CO-NEIGHBOR MULTI-VIEW SPECTRAL EMBEDDING FOR MEDICAL CONTENT-BASED RETRIEVAL

Hangyu Che[1], Sidong Liu[1], Student Member, IEEE, Weidong Cai[1], Member, IEEE, Sonia Pujol[2], Ron Kikinis[2], Dagan Feng[1], Fellow, IEEE

[1] BMIT Research Group, School of IT, University of Sydney, Australia
[2] Surgical Planning Laboratory, Brigham & Women's Hospital, Harvard Medical School, Boston, USA

ABSTRACT

Multimodal medical data from various information sources are often used to depict patients. We refer to each source as a 'view'. Multi-view features could provide complementary information to each other; thus by fusing the multi-view features, we could greatly enhance the current medical content-based retrieval framework. In this paper, we propose a Co-neighbor Multi-view Spectral Embedding (CMSE) algorithm, which is an advanced feature fusion method based on the multi-view spectral analysis. CMSE aims to seek a smooth embedding for the multi-view features by maximizing the neighborhood affinity across all feature spaces. We evaluated the proposed CMSE algorithm using a freely available neuroimaging database, ADNI, with 331 pre-diagnosed subjects. Totally, 9 views of features were extracted for validation, and an improved retrieval performance was achieved over other state-of-art feature fusion methods.

Index Terms— Neuroimaging, dimensionality reduction, spectral embedding, multiple views

1. INTRODUCTION

Medical Content-Based Retrieval (MCBR) is at the intersection of computer vision, database and medical informatics [1, 2]. The underlying function of MCBR systems is to allow us to access a large database of pre-diagnosed patients and return the most similar images with the query for clinical decision support. Currently, the medical data of various sources are often used in patient diagnosis; we call each source as a 'view'. Take neurodegenerative disease as an example, assuming there are multiple available biomarkers sensitive to the disease, thus the multi-view data extracted from different biomarkers could provide complementary information to each other which is beneficial to a more accurate diagnosis. If multi-view data can be fully utilized by optimally fusing them, the performance of the MCBR and other problem domains will be greatly enhanced compared to using single view data.

As mentioned in [3], there are roughly two categories of multi-view data fusion methods. One category is *feature fusion*, which combines or aligns multiple feature spaces into one for further analysis [4]. The other kind is *decision fusion* which first sends each view as input to an analysis unit (*e.g.* classifier) and then derives the final result (*e.g.* classification label) by selecting the most prevailing decision among all views [5]. It is usually preferable to adopt feature fusion, since decision fusion requires repetitive training for different analysis units.

A simple way to fuse multi-view features is to select the most discriminant dimensions based on the feature selection criteria, such as Elastic Net (EN) [6], and form a new basis using the selected dimensions. Another way is to embed the high-dimensional concatenated feature space into a low-dimensional space using the dimensionality reduction algorithms, like ISOMAP [7]. However using these methods may submerge the characteristic of the individual view.

Several multi-view feature fusion methods using spectral embedding have been reported recently. In [8], the distributed spectral embedding (DSE) algorithm was proposed. The basic idea of DSE is to optimize the final low-dimensional embedding through minimizing the sum of its distances from the original feature spaces. Another study signed a new solution, multi-view spectral embedding (MSE) [9]. MSE uses k nearest neighbors to help keep the important natural information of each view and combine multiple views by adding relative weights.

In this paper, we demonstrate an enhanced unsupervised solution based on MSE, called Co-neighbor Multi-view Spectral Embedding (CMSE). We consider multimodal superiority at neighbor searching stage with an assumption that two instances are more likely to be in the same class if they are neighbors across multiple views. The proposed CMSE was validated using the Alzheimer's Disease Neuroimaging Initiative (ADNI) [10] baseline cohort with 331 pre-diagnosed subjects in 4 disorder groups. There were totally 9 views of features extracted. The preliminary results showed an improved retrieval performance compared to other state-of-the-art multi-view feature fusion methods.

