Face recognition via synthetic discriminants: A relative comparison of accuracy and efficiency

William Hilton

Department of Mathematics
University of Central Florida

Motivation

- Linear optical pattern recognition is now as relevant as ever for personal use, state interests, and commercial profits.
- Linear flavor still surprisingly useful due to today's relatively small processing-power-to-data-size ratio.
- Examples include recognition (social media, military GIS, tolls), identification (security devices, signal searching), classification (tracking, discovery, matching), etc

Scope & Agenda

- Briefly review synthetic discriminant functions (SDFs)
- Explore 'applied' considerations w.r.t. face recognition
- Discuss construction, execution, and comparison of three implemented SDFs

History: Constrained SDFs

- Hester and Casasent[1] take a first bite with Equal correlation plane synthetic discriminant function.
- Kumar[2] incorporates noise tolerance with Minimum variance synthetic discriminant function. Is useless due to difficulty in inverting noise power spectral density.
- Mahalanobis[3] controls whole plane with Minimum average correlation filters.
- Refregier[4] combines all in optimal trade-off filter.

[1] C.F. Hester, D. Casasent, Multivariate technique for multiclass pattern recognition, Appl. Opt. 19 (1980) 1758-1761.

[2] B.V.K. Vijaya Kumar, Minimum variance synthetic discriminant functions, J. Opt. Soc. Am (1986) 1579-1589.

[3] A. Mahalanobis, B.V.K. Vijaya Kumar, D. Casasent, Minimum average correlation filters, Appl. Opt. 26 (1987) 3630-3633.

[4] P. Refregier, Filter Design for optical pattern recognition: Multi-criteria optimization approach, Opt. Let. 15 (1990) 854-856.

History: Unconstrained and Beyond

- Mahalanobis[5] eliminates counterproductive restrictions with unconstraied MACE and OTF.
- Kumar[6] then synthesized results into new distance classifier filters.
- Savvides[7] adds false-class training to DCCF, dubbed minimax distance transform correlation filter.
- Tweaks continue, as recently as July 2017 with Thirunavukkarsu's[8] disease classifier.

[5] A. Mahalanobis, B.V.K. Vijaya Kumar, D. Casasent, Unconstrained correlation filters, Appl. Opt. 33 (1994) 3659-3751

[6] B.V.K. Vijaya Kumar, D. Casasent, A. Mahalanobis, Distance- classifier correlation filters for multiclass recognition, Appl. Opt. 35 (1996) 3127-3133.

[7] M. Savvides, B.V.K. Vijaya Kumar, P.K. Khosla, Two-class minimax distance transform correlation filter, Appl. Opt. 41 (2002) 6829-6840.

[8] Thirunavukkarsu, K. S. A Fast Correlation Filter Based Gradient Boosting Classifier for Disease Diagnosis. Intl. Journ. of Adv. Re. in Comp. Sci. 8 (2017) 900-909.

Notation

- We restrict our attention to one-tuple optical signals (i.e. black and white images) in a spatially discretized dimension of size $d_1 d_2 = d$.
- In a given family of the two-dimensional fast Fourier transforms of some images of faces, we let $T_{C,i} \in \mathbb{C}^d$ be the i^{th} column vector of lexicographically ordered images of size d be longing to person C and we let $F_{C,i} \in \mathbb{C}^d$ be the same, except *not* belonging to person C.
- For notational convenience, given a matrix A, we let $A(\mathrm{e})$ represent its diagonalization. Since the result is sparse, the equivalent computational object is usually stored compactly.

