Informative Data Fusion:
Beyond Canonical Correlation Analysis

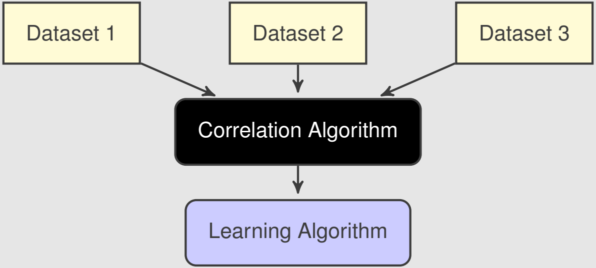
Nicholas Asendorf

asendorf@umich.edu

Department of Electrical Engineering and Computer Science
University of Michigan

Dissertation Defense

May 4, 2015



Thesis Goal

Develop theoretically justified, robust correlation algorithms for multi-dataset fusion

Multiple Datasets

- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.

A Myriad of Applications

Multiple Datasets

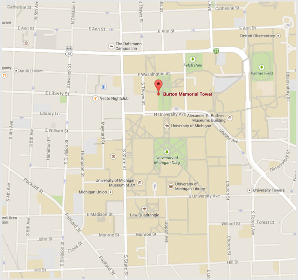
- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.



A Myriad of Applications

Multiple Datasets

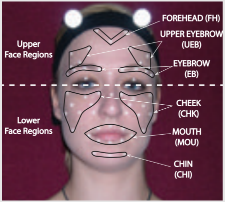
- Audio-Video
- Audio-Audio

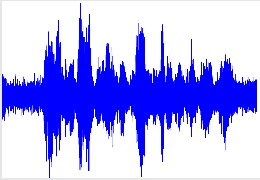
Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.





Multiple Datasets

- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.

A Myriad of Applications

Multiple Datasets

- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.

- disco influences
- danceable grooves
- repetitive melodic phrasing
- extensive vamping
- minor key tonality

Canonical Correlation Analysis

What is it?

- Dimensionality reduction algorithm for exactly 2 datasets, X, Y
- Correlation coefficients, linear transformations

What is it not?

- Data fusion algorithm

Covariance matrices

- $R_{xx} = \mathbb{E}\left[xx^H\right]$
- $R_{yy} = \mathbb{E}\left[yy^H\right]$
- $R_{xy} = \mathbb{E}\left[xy^H\right]$

Optimization problem

$$\begin{aligned} \underset{w_x, w_y}{\operatorname{argmax}} \quad & \rho = w_x^H R_{xy} w_y \\ \text{subject to} \quad & w_x^H R_{xx} w_x = 1 \\ & w_y^H R_{yy} w_y = 1 \end{aligned}$$

Variable Transformation

- $f = R_{xx}^{1/2} w_x$
- $g = R_{yy}^{1/2} w_y$

What is it?

- Dimensionality reduction algorithm for exactly 2 datasets
- Correlation coefficients, linear transformations

What is it not?

- Data fusion algorithm

Optimization problem

$$\underset{f,g}{\operatorname{argmax}} \quad \rho = f^H \underbrace{R_{xx}^{-1/2} R_{xy} R_{yy}^{-1/2}}_{C_{\text{cca}}} g$$

$$\text{subject to} \quad \|f\|_2 = 1\,,\ \|g\|_2 = 1$$

Canonical Vectors

- $w_x = R_{xx}^{-1/2} f$
- $w_y = R_{yy}^{-1/2} g$

Insight

$$\#\ \text{correlated signals} = k = \operatorname{rank}(C_{\text{cca}})$$

Empirical CCA

Training Datasets

- $X = [x_1, \ldots, x_n]$
- $Y = [y_1, \ldots, y_n]$

Sample Covariance Matrices

- $\widehat{R}_{xx} = \frac{1}{n} XX^H$
- $\widehat{R}_{yy} = \frac{1}{n} YY^H$
- $\widehat{R}_{xy} = \frac{1}{n} XY^H$

Estimate

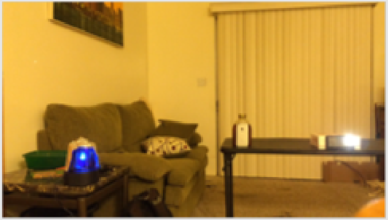
$$\boxed{\begin{aligned} \widehat{C}_{\text{cca}} &= \widehat{R}_{xx}^{-1/2} \widehat{R}_{xy} \widehat{R}_{yy}^{-1/2} \\ &= \widehat{F} \widehat{K} \widehat{G}^H \end{aligned}}$$

Data SVDs

- $X = \widehat{U}_x \widehat{\Sigma}_x \widehat{V}_x^H$
- $Y = \widehat{U}_y \widehat{\Sigma}_y \widehat{V}_y^H$
- $\sigma \left(\widehat{C}_{\text{cca}} \right) = \sigma \left(\widehat{V}_x^H \widehat{V}_y \right)$

