Multimodal Biometrics

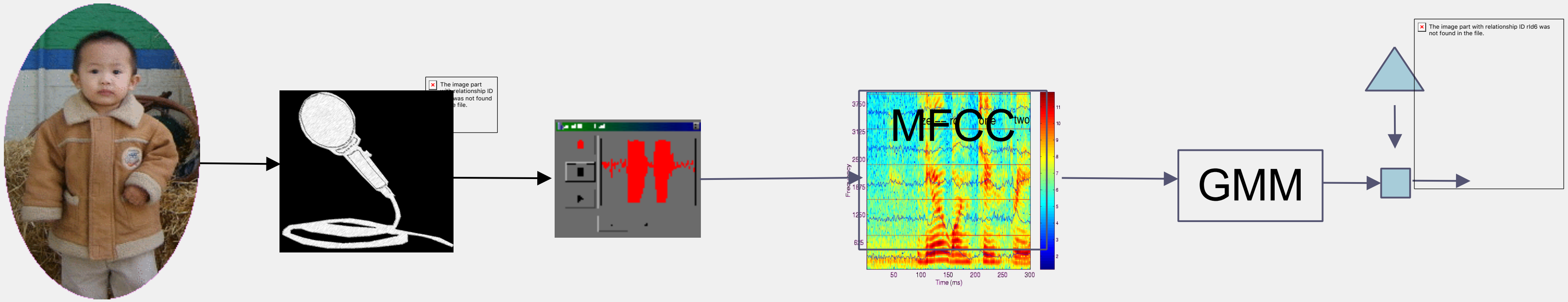
Josef Kittler

Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford GU2 7XH

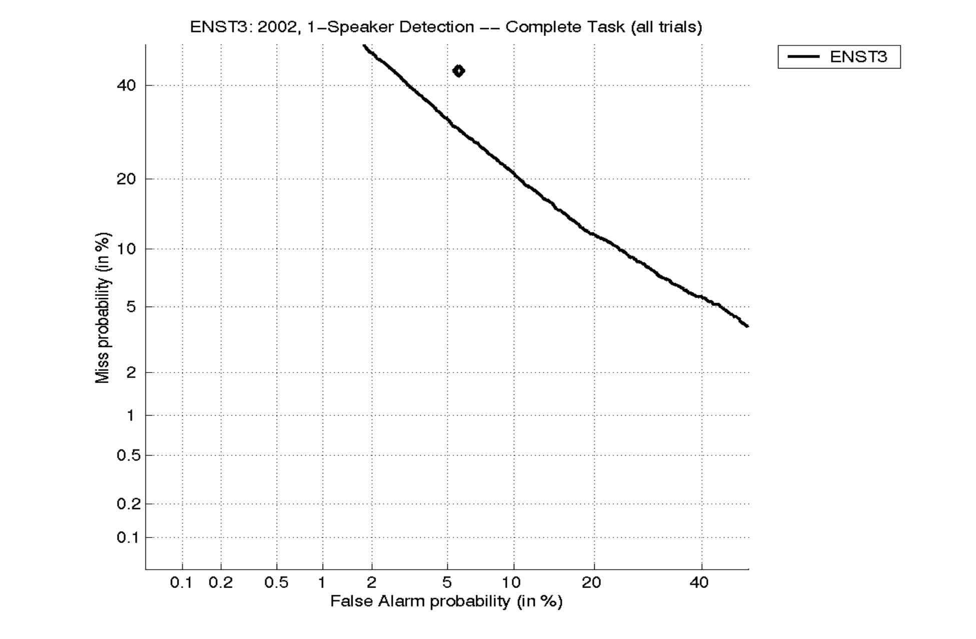
J.Kittler@surrey.ac.uk

Acknowledgements: Dr Norman Poh

Biometric authentication and Performance characterisation



- False rejection
- False acceptance
- Total error rate/Half total error rate
- Operating point
  - Equal error rate (civilian)
  - Zero false acceptance (high security forensic)
  - Zero false rejection (low risk banking)



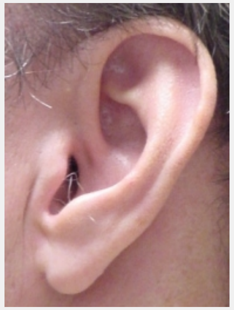
UNIVERSITY OF SURREY

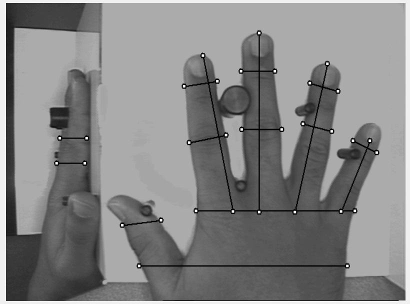
Multimodal biometrics

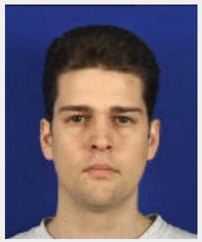
- Different biometric modalities developed
  - –finger print
  - –iris
  - –face (2D, 3D)
  - –voice
  - –hand
  - –lips dynamics
  - –gait

Different traits- different properties

- •usability
- •acceptability
- •performance
- •robustness in changing environment
- •reliability
- •applicability (different scenarios)







Benefits of multimodality

- Motivation for multiple biometrics
  - To enhance performance
  - To increase population coverage by reducing the failure to enroll rate
  - To improve resilience to spoofing
  - To permit choice of biometric modality for authentication
  - To extend the range of environmental conditions under which authentication can be performed

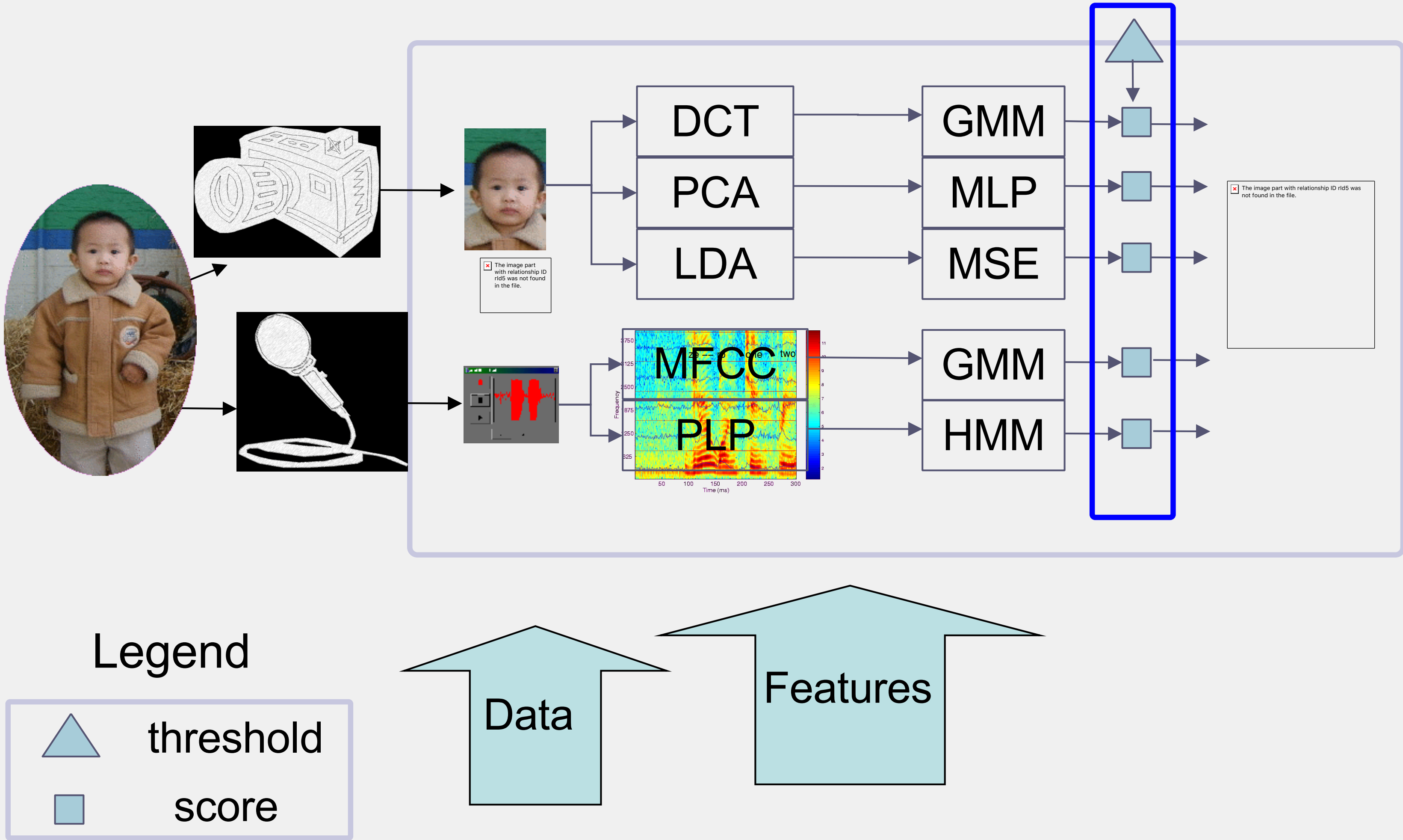
OUTLINE

- Fusion architectures
- Score level fusion: Problem formulation
- Estimation error
- Multiple expert paradigm
- Quality based fusion of biometric modalities
- Discussion and conclusions

Fusion architectures

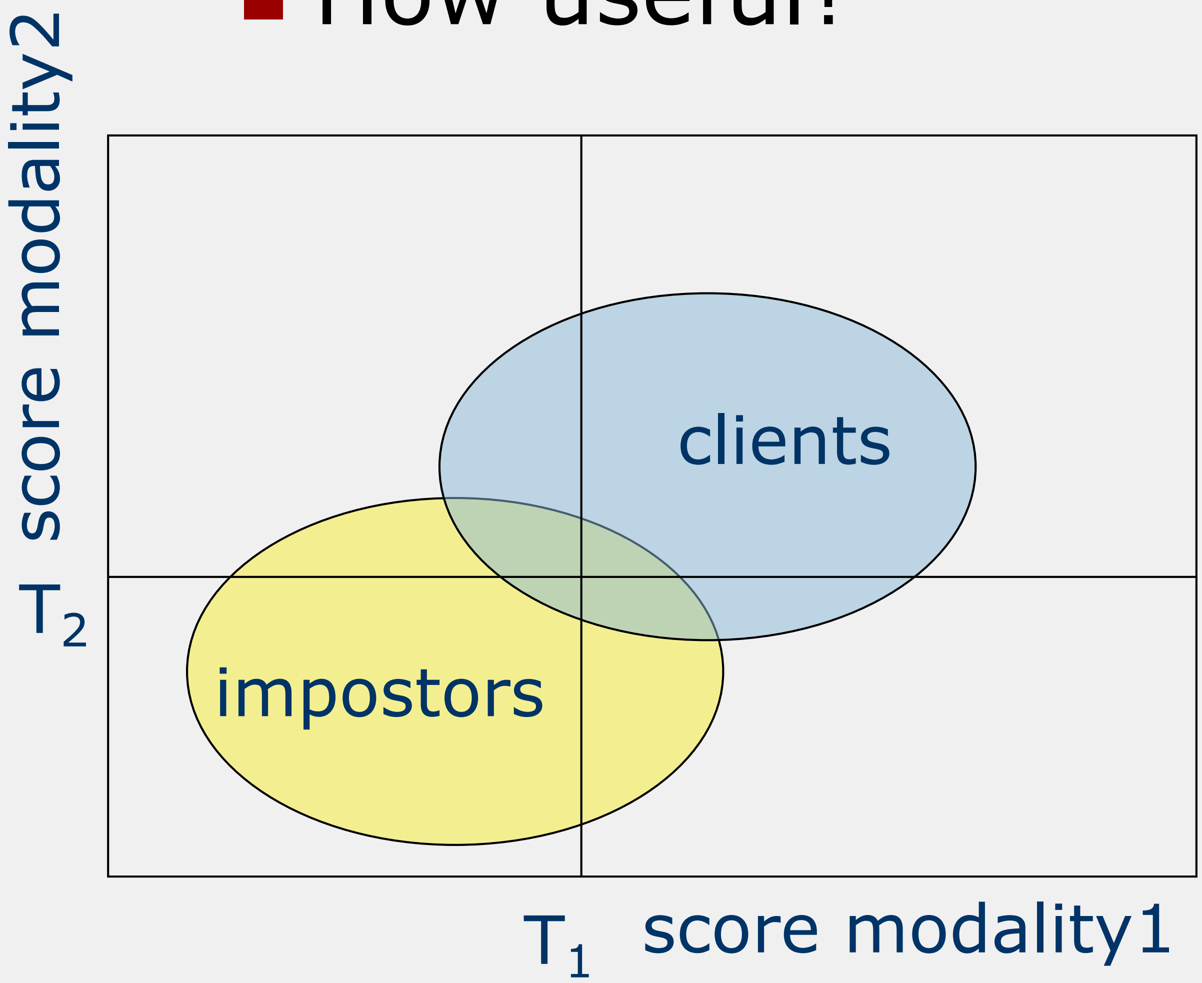
- Integration of multiple biometric modalities
- Sensor (data) level fusion
  - Linear/nonlinear combination of registered variables
  - Representation space augmentation
- Feature level fusion
- Soft decision level fusion
- Decision level fusion

