model. Experiments demonstrate FBA in the contexts of difficult abnormal MR brain images and highly variable CT body scans, and compare with three state-of-the-art image alignment algorithms: FLIRT, RBM and ELASTIX. In the context of brain MRI, FBA achieves alignment accuracy similar to other approaches, while offering more robust, global solutions at a fraction of the cost in computation and memory. FBA leads to a system for general alignment of CT body scans, capable of coping with significant inter-subject variability and truncation, where other approaches break down without manual initialization. A number of model-to-image alignment techniques were investigated other than FLIRT, RBM and ELASTIX, however results were either poor or required a high degree of manual supervision, highlighting the difficulty of the experimental data.

The most immediate practical use of FBA is as a means of efficiently and robustly initializing registration or segmentation procedures in domains where training images are available. Memory requirements are reduced by more than an order of magnitude in comparison to standard image registration. For example a 256x256x176 brain volume of size 11MB (floating point) can be represented by a 0.4MB feature set, and a 0.1MB feature-based brain atlas containing the 1000 most frequently occurring features can be used to successfully align the set of testing images used in Section 3.2. Furthermore, while feature extraction imposes a one-time computational cost, e.g. approximately 20 seconds for brain images, the resulting feature set can be aligned extremely efficiently in sub-second time. This will be particularly important in applications in which images are aligned multiple times, e.g. group-wise registration [85] or content-based image retrieval.

FBA leads to a number of future research directions. Feature correspondences may be useful in computing more accurate deformable or articulated model-to-image transforms throughout the image. Neighborhood constraints may prove effective at modeling features arising from repeating image structures, e.g. vertebrae or airway structures. While 3D scale-invariant feature correspondences can be identified over a range of affine and non-linear deformations, 3D affine-invariant features [52] may prove effective in coping with voxel anisotropy. Orientation-sensitive correspondences resulting from FBA could prove useful in analyzing and classifying subjects according to factors such as disease [77]. FBA could be used to align complementary data

sources, e.g. CT and positron emission tomography (PET) volumes [13], by developing inter-modality scale-invariant feature extraction and matching techniques. FBA could serve as a useful object detection framework for non-medical 3D scanning systems [30]. Finally, FBA could serve as an effective means of studying inter-species alignment or modeling growth in temporal image sequences, where one-to-one image correspondence may not exist throughout the image.

5. Acknowledgements

We would like to thank Dominik Meier, Ron Kikinis and Dan Kacher for helpful suggestions regarding this work. This work was supported by NIH grants P41-RR-013218, P41-EB-015902 and R01-HD-057963.

References

[1] Akgul, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., Acar, B., 2011. Content-based image retrieval in radiology: Current status and future directions. Journal of Digital Imaging 24 (2), 208–222.

[2] Allaire, S., Kim, J., Breen, S., Jaffray, D., Pekar, V., 2008. Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis. In: MMBIA.

[3] Amit, Y., Kong, A., 1996. Graphical templates for model registration. IEEE TPAMI 18 (3), 225–236.

[4] Archip, N., Clatz, O., Whalen, S., Kacher, D., Fedorov, A., Kot, A., Chrisochoides, N., Jolesz, F., Golby, A., Black, P., Warfield, S., 2007. Non-rigid alignment of pre-operative mri, fmri, and dt-mri with intra-operative mri for enhanced visualization and navigation in image-guided neurosurgery. Neuroimage 35 (2), 609–624.

[5] Arsigny, V., Pennec, X., Ayache, N., 2005. Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations - application to the registration of histological slices. MIA 9 (6), 507–523.

[6] Avants, B. B., Epstein, C. L., Grossman, M., Gee, J. C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. MIA 12 (1), 26–41.

[7] Baiker, M., Milles, J., Dijkstra, J., Henning, T. D., Weber, A. W., Que, I., Kaijzel, E. L., Lwik, C. W., Reibera, J. H., Lelieveldt, B. P., 2010. Atlas-based whole-body segmentation of mice from low-contrast Micro-CT data. Medical Image Analysis 14 (6), 723–737.

[8] Barber, D., 1976. Digital computer processing of brain scans using principal components. Phys. Med. Biol. 21 (5), 792–803.

[9] Beichel, R., Bischof, H., Leberl, F., Sonka, M., September 2005. Robust active appearance models and their application to medical image analysis. IEEE TMI 24 (9), 1151–1169.

[10] Beis, J. S., Lowe, D. G., 1997. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: CVPR. pp. 1000–1006.

