# Bag of Features for Automatic Classification of Alzheimer's Disease in Magnetic Resonance Images

Andrea Rueda, John Arevalo, Angel Cruz, Eduardo Romero, and Fabio A. González

BioIngenium Research Group, Universidad Nacional de Colombia, Bogotá, Colombia
{adruedao,jearevaloo,aacruzr,edromero,fagonzalezo}@unal.edu.co

**Abstract.** The goal of this paper is to evaluate the suitability of a bag-of-feature representation for automatic classification of Alzheimer's disease brain magnetic resonance (MR) images. The evaluated method uses a bag-of-features (BOF) to represent the MR images, which are then fed to a support vector machine, which has been trained to distinguish between normal control and Alzheimer's disease. The method was applied to a set of images from the OASIS data set. An exhaustive exploration of different BOF parameters was performed, i.e. feature extraction, dictionary construction and classification model. The experimental results show that the evaluated method reaches competitive performance in terms of accuracy, sensibility and specificity. In particular, the method based on a BOF representation outperforms the best published result in this data set improving the equal error classification rate in about 10% (0.80 to 0.95 for Group 1 and 0.71 to 0.81 for Group 2).

## 1 Introduction

Alzheimer's disease (AD) is a slow, progressive and incurable brain disease, characterized by a progressive memory impairment together with reasoning, planning, language, and perception disturbances. The clinical onset is still poorly understood and the diagnosis is mainly accomplished using psychological tests that become positive when the disease is practically irreversible. Yet MR (Magnetic resonance) images are a useful information source to study the neurodegenerative process in specific regions [13], their role is quite blurry in the early disease stages. Lately, several researchers have focused on automatic detection of the Alzheimer's disease in MR, aiming to identify the early stages by investigating certain subtle anatomical changes of the primary cortex, looking for morphological biomarkers.

The main problem regarding a proper AD characterization is the large brain anatomical variability. This noise is increased with some physiological processes, for instance the aging itself leads to a natural neurodegenerative process. An opportune diagnosis is crucial towards delaying the neurodegenerative process and depends upon the possibility of establishing objective differences among brains, a very up-to-date research problem known as morphometry. This brain comparison has been previously approached by characterizing spatial regions with different estimations of the differences between a particular brain and a statistical brain template: voxel-based morphometry [1,18], tensor-based morphometry [2,16,6,15,10] and object-based morphometry [11,14]. Recently, it has been proposed to compare relevant brain features rather than anatomical

L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 559–566, 2012.

structures, a technique known as Feature-Based Morphometry (FBM) [17]. This method combines local, spatial and scale invariant features and a probabilistic model to characterize the relative differences between groups, reporting an Equal Error Classification Rate (EER) of about 0.80. The objective of these methods is to diminish the amount of recorded noise by determining a set of zones that may be invariant under different noise conditions.

This article establishes brain similarities by calculating the differences of a part-based brain representation, the Bag Of Features (BOF). The basic idea behind this method is to represent the image visual content as a probability distribution (histogram) of local features (visual words) and collect a knowledge base from a set of images, previously labeled. In fact, this approach has been successfully used in medical image analysis. [9] used a BOF approach in classification of histopathology images using local features (SIFT, raw-patches) and a new strategy that separated semantically the basic stain components. Cruz-Roa et al. in [7] proposed a strategy for automatic visual mining of histopathology images using a BOF representation. The results showed that BOF is a good alternative for representing histology images. It allows to extract implicit patterns and use them to perform automatic annotation, reaching a F-score of 90%. This method was extended using a non-negative matrix factorization from a BOF image representation [8].

This paper evaluates the BOF approach to represent 3D MRI images for automatic Alzheimer's disease classification. For doing so, an exhaustive exploration of different variations from classical BOF approaches was performed and compared with our proposed method. Main contributions of this paper are: an adapted BOF method, an exhaustive evaluation of each BOF stage (feature detection and description, dictionary construction, classification method) from different BOF approaches for 2D and 3D images, and a new strategy from automatic classification of Alzheimer's disease.

The rest of paper is organized as follows: Section 2 presents the proposed method, Section 3 presents the MRI Alzheimer data set and the experimental setup used to evaluate the classification performance of our method. Section 4 shows the experimental results and discussion. Finally Section 5 presents the conclusions and future work.