Decision level fusion



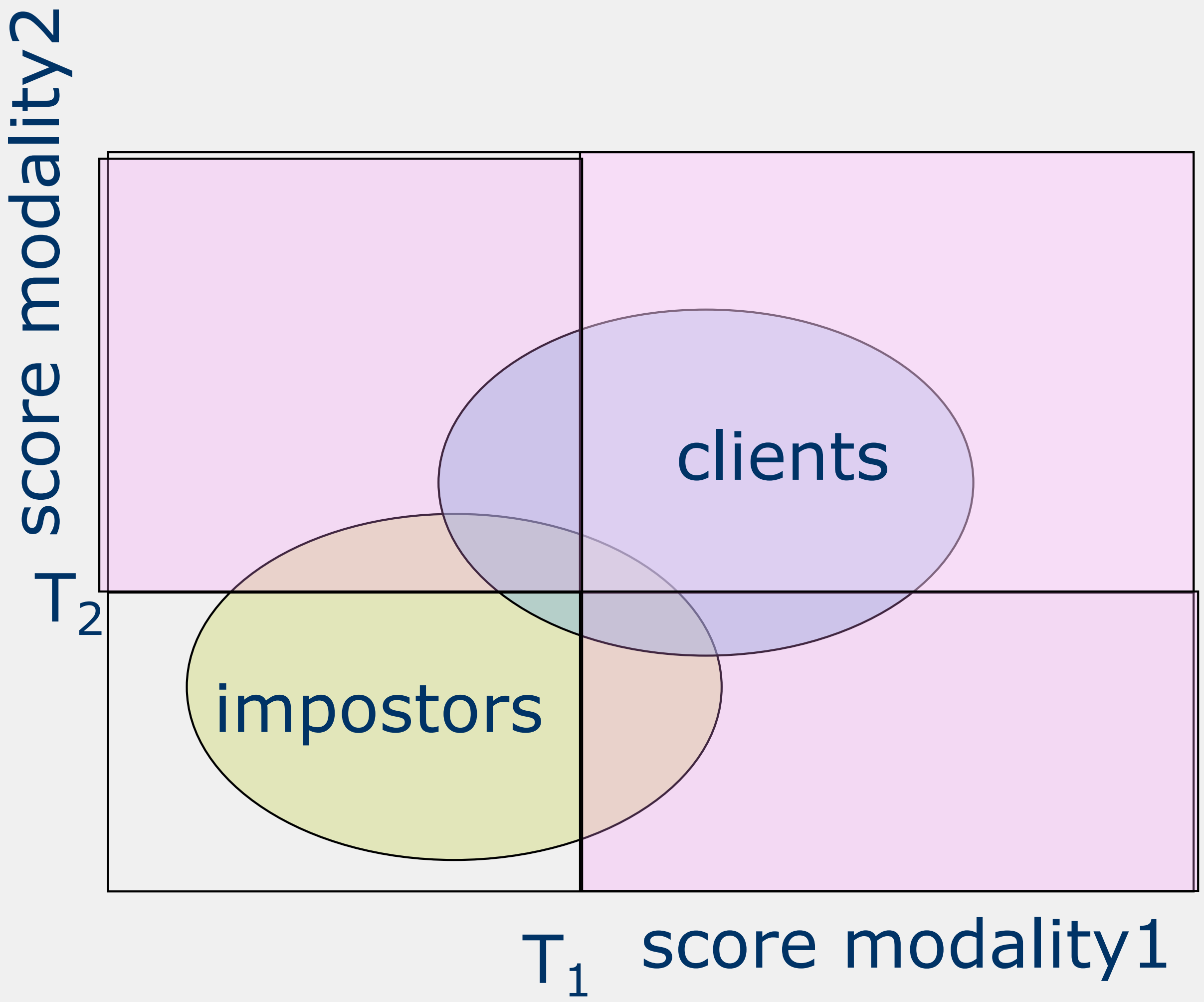
Decision-level fusion

- How useful?



