CONSECUTIVE INDEPENDENCE AND CORRELATION TRANSFORM FOR MULTIMODAL FUSION: APPLICATION TO EEG AND FMRI DATA

Mohammad A. B. S. Akhonda[1], *Yuri Levin-Schwartz*[2], *Suchita Bhinge*[1], *Vince D. Calhoun*[3,4], *and Tülay Adalı*[1]
[1]Dept. of CSEE, University of Maryland, Baltimore County, MD 21250
[2]Icahn School of Medicine at Mount Sinai, New York, NY 10029
[3]The Mind Research Network and [4]Dept. of ECE, University of New Mexico, Albuquerque, NM 87131

ABSTRACT

Methods based on independent component analysis (ICA) and canonical correlation analysis (CCA) as well as their various extensions have become popular for the fusion of multimodal data as they minimize assumptions about the relationships among multiple datasets. Two important extensions that are widely used, joint ICA (jICA) and parallel ICA (pICA), make a number of simplifying assumptions that might limit their usefulness such as identical mixing matrices for jICA, and the requirement for the same number of components for jICA and pICA. In this paper, we propose a new, flexible hybrid method for fusion based on ICA and CCA, called consecutive independence and correlation transform (C-ICT), which relaxes the main limitations of jICA and pICA. We demonstrate performance advantages of C-ICT both through simulations and application to real medical data collected from schizophrenia patients and healthy controls performing an auditory oddball task (AOD).

Index Terms— Data Fusion, EEG, FMRI, Independent Component Analysis, Canonical Correlation Analysis

1. INTRODUCTION

In fusion studies, the collection of data from different sensors or modalities is becoming increasingly popular, since each modality is expected to provide unique, yet complementary, information about the system of interest, such as the brain [1, 2, 3]. Maximizing the utilization of the joint information available in data from different modalities is the fundamental motivation for performing a joint analysis on multimodal data [4, 5, 6]. Since the relationship among modalities are not well understood, it is important to reduce the assumptions placed on the data. This has led to the use of data-driven blind source separation (BSS) techniques for the analysis of multimodal data, especially those based upon independent component analysis (ICA) [1], canonical correlation analysis (CCA) [7] and its extension multiset CCA (M-CCA) [8, 9]. Though, CCA-based methods are popular in fusion studies and successfully implemented for many applications [10], they only exploit second-order statistics (SOS) to enable fusion, as opposed to ICA that can take all order statistical information into account. Furthermore, ICA provides independent components—sources—that enable easy interpretation, critical in real-world applications. This has been one of the reasons for the popularity of ICA-based techniques for fusion of medical imaging data [11, 12, 13].

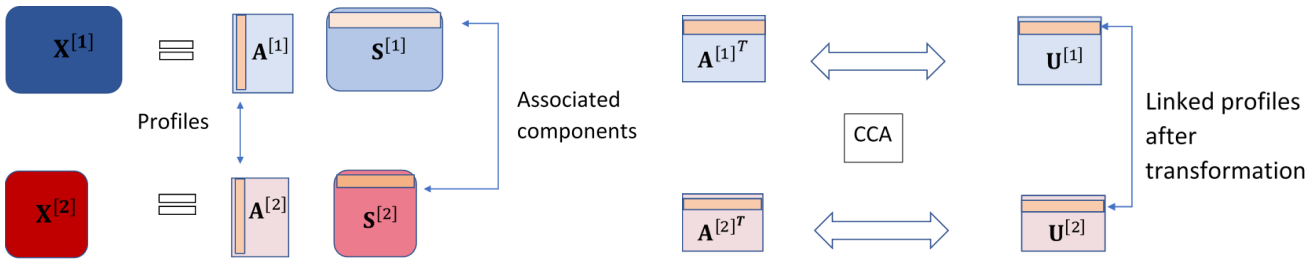
Since ICA formulation is originally for a single dataset, several extensions have been proposed to enable fusion of multiple datasets. By concatenating each dataset together and applying a single ICA on the 'joint' dataset with more samples, joint ICA (jICA) has found wide application in the fusion of medical imaging data such as electroencephalogram (EEG), functional magnetic resonance imaging (fMRI) and structural MRI (sMRI) [1, 14, 15]. However, jICA assumes a common mixing matrix for all the modalities [16], which is a strong constraint. Additionally, jICA assumes that a single probability density function can describe the distributions of the latent sources from multiple modalities at the same time, again another strong assumption that limits its adaptability to many real-world applications. Another important extension of ICA, parallel ICA (pICA) [11, 17], consists of the application of separate ICAs on each dataset, while maximizing correlation of subject profiles, i.e., columns of mixing matrix at each iteration, thus, alleviating the constraint in jICA of a single mixing matrix for all modalities. However, the performance of pICA suffers when the number of subjects is low, due to errors in the correlation estimation. Additionally, this method relies on several user-defined parameters to ensure convergence and optimize performance, decreasing its practicality and potentially causing the final estimated components to not be as independent as in a regular ICA decomposition.