Notation

Taking N_T, N_F to be the number of true/false images, we define

$$M_{C,t} = \frac{1}{N_T}\sum_{k=0}^{N_T} T_{C,k} \quad M_{C,f} = \frac{1}{N_F}\sum_{k=0}^{N_F} F_{C,k} \tag{1}$$

$$D_{C,t} = \frac{1}{N_T}\sum_{k=0}^{N_T} \|T_{C,k}(\mathrm{e})\|_2^2 \quad D_{C,f} = \frac{1}{N_F}\sum_{k=0}^{N_F} \|F_{C,k}(\mathrm{e})\|_2^2 \tag{2}$$

$$S_{C,t} = \frac{1}{N_T}\sum_{k=0}^{N_T} (T_{C,k}(\mathrm{e}) - M_{C,t}(\mathrm{e}))^*(T_{C,k}(\mathrm{e}) - M_{C,t}(\mathrm{e})) \tag{3}$$

$$S_{C,f} = \frac{1}{N_F}\sum_{k=0}^{N_F} (F_{C,k}(\mathrm{e}) - M_{C,f}(\mathrm{e}))^*(F_{C,k}(\mathrm{e}) - M_{C,f}(\mathrm{e})) \tag{4}$$

and we may interpret M as the frequency mean, D as the power spectral density, and S as the spectral variance of the true and false classes of person C, respectively.

THEORY

Taking $\vec{u}$ to be an N by 1 column vector, we have the filters

$$h_{\mathrm{ECPSDF;\,C}} = T_C \cdot (T_C^+ \cdot T_C)^{-\mathbf{1}} \cdot \vec{u} \tag{5}$$

$$h_{\mathrm{MACE;\,C}} = D_{C,t}^{-1} T_C \cdot (T_C^+ \cdot D_{C,t}^{-1} T_C)^{-\mathbf{1}} \cdot \vec{u} \tag{6}$$

$$h_{\mathrm{UMACE;\,C}} = D_{C,t}(\mathrm{e})^{-1} M_{C,t} \tag{7}$$

$$h_{\mathrm{DCCF;\,C}} = S_t(\mathrm{e})^{-1} M_{C,t} \tag{8}$$

$$h_{\mathrm{mMDTCF;\,C}} = (S_t(\mathrm{e}) + S_f(\mathrm{e}))^{-1} (M_{C,t} - M_{C,f}) \tag{9}$$

corresponding to the C^{th} class and where, notably, the first two filters require true matrix operations (denoted with boldface operators and written product dots), while the latter three require no traditional matrix operations.

Synthetic Discriminant Function Family

$$h_{\text{ECPSDF}} = X(X^{+}X)^{-1}\vec{u}$$

Unconstrained-types

$$h_{\text{UMACE}} = D^{-1}M$$

$$h_{\text{OTF}} = (\alpha D + \sqrt{1-\alpha^2}C)^{-1}X(X^{+}(\alpha D + \sqrt{1-\alpha^2}C)^{-1}X)^{-1}\vec{u}$$

Distance Classifiers

$$h_{\text{DCCF}} = S^{-1}M$$

$$h_{\text{MDTCF}} = S_t^{-1}(R - M_f)$$

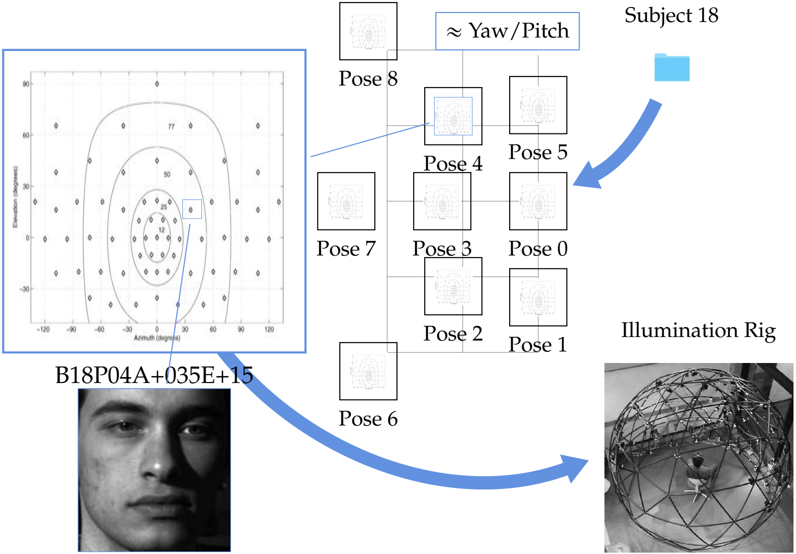
$$h_{\text{mMDTCF}} = (S_t + S_f)^{-1}(M_t - M_f)$$

Focus: Application

- We implement ECP, MACE, UMACE, DCCF, MDTC, and mMDTC on the first ten subjects of the Extended Yale Face Database B[9].