Motivational Example - Flashing Light Video

Left Camera



Right Camera

System parameters

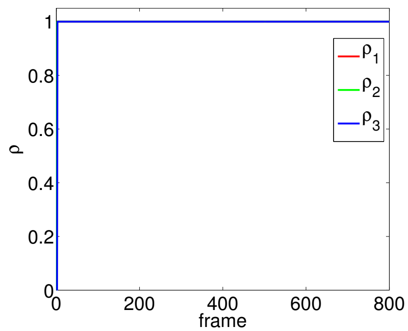
- Vectorize 135×240 image $\Rightarrow p = q = 32400$ pixels
- 30 fps @ 30 seconds $\Rightarrow n = 900$ frames

Goal

Identify correlated pixels between camera views

Empirical CCA Results - Canonical Correlations

Top Canonical Correlations

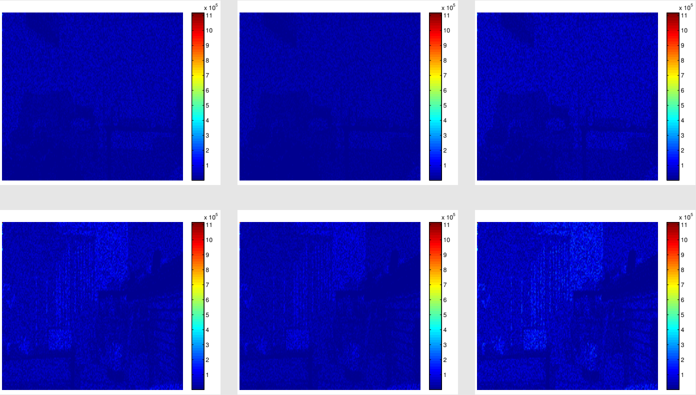


* After 900 frames = 30 seconds of video

First Second Third

Left

Right



Previous Observations

"Therefore, the empirical canonical correlations are defective and may not be used as estimates of canonical correlations between random variables."

- A. Pezeshki, L.L. Scharf et al.
Asilomar Conference on Signals Systems and Computer, 2004

"In conclusion, CCA provide(s) reliable information about spatial correlations existing among pairs of data sets only when SNRs ... are reasonably high, and the sample support is significantly larger than the data dimensions."

- H. Ge et al.
ICASSP, 2009

Correlation Algorithm Wish List

1. **Reliable in the sample deficient regime**
   - meaningful correlations
   - meaningful canonical vectors
2. **Statistical test for correlations**
   - consistency analysis
3. **Robust**
   - non-gaussian data
   - missing data
4. **Extends to more than 2 datasets**

Thesis Outline

1. Introduction
2. Performance of Matched Subspace Detectors Using Finite Training Data
3. Extensions of Deterministic MSD to Missing Data and Useful Subspace Components
4. Using CCA and ICCA to Detect Correlations in Low-Rank Signal-Plus-Noise Datasets
5. On Estimating Population Canonical Vectors
6. The Top Singular Values of XY^H
7. The Largest Singular Values of a Random Projection of a Low-Rank Perturbation of a Random Matrix
8. CCA and ICCA for Regression and Detection
9. Content Based Image Retrieval and Automatic Image Annotation Using Correlations Methods
10. Multiset CCA (MCCA)

Two-Dataset Model

Linear Subspace Model

$$x_i = U_x s_{x,i} + z_{x,i}$$
$$y_i = U_y s_{y,i} + z_{y,i}$$

Parameters

- $U_x^H U_x = I_{k_x}$, $U_y^H U_y = I_{k_y}$
- $z_{x,i} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_p)$, $\;z_{y,i} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_q)$
- $\mathbb{E}\left[\begin{bmatrix} s_{x,i} \\ s_{y,i} \end{bmatrix}\begin{bmatrix} s_{x,i}^H & s_{y,i}^H \end{bmatrix}\right] = \begin{bmatrix} \Theta_x & K_{xy} \\ K_{xy}^H & \Theta_y \end{bmatrix}$
- $K_{xy} = \Theta_x^{1/2} P_{xy} \Theta_y^{1/2}$
- $\Theta_x = \textbf{diag}\left(\left(\theta_1^{(x)}\right)^2, \ldots, \left(\theta_{k_x}^{(x)}\right)^2\right)$, $\;\Theta_y = \textbf{diag}\left(\left(\theta_1^{(y)}\right)^2, \ldots, \left(\theta_{k_y}^{(y)}\right)^2\right)$
- P_{xy} contains correlations ρ_{kj} between signals of x_i and y_i
- $\widetilde{K}_{xy} = \left(\Theta_x + I_{k_x}\right)^{-1/2} K_{xy} \left(\Theta_y + I_{k_y}\right)^{-1/2}$, with singular values $\kappa_1, \ldots, \kappa_{\min(k_x, k_y)}$

Statistical Test for Empirical CCA Correlations

Canonical Correlations

- $\widehat{\rho}_{\text{cca}}^{(i)}$ are singular values of $\widehat{C}_{\text{cca}} = \widehat{R}_{xx}^{-1/2}\widehat{R}_{xy}\widehat{R}_{yy}^{-1/2}$