## 2 Automatic BOF Classification of Alzheimer's Disease

The proposed method, illustrated in Figure 1, is composed of two main phases: first, a BOF image representation of MRI images, and, second, an AD automatic classifier, i.e. a Support Vector Machine (SVM) with a RBF (Radial Basis Function) Kernel trained to distinguish between normal controls or subjects diagnosed with Alzheimer's disease.

### 2.1 BOF Image Representation

A BOF represents an image as a frequency histogram of an unordered collection of individual regions (patches or blocks). This collection is constructed by selecting the most common regions in a whole image collection. Extracted regions correspond to squared image pieces, known as image patches or local image features, which are vectorized (linearized) by concatenating each patch row one after the other. With these patches,

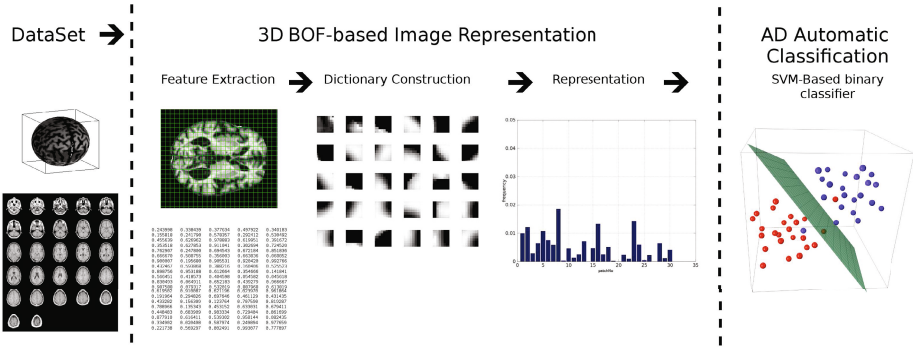



**Fig. 1.** The proposed method with BOF MRI scans and a Binary SVM Classifier

a visual dictionary is built by applying a clustering algorithm, and each cluster centroid will represent the set of patches as a visual word. The complete representation procedure is described as follows:

1. *Local feature detection and description:* In the first step, the main goal is to detect visual patterns and to describe each image in terms of their visual appearance. In our particular case, the whole 3D MRI volume is processed in a slice-by-slice basis, following the volume acquisition direction. The detection method starts by defining a regular 2D grid on the whole slice, and extracting every patch, which is then linearized as a vector with the gray pixel intensities. Patch description is enriched with spatial information, by also concatenating each vector 3D coordinates (x, y, z). As the original BoF representation does not take into account the spatial location of each patch, including this information into the descriptor could be useful in posterior stages. Patches belonging to the background (black homogeneous patches) are discarded, i.e., patches with null mean and variance. In addition, following the proposed scheme in [3], a normalization process is performed on each patch, in order to deal with luminance variations of MRI volumes, through a transformation for each point $x_i$ in a patch $X$ with null mean $\mu$ and unitary standard deviation $\sigma$ as follows:
$$x_i' = \frac{(x_i - \mu)}{\sigma}$$
2. *Dictionary construction*: The second step comprises a clustering k-means algorithm on the extracted patches, allowing to construct a visual dictionary by finding the most representative patches in the image collection. Each cluster centroid is then considered as a visual word of the dictionary.
3. *Histogram image representation*: As mentioned before, the idea behind the BOF model is to represent an image using the most common regions in the whole image collection. For a given dictionary, it is possible to describe an image as the frequency of visual words. Then, the final image representation will be a histogram with as many elements as dictionary visual words. The complete 3D representation of a MR volume is obtained by merging each 2D slice histograms as a single 3D histogram.

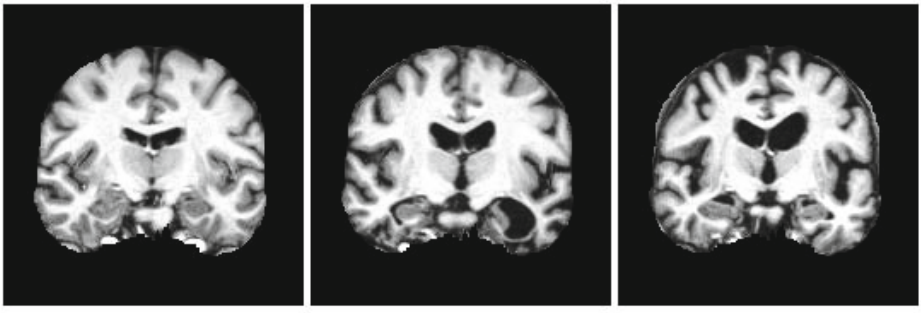


**Fig. 2.** Coronal example slices of brain MR volumes from OASIS data set. Left: normal control subject, Middle: patient diagnosed with very mild AD, Right: patient diagnosed with mild AD.

