Fusion of Multimodal Biometrics

Josef Kittler

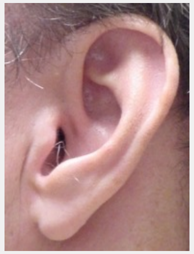
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford GU2 7XH

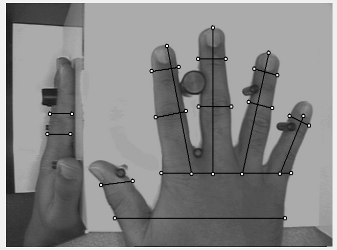
J.Kittler@surrey.ac.uk

Acknowledgements: Dr Norman Poh

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 enrol 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
  - To enable seamless switching/fusion of different biometrics in dynamic acquisition scenarios

OUTLINE

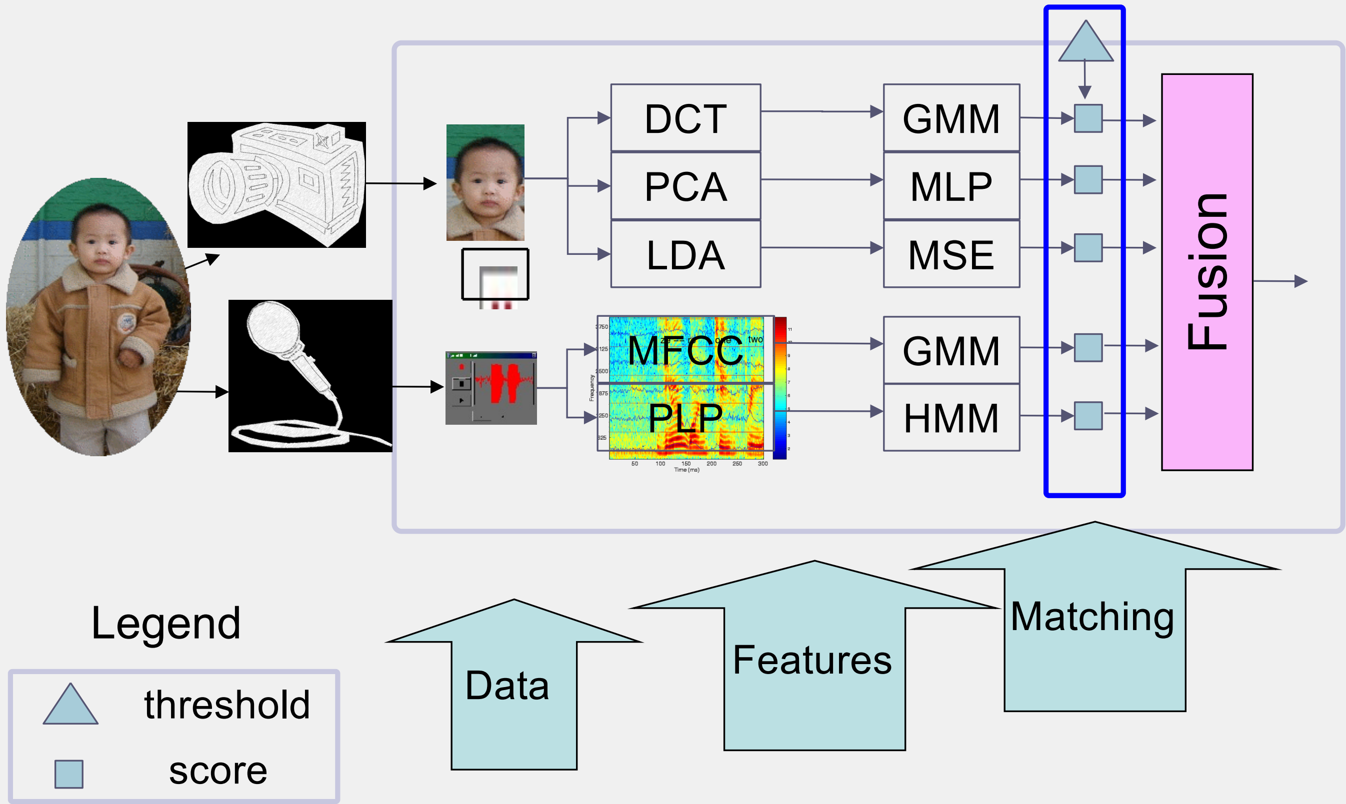
- Fusion architectures
- Problem formulation
- Estimation error
- Case study: Multimodal and cross-modal person re-identification
- Conclusions

The aim: *To discuss the purpose of multimodal biometrics fusion, and to introduce basic fusion architectures and underlying mathematical models*

Fusion architectures

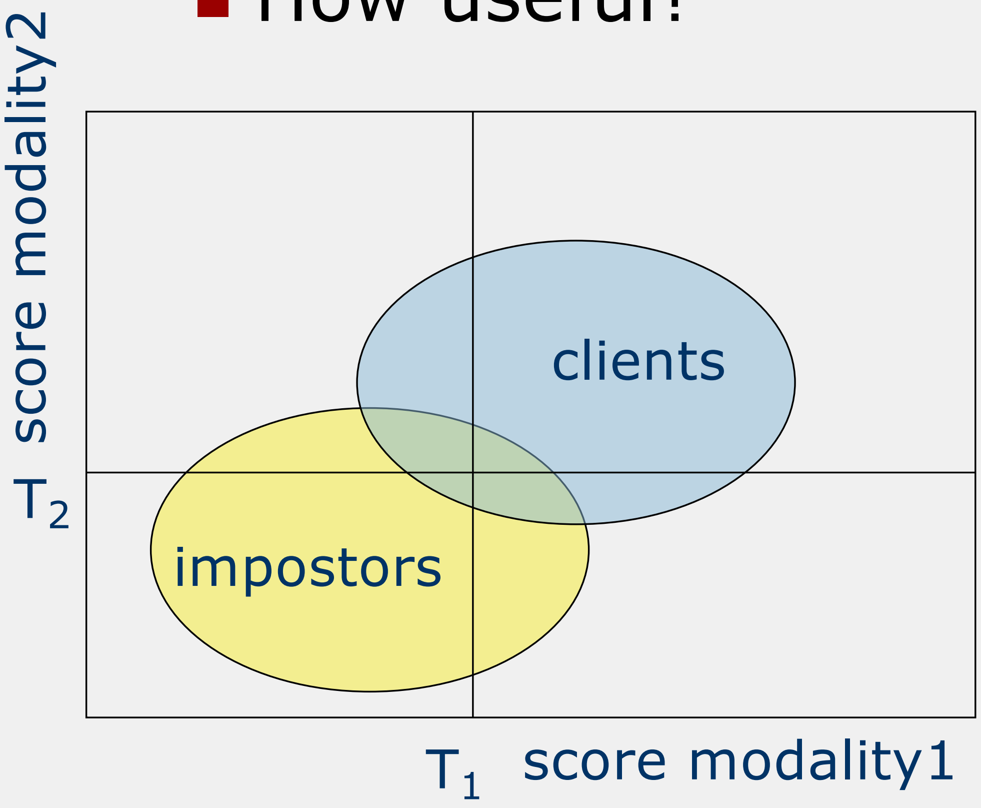
- Integration of multiple biometric modalities
- Sensor (data) level fusion
  - Linear/nonlinear combination of registered variables
  - Representation space augmentation
- Feature level fusion
- Soft decision level fusion
- Decision level fusion

Decision level fusion



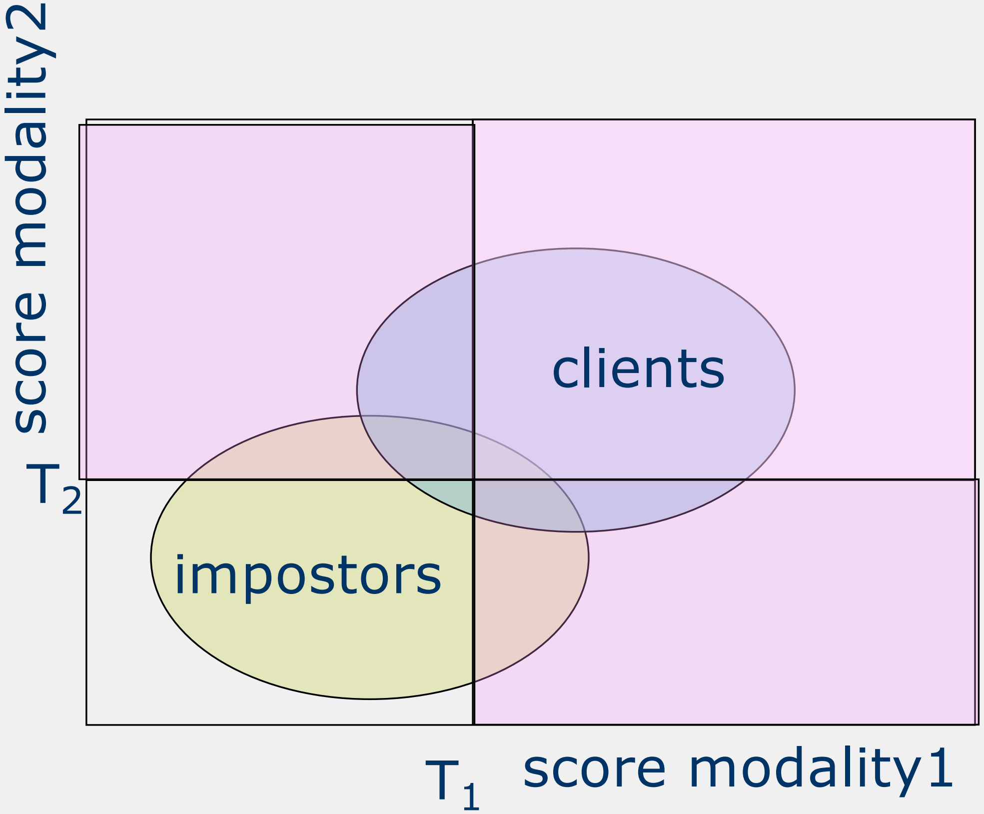
Decision-level fusion

- How useful?



