FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

MR Image texture analysis applied to the diagnosis of Alzheimer's Disease

Mathias Bjørn Jørgensen & Mirza Hasanbasic

6th June 2016

Contents

1 Introduction **3**
1.1 Problem Definition . . . 3

2 Data **4**

3 Method **5**
3.1 Image texture analysis methods . . . 5
3.1.1 Co-occurrence matrix . . . 5
3.1.2 Texture features from co-occurrence matrix . . . 7
3.2 Machine learning methods . . . 7
3.2.1 10-fold cross-validation . . . 7
3.2.2 Feature selection . . . 7
3.2.2.1 Naive . . . 7
3.2.2.2 Sequential Forward Feature . . . 7
3.2.3 K-nearest neighbors algorithm . . . 7
3.3 Erode . . . 8

4 Implementation **10**

5 Result **11**

6 Discussion **12**

7 Conclusion **13**

Appendices **14**

A Co occurrence matrix derivation features **15**

Litteratur **17**

List of Figures

3.1 Example of the offsets for the 2D 5

3.2 Example how the values in the GLCM are calclulated of the 4-by-5 image I. Element (1, 1) in the GLCM contains the value 1 because there is only one instance inf the image where two horizontally adjacent pixels have the values 1 and 1. . 6

3.3 Example of the offsets for the 3D 7

3.5 Text 9

3.7 Text 9

List of Tables

List of Corrections

Note: find reference 5
Note: level or intensity? 5
Note: Lave 3x3 matrice med offsets. 6
Note: show 3x3 matrix with angles clearly visible 6
Note: Show example on a 5x5 matrix with GI 8, offset [1 0] and [0 1] 6
Note: 3 gange 3 gange, farv de ønskede dele eller noget 6
Note: how does this remove overfitting? 7

Mathias Bjørn Jørgensen & Mirza Hasanbasic

Abstract

This report will examine MRI scans of brains, using image texture analysis and machine learning

Chapter 1

Introduction

Alzherimer's Disease (AD) is the most common cause of dementia among people and is a growing problem in the aging populations. It has a big impact on health services and society as life expectancy increases. In 2010 the total global costs of dementia was estimated to be about 1% of the worldwide gross domestic product[1]. AD is the cause in about 60%-70% of all cases of dementia[1] and about 70% of the risk is believed to be genetic [2]. Currently there are no way to cure dementia or to alter the progressive course. But however, much can be done to support and improve the lives if AD is diagnosed in the early stage of progression [1]

In this report we will examine MRI data of the hippocampus using image texture analysis and apply machine learning in order to diagnose AD in patients. Our dataset contain 100 patients 50 control and 50 with AD.

We will be using two different image texture analysis method, one which will me in 2D[3][4] and the other one will be in 3D[5], from which we calculate the data to the gray level co-occurrence matrix (GLCM).

1.1 Problem Definition

Is it possible to classify MRI data of the hippocampus into groups of healthy controls vs Alzheimer's patient, using a predefined set of image texture metrices, with an accuracy greater than 80%?

Is there a difference in diagnosing AD successfully by calculating the co-occurence matrix in 3D compared to 2D.

[1]With the terms as of direct medical costs, direct social costs and costs of informal care

Chapter 2

Data

The MRI head scans were acquired on a General Electric 3-T for all our 100 patients, were we got 50 normal subjects and 50 AD patients.

Chapter 3

Method

3.1 Image texture analysis methods

3.1.1 Co-occurrence matrix

The co-occurence matrix (COM) is second-order statistics methods, which is based on information about gray levels in pair of pixes. The matrix is defined over the image with distribution values at a given offset. Mathematically we have a COM matrix **C** which is defined over an $n \times m$ image **I**, with $\Delta x, \Delta y$ being the parameterized offset, is calculated by

$$C_{\Delta x, \Delta y}(i,j) = \Sigma_{p=1}^{n} \Sigma_{q=1}^{m} \begin{cases} 1, & \text{if } I(p,q) = i \text{ and } I(p+\Delta x, q+\Delta y) = j \\ 0, & \text{otherwise} \end{cases}$$

FiXme Note: find reference

The element (5,4) in the COM can be translated to meaning how many times there exist an element in the image with GI **FiXme Note: level or intensity?** 5 and another element offset $\Delta x, \Delta y$ from the originial with greyscale intensity (GI) 4, i.e. if the offset is (,1) and the first element is (x,y)(4,3) with GI 5 it would mean that element (x,y)(5,3) would have GI 4. If COM(4,4) is ten, it translates into there being ten instances with element (x,y) = 5 and (x+Δx,y+Δy) = 4.

FiXme Note: find reference
FiXme Note: level or intensity?

