A Statistical Parts-based Appearance Model of Inter-subject Variability

Matthew Toews[1], D. Louis Collins[2], and Tal Arbel[1]

[1] Centre for Intelligent Machines, McGill University, Montréal, Canada, (mtoews,arbel)@cim.mcgill.ca
[2] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montréal, Canada, louis@bic.mni.mcgill.ca

Abstract. In this article, we present a general statistical parts-based model for representing the appearance of an image set, applied to the problem of inter-subject MR brain image matching. In contrast with global image representations such as active appearance models, the parts-based model consists of a collection of localized image parts whose appearance, geometry and occurrence frequency are represented statistically. Given a typical set of brain MR images, the parts-based approach explicitly addresses the problem of missing features due to anatomical differences between subjects. Model learning automatically discovers parts that appear with statistical regularity in a large collection of subject images with little or no manual help, in the presence of significant inter-subject variability. The parts can be robustly represented by generic scale-invariant features, and therefore, the framework can be applied to a wide variety of image modalities. Experimentation shows that a parts-based model can be learned from a large set of MR brain images, and used to determine parts that are common within the group of subjects. The framework can potentially be used to study the commonality of anatomical parts, or to group subjects according to the parts they share. Images of new subject can be fit to the model in a robust manner. Preliminary results indicate that the model can be used to automatically identify useful features for inter-subject image registration despite large changes in appearance.

1 Introduction

In this article, we consider the task of inter-subject registration, i.e. determining correspondence between images of different subjects of a population. Inter-subject registration is a task of great importance to the medical imaging community, as it lies at the heart of understanding how individuals vary within a population, and poses a significant challenge due to problem of inter-subject variability. No two subjects are identical - structures may exhibit significant variation from one subject to the next or may simply not exist in all subjects.

Given the issue of inter-subject variability, it is reasonable to expect that registration attempting to determine a one-to-one mapping between different subjects may perform poorly in locations where such a mapping may not exist. Authors have attempted to identify points or regions where registration can be expected to perform well. Salient feature detection approaches are an option [8, 9], but are generally ineffective due to inter-subject variability - features extracted in one subject may not exist in the other. Similar features can be extracted over a set of aligned subjects [5], but again one cannot know if these features will be relevant to new images. We hypothesize that it is reasonable to expect that, in order to determine the image regions likely to register well, one must first see the images to be matched, in addition to being acquainted with the image domain.

In this paper, we present a new probabilistic approach to automatically identify image regions likely to result in meaningful inter-subject registration. Our approach is two-fold. First, off-line, we present a learning approach to building a parts-based appearance model of the image class in question (e.g. MR brain images). The model statistically quantifies the appearance, geometry and occurrence frequency of localized image parts which occur with statistical regularity over a set of training subjects. Second, we fit this model to new images to be registered, thereby identifying instances of learned model parts in the images. Our image regions arising from the same model parts in both images are good candidates for registration.

The main contribution of this framework is a new parts-based statistical model of appearance that is based on localized image regions, as opposed to the global modes of image variation popular in the medical imaging literature [3, 10]. The local nature of the method makes it possible to explicitly account for occlusion arising from inter-subject variation on a local scale, as model parts are not expected to (and typically do not) occur in all images. In addition, the model can be efficiently learned over a large set of training images with no manual interaction, and fit to new images, all in the presence of significant inter-subject variation. The second contribution is a method of using this model to automatically determine the regions most likely to lead to meaningful registration in new images.

Our model is based on generic scale-invariant image features, and is general enough to be applied to a wide variety of image modalities. We present experiments based on T1-weighted MRI brain images from the ICBM152 data set [2], which consists of 152 volumes of 88 male and 66 female normal subjects, aged 24.6 ± 4.8 years. Preliminary results indicate that the system is able to automatically locate parts and quantify their commonality in set of subject brains with large appearance variation.

The remainder of this paper is organized as follows: we describe the model in Section 2, model learning and fitting to new data in Section 3, experimentation involving inter-subject registration in section 4 and a discussion in Section 5.

2 A Statistical Parts-based Appearance Model

Identifying the regions most likely to match well between two different subjects requires first learning the image domain itself, in order to know what image structure occurs with regularity, and the range of geometrical and appearance variability one can expect. To this end, we present a statistical model based on localized image parts. The parts of our model are based on generic, automatically detected scale-invariant features - a model based on such features is attractive because it can be automatically applied in a wide variety of contexts, as opposed to approaches based on special-purpose detectors for specific image structures (i.e. a particular sulcus) or manual landmark selection, which is tedious for large data sets and prone to inter-rater variation, as a human must decide which landmarks are optimal, how many landmarks must be used, etc.