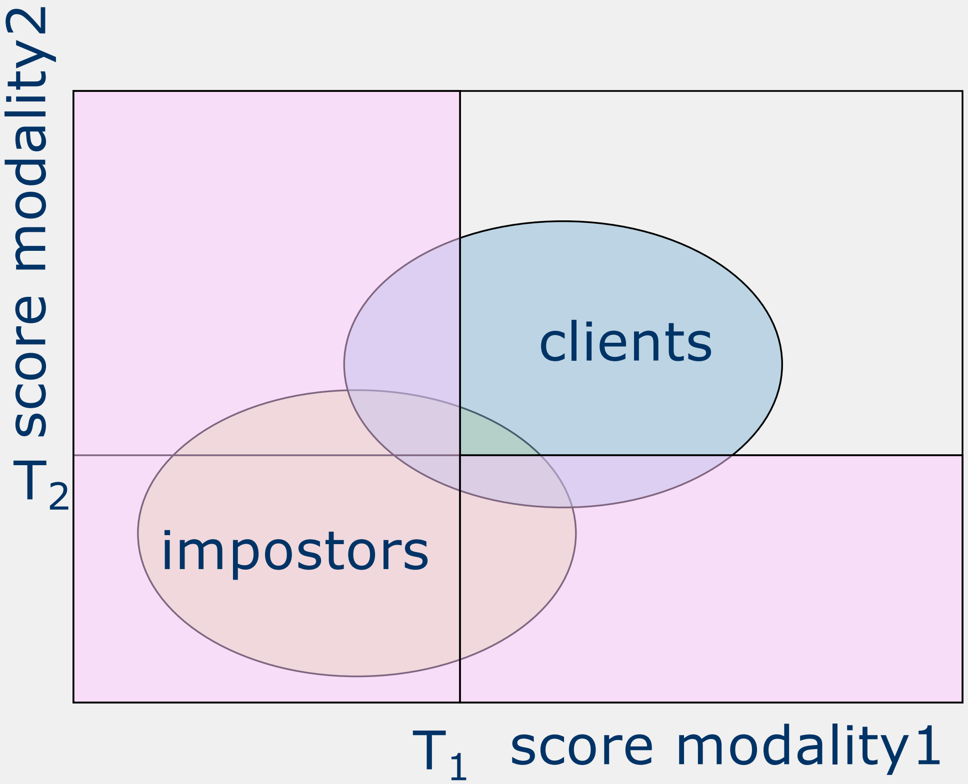
Decision-level fusion

- Accepted by either modality



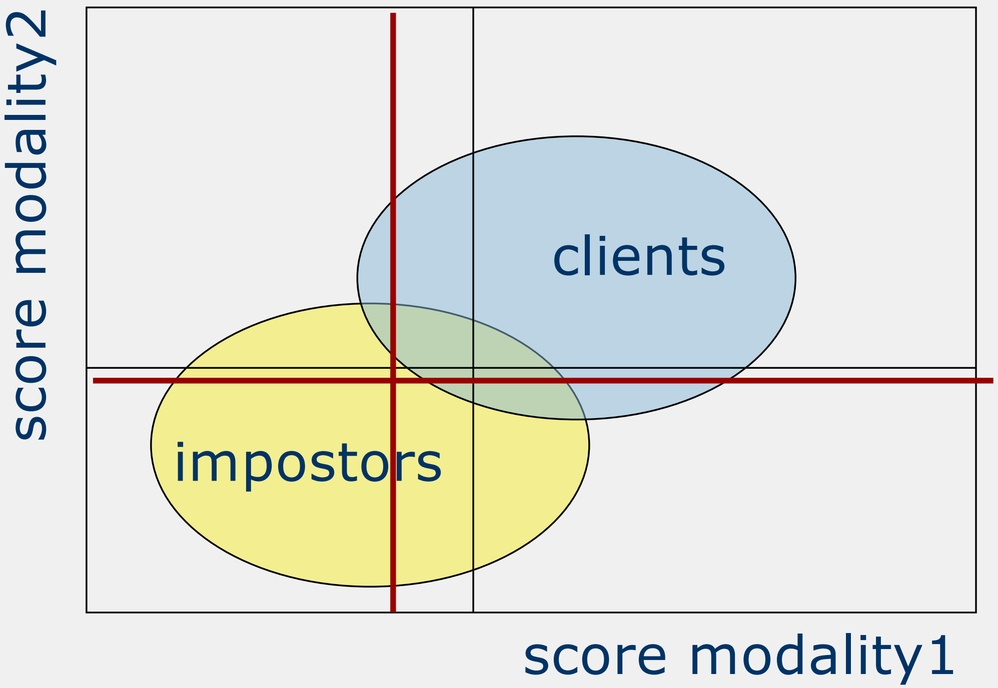
Decision-level fusion

- Accepted by both



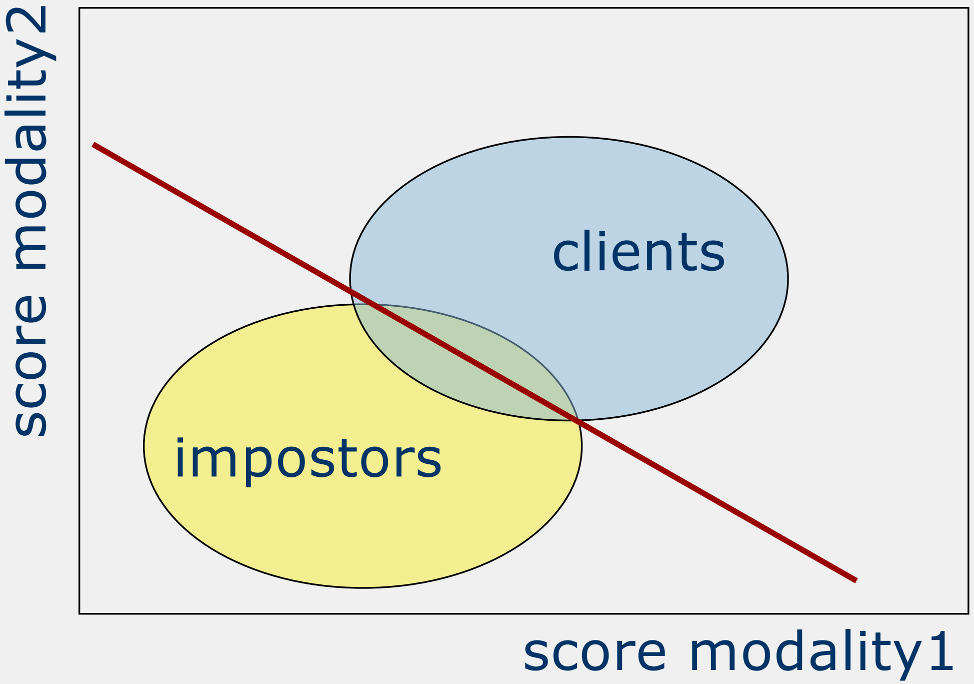
Decision-level fusion

Better performance by adapting the thresholds

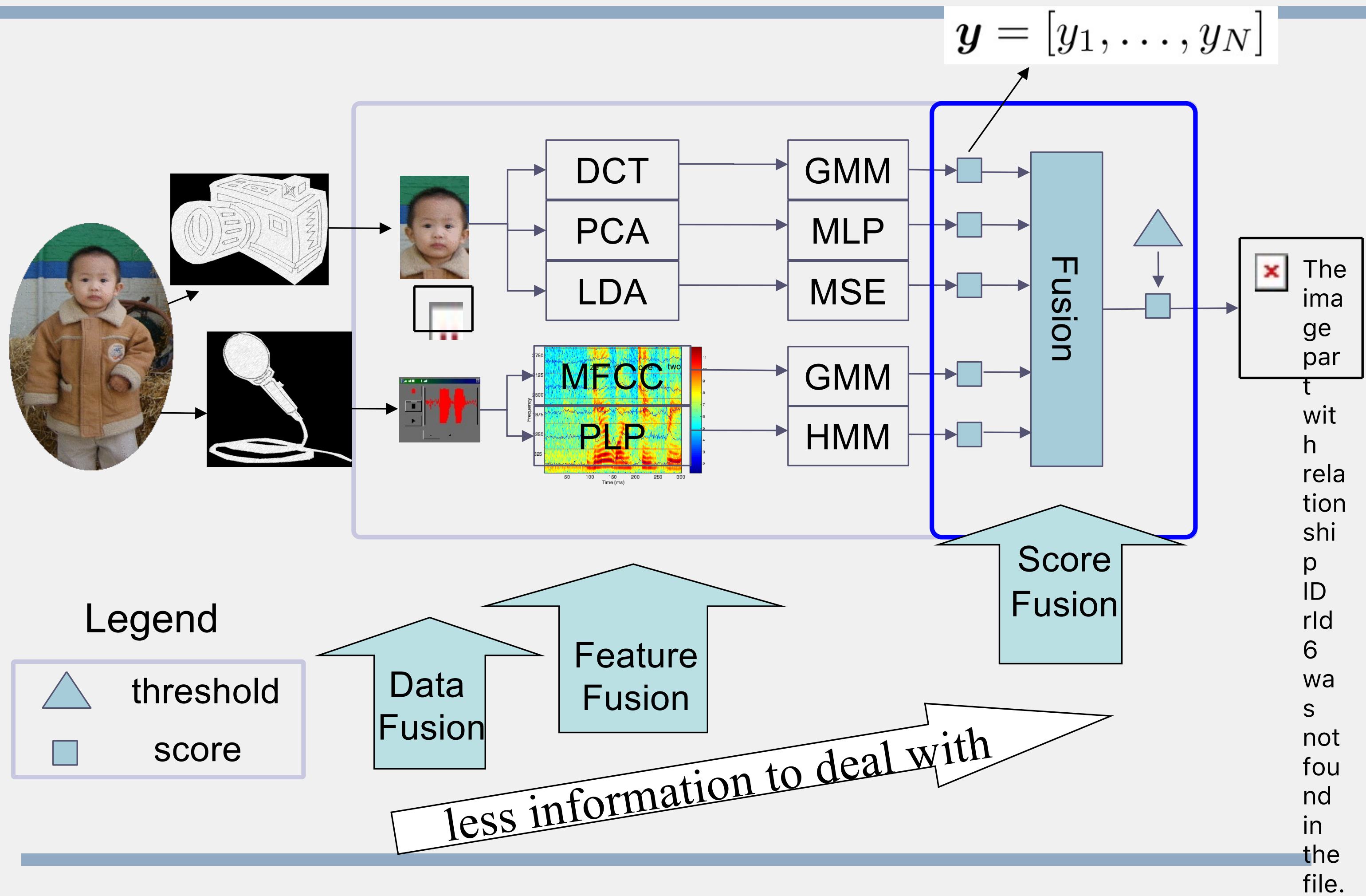


Score-level fusion

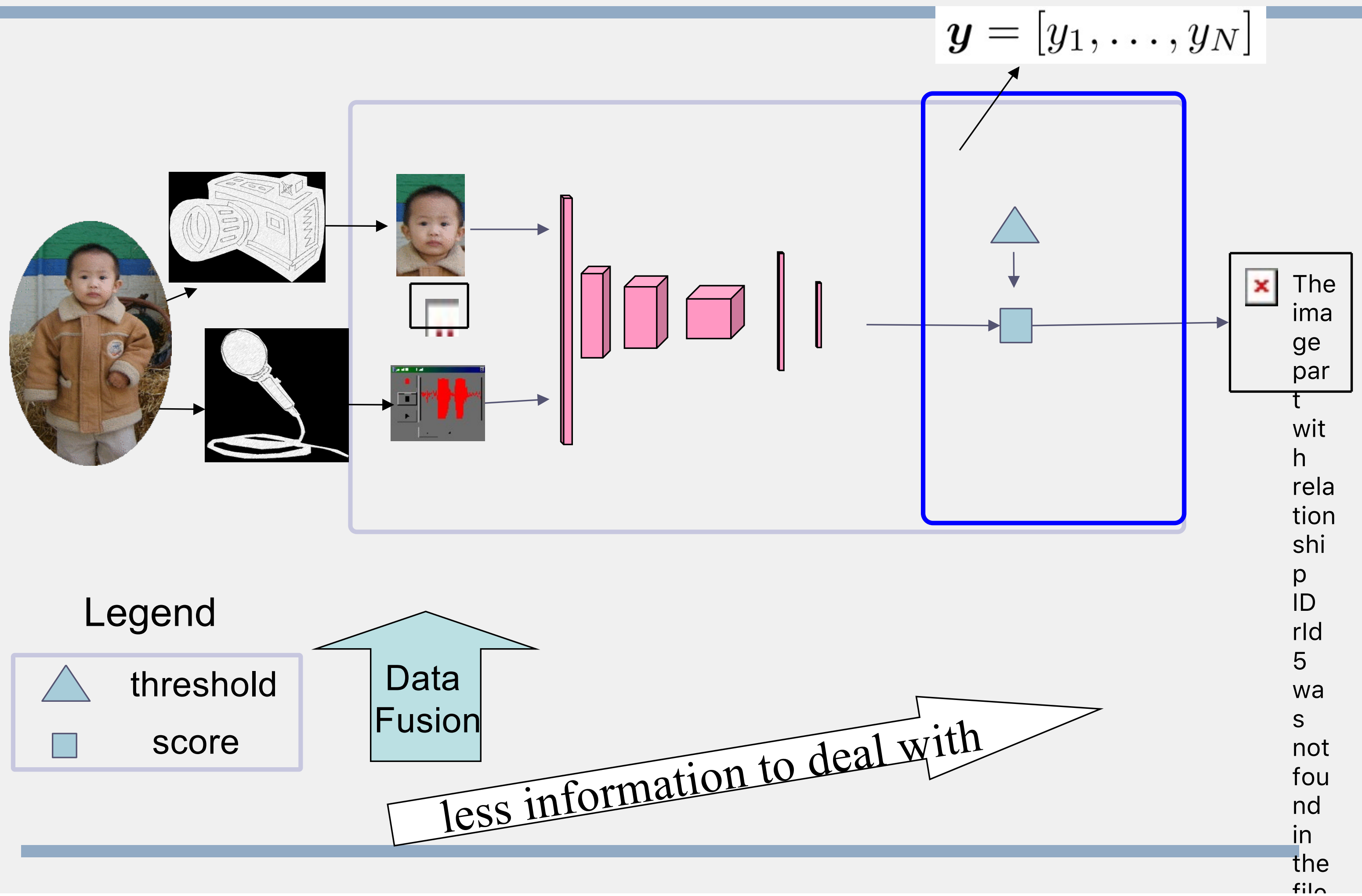
- Should improve performance



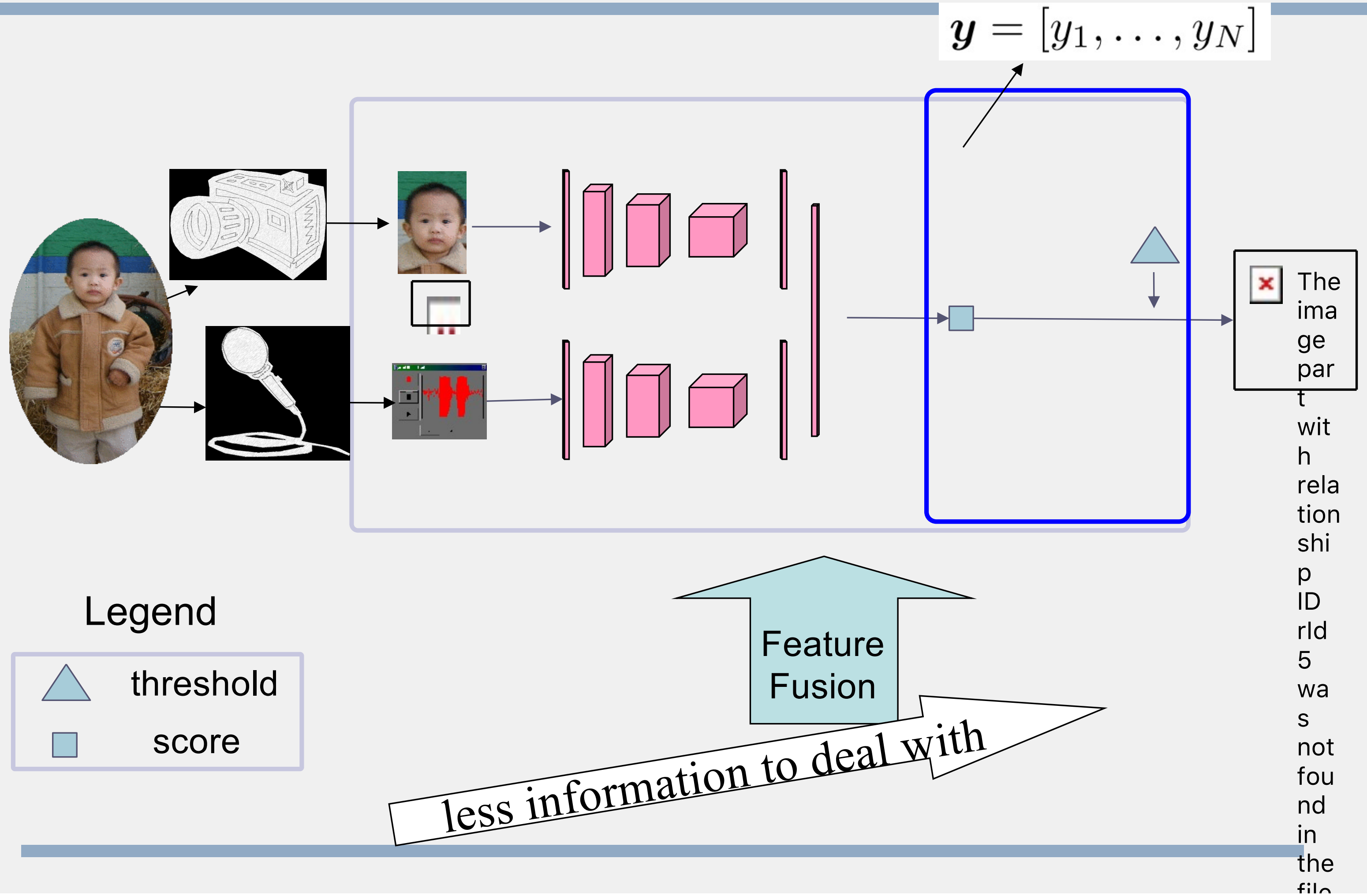
Levels of Fusion



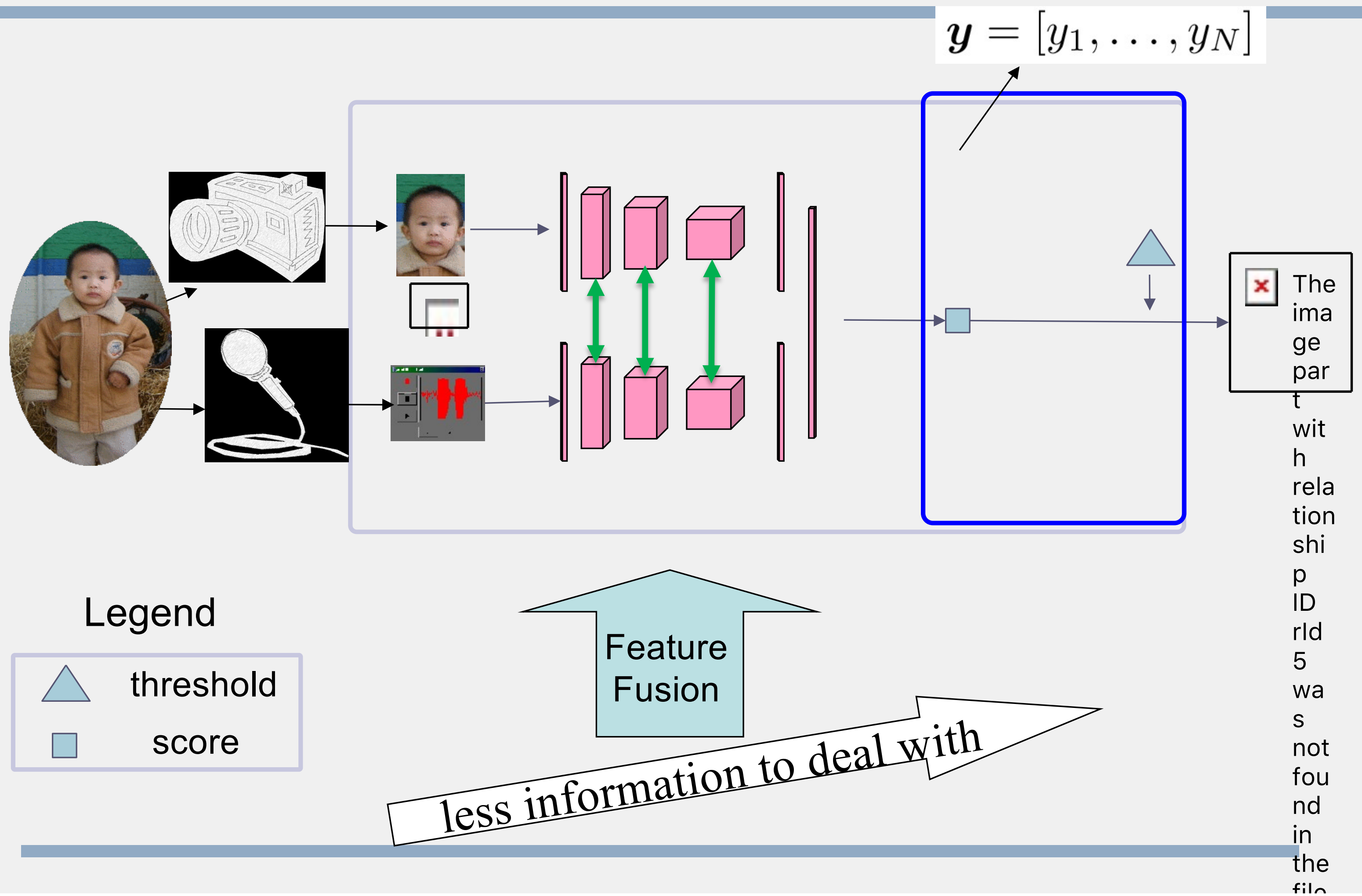
Data level fusion



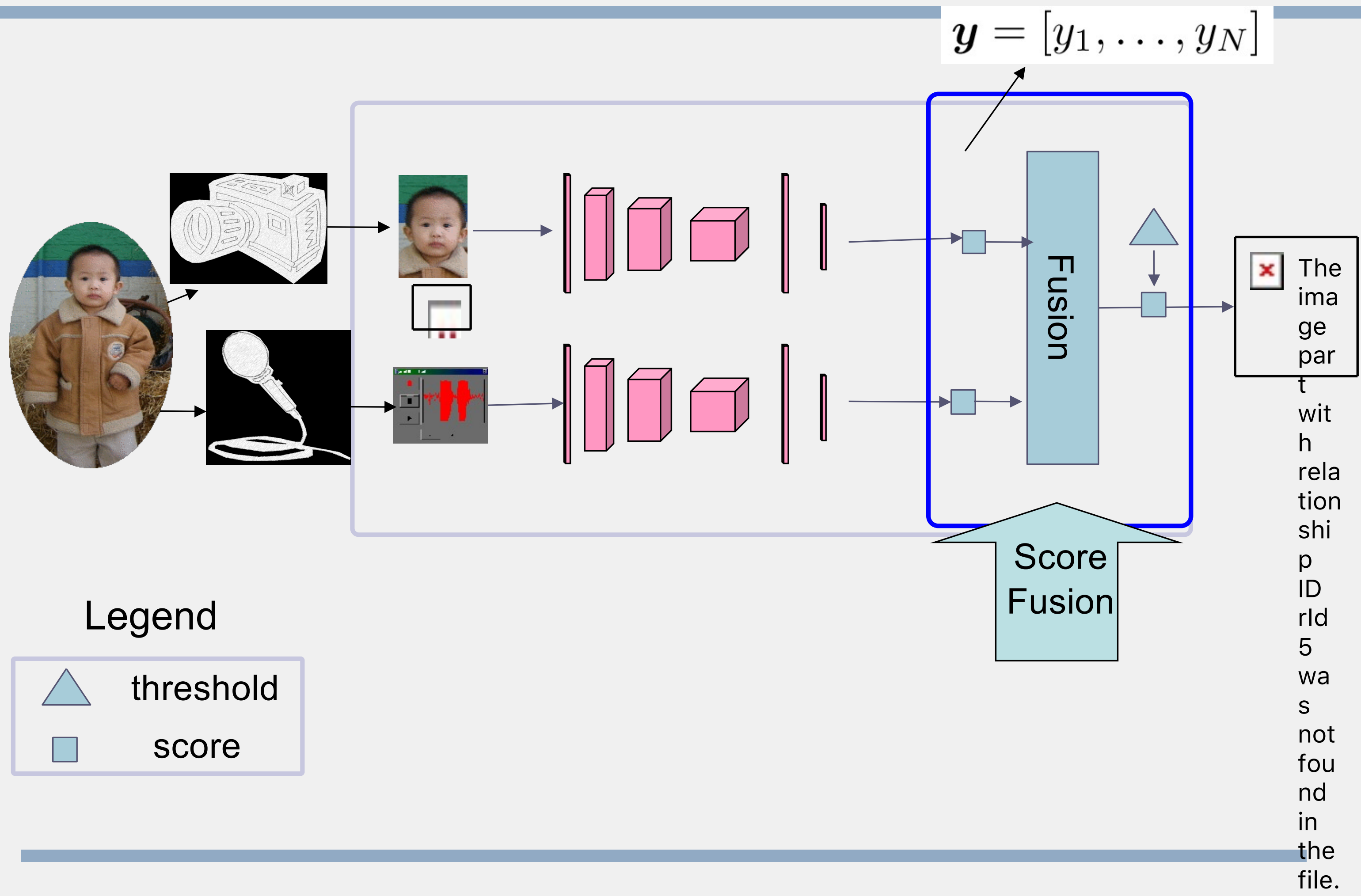
Feature level fusion



Feature level fusion



Score level fusion



17