2. METHOD

2.1 Data, Prepressing and Multi-view Feature Extraction

Neuroimaging data obtained from the Alzheimer's Disease

Neuroimaging Initiatives (ADNI) database were used in this study. There are 331 patients, each with a T1-weighted magnetic resonance imaging (MRI) image and a positron emission tomography (PET) image [11]. The patients were diagnosed as one of the two types of neurodegenerative disorders – Alzheimer's disease (AD) and mild cognitive impairment (MCI), or labeled as normal control (NC). MCI is usually considered as a prodromal stage of AD, and was further classified as MCI non-converter (*nc*MCI) and MCI converter (*c*MCI) depending on whether patients converted to AD in 3 years.

All the 3D MRI and PET images were initially converted to ADNI format [12]. As shown in Fig.1, (1) The PET images were aligned to the corresponding MRI images [13]. (2) We registered the MRI images to the ICBM_152 template [14] using image registration toolkit (IRTK) [15]. (3) The output coefficients from IRTK were applied to wrap the aligned PET images into the template space. Finally all 83 brain functional regions were mapped to each registered MRI and PET image using multi-atlas propagation with enhanced registration [16].

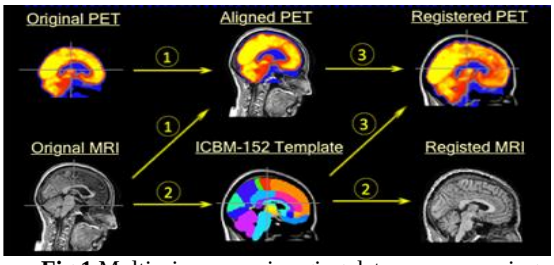

Fig.1 Multi-view neuroimaging data pre-processing

Totally 9 views of features were extracted from each brain region of interest (ROI). MRI features include the brain region volume, convexity and solidity [17]. PET features include the cerebral metabolism rate of glucose (CMRGLC) [18], mean index ratio [19], fisher discriminant ratio [20], lesion area, contrast and mean value extracted using difference of Gaussian (DOG) operator [21, 22].

2.2 Part Optimization with Co-neighbor

Table 1: some important notations used in the paper

Notation	Explanation
X	Matrices of multiple views
$[X]_{ij}$	The (i,j)th entry of X
x_i, y_i	Feature vector of the ith instance, embedded feature vector of ith instance
$(x)_i$	The ith element of vector x
$X^{(i)}, x^{(i)}$	Data from the ith view
k	Number of neighbors to search
$k_j^{(i)}$	Number of valid neighbors of the jth element in the ith view
m	Number of views
m_i	Dimensionality of feature from the ith view
n	Number of instances
d	Final dimensionality
v_{jl}	Vector of view indexes where lth instance is neighbor of jth instance
Y	The embedded matrix(final result of CMSE)
$\dim(\cdot)$	Dimensionality of a vector
$tr(\cdot)$	Trace of a square matrix

Table 1 lists the important notations will be used in the rest of this paper; some notations are inherited from MSE algorithm to make it easily understood.

The traditional MSE is based on the assumption that neighbors in the feature space are more likely with the same label. We conceive an improved assumption as: if two instances are neighbors across multiple views, the possibility that they are in the same class is much higher than being neighbors in a single view. Therefore, we perform a voting among multiple views in neighbor selection, only the neighbors who satisfy the voting criteria ($\geq$ half views vote for the neighbor) will be kept. The problem can be refined as: given a set of matrices $X = \{X^{(i)} \in \Re^{m_i \times n}\}_{i=1}^{m}$, where $X^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}]$, CMSE aims to find a low-dimensional embedding $Y \in \Re^{d \times n}$ of X.

Given the ith view $X^{(i)}$, consider an arbitrary instance $x_j^{(i)}$ and its k nearest neighbors; suppose $x_l^{(i)}$ is one of k nearest neighbors of $x_j^{(i)}$ in the ith view, 'Co-neighbor' method finds all views where x_l is one of k nearest neighbors of x_j and use the indexes of these views to build up the vector v_{jl}, if $dim(v_{jl}) \geq \frac{m}{2}$, x_l is kept as a neighbor of x_j in views whose index appears in v_{jl}, otherwise we discard this neighborhood relationship in all views. To increase the effectiveness of the assumption, we add the weights to distances from the target to the neighbors