Scale-invariant features [8, 9] offer an improvement over simple features such as corners [6], as they localize salient image patterns in scale in addition to image translation. They can be extracted in a wide variety of image contexts, and are robust to reasonable variation in intensity, in addition to geometrical deformations such as translation, orientation, scale and affine transformations. Feature detection and correspondences based on scale-invariance features are both fast: features can be extracted efficiently from image pyramids and matched by normalizing feature image content with respect to feature geometry, removing the need to perform an explicit search over deformation parameters. Feature geometrical parameters such as location x, orientation θ and scale σ recovered in the extraction process can be used formulate multiple independent hypotheses as to the geometrical transform between images, leading to occlusion and noise-resistant correspondence. In addition, scale-invariant features can be extracted from a variety of different image properties such as blobs [8], edges [9], phase [1] and entropy [7], and used in conjunction with each other.

Despite their attractive properties, scale-invariant features are rarely used for inter-subject registration because of the fact that, in general, the same features cannot be extracted in images of different subjects due to inter-subject variability. As a result, the number of features located is typically insufficient for reliably determining correspondence between different subjects. Although this fact can be considered a shortcoming of automatic feature detection, in this paper, we argue that it merely reflects the difficulty of inter-subject registration and serves as an important indication of occlusion. Recently, statistical models of features learned from a set of images have emerged in order to address these difficulties [4, 12]. The parts-based model presented in this section is based on statistically quantifying the occurrence frequency, appearance and geometry of features over a large set of subjects, thereby learning a set of parts that can be reliably matched between images of different subjects. We present the components and the probabilistic formulation of the model, and describe how such a model can be learned from training data, and fit to a new image instances.

2.1 Model Components

Our parts-based model consists of a set of parts m_i with respect to a common reference frame o. Model parts are scale-invariant features denoted as $m_i : \{m_i^b, m_i^g, m_i^a\}$ representing the occurrence, geometry and appearance of a scale-invariant feature within an image. Feature occurrence m_i^b is a binary random variable representing the probability of feature presence (or absence) in an image. Feature geometry $m_i^g : \{x_i, \theta_i, \sigma_i\}$ is an oriented region in $\mathbb{R}^N$ image space, represented by N-parameter location x_i, an $N-1$ parameter orientation θ_i, and a scale σ_i. Feature appearance m_i^a describes the image content at region m_i^g, and can generally be parameterized in a number of ways, i.e principle components [13].

Reference frame $o : \{o^b, o^g\}$ represents the occurrence and geometry of a common reference frame relating parts. o^b is a binary random variable indicating the presence/abscene of the reference frame, who's significance will be made clear in the discussion of model learning and fitting. o^g is parameterized in the same manner as scale-invariant feature geometry m^g, and serves as a common geometrical frame within which different features m_i and m_j can be considered as statistically independent (i.e. knowing o^g, feature variation is localized to a neighborhood around an expected value of m_i). Within the context of MRI brain registration, a well-known definition of o^g is the midplane line defining the Talairach stereotaxic space [11], which passes from the superior aspect of the anterior commissure to the inferior aspect of the posterior commissure. Figure 1 illustrates the relationship between o and m_i.

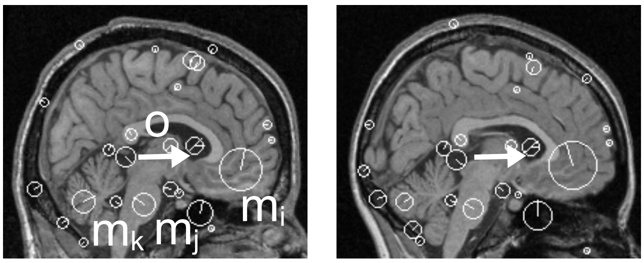

Fig. 1. Scale-invariant features and reference frames in sagittal slices of T1-weighted MR brain images of two subjects. Features m_i, illustrated as white circles inset by radial lines, are oriented regions consisting of a location x_i, orientation θ_i and scale σ_i. Reference frame o, illustrated as a white arrow, represents the projection of the Talairach AC-PC line onto the slice. Feature occurrence, appearance and geometry with respect to the reference frame can be quantified statistically via a parts-based model over a large set of subjects.