Biometric system

Pattern recognition problem

N – number of classes

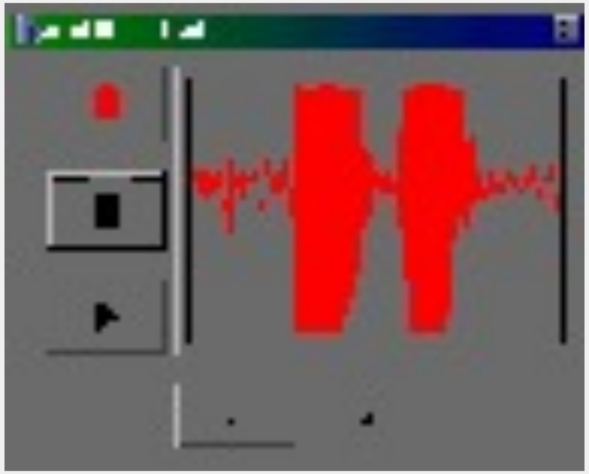
b - biometric trait

x - feature vector

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

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

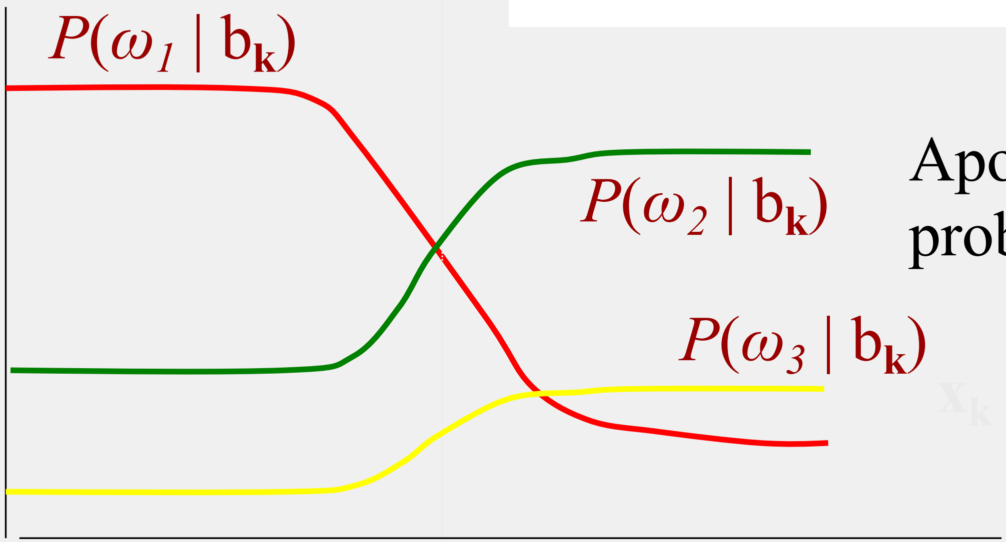
Pattern representation



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

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

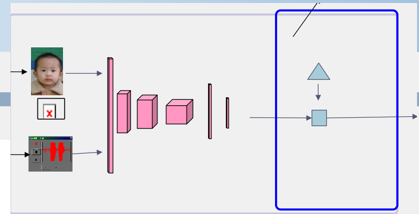
- Bayes decision rule
  - *Assign* subject *to class* ω *if*

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

- Note

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

Fusion options