In this study, we propose a new hybrid technique, consecutive independence and correlation transform (C-ICT), which is able to exploit the strengths of both ICA and CCA for joint analysis of multimodal data. The main advantage of this method over others is that it does not impose any restrictions on the mixing matrices like jICA and allows for each modality to have a different number of factors. Different ICA algorithms can be applied on each modality depending on the nature of the data, a flexibility not available either in jICA or pICA. However, to avoid inaccurate estimation of the canonical vectors, which results in overestimation of the sample canonical correlations [18], CCA requires a large number of samples. Hence, we address this issue by using only the informative components in the computation of sample correlations as in [19, 20] enabling the application of C-ICT to scenarios with both low as well as high sample sizes. We evaluate C-ICT with simulated multimodal data as well as with real EEG and fMRI data taken from healthy subjects as well as patients with schizophrenia doing an auditory oddball task (AOD) and demonstrate the advantages it offers over jICA and pICA.

2. BACKGROUND

Direct fusion of data from different modalities is challenging due to the difficulty of obtaining a shared dimension across modalities. Therefore, reducing each modality to a feature, a lower dimensional multivariate representation of data, for each subject is an effective approach [4, 21]. This reduction to feature space enables the examination of the association between the modalities through a com-

This work was supported in part by NSF-CCF 1618551 and NIH-NIBIB R01 EB 005846.



(a) ICA step (b) CCA step

Fig. 1: Model of C-ICT for $K = 2$, (a) ICA on each dataset separately factorizes them into a set of subject profiles and independent associated components, (b) CCA on transpose profiles provides maximally correlated profiles and corresponding combination of components common across modalities.

mon subject dimension, $i.e.$, variations across individuals. Consider $\mathbf{X}^{[k]} \in \mathbb{R}^{M \times T_k}$ $k = 1, 2, \ldots, K$, are K feature datasets constructed from K modalities, where mth row of each feature dataset is formed by extracting one feature from the mth subject and the number of subjects M is common across datasets. Then the noiseless ICA model on kth dataset is,

$$\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]}, \tag{1}$$

where each dataset is a linear mixture of M latent sources, $\mathbf{S}^{[k]} \in \mathbb{R}^{M \times T_k}$ through an invertible mixing matrix, $\mathbf{A}^{[i]} \in \mathbb{R}^{M \times M}$. ICA estimates a demixing matrix $\mathbf{W}_{\text{ica}}^{[k]}$ such that the estimated source components are statistically independent within each dataset and can be computed using $\hat{\mathbf{S}}^{[k]} = \mathbf{W}_{\text{ica}}^{[k]}\mathbf{X}^{[k]}$. Due to the fact that we seek to maximize independence between the components, $\hat{\mathbf{s}}_j^{[k]}$, $j = 1, .., M$, estimation of demixing matrices can be accomplished through the minimization of mutual information, written as

$$I\left(\hat{\mathbf{S}}^{[k]}\right) = \sum_{j=1}^{M} H(\hat{\mathbf{s}_j}^{[k]}) - \log|\det\left(\mathbf{W}_{\text{ica}}^{[k]}\right)| - H(\mathbf{X}^{[k]}) \tag{2}$$

where $H(\cdot)$ is the differential entropy.

