Improving Multiset Canonical Correlation Analysis in High Dimensional Sample Deficient Settings

Nicholas Asendorf
Department of Electrical and Computer Engineering
University of Michigan
Ann Arbor, Michigan 48105
Email: asendorf@umich.edu

Raj Rao Nadakuditi
Department of Electrical and Computer Engineering
University of Michigan
Ann Arbor, Michigan 48105
Email: rajnrao@umich.edu

***Abstract*—We consider the problem of inferring and learning latent correlations present in multiple noisy matrix-valued datasets using multiset canonical correlation analysis (MCCA). We show that empirical MCCA will provably fail to infer the presence of latent correlations when the sample size is less than a threshold that is completely specified by the dimensionality of the datasets. For the setting where the individual noisy data matrices are structured as low-rank-plus-noise, we propose a simple modification of MCCA, which we label Informative MCCA (IMCCA). We show, on both synthetic and real-world datasets, that IMCCA reliably infers and learns latent correlations.**

I. Introduction

Multi-modal data fusion is a relevant problem for many emerging machine learning applications. In these applications, we have access to multiple datasets, possibly collected using different sensing modalities, each of which describe some feature of the system. Such datasets have been encountered in machine learning [1]–[3], medical imaging [4]–[8], and multi-temporal hyperspectral imaging [9], to name a few. The goal is to infer and learn the latent correlations between datasets; dimensionality reduction algorithms play an important role in this context.

Canonical correlation analysis (CCA) is a classical joint multidimensional dimensionality reduction algorithm for the two dataset setting [10]. CCA learns a linear transformation for each dataset such that the transformed features have maximal correlation. The extension to the multi-dataset setting is referred to as multiset canonical correlation analysis (MCCA) and the underlying theory has evolved significantly over the past few decades [11]–[15]. The five objective functions (notions of multiset correlation) posed by Kettenring and four constraint functions on canonical vectors posed by Nielsen give rise to twenty different formulations of MCCA.

In this paper, we consider the 'MAXVAR' formulation as the 'most natural 'extension of CCA to multiple datasets in a sense that we will return to shortly. We begin by deriving the theoretical solution to MAXVAR assuming known covariance matrices and then consider the empirical setting where the covariance matrices are unknown and have to be estimated from a finite number of samples. Substituting the sample covariance matrices into the oracle MCCA formulation yields the empirical MCCA algorithm. We show that the empirical MCCA solution can be computed directly by considering the SVD of a block-structured matrix where the blocks are composed of the pairwise product of the right singular vectors of the individual data matrices. This leads to the formulation of the informative MCCA (IMCCA) algorithm, which forms the block-structured matrix using a smaller subset of these singular vectors corresponding to the 'informative components' associated with the latent low rank matrices. We then show that the IMCCA algorithm can reliably infer and learn latent correlations in settings where the empirical MCCA algorithm provably cannot.

II. Multiset Canonical Correlation Analysis

Let $y_1 \in \mathbb{C}^{d_1\times 1}, \ldots, y_m \in \mathbb{C}^{d_m\times 1}$ model observations from each of the m datatsets. Define the covariance matrix between all datasets as

$$R \quad = \mathbb{E}\left[\left[y_i y_j^H\right]_{i,j=1}^m\right] = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ R_{m1} & R_{m2} & \cdots & R_{mm} \end{bmatrix}$$

and the block diagonal matrix $R_D = \textbf{blkdiag}(R_{11}, R_{22}, \ldots, R_{mm})$. Both R and R_D are of size $d \times d$ where $d = \sum_{j=1}^m d_j$.