### 2.2 SVM-Based Binary Classifier

A first approximation to the problem of classifying Alzheimer's cases is binary since the clinical meaningful task consists in detecting the early stages, i.e., when no clinical sign is present at all and the disease is still hidden. At this point, anatomical information may be valuable to determine any abnormality. From a machine learning perspective, for the binary classification problem, patients classified with Alzheimer's disease are modeled as the positive class, while the normal controls corresponds to the negative class. Every histogram bin is considered as a model feature, trained by an RBF Kernel, already used in some biomedical problems [5].

## 3 Experimental Setup

A set of brain MR images from healthy and pathological subjects were extracted from the OASIS (Open Access Series of Imaging Studies) database [12]. Each subject has been previously analyzed with a Mini-Mental State Examination (MMSE) and a Clinical Dementia Rating (CDR), and diagnosed as normal controls (NC) or with probable Alzheimer's disease (AD) using the scores obtained in the MMSE and CDR tests. The OASIS database provides a number of images per subject, from which we have selected the skull-stripped gain-field corrected atlas-registered image to the 1988 atlas space of Talairach and Tournoux [4]. To evaluate the performance of the proposed approach, results are reported on a subset of the available images composed by two different divisions:

1. *Group 1*: 86 subjects aged 60-80 years, mild AD ($CDR = 1$): 20 AD, 66 NC
2. *Group 2*: 136 subjects aged 60-80 years, mild and very mild AD ($CDR = \{1, 0.5\}$): 70 AD, 66 NC

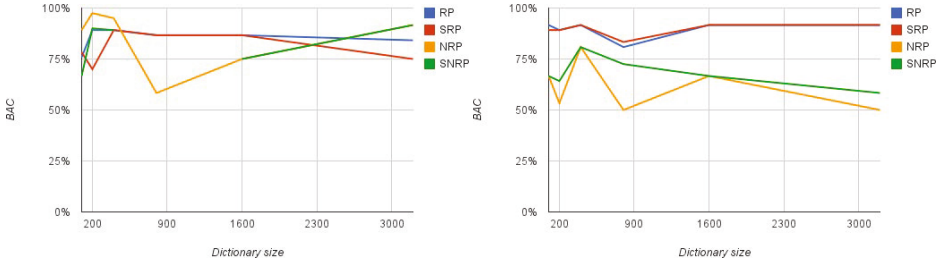
Figure 2 shows coronal example slices of MR volumes from a normal subject, a patient clinically diagnosed with very mild AD and a patient clinically diagnosed with mild AD.

The local feature detection was performed on a regular grid (without patch overlapping) with patch sizes of $8 \times 8$ and $16 \times 16$ pixels. The used descriptor consists of a linearized patch, known as *raw patch*, in four different configurations according to the inclusion of spatial information and the patch normalization: Raw Patch (RP), Normalized Raw Patch (NRP), Spatial Raw Patch (SRP) and Spatial Normalized Raw Patch (SNRP). Dictionary construction was carried out with a $k$-means algorithm for different dictionary sizes, $k$ = 100, 200, 400, 800, 1600 and 3200 visual words. For the classification phase, each group was divided into two sets: training (70%) and test (30%), maintaining the class proportions within each set. With the training set, SVM parameters ($C$ and $\gamma$ in RBF) were adjusted using a 10-fold cross-validation, and the best parameter combination was used to train the SVM model. Finally, images in the test set were classified using the trained model. Classification performance evaluation was calculated using standard classification measures, such as sensibility (SEN), specificity (SPC), balanced accuracy (BAC) and equal error classification rate (EER). Following [17], the EER is defined as the value of the true positive rate corresponding to the point of the ROC curve where the false positive rate is equal to the false negative rate.