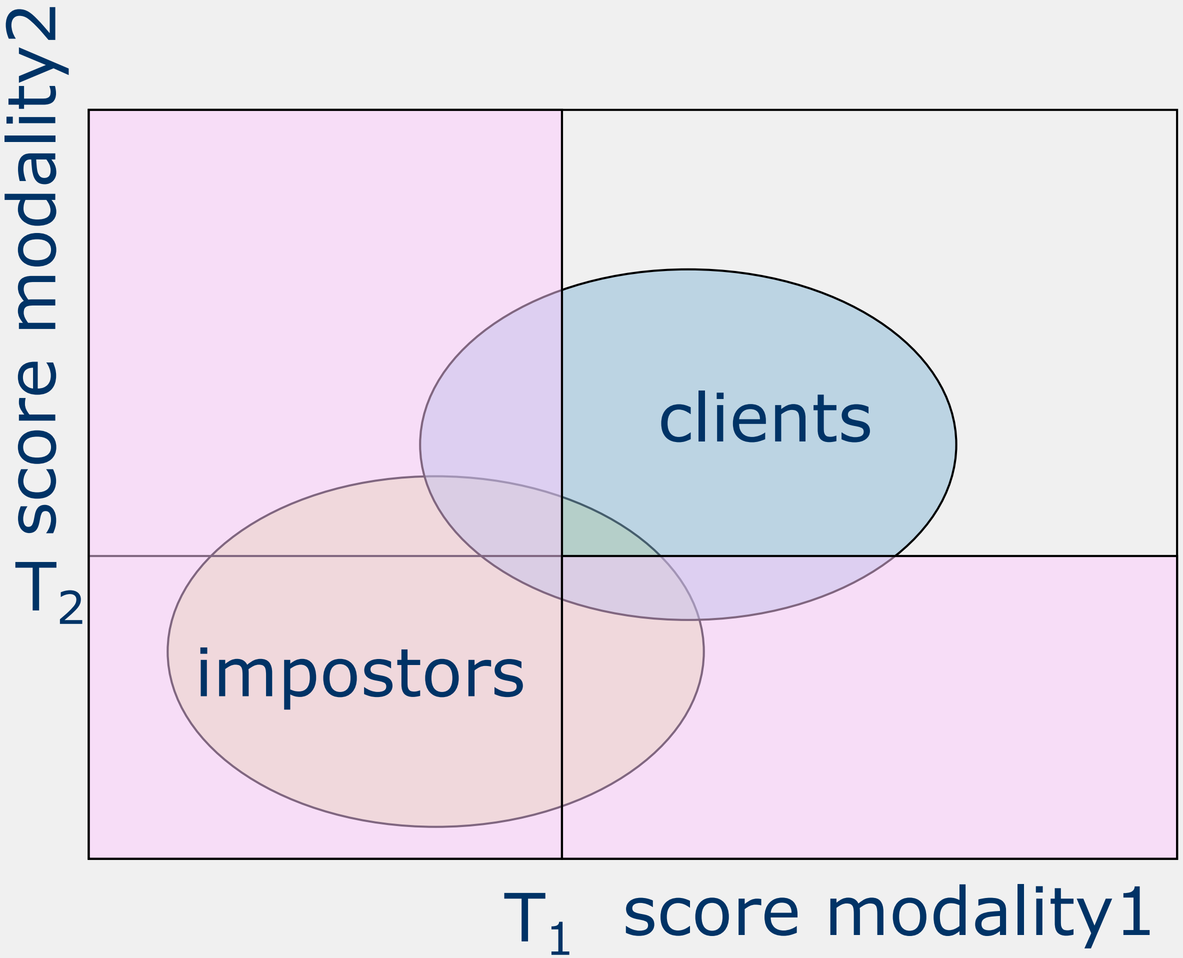
Decision-level fusion

- Accepted by either modality



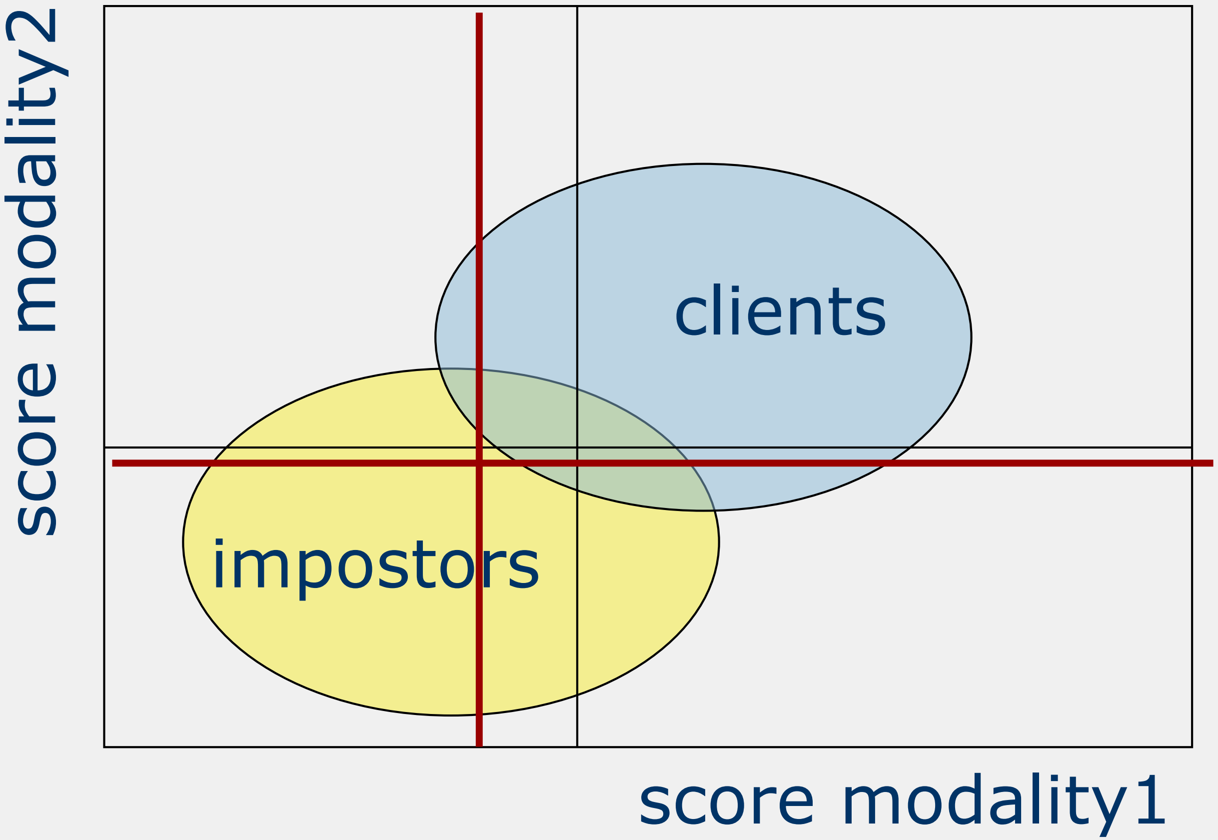
Decision-level fusion

- Accepted by both



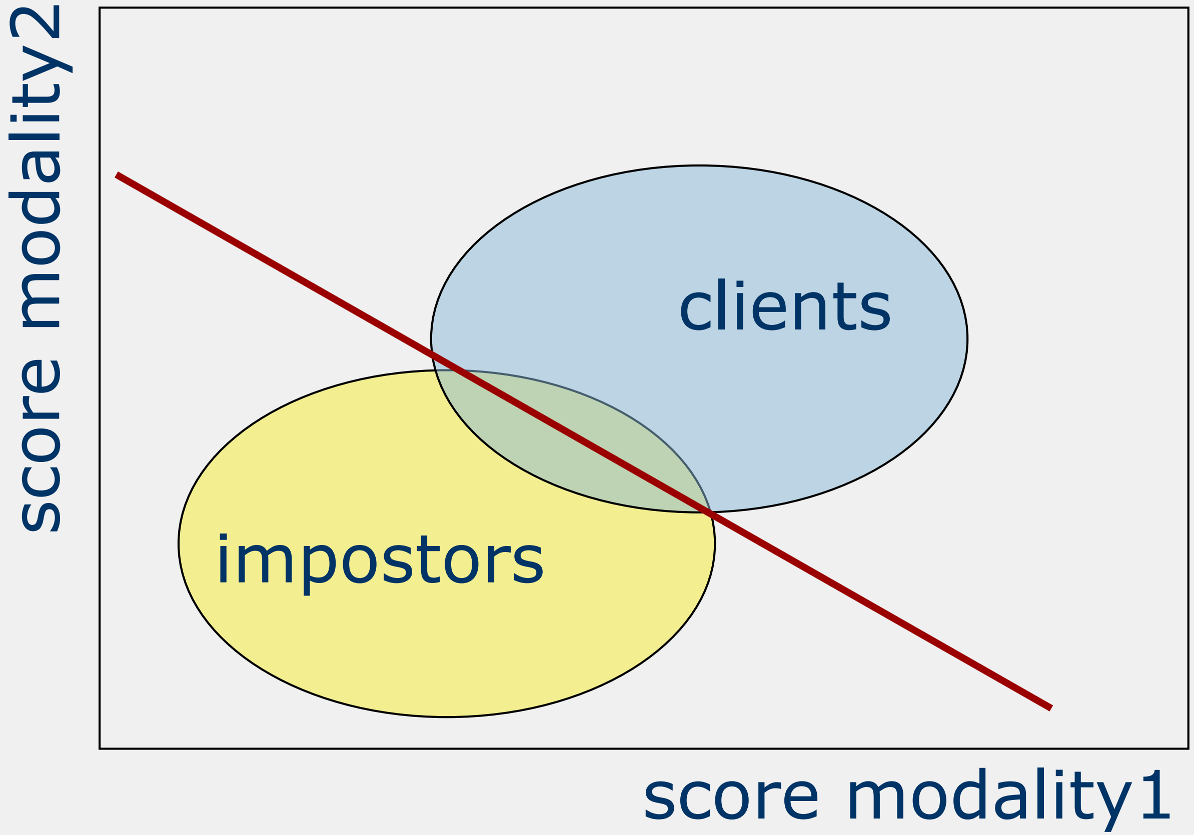
Decision-level fusion

Better performance by adapting the thresholds

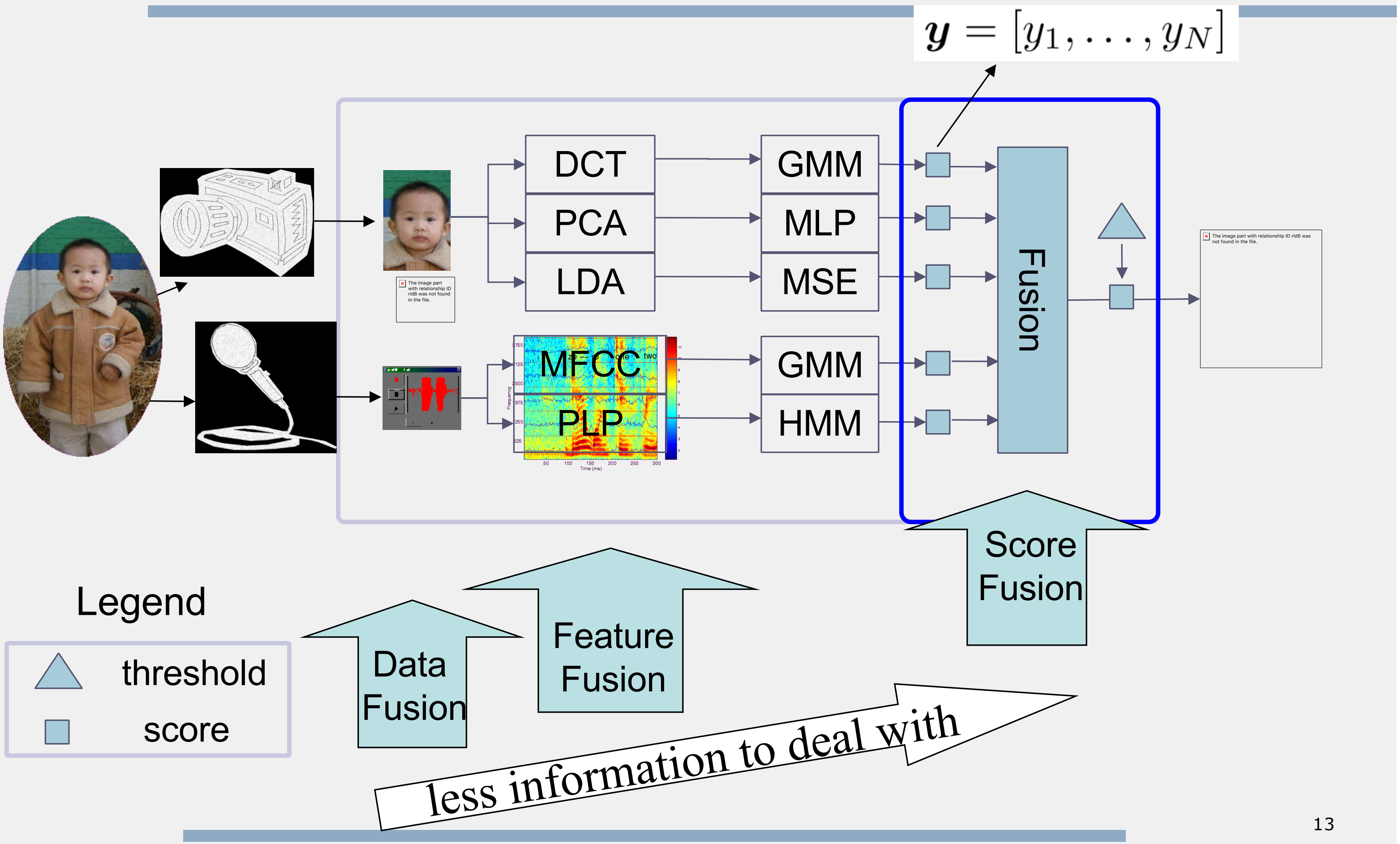


Score-level fusion

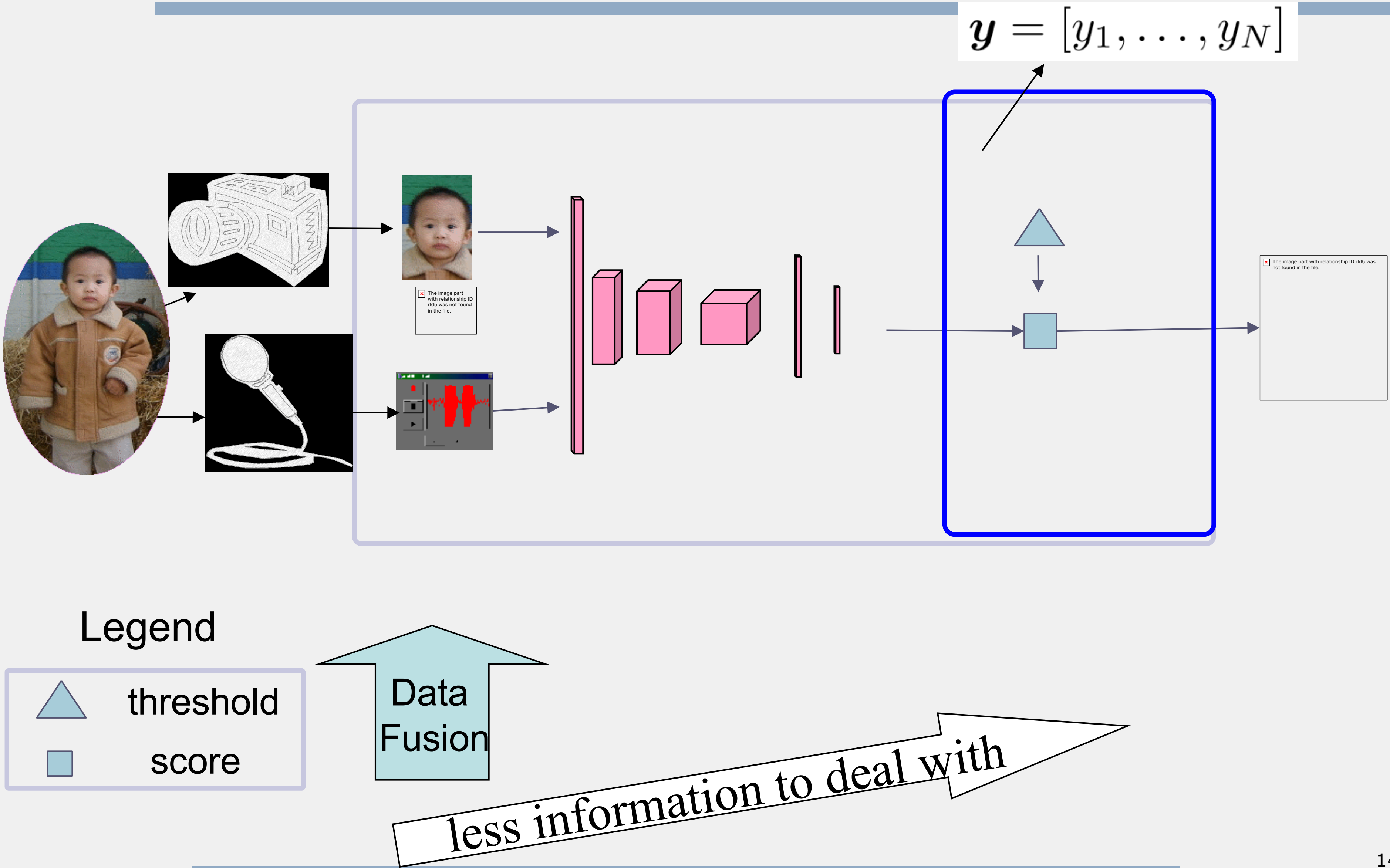
- Should improve performance



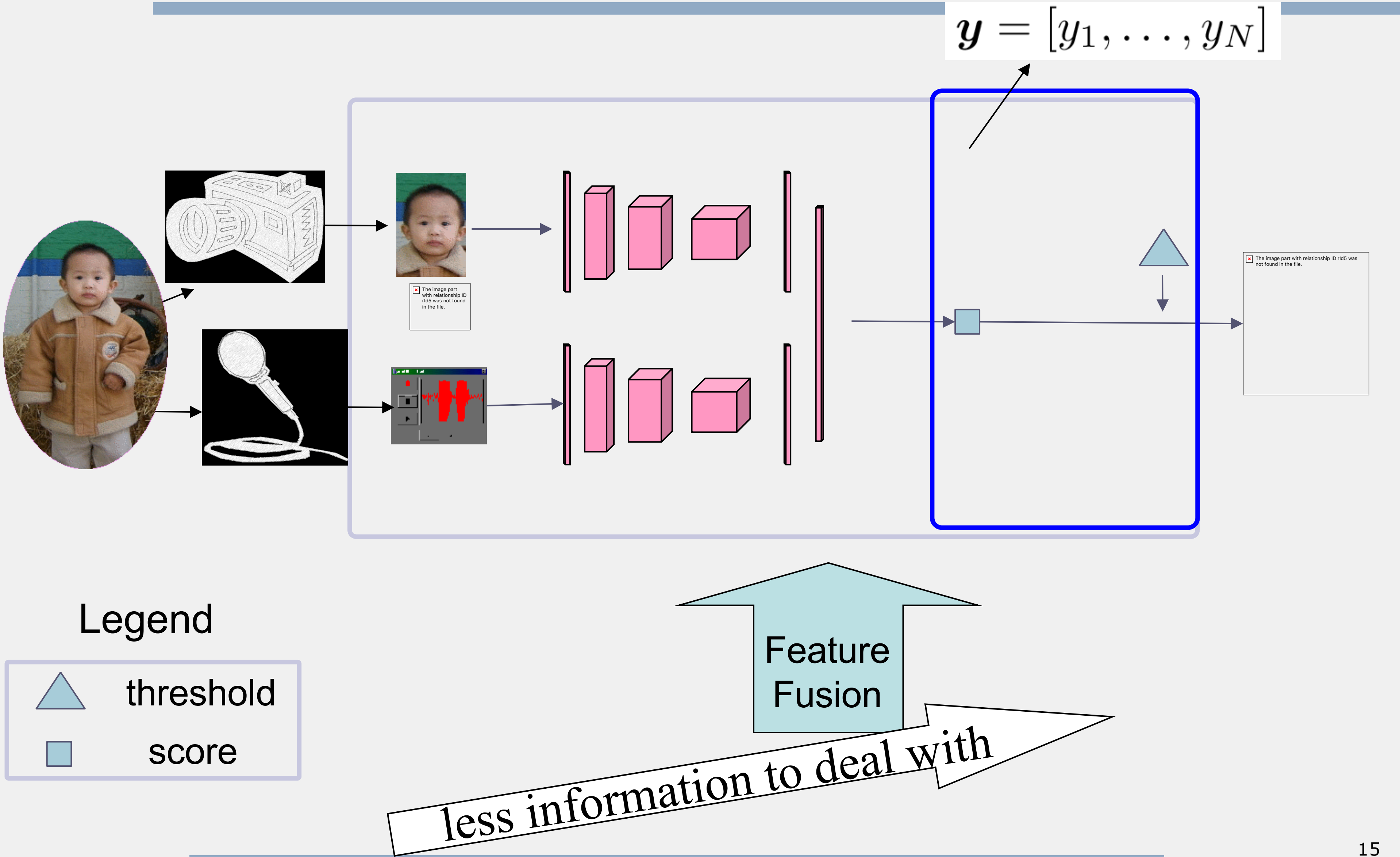
Levels of Fusion



Data level fusion

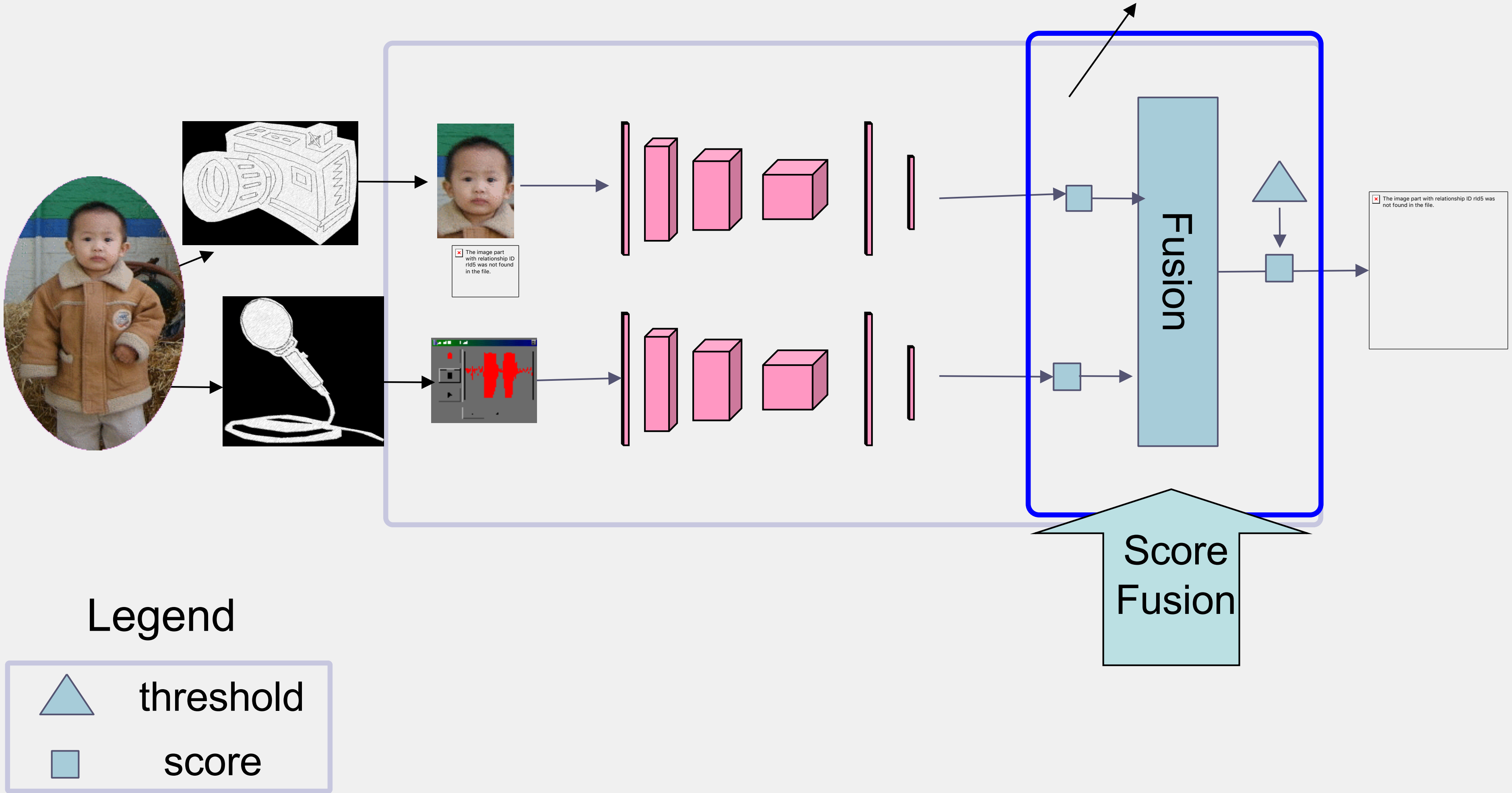


Feature level fusion



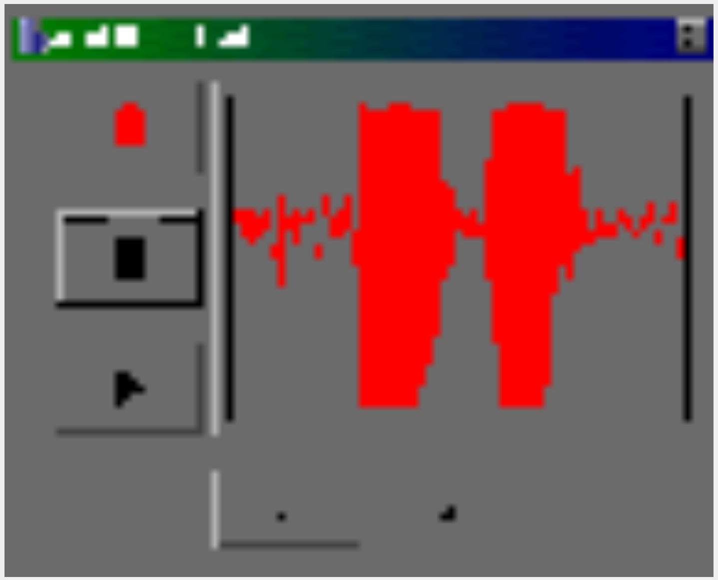
Score level fusion

$$\boldsymbol{y} = [y_1, \ldots, y_N]$$



Biometric system

Pattern representation



Pattern recognition problem

N – number of classes

b - biometric trait

x - feature vector

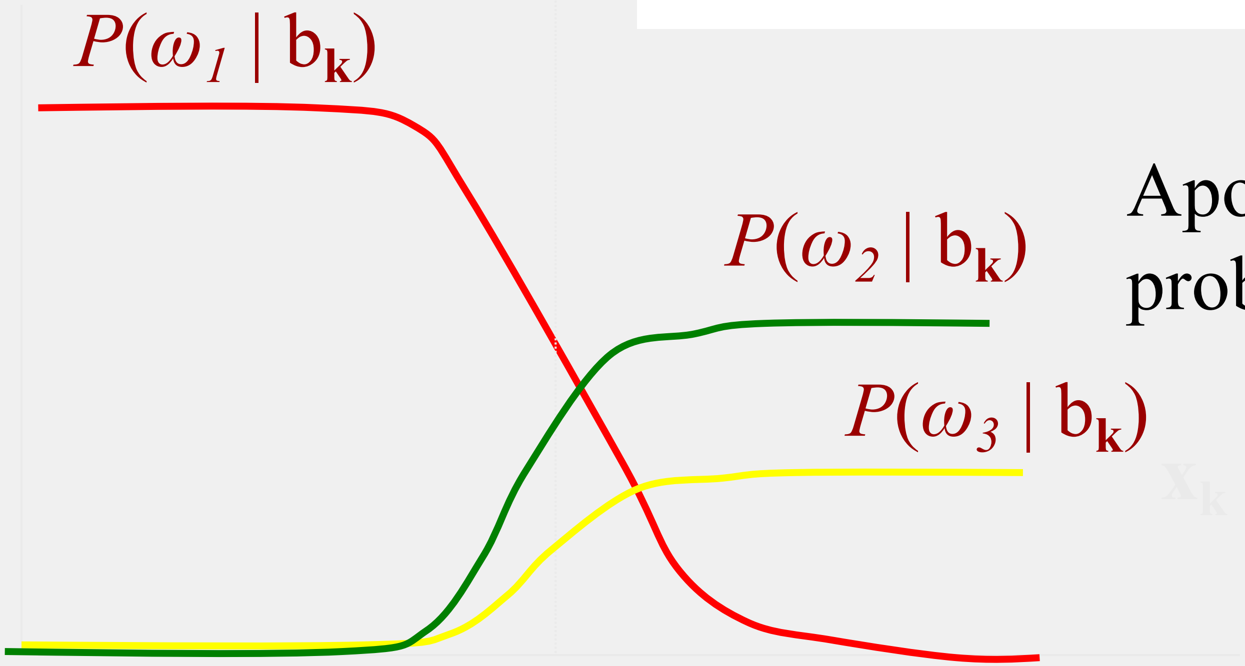
$P(\theta)$ -priori probability of class θ

$p(x_k|\theta)$, $p(b_k|\theta)$ -measurement distri-butions of patterns in class θ

Bayesian decision making

Bayes minimum Error rule

$$b_k \rightarrow \omega \ if$$
$$P(\omega|b_k) = max_\theta P(\theta|b_k)$$



Aposteriori class probabilities

Problem formulation

- Given