[11] Blezek, D. J., Miller, J. V., 2006. Atlas stratification. In: MICCAI.

[12] Brummer, M. E., 1991. Hough transform detection of the longitudinal fissure in tomographic head images. IEEE TMI 10 (1), 74–81.

[13] Camara, O., Delso, G., Colliot, O., Antonio, M.-I., 2007. Explicit Incorporation of Prior Anatomical Information Into a Nonrigid Registration of Thoracic and Abdominal CT and 18-FDG Whole-Body Emission PET Images. IEEE TMI 26 (2), 164–178.

[14] Carneiro, G., Georgescu, B., Good, S., Comaniciu, D., 2008. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE TMI 27 (9), 1342–1355.

[15] Carneiro, G., Jepson, A., 2003. Multi-scale phase-based local features. In: CVPR. Vol. 1. pp. 736–743.

[16] Cheung, W., Hamarneh, G., 2009. N-sift: N-dimensional scale invariant feature transform. IEEE TIP 18 (9), 2012–2021.

[17] Chung, M. K., 2006. Heat kernel smoothing on unit sphere. In: ISBI.

[18] Clatz, O., Delingette, H., Talos, I.-F., Golby, A. J., Kikinis, R., Jolesz, F. A., Ayache, N., Warfield, S. K., 2005. Robust nonrigid registration to capture brain shift from intraoperative MRI. IEEE TMI 24 (11), 1417–1427.

[19] Cleary, K., 2010. National library of medicine: Liver ct. insightjournal.org/midas.

[20] Collins, D. L., Neelin, P., Peters, T. M., Evans, A. C., 1994. Automatic 3D inter-subject registration of MR volumetric data in standardized talairach space. Journal of Computer Assisted Tomography 18 (2), 192–205.

[21] Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE TPAMI 24 (5), 603–619.

[22] Cootes, T., Edwards, G., Taylor, C., June 2001. Active appearance models. IEEE PAMI 23 (6), 681–684.

[23] Criminisi, A., Shotton, J., Bucciarelli, S., 2009. Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. In: MICCAI Workshop on Probabilistic Models for Medical Image Analysis.

[24] Duda, R. O., Hart, P. E., Stork, D. G., 2001. Pattern classification, 2nd Edition. Wiley.

[25] El-Naqa, I., Yang, Y., Galatsanos, N., Nishikawa, R., Wernick, M., 2004. A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE TMI 23 (10), 1233–1244.

[26] Evans, Hastings, Peacock, 1993. Statistical Distributions, 2nd Edition. John Wiley and Sons.

[27] Felzenszwalb, P. F., Huttenlocher, D., 2005. Pictorial structures for object recognition. IJCV 61 (1), 55–79.

[28] Fergus, R., Perona, P., Zisserman, A., 2006. Weakly supervised scale-invariant learning of models for visual recognition. IJCV 71 (3), 273–303.

[29] Fischler, M., Elschlager, R., 1973. The representation and matching of pictorial structures. IEEE Trans. on Computers 22 (1), 67–92.

[30] Flitton, G., Breckon, T. P., Megherbi, N., 2010. Object Recognition using 3D SIFT in Complex CT Volumes. In: BMVC.

[31] Freund, Y., Schapire, R. E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55 (1), 119–139.

[32] Grimson, W. E. L., Lozano-Perez, T., 1987. Localizing overlapping parts by searching the interpretive tree. IEEE TPAMI 9 (4), 469–482.

[33] Guld, M. O., Kohnen, M., Keysers, D., Schubert, H., Wein, B., Bredno, J., Lehmann, T. M., 2002. Quality of dicom header information for image categorization. In: Int. Symposium on Medical Imaging. Vol. 4685. SPIE, pp. 280–287.

[34] Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference. pp. 147–151.

[35] Jager, F., Hornegger, J., 2009. Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. IEEE TMI 28 (1), 137–150.

[36] Jenkinson, M., Smith, S. M., 2001. A global optimisation method for robust affine registration of brain images. MIA 5 (2), 143–156.

[37] Jian, B., Vemuri, B., 2011. Robust point set registration using gaussian mixture models. IEEE TPAMI 33 (8).

[38] Johnson, C., Chen, M., Toledano, A., Heiken, J., Dachman, A., Kuo, M., Menias, C., Siewert, B., Cheema, J., Obregon, R., Fidler, J., Zimmerman, P., Horton, K., Coakley, K., Iyer, R., Hara, A., Halvorsen, R. J., Casola, G., Yee, J., Herman, B., Burgart, L., Limburg, P., 2008. Accuracy of CT colonography for detection of large adenomas and cancers. New England Journal of Medicine 359 (12), 1207–1217.

[39] Kadir, T., Brady, M., 2001. Saliency, scale and image description. IJCV 45 (2), 83–105.

[40] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., Parsey, R. V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. NeuroImage 46 (3), 786–802.

[41] Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J., 2010. elastix: a toolbox for intensity based medical image registration. IEEE TMI 29 (1), 196–205.

[42] Kloppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., Fox, N. C., Jack Jr, C. R., Ashburner, J., Frackowiak, R. S. J., 2008. Automatic classification of mr scans in alzheimer's disease. Brain 131, 681–689.

[43] Laptev, I., 2005. On space-time interest points. IJCV 64 (2/3), 107–123.

[44] Lehmanna, T. M., Gld, M. O., Deselaers, T., Keysers, D., Schubert, H., Spitzer, K., Ney, H., Wein, B., 2005. Automatic categorization of medical images for content-based retrieval and data mining. Computerized Medical Imaging and Graphics 29, 143–155.

[45] Leibe, B., Leonardis, A., Schiele, B., 2004. Combined object categorization and segmentation with an implicit shape model. In: ECCV Workshop on Statistical Learning in Computer Vision.

[46] Li, B., Christensen, G. E., Hoffman, E. A., McLennan, G., Reinhardt, J. M., 2003. Establishing a Normative Atlas of the Human Lung: Intersubject Warping and Registration of Volumetric CT Images. Academic Radiology 10 (3), 255–265.

[47] Lindeberg, T., 1998. Feature detection with automatic scale selection. IJCV 30 (2), 79–116.

[48] Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. IJCV 60 (2), 91–110.

[49] Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R., 2007. Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented and demented older adults. Journal of Cognitive Neuroscience 19, 1498–1507.

[50] Marr, D., Poggio, T., 1977. A theory of human stereo vision. In: MIT AI Memo. Vol. 451.

[51] Mikolajczyk, K., Schmid, C., 2004. Scale and affine invariant interest point detectors. IJCV 60 (1), 63–86.

[52] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L. V., 2005. A comparison of affine region detectors. IJCV 65, 43–72.

[53] Moravec, H. P., 1979. Visual mapping by a robot rover. In: Proc. of the 6th International Joint Conference on Artificial Intelligence. pp. 598–600.

[54] Murphy, K., van Ginneken, B., Reinhardt, J. M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G. E., Garcia, V., Vercauteren, T., Ayache, N., Commowick, O., Malandain, G., Glocker, B., Paragios, N., Navab, N., Gorbunova, V., Sporring, J., de Bruijne, M., Han, X., Heinrich, M. P., Schnabel, J. A., Jenkinson, M., Lorenz, C., Modat, M., McClelland, J. R., Ourselin, S., Muenzing, S. E. A., Viergever, M. A., de Nigris, D., Collins, D. L., Arbel, T., Peroni, M., Li, R., Sharp, G. C., Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Smeets, D., Loeckx, D., Song, G., Tustison, N., Avants, B., Gee, J. C., Staring, M., Klein, S., Stoel, B. C., Urschler, M., Werlberger, M., Vandemeulebroucke, J., Rit, S., Sarrut, D., Pluim, J. P. W., 2011. Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge. IEEE TMI 30 (11), 1901–1920.

[55] National Alliance for Medical Image Computing (NAMIC), 2011. NAMIC traumatic brain injury. www.na-mic.org/Wiki/index.php/DBP3:UCLA.

[56] Ni, D., Qu, Y., Yang, X., Chui, Y. P., Wong, T.-T., Ho, S., Heng, P. A., 2008. Volumetric ultrasound panorama based on 3D SIFT. In: MICCAI.

[57] Niemeijer, M., Garvin, M. K., Lee, Kyungmoo van Ginneken, B., Abramoff, M. D., Sonka, M., 2009. Registration of 3D spectral OCT volumes using 3D SIFT feature point matching. In: Medical Imaging 2009: Image Processing. Vol. 7259. SPIE.

[58] Nister, D. Stewenius, H., 2006. Scalable recognition with a vocabulary tree. In: CVPR. pp. 2161–2168.

[59] Nowinski, W., Qian, G., Bhanu, P. K., Hu, Q., Aziz, A., 2006. Fast talairach transformation for magnetic resonance neuroimages. J. Comp. Assist Tomogr. 30 (4), 629–641.

[60] Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C., 2011. Dramms: Deformable registration via attribute matching and mutual-saliency weighting. Medical Image Analysis 15 (4), 622–639.

[61] Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N., 2001. Reconstructing a 3d structure from serial histological sections. Image and Vision Computing 19 (1–2), 25–31.

[62] Pennec, X., Ayache, N., Thirion, J.-P., 2000. Landmark-Based Registration Using Features Identified Through Differential Geometry. Academic Press, Ch. 31, pp. 499–513.

[63] Periaswamy, S., Farid, H., 2006. Medical image registration with partial data. Medical Image Analysis 10, 452–464.

[64] Potesil, V., Kadir, T., Platsch, G., Brady, M., 2010. Improved anatomical landmark localization in medical images using dense matching of graphical models. In: BMVC.

[65] Rasmussen, C. E., 2001. The infinite gaussian mixture model. In: Neural Information Processing Systems. pp. 554–560.

[66] Rohr, K., 1997. On 3D differential operators for detecting point landmarks. Image and Vision Computing 15 (3), 219–233.

[67] Rohr, K., Stiehl, H. S., Sprengel, R., Buzug, T. M., Weese, J., Kuhn, M. H., 2001. Landmark-based elastic registration using approximating thin-plate splines. IEEE TMI 20 (6), 526–534.

[68] Romeny, B. t. H., 2003. Front-End Vision and Multi-Scale Image Analysis. Kluwer Academic Publisher.

[69] Scovanner, P., Ali, S., Shah, M., 2007. A 3-dimensional sift descriptor and its application to action recognition. In: International conference on Multimedia. pp. 357–360.

[70] Stewart, C. V., Tsai, C.-L., Roysam, B., 2003. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE TMI 22 (11), 1379–1394.

[71] Talairach, J., Tournoux, P., 1988. Co-planar Stereotactic Atlas of the Human Brain: 3-Dimensional Proportional System: an Approach to Cerebral Imaging. Georg Thieme Verlag, Stuttgart.

[72] Tao, Y., Peng, Z., Krishnan, A., Zhou, X. S., 2011. Robust learning-based parsing and annotation of medical radiographs. IEEE TMI 30 (2), 338–350.

[73] Toews, M., Arbel, T., 2007. A statistical parts-based appearance model of anatomical variability. IEEE TMI 26 (4), 497–508.

[74] Toews, M., Arbel, T., 2009. Detection, localization and sex classification of faces from arbitrary viewpoints and under occlusion. IEEE TPAMI 31 (9), 1567–1581.

[75] Toews, M., Wells III, W. M., 2009. Sift-rank: Ordinal descriptors for invariant feature correspondence. In: CVPR.

[76] Toews, M., Wells III, W. M., 2010. A mutual-information scale-space for image feature detection and feature-based classification of volumetric brain images. In: MMBIA.

[77] Toews, M., Wells III, W. M., Collins, D. L., Arbel, T., 2010. Feature-based morphometry: Discovering group-related anatomical patterns. NeuroImage 49 (3), 2318–2327.

[78] Tomasi, C., Shi, J., 1994. Good features to track. In: CVPR. pp. 593–600.

[79] Turk, M., Pentland, A. P., 1991. Eigenfaces for recognition. CogNeuro 3 (1), 71–96.

[80] Urschler, M., Zach, C., Ditt, H., Bischof, H., 2006. Automatic Point Landmark Matching for Regularizing Nonlinear Intensity Registration: Application to Thoracic CT Images. In: MICCAI. pp. 710–717.

[81] Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: CVPR. pp. 511–518.

[82] Wells III, W. M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of mutual information. MIA 1 (1), 35–52.

[83] Wells III, W. M. 1997. Statistical Approaches to Feature-Based Object Recognition. IJCV 21, 63–98.

[84] Wels, M., Zheng, Y., Carneiro, G., Huber, M., Hornegger, J., Comaniciu, D., 2009. Fast and Robust 3-D MRI Brain Structure Segmentation. In: MICCAI.

[85] Wu, G., Wang, Q., Jia, H., Shen, D., 2012. Feature-based groupwise registration by hierarchical anatomical correspondence detection. Human Brain Mapping 33 (2), 253–271.

[86] Yang, J., Williams, J. P., Sun, Y., Blum, R. S., Xu, C., 2011. A robust hybrid method for nonrigid image registration. Pattern Recognition 44, 764–776.

[87] Zhang, P., Cootes, T. F., 2012. Automatic construction of parts+geometry models for initializing groupwise registration. IEEE TMI 31 (2), 341–358.