$$d^*(x_j^{(i)}, x_l^{(i)}) = \|x_j^{(i)}, x_l^{(i)}\|^2 / dim(v_{jl})^{\beta}, \quad (i \in v_{jl}) \tag{1}$$

where β is a parameter to exaggerate the effect of the resemblance between $x_j^{(i)}$ and $x_l^{(i)}$. After all the neighbors of $x_j^{(i)}$ are selected, we define the patch of $x_j^{(i)}$ as $X_j^{(i)} = [x_j^{(i)}, x_{j1}^{(i)}, \dots, x_{jk_j^{(i)}}^{(i)}] \in \Re^{m_i \times (k_j^{(i)}+1)}$, where $x_{jl}^{(i)}$ is the lth neighbor of jth instance in the ith view, for each $X_j^{(i)}$ there is a part mapping $f_j^{(i)}: X_j^{(i)} \to Y_j^{(i)}$, wherein $Y_j^{(i)} = [y_j^{(i)}, y_{j1}^{(i)}, \dots, y_{jk_j^{(i)}}^{(i)}] \in \Re^{d \times (k_j^{(i)}+1)}$. In order to maintain the important information after dimensionality reduction, the locality of the low-dimensional space should be preserved from the original space. For the jth instance in the ith view, we have the part optimization function:

$$\underset{Y_j^{(i)}}{\operatorname{argmin}} \sum_{l=1}^{k_j^{(i)}} \|y_j^{(i)} - y_{jl}^{(i)}\|^2 (w_j^{(i)})_l, \tag{2}$$

$w_j^{(i)}$ is a $k_j^{(i)}$- dimensional weight vector defined as:

$$(w_j^{(i)})_l = \exp\left(-d^*\left(x_j^{(i)}, x_l^{(i)}\right)\right). \quad (3)$$

Using Eq. (3), we can reformulate Eq. (2) as below:

$$arg\min_{Y_j^{(i)}} tr\left(Y_j^{(i)} L_j^{(i)} \left(Y_j^{(i)}\right)^T\right), \quad (4)$$

where $L_j^{(i)} \in \Re^{\left(k_j^{(i)}+1\right)\times\left(k_j^{(i)}+1\right)}$ encrypt the objective function of the jth patch on the ith view,

$$L_j^{(i)} = \begin{bmatrix} \sum_{l=1}^{k_j^{(i)}} \left(w_j^{(i)}\right)_l & -\left(w_j^{(i)}\right)^T \\ -w_j^{(i)} & diag\left(w_j^{(i)}\right) \end{bmatrix}. \quad (5)$$

To take advantage of statistical property of individual view, we assign a nonnegative weight to each view. $\alpha = [\alpha_1, \dots, \alpha_m]$ represents the weights vector of m views. Value of α_i implies the importance of $X^{(i)}$.

$$\underset{Y=\left\{Y_j^{(i)}\right\}_{i=1}^m, \alpha}{\operatorname{argmin}} \sum_{i=1}^{m} \alpha_i tr\left(Y_j^{(i)} L_j^{(i)} \left(Y_j^{(i)}\right)^T\right) \quad (6)$$

2.3 Global Alignment and Optimization

$Y_j^{(i)} = \left[y_j^{(i)}, y_{j1}^{(i)} \dots, y_{jk_j^{(i)}}^{(i)}\right]$ is a low-dimensional embedding for each patch $X_j^{(i)}$, we can assume that every $Y_j^{(i)}$ is selected from a global coordinate $Y = [y_1, \dots, y_n]$, thus we can express $Y_j^{(i)}$ as $YS_j^{(i)}$, wherein $S_j^{(i)} \in \Re^{n\times k_j^{(i)}}$ is the selection matrix. Then we optimize the global coordinate alignment by summing all part optimization:

$$arg\min_{Y,\alpha} \sum_{i=1}^{m} \alpha_i tr\left(YL^{(i)}Y^T\right), \quad (7)$$

$L^{(i)} \in \Re^{n\times n}$ is the Laplacian graph matrix for the ith view which represents an undirected graph:

$$L^{(i)} = \sum_{j=1}^{n} S_j^{(i)} L_j^{(i)} \left(S_j^{(i)}\right)^T = D^{(i)} - W^{(i)}, \quad (8)$$

where $W^{(i)} \in \Re^{n\times n}$ is a symmetric matrix contains mutual relationship between every pair of instances. $\left[W^{(i)}\right]_{pq}$ represents the weighted distance between $x_p^{(i)}$ and $x_q^{(i)}$. There are three possible relationships between $x_p^{(i)}$ and $x_q^{(i)}$:

1. $x_q^{(i)}$is a neighbor of the $x_p^{(i)}$, but $x_p^{(i)}$ is not a neighbor of $x_q^{(i)}$;
2. $x_p^{(i)}$ and $x_p^{(i)}$are neighbors to each other;
3. $x_p^{(i)}$ and $x_p^{(i)}$are not neighbors.

For the above three situations, the corresponding expression for $\left[W^{(i)}\right]_{pq}$ ($\left[W^{(i)}\right]_{qp} = \left[W^{(i)}\right]_{pq}$):

1. $\left[W^{(i)}\right]_{pq} = \exp\left(-d^*\left(x_p^{(i)}, x_q^{(i)}\right)\right)$
2. $\left[W^{(i)}\right]_{pq} = \exp\left(-\min\left(d^*\left(x_p^{(i)}, x_q^{(i)}\right), d^*\left(x_q^{(i)}, x_p^{(i)}\right)\right)\right)$
3. $\left[W^{(i)}\right]_{pq} = 0$

$D^{(i)}$ is a diagonal matrix $\left[D^{(i)}\right]_{jj} = \sum_l \left[W^{(i)}\right]_{jl}$. Since $L^{(i)}$ is not normalized, we normalize it as follow:

$$L_{(nor)}^{(i)} = \left(D^{(i)}\right)^{-\frac{1}{2}} L^{(i)} \left(D^{(i)}\right)^{-\frac{1}{2}} \; s.t. YY^T = I \quad (9)$$

We set $\alpha_i \to \alpha_i^\gamma, \gamma > 1$ as in [23]. Then, we use the alternating optimization [24] to optimize α and Y. First, we fix Y to update α as:

$$\alpha_i = \frac{\left(1/tr\left(YL_{(nor)}^{(i)}Y^T\right)\right)^{1/(\gamma-1)}}{\sum_{i=1}^{m}\left(1/tr\left(YL_{(nor)}^{(i)}Y^T\right)\right)^{1/(\gamma-1)}} \quad (10)$$

Next, we update Y using

$$\underset{Y}{\operatorname{argmin}}\, tr(Y\Lambda Y^T), \quad (11)$$

where $\Lambda = \sum_{i=1}^{m} \alpha_i{}^\gamma L_{(nor)}^{(i)}$, Y is optimized as the eigenvectors with d smallest eigenvalues of the matrix Λ.

The time complexity does not increase compared with MSE, which is still $O(n^3)$. According to the aforementioned information, the algorithm can be summarized as below:

Input: multi-view data $\boldsymbol{X} = \{\boldsymbol{X}^{(i)} \in \Re^{\boldsymbol{m_i}\times\boldsymbol{n}}\}_{\boldsymbol{i=1}}^{\boldsymbol{m}}$, final dimensionality $\boldsymbol{d}$, $\boldsymbol{\gamma > 1}$ and number of neighbors k
Output: a low-dimensional embedding $\boldsymbol{Y} \in \Re^{\boldsymbol{d}\times\boldsymbol{n}}$
Algorithm:

1. Calculate $\boldsymbol{L}_{(\boldsymbol{nor})}^{(\boldsymbol{i})}$ using Eq. (9): only preserve the neighbors appear in equal to or more than half views and calculate the $\boldsymbol{d}^*$using Eq. (1).
2. Initialize $\boldsymbol{\alpha} = [\boldsymbol{1/m}, \dots, \boldsymbol{1/m}]$
3. **Repeat**
4. $\boldsymbol{Y} = \boldsymbol{U}^{\boldsymbol{T}}$, where $\boldsymbol{U} = [\boldsymbol{u_1}, \dots, \boldsymbol{u_d}]$ are the eigenvectors associated with the smallest d eigenvalues of the matrix $\boldsymbol{\Lambda}$ defined as Eq. (11)
5. Update $\boldsymbol{\alpha}$ using Eq. (10)
6. **Until convergence**

Fig. 2 The pseudo code for CMSE algorithm

3. EXPERIMENTS AND RESULTS

3.1 Optimal Features Combination

There is an optimal combination among 9 views of features, because multiple views are complementary. We searched all the combinations and found out the best group of features based on MCBR result, namely brain region volume and solidity from MRI, CMRGLC, mean index ratio, lesion area, lesion contrast and lesion mean value from PET. The best feature group was used for all methods in the experiments.