2.2 Probabilistic Model Formulation

Our model consists of a set of N model parts $\{m_i\}$, when observed in a new image can be used infer the reference frame o. Assuming that parts m_i are conditionally independent given o, the posterior probability of o given m_i can be expressed using Bayes rule as:

$$p(o|\{m_i\}) = \frac{p(o)p(\{m_i\}|o)}{p(\{m_i\})} = \frac{p(o)\prod_i^N p(m_i|o)}{p(\{m_i\})}, \tag{1}$$

where $p(o)$ is a prior over reference frame geometry and occurrence and $p(m_i|o)$ is the likelihood of feature m_i given o. Our model focuses principally on the likelihood term $p(m_i|o)$, which can be expressed as:

$$p(m_i|o) = p(m_i^a, m_i^b|o)p(m_i^g|o) = p(m_i^a|m_i^b)p(m_i^b|o^b)p(m_i^g|o^b, o^g), \quad (2)$$

under the assumptions that m^a and m^b are statistically independent of m^g given o, and that m^a and o are statistically independent given m^b.

Appearance likelihood $p(m_i^a|m_i^b)$ is represented as a multivariate Gaussian distribution in an appearance space and parameterized by mean and covariance μ_i^a, Σ_i^a. $p(m_i^b|o^b)$ is the probability of feature occurrence given reference frame occurrence, represented as a discrete multinomial distribution with event count parameters $\pi_i = \{\pi_i^1, \ldots, \pi_i^4\}$. Geometry likelihood $p(m_i^g|o^b, o^g)$ models the residual error of a linear transform from feature to reference frame geometry $m_i^g \rightarrow o^g$, and is represented as a Gaussian distribution with mean and covariance parameters μ_i^g, Σ_i^g. In order to characterize geometrical error in a scale-invariant manner, scale is transformed logarithmically, and translation is normalized by reference frame scale.

3 Model Learning and Fitting

Model learning involves estimating the model parameters of a set of N features $\{m_i\}$ introduced in the previous section from a set of training images. Learning proceeds based on a set of data vectors of the form $\{m_i^a, m_i^g, o^g\}$, where m_i^a and m_i^g are automatically extracted features and o^g is the labeled reference frame. Labeling o^g can be done by manually by defining a line segment corresponding to o^g in images, or in an approximate manner via linear registration of MR volumes into the same stereotactic space - we adopt the latter approach. Features are extracted and represented using the SIFT (scale-invariant feature transform) technique [8], based on an efficient implementation available online from the author, although a variety of other techniques could be used. Briefly, SIFT features are extracted as maxima/minima in a difference-of-Gaussian scale space pyramid, determining feature geometry m_i^g. The SIFT appearance representation m_i^a is a 128-value vector, corresponding to bins of a histogram of image first derivatives quantized into 8x4x4=128 bins over orientation and (x,y) position.

Prior to learning, data vectors are normalized spatially wrt the reference frame $\{m_i^a, m_i^g, o^g\} \rightarrow \{m_i^a, \bar{m}_i{}^g\}$. Learning begins by clustering data vectors according to normalized geometry, determining μ_i^g, Σ_i^g. Occurrence and appearance parameters π_i and μ_i^a, Σ_i^a are then estimated simultaneously such that the likelihood ratio $\frac{p(m_i^{b=1}|o^{b=1})}{p(m_i^{b=1}|o^{b=0})}$ is maximized. Note that this ratio represents a measure of the distinctiveness of a particular feature within the reference frame. As the number of clusters N is unknown *a priori*, clustering is achieved by growing variance estimates around individual vectors instead of algorithms such as K-means. After learning, features can be automatically ranked according to their likelihoods in sagittal slices of 102 brains from the ICBM152 data set [2], see Figure 2. Having quantified feature appearance, geometry and occurrence frequency,

