DIAGNOSTIC FEATURES OF ALZHEIMER'S DISEASE EXTRACTED FROM FDG PET IMAGES

A Sayeed[1], M Petrou[1] and R Maguire[2]
[1] School of Electronics, Computing and Mathematics, University of Surrey, Guildford, GU2 7XH,
[2] GNIP Project, University Hospital Groningen, NL

Abstract— **FDG-PET images of patients suffering from Alzheimers disease (AD) were obtained from Paul Scherre Institute, Villingen, Switzerland. The data were from a CTI/Siemens ECAT 933/04-16 scanner, comprising of 7 image slices 128×128 pixels. The study included 48 Clinically diagnosed AD patients and 73 normal controls. Using an invariant feature extraction method features were extracted. The features are invariant to translation and rotation of object(s) within the image. The patients are separated into two groups one for training (24 AD and 37 normal controls) and one cross validation testing (24 AD and 36 normal controls). Discriminant function analysis yielded a classification accuracy of 88% sensitivity and 86% specificity, when these features were used.**

Keywords— **Alzheimers, FDG PET, Invariant features**

I. Introduction

In this paper we utilise the method for constructing invariant features for gray scale images proposed by Schulz-Mirbach [4]. The method produces features invariant to rotation and translation, but not invariant to scaling, thus its use is limited. Essentials concerning the action of transformation groups on gray scale images and the basic concepts for calculating invariant gray scale features by evaluating necessary integrals over the image will be introduced.

II. Theory of invariant gray scale features

Let $\mathbf{M}$ be a gray scale image, where $\mathbf{M}[x, y]$ is the gray value at pixel coordinates (x, y). In order to formulate the theory both continuous and discrete cases are considered. In the discrete case the pixel coordinates (x, y) are integers in the range $0 \leq x < N_x$, $0 \leq y < N_y$ where N_x and N_y are the dimensions of the image. In the continuous case the pixel coordinates can be real numbers.

Rotation and translation will be described by the action of the transformation group G with elements $g \in G$ on the images. So for an image $\mathbf{M}$ and a group element $g \in G$ the transformed image is denoted by $g\mathbf{M}$ [7]. So for an image translated by $\mathbf{t} = (t_x, t_y)^T \in \Re$ and rotated by angle $\phi \in [0, 2\pi]$ there exists

$$(g\mathbf{M})[x, y] \equiv \hat{\mathbf{M}}[x, y] = \mathbf{M}[k, l] \qquad \textit{with}$$
$$\begin{pmatrix} k \\ l \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{1}$$

All indices are modulo N. Due to the periodic boundary conditions the range of the components of the translation vector $\mathbf{t}$ is restricted to $0 \leq t_x \leq N_x$, $0 \leq t_y \leq N_y$ which is the size of the image. In the discrete formulation pixel coordinates are restricted to integers. Since vector $(k, l)^T$ in equation 1 is likely to have non integer values, appropriate rounding or interpolation is necessary.

An invariant image feature is a function $F(\mathbf{M})$ which is invariant to the action of the transformation group on the images i.e

$$F(g\mathbf{M}) = F(\mathbf{M}) \qquad \forall g \in G. \tag{2}$$

So feature F will remain constant even if image $\mathbf{M}$ is transformed by g.

The transformation law (1) states that "an image transformation consists of a rotation around the rotation centre followed by translation. This rotation centre is not known a priori and it does not necessarily fall inside the image. However, by applying an appropriate translation it is possible to bring the coordinate origin to the rotation centre. Since we are seeking features which are invariant both to rotation and translation the position of the rotation centre does not matter." [6]

A. Constructing invariant features

According to Schulz-Mirbach [4], [7] it is possible to construct an invariant feature $F(\mathbf{M})$ by integrating $f(g\mathbf{M})$ over the transformation group G:

$$F(\mathbf{M}) = A[f](\mathbf{M}) = \int_G f(g\mathbf{M}) dg \tag{3}$$

where $A[f]$ is called the average of f. This averaging technique is described in greater detail in [5]. Since we are considering the group of image rotations and translations with cyclic boundary conditions, the integration over the transformation group can be written as

$$A[f](\mathbf{M}) = \frac{1}{2\pi N_x N_y} \int_{t_y=0}^{N_y} \int_{t_x=0}^{N_x} \int_{\phi=0}^{2\pi} f(g\mathbf{M}) d\phi dt_x dt_y \tag{4}$$

Therefore if the function $f(\mathbf{M})$ is already invariant i.e. $f(\mathbf{M}) = f(g\mathbf{M})$ it remains unaltered by the group averaging. So $A[f](\mathbf{M}) = f(\mathbf{M})$.