$$\begin{array}{ll} \text{biometric traits:} & [b_1, ...b_K] \\ \text{biometric features:} & [x_1, ...x_K] \\ \text{identities:} & [\theta_1, ..., \theta_R] \end{array}$$

- Bayes decision rule
  - *Assign* subject *to class* θ *if*

  $P(\omega| b_1,\ldots, b_K) = \max P(\theta | b_1,\ldots, b_K)$

- Note

$$P(\omega|b_1, ..., b_K) \propto \frac{p(b_1, ..., b_K|\omega)P(\omega)}{normalisation\ factor}$$

Fusion options

- **Signal level fusion**

$$
\begin{aligned}
p(b_1, ..., b_K|\omega) &\propto \int_x p(x, b_1, ..., b_K, \omega) \propto \\
&\propto \int_x P(\omega|x) p(x|b_1, ..., b_K) \\
&\propto \hat{P}(\omega|x)
\end{aligned}
$$

- The integration over x is marginalisation over the distribution $p(x|b_1, ..., b_K)$
  - x is a feature vector determined by all traits
  - Implicitly a multiple classifier fusion
    - Bagging, boosting, drop out, hard sample mining
  - Marginalised estimate of class posterior $\hat{P}(\omega|x)$

Fusion options

- Feature level fusion

$$
\begin{aligned}
p(b_1,..,b_K|\omega) \quad & \propto \int_{x_1,..,x_K} p(x_1,..,x_K,b_1,..,b_K,\omega) \propto \\
& \propto \int_{x_1,..,x_K} P(\omega|x_1,..,x_K)p(x_1,..,x_K|b_1,..,b_K) \\
& \propto \hat{P}(\omega|x_1,...,x_K)
\end{aligned}
$$