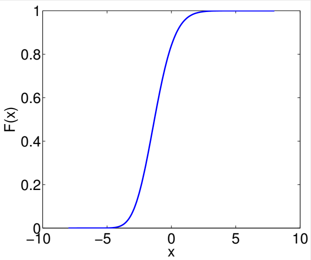
Estimate of # of Correlated Signals

$$\widehat{k}_{\text{cca}} = \sum_{i=1}^{\min(p,q)} \mathbb{1}_{\left\{\left(\widehat{\rho}_{\text{cca}}^{(i)}\right)^2 > \tau_{\text{cca}}^{\alpha}\right\}}$$

Setting the threshold

- F_{cca} is the cdf of largest singular values of $\widehat{C}_{\text{cca}}$ when X and Y are uncorrelated
- $\tau_{\text{cca}}^{\alpha} = F_{\text{cca}}^{-1}(1-\alpha)$
- $\tau_{\text{cca}}^{\alpha} \approx \sigma_{n,p,q}\text{TW}_{\mathbb{C}}^{-1}(1-\alpha) + \mu_{n,p,q}$

Tracy-Widom Distribution



Proposition (Empirical CCA Consistency, Bao et al.)

Let $p, q, n \to \infty$ with $p/n \to c_x$ and $q/n \to c_y$. Given the above linear subspace data model,

$$\widehat{k}_{cca} \xrightarrow{a.s.} k \qquad \text{if } n > p + q \text{ and } \kappa_k^2 > r_c$$

where

$$r_c = \frac{c_x c_y + \sqrt{c_y c_y (1 - c_x)(1 - c_y)}}{(1 - c_x)(1 - c_y) + \sqrt{c_x c_y (1 - c_x)(1 - c_y)}}.$$

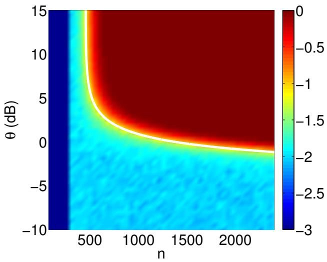
Recall

- $\widetilde{K}_{xy} = \left(\Theta_x + I_{k_x}\right)^{-1/2} \Theta_x P_{xy} \Theta_y \left(\Theta_y + I_{k_y}\right)^{-1/2}$
- Singular values $\kappa_1, \ldots, \kappa_{\min(k_x, k_y)}$

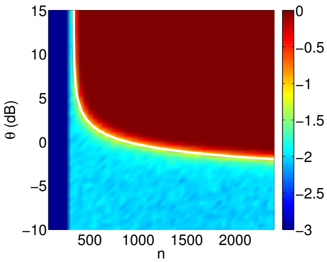
Simulation parameters

- $p = q = 150$, $k = 1$, $\theta = \theta_x = \theta_y$, $\alpha = 0.01$

$\rho = \mathbf{0.7}$



$\rho = \mathbf{0.9}$



Problems!

- Degenerate case: $\rho = 1$ when $n < p + q$ (Pezeshki 2004)
- Consistency boundary is dependent on correlation

Informative CCA (ICCA)

Not all singular vectors are informative! (Nadakuditi, 2011)

- $\sigma_i\left(\widehat{C}_{\text{cca}}\right) = \sigma_i\left(\widehat{V}_x^H \widehat{V}_y\right)$
- Trim data SVD's to only use informative components

1. Trim data SVD's: $X = \widehat{U}_x \widehat{\Sigma}_x \widehat{V}_x^H$ and $Y = \widehat{U}_y \widehat{\Sigma}_y \widehat{V}_y^H$
   - $\mathring{U}_x = \widehat{U}_x\left(:, 1:\widehat{k}_x\right)$, $\mathring{U}_y = \widehat{U}_y\left(:, 1:\widehat{k}_y\right)$
   - $\mathring{V}_x = \widehat{V}_x\left(:, 1:\widehat{k}_x\right)$, $\mathring{V}_y = \widehat{V}_y\left(:, 1:\widehat{k}_y\right)$
2. Form $\widehat{C}_{\text{icca}} = \mathring{U}_x \mathring{V}_x^H \mathring{V}_y \mathring{U}_y$
3. Take SVD: $\widehat{C}_{\text{icca}} = \widetilde{F}\widetilde{K}\widetilde{G}^H$
4. $\widehat{\rho}_{\text{icca}}^{(i)} = \widetilde{k}_i$
5. $\widetilde{w}_x^{(i)} = \widehat{R}_{xx}^{-1/2}\widetilde{f}_i$
6. $\widetilde{w}_y^{(i)} = \widehat{R}_{yy}^{-1/2}\widetilde{g}_i$

Statistical Test for ICCA Correlations

Estimate of # of Correlated Signals

$$\widehat{k}_{\text{icca}} = \sum_{i=1}^{\min(\widehat{k}_x,\widehat{k}_y)} \mathbb{1}_{\left\{\left(\widehat{\rho}_{\text{icca}}^{(i)}\right)^2 > \tau_{\text{icca}}^{\alpha}\right\}}$$