Note that the columns of the estimated mixing matrices, $\hat{\mathbf{a}}_j^{[k]}$, $j = 1, .., M$ are referred to as the subject covariations/profiles that establish connection across modalities [7, 16, 21]. However, ICA can only analyze one dataset at a time, limiting its applicability for multimodal fusion. Also an approach that combines the results of separate ICAs does not effectively make use of the joint information and results in suboptimal performance [22]. Thus, several extensions of ICA have been proposed to jointly analyze multiple datasets. Two popular extensions are jICA and pICA.

2.1. JICA and PICA

In order to enable joint analysis of multimodal data, jICA concatenates all the datasets together and applies a single ICA to the "joint" dataset. For the datasets given in (1),

$$\left[\mathbf{X}^{[1]} \ldots . \ \mathbf{X}^{[K]}\right] = \mathbf{A}[\mathbf{S}^{[1]} \ldots . \ \mathbf{S}^{[K]}] = \mathbf{AS}, \tag{3}$$

where, $\mathbf{S} \in \mathbb{R}^{M \times (T_1 + \cdots + T_k)}$ represents the joint source signals. Here, the connection is determined through the common mixing matrix for all datasets, which is a very strong assumption. Additionally, $\mathbf{S}^{[i]}$ contains latent sources from each modality by assuming that a single density function can describe the latent sources for each modality at the same time, which is another strong constrain.

To enhance inter-modality association, pICA maximizes the correlation between the subject profiles at each iteration by adding an extra term to the ICA cost function [11]. In contrast to jICA, pICA performs separate ICAs on each modality simultaneously with a constraint as in

$$\arg\max_{\lambda} I\left(\hat{\mathbf{S}}^{[1]}\right) + I\left(\hat{\mathbf{S}}^{[2]}\right) + \lambda f\left(\hat{\mathbf{A}}^{[1]}, \hat{\mathbf{A}}^{[2]}\right) \tag{4}$$

where the linking term, $f(\cdot)$, introduces dataset correlation through profiles and λ is dynamically adjusted to balance independence and the dataset correlation. Encouraging profile correlation at each iteration step may result in estimated components that are not as independent as in a single ICA decomposition. Additionally, as the algorithm not only looks to maximize independence but also correlation, optimization of pICA depends on many user defined parameters which may limit its practicality.

3. METHODOLOGY

3.1. C-ICT

Both jICA and pICA make assumptions that might limit their performance and practical applicability. This motivates the development of a new method that places very limited assumptions on the datasets, exploits the full strength of ICA and at the same time, maximizes the correlation between subject profiles across modalities. We propose a hybrid model based on ICA and CCA to factor and fuse multimodal data, called C-ICT. The generative model of C-ICT is given in Figure 1. We perform C-ICT in two steps: an ICA step to estimate independent features and associated subject covariations followed by a CCA step in order to exploit the complementary information across datasets.

3.1.1. ICA

Consider two feature datasets, $\mathbf{X}^{[1]} \in \mathbb{R}^{M \times T_1}$ and $\mathbf{X}^{[2]} \in \mathbb{R}^{M \times T_2}$, where each can be separately decomposed into a mixing matrix and associated source components using (1). The mixing matrix and the source component matrix for the kth dataset is denoted as $\mathbf{A}^{[k]} \in \mathbb{R}^{M \times T_k}$ and $\mathbf{S}^{[k]} \in \mathbb{R}^{M \times T_k}$ respectively. ICA estimates a demixing matrix for the kth dataset, $\mathbf{W}_{\text{ica}}^{k}$, and we compute the estimated sources using $\hat{\mathbf{S}}^{[k]} = (\mathbf{W}_{\text{ica}}^{[k]})^T\mathbf{X}^{[k]}$. The columns of the estimated mixing matrices, $\hat{\mathbf{A}}^{[1]}$ and $\hat{\mathbf{A}}^{[2]}$, are the subject covariations that represent the relative weights for the corresponding source estimates for each subject.

