Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2013 December 31; 2013: 744–747. doi:10.1109/ISBI.2013.6556582.

CLASSIFYING MEDICAL IMAGES USING MORPHOLOGICAL APPEARANCE MANIFOLDS

Erdem Varol, Bilwaj Gaonkar, and Christos Davatzikos

University of Pennsylvania, Section of Biomedical Image Analysis, Department of Radiology, 3600 Market Street, Philadelphia, PA, 19104, USA

Abstract

Input features for medical image classification algorithms are extracted from raw images using a series of pre processing steps. One common preprocessing step in computational neuroanatomy and functional brain mapping is the nonlinear registration of raw images to a common template space. Typically, the registration methods used are parametric and their output varies greatly with changes in parameters. Most results reported previously perform registration using a fixed parameter setting and use the results as input to the subsequent classification step. The variation in registration results due to choice of parameters thus translates to variation of performance of the classifiers that depend on the registration step for input. Analogous issues have been investigated in the computer vision literature, where image appearance varies with pose and illumination, thereby making classification vulnerable to these confounding parameters. The proposed methodology addresses this issue by sampling image appearances as registration parameters vary, and shows that better classification accuracies can be obtained this way, compared to the conventional approach.

1. INTRODUCTION

The problem of classifying an individual into one of two or more classes (e.g. normal and pathologic) using a medical scan is important for diagnosis and prognosis. Over the last ten years the neuroimaging community, amongst others, has attempted to address this problem using various methods. An underlying theme in most of these methods is to take a two step approach where: the first step involves the use of existing medical image analysis technology such as segmentation and registration to extract features from raw images, and, the second step involves using these features as an input to a classification algorithm typically borrowed from machine learning literature to do the actual disease detection. Registration algorithms used in the first step transform the input samples into a common template space, thereby providing a standardized coordinate system to be used for building classifiers. However, most automated registration methods are parameter and template dependent. One of the most important parameters is the regularization parameter(s) that influence the smoothness of deformable registration algorithms, which is the main focus of our approach herein.

In the two step classification paradigm described above, parameter and template selection for the registration methods is agnostic to the goal of classification. Parameters that maximize commonly used registration accuracy criteria (such as Dice coefficients and the like) are by no means optimal for classification or even for group comparisons. Attempting to find parameters to maximize classifier performance is an extremely difficult and computationally intensive task. This task is further complicated by the fact that a different parameter setting might be optimal for every individual subject [1, 2, 3]. The proposed methodology is motivated by work in both the computer vision [4, 5, 6] and medical imaging literatures [7] that has shown that understanding the image appearance manifold

formed by sampling the parameter space (in our case, the parameters affecting registration accuracy) improves the richness of the representation of each object, and can therefore increase our ability to recognize an individual in a way that is robust relative to these parameters. The remaining manuscript is divided into three sections. Section 2 describes the methodology, Section 3 presents the results and Section 4 concludes the manuscript.

2. METHOD

Motivation

Registration aims to help characterize anatomical differences between a subject S_i and a template T by mapping the template space Ω_T to the subject space Ω_{S_i} through a diffeomorphism, $h_\lambda \in \mathcal{H}_{S_i} : \Omega_T \rightarrow \Omega_{S_i}, \mathbf{x} \mapsto h_\lambda(\mathbf{x})$. $\mathcal{H}_{S_i}$ is the set of all diffeomorphic transformations, h_λ , between these two spaces that maximize a similarity criterion between T and registered subject $h_\lambda(S_i)$ with λ as a parameter that controls regularization. While the mapping h_λ aims to transfer subject information to template space, the constraints set by the regularization parameter λ coupled with the requirement that the mapping is diffeomorphic results in imperfect registration and consequently a residual:

$$R_{h_\lambda}(\mathbf{x}) \triangleq T(\mathbf{x}) - S(h_\lambda(\mathbf{x})), \mathbf{x} \in \Omega_T \quad (1)$$

Although it is possible to achieve a mapping that results in extremely small residuals, such a transformation might not necessarily be the best way of representing brain morphology, from the perspective of understanding group differences and classifying individuals. Registering images to a template only aids in classification if the subtle features that bring forth maximum discrimination between classes are captured fully. Because we do not know a priori which features are the most discriminative, we cannot pick the best registration parameter that will capture these features.

To circumvent the problem of finding the optimal registration, we build upon the *Anatomical Equivalence Class* (AEC) concept introduced by Makrogiannis et. al. [1]. The *Complete Morphological Descriptor* (CMD) defined by,

$$\mathcal{M}_{h_\lambda} \triangleq (h_\lambda(\cdot), R_{h_\lambda}(\cdot)) \quad (2)$$

allows us to losslessly capture the full set of anatomical information supplied by the sample, regardless of the registration accuracy: different transformations are used in conjunction with different residuals, each time exactly reconstructing the original image, hence lossless. The AEC is then defined as the set of all possible CMD's of the sample by varying the confounding parameter, λ , and the resulting mapping h_λ : (Figure 1)

$$\mathcal{A}(S_i) \triangleq \{\mathcal{M}_{h_\lambda}(S_i), \forall \lambda \in \mathbb{R}^+, h_\lambda \in \mathcal{H}_{S_i}\} \quad (3)$$

Since an AEC potentially consists of infinitely many unique appearance models of a subject's anatomy, we propose to sample multiple morphological appearances (CMD's) of subjects from AEC's as a means of enriching training sets for medical image classification.