Setting the threshold

- F_{icca} is the cdf of largest singular values of $\widehat{C}_{\text{icca}}$ in the null setting of no correlation
- $\tau_{\text{icca}}^{\alpha} = F_{\text{icca}}^{-1}(1-\alpha)$
- $\tau_{\text{icca}}^{\alpha} \approx \sigma_{n,\widehat{k}_x,\widehat{k}_y} \text{TW}_{\mathbb{C}}^{-1}(1-\alpha) + \mu_{n,\widehat{k}_x,\widehat{k}_y}$

Theorem (ICCA Consistency)

Let $p, q, n \to \infty$ with $p/n \to c_x$ and $q/n \to c_y$. Given the linear subspace data model,

$$\widehat{k}_{icca} \xrightarrow{a.s.} k \quad \text{if } \min_{i=1,\ldots,k_x} \theta_i^{(x)} > c_x^{1/4} \text{ and } \min_{i=1,\ldots,k_y} \theta_i^{(y)} > c_y^{1/4}$$

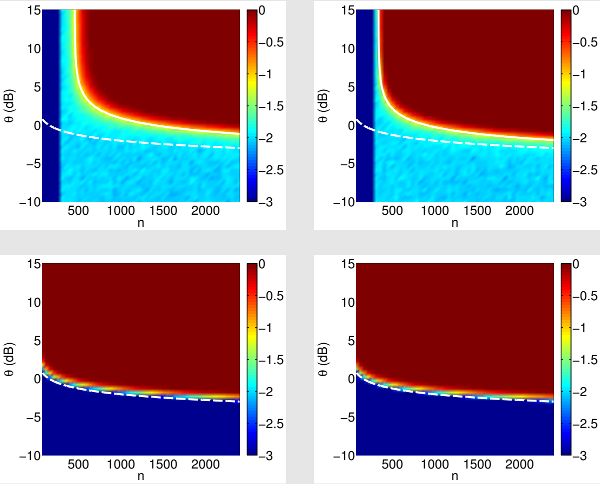
Empirical CCA and ICCA Consistency

- $p = q = 150$, $k = 1$, $\theta = \theta_x = \theta_y$, $\alpha = 0.01$

$\rho = 0.7$ $\rho = 0.9$

CCA

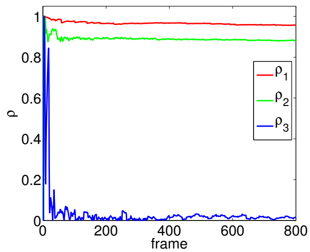
ICCA

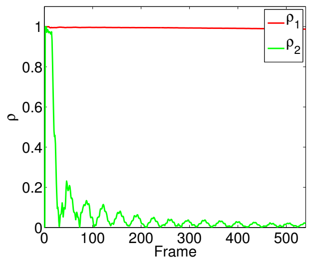


Empirical CCA and ICCA Demonstration

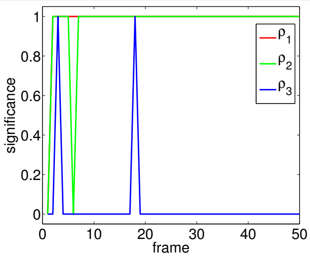
Video-Video | **Audio-Video**

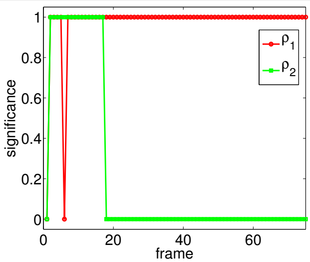
Correlations





Significance





Correlation Analysis Wish List

Correlation Algorithm Wish List

1. **Reliable in the sample deficient regime**
   - ✳ meaningful correlations ✓
   - ✳ meaningful canonical vectors
2. **Statistical test for correlations ✓**
   - ✳ consistency analysis ✓
3. **Robust**
   - ✳ non-gaussian data
   - ✳ missing data
4. **Extends to more than 2 datasets**

Missing Data Model

Matrix model

- $V_x = \left[s_{x,1}, \ldots, s_{x,n}\right]$, $V_y = \left[s_{y,1}, \ldots, s_{y,n}\right]$
- $Z_x = \left[z_{x,1}, \ldots, z_{x,n}\right]$, $Z_y = \left[z_{y,1}, \ldots, z_{y,n}\right]$

$$\boxed{\begin{aligned} X &= \left(U_x V_x^H + Z_x\right) \odot M_x \\ Y &= \left(U_y V_y^H + Z_y\right) \odot M_y \end{aligned}}$$

$$M_{ij}^x = \begin{cases} 1 & \text{with probability } \gamma_x \\ 0 & \text{with probability } 1 - \gamma_x \end{cases} \qquad M_{ij}^y = \begin{cases} 1 & \text{with probability } \gamma_y \\ 0 & \text{with probability } 1 - \gamma_y \end{cases}$$

- $\odot$ denotes the Hadamard or element-wise product.