3.1.2. CCA

In order to establish a connection across datasets, we perfom CCA on the estimated subject covariations to maximize the correlation between their transformations. CCA finds a pair of weighthing vectors to transform $\hat{\mathbf{a}}^{[1]}$ and $\hat{\mathbf{a}}^{[2]}$ such that the normalized correlation between their transformations, $\mathbf{u}^{[1]} = (\mathbf{w}_{\text{cca}}^{[1]})^T \hat{\mathbf{a}}^{[1]}$, $\mathbf{u}^{[2]} = (\mathbf{w}_{\text{cca}}^{[2]})^T \hat{\mathbf{a}}^{[2]}$, is maximized. Therefore, performing CCA produces combination of profiles whose transformations are maximally correlated across the datasets and therefore, the combination of estimated components that are associated with those profiles. Note that, the use of CCA enables the extraction of different number of components for each modality. This solution can be expanded to multiple datasets by using M-CCA [8]. By calculating the informative components present in the datasets and estimating canonical vectors only for those components, a solution given in [19, 20] for CCA, C-ICT can be applicable to both high and low sample size applications.

3.2. Order and Algorithm Selection

Determining the signal subspace from the observed data is a crucial step, since many problems are overdetermined in nature. Determining the signal subspace by using model order selection significantly improves the performance of BSS algorithm [23]. Performing ICA only on the signal subspace is vital for the performance of overall fusion and provides robustness. In this paper, we use the order selection method described in [23]. Additionally, algorithm used for ICA decomposition also plays an important role in its performance. To our best knowledge, Infomax is the only algorithm that has been used to implement pICA [11, 17]. Though Infomax is a popular choice for ICA on fMRI data, it uses a fixed nonlinearity, which tends to emphasize sources that are highly super-Gaussian. A more robust algorithm based on a flexible density model, ICA-EBM [24] can estimate sources from wide range of distributions leading to better maximization of independence than Infomax [1]. For our work, we use ICA-EBM for both C-ICT and jICA and Infomax for pICA. Because of the iterative nature of both Infomax and ICA-EBM, we make sure that we evaluate the consistency of a both algorithm using multiple runs. We use minimum spanning tree method available in the group ICA for FMRI toolbox (GIFT) [25] to identify a result that represents a consistent estimate over multiple runs to use as the final output.

4. IMPLEMENTATION AND RESULTS

4.1. Data Modalities and Extracted Features

The data used in this study are fMRI and EEG data from 16 schizophrenia patients as well as 22 healthy controls during the performance of an auditory oddball task. The task involved the subjects listening to three different types of auditory stimuli and pressing a button when they hear a target sound [12]. Features are extracted from the fMRI data by computing task related spatial maps for each subject using the general linear model-based regression approach available in the statistical parametric mapping (SPM) toolbox [26]. For the EEG data, we derive event related potentials (ERP) by averaging small windows around the target tone across repeated instances of the task for each subject. For our work, we use ERPs from the Cz channel. We construct a matrix of 38 subjects by 60186 voxels for fMRI and 38 subjects by 451 time points for EEG. For jICA, to balance the contribution from each modality, we follow the procedure given in [1], we repeat the EEG data 100 times resulting 45100 time points.

4.2. Simulation Example

In this section, we present a simulation example of two datasets to compare the relative performances of C-ICT, jICA and pICA. We use Infomax for pICA and ICA-EBM for C-ICT and jICA. Infomax tends to estimate sources that are high super-Gaussian where ICA-EBM is a flexible algorithm. So, to be consistent with both Infomax and ICA-EBM, for each dataset, we generate 10 sources each of 1,000 independent and identically distributed (i.i.d.) samples drawn from a zero-mean Laplacian distribution with a standard deviation of 4. Using the same notation as in Section 2, latent sources of each dataset are linearly mixed using $M \times 10$ mixing matrices. The columns of mixing matrix, $\mathbf{A}^{[i]} \in \mathbb{R}^{M \times 10}$ are all from zero-mean Gaussian distribution. To establish a link between datasets, correlation is introduced between corresponding columns/profiles of the mixing matrices. Correlated profiles of each dataset are of unit standard deviation and uncorrelated ones have 0.5 standard deviation. We use an order 10 for dimensionality reduction resulting in $\mathbf{X}^{[i]} \in \mathbb{R}^{10 \times 1000}$ for C-ICT and pICA, while $\mathbf{X} \in \mathbb{R}^{10 \times 2000}$ for jICA. Results are averaged over 50 independent runs.