## 4 Results and Discussion

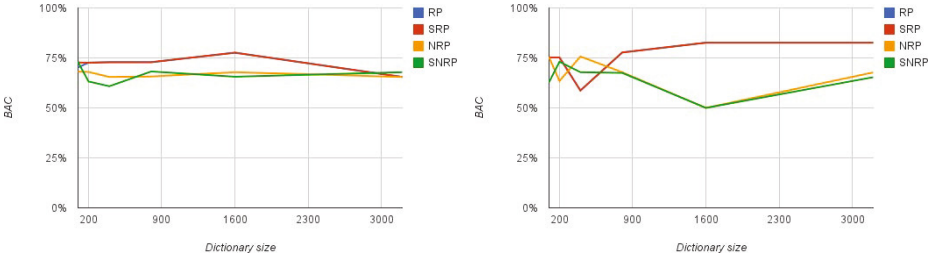
Figures 3 and 4 presents the obtained performance in terms of BAC measure for each configuration in Group 1 and Group 2, respectively. These results suggest that normalization process of patches worsens the discrimination capability of BOF representation. This could be explained because the visual variability of local patches is reduced whereas without normalization the visual appearance of patches is wider.

Other interesting results are that spatial information does not seem to contribute significantly to distinguish between normal control and Alzheimer in this image representation approach. A probable explanation is that BOF is a histogram of local patterns occurrences, given by the set of visual words of the dictionary, which captures the representative visual variations to represent the MRI volume. However the fact of including spatial patch description information increases the number of possible spatially located




(a) $8 \times 8$ patch size (b) $16 \times 16$ patch size

**Fig. 3.** Performance comparison, in terms of BAC, in Group 1 for each feature descriptor: Raw Patch (RP), Normalized Raw Patch (NRP), Spatial Raw Patch (SRP) and Spatial Normalized Raw Patch (SNRP), varying the dictionary size for patch sizes of $8 \times 8$ (left) and $16 \times 16$ (right)




(a) $8 \times 8$ patch size

(b) $16 \times 16$ patch size. RP and SRP have the same values.

**Fig. 4.** Performance comparison, in terms of BAC, in Group 2 for each feature descriptor: Raw Patch (RP), Normalized Raw Patch (NRP), Spatial Raw Patch (SRP) and Spatial Normalized Raw Patch (SNRP), varying the dictionary size for patch sizes of $8 \times 8$ (left) and $16 \times 16$ (right)

**Table 1.** Summary of best results in terms of EER for the two evaluated groups. Value within parenthesis refers to the best dictionary size for the corresponding configuration. The baseline is reported at top, Feature Based Morphometry (FBM), and the different feature descriptors combined with patch size are reported in the following rows: Raw Patch (RP), Normalized Raw Patch (NRP), Spatial Raw Patch (SRP) and Spatial Normalized Raw Patch (SNRP).

| | | Group 1 | Group 2 |
|---|---|---|---|
| Baseline (FBM) | | 0.80 | 0.71 |
| $8 \times 8$ | RP | 0.90 (1600) | **0.81 (400)** |
| | NRP | **0.95 (200)** | 0.71 (400) |
| | SRP | 0.90 (200) | **0.81(400)** |
| | SNRP | **0.95 (100)** | 0.71 (400) |
| $16 \times 16$ | RP | **0.95 (1600)** | **0.81 (100)** |
| | NRP | 0.90 (1600) | 0.76 (400) |
| | SRP | **0.95 (1600)** | **0.81 (400)** |
| | SNRP | 0.80 (400) | 0.76 (400) |

visual patterns. Then the dictionary size required when spatial information is added must be larger.

Under these results results, it is clear that the improvement obtained when the dictionary size increases (two times) is mild, whereas patch sizes of $16 \times 16$ provide a better performance than $8 \times 8$ patch size. These results suggest that a moderate wider local region is better to detect the local visual changes that characterize the Alzheimer of the normal controls.

Table 1 reports the EER for different configurations using the best dictionary size. At the top, the first classification results on the OASIS dataset, reported by Toews et al. with the FBM approach are included as a baseline [17]. It is important to notice that Toews' result were produced using a leave-one-out evaluation setup, different to the setup used in this work, which has a test data set with a stratified sample of the 30% of the data. The leave-one-out strategy generally overestimates the performance of the

evaluated model, in fact some of the configurations got 98% EER when evaluated using leave-one-out cross validation, however we chose to report the results with our setup since it generates a better estimation of the classifier performance. In both groups, the proposed strategy outperforms the baseline.

In general, the results suggest that the best classification results can be obtained by using a patch size of $16 \times 16$ and a dictionary size between 100 and 400. The inclusion of spatial information into the descriptor vector did not show a clear advantage, but normalization seems to degrade the results.