Empirical CCA and ICCA Consistency in Missing Data

Theorem (Missing data consistency)

Let $p, q, n \to \infty$ with $p/n \to c_x$ and $q/n \to c_y$ and assume a low-coherence condition on the signal vectors. Given a linear subspace data model with missing data entries,

$$\widehat{k}_{cca} \xrightarrow{a.s.} k \qquad \text{if } \min_{i=1,\ldots,k} \mathring{\kappa}_i^2 > r_c \text{ and } n > p + q$$

$$\widehat{k}_{icca} \xrightarrow{a.s.} k \qquad \text{if } \min_{i=1,\ldots,\widehat{k}_x} \theta_i^{(x)} > \frac{c_x^{1/4}}{\sqrt{\gamma_x}} \text{ and } \min_{i=1,\ldots,\widehat{k}_y} \theta_i^{(y)} > \frac{c_y^{1/4}}{\sqrt{\gamma_y}}$$

where $\mathring{\kappa}_i$ are the singular values of

$$\left(\gamma_x \Theta_x + I_{k_x}\right)^{-1/2} \left(\gamma_x \Theta_x\right)^{1/2} P_{xy} \left(\gamma_y \Theta_y\right)^{1/2} \left(\gamma_y \Theta_y + I_{k_y}\right)^{-1/2}.$$

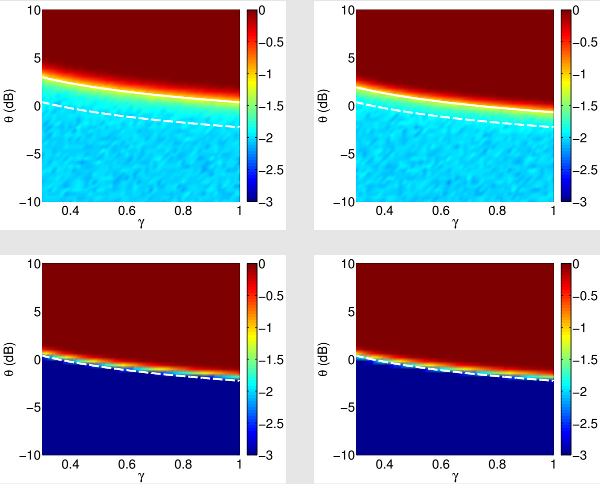
Empirical CCA and ICCA Consistency in Missing Data

- $p = q = 150$, $n = 1200$, $k = 1$, $\theta = \theta_x = \theta_y$, $\alpha = 0.01$

$\rho = 0.7$ $\rho = 0.9$

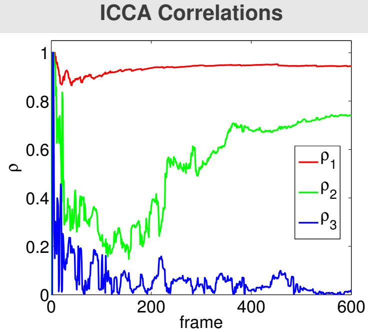
CCA

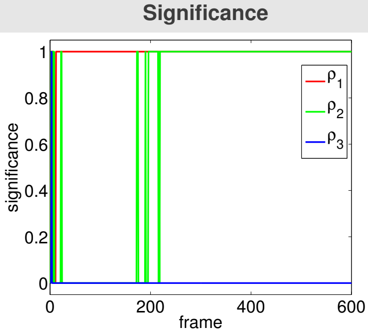
ICCA



Missing Data Demonstration

- $p = q = 32400$ pixels
- $n = 600$ frames
- $\gamma_x = \gamma_y = 0.75$





Non-Gaussian Data

Relax Noise Assumptions

- $z_{x,i}$ and $z_{y,i}$ are zero mean with bounded fourth moment

ICCA consistency theorem still holds

$$\boxed{\widehat{k}_{\text{icca}} \xrightarrow{\text{a.s.}} k \text{ if } \min_{i=1,\ldots,\widehat{k}_x} \theta_i^{(x)} > c_x^{1/4} \text{ and } \min_{i=1,\ldots,\widehat{k}_y} \theta_i^{(y)} > c_y^{1/4}}$$

General Form

- $z_{x,i}$ and $z_{y,i}$ are drawn from noise distributions μ_{Z_x} and μ_{Z_y}
- consistency dependent on D-transforms of noise distributions

Correlation Analysis Wish List

Correlation Algorithm Wish List

1. **Reliable in the sample deficient regime**
   - meaningful correlations ✓
   - meaningful canonical vectors
2. **Statistical test for correlations ✓**
   - consistency analysis ✓
3. **Robust ✓**
   - non-gaussian data ✓
   - missing data ✓
4. **Extends to more than 2 datasets**

MCCA - Multiset CCA

What if we have m datasets $Y_1, \ldots, Y_m$?