Equation 4 can be implemented by a two step strategy where in the first step f is calculated for each pixel, and in the second step the integral of all these results is computed. In the discrete domain it is simply the sum of all the results obtained by evaluating f. Figure 1 describes this process schematically.

If we consider the monomial example $f(\mathbf{M}) = \hat{\mathbf{M}}[0, 0]$, we can deduce $\hat{\mathbf{M}}[0, 0] = \mathbf{M}[t_x, t_y]$ from equation 1. Then the group average, which is the feature, is given by

$$A[f](\mathbf{M}) = \frac{1}{N^2} \int_{t_y=0}^{N_y} \int_{t_x=0}^{N_x} \mathbf{M}[t_x, t_y] dt_x dt_y \tag{5}$$

In this case the result is simply the average gray value of the image.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Diagnostic Features of Alzheimer's Disease Extracted From FDG Pet Images	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) School of Electronics Computing and Mathematics University of Surrey Guildford, GU2 7XH	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

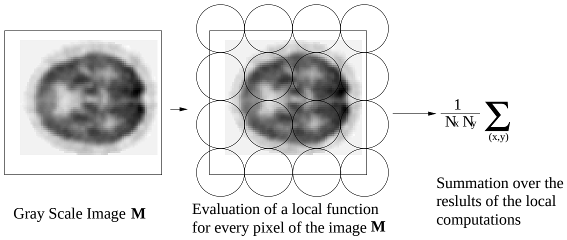

Fig. 1. Calculating invariant features

If we consider the monomial example $f(\mathbf{M}) = \hat{\mathbf{M}}[0,0]\hat{\mathbf{M}}[5,0]$, again we can deduce $\hat{\mathbf{M}}[0,0]\hat{\mathbf{M}}[5,0] = \mathbf{M}[t_x, t_y]\mathbf{M}[5\cos(\phi) + t_x, 5\sin(\phi) + t_y]$ from equation 1. Thus the feature is given by

$$A[f](\mathbf{M}) = \frac{1}{2\pi N_x N_y} \int_{t_y=0}^{N_y} \int_{t_x=0}^{N_x} \int_{\phi=0}^{2\pi} \mathbf{M}[t_x, t_y] \\ \mathbf{M}[5\cos(\phi) + t_x, 5\sin(\phi) + t_y] d\phi dt_x dt_y \tag{6}$$

This equation can be described by the two step strategy where the local function is

$$\int_{\phi=0}^{2\pi} \mathbf{M}[t_x, t_y]\mathbf{M}[5\cos(\phi) + t_x, 5\sin(\phi) + t_y] d\phi dt_x dt_y \tag{7}$$

Here the kernel operates at a neighbourhood of radius 5 pixels. We then sum all the local computations. This process if explained in more detail in section III-B.

B. Monomial properties

The method described not only allows invariant features which are invariant with respect to global image transformation, where a single angle and translation vector describe the transformation of the image, but also several local transformations. This is viable if there is only moderate overlap between the local transformation regions, i.e. as long as the object separation in the scene is greater than the kernel size of the monomials, [6].

This property is especially beneficial when we are considering images of the brain. Certain regions of the brain are affected with the onset of AD, however the position of these regions varies slightly between patients. Typically one would register the images to determine the position of these regions for analysis. By utilising the position invariance of this method one does not necessarily need to register the images. The effect of each of these regions (objects) will impact on the calculated feature irrespective of their position within the brain.

Furthermore, Schulz-Mirbach *et al* [6] have shown that provided objects do not overlap, the invariant features are approximately additive. This means that if we obtain features for two given objects and then in a scene both are present, the feature value here will be approximately the sum of the independent feature values.

The effect of AD on each region of the brain varies. For a given patient not all regions are affected. The additive nature of the features will enable the features to have the cumulative effects of all the regions that exhibit the effects of AD.

III. Experimentation

A. Data acquisition

FDG-PET images of patients suffering from AD and normal controls were obtained from the Paul Scherrer Institute (PSI), Villingen, Switzerland. The data were acquired with a CTI Siemens ECAT 933/04-16 scanner, over a period of 2 years. The scanning protocol remained the same during the duration of this study to eliminate the chance of any systematic errors being introduced. The data supplied had been reconstructed using filtered back projection and consisted of the first 16 frames of a dynamic scan, comprising of 7 image slices taken axially. The slice separation was approximately $8mm$ resulting in a total field of view of $56mm$ and each slice was 128×128 pixels.

The data comprised 73 normal controls and 48 patients clinically diagnosed with AD using the criteria of the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorder Association (NINCDS- ADRDA) [1]. For this study the dynamic nature of the data was not necessary thus the 16 frames for each subject were summed to enforce the signal.

The data was separated into two groups one for training purposes and one to be used for testing. For the training phase there were 37 normal controls and 24 AD and for the testing phase there were 36 normal controls and 24 AD.