## 5 Conclusion and Future Work

This paper presents a BOF image representation scheme for brain MR images, which, combined with a SVM, allows classification of normal controls and patients clinically diagnosed with Alzheimer's disease. The experimental results shows that the proposed method has a competitive performance that improves results reported by state-of-art methods. The results are encouraging and suggest that the BOF representation has the ability to capture visual patterns useful for discriminating healthy MR brain volumes from those exhibiting the Alzheimer's disease. Our future work includes a thorough evaluation including additional data sets, an exploration of feature descriptors such as SIFT, HOG and feature combinations, among others, the analysis of the visual dictionary to find the most discriminating visual words that characterize the Alzheimer's disease, and improvements to the method to enhance its interpretability paving the way to automatic methods that could effectively support clinical diagnosis.

**Acknowledgements.** This work was partially funded by the project "Anotación Automática y Recuperación por Contenido de Imágenes Radiológicas usando Semántica Latente" number 110152128803 and project "Sistema para la Recuperación de Imágenes Médicas utilizando Indexación Multimodal" number 110152128767 by Convocatoria Colciencias 521 de 2010.

## References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry–the methods. Neuroimage 11(6), 805–821 (2000)
2. Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K.: et al. Identifying global anatomical differences: deformation-based morphometry. Human Brain Mapping 6(5-6), 348–357 (1998)
3. Avni, U., Greenspan, H., Sharon, M., Konen, E., Goldberger, J.: X-ray image categorization and retrieval using patch-based visualwords representation. In: IEEE International Symposium on Biomedical Imaging, ISBI 2009, pp. 350–353. IEEE (2009)
4. Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z.: A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 23(2), 724–738 (2004)

5. Caicedo, J.C., Cruz, A., Gonzalez, F.A.: Histopathology Image Classification Using Bag of Features and Kernel Functions. In: Combi, C., Shahar, Y., Abu-Hanna, A. (eds.) AIME 2009. LNCS, vol. 5651, pp. 126–135. Springer, Heidelberg (2009)
6. Chung, M.K., Worsley, K.J., Paus, T., Cherif, C., Collins, D.L., Giedd, J.N., Rapoport, J.L., Evans, A.C.: A unified statistical approach to deformation-based morphometry. NeuroImage 14(3), 595–606 (2001)
7. Cruz-Roa, A., Caicedo, J.C., González, F.A.: Visual pattern mining in histology image collections using bag of features. In: Artificial Intelligence in Medicine (2011)
8. Cruz-Roa, A., Díaz, G., Romero, E., González, F.A.: Automatic Annotation of Histopathological Images Using a Latent Topic Model Based On Non-negative Matrix Factorization. J. Path Inform. 2(1), 4 (2011)
9. Diaz, G., Romero, E.: Micro-structural tissue analysis for automatic histopathological image annotation. Microscopy Research and Technique, pp. 343–358 (2011)
10. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 21(1), 46–57 (2004)
11. Mangin, J.F., Riviere, D., Cachia, A., Duchesnay, E., Cointepas, Y., Papadopoulos-Orfanos, D., Collins, D.L., Evans, A.C., Régis, J.: Object-based morphometry of the cerebral cortex. IEEE Transactions on Medical Imaging 23(8), 968–982 (2004)
12. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: 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(9), 1498–1507 (2007)
13. Nestor, P.J., Scheltens, P., Hodges, J.R.: Advances in the early detection of alzheimer's disease. Nature Reviews Neuroscience (2004)
14. Pohl, K.M., Sabuncu, M.R.: A Unified Framework for MR Based Disease Classification. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 300–313. Springer, Heidelberg (2009)
15. Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V.: Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change. IEEE Transactions on Medical Imaging 25(5), 626–639 (2006)
16. Thompson, P.M., Giedd, J.N., Woods, R.P., MacDonald, D., Evans, A.C., Toga, A.W., et al.: Growth patterns in the developing brain detected by using continuum mechanical tensor maps. Nature 404(6774), 190–192 (2000)
17. Toews, M., Wells III, W., Collins, D.L., Arbel, T.: Feature-based morphometry: Discovering group-related anatomical patterns. NeuroImage 49(3), 2318–2327 (2010)
18. Toga, A.W., Thompson, P.M., Mega, M.S., Narr, K.L., Blanton, R.E.: Probabilistic approaches for atlasing normal and disease-specific brain variability. Anatomy and Embryology 204(4), 267–282 (2001)