We compare the performance of the different methods by either changing the number of subjects or the correlation value introduced between the profiles. In the first case, we make three corresponding profiles from each dataset correlated with a correlation matrix $\mathbf{C} = [0.9\ 0.7\ 0.5]$ and change the number of subjects from 50 to 300 with a step size 50. That means the first columns of $\mathbf{A}^{[1]}$ and $\mathbf{A}^{[2]}$ has a correlation of 0.9, second columns has a correlation of 0.7 and the third ones 0.5. For the second case, in order to probe the robustness of each method with a reduced strength of the connection between the datasets, we reduce the correlation value of one pair of corresponding profiles from 0.9 to 0.4 with a step size of 0.1, while keeping the subject count fixed at 50.

For both cases, we show the average correlation between estimated and true components whose associated profiles are correlated across datasets and the average correlation estimation error between the estimated correlation matrix $\hat{\mathbf{C}}$ to the true correlation matrix $\mathbf{C}$. Note that, to make a valid comparison with jICA, where all the modalities share a single mixing matrix, we look for estimated profiles that has the highest correlation with the original profiles from each dataset and calculate correlation of corresponding estimated and true source components. The results of simulation examples are given in Figure 2.

In the Figure 2, when subject count or correlation introduced is low, C-ICT provides better performance than both pICA and jICA and has a lower correlation estimation error than pICA. But as the subject count or correlation value increases, pICA has lower correlation estimation error than C-ICT. This makes sense as the pICA algorithm maximizes the correlation between profiles when it is high. But when correlation or number of samples is low, pICA incorrectly estimates the correlation and thus biases the results in future iterations, resulting in worse performance. We also see that jICA has the worst performance of the three methods, due to the restrictive common mixing matrix assumption.

4.3. Results and Discussion

Due to the spatial smoothness in fMRI and strong temporal correlation in EEG data, using traditional information theoretic criteria directly as in [27] would yield a highly overestimated order, due to the fact that the independent and identically distributed samples assumption is violated. For order selection of fMRI and EEG data, we use a recent approach based on entropy rate given in [24], which addresses this issue by directly modeling the sample dependence to

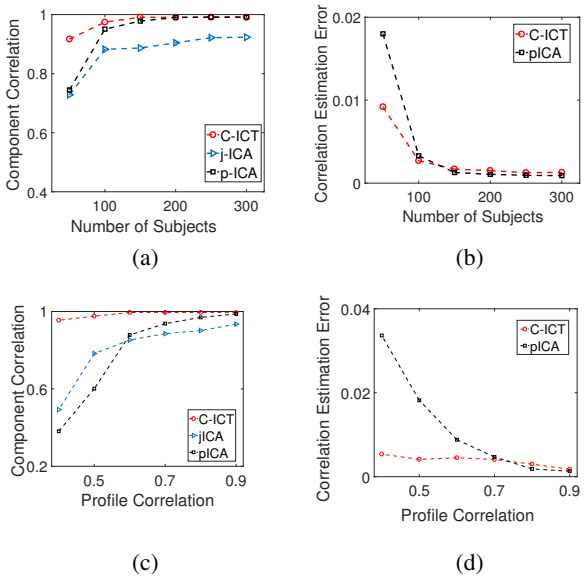
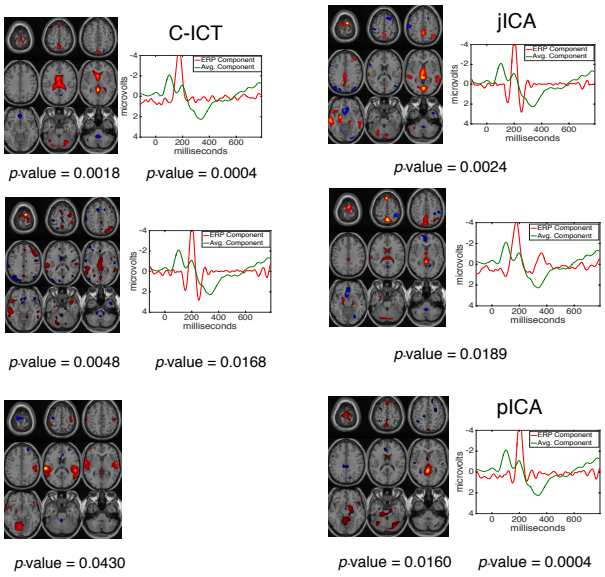


Fig. 2: Estimation performance of common component with two datasets as (a) and (b) with different number of subjects, (c) and (d) with different correlation value. Note that jICA is not present in (b) and (d). because of the common mixing matrix assumption. Here component correlation is the average correlation between the true and estimated components and correlation estimation error is the average mean square error calculated between the original profiles correlation and estimated profiles correlation.