The goal of MCCA is to learn canonical coefficient vectors, $x_j \in \mathbb{C}^{d_j\times 1}$ for $j = 1, \ldots, m$, such that the canonical variates, $w_j = x_j^H y_j$, are optimal with respect to an objective function $J(\cdot)$ and constraint function $h(\cdot)$. Define the vector of canonical vectors as $x = \left[x_1^H, \ldots, x_m^H\right]^H \in \mathbb{C}^{d\times 1}$ and the vector of canonical variates as $w = [w_1, \ldots, w_m]^H \in \mathbb{C}^{m\times 1}$. The covariance matrix of w is

$$\Phi(x) = \mathbb{E}\left[ww^H\right] = \begin{bmatrix} x_1^H R_{11} x_1 & \ldots & x_1^H R_{1m} x_m \\ \vdots & \ddots & \vdots \\ x_m^H R_{m1} x_1 & \ldots & x_m^H R_{mm} x_m \end{bmatrix}.$$

Using this notation, the MCCA optimization problem is

$$\begin{aligned} \underset{x}{\text{optimize}} \quad & J(\Phi(x)) \\ \text{subject to} \quad & h(x, R). \end{aligned} \tag{1}$$

In [14], Kettenring summarizes five commonly used objective functions and Nielson summarizes four commonly used constraint functions in [15]. In this paper we consider the maximum variance (MAXVAR) objective function, which seeks to maximize the largest eigenvalue of $\Phi(x)$, and the constraint functions $x_j^H R_{jj} x_j = 1$, $j = 1, \ldots, m$, which require the canonical variates to be unit norm. Therefore, the MCCA optimization problem considered herein is

$$\begin{aligned} &\underset{x}{\operatorname{argmax}} && \lambda_1\left(\Phi(x)\right) \\ &\text{subject to} && x_j^H R_{jj} x_j = 1,\, j = 1, \ldots, m, \end{aligned} \tag{2}$$

where $\lambda_1(M)$ is the largest eigenvalue of a matrix M.

A. Solution to MAXVAR

We rewrite the MAXVAR optimization problem in (2) as

$$\begin{aligned} &\underset{x}{\operatorname{argmax}} && \lambda \\ &\text{subject to} && x_i^H R_{ii} x_i = 1 \;, 1 \leq i \leq m, \\ & && \Phi(x) a = \lambda a, \quad a^H a = 1. \end{aligned} \tag{3}$$

Writing $\Phi(x) = X^H R X$, where $X = \mathbf{blkdiag}(x_1, \ldots, x_m)$, and making the transformation $\widetilde{a} = R_D^{1/2} X a$, (3) becomes

$$\begin{aligned} &\underset{\widetilde{a}}{\operatorname{argmax}} && \lambda \\ &\text{subject to} && \widetilde{a}^H R_D^{-1/2} R R_D^{-1/2} \widetilde{a} = \lambda, \quad \widetilde{a}^H \widetilde{a} = 1. \end{aligned} \tag{4}$$

The solution to this formulation is immediately solved by examining the matrix

$$C_{\text{mcca}} = R_D^{-1/2} R R_D^{-1/2}. \tag{5}$$

The largest eigenvalue of C_{mcca} is exactly λ in (4) and $\widetilde{a}$ is exactly the corresponding eigenvector. To recover the canonical vectors, we simply invert our original transformation to obtain

$$x_j = \frac{R_{jj}^{-1/2} \widetilde{a}_j}{\|\widetilde{a}_j\|}, \tag{6}$$

where $\widetilde{a}_j \in \mathbb{C}^{d_i}$ is the component of $\widetilde{a}$ corresponding to dataset j. We may obtain higher order canonical vector and canonical correlation pairs by taking successive eigenvalue-eigenvector pairs of C_{mcca}. We denote the ith order canonical vector with the superscript $x_j^{(i)}$ corresponding to λ_i.

Examining (5), we see that if all of the datasets are uncorrelated ($R_{ij} = 0$), $C_{\text{mcca}} = I_d$ and all eigenvalues of C_{mcca} are equal to 1. This motivates the use of the statistic

$$\lambda_i - 1, \tag{7}$$

to infer the presence of latent correlations in multiple datasets.