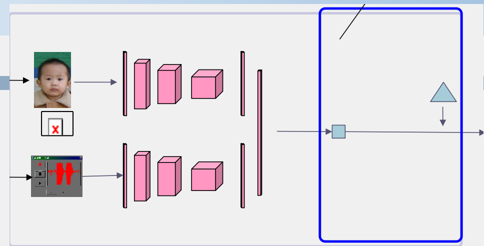
- **Signal level fusion**

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

- The integration over $\hat{x}$ is marginalisation over the distribution $p(\hat{x}|b_1, ..., b_K)$
  - 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 $P(\omega|x)$

UNIVERSITY OF SURREY

Fusion options

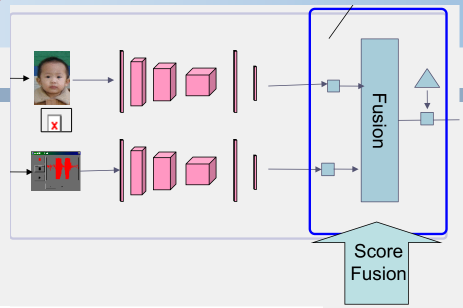


- Feature level fusion

$$\begin{aligned} p(b_1,..,b_K|\omega) &\propto \int_{\hat{x}_1,..,\hat{x}_K} p(\hat{x}_1,..,\hat{x}_K,b_1,....,b_K|\omega) \\ &\propto \int_{\hat{x}_1,..,\hat{x}_K} \hat{P}(\omega|\hat{x}_1,..,\hat{x}_K)\prod_i p(\hat{x}_i|b_i) \\ &\propto P(\omega|x_1,..,x_K) \end{aligned}$$