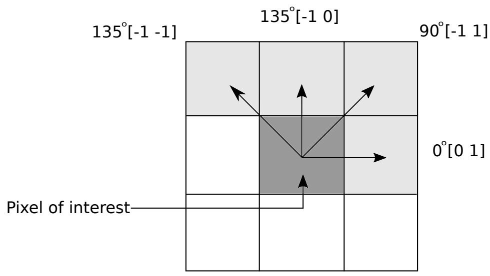


Figure 3.1: Example of the offsets for the 2D

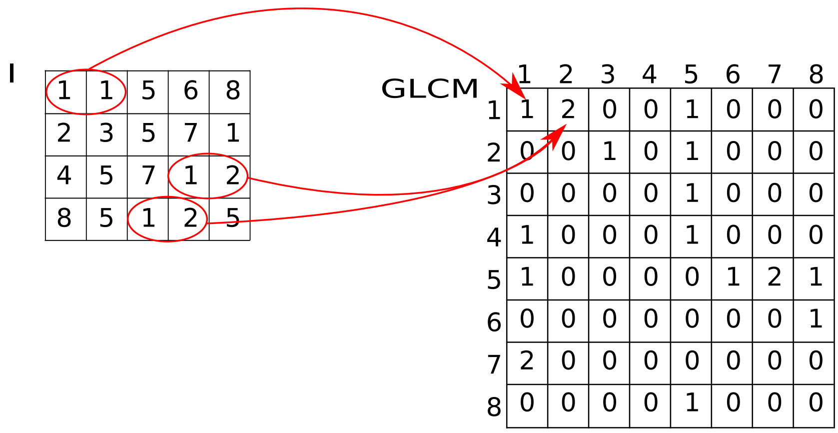


Figure 3.2: Example how the values in the GLCM are calclulated of the 4-by-5 image I. Element (1, 1) in the GLCM contains the value 1 because there is only one instance inf the image where two horizontally adjacent pixels have the values 1 and 1.

A single image have multiple COMs as different offsets creates different relations. Consider a 3 × 3 matrix looking at element (2,2) we can then create eight different offsets, (1,0),(1,1),(0,1)),(-1,1),(-1,0),(-1,-1),(0,-1),(1,-1), however they are not unique. **FiXme Note: Lave 3x3 matrice med offsets.**

FiXme Note: Lave 3x3 matrice med offsets.

Focusing on the two offsets (0,1), (0,-1) in element (2,2) and (1,2) with GI 1 and 2 respectfully increases the entry $\text{COMs}_{1,0}(1,2)$ and $\text{COMs}_{-1,0}$ (1,1) with one, showing that ${\text{COMs}_{0,1}}^{\mathbf{T}}$ = $\text{COMs}_{0,-1}$. There exist the same relation between (1,1)(-1,-1), (0,1)(0,-1), and (-1,1)(1,-1). This leaves four different offsets for analysis (0,1),(-1,1), (-1,0)),(-1,-1) in general (0,d),(-d,d),(-d,0),(-d,-d) where d is the distance which are commonly named angles 0°, 45°, 90° and 135°. **FiXme Note: show 3x3 matrix with angles clearly visible**

FiXme Note: show 3x3 matrix with angles clearly visible

The co-occurrence matrix is quadratic with the number of rows and columns equal to the amount of GI, for example if we have 256 GI we get a 256 × 256 COM. nd [0 1]]FiXme Note: Show example on a 5x5 matrix with GI 8, offset [1 0] and [0 1]**FiXme Note: Show example on a 5x5 matrix with GI 8, offset [1 0] and [0 1]**

a

Extending this method to three-dimensions it is necessary to look on how the offsets are defined because the size of the COM is defined by the amount of GIs and not by the images it is derived from. Considering a 3 × 3 × 3 matrix we have a possible of 26 offsets. In two-dimensions it is possible to eliminate half of the offsets because of the relation $\text{COM}^{\mathbf{T}}_{d,d} = \text{COM}_{-d,-d}$, and it is the same case in three-dimensions with the relation being $\text{COM}^{\mathbf{T}}_{d,d,d} = \text{COM}_{-d,-d,-d}$. This leaves 13 offsets which are illustrated below.
FiXme Note: 3 gange 3 gange, farv de ønskede dele eller noget