- Each modality has its own set of features x_i
- Score is a function of all x_i jointly
- Fusion process marginalisation is over the joint distribution of all modalities
- In addition, there could be modality specific marginalisation at the feature extraction level

Fusion options

- Score level fusion

$$\begin{aligned} p(b_1, ..., b_K|\omega) &\propto \prod_i \int_{x_i} p(x_i, b_i, \omega) \propto \\ &\propto \prod_i \int_{x_i} P(\omega|x_i)p(x_i|b_i) \\ &\propto \prod_i \hat{P}(\omega|x_i) \end{aligned}$$

- Each modality has its own set of features x_i
- The fused score is a product of individual modality specific scores
- Fusion process marginalisation is over modality specific distributions

Problem formulation: comments

- basic score level fusion is by product
- product can be approximated by a sum if $\hat{P}(\theta|x_k)$ does not deviate much from $P(\theta)$
  i.e. $\hat{P}(\theta|x_k) = P(\theta) + \Delta_k$
- the resulting decision rule becomes

$$p(b_1, ..., b_K|\omega) \quad \propto \prod_k \hat{P}(\omega|x_k)$$
$$\propto \sum_k \hat{P}(\theta|x_k)$$

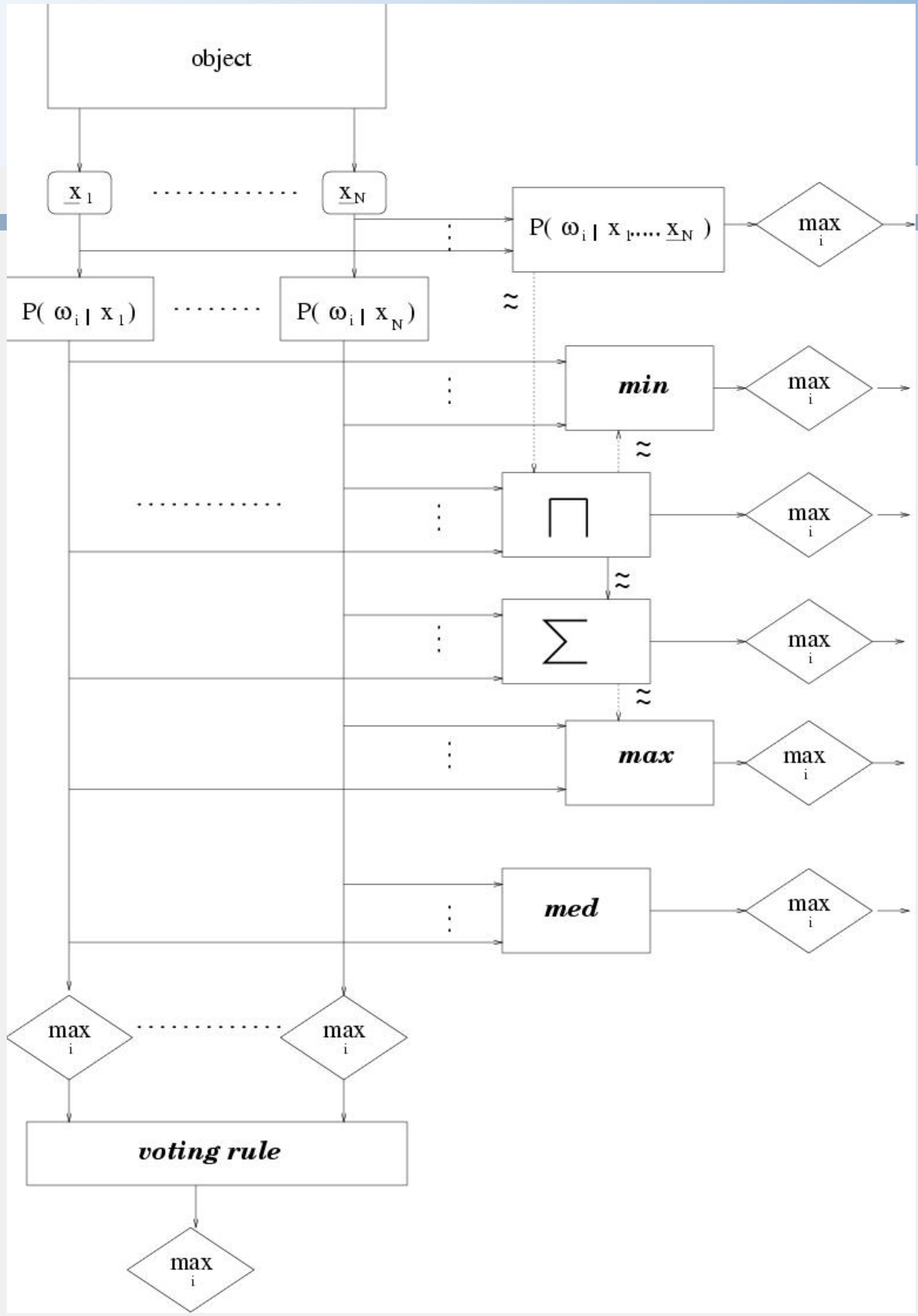
Fusion options

- Decision level fusion
  - Builds on score level fusion
  - Different fusion rules (rank, vote, ect)
- Example: Vote fusion
  - Each modality produces a hard decision

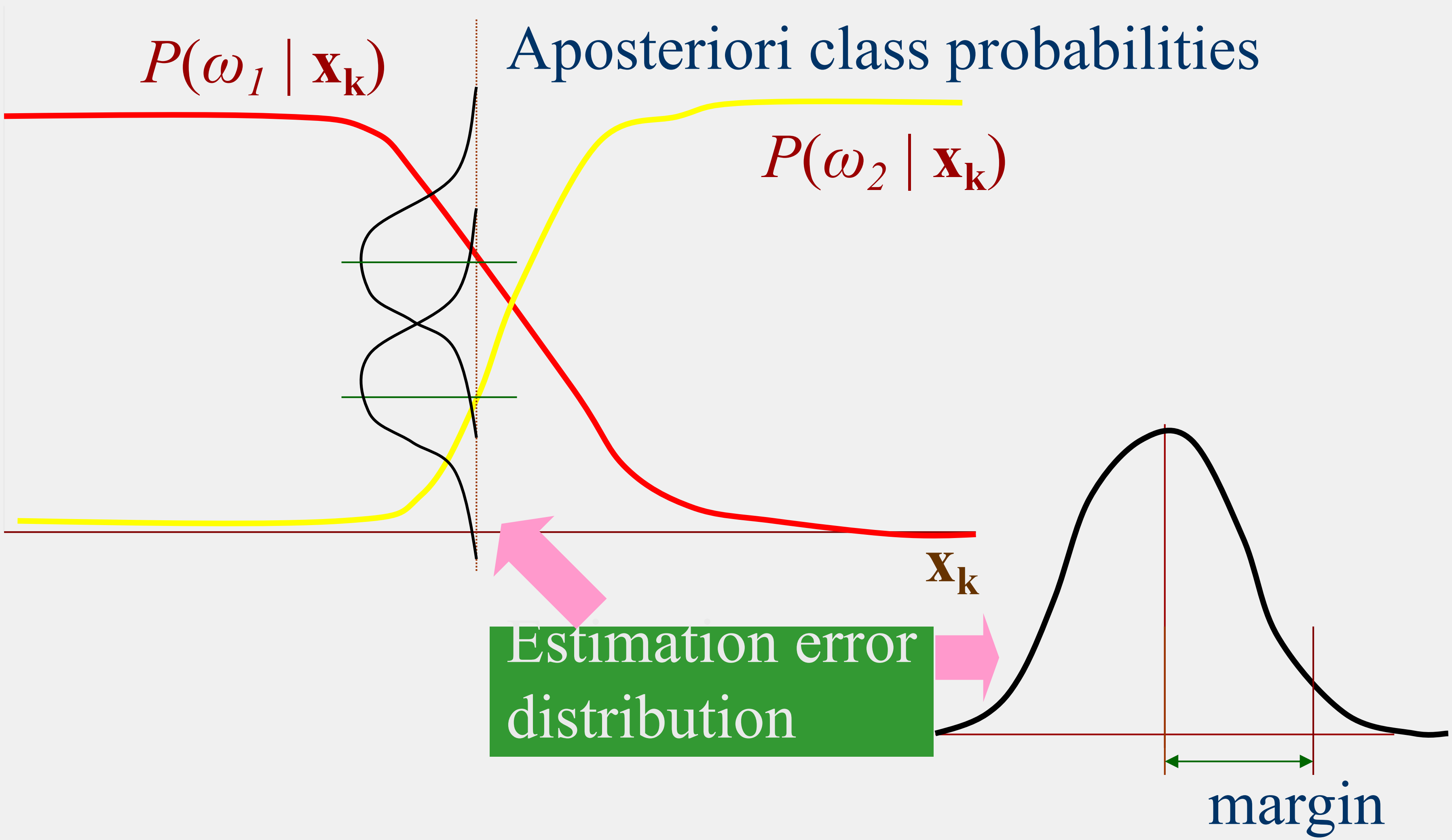
$$b_i \rightarrow \theta(i) \;\; if \;\; \theta(i) = \max \arg_\gamma \hat{P}(\gamma | x_i)$$

  - n_θ - the count of modalities outputting θ
  - Final decision $[b_1, \ldots, b_K] \rightarrow \omega \;\; if \;\; n_\omega = \max_\theta n_\theta$
  - In a two class case, a hard decision is made by comparing the score against a threshold

Fixed fusion strategies



Effect of estimation errors

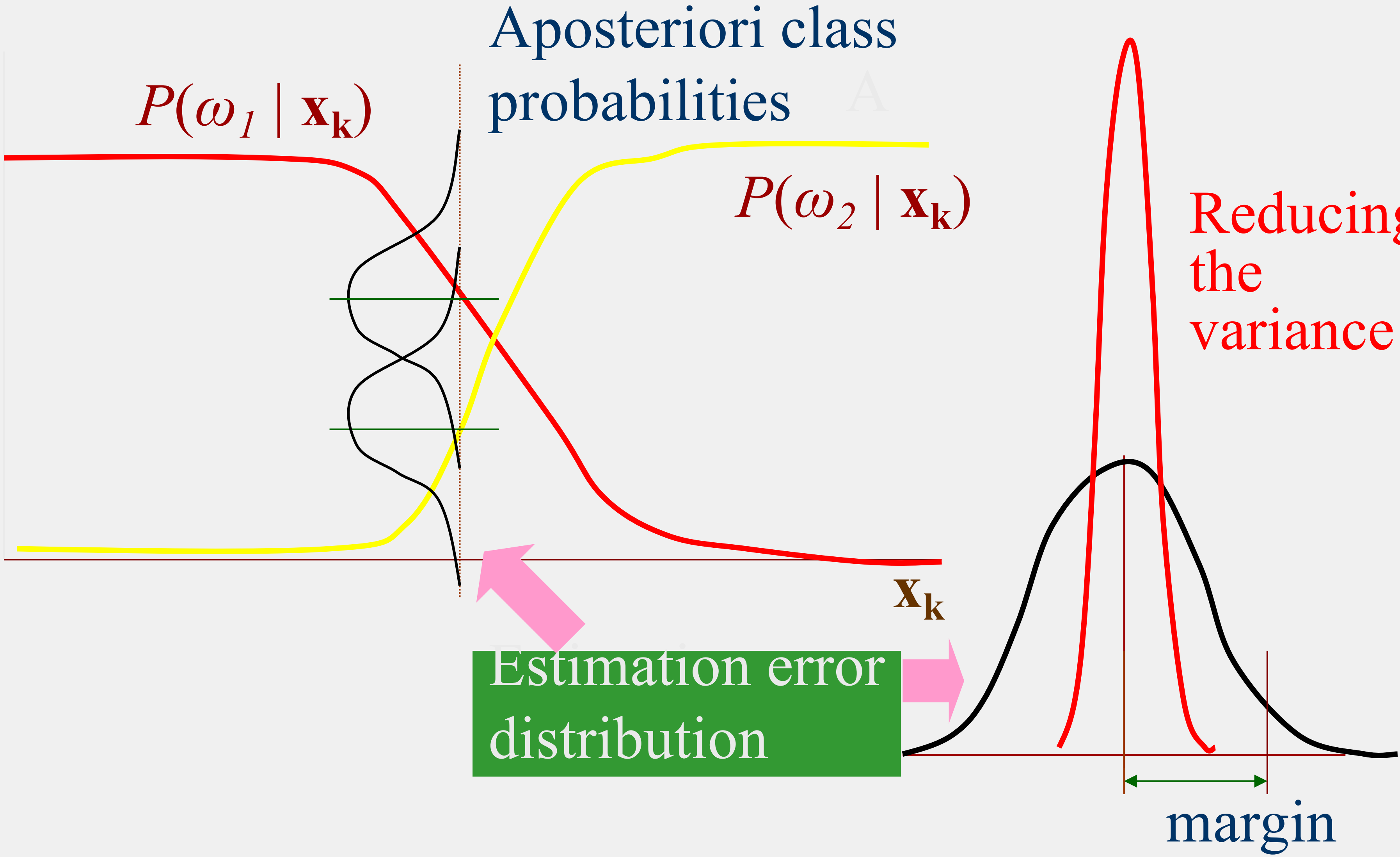


Sources of estimation errors

$$\tilde{P}(\omega|\mathbf{x}_i) = \int\int P(\omega|\mathbf{x}_i, X_i, M, \gamma_i)p(M)p(\gamma_i)dMd\gamma_i$$