Densely Sampling AEC's To Generate Training Sets

Usually, support vector machine based medical image classification methods use a single mapping of subjects' anatomy, $(h_{\lambda^*}(S); R_{h_{\lambda^*}}(S))$ as input samples. Features are then

extracted from these inputs via a function ϕ and then the optimal separating hyperplane is determined by solving the SVM optimization problem:

$$\arg \min_{w \in H, b \in \mathbb{R}} \frac{1}{\ell} \sum_{i=1}^{\ell} L_h(y_i, w^T \phi(\mathcal{M}_{h_{\lambda^*}}(S_i)) + b) + \mu \|w\|^2 \quad (4)$$

Here, H is a reproducing kernel Hilbert space (RKHS) of a kernel $k : \Phi \times \Phi \rightarrow \mathbb{R}$, μ is free regularization parameter, and $L_h : \mathbb{R} \rightarrow [0; \infty)$ is the so called hinge loss function of the SVM. Such an approach finds an optimal hyperplane in a localized region on each subject's AEC (Figure 2).

In our framework, we extend the domain of the hyperplane to be the entire AEC's of the subjects. In this sense, the optimization problem becomes:

$$\arg \min_{w \in H, b \in \mathbb{R}} \frac{1}{\ell m} \sum_{i=1}^{\ell} \sum_{j=1}^m L_h(y_i, w^T \phi(\mathcal{M}_{h_{\lambda_j}}(S_i)) + b) + \mu \|w\|^2 \quad (5)$$

Here, m is the number of points that are sampled from each AEC. With a criteria that ensures that samplings are unique and distinct, it is possible to better approximate the structure of the AEC's as m increases. Such criteria would be:

$$p \neq q \iff \mathcal{M}_{h_{\lambda_p}}(S_i) \neq \mathcal{M}_{h_{\lambda_q}}(S_i) \forall p: \|\mathcal{M}_{h_{\lambda_p}}(S_i) - \mathcal{M}_{h_{\lambda_{p+1}}}(S_i)\| > \varepsilon_m \quad (6)$$

We expect the hyperplane found in Eq. (5) to generalize better than a hyperplane found using Eq. (4) because a classifier determined via Eq. (5) learns using information captured from S_i at m different values of λ . Our experiments confirmed this. Further it is also prudent to expect that the classifier performance would improve as m increases. This also happened to be true for our experiments.

3. EXPERIMENTS AND RESULTS

Data

We validated our framework using T1-weighted MR imaging data obtained for the study of two distinct diseases: Alzheimer's disease (AD) and Autism Spectrum Disorder (ASD). The AD dataset (139 patients / 178 controls) contained T1-weighted MR scans from 1.5T scanners acquired sagittally using volumetric 3D MPRAGE with 1.25×1.25 mm in plane spatial resolution and 1.2 mm thick sagittal slices. The ASD dataset (105 patients / 101 controls) were T1-weighted MR scans from a 3T scanner, acquired sagittally using volumetric 3D MPRAGE with $0.8 \text{ mm} \times 0.8 \text{ mm}$ in plane resolution and 0.9 mm thick sagittal slices.

Preprocessing and Construction of AEC's

Image preprocessing involved bias correction, skull removal and segmentation into grey matter/white matter/ cerebrospinal fluid. The AEC for a specific subject was generated by registering it to a template using several different levels of smoothness (where smoothness is controlled by a parameter we denote as λ). Registration was performed using the diffeomorphic demons algorithm [8]. The Jacobian determinant of the deformation field and the residual were computed and concatenated together in a vector which constitutes a CMD: $(J_{h_\lambda}, R_{h_\lambda})$. The collection of vectors obtained by generating CMDs using various λ 's

constitute an approximate AEC. Thus, an AEC might contain only one CMD or might contain several CMD's. The number of CMD's that an AEC descriptor contains depends on the number of unique values of λ at which registration between a given subject and the template has been performed. We refer to this number as the 'sampling rate' of the AEC. For instance if registrations were performed using $\lambda = 0$, $\lambda = 50$ and $\lambda = 100$ then the AEC for that subject would be a collection of 3 distinct vectors and have a 'sampling rate' of 3. Note that in all our experiments we use AEC's with the same sampling rate for both training and testing.

Classification methodology

Training: CMD's that constitute the AEC's of all the training subjects were used with the corresponding subject label to train a linear SVM. The training process which involves solving Eq. (5) gives us an SVM model which we denote by w^*, b^* . **Testing:** We use our learnt SVM model to obtain a decision score ($w^{*T} x_{test, CMD} + b^*$) for every CMD corresponding to a given test subject. To generate a single decision score for the test subject we simply add the decision scores of all the CMD's corresponding to it.