- 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

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

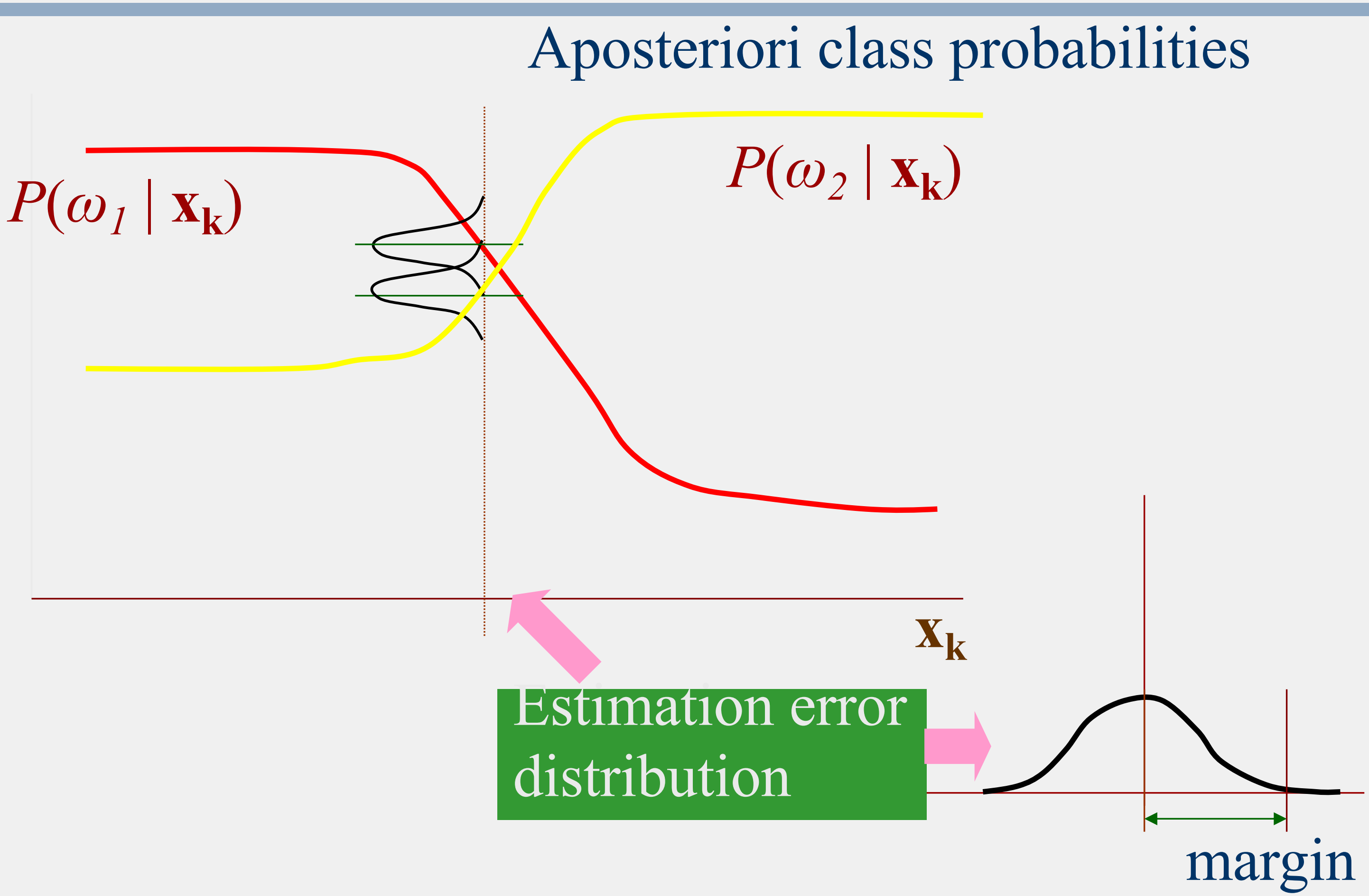
- Each modality has its own set of features x_i
- The fused score is a product of individual modality specific scores
- Fusion process marginalisation is over modality specific distributions

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

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

Effect of estimation errors

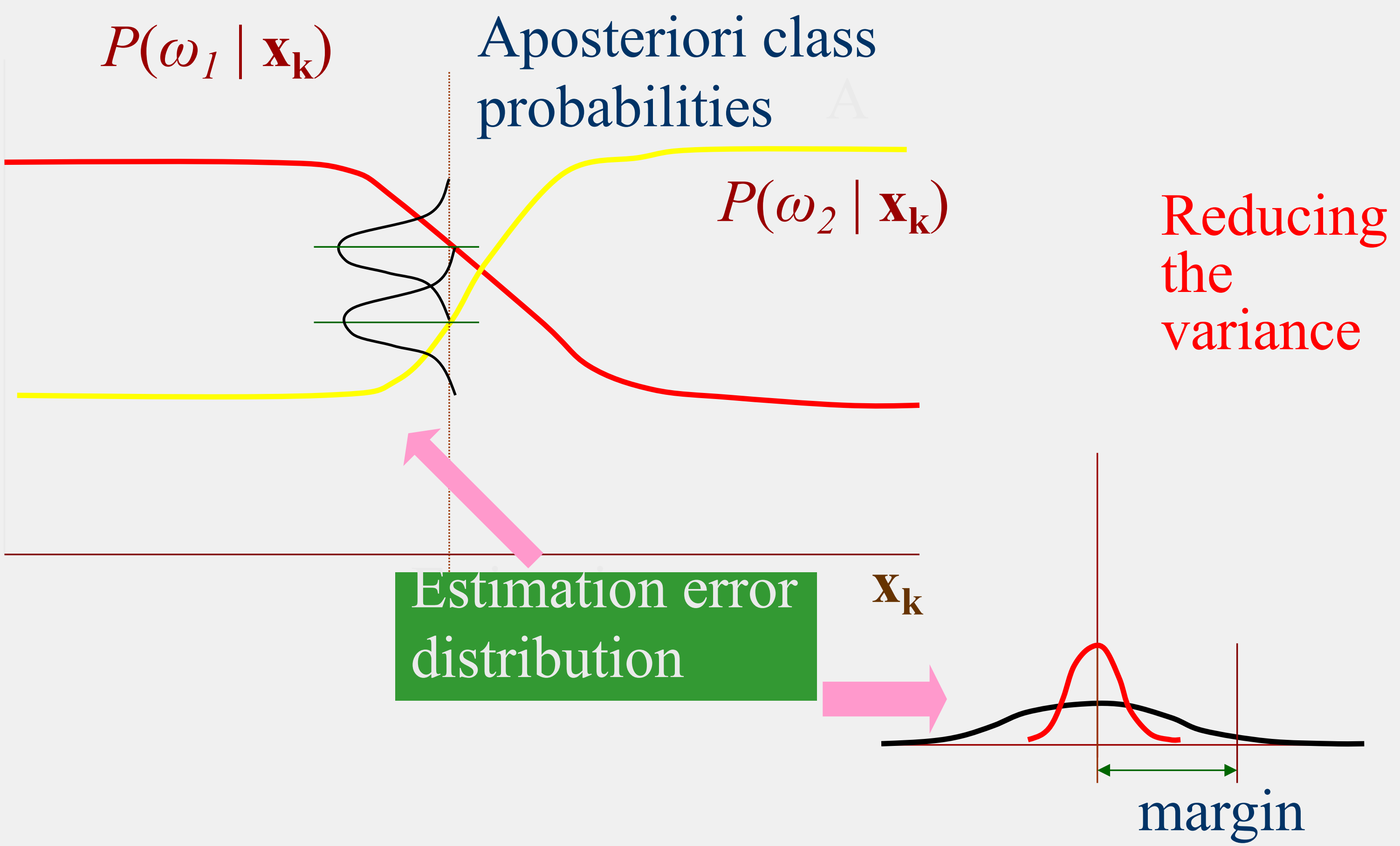


Sources of estimation errors

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

$\mathbf{x}_i$	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



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

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

UNIVERSITY OF SURREY

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

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

- Median

$$\hat{s} = \frac{s - median\ \ s}{MAD}$$

$$MAD = median|s - median(s)|$$

- Double sigmoid

$$\hat{s} = \frac{1}{1 + \exp\{-2(\frac{s-t}{r})\}}$$

r has different values for scores greater/smaller than threshold t

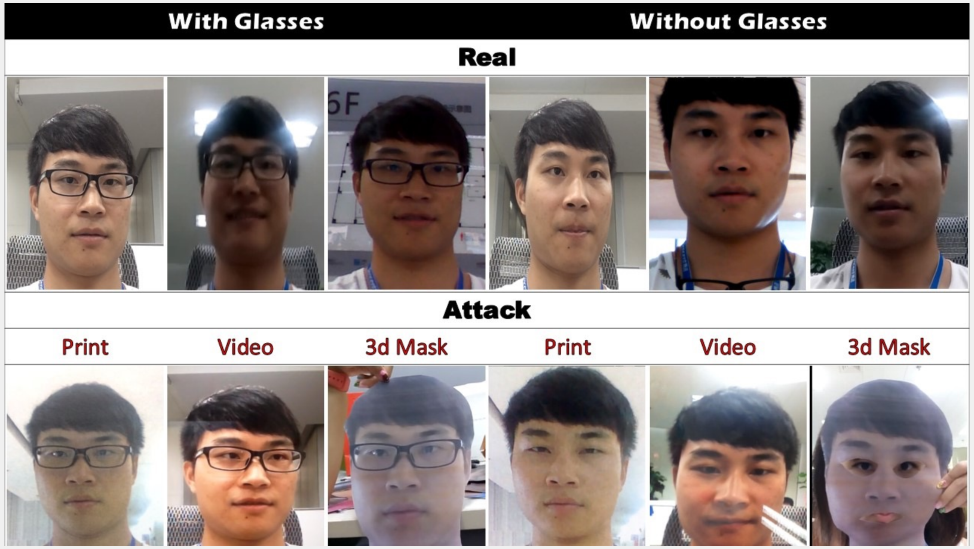
- Tanh

$$\hat{s} = 0.5[\tanh\{0.01\frac{s-\mu}{\sigma}\} + 1]$$

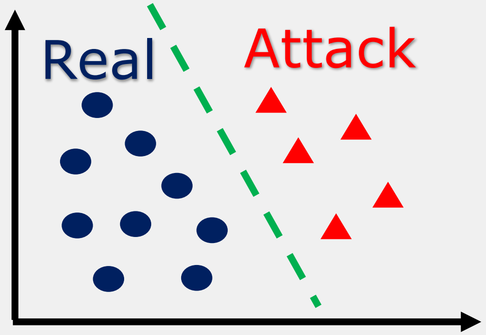
- Min-max, Z-score and tanh are efficient, median, double-sigmoid and tanh are robust