$\mathbf{x}_i$	Feature vector output by sensor i
X_i	Training set for the i-th expert
M	Classifier model
$p(M)$	Distribution of models
γ_i	Parameters for expert i
$p(\gamma_i)$	Distribution of expert i parameter

Coping with estimation errors



Variance reduction

- Consider a vector of normalised scores

$$\hat{\mathrm{s}} = [\hat{s}_1, \hat{s}_2, ...\hat{s}_R]^T$$

- with mean

$$\mu_k = [\mu_{k1}, \mu_{k2}, ..., \mu_{kR}]^T, \; k = c, i$$

- and covariance matrix

$$\Sigma_k = \begin{bmatrix} \sigma_{k11} & \cdot & \cdot & \cdot & \sigma_{k1R} \\ \sigma_{k21} & \sigma_{k22} & \cdot & \cdot & \sigma_{k2R} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \sigma_{kR1} & \sigma_{kR2} & \cdot & \cdot & \sigma_{kRR} \end{bmatrix}$$

Variance reduction

- Fuse scores by $\widehat{S} = \frac{1}{R}\sum_{j=1}^{R} \widehat{s}_j$
- Average class conditional variance

$$\bar{\sigma}_k = \frac{1}{R}\sum_{j=1}^{R} \sigma_{kjj}$$

- Variance of fused score

$$\widehat{\sigma}_k = E\{(\widehat{S} - \widehat{\mu}_k)^2\} = E\{[\frac{1}{R}\sum_{j=1}^{R}(\widehat{s}_{kj} - \mu_{kj})]^2\}$$

Variance reduction

- Rearranging

$$\hat{\sigma}_k = E\{\frac{1}{R^2}[\sum_{j=1}^{R}(\hat{s}_{kj}-\mu_{kj})^2+2\sum_{j=1}^{R}\sum_{i>j}^{R}(\hat{s}_{kj}-\mu_{kj})(\hat{s}_{ki}-\mu_{ki})]\}$$

- Variance can be bounded

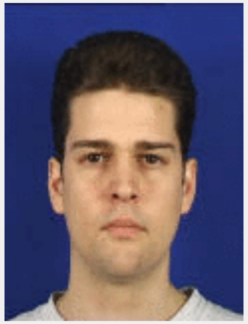
$$\frac{1}{R}\bar{\sigma}_k \leq \hat{\sigma}_k \leq \bar{\sigma}_k \qquad 0 \leq \hat{\sigma}_k \leq \bar{\sigma}_k$$

  - For uncorrelated scores - variance reduces by a factor of R
  - For negatively correlated scores – variance can be brought to zero
  - For negatively correlated scores the variance drops most when

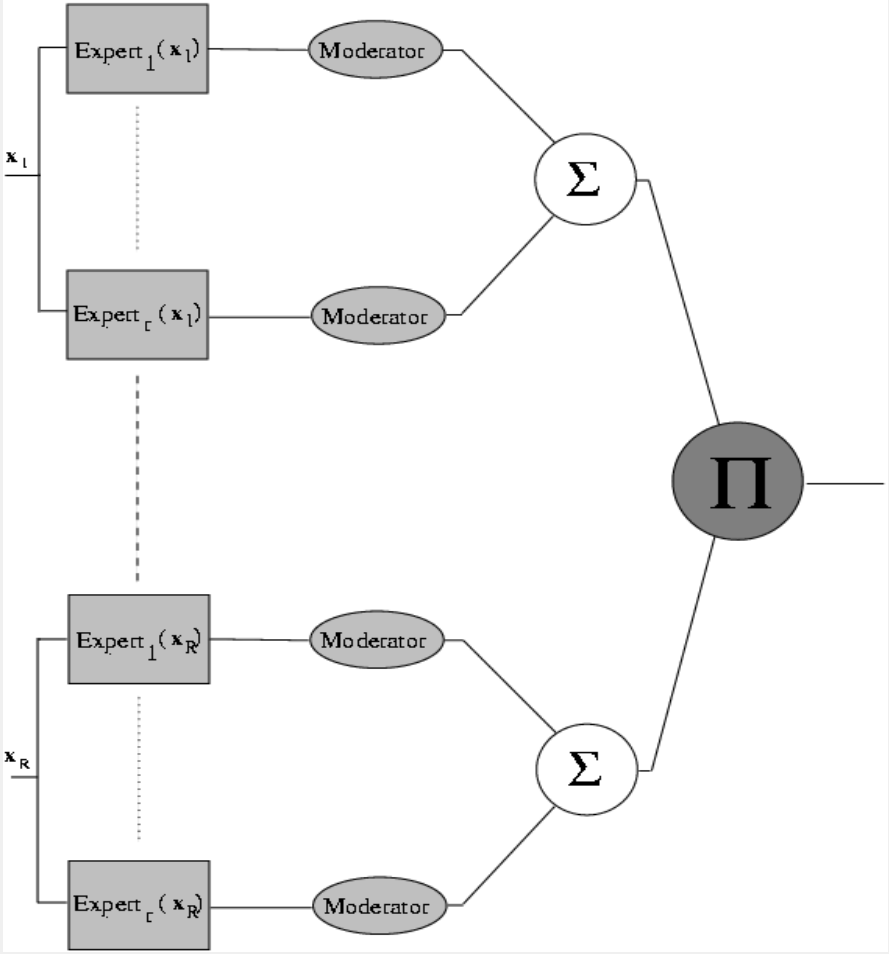
$$\sigma_{ii} = \sigma_{jj} \quad \forall j$$

Biometric Personal Identity Authentication

FACE



VOICE



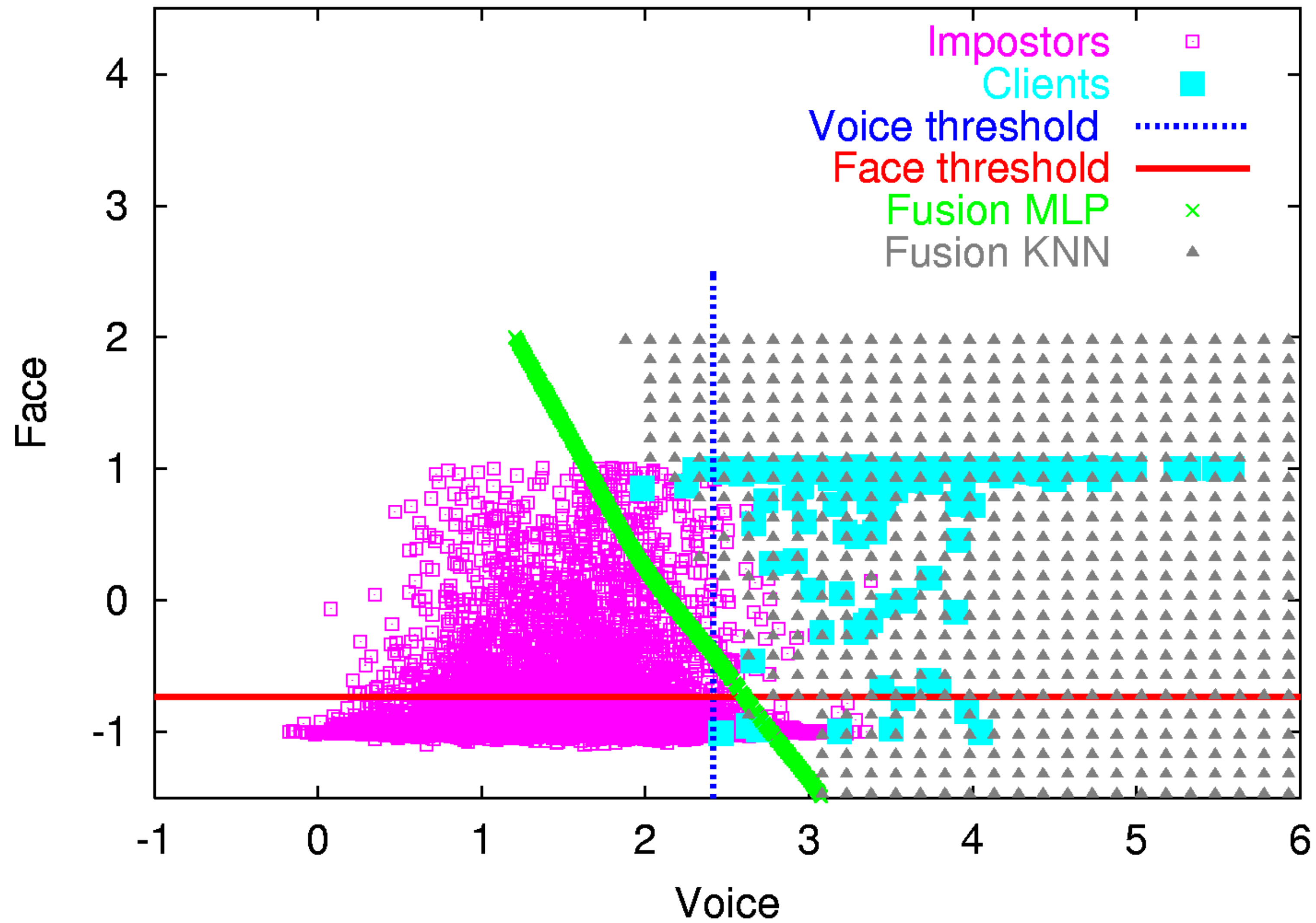
Fusion of face and voice

Performance of individual and fused experts

Toy example

Modalities	Performance		
	FAR	**FRR**	HTER
Face	**1.75**	**2.00**	1.88
Voice	**1.47**	**1.00**	1.23
Fusion SVM	**0.32**	**0.25**	0.28
Fusion MLP	**0.34**	**0.25**	0.29

Merits of multimodal fusion



Fusion strategies

- simple rules (sum, product, max, min, rank)
- trained fusion rule (logistic regression, decision templates, sparse based representation, svm, deep architectures)
- multistage systems (stacking)
- machine learning tools
  - Separability measures
  - Feature selection
  - Clustering
  - Distance metric
  - Classification

Direct score fusion: score normalisation

- Aposteriori class probabilities are automatically normalised to [0,1]
- Some systems compute a matching score s_i, rather than $P(\omega_i|\mathbf{x})$
- Scores have to be normalised to facilitate fusion by simple rules
  - aposteriori probability estimate

$$P(\omega_i|s) = \frac{p(s|\omega_i)P(\omega_i)}{\sum_{k=1}^{R} p(s|\omega_k)P(\omega_k)}$$

Score normalisation (cont)

- Motivation for score normalisation
  - Non-homogeneous scores (distance, similarity)
  - Different ranges
  - Different distributions
- Desirable properties
  - Robustness
  - Efficiency
- Most effective methods
  - Nonlinear mapping with saturation for very large/small scores
  - Increased sensitivity near the boundaries (Ross and Jain)

Score normalisation (cont)