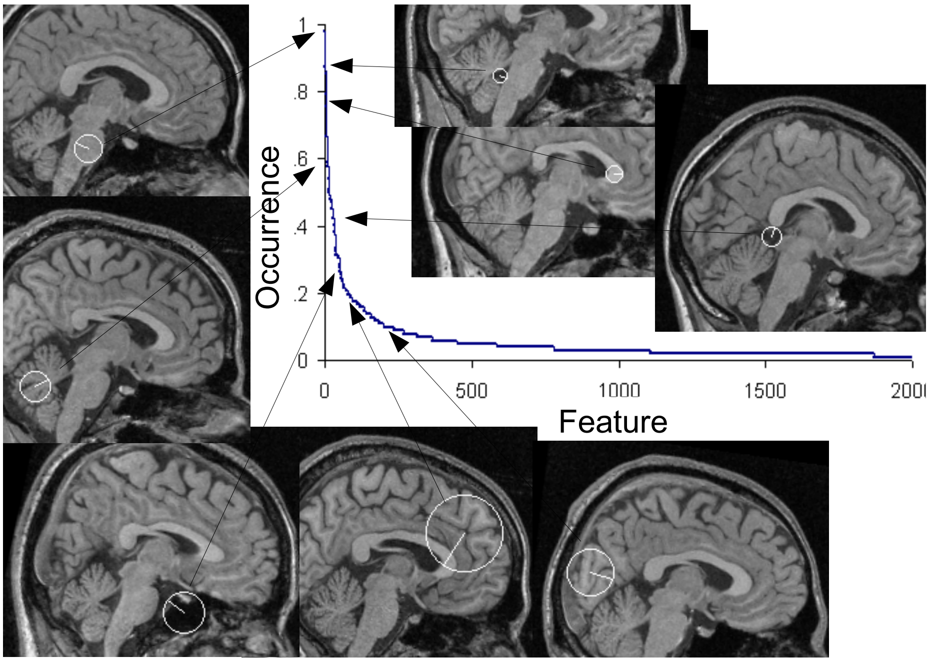

Fig. 2. A graph of features sorted by occurrence frequency, i.e. $p(m_i^{b=1}|o^{b=1})$. The images contain features extracted by the indicated frequency within a sample of 102 brains from the ICBM152 data set [2]. Note that feature occurrence drops off sharply, indicating that a relatively small number of features are common to all brains, whereas a large number of features are specific to a small number of brains.

features observed in new images can be fit to the model and assessed as to their usefulness for tasks such as inter-subject registration - highly distinctive features indicate good candidate regions for registration, whereas poorly distinctive features are typically related to ambiguous or subject-specific characteristics.

Once the model has been learned, it can be fit to a new image to localize the reference frame o. Unlike other registration/fitting techniques based on iterative algorithms which tend to go awry when started outside a 'capture radius' of the optimal solution, our model can be fit globally. In order to determine o in a new image, features extracted from a new image are matched to those in the model. A particular pairing of model features to image features results in a hypothesis as to o^g in the new image. We are interested in evaluating whether this hypothesis is the result of a true model instance or random noise, i.e. $o = \{o^g, o^{b=1}\}$ or $\overline{o} = \{o^g, o^{b=0}\}$. These two possibilities can be compared via a Bayes decision ratio:

$$\gamma(o) = \frac{p(o|\{m_i\})}{p(\overline{o}|\{m_i\})} = \frac{p(o)}{p(\overline{o})} \prod_{i=1}^{N} \frac{p(m_i|o)}{p(m_i|\overline{o})}, \tag{3}$$

where high $\gamma(o)$ indicates the presence of a model, and $\frac{p(o)}{p(\bar{o})}$ is a constant expected ratio of true to false model instance. Fitting is performed by identifying o^g maximizing $\gamma(o)$, via an efficient search process considering small subsets of model features $\{m_i\}^k \subseteq \{m_i\}$ that agree on discrete o^g.

4 Application to Inter-subject Registration

Given a new pair of subject brain images, we wish to explore how the features derived from the model can be used for inter-subject registration. Intuitively, fitting the model to a new image can be seen as describing the new image in terms of a unique combination of learned model parts. Images of different subjects can be compared in terms of the model features that they share, the key notion being that any given pair of subjects shares a unique set of image parts - this information can potentially be used in a number of ways - i.e. to drive inter-subject registration in regions where images are known to have statistically similar content, or to cluster subject images that share similar image content.

