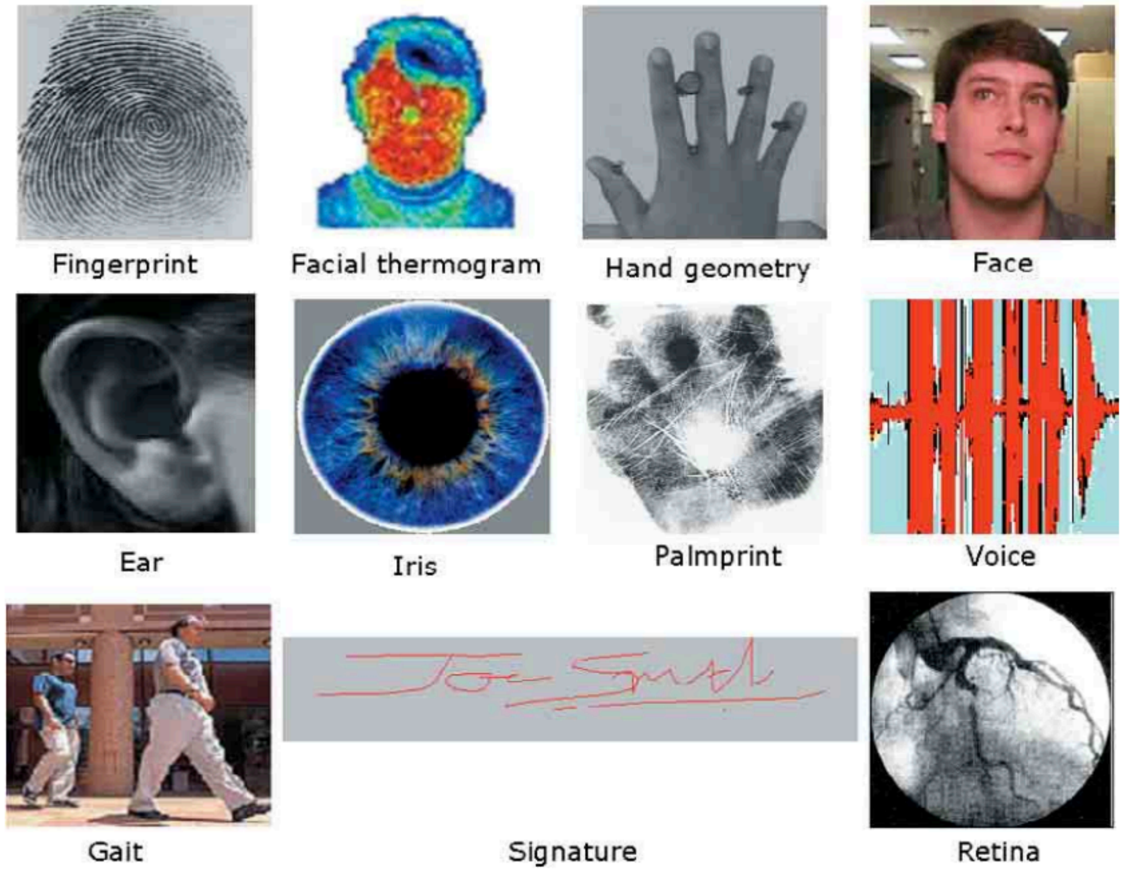
Multimodal Biometrics

Slides are prepared from the resources on the web.