- Min-max

$$\hat{s} = \frac{s - \min s}{\max s - \min s}$$

- Scaling

$$\hat{s} = \frac{s}{\max s}$$

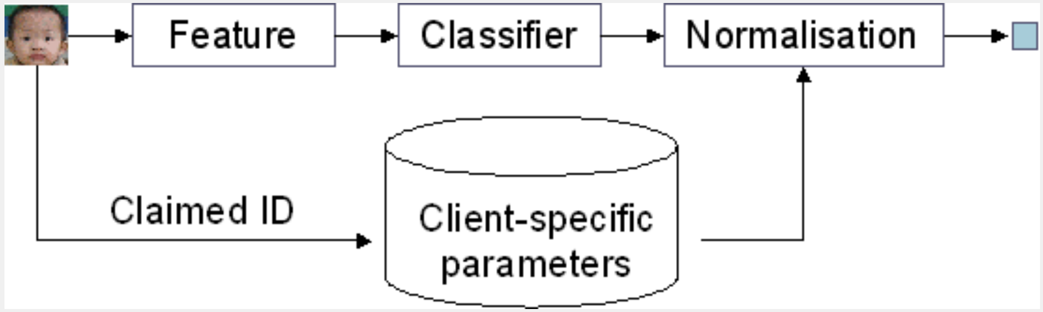
- Z-score

$$\hat{s} = \frac{s - \mu}{\sigma}$$

μ and σ are the mean and standard deviation of the score distribution,

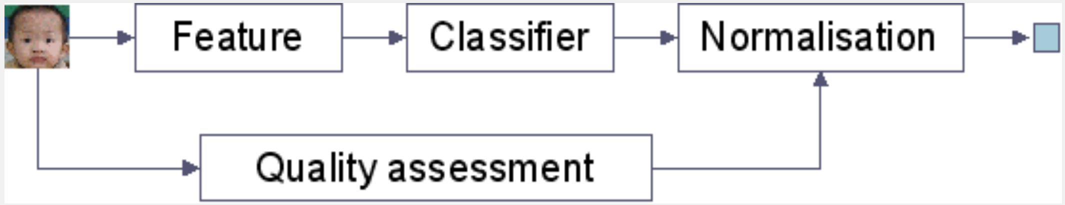
Information sources

Client/user-specific normalization (offline)



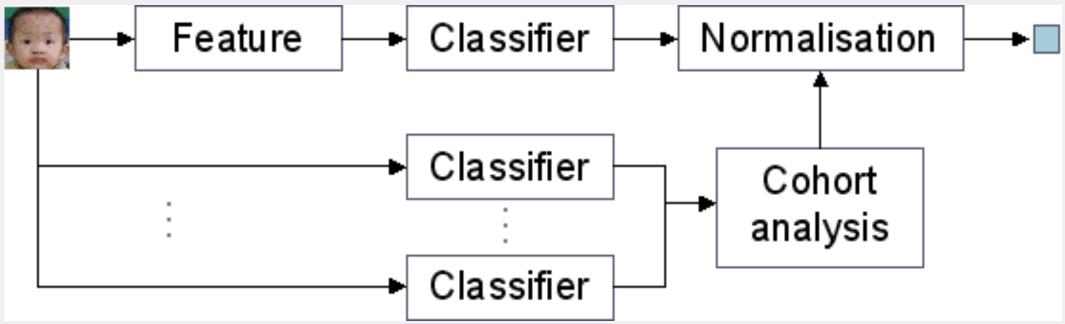
User-dependent score characteristics

Quality-based normalization



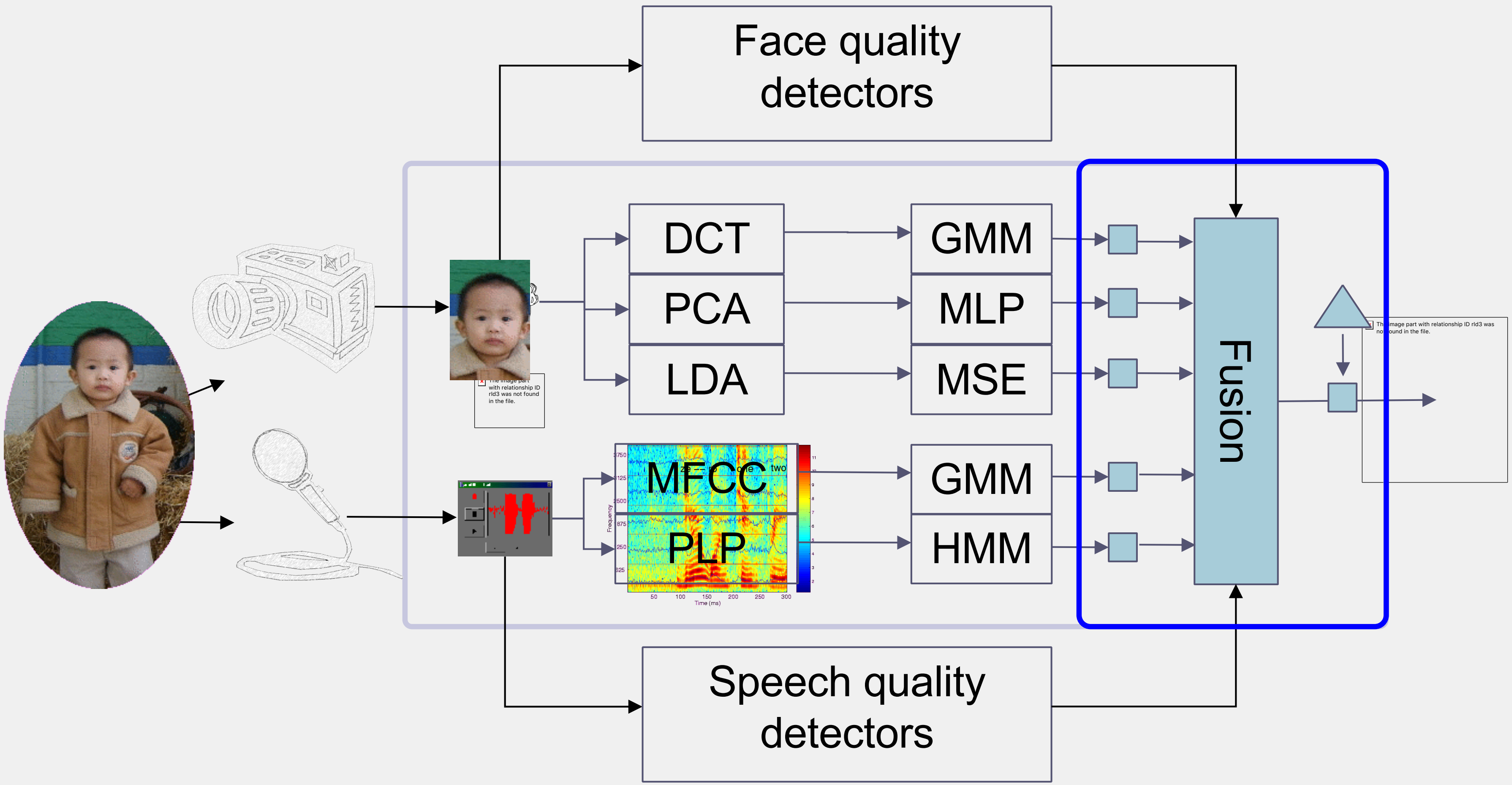
Changing signal quality

Cohort-based normalization (online)



Changing signal quality

Confidence-based Fusion Algorithms



Generative & Discriminative Approaches in QDF

Generative	$y_{com}^{llr} \equiv \boxed{f^{llr}(x,q)} = \log \frac{p(x,q\|C)}{p(x,q\|I)}$	e.g. GMM
Discriminative (probability-based)	$y_{com}^{prob} \equiv \boxed{f^{prob}(x,q)} = P(C\|x,q)$	e.g. MLP logistic regression
Discriminative (function-based)	$P(C\|x,q) \approx \mathrm{sigmoid}\big(f^{disc}(x,q)\big) = \frac{1}{1+\exp\left(-\boxed{f^{disc}(x,q)}\right)}$	e.g. SVM, MLP

Algorithm used in experiments
x and *q* are vectors

Case study in multimodal soft biometric fusion

- Multimodal biometric traits
- Multimodal sensing of the same biometric trait
  - Different spectral bands
  - Voice/image sensed lips dynamics
  - Visual/language modalities for person re-identification

Background and motivation

- Video surveillance very important tool for crime prevention and detection
  - Watch list
  - Forensic video analysis
- Hard biometrics (face) not always available
- Other video analytics tools are useful alternatives
  - Soft biometrics (clothing, gait)
  - Tracking

Soft biometrics and re-identification

- Person Re-Identification
  - Recognising a person from non-overlapping cameras
- Formulated as a ranking problem

Re-ID with V&L

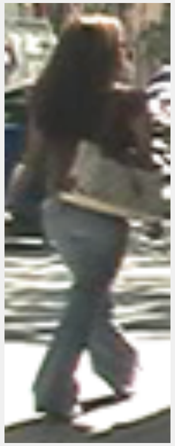
- The majority of existing methods are vision only
  - Images or videos
- Joint vision and language modelling
  - Image and video captioning, Visual question answering, Image synthesis from language, …
- Can language help vision in Re-ID?

Language annotation

- Augmenting existing datasets
  - CUHK03: ~2700 descriptions
  - VIPeR: ~1300 descriptions
- Crowd-sourced, 8 annotators
- Annotation
  - Free style sentences, not attributes
  - Encouraged to cover details
  - On average 45 words per description
  - Per image rather than per identity

Language annotation

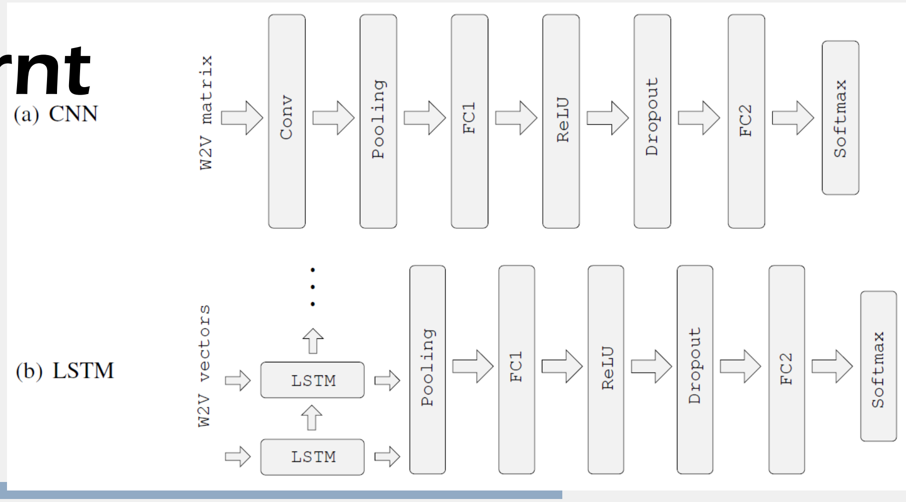
A front profile of a young, slim and average height, black female with long brown hair. She wears sunglasses and possibly earrings and necklace. She wears a brown t-shirt with a golden coloured print on its chest, blue jeans and white sports shoes.



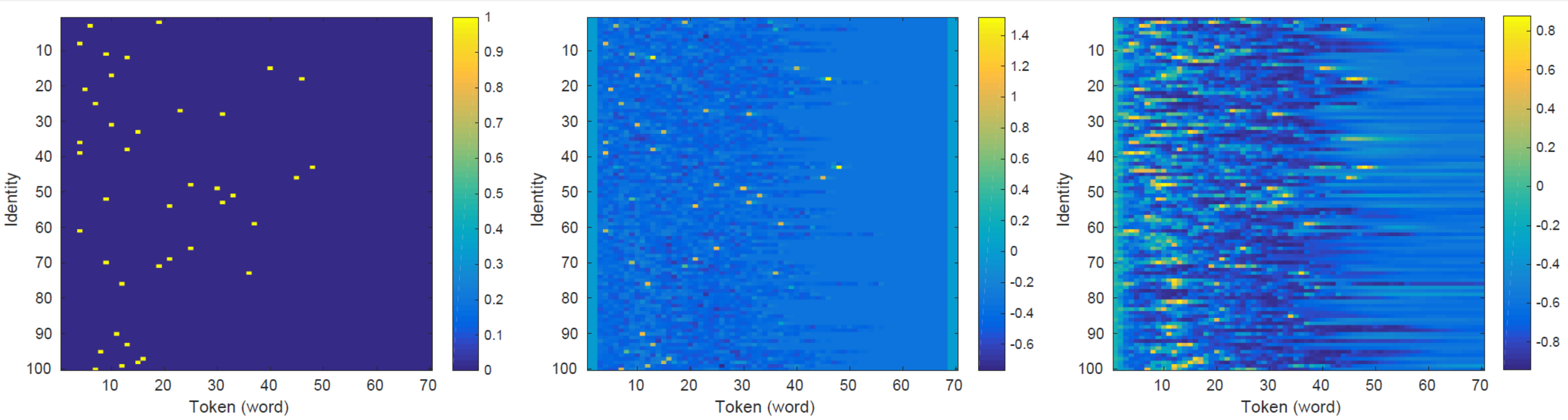
A short and slim young woman carrying a tortilla coloured rectangular shoulder bag with caramel straps, on her right side. She has a light complexion and long, straight auburn hair worn loose. She wears a dark brown short sleeved top along with bell bottomed ice blue jeans and her shoes can't be seen but she might be wearing light colored flat shoes.