B. Empirical MAXVAR

In a practical setting, the covariance matrices used to form R and R_D needed for C_{mcca} in (5) are unknown and hence have to be learned from training data. We stack the n observations from each dataset to form the m training data matrices

$$Y_j = \left[y_j^{(1)}, \ldots y_j^{(n)}\right] \in \mathbb{C}^{d_j \times n}, j = 1, \ldots, m. \tag{8}$$

Substituting the sample covariance matrices (which are the maximum-likelihood estimates under a Gaussian prior), $\widehat{R}_{ij} = \frac{1}{n} Y_i Y_j^H$, into (5) yields the plug-in estimate $\widehat{R}_{\text{mcca}}$. Henceforth, we shall label canonical vectors obtained from an analysis of the eigen-structure of $\widehat{R}_{\text{mcca}}$ as empirical MAXVAR or empircal MCCA.

To explore the structure of $\widehat{R}_{\text{mcca}}$, denote the SVD of each training dataset as $Y_j = \widehat{U}_j \widehat{\Sigma}_j \widehat{V}_j^H$ and form the block matrices $\widehat{U} = \mathbf{blkdiag}(\widehat{U}_1, \ldots, \widehat{U}_m) \in \mathbb{C}^{d \times d}$, $\widehat{\Sigma} = \mathbf{blkdiag}(\widehat{\Sigma}_1, \ldots, \widehat{\Sigma}_m) \in \mathbb{C}^{d \times nm}$, and $\widehat{V} = [\widehat{V}_1, \ldots, \widehat{V}_m] \in \mathbb{C}^{n \times nm}$. With these definitions, we form $\widehat{R} = \widehat{U}\widehat{\Sigma}\widehat{V}^H \widehat{V} \widehat{\Sigma}^H \widehat{U}^H$ and $\widehat{R}_D = \widehat{U}\widehat{\Sigma}\widehat{\Sigma}^H \widehat{U}^H$. After some straightforward algebraic manipulations, we note that

$$\widehat{R}_{\text{mcca}} = \widehat{R}_D^{-1/2} \widehat{R} \widehat{R}_D^{-1/2} = \widetilde{U} \widetilde{V}^H \widetilde{V} \widetilde{U}^H, \tag{9}$$

where

$$\begin{aligned} \widetilde{U} &= \mathbf{blkdiag}(\widehat{U}_1(:, 1 : \min(n, d_1)), \ldots, \widehat{U}_m(:, 1 : \min(n, d_m))) \\ \widetilde{V} &= [\widehat{V}_1(:, 1 : \min(n, d_1)), \ldots, \widehat{V}_m(:, 1 : \min(n, d_m))]. \end{aligned}$$

From this, we see that the eigenvalues of $\widehat{R}_{\text{mcca}}$ are exactly the eigenvalues of $\widetilde{V}^H \widetilde{V}$. Let $\widehat{\lambda}_i$ be the eigenvalues of $\widehat{R}_{\text{mcca}}$. Thus the statistic

$$\widehat{\lambda}_i - 1, \tag{10}$$

might be used to infer the presence of latent correlations. Note that (10) is the plug-in estimate of the statistic in (7).

III. New Theoretical Result

A key result in two-dataset empirical CCA is that when the number of samples is less than the combined dimension of the datasets ($n < d_1 + d_2$) then the largest eigenvalue of $\widehat{R}_{\text{mcca}}$ is deterministically equal to 2 [16], regardless of whether an underlying correlation actually exists between the datasets. We now provide an analogous statement for the MAXVAR formulation of MCCA for the setting where $m > 2$.

Theorem III.1. *Let $m > 2$ be the number of datasets to use in MCCA. If $2n < \min_{i \neq j \neq k}(d_i + d_j + d_k)$ then the largest eigenvalue of $\widehat{R}_{mcca}$ is equal to m.*

In high-dimensional settings, because of computational complexity considerations, it is often prudent to learn correlations only when they are detected to be statistically significant. Theorem III.1 establishes conditions when the empirical MAXVAR test statistic in (10) will provably fail to infer the presence of correlations despite the presence of correlations.