The background of the image influences the local calculations of f. It is claimed that if the background is homogeneous then the impact on the calculation is insignificant [6]. In our image the background does not appear to be homogeneous, therefore it becomes necessary to extract the object from the background. Thus the images were manually segmented from the noisy background. Once this was done the invariant features were calculated.

B. Monomials

The following monomials, which were found to perform well by Schael *et al* [3] were used for this study. Although Schael *et al* used these for an industrial inspection task to recognise defects in textures, they seem to perform well in the task at hand compared to other arbitrarily chosen monomials.

$$f_1(\mathbf{M}) = \hat{\mathbf{M}}[0,0]\hat{\mathbf{M}}[5,0]$$
$$f_2(\mathbf{M}) = \hat{\mathbf{M}}[3,0]\hat{\mathbf{M}}[0,8]$$
$$f_3(\mathbf{M}) = \hat{\mathbf{M}}[0,0]\hat{\mathbf{M}}[5,0]\hat{\mathbf{M}}[0,10]$$

The monomials can be thought of as producing a local window inside which all calculations are performed. If we consider the f_1 monomial, this will produce a window of radius 5 pixels (see figure 2). In the window the first monomial will select the central pixel, this will be multiplied by a pixel at distance 5 pixels determined by the second monomial. This is repeated for angles $0 - 2\pi$ and the results summed. This is essentially the local part which will be repeated for all pixels to produce the feature.

In practice, to compute f_1 we consider a pixel and first multiply it with all pixels at distance 5 from it. Then we find the average value of these pairwise multiplications. This is equivalent to integrating over all rotation angles. This way we have a rotation invariant number assigned to each pixel. Finally we average all these numbers to find a single translation invariant number that characterises the whole region.

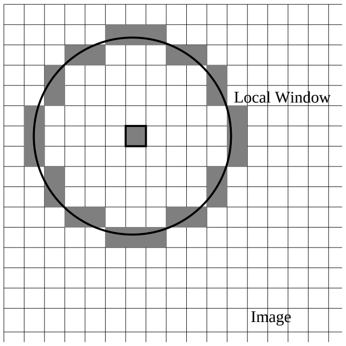

Fig. 2. Monomial f_1- Action of a local kernel

In order to compute f_2 we consider a pixel, at distance 3 pixels from the centre of the window. This is multiplied with pixels a distance 8 pixels from the centre of the window. The two pixels are 90 degrees to each other with respect to the centre of the window. The average of this is calculated for all pairwise multiplications to produce the rotation invariant number (see figure 3). Then the average of all these local windows is taken to produce a translation invariant number, which characterises the whole region.

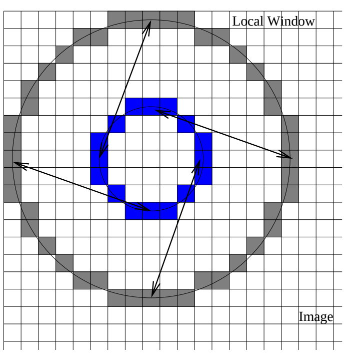

Fig. 3. Monomial f_2- Action of a local kernel; arrows indicate pixels that are multiplied pairwise

The computation of f_3 requires a much larger local window, namely one of 10 pixels. Here as in feature f_1 the central pixel in the local window is multiplied by a pixel at distance 5, which in turn is multiplied with a pixels a distance 10 away. The pixel form the vertices of a right angle triangle (see figure 4). The average of all these multiplications for all angles produces the rotation invariant number. This is done for each pixel and averaged to produce the translation invariant number, which characterises the whole region.

Applying these 3 monomials resulted in 21 (7 slices $\times$ 3) features in total per patient. These invariant features were used to perform discriminant function analysis on the training data and produce discriminant functions using the package Statistica [8]. Individual features were tested as well as combinations of many features to gain good

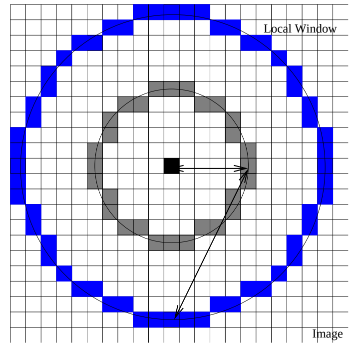

Fig. 4. Monomials f_3- Action of a local kernel, arrows indicate pixels that are multiplied together

classification accuracy. In order to test the accuracy the discriminant function obtained from this training phase was used to classify the test data.

IV. Results and Discussion

Using all 21 features for classification yielded a classification accuracy of 97% during the training phase. The classification accuracy reduced to 80% on the test data, (see table I).