Notation

- covariance matrices: $R_{ij} = \mathbb{E}\left[y_i y_j^H\right]$
- canonical vectors: $x_1, \ldots, x_m$
- canonical variates: $w_1, \ldots, w_m$, $w_i = x_i^H y_i$

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

Optimization Problem

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

MCCA - Maximum Variance (MAXVAR)

MAXVAR Optimization Problem

$$\begin{aligned} \underset{x_1,\ldots,x_m}{\operatorname{argmax}} \quad & \lambda_1\left(\Phi(x)\right) \\ \text{subject to} \quad & x_i^H R_{ii} x_i = 1 \end{aligned}$$

Empirical MAXVAR

- Let V_i be the right singular vectors of dataset Y_i
- $\lambda_j\left(\widehat{\Phi}(x)\right) = \lambda_j\left(\widehat{C}_{\text{mcca}}\right)$

$$\widehat{C}_{\text{mcca}} = \begin{bmatrix} I_{d_1} & V_1^H V_2 & \cdots & V_1^H V_m \\ V_2^H V_1 & I_{d_2} & \cdots & V_2^H V_m \\ \vdots & \vdots & \ddots & \vdots \\ V_m^H V_1 & V_m^H V_2 & \cdots & I_{d_m} \end{bmatrix}$$

Proposed correlation statistic

$$\widehat{\rho}_{\text{mcca}}^{(j)} = \lambda_j\left(\widehat{C}_{\text{mcca}} - I\right)$$

Informative MCCA

Idea - Trim right singular vectors

- $\mathring{V}_i = \widehat{V}_i \left(:, 1 : \widehat{k}_i \right)$
- $\mathring{V} = \left[\mathring{V}_1, \ldots \mathring{V}_m \right]$

New IMCCA Matrix

$$\boxed{\widehat{C}_{\text{imcca}} = \mathring{V}^H \mathring{V}}$$

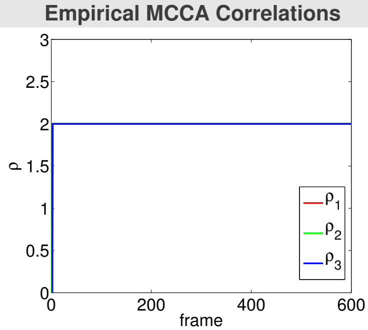
Proposed correlation statistic

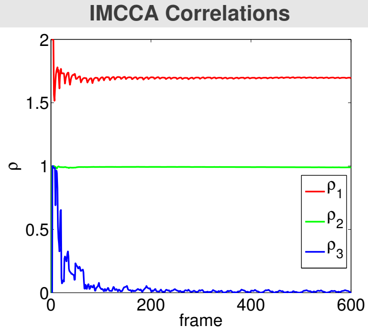
$$\boxed{\widehat{\rho}^{(j)}_{\text{imcca}} = \lambda_j \left(\widehat{C}_{\text{imcca}} - I \right)}$$

Insights

- Eigenvalue detects correlation
- Eigenvector reveals structure

Empirical MCCA and IMCCA Demonstration





Correlation Algorithm Wish List

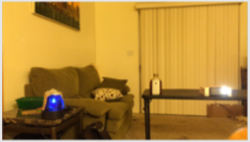
1. **Reliable in the sample deficient regime**
   - ✳ meaningful correlations ✓
   - ✳ meaningful canonical vectors
2. **Statistical test for correlations ✓**
   - ✳ consistency analysis ✓
3. **Robust ✓**
   - ✳ non-gaussian data ✓
   - ✳ missing data ✓
4. **Extends to more than 2 datasets ✓**

Data SVDs

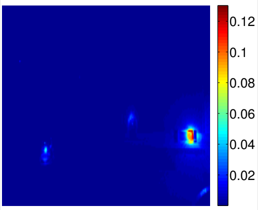
- $X = \widehat{U}_x \widehat{\Sigma}_x \widehat{V}_x^H$ with trimmed versions $\mathring{U}_x$, $\mathring{\Sigma}_x$, $\mathring{V}_x$
- $U_{\widetilde{K}}$ left singular vectors of $\widetilde{K}_{xy}$
- $\widehat{U}_{\widetilde{K}}$ left singular vectors of $\widehat{C}_{\text{cca}}$
- $\mathring{U}_{\widetilde{K}}$ left singular vectors of $\widehat{C}_{\text{icca}}$

Estimate	
Population	$W_x = U_x \left(\Theta_x + I_{k_x}\right)^{-1/2} U_{\widetilde{K}}$
Empirical CCA	$\widehat{W}_x^{\text{cca}} = \widehat{U}_x \left(\widehat{\Sigma}_x\right)^{-1} \widehat{U}_{\widetilde{K}}$
ICCA	$\widehat{W}_x^{\text{icca}} = \mathring{U}_x \mathring{\Sigma}_x^{-1} \mathring{U}_{\widetilde{K}}$
ICCA+	$\widehat{W}_x^{\text{icca+}} = \mathring{U}_x \Lambda_x^{\text{opt}} \mathring{U}_{\widetilde{K}}$