[9] Lee, Kuang-Chih, Jeffrey Ho, and David J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions On Pattern Analysis And Machine Intelligence no. 5 (2005): 684.

EXT. YALE FACE DATABASE B ARCHITECTURE

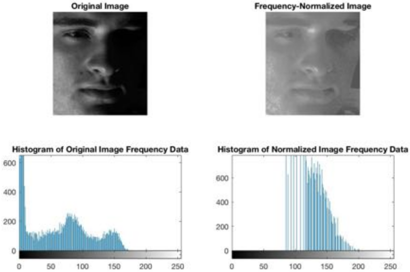


Database Normalization Part 1

- In line with standard practice, we normalize the database to eliminate bias in the comparative filter analysis.
- Manually identify and uniformly discard all ‘bad’ (i.e., ambient and/or no flash) photos.
- Manually record face locations.
- Algorithmically crop images based on the average face size and location.

Database Normalization Part 2

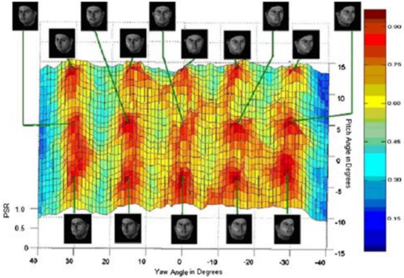
- Normalize color map via Retinex Method[10].
- Essentially, frequency distribution of image is mapped to a normal curve, as in following figure.



[10]Park, Young Kyung, Seok Lai Park, and Joong Kyu Kim. Retinex method based on adaptive smoothing for illumination invariant face recognition. Signal Processing 88.8 (2008): 1929-1945.

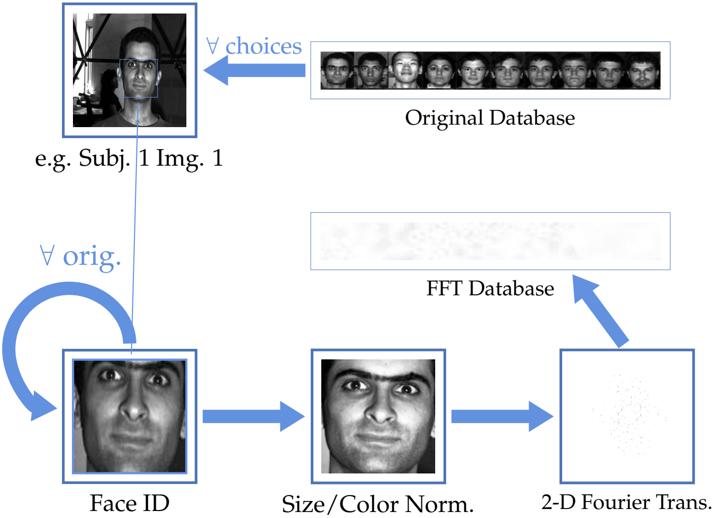
Database Normalization Bonus Part

- Normalize relative distance between facial features (which is inconsistent due to yaw and pitch of bust among poses) via projective mappings and incorporate noise adjustments



- Implement in a future iteration

Database Normalization Architecture



Filter Design and Testing

- Randomly choose N images (per-person) for filter design, i.e. training data. Remaining $N_{\text{test}} = 9 \cdot 10 \cdot 59 - N$ images used for testing.