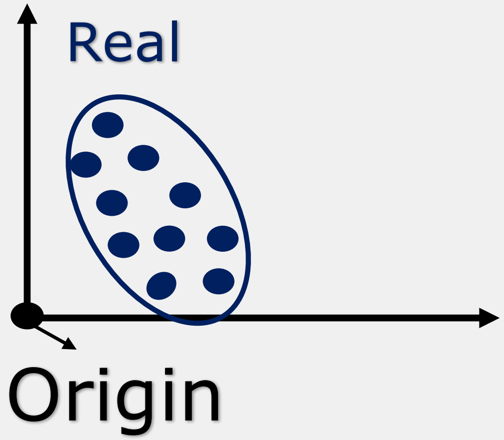
Face spoofing attack detection

- The problem

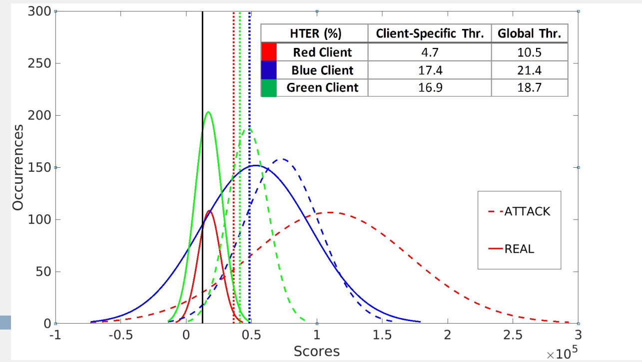


- The approach – anomaly detection



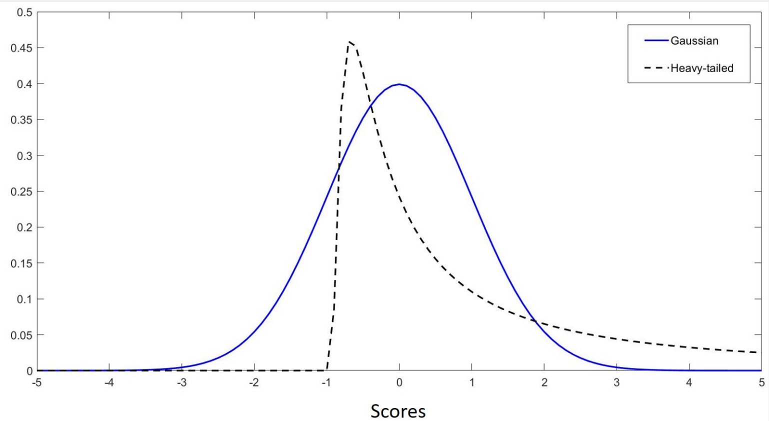


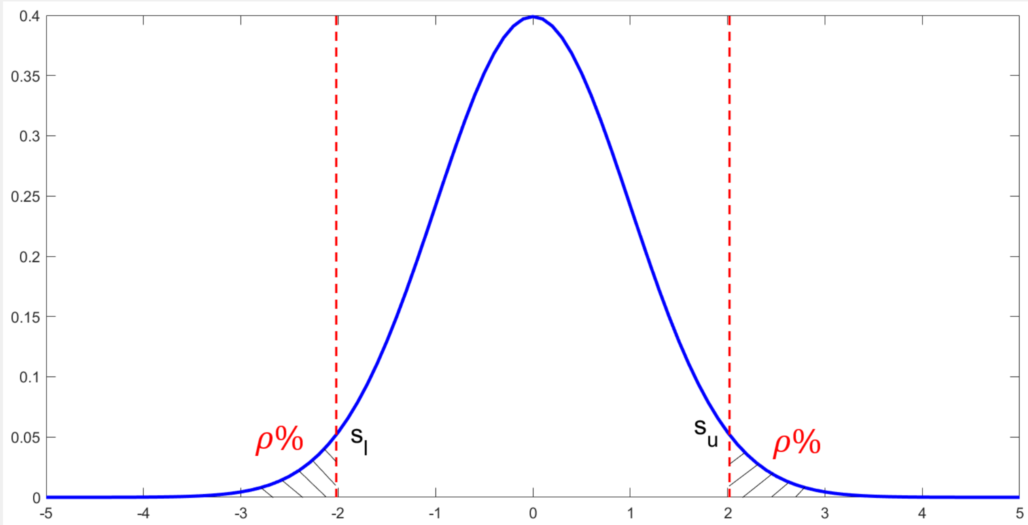
- Client specific solution



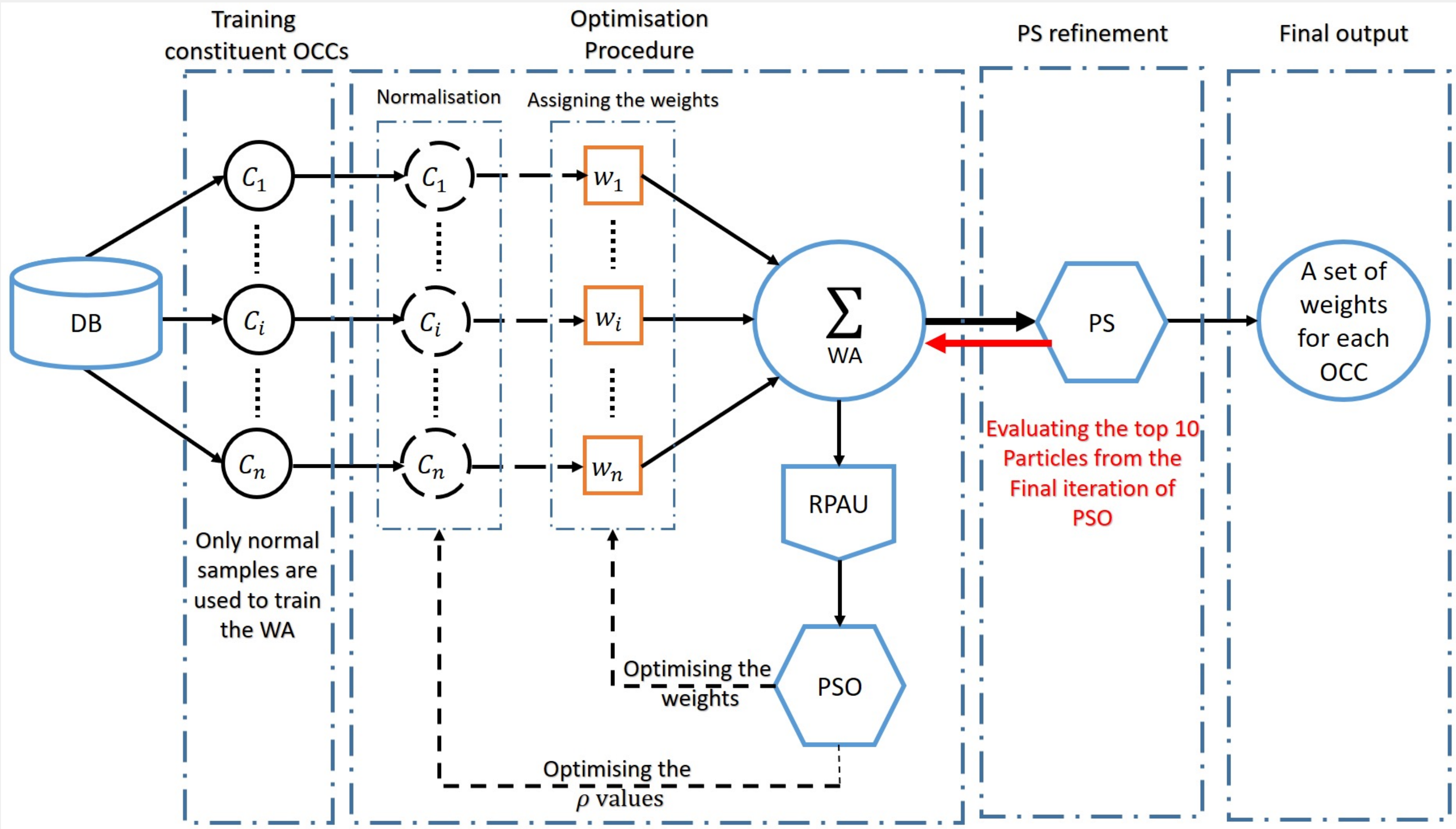
One class normalisation method

- Two-sided normalization
  - ρ% tail cut-off
  - cut-off points mapped to [0,1]
- Heavy tail distribution