IV. New Algorithm: Informative MCCA

Empirical MCCA places no model on the observations and so we simply stack the observations in data matrices as in (8). We now consider a setting where the observations of each dataset have latent low-rank plus noise structure. We denote the ith observation from the jth dataset as $y_j^{(i)}$ as in (8) and model them as

$$y_j^{(i)} = U_j s_j^{(i)} + z_j^{(i)}, \tag{11}$$

where $U_j^H U_j = I_{k_j}$ model the low-rank signal subspace of each dataset and $z_j^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_{d_j})$ model the noise. Furthermore, assume that signals are modeled as

$$s_j^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, \Theta_j)$$

where $\Theta_j = \mathbf{diag}\left(\left(\theta_j^{(1)}\right)^2, \ldots, \left(\theta_j^{(k_j)}\right)^2\right)$. Assume that for all j and i, $z_j^{(i)}$ are mutually independent and independent from all $s_j^{(i)}$. Finally, assume that the signal covariance is modeled as

$$\mathbb{E}\left[s_j s_\ell^H\right] =: K_{j\ell} = \Theta_j^{1/2} P_{j\ell} \Theta_\ell^{1/2}$$

where the entries of $P_{j\ell}$ are between -1 and 1 and represent the correlation between the entries of s_j and s_ℓ. Under this model, the covariance matrices are

$$\begin{aligned} \mathbb{E}\left[y_j y_j^H\right] &= U_j \Theta_j U_j^H + I_{d_j} =: R_{jj} \\ \mathbb{E}\left[y_j y_\ell^H\right] &= U_j K_{j\ell} U_\ell^H =: R_{j\ell}. \end{aligned} \tag{12}$$

A. IMCCA

Empirical MCCA examines the eigen-decomposition of the matrix $\widehat{R}_{\text{mcca}} = \widetilde{U}\widetilde{V}^H\widetilde{V}\widetilde{U}^H$. This matrix uses the entire spectrum of all the datasets. Under the low-rank signal-plus-noise model, this is incorrect as each dataset is low-rank.

Motivated by this observation from the low-rank data model in (11), we expect the principal components (or top few singular vectors) of the SVD of these matrices to be 'informative' (*i.e.*, correlated with the latent vectors in (11). Thus our idea is to form the matrix $\widehat{R}_{\text{mcca}}$ from just the principal 'informative' singular vectors instead of all the singular vectors. This leads to our so-called informative version of MAXVAR.

Let $\widehat{k}_j$ be estimates of the number of the latent rank of the low-rank matrix in (11) or equivalently, of the number of principal components used from each dataset. We then form the matrices

$$\mathring{U}_j = \widehat{U}_j\left(:, 1:\widehat{k}_j\right), \mathring{V}_j = \widehat{V}_j\left(:, 1:\widehat{k}_j\right).$$

Let

$$\mathring{U} = \mathbf{blkdiag}(\mathring{U}_1, \ldots, \mathring{U}_m), \quad \mathring{V} = \left[\mathring{V}_1, \ldots \mathring{V}_m\right],$$

and define the informative MAXVAR (IMCCA) matrix as

$$\widehat{R}_{\text{imcca}} = \mathring{U}\mathring{V}^H\mathring{V}\mathring{U}^H. \tag{13}$$

Denote the eigenvalues of this IMCCA matrix by $\widetilde{\lambda}_i$. As in, (7) we propose using

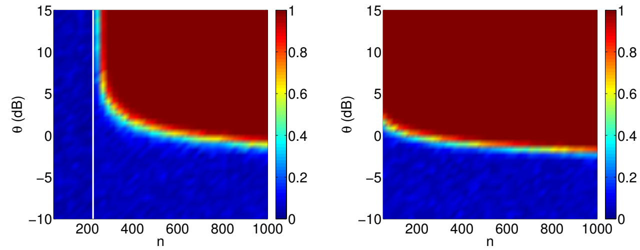
$$\widetilde{\lambda}_i - 1, \tag{14}$$

as a test statistic to infer the presence of latent correlations. The IMCCA canonical vector estimate is found by substituting the eigenvectors of $\widehat{R}_{\text{imcca}}$ in (6) along with the covariance matrix estimates found in (12). We note that the statistic in (14) is computationally much more tractable to compute than (10) when $\max_j \widehat{k}_j \ll d_j$.