FiXme Note: 3 gange 3 gange, farv de ønskede dele eller noget

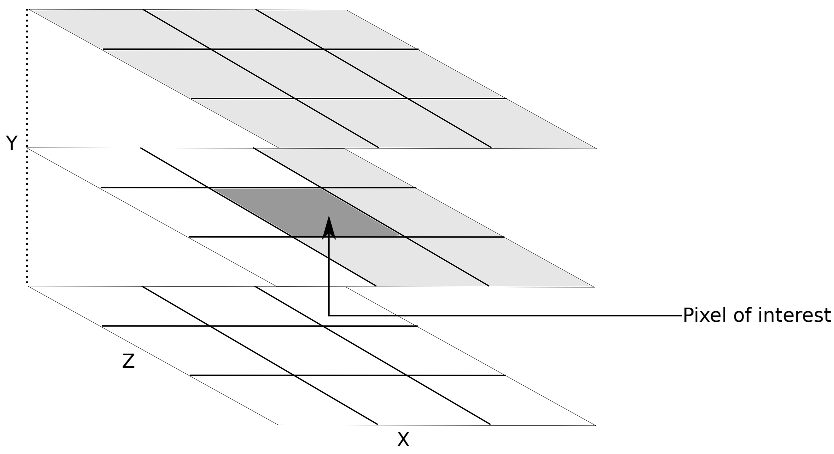


Figure 3.3: Example of the offsets for the 3D

3.1.2 Texture features from co-occurrence matrix

3.2 Machine learning methods

3.2.1 10-fold cross-validation

Given a model with unknown paramers and a training set to which the model can be fit then the fitting process optimizes the models parameters to fit the training data. Validating this model against independent data (test data) from the same data pool as the training, it will generally turn out that the model does not fit the test data as well as the training data. It is known as overfitting and is a problem when the size of the training data set is small. Cross-validation is used to counteract overfitting.

Dividing the entire data set into 10 groups at random, one subsample is saved for testing and the reamining nine is used to fit the model. The procedures is done so all subsamples get to be used to validate exactly one time and the validation result can then be averaged to produce a single estimation. This solves the problem of overfitting, as the validation data set is never used in to fit the model. **FiXme Note: how does this remove overfitting?**

FiXme Note: how does this remove overfitting?

3.2.2 Feature selection

3.2.2.1 Naive

3.2.2.2 Sequential Forward Feature

3.2.3 K-nearest neighbors algorithm

k-NN for short is a method that is used for classification and regression. Where the output is a class and member of this class, and this object is classified by its neighbors. For instance, if we chose k to 1, then the object will be assigned to the class of the single nearest neighbor.

The algorithm consist of training examples, that are vectors in multidimensional space, with each its label. The most used distance metric is Euclidean distance.

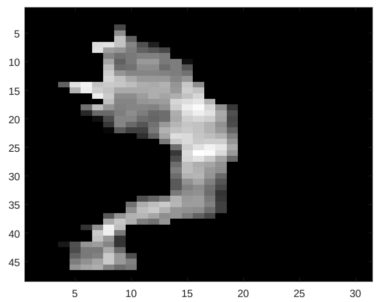
The drawback of k-NN is that classification can be skewed in that way, that the more frequent class tend to dominate the prediction of new examples, because they tend to be common among the k-NN due to their large number.

The way we wish to implement the k-NN in matlab is, first we handle the data, then we will calculate the distance between two data instances and after that, we can locate k most similar data instances and generate a response from a set. After all this is done, we have to summarize the accuracy of predictions.

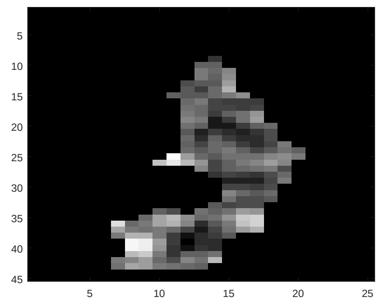
Dette vil være en lille introduktion til de tools vi bruger til at analysere vores data med. Da vores data er MRI skanninger af hjernen, som er nogen voxels[1] som bliver repræsenteret i 3D.

3.3 Erode

Each patients hippocampus has been segmented in the MRI scan. The problems we can run into are that the background will blur with the segmentation i.e. the hippocampus. Performing a erosion can solve this problem and we can focus on the hippocampus, with the maximum number of details. As seen in figure 3.4a the erosion has not been performed yet, but we might have some problems with data surrounding the hippocampus is blurring out the edges of the hippocampus. To solve this, we create a mask to separate the hippocampus from the background data and end up with what is left in figure 3.4b



(a) Hippocampus at slice 10 on the X-axis



(b) Hippocampus at slice 10 on the X-axis after erosion has been performed

The erosion we have used is called a city-block metric and can be seen in figure 3.5.