Note that our filters are designed to maximize the so-called peak-to-sidelobe ratio (PSR) of the spatial correlation between the filter and image. In pseudocode, we define

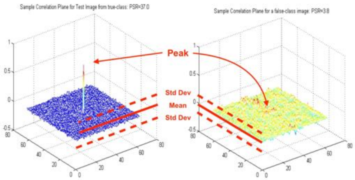
```
% PSR for correlation plane C
PSR = (max(C)-mean(C))/std(C);
```

Filter Design and Testing Architecture

Normalized FFT Database → Testing Database

Normalized FFT Database → Training Database → Filter Design → $h_{1,2,3;C}$ → $\otimes$

Testing Database → $\otimes$ → Spatial Correlation Plane



Comparative Filter Performance: Method 1

- "Filter response" to true/false class on previous slide is merely ideal
- For convenient relative metric, choose maximum PSR (per filter, per image) to be positive identification.

In particular, we map PSR responses to predicted subject indices by

$$\begin{pmatrix} \mathrm{PSR}(h_{\mathrm{MACE};\,1} \otimes \tau_1) & \cdots & \mathrm{PSR}(h_{\mathrm{MACE};\,10} \otimes \tau_1) \\ \vdots & \ddots & \vdots \\ \mathrm{PSR}(h_{\mathrm{MACE};\,1} \otimes \tau_{N_{\mathrm{test}}}) & \cdots & \mathrm{PSR}(h_{\mathrm{MACE};\,10} \otimes \tau_{N_{\mathrm{test}}}) \end{pmatrix} \overset{\max}{\longmapsto} \begin{pmatrix} C(\tau_1) \\ \vdots \\ C(\tau_{N_{\mathrm{test}}}) \end{pmatrix}$$

for each filter, and we condense the record into a confusion matrix.

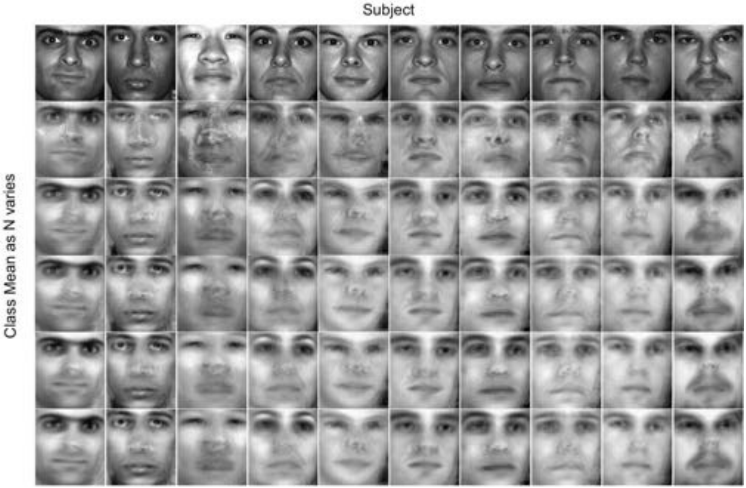
Comparative Filter Performance: Method 2

- Also consider an alternative performance metric:
- Compute integral of the receiver operating characteristic (ROC) curve for list of PSR scores corresponding to each filter, which generates an absolute accuracy statistic.
- Record total filter computation time (per filter), an absolute efficiency statistic.
- Consider imposing a scaled l_2 norm of the accuracy, efficiency two-tuple and then rank order the filters.