3.2 Evaluation Strategy

We used the query-by-example paradigm in this study. We evaluated the performance using the Mean Average

Precision (MAP) [25]. The number of relevant document was set to 5 and the relevance criteria were set as below:

1. rel(AD,AD)=rel(cMCI, cMCI)=rel(ncMCI, ncMCI)=1
2. rel(AD, cMCI)=rel(cMCI, ncMCI)=rel(NC, ncMCI)=0.25
3. rel(AD, ncMCI)=rel(ncMCI, AD)=0.25
4. rel(AD,NC)=0

3.3 Results

Table 2: MCBR MAP performance (%) of CMSE compared to best single view, Elastic Net (EN), ISOMAP and MSE

Algorithm	NC	*nc*MCI	*c*MCI	AD	Average
Best single view (volume)	67.00	68.59	59.01	64.84	65.32
Elastic Net (EN)	66.65	71.15	62.73	**68.91**	67.83
ISOMSP (d=100,k=40)	59.20	70.77	64.04	64.09	65.00
MSE (d=50,γ=7,k=25)	64.15	71.22	**65.09**	63.59	66.37
CMSE (d=50,γ=7,k=40, $\beta = 3$)	**69.68**	**71.59**	63.70	68.14	**68.66**

From the results shown in Table 2, CMSE achieved the best performance for NC (69.68%) and *nc*MCI (71.59%) as well as the highest average accuracy of all 4 groups (68.66%). The best single view among 7 views is "volume", which has a similar performance as ISOMAP at around 65.00%. For *c*MCI, CMSE did not achieve the best result, because *c*MCI is the class with the fewest instances, which means that the number of co-neighbors is relatively small. The supervised Elastic Net obtained the best performance of AD; whereas the performance of CMSE is very close to EN (68.14% vs. 68.91%). Overall, CMSE is very robust in all classes and might be further improved as the number of subjects grows.

4. CONCLUSION

Multi-view medical data have been increasingly used to represent the patients because multiple views are complementary to each other. In this study, we proposed a CMSE algorithm, which can search co-neighbors across multi-views. Through preliminary experiments, CMSE achieved better performance than some wildly used methods, including the supervised method, Elastic Net. Comparing with supervised methods, CMSE has a congenital advantage that CMSE processes unlabeled data, which are cheap and easy to get. Especially for MCBR, unsupervised methods could be generalized much better than supervised methods. In conclusion, our proposed 'Co-neighbor' method helps to keep the complementary information among multiple views and leads to more robust content-based retrieval.

5. ACKNOWLEDGEMENTS

This work was supported in part by the ARC, AADRF, NA-MIC (NIH U54EB005149), and NAC (NIH P41EB015902).

REFERENCES

[1] H. Müller, N. Michoux, *et al.*, "A Review of Content-based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions," Int'l J. Med. Info., vol. 73, pp. 1-23, 2004.

[2] W. Cai, J. Kim, D. Feng, "Content-Based Medical Image Retrieval", in "Biomedical Information Technology", Chapter 4, pp. 83-113, Edited by D. Feng, Elsevier, 2008.

[3] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal Fusion for Multimedia Analysis: A Survey," Multimedia Systems, vol. 16, pp. 345-379, 2010.

[4] S. Zhang, M. Yang, T. Cour, K. Yu, D. Metacas, "Query Specific Fusion for Image Retrieval," ECCV 2012, pp. 660-673, 2012.

[5] S. Liu, Y. Song, W. Cai, S. Pujol, R. Kikinis, X. Wang, D. Feng, "Multifold Bayesian Kernelization in Alzheimer's Diagnosis", MICCAI 2013, LNCS 8150, pp.303–310, 2013.