[1]noget her

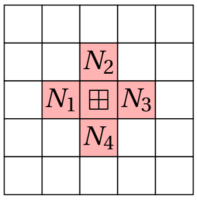
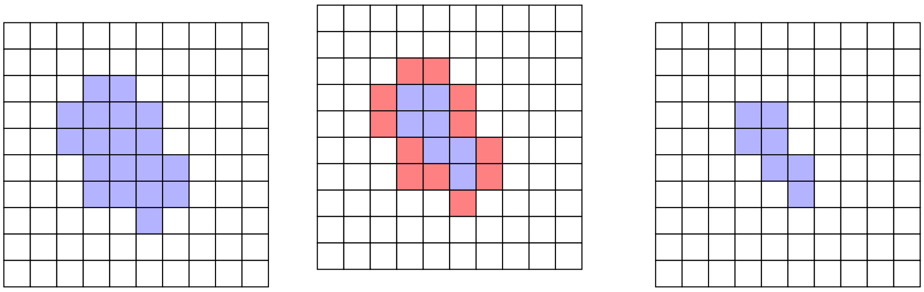


Figure 3.5: Text

To give an example of how the erosion works, it will be illustrated and we will use the city-block metric for this purpose. So we will use figure 3.5 and erode the image in figure **??**.



(a) Original image

(b) Performing erode on the image

(c) The image after erosion

As seen in figure **??** the noise (background) have been removed. This is an example in 2D. Now we wish to extend the erosion city-block to 3D. As seen in figure 3.5 it have 4 neighbours and when we extend this to 3D we will end up with 6 neighbours instead as seen in figure 3.7 and the concept is still the same as in 2D

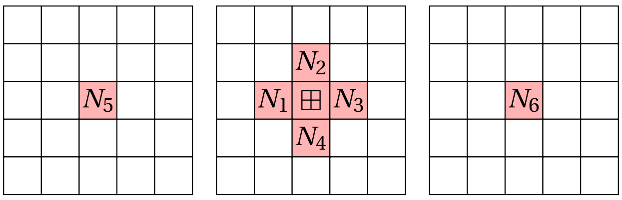


Figure 3.7: Text

Chapter 4

Implementation

Chapter 5

Result

Chapter 6

Discussion

Chapter 7

Conclusion

Appendices

Appendix A

Co occurrence matrix derivation features

$$C_x(i) = \sum_{j=1}^{N} C(i, j)$$

$$C_y(i) = \sum_{i=1}^{N} C(i, j)$$

$$C_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{N} \sum_{j=1}^{N}, \quad \text{k = 2, 3, ... , 2N}$$

$$C_{x+y}(k) = \sum_{\substack{i=1 \\ |i-j|=k}}^{N} \sum_{j=1}^{N}, \quad \text{k=0,1,...,N-1}$$

$$f_1 = \sum_{i=1}^{N} \sum_{j=1}^{N} \{C(i, j)\}^2 \tag{A.0.1}$$

$$f_2 = \sum_{n=0}^{N-1} n^2 \left\{C_{x+y}(k)\right\} \tag{A.0.2}$$

$$f_3 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n} ijC(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{A.0.3}$$

$$f_4 = \sum_{i=1}^{N} \sum_{j=1}^{N} (i - \mu)^2 C(i, j) \tag{A.0.4}$$

$$f_5 = \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{1}{1 + (i - j)^2} C(i, j) \tag{A.0.5}$$

$$f_6 = \sum_{i=2}^{2N} i C_{x+y}(i) \tag{A.0.6}$$

$$f_7 = \sum_{i=2}^{2N} (i - f_6)^2 C_{x+y}(i) \tag{A.0.7}$$

$$f_8 = \sum_{i=2}^{2N} C_{x+y}(i) \log(C_{x+y}(i)) \tag{A.0.8}$$

$$f_9 = -\sum_{i=1}^{N} \sum_{j=1}^{N} C(i,j) \log(C(i,j)) \tag{A.0.9}$$

$$f_{10} = \text{variance of } C_{x-y} \tag{A.0.10}$$

$$f_{11} = -\sum_{i=0}^{N-1} C_{x-y}(i) \log(C_{x-y}(i)) \tag{A.0.11}$$

Bibliography

[1] Dementia. http://www.who.int/mediacentre/factsheets/fs362/en/. Accessed: 2016 April.

[2] Anne Corbett Carol Brayne Dag Aarsland Emma Jones Clive Ballard, Serge Gauthier. Alzheimer's disease. *Lancet*, page 13, March 2011.

[3] Peter A. Freeborough & Nick C. Fox. Mr image texture analysis to the diagnosis and tracking of alzheimer's disease. *IEEE*, 17(3):5, June 1998.

[4] L.M. Li F. Cendes G. Castellano, L. Bonilha. Texture analysis of medical images. *Neuroimage Laboratory*, page 9, April 2004.

[5] Sanjay Kalra Rouzbeh Maani, Yee Hong Yang. Voxel-based texture analysis of the brain. *Plos One*, page 19, March 2015.