Biometric Modalities



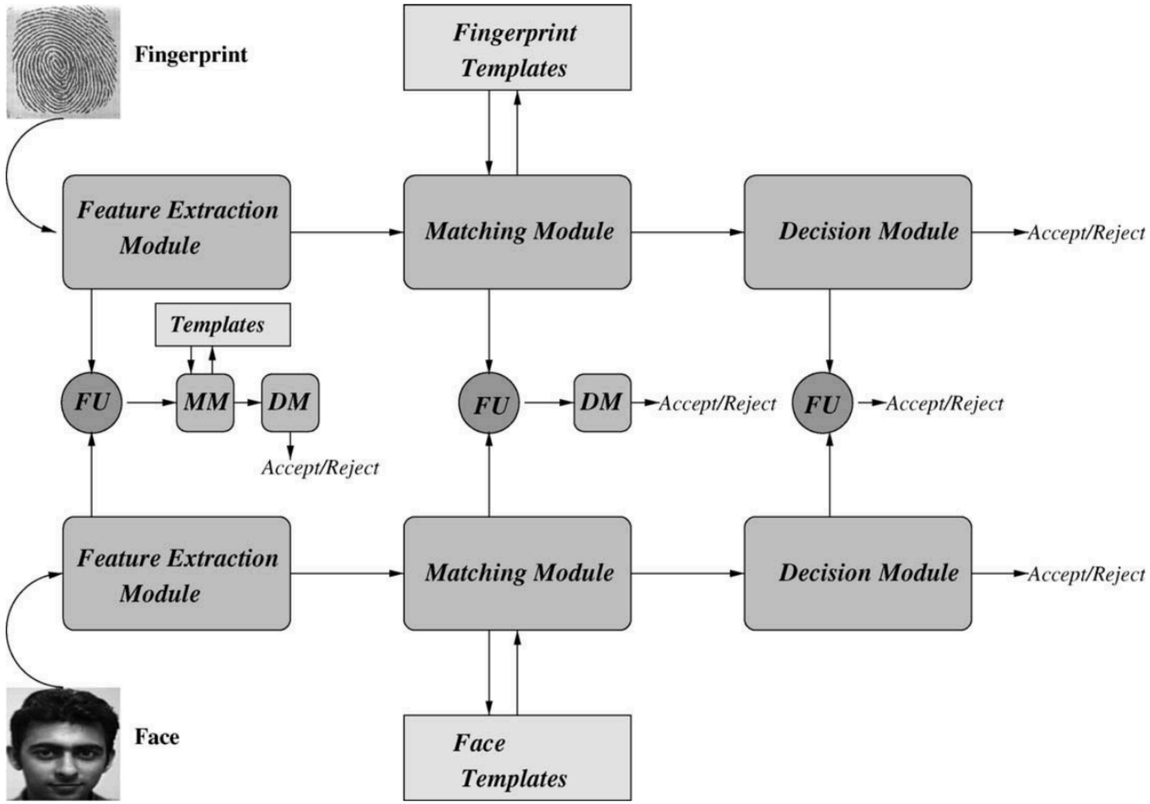
Why do we need multimodal biometrics?