Execution & Results: Parameter Control

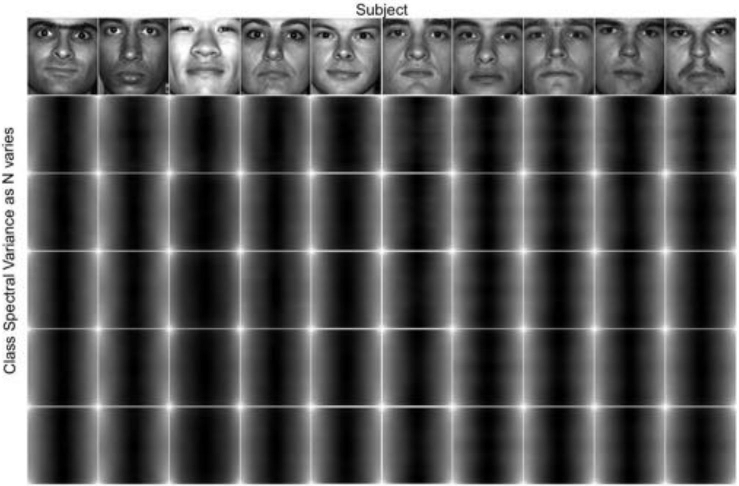
- Execute for $N \in \{1, 5, 10, 15, 20\}$ per pose, per subject, since pose normalization is not considered.
- Equals $N_{\text{Train}} \in \{9, 45, 90, 135, 180\}$ per subject. This implies a training-to-testing ratio of roughly $\frac{1}{100}, \frac{1}{10}, \frac{1}{5}, \frac{1}{4}$, and $\frac{1}{3}$, respectively.
- At each N, randomly select among the 59 illumination-varying images, since illumination was normalized.

Class Means for $N_{\text{Train}} \in \{9, 45, 90, 135, 180\}$



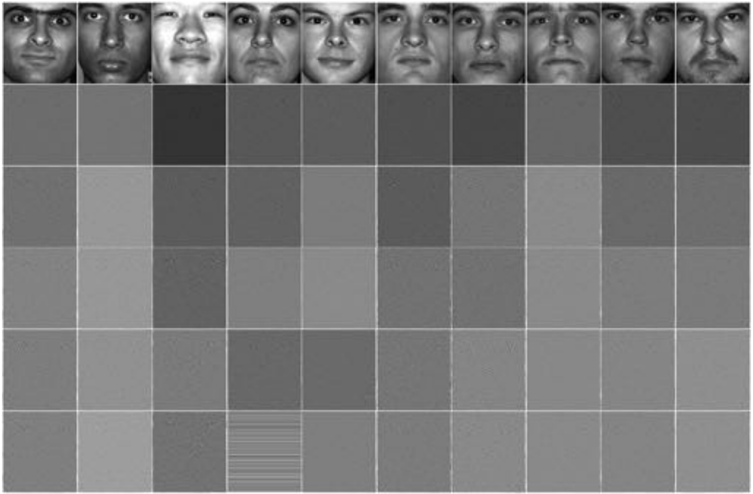
Notice excellent illumination reconciliation and poor pose reconciliation, especially for small N.

Class Spec. Var. for $N_{\text{Train}} \in \{9, 45, 90, 135, 180\}$



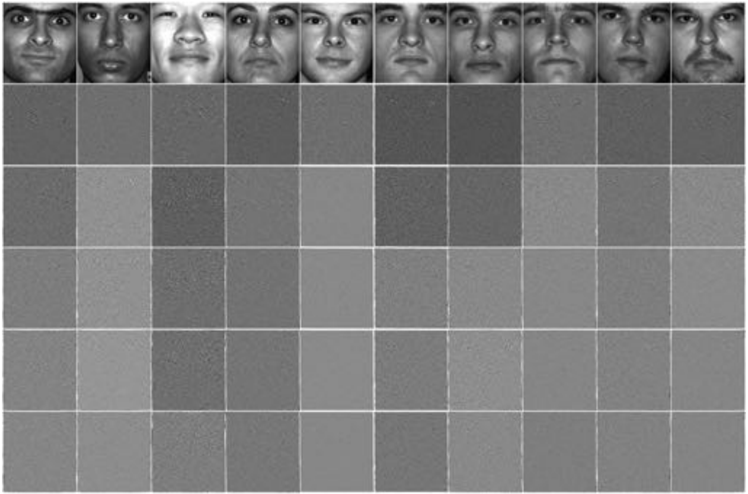
Note low variance in Subject 3.