[6] L. Shen, S. Kim, Y. Qi, M. Inlow, S. Swaminathan, K. Nho, *et al.*, "Identifying Neuroimaging and Proteomic Biomarkers for MCI and AD via the Elastic Net," MBIA 2010, pp. 27-34, 2010.

[7] H. Park, "ISOMAP Induced Manifold Embedding and Its Application to Alzheimer's Disease and Mild Cognitive Impairment," Neuroscience Letters, vol. 513, pp. 141-145, 2012.

[8] B. Long, S. Y. Philip, and Z. Zhang, "A General Model for Multiple View Unsupervised Learning," SDM 2008, pp. 822-833, 2008.

[9] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview Spectral Embedding," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 40, pp. 1438-1446, 2010.

[10] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods," J. MRI, vol. 27, pp. 685-691, 2008.

[11] W. Cai, D. Feng, R. Fulton, "Content-Based Retrieval of Dynamic PET Functional Images", IEEE Trans. Info. Tech. Biomed., vol.4, no.2, pp.152-158, 2000.

[12] W. J. Jagust, D. Bandy, K. Chen, *et al.*, "The Alzheimer's Disease Neuroimaging Initiative Positron Emission Tomography Core," Alzheimer's & Dementia, vol. 6, pp. 221-229, 2010.

[13] M. Jenkinson, P. Bannister, et al., "Improved Optimization for the Robust And Accurate Linear Registration and Motion Correction Of Brain Images," Neuroimage, vol. 17, pp. 825-841, 2002.

[14] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, *et al.*, "A Probabilistic Atlas and Reference System for the Human Brain: International Consortium for Brain Mapping (ICBM)," Phil. Trans. of Royal Soci. London. B: Biol. Sci., vol. 356, pp. 1293-1322, 2001.

[15] J. A. Schnabel, D. Rueckert, et al., "A Generic Framework for Non-rigid Registration based on Non-uniform Multi-level Free-form Deformations," MICCAI 2001, 2001, pp. 573-581.

[16] R. A. Heckemann, S. Keihaninejad, P. Aljabar, *et al.*, "Automatic Morphometry in Alzheimer's Disease and Mild Cognitive Impairment," Neuroimage, vol. 56, pp. 2024-2037, 2011.

[17] P. G. Batchelor, A. D. Castellano Smith, *et al.*, "Measures of Folding Applied to the Development of the Human Fetal Brain," IEEE Trans. Med. Imag., vol. 21, pp. 953-965, 2002.

[18] W. Cai, S. Liu, L. Wen, S. Eberl, M. J. Fulham, and D. Feng, "3D Neurological Image Retrieval with Localized Pathology-Centric CMRGlc Patterns," ICIP 2010, pp. 3201-3204, 2010.

[19] S. Batty, J. Clark, T. Fryer, and X. Gao, "Prototype System for Semantic Retrieval of Neurological PET Images," Med. Imag. Info., pp. 179-188, 2008.

[20] S. Liu, W. Cai, L. Wen, S. Eberl, M. J. Fulham, and D. D. Feng, "Generalized Regional Disorder-Sensitive-Weighting Scheme for 3D Neuroimaging Retrieval," EMBC 2011, pp. 7009-7012, 2011.

[21] M. Toews, W. Wells III, D. L. Collins, and T. Arbel, "Feature-based Morphometry: Discovering Group-related Anatomical Patterns," NeuroImage, vol. 49, pp. 2318-2327, 2010.

[22] W. Cai, S. Liu, Y. Song, S. Pujol, et al., "A 3D-Difference of Gaussian based Lesion Detector for Brain PET," ISBI 2014, 2014.

[23] M. Wang, X.-S. Hua, X. Yuan, Y. Song, and L.-R. Dai, "Optimizing Multi-graph Learning: Towards a Unified Video Annotation Scheme," ICMM 2007, pp. 862-871, 2007.

[24] J. C. Bezdek and R. J. Hathaway, "Some Notes on Alternating Optimization," AFSS 2002, pp. 288-300, 2002.

[25] S. Liu, W. Cai, L. Wen, *et al.*, "Multi-Channel Brain Atrophy Pattern Analysis in Neuroimaging Retrieval." ISBI 2013, pp. 206-209, 2013.