- Universality
- Uniqueness
- Improved discrimination power

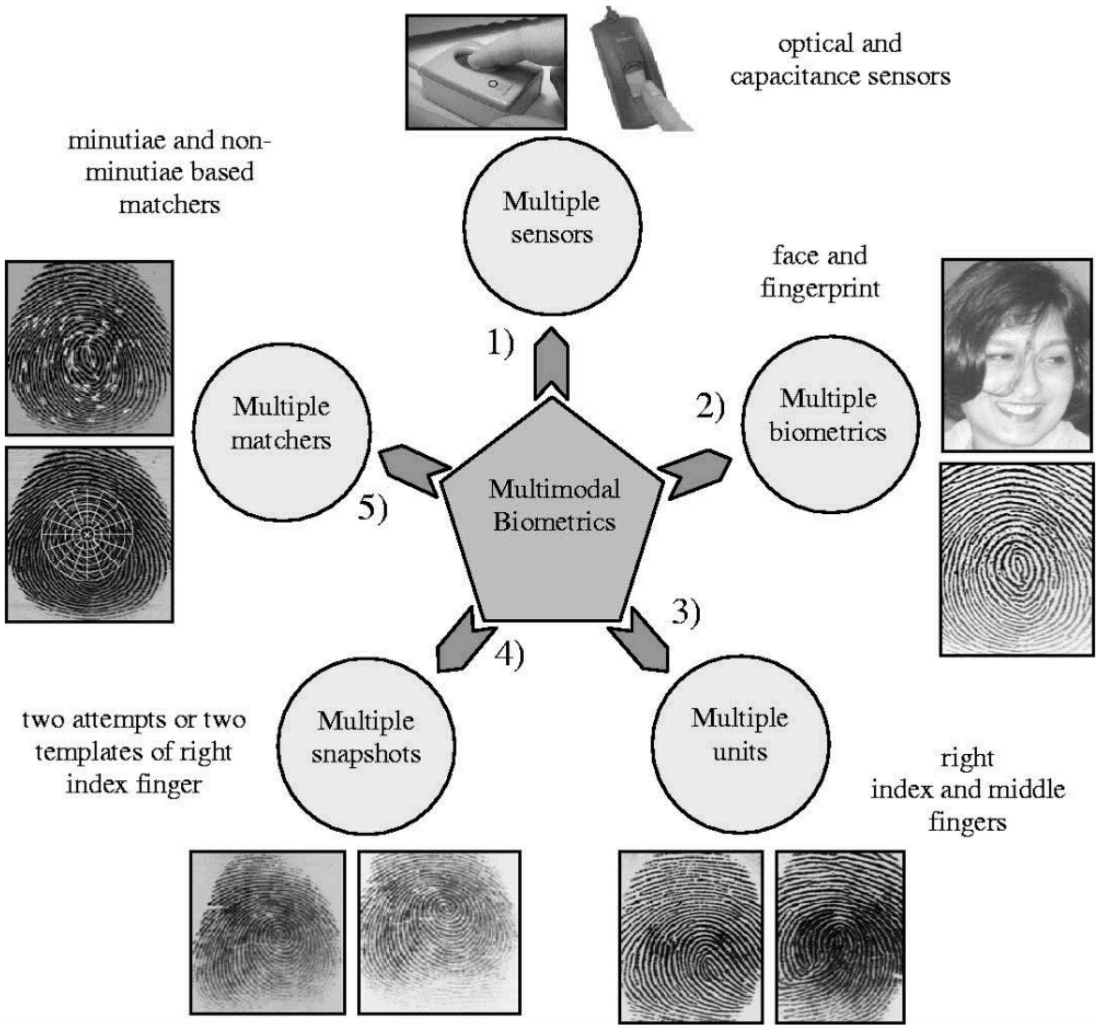
Biometric Fusion

A bimodal biometric system showing the three levels of fusion

(FU: fusion module, MM: matching module, DM: decision module)