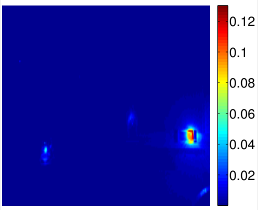
Original Scene



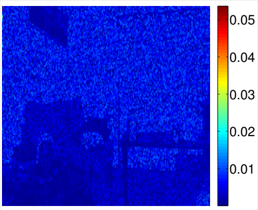
ICCA



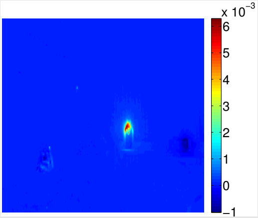
ICCA+



Empirical CCA



ICCA - ICCA+



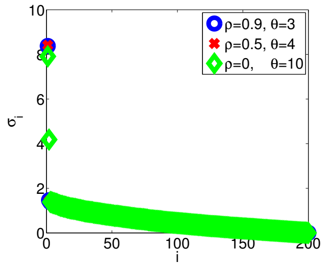
Correlation Algorithm Wish List

1. **Reliable in the sample deficient regime** ✓
   - ✳ meaningful correlations ✓
   - ✳ meaningful canonical vectors ✓
2. **Statistical test for correlations** ✓
   - ✳ consistency analysis ✓
3. **Robust** ✓
   - ✳ non-gaussian data ✓
   - ✳ missing data ✓
4. **Extends to more than 2 datasets** ✓

Recall key insight

- $k = \textbf{rank}(C_{\text{cca}}) = \textbf{rank}(R_{xy})$

Simulation of $\sigma_i\left(XY^H\right)$ with noise



Use SNR estimates

- Pre-whiten the datasets
- This is exactly CCA/ICCA!!

Contributions

- SVD of XY^H is solution to regularized CCA as $\eta \to \infty$
- Almost sure limit of $\sigma_i\left(XY^H\right)$

SVD of Random Projections - Chpt. 7

Signal-plus-noise Matrix

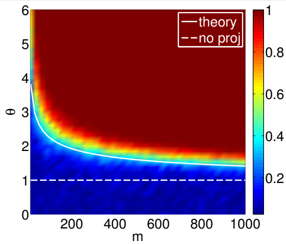
$$\widetilde{X}_n = \sum_{i=1}^{r} \theta_i u_i v_i^T + X_n.$$

Projections

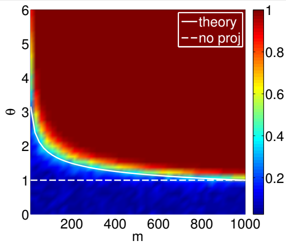
$$Y_n = P_n^H \widetilde{X}_n$$

- $P_n = G_n \in \mathbb{C}^{n \times m}$ with i.i.d. $\mathcal{CN}(0,1)$ entries
- $P_n = Q_n \in \mathbb{C}^{n \times m}$ s.t. $Q_n^H Q_n = I_m$

Gaussian



Unitary



Content Based Image Retrieval

Training Data

Caption 1

Caption 2

⋮ ⋮

Caption *n*

Text Query

snowy mountains

⇓

Automatic Image Annotation

Training Data

→ Caption 1

→ Caption 2

⋮ ⋮

→ Caption *n*

Image Query

⇓

snowy mountains

Correlation Based Approaches

Intuition

- Image features are correlated to words
- CCA/ICCA seem like natural algorithms to identify correlations

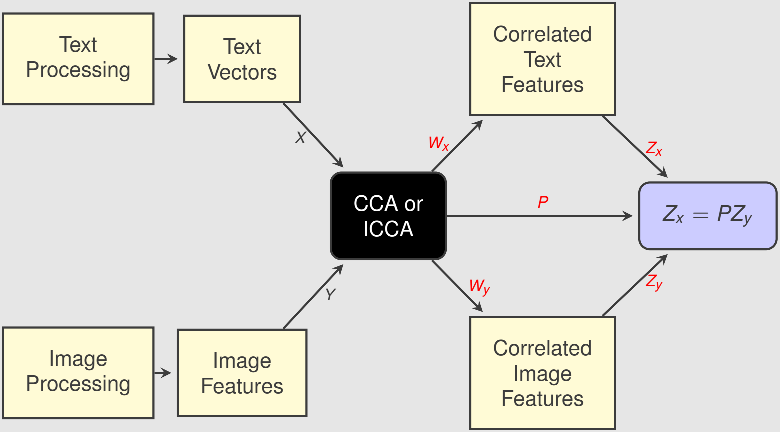
Text Processing - *tf*idf* feature vectors**

- term frequency (*tf*): frequent words are important
- inverse document frequency (*idf*): unique document words are important
- optional Porter stemming and stopword removal