Re-ID with language

- ResNet-50 for visual information
- Word2Vec embedding
- Neural networks: CNN and LSTM
- Multi-class setting, 2 examples per class (identity)
- Data augmentation
- Metric learning with learnt representations (XQDA)
- Canonical Correlation



Re-ID with language



- Detecting the concept of "spectacles"
  - "bespectacled", "glasses", "eye-glasses", ...
  - GT, CNN, LSTM
  - One channel becomes "spectacles" detector during training
  - Good representation learnt from unstructured data

Canonical correlation analysis

- Consider features x and y extracted from two biometric modalities
- Basic principle – find direction in the respective feature spaces that yield maximum correlation
  - Gauging linear relationship between two multidimensional random variables (feature vectors of two biometric modalities)
  - Finding two sets of basis vectors such that the projection of the feature vectors onto these bases is maximised
  - Determine correlation coefficients

CAA problem formulation

- Training set of pairs of vectors $(x_i, y_i), \ \ i = 1, n$
- Maximisation of the correlation of the projections

$$\begin{aligned} \max_{w_x, w_y} E\{w_x^T x y^T w_y\} &= \max_{w_x, w_y} w_x^T C_{xy} w_y \ \ s.t. \\ E\{w_x^T x x^T w_x\} &= w_x^T C_{xx} w_x = 1 \\ E\{w_y^T y y^T w_y\} &= w_y^T C_{yy} w_y = 1 \end{aligned}$$

- Leads to an eigenvalue problem

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} =$$

$$= \lambda \begin{bmatrix} (1-\kappa)C_{xx} + \kappa I & 0 \\ 0 & (1-\kappa)C_{yy} + \kappa I \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix}$$

- With cov matrices regularised by κI

Re-ID with V&L

- Three sets:
  - Training, query, gallery
  - Training: image and language pairs
- Various settings, query x gallery:
  - V x V, L x L, V x L, V x VL, VL x VL
- Asymmetric settings:
  - Transfer language info. With CCA
- XQDA as metric learning

UNIVERSITY OF SURREY

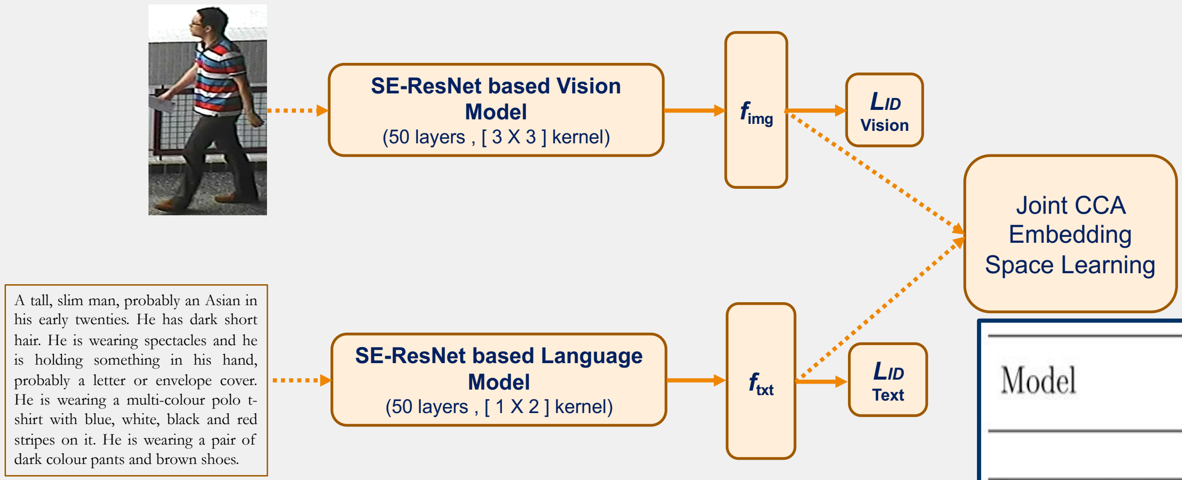
Re-ID with V&L

	Gated CNN	68.1	88.1	94.6
Ours	V x V	70.3	93.2	96.6
	L x L	41.1	69.8	82.5
	V x L	17.7	48.5	66.0
	V x VL	73.5	94.5	97.7
	VL x VL	**81.8**	**98.1**	**99.3**

- Results on CUHK03, R1, R5, R10
- LxL worse than VxV: more information in vision
- VxVL 3.2 points higher than VxV
- VLxVL 12.7 points higher than VxV, better than state-of-the-art
- Language helps

Person Re-ID

- Crossmodal & multimodal matching facilitated by CAA



- Performance gain due to
  - Joint training
  - Fusion of modalities

Model	Rank@1 (%)	Rank@5 (%)	Rank@10 (%)	mAP (%)	medR
V × V					
Separately Train	59.91	80.5	85.7	64.45	1
Jointly Train + CCA	82.05	94.3	96.8	84.75	1
L × V					
Separately Train	13.6	32.99	43.04	18.5	15
Jointly Train + CCA	27.9	50.6	60.7	33.4	5
VL × V					
Separately Train	65.87	84.19	88.9	64.8	1
Jointly Train + CCA	84.7	95.0	97.1	84.1	1
VL × VL					
Separately Train	68.0	84.7	89.58	71.8	1
Jointly Train + CCA	80.86	94.16	96.6	83.85	1

Take home message

- Role of multimodal biometrics
- Fusion levels
- Math formulation of different alternatives
- The concept of marginalisation/multiple classifier systems
- Notion of quality based, user specific and cohort based extensions of fusion
- Multimodal sensing and fusion of a single biometric
- Example: fusion of vision/language modalities for soft biometrics

References

- N. Poh, A. Martin, and S. Bengio. Performance Generalization in Biometric Authentication Using Joint User-Specic and Sample Bootstraps. IEEE Trans. Pattern Analysis and Machine Intelligence, 29(3):492-498, 2007.
- N. Poh and J. Kittler. Incorporating Variation of Model-specic Score Distribution in Speaker Verication Systems. IEEE Transactions on Audio, Speech and Language Processing, 16(3):594-606, 2008.
- N. Poh, T. Bourlai, J. Kittler and al. A Score-level Quality-dependent and Cost-sensitive Multimodal Biometric Test Bed. Pattern Recognition, 43(3):1094{1105, 2010.
- N. Poh, T. Bourlai, J. Kittler and al. Benchmarking Quality-dependent and Cost-sensitive Multimodal Biometric Fusion Algorithms. IEEE Trans. Information Forensics and Security,4(4):849{866, 2009.
- N. Poh, J. Kittler and T. Bourlai, Quality-based Score Normalisation with Device Qualitative Information for Multimodal Biometric Fusion", IEEE Trans. on Systems, Man, Cybernatics Part A : Systems and Humans, 40(3):539{554, 2010.
- Tresadern, P., et al., *Mobile Biometrics: Combined Face and Voice Verification for a Mobile Platform.* Pervasive Computing, IEEE, 2013. **12**(1): p. 79-87.
- Poh, N. and S. Bengio, *F-ratio Client-Dependent Normalisation on Biometric Authentication Tasks*, in *IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP)*2005: Philadelphia. p. 721-724.
- Poh, N. and J. Kittler, *Incorporating Variation of Model-specific Score Distribution in Speaker Verification Systems.* IEEE Transactions on Audio, Speech and Language Processing, 2008. **16**(3): p. 594-606.

- Poh, N., et al., *Group-specific Score Normalization for Biometric Systems*, in *IEEE Computer Society Workshop on Biometrics, CVPR2010*. p. 38-45.
- Poh, N. and M. Tistarelli. *Customizing biometric authentication systems via discriminative score calibration*. in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 2012. IEEE.
- Poh, N. and J. Kittler, *A Biometric Menagerie Index for Characterising Template/Model-specific Variation*, in *Proc. of the 3rd Int'l Conf. on Biometrics*2009: Sardinia. p. 816-827.
- Poh, N. and J. Kittler, *A Methodology for Separating Sheep from Goats for Controlled Enrollment and Multimodal Fusion*, in *Proc. of the 6th Biometrics Symposium*2008: Tampa. p. 17-22.
- Poh, N., et al., *A User-specific and Selective Multimodal Biometric Fusion Strategy by Ranking Subjects.* Pattern Recognition Journal, 46(12): 3341-57 , 2013:
- Poh, N., G. Heusch, and J. Kittler, *On Combination of Face Authentication Experts by a Mixture of Quality Dependent Fusion Classifiers*, in *LNCS 4472, Multiple Classifiers System (MCS)*2007: Prague. p. 344-356.
- Poh, N. and J. Kittler, *A Unified Framework for Multimodal Biometric Fusion Incorporating Quality Measures.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012. **34**(1): p. 3-18.

- Poh, N., A. Rattani, and F. Roli, *Critical analysis of adaptive biometric systems.* Biometrics, IET, 2012. **1**(4): p. 179-187.
- Merati, A., N. Poh, and J. Kitter, *Extracting Discriminative Information from Cohort Models*, in *IEEE 3rd Int'l Conf. on Biometrics: Theory, Applications, and Systems (BTAS)*2010. p. 1-6.
- Poh, N., A. Merati, and J. Kitter, *Making Better Biometric Decisions with Quality and Cohort Information: A Case Study in Fingerprint Verification*, in *Proc. 17th European Signal Processing Conf. (Eusipco)*2009: Glasgow. p. 70-74.
- Merati, A., N. Poh, and J. Kittler, *User-Specific Cohort Selection and Score Normalization for Biometric Systems.* Information Forensics and Security, IEEE Transactions on, 2012. **7**(4): p. 1270-1277.
- Poh, N., A. Merati, and J. Kittler. *Heterogeneous Information Fusion: A Novel Fusion Paradigm for Biometric Systems*. in *International Joint Conference on Biometrics*. 2011.
- Poh, N., A. Martin, and S. Bengio, *Performance Generalization in Biometric Authentication Using Joint User-Specific and Sample Bootstraps.* IEEE Trans. on Pattern Analysis and Machine, 2007. **29**(3): p. 492-498.
- Poh, N. and J. Kittler, *A Method for Estimating Authentication Performance Over Time, with Applications to Face Biometrics*, in *12th IAPR Iberoamerican Congress on Pattern Recognition (CIARP)*2007. p. 360-369.

- Poh, N. and S. Bengio, *Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments?*, in *LNCS 3869, 2nd Joint AMI/PASCAL/IM2/M4 Workshop on Multimodal Interaction and Related Machine Learning Algorithms MLMI*2005: Edinburgh. p. 87-100.
- Poh, N., et al., *Benchmarking Quality-dependent and Cost-sensitive Score-level Multimodal Biometric Fusion Algorithms.* IEEE Trans. on Information Forensics and Security, 2009. **4**(4): p. 849-866.
- Poh, N., et al., *An Evaluation of Video-to-video Face Verification.* IEEE Trans. on Information Forensics and Security, 2010. **5**(4): p. 781-801.
- M. Tistarelli, Y. Sun, and N. Poh, On the Use of Discriminative Cohort Score Normalization for Unconstrained Face Recognition, IEEE Trans. on Information Forensics and Security, 2014.