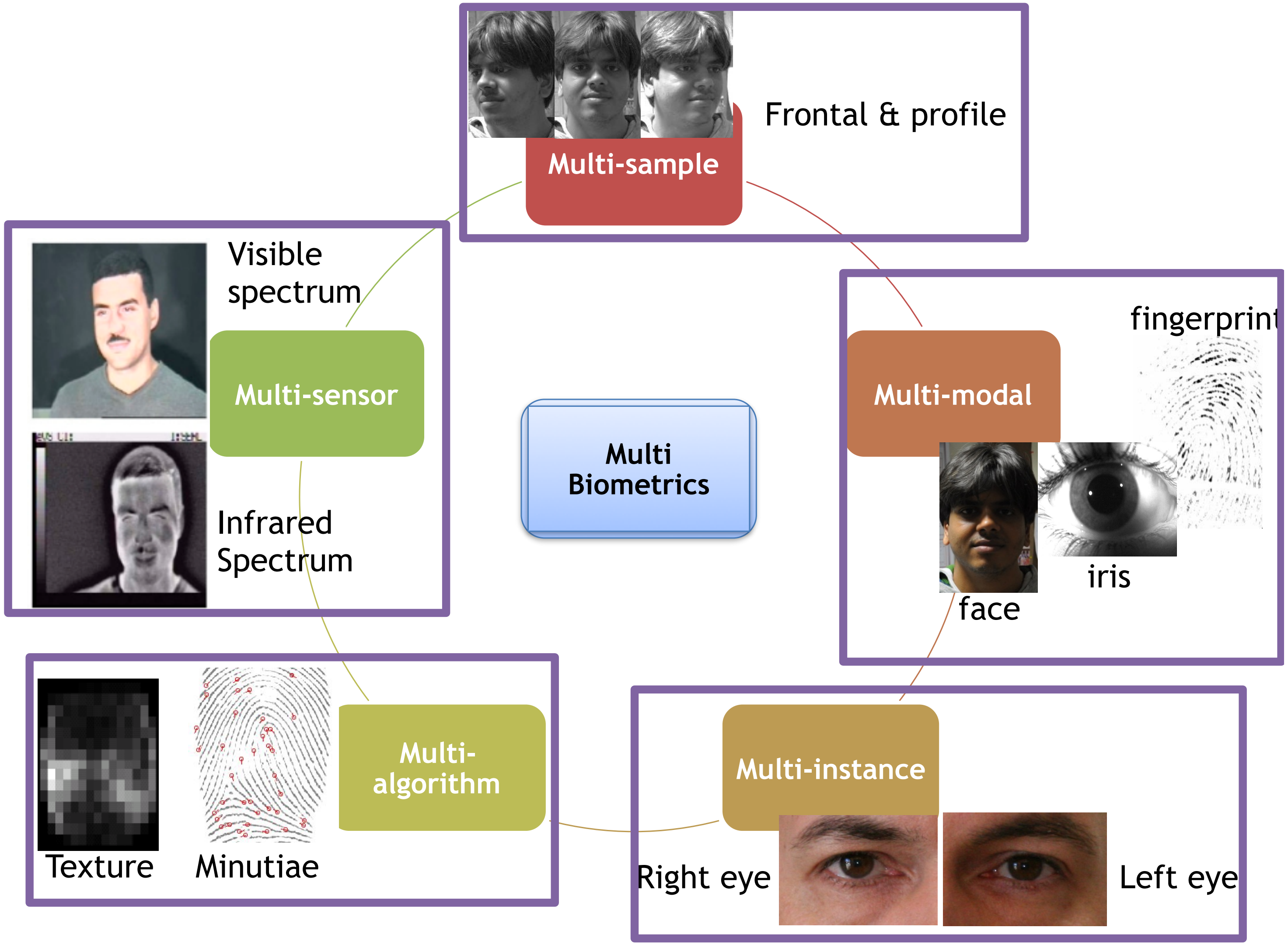


Different kinds of fusion



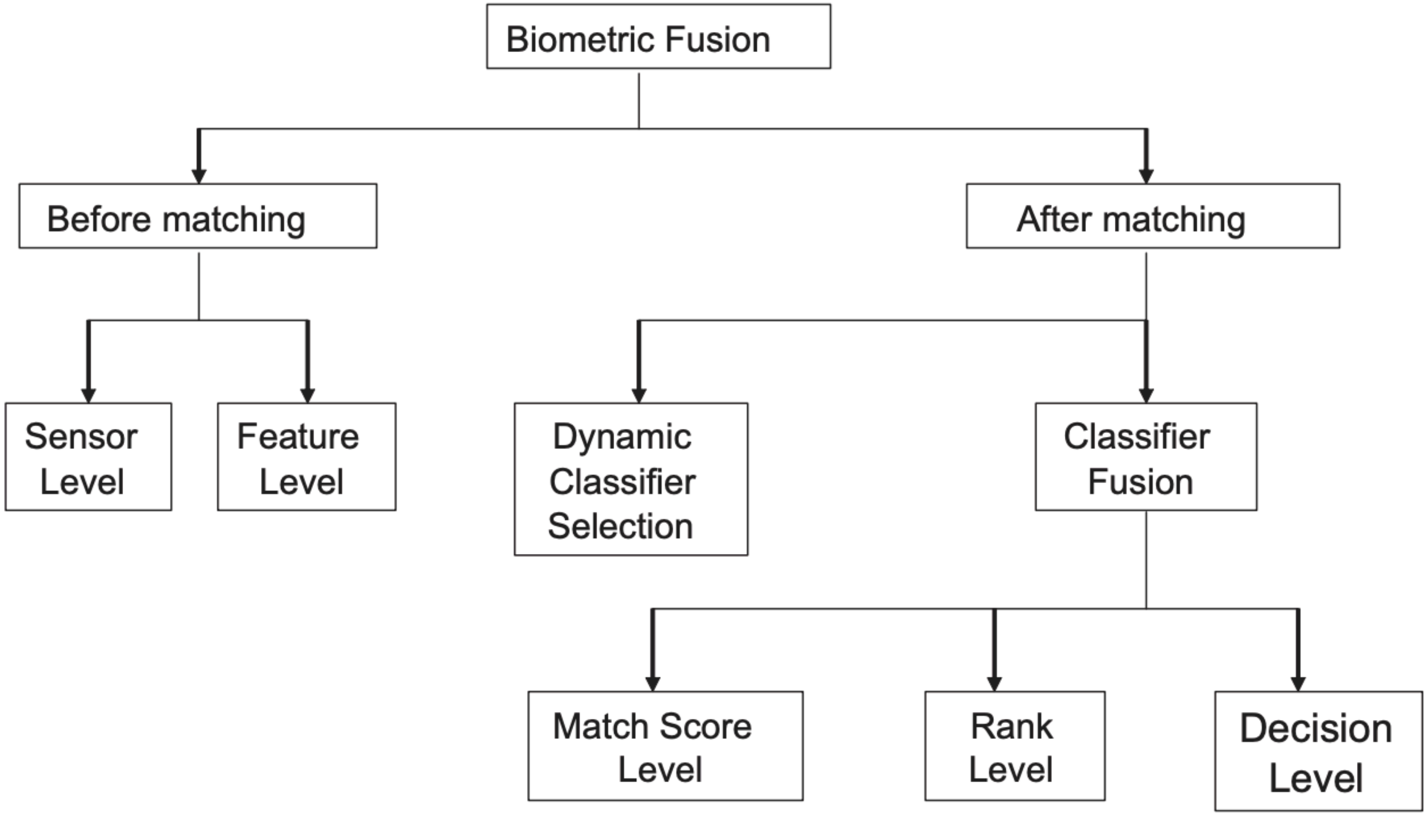
Information fusion in biometrics A Ross, A Jain
Pattern recognition letters 24 (13), 2115-2125