Fusion of anomaly detectors

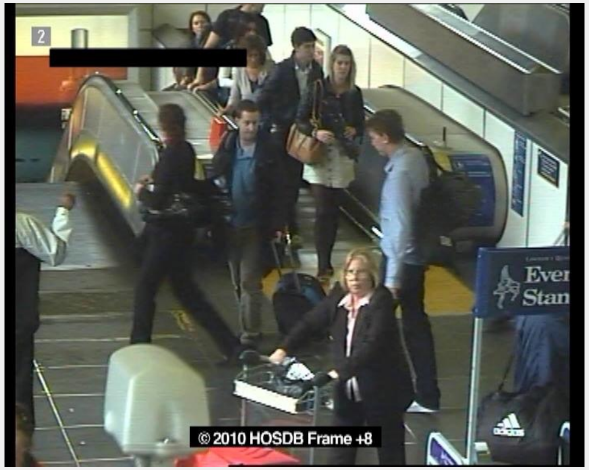


Case study in multimodal soft biometric fusion

- 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

UNIVERSITY OF SURREY

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

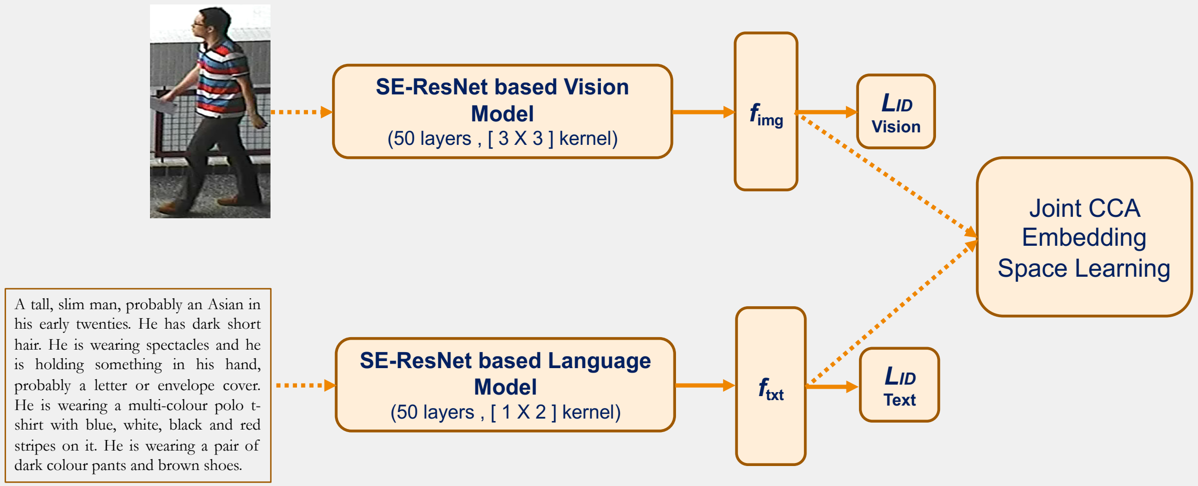
Language annotation

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.

Person Re-ID

- Crossmodal & multimodal matching facilitated by CAA



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

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

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

- 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

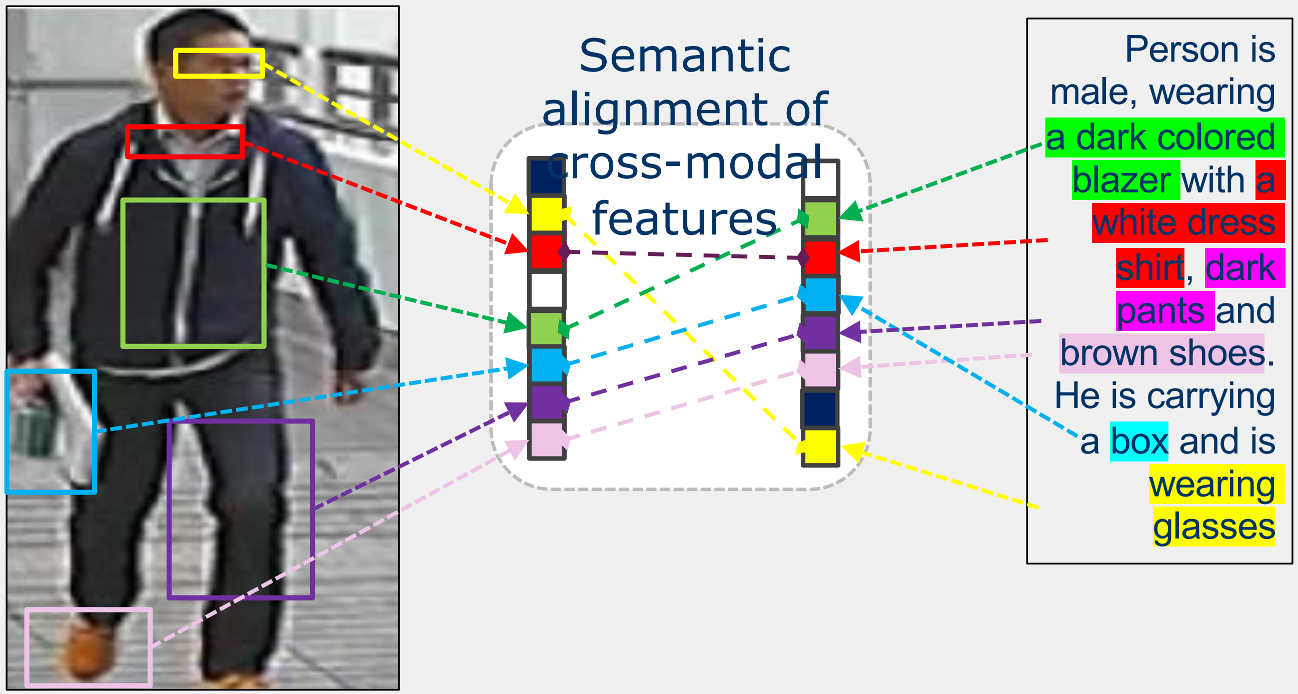
Re-ID with V&L

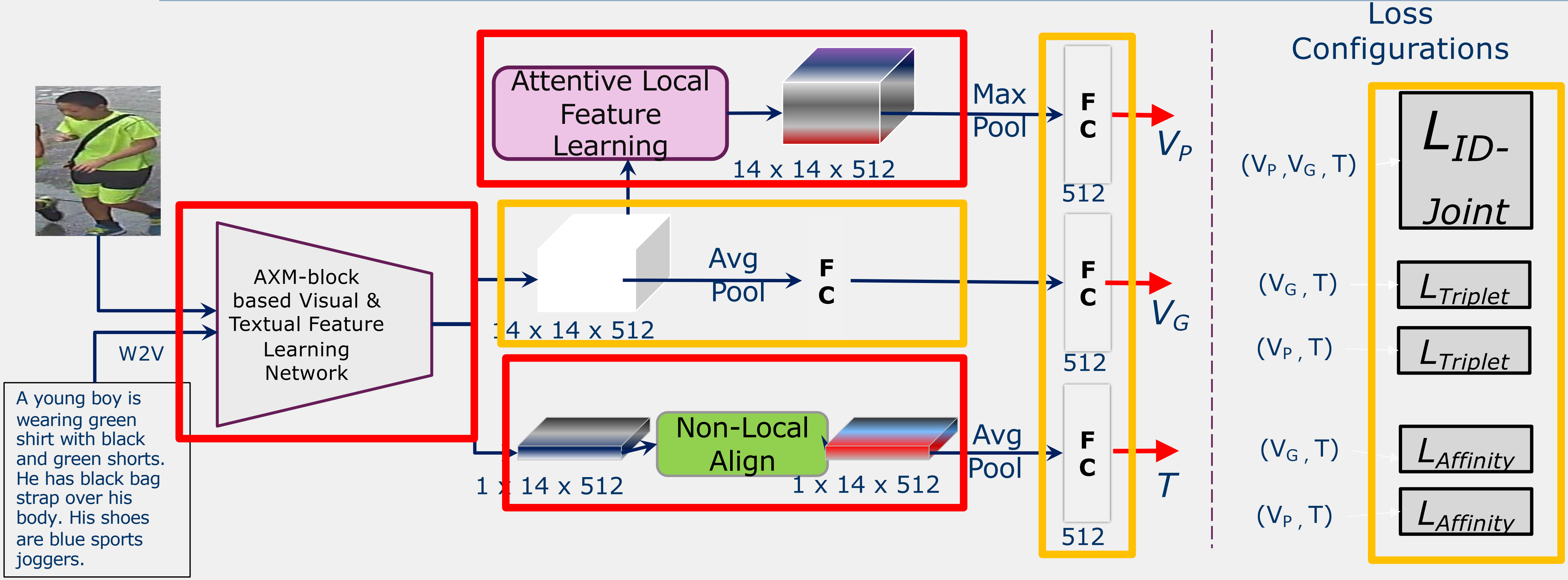
- Three sets:
  - Training, query, gallery
  - Training: image and language pairs
- Various settings, query x gallery:
  - V x V, L x L, V x L, V x VL, VL x VL
- Asymmetric settings:
  - Transfer language info. With CCA
- XQDA as metric learning

Multimodal and cross-modal image retrieval

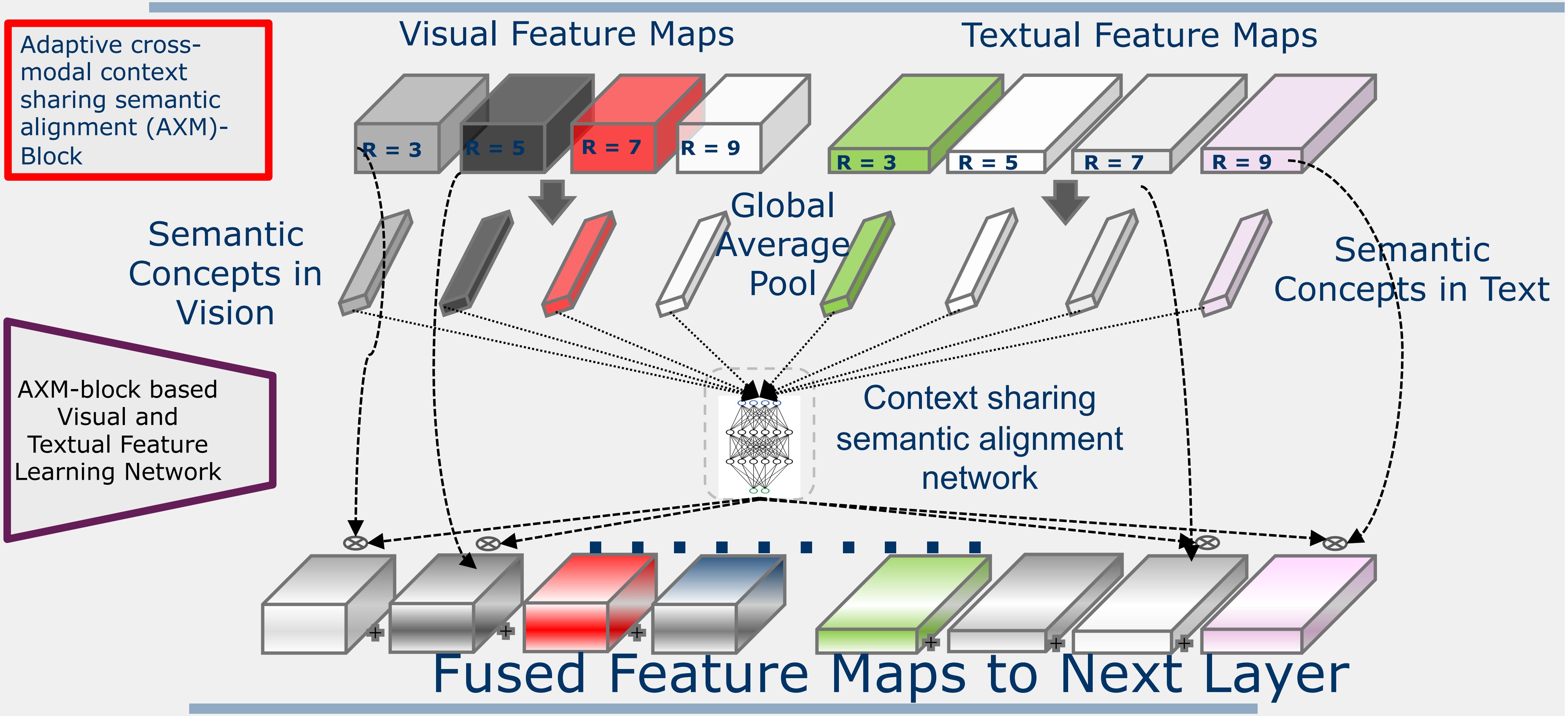
AXM-Net: Semantic Alignment and Context Sharing for Cross-Modal Person Re-identification