V. Results

A. Numerical Simulations

To validate Theorem III.1 and to compare the ability of the empirical MCCA and IMCCA algorithms to infer the presence of correlations, we consider a rank-1 setting. We set $d_1 = d_2 = d_3 = 150$, $k_1 = k_2 = k_3 = 1$, and $P_{12} = P_{13} = P_{23} = 0.9$. We then sweep over n and $\theta = \theta_1 = \theta_2 = \theta_3$. For each value of n and θ, we generate data matrices from data modeled in (11), form $\widehat{R}_{\text{mcca}}$ in (9) or $\widehat{R}_{\text{imcca}}$ in (13), and compute its largest eigenvalue. We then generate $\widehat{R}_{\text{mcca}}$ (or $\widehat{R}_{\text{imcca}}$) from purely noise only matrices where the noise is modeled exactly as in (11). We repeat this for 500 trials to obtain 500 eigenvalues from signal bearing $\widehat{R}_{\text{mcca}}$ (or $\widehat{R}_{\text{imcca}}$) and 500 eigenvalues from noise only $\widehat{R}_{\text{mcca}}$ (or $\widehat{R}_{\text{imcca}}$). Figure 1 plots the Kolmorgorv-Smirnov (KS) statistic (significance value of 0.05) used to test if the distributions between the signal-bearing and noise-only eigenvalues are statistically different. A value of 1 represents the distributions are statistically different while a value of 0 represents the distributions are identical. For IMCCA, we set $\widehat{k}_1 = \widehat{k}_2 = \widehat{k}_3 = 1$.



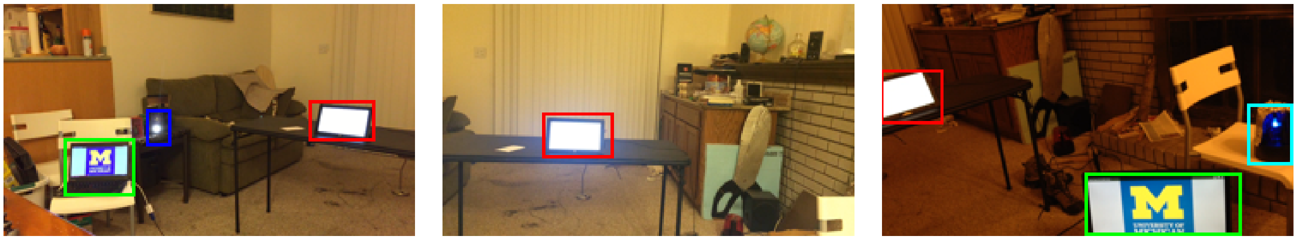
(a) Empirical MCCA (b) IMCCA

Fig. 1. Heatmaps for detection of a correlated signal for a rank-1 setting of $d_1 = d_2 = d_3 = 150$, $k_1 = k_2 = k_3 = 1$, and $P_{12} = P_{13} = P_{23} = 0.9$. We generate data from (11) sweeping over n and $\theta = \theta_1 = \theta_2 = \theta_3$. We take the largest eigenvalue of $\widehat{R}_{\text{mcca}}$ in (9) and $\widehat{R}_{\text{imcca}}$ (13) for both the signal and noise only cases and repeat for 500 trials. We plot the Kolmorgov-Smirnov test statistic between the eigenvalue distributions of the signal and noise only cases. (a) Results for empirical MCCA. The white line is minimum number of samples needed for non-deterministic eigenvalues as predicted by Theorem III.1. When we have fewer samples than this limit, the empirical MCCA test statistic cannot infer the presence of the correlated signal. (b) Results for IMCCA. The IMCCA test statistic is able to infer the presence of the correlated signal in the small-sample regime.