Taxonomy of Fusion

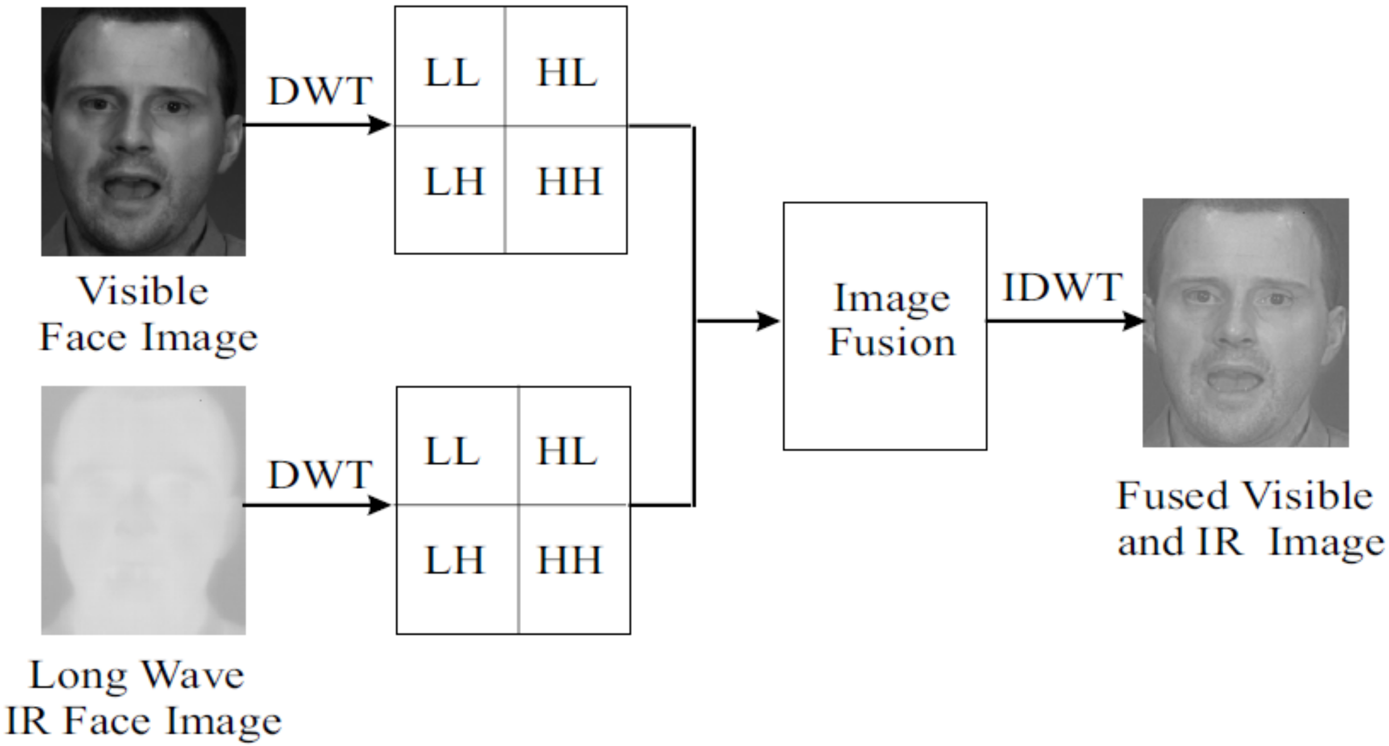