3.1. Results

We perform leave one out cross validation experiments using both the AD and the ASD datasets and by using AEC sampling rates of 1,3,5 and 9. The results are summarized in Table 1. The column denoted by 'AEC' records the average leave one out AUC obtained using the method described in the manuscript. The AUC's of classifiers trained only using CMD's that corresponded to a single λ can be found in the columns denoted by $\lambda = value$. The table shows that the SVM trained using the entire AEC outperforms the SVM trained using only a single CMD per subject. Furthermore, increased AEC sampling rate consistently improves classification performance(Figure 3,Table 1).

4. CONCLUSION AND FUTURE DIRECTIONS

Here, we demonstrate that incorporating information from a larger number of uniquely parameterized registrations, leads to a better classifier performance in terms of AUC. We have demonstrated the concept empirically by varying deformation field regularization in the demons algorithm. However, it is plausible that the concept is generic and is applicable to any registration method and any parameter that is an input to such a method. Perhaps, this concept can be extended beyond registration and applied to other 'parameter dependent' image pre processing machinery such as bias correction or skull removal or segmentation. It's possible that different combinations of parameter settings in preprocessing capture different aspects of a subject's brain morphology. Future work in this direction is necessary to ascertain these possibilities.

Finally, we note that an important extension of our approach would be to include registrations obtained via different atlases/templates. The dependence on a single template, arbitrary or group-average, introduces potentially harmful biases emanating from the relative position of an individual from whatever template is used or estimated. Extending the AEC's of individuals to include the utilization of a representative set of templates largely eliminates this bias and can be readily handled in our framework.

In conclusion, we have proposed a new paradigm that can be exploited to improve medical imaging based pattern classification.

REFERENCES

1. Makrogiannis S, et al. Anatomical equivalence class: A morphological analysis framework using a lossless shape descriptor. *Medical Imaging, IEEE Transactions on*. 2007; vol. 26(no. 4)
2. Baloch S, et al. An anatomical equivalence class based joint transformation-residual descriptor for morphological analysis. *Information Processing in Medical Imaging*, vol. 4584 of *Lecture Notes in Computer Science*. 2007
3. Baloch S, et al. Morphological appearance manifolds in computational anatomy: Groupwise registration and morphological analysis. *NeuroImage*. 2009; vol. 45 Supplement 1(no. 1):S73–S85. [PubMed: 19061962]
4. Wakin, MB. Ph.D. thesis. Rice University; 2006. *The Geometry of Low-dimensional Signal Models*.
5. Vasilescu MAO, et al. Multilinear subspace analysis of image ensembles. *Computer Vision and Pattern Recognition*, 2003. *Proceedings. 2003 IEEE Computer Society Conference on*. 2003
6. Wright J, et al. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*. 2010; vol. 98
7. Trounev A, et al. Metamorphoses through lie group action. *Foundations of Computational Mathematics*. 2005
8. Vercauteren T, et al. Diffeomorphic demons: Efficient nonparametric image registration. *NeuroImage*. 2009; vol. 45

Fig. 1.

Top: The least smooth registration with lower residual Bottom: The smoothest registration with large residual

Morphological Appearance Manifolds

 λ_0 Hyperplane

Manifold Hyperplane

Fig. 2.

(a) 3-dimensional embedding of CMD's at various λ 's are represented by points
 (b) Classification hyperplane found using CMDs from a single λ (c) Classification hyperplane found using proposed method

Fig. 3. ROC curves for AD (upper) and ASD (lower) classification. Blue indicate the ROC curves for classifiers trained with one CMD per individual for varying λ . Red indicate the classifier performance when the training and testing sets are extended to include all CMD's per individual, and testing classification aggregated through unweighted voting

Table 1

Blue values highlight the performance of the proposed method. Red values indicate the best possible performance using a single λ for preprocessing. Note that AUC of classification by increased AEC sampling rates (SR). The test data are entire AEC's, and decisions on CMD's are combined through voting. The columns with λ indicate the classifier performance after training with the corresponding CMD's only. Empty boxes indicate the CMD's not included in the level of AEC sampling.

		Area under ROC curve (AUC)									
Data	SR	AEC	$\lambda = 0$	$\lambda = 12.5$	$\lambda = 25$	$\lambda = 37.5$	$\lambda = 50$	$\lambda = 67.5$	$\lambda = 75$	$\lambda = 87.5$	$\lambda = 100$
AD	1	0.7497	0.7497	×	×	×	×	×	×	×	×
	3	0.8281	0.7497	×	×	×	0.7925	×	×	×	0.8043
	5	0.8499	0.7497	×	0.7691	×	0.7925	×	0.8028	×	0.8043
	9	0.8577	0.7497	0.7552	0.7691	0.7810	0.7925	0.7989	0.8028	0.8022	0.8043
ASD	1	0.648	0.648	×	×	×	×	×	×	×	×
	3	0.7487	0.648	×	×	×	0.6071	×	×	×	0.6195
	5	0.7491	0.648	×	0.6256	×	0.6071	×	0.6131	×	0.6195
	9	0.7569	0.648	0.6479	0.6256	0.6154	0.6071	0.6086	0.6131	0.6115	0.6195