Ideally, we wish the distributions of the top eigenvalues of signal bearing and noise-only matrices to be different in all parameter regimes. However, as evident in Figure 1(a), in the small-sample, high-dimensional regime, the distributions of the largest eigenvalue used by empirical MCCA are identical whether the datasets have latent correlations or not. This is undesirable and verifies Theorem III.1. However, IMCCA is able to overcome this undesirable performance to reliably infer the presence of correlations in multiple datasets in the low-sample, high-dimensionality regime.

B. Real World Video Example

To verify the effectiveness of IMCCA for real world applications and to showcase the extreme sub-optimality of empirical MCCA, we setup a controlled experiment consisting of four stationary flashing lights and three stationary iPhone cameras. Figure 2 manually identifies each source in each camera view by drawing a colored box around it. Based on our setup, all cameras share a flashing tablet, while the left and right views share a flashing laptop. The left and right views also each have an independent source in their view.



(a) Left Camera (b) Middle Camera (c) Right Camera

Fig. 2. Manual identification of each source in each camera. All three sources share a common flashing tablet, outlined in red. The left and right camera views share a common flashing laptop screen, outlined in green. The left camera has an independent flashing phone light, outlined in dark blue. The right camera has an independent flashing police light, outlined in cyan.

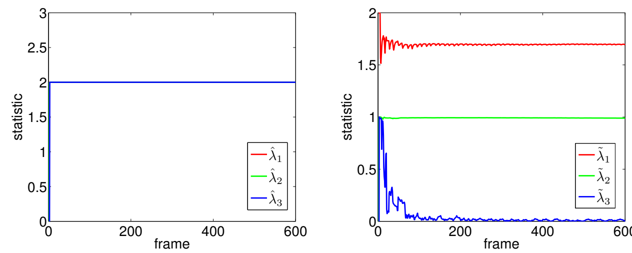
To synchronize the cameras we used the RecoLive MultiCam iPhone app [1]. After turning on all light sources, we recorded 20 seconds of video at 30 frames per second, which we converted to grayscale and downsampled to a resolution of 240×135. We then vectorized each image and stacked the 600 frames into data matrices, all of dimension 32400×600.

We run empirical MCCA and IMCCA after every frame. Specifically, for frame ℓ, we construct the $32400 \times \ell$ submatrices Y^{ℓ}_{left}, Y^{ℓ}_{middle}, and Y^{ℓ}_{right} by taking the matrix of the first ℓ vectorized frames in each view and then subtracting the mean of the resulting submatrix. We then use these resulting submatrices as inputs to empirical MCCA and IMCCA in equations (9) and (13), respectively. Using our knowledge of the number of sources in each camera, we set $\widehat{k}_{\text{left}} = 3$, $\widehat{k}_{\text{middle}} = 1$, and $\widehat{k}_{\text{right}} = 3$. Figure 3 plots the top 3 correlation statistics returned by empirical MCCA and IMCCA, as defined in (10) and (14), respectively. As expected due to the extreme sample deficient regime, the empirical MCCA statistics are equal to $2 = m - 1$, which incorrectly identifies the top three canonical vectors as being perfectly correlated. However, the IMCCA statistic correctly identifies two correlations that exist. The third correlation statistic returned by IMCCA is essentially zero. Figure 4 plots the canonical vectors associated with the top two correlation statistics for empirical MCCA and IMCCA as computed by (6). The learned MCCA canonical vectors appear extremely random and noisy while the learned IMCCA canonical vectors correctly identify the two sources of correlation in our video.

VI. Conclusion

We considered the problem of inferring, identifying and learning latent correlations in more than two datasets. We provided sufficient conditions for the deterministic failure of the empirical MCCA algorithm to infer latent correlations and showed that these will generically occur in the low-sample, high dimensional regime. We then considered the setting where the individual datasets may be modeled as low-rank-signal-plus-noise matrices and developed an algorithm (IMCCA) that functions well in the low-sample, high dimensional regime. The IMCCA algorithm uses the fact that the principal singular vectors are more informative than the other singular vectors. We demonstrated its improved performance relative to empirical MCCA on a synthetic and real world video dataset.