MACE Filters for $N_{\text{Train}} \in \{9, 45, 90, 135, 180\}$



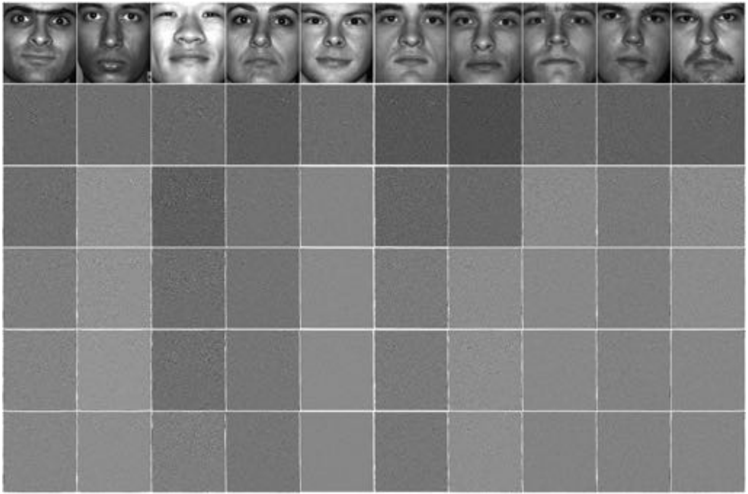
Erroneous result occured when matrix inversion failed.

UMACE FILTERS FOR $N_{\text{TRAIN}} \in \{9, 45, 90, 135, 180\}$



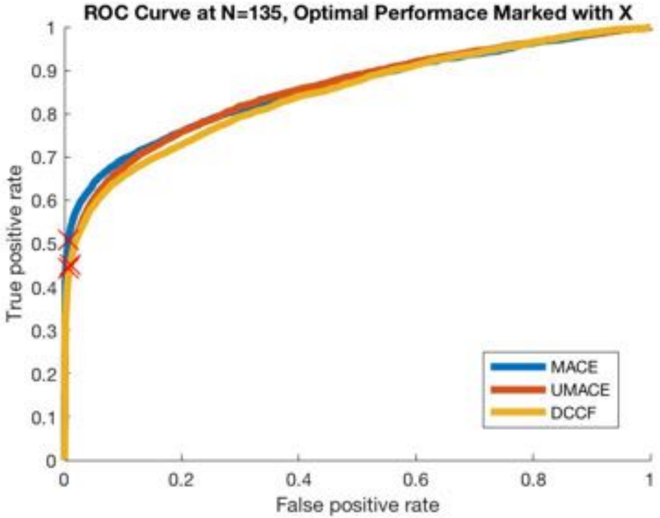
Note more uniform filter energy.

DCCF Filters for $N_{\text{Train}} \in \{9, 45, 90, 135, 180\}$

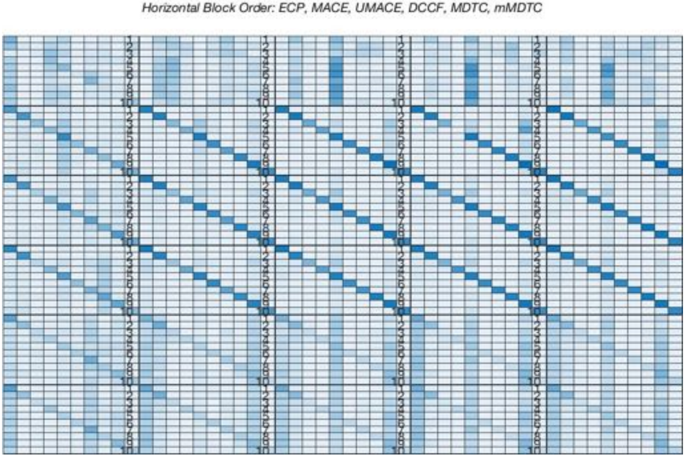


Filter energy appears to be less uniform than UMACE.