(a) C-ICT (b) jICA and pICA

Fig. 3: Components generated using all three methods. (a) Three fMRI and two ERP components are found significant using C-ICT where the first two pairs of components have correlation between them of 0.64 and 0.47, (b) For jICA two components and for pICA one component from each modality are found significant.

write the likelihood. For C-ICT and pICA, we perform order selection on individual datasets and and chose the order we get for fMRI dataset as we find estimated components are more stable at this order. For jICA, since order selection is performed after concatenating fMRI and EEG data, we perform order selection on the concatenated dataset. We test the stability of the results for the orders given by [24] for all three methods and select $N = 20$ for pICA and C-ICT, $N = 15$ for jICA.

Since the data studied in this paper are from two groups, patients with schizophrenia and healthy controls, a natural objective is to find factors in both datasets that can distinguish between the two groups. This is accomplished by performing a two sample t-test on the subject covariations to identify profiles that show a significant ($p < 0.05$) group difference. The associated components of these profiles are referred to as *biomarkers*. Note that for C-ICT, As the order selected ($N = 20$) is close to the dimensionality of the datasets ($M = 38$), we select the biomarkers after the ICA step and use only those for the CCA step.

For all three methods, we choose only those components whose associated profiles show significant group differences. For the fMRI Z-maps, thresholded at $Z = 2.7$, red, orange, yellow refer to an increase in controls over patients and blue denotes a decrease in controls versus patients. As shown in Figure 3, all three methods capture components that can differentiate between healthy controls and patients. JICA captures two components that are significant, while pICA estimates only one significant component from each modality. For C-ICT, shown in Figure 3 (a), we obtain three fMRI components and two ERP components that are significant. The first two pairs of components have correlation of 0.64 and 0.47. The top ERP component in C-ICT reports N2/P3 complex where the associated fMRI component shows higher activation in healthy controls in the parietal region. The second fMRI component shows activation in the sensorimotor region where the ERP components reports N2. The third fMRI components show activation in auditory region. The fMRI components in jICA report a higher activation in healthy controls in the sensorimotor region and part of the auditory region and associated ERP components are similar to C-ICT. The ERP component in pICA also shares similarities with those found using jICA and C-ICT, since it is also shows temporal activation in the N2/P3 complex. The fMRI component from pICA shows activity mainly in the sensorimotor region. Comapring to pICA and jICA, C-ICT in general picking up more significant biomarkers and also showing the degree of association between the linked components through correlation.

5. CONCLUSION

In this paper, we introduce a novel method to analyze multimodal data, C-ICT and compare it with two existing methods, jICA and pICA, using both simulations as well as real multimodal data obtained from patients with schizophrenia as well as healthy controls. We find that C-ICT, while easing the limiting assumptions imposed by jICA and pICA, also has superior performance to both methods on simulated data. We also find that using real fMRI and ERP data, C-ICT in general can find components that are better able to differentiate between patients with schizophrenia and healthy controls than either pICA or jICA. The success of the C-ICT in this context motivates its use for the fusion in other problems, and also of more than two modalities, using the multiset extension of CCA, MCCA, to identify combinations of components that are associated across multiple datasets.

6. REFERENCES

[1] T. Adalı, Y. Levin-Schwartz, and V. D. Calhoun, "Multi-modal data fusion using source separation: Application to medical imaging," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1494 – 1506, 2015.

[2] B. Mijović, M. D. Vos, K. Vanderperren, B. Machilsen, S. Sunaert, S. V. Huffel, and J. Wagemans, "The dynamics of contour integration: A simultaneous eeg–fmri study," *NeuroImage*, vol. 88, pp. 10–21, 2014.

[3] R. Leeb, H. Sagha, R. Chavarriaga, and J. del. R. Milln, "Multimodal fusion of muscle and brain signals for a hybrid-bci," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug 2010, pp. 4343–4346.