[1] http://recolive.com/en/



(a) MCCA (b) IMCCA

Fig. 3. Top 3 correlation statistics returned over the 600 frames of video. (a) Empirical MCCA statistic as defined in (10). (b) IMCCA statistic as defined in (14).

Acknowledgment

This work was supported by ONR Young Investigator Award N000141110660, ONR Award N00014-15-1-2141, AFOSR Young Investigator Award FA9550-12-1-0266, NSF award CCF-1116115, ARfO MURI grant W911NF-11-1-0391, and ARO MURI grant W911NF-15-1-0479.

References

[1] D. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

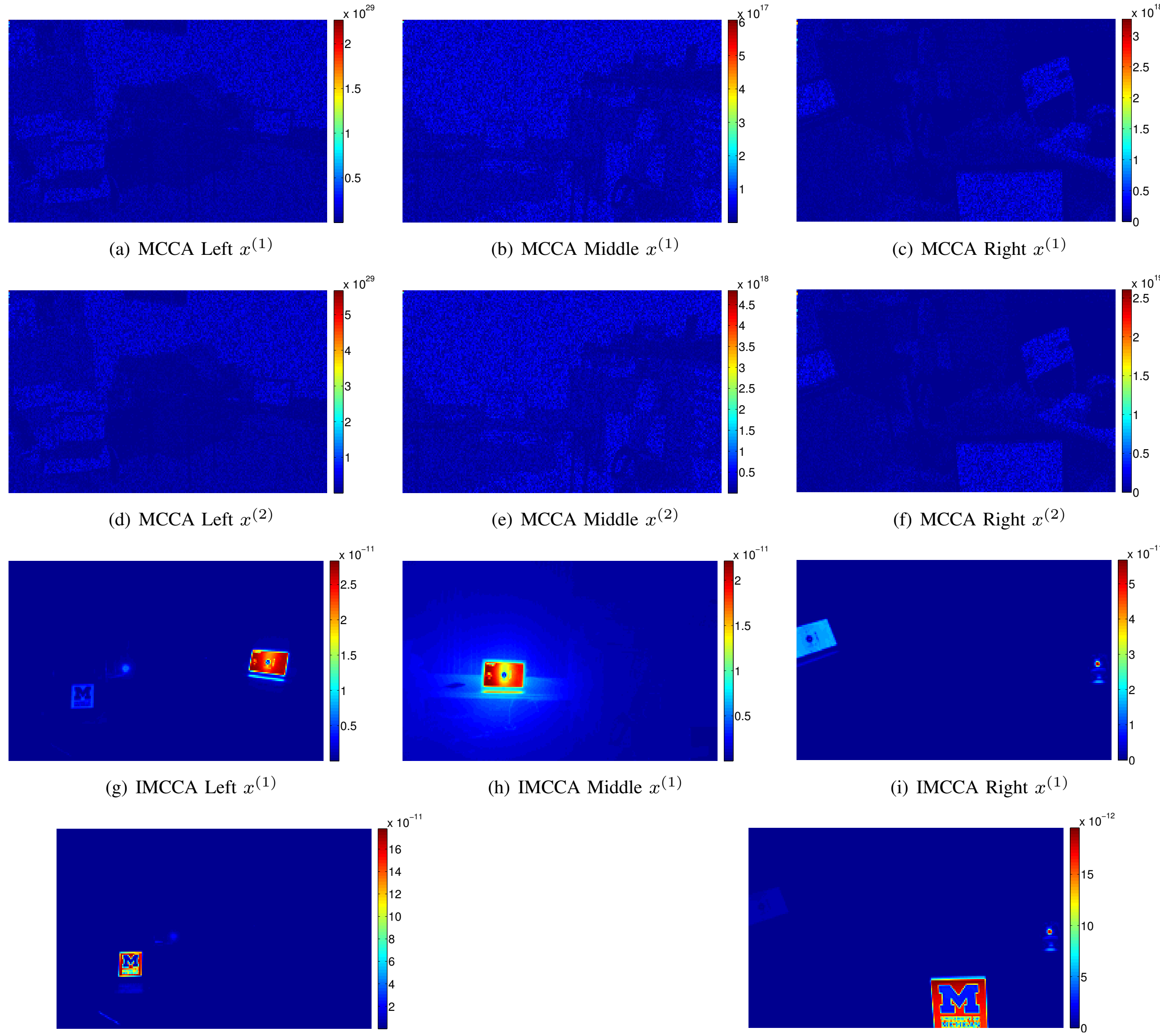
[2] P. Dhillon, D. P. Foster, and L. H. Ungar, "Multi-view learning of word embeddings via cca," in *Advances in Neural Information Processing Systems*, 2011, pp. 199–207.