ROC Curve Example

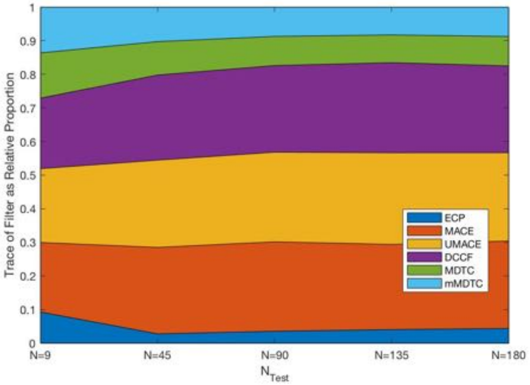


Confusion Mats. for $N_{\text{Train}} \in \{9, 45, 90, 135, 180\}$



Note effect of matrix inversion failure in block $(2, 4)$.

TRACE PROPORTION FOR $N_{\text{TRAIN}} \in \{9, 45, 90, 135, 180\}$



Demonstrates that MACE, UMACE, and DCCF are superior to MDTC and mMDTC, which are superior to ECP. We use Method 2 to compare behavior more closely.

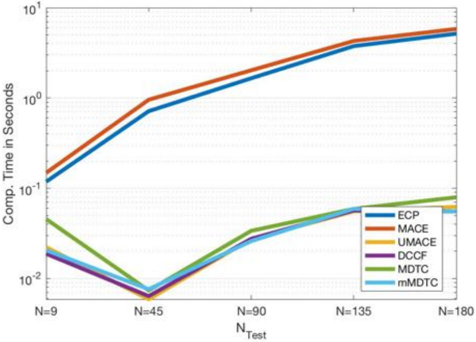
Comparative Limitations Remark

We would like to find some real mapping $\mathcal{F} : \mathbb{R}^2 \to \mathbb{R}$ that measures accuracy and efficiency pairs. Generally, we would have some l_2 norm-like mapping

$$\sqrt{\mathcal{F}_1(\text{Acc})^2 + \mathcal{F}_2(\text{Eff})^2} \tag{10}$$

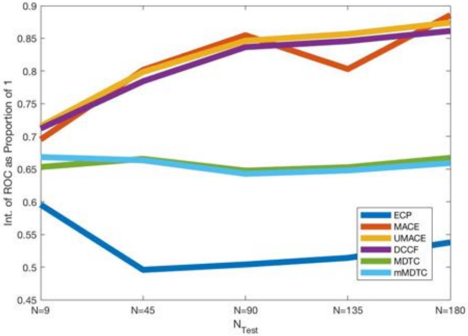
which will measure of the extent of inaccuracy and inefficiency of the filters. As desired, this would impose a well-ordering of the filters, but it would (and must) relate accuracy and efficiency in a manner that is not necessarily prescribed within the context of our problem. To proceed, we use possibly fuzzy judgment.

Method 2 Data: Efficiency



Assessment: Matrix-dependent filters enjoy efficiency linearly dependent upon training data size. Parallel processing would eliminate this comparative inefficiency at a cost which is unlikely to exceed *comparative* benefits.

Method 2 Data: Accuracy



Assessment: ECP cannot negotiate large training sizes, MDTC and mMDTC cannot compactify large false-class metrics, MACE is sensitive to matrix computation issues, and UMACE is slightly superior to DCCF.

Conclusion

- In consideration of the absolute and comparative metrics of accuracy, efficiency, and their norms across various N_{Train}, we support the claim that UMACE is followed by DCCF and then MACE in terms of robustness with respect to face recognition under varying illumination and pose. The remaining filters, ECP, MDTC, and mMDTC, are said to be inappropriate for face recognition with large or highly varying false-classes.
- Our finding is upheld in the literature.[11]

[11] Levine, Martin David and Yu, Yingfeng. Face recognition subject to variations in facial expression, illumination conditions and poses using correlation filters. IEEE Transactions Computer Vision and Image Understanding 104 (2006) 1-15

End

- Future consideration: To incorporate feature data into linear processing paradigm and consequently improve recognition performance, it may be advantageous to extend signal data into binary indicial spaces representing feature on/off status.
- Thank you for listening.