Person is male, wearing a dark colored blazer with a white dress shirt, dark pants and brown shoes. He is carrying a box and is wearing glasses


Semantic alignment of cross-modal features
Person is male, wearing a dark colored blazer with a white dress shirt, dark pants and brown shoes. He is carrying a box and is wearing glasses



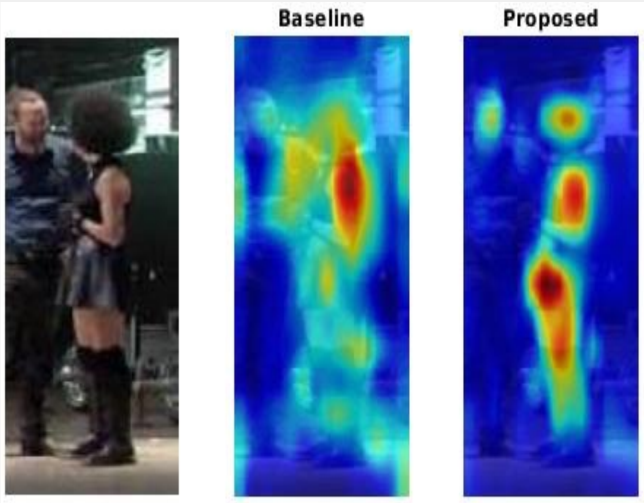
Ammarah Farooq, etal., "**AXM-Net: Cross-Modal Context Sharing Attention Network for Person Re-ID**", Arxiv, 2021.

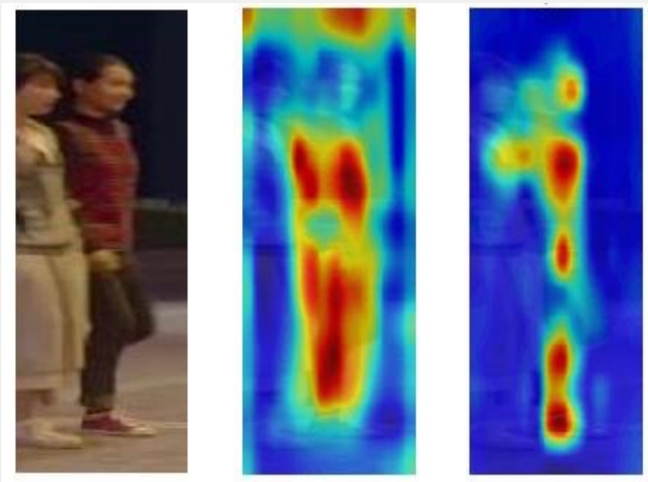

Adaptive cross-modal context sharing semantic alignment (AXM)-Block
Visual Feature Maps
Textual Feature Maps
R = 3
R = 5
R = 7
R = 9
R = 3
R = 5
R = 7
R = 9
Semantic Concepts in Vision
Global Average Pool
Semantic Concepts in Text
AXM-block based Visual and Textual Feature Learning Network
Context sharing semantic alignment network
Fused Feature Maps to Next Layer

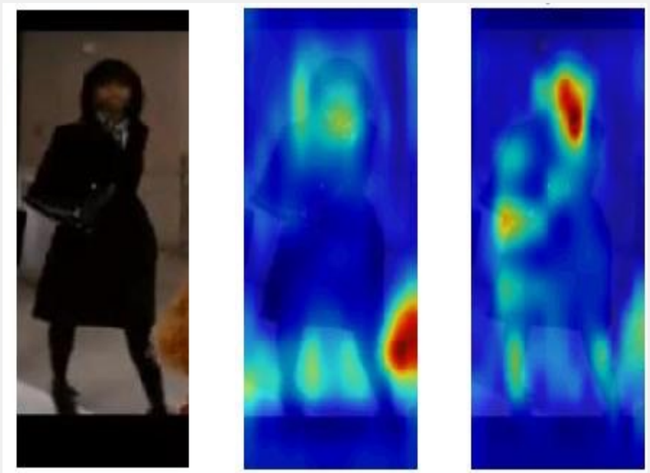
Model	CrossRe-ID						CUHK-SYSU					
	V → V		T → V		VT → V		V → V		T → V		VT → V	
	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP
JT + CCA [12]	86.77	88.90	33.61	39.40	88.59	87.95	74.13	77.16	11.37	15.78	77.68	75.8
AXM-Net + joint ID + affinity	95.14	96.04	44.66	50.49	95.26	95.22	86.00	87.75	19.93	24.82	88.72	87.02
AXM-Net + joint ID + triplet	95.02	96.00	47.33	52.58	95.75	95.41	86.24	88.02	20.93	26.04	87.86	86.40
AXM-Net + joint ID + affinity + triplet	94.29	98.9	46.48	52.21	94.05	93.93	85.86	87.70	21.44	26.77	88.62	86.73

Table 3. Performance comparison on cross-modal Re-ID. Query → Gallery

Rejection of noisy information





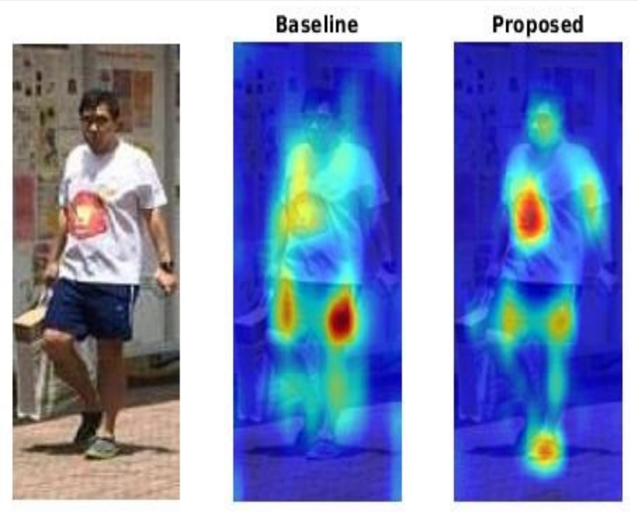


The woman wears a black sleeveless top. She wears a black leather skirt with black boots and has a curly brown afro.

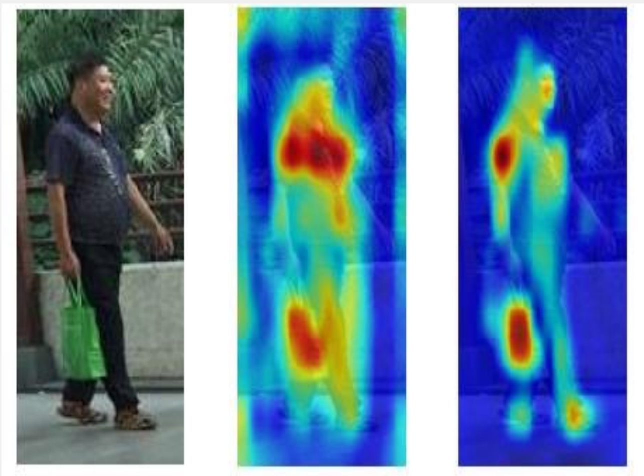
This person is wearing a black and red tartan sweatshirt, cuffed jeans, and red converses.

The lady wears a black long jacket and black boots. She is carrying a black should bag.

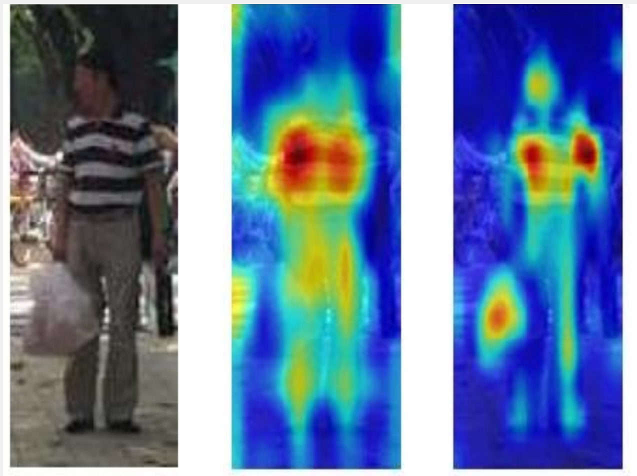
Focus on discriminative information



A man carries a brown package inside a white tote bag with green graphics while wearing a white t-shirt with a ***red-and-yellow animal face centered on the front*** over blue shorts with a white stripe on the sides and gray shoes.



The man is wearing a navy blue shirt with black pants. He has on brown shoes. He is carrying a ***green bag***.



The man is looking over his shoulder to his right. He has short cut black hair. He is wearing a horizontally striped short sleeved short with khaki pants and dark shoes. The man is holding a ***white shopping bag in his right hand***.

Conclusions and future research

- Conclusions
  - We have provided an information theoretic underpinning of machine learning
  - The properties of information measures impact on performance
    - Function properties of measures, data distribution models

- Future directions of research

- Training distribution
  - augmentation
  - balancing distribution biases
  - feature distribution augmentation
  - boosting
  - unlabelled data
- Parameter distribution
- Domain adaptation/shift
- Testing and evaluation
- Quality dependent distributions

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.