[3] K. Chaudhuri, S. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 129–136.

[4] N. Correa, T. Adali, Y. Li, and V. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *Signal Processing Magazine, IEEE*, vol. 27, no. 4, pp. 39–50, 2010.

[5] F. Deleus and M. Van Hulle, "Functional connectivity analysis of fmri data based on regularized multiset canonical correlation analysis," *Journal of Neuroscience Methods*, 2011.

[6] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in ssvep-based bci using multiset canonical correlation analysis," *International journal of neural systems*, vol. 24, no. 04, 2014.



(a) MCCA Left $x^{(1)}$ (b) MCCA Middle $x^{(1)}$ (c) MCCA Right $x^{(1)}$

(d) MCCA Left $x^{(2)}$ (e) MCCA Middle $x^{(2)}$ (f) MCCA Right $x^{(2)}$

(g) IMCCA Left $x^{(1)}$ (h) IMCCA Middle $x^{(1)}$ (i) IMCCA Right $x^{(1)}$

(j) IMCCA Left $x^{(2)}$ (k) IMCCA Right $x^{(2)}$

Fig. 4. Top 2 canonical vectors for the video experiment as defined in (6). (a)-(c) First empirical MCCA canonical vectors. (d)-(f) Second empirical MCCA canonical vectors. (g)-(i) First IMCCA canonical vectors. (j)-(k) Second IMCCA canonical vectors. There is no second middle camera IMCCA canonical vector as $\widehat{k}_{\text{middle}} = 1$. In the sample deficient regime, empirical MCCA returns random pixels as the canonical vectors are random. IMCCA correctly identifies the shared light in all three views with the first canonical vector and the shared laptop light in the left and right views with second canonical vector.

[7] P. Guccione, L. Mascolo, G. Nico, P. Taurisano, G. Blasi, L. Fazio, and A. Bertolino, "Functional brain networks and schizophrenia analysis with fmri by multiset canonical correlation analysis," in *Advances in Biomedical Engineering (ICABME), 2013 2nd International Conference on*. IEEE, 2013, pp. 207–210.

[8] X. Chen, C. He, and H. Peng, "Removal of muscle artifacts from single-channel eeg based on ensemble empirical mode decomposition and multiset canonical correlation analysis," *Journal of Applied Mathematics*, vol. 2014, 2014.

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

[10] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[11] B. Vinograde, "Canonical positive definite matrices under internal linear transformations," *Proceedings of the American Mathematical Society*, vol. 1, no. 2, pp. 159–161, 1950.

[12] R. G. Steel, "Minimum generalized variance for a set of linear functions," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 456–460, 1951.

[13] P. Horst, "Generalized canonical correlations and their applications to experimental data," *Journal of Clinical Psychology*, vol. 17, no. 4, pp. 331–347, 1961.

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

[15] A. Nielsen, "Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data," *Science*, vol. 21, no. 4, pp. 555–567, 1994.

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