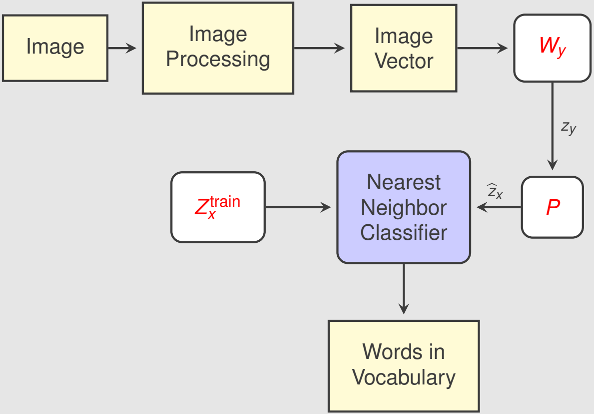
Image Processing - Visual Words

1. create SIFT features for all training images
2. k-mean clustering of all SIFT features to “visual words”
3. assign each SIFT keypoint to closest visual word
4. bag of words count the occurrences of visual word in each image

Training Pipeline



Automatic Caption Annotation Pipeline



Pascal Dataset

Pascal Dataset

- 1000 images each with 5 captions
- Vocabulary: 2393 words

- A D-ERFW-6 in flight.
- An army green plane flying in the sky.
- An old fighter plane flying with German military markings.
- A small green and yellow plane in the sky.
- A WWII fighter plane with its landing gear down.

Pascal Dataset - Image Retrieval

Text Query: airplane

CCA

ICCA

Pascal Dataset: Image Annotation

Image Query

CCA Annotation	**ICCA Annotation**
1. hairless	1. plane
2. buddi	2. ship
3. swan	3. cruis
4. leaf-less	4. fly
5. bnsf	5. blue
6. desert	6. jet
7. fluffi	7. airplan
8. salad	8. dock
9. majest	9. fighter
10. memorabilia	10. through

Correlation Algorithm Wish List

1. **Reliable in the sample deficient regime** ✓
   - meaningful correlations ✓
   - meaningful canonical vectors ✓
2. **Statistical test for correlations** ✓
   - consistency analysis ✓
3. **Robust** ✓
   - non-gaussian data ✓
   - missing data ✓
4. **Extends to more than 2 datasets** ✓

Future Work

Kernel CCA

- analysis for nonlinear low-rank signal-plus-noise model
- analysis of choice of kernel
- analysis of choice of regularization parameter

MCCA

- MAXVAR theorem for deterministic correlation when $n < \sum_i d_i$
- null distribution for statistical test to estimate number of correlated signals
- consistency analysis for estimate of number of correlated signals

Image Annotation and Retrieval

- clever feature engineering with NLP techniques
- extension to nonlinear algorithms

Acknowledgments

Committee

- Prof. Hero, Prof. Laura, Prof. Rada
- Prof. Raj

Colleagues

- Nadakuditi group
- GSC and SPeecs committees
- 3M ATG

Friends

- Hereford
- University of Maryland
- University of Michigan

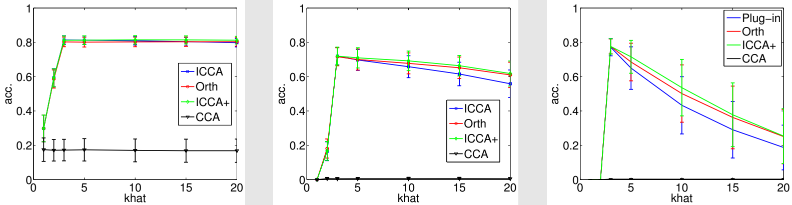
Family

ICCA+ Robustness to $\widehat{k}_x$ overestimation

- $k_x = k_y = 3$, $p = 200$, $q = 250$, $n = 1000$, $\Theta_x = \Theta_y = \textbf{diag}(3, 2, 1)$, $P_{xy} = \textbf{diag}(0.9, 0.5, 0.3)$, $V_k = I_3$

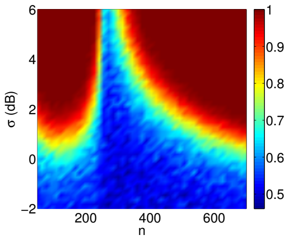
$$U_k = \left[\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right].$$

$w_x^{(1)}$ $w_x^{(2)}$ $w_x^{(3)}$

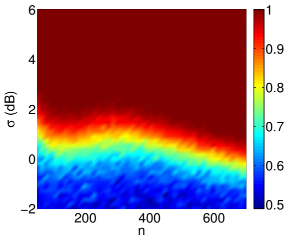


Limit of RCCA

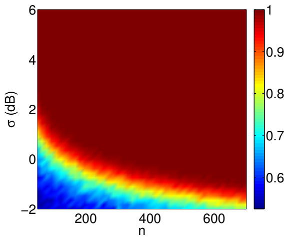
✳ AUC for signal vs. noise, $k = 1$, $p = 100$, $q = 150$, $\rho = 0.9$



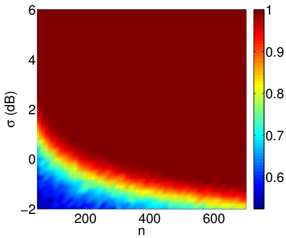
(a) $\eta = 0.0001$



(b) $\eta = 0.01$



(c) $\eta = 10$



(d) $\eta = 1000$