[4] D. Lahat, T. Adalı, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[5] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information Fusion*, vol. 19, pp. 4–19, 2014.

[6] F. Savopol and C. Armenakis, "Merging of heterogeneous data for emergengy mapping: Data integration or data fusion?," in *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*. 2002, vol. 34, pp. 668–674, Presented at 34th Conference for International Society for Photogrammetry and Remote Sensing (ISPRS), Ottawa, Canada.

[7] N.M. Correa, T. Adalı, Yi-Ou Li, and V. D. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, 2010.

[8] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[9] A.A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 293–305, March 2002.

[10] Q.-S. Sun, S.-G. Zeng, P.-A. Heng, and D.-S. Xia, "Feature fusion method based on canonical correlation analysis and handwritten character recognition," in *ICARCV 2004 8th Control, Automation, Robotics and Vision Conference,*, Dec 2004, vol. 2.

[11] J. Liu, G. Pearlson, A. Windemuth, G. Ruano, N. I. Perrone-Bizzozero, and V. D. Calhoun, "Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA," *Human Brain Mapping*, vol. 30, no. 1, pp. 241–255, 2009.

[12] V. D. Calhoun, T. Adalı, G. Pearlson, and K. Kiehl, "Neuronal chronometry of target detection: Fusion of hemodynamic and event related potential data," *NeuroImage*, vol. 30, pp. 544–553, 2006.

[13] A. R Groves, C. F Beckmann, S. M Smith, and M. W Woolrich, "Linked independent component analysis for multimodal data fusion," *NeuroImage*, vol. 54, no. 3, pp. 2198–2217, 2011.

[14] J. Mangalathu-Arumana, S.A. Beardsley, and E. Liebenthal, "Within-subject joint independent component analysis of simultaneous fMRI/ERP in an auditory oddball paradigm," *NeuroImage*, vol. 60, no. 4, pp. 2247 – 2257, 2012.

[15] B. Mijovic, K. Vanderperren, N. Novitskiy, B. Vanrumste, P. Stiers, B. V. Bergh, L. Lagae, S. Sunaert, J. Wagemans, S. V. Huffel, and M. D. Vos, "The "why" and "how" of joint ICA: Results from a visual detection task," *NeuroImage*, vol. 60, no. 2, pp. 1171 – 1185, 2012.

[16] T. Adalı, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1478 – 1493, 2015.

[17] G. D Pearlson, V. D Calhoun, and J. Liu, "An introductory review of parallel independent component analysis (p-ica) and a guide to applying p-ica to genetic data and imaging phenotypes to identify disease-associated biological pathways and systems in common complex disorders," *Frontiers in genetics*, vol. 6, pp. 276, 2015.

[18] A. Pezeshki, L. L. Scharf, M. R. Azimi-Sadjadi, and M. Lundberg, "Empirical canonical correlation analysis in subspaces," in *2004 Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, Nov. 2004, vol. 1, pp. 994–997.

[19] N. Asendorf and R. R. Nadakuditi, "Improved estimation of canonical vectors in canonical correlation analysis," in *49th Asilomar Conference on Signals, Systems and Computers*, Nov 2015, pp. 1806–1810.

[20] Z. Bao, J. Hu, G. Pan, and W. Zhou, "Canonical correlation coefficients of high-dimensional normal vectors: finite rank case," *ArXiv e-prints*, July 2014.

[21] V. D. Calhoun and T. Adalı, "Feature-based fusion of medical imaging data," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 711–720, Sep. 2009.

[22] Y. Levin-Schwartz, V. D. Calhoun, and T. Adal, "Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1385–1395, July 2017.

[23] G.S. Fu, M. Anderson, and T. Adalı, "Likelihood estimators for dependent samples and their application to order detection," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4237–4244, Aug. 2014.

[24] X. L. Li and Adalı, "Independent component analysis by entropy bound minimization," .

[25] GIFT, "Group ICA of fMRI Toolbox (GIFT)," http://mialab.mrn.org/software/gift/index.html, 2011.

[26] SPM5, "Statistical Parametric Mapping," http://www.fil.ion.ucl.ac.uk/spm/software/spm5, 2011.

[27] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech and Signal Process.*, vol. 33, no. 2, pp. 387–392, April 1985.