TABLE I

Classification accuracy using all 21 invariant features

a. Training

True	Predicted AD	Predicted Normal	correct
AD	24	0	100%
Normal	2	35	95%
		Total	97%

b. Testing

True	Predicted AD	Predicted Normal	correct
AD	17	7	71%
Normal	4	33	89%
		Total	80%

Although the training classification accuracy was extremely high, the resultant testing accuracy was not as good. This can be explained by the fact that using so many features resulted in over training thus the discriminant function essentially described the training data too well. However, the testing data being slightly different were not classified so well.

The best 3 and the best 7 features were also used to classify the cases, (see tables II and III). The best features were selected by Statistica using a measure of F to enter. Here, the program selects for inclusion in the feature set, the feature that makes the most significant additional contribution to the discrimination between groups; that is, the program chooses the variable with the largest F value. The F value for a variable indicates its statistical significance in the discrimination between groups [8].

What we observe here is that reducing the number of features used in the classification reduces the classification

TABLE II

CLASSIFICATION ACCURACY USING THE BEST 7 INVARIANT MONOMIAL FEATURES, $(f_1(1), f_3(1), f_1(7), f_2(1), f_2(6), f_1(3), f_1(4))$ (THE NUMBER IN BRACKETS CORRESPONDS TO IMAGE SLICE NUMBER)

a. Training

True	Predicted AD	Predicted Normal	correct
AD	23	1	96%
Normal	3	34	92%
		Total	94%

b. Testing

True	Predicted AD	Predicted Normal	correct
AD	20	4	83%
Normal	7	30	81%
		Total	82%

TABLE III

CLASSIFICATION ACCURACY USING THE BEST 3 INVARIANT MONOMIAL FEATURES, $(f_1(1), f_3(1), f_1(7))$ (THE NUMBER IN BRACKETS CORRESPONDS TO IMAGE SLICE NUMBER)

a. Training

True	Predicted AD	Predicted Normal	correct
AD	21	3	88%
Normal	3	34	92%
		Total	90%

b. Testing

True	Predicted AD	Predicted Normal	correct
AD	21	3	88%
Normal	5	32	86%
		Total	87%

accuracy of the training data. This is desirable as we do not really want to over-fit the data. As a result, the classification accuracy on the testing data, which is the most important one, improves even only slightly. The classification accuracy between training and testing is much more similar when using fewer features. In addition, notice that using many features biases the system towards a false negative thus reducing its sensitivity, which is very important in a real diagnostic situation.

Further reduction in the number of features used does not improve the performance but in fact deteriorates it. This leads to the conclusion that using too many features over-fits the data, so the predictive ability of the system for new data is somewhat limited. On the other had using too few features does not describe the classes adequately hence its performance is low in training and testing. The best performance appears to be when combining 3 features. These features are the values of the monomials f_1 and f_3 computed for slices 1 and 7.

It is a well known fact that AD affects the hippo-campus and cerebral cortex most vigorously [2]. The best 3 features correspond to slices 1 and 7 which depict these parts of the brain.

Although AD affects the brain globally, various regions are more affected than others. The position of these regions in the brain vary slightly between patients. A fully global feature extraction method will not take into account this. However, the local rotation and translation invariance of the monomial method would, and this explains its performance.

ACKNOWLEDGMENTS

An EPSRC studentship supporting this project is gratefully acknowledged.

REFERENCES

[1] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M Stadlan. Clinical Diagnosis of Alzheimer's Disease; Report of the NINCDS-ADRDA Work Group. *Neurology*, 34:939–944, 1984.

[2] National Institute of Aging and National Institute of Health. *Progress Report on Alzheimer's Disease*, 1999. http://www.alzheimers.org/.

[3] M. Schael and H. Burkhardt. Error Detection on Textures using Invariant Grey-scale Features. In *Workshop on Texture Analysis*, pages 165–179, Germany, 1998.

[4] H. Schulz-Mirbach. On the Existence of Complete Invariant Feature Spaces in Pattern Recognition. In *In Proc. of the 11th International Conference on Pattern Recognition, Conference B: Pattern Recognition Methodology and Systems*, volume 2, pages 178–182, 1992.

[5] H. Schulz-Mirbach. Constructing Invariant Features by Averaging Techniques. In *Proc of the 126th International Conference on Pattern Recognition*, volume 2, pages 387–390, 1994.

[6] H. Schulz-Mirbach. Invariant features for gray scale images. In *DAGM - Symposium "Mustererkennung"*, pages 1–14, Institud für Technische Informatick I 21071 Hamburg, Germany, 1995.

[7] H. Schulz-Mirbach. Invariant gray scale Features. Internal Report 8/96. Technical report, Technical University of Hamburg-Harburg, Technische Informatick I 21071 Hamburg, Germany, 1996.

[8] StatSoft. *Statistica for Windows, Vol II, Manual*, 2^{nd} edition.