In order to explore the possibility applying the framework to inter-subject image registration, we examine a few cases where the feature parts automatically found through training are used to match one subject brain to another. Two trials are performed in order to compare 1) selecting likely model features of one subject similar to the strategy proposed by [5] with 2) selecting likely model features common to both subjects. Registration is performed by determining feature displacement from one image to the next, based on a sum-of-squared-differences similarity measure regularized by an elastic prior between features. On a data set of normal subjects, both feature selection techniques result in reasonable registration, as most subjects share similar features. In the presence of significant inter-subject variability however, selecting features common to both subjects allows registration to avoid regions in which a valid solution may not exist, as illustrated in Figure 3.

5 Discussion

In this paper, we presented a parts-based approach to statistical appearance modeling, where a set of images is represented by a collection of automatically-extracted generic parts. The parts-based approach is unique in that it models the variation of local structures, and as such is able to explicitly model the notion of occlusion. Preliminary experimentation shows that model parts can be identified and quantified probabilistically, and robustly fit to new image, all in the presence of significant inter-subject variability. In addition, the model can be used to identify statistically significant parts common to images of different subjects. Such parts indicate image structure shared common to both subjects, and can be matched with higher reliability from one subject to another. These encouraging results should naturally lead to the development of robust scale-invariant feature detectors in 3-D and 4-D, applicable in a wide variety of contexts including (1) intra-subject image registration and (2) the study of anatomical structures common to medical images of many subjects, an important area of computational anatomy.

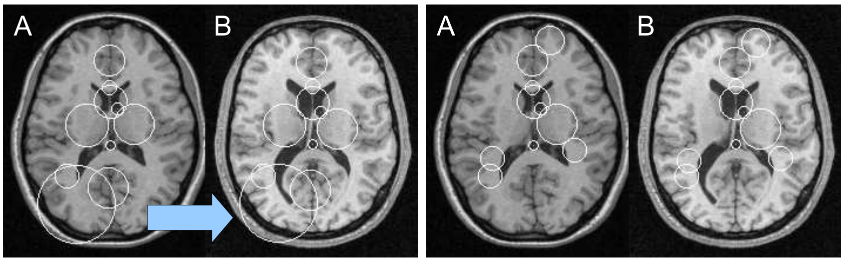

Fig. 3. Illustrating model-based feature selection for registration. In each image pair, 10 features are selected and registered between subjects A and B. On the left, the most likely model features of subject A are used, on the right the most likely features common to both subjects A and B are used. Notice that a valid registration solution may not exist in the lower left region, due to an enlarged ventricle of subject B. Basing registration on features common to both subjects, this ambiguous region can be avoided, resulting in more meaningful registration in the presence of inter-subject variability.

References

1. G. Carneiro and A.D. Jepson, *Multi-scale phase-based local features*, CVPR03, vol. 1, 2003, pp. 736–743.
2. D. L. Collins, N.J Kabani, and A.C. Evans, *Automatic volume estimation of gross cerebral structures*, 4th International Conference of Functional Mapping of the Human Brain (A. Evans, ed.), 1998.
3. T.F. Cootes, G.J. Edwards, and C.J. Taylor, *Active appearance models*, Pattern Analysis and Machine Intelligence **23** (2001), no. 6, 681–684.
4. R. Fergus, P. Perona, and A. Zisserman, *Object class recognition by unsupervised scale-invariant learning*, CVPR, 2003, pp. 264–271.
5. Wu Guorong, Feihu Qi, and Dinggang Shen, *Learning best features for deformable registration of mr brains*, MICCAI05, 2005.
6. C. Harris and M. Stephens, *A combined corner and edge detector*, Proceedings of the 4th Alvey Vision Conference, 1988, pp. 147–151.
7. T. Kadir and M. Brady, *Saliency, scale and image description*, IJCV **45** (2001), no. 2, 83–105.
8. David G. Lowe, *Distinctive image features from scale-invariant keypoints*, IJCV **60** (2004), no. 2, 91–110.
9. K. Mikolajczyk and C. Schmid, *Scale and affine invariant interest point detectors*, IJCV **60** (2004), no. 1, 63–86.
10. Daniel Rueckert, Alejandro F. Frangi, and Julia A. Schnabel, *Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration*, TMI **22** (2003), no. 8, 1014–1025.
11. J. Talairach and P. Tournoux, *Co-planar stereotactic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging*, Georg Thieme Verlag, Stuttgart, 1988.
12. M. Toews and T. Arbel, *Detection over viewpoint via the object class invariant*, ICPR, 2006.
13. M. Turk and A. P. Pentland, *Eigenfaces for recognition*, CogNeuro **3** (1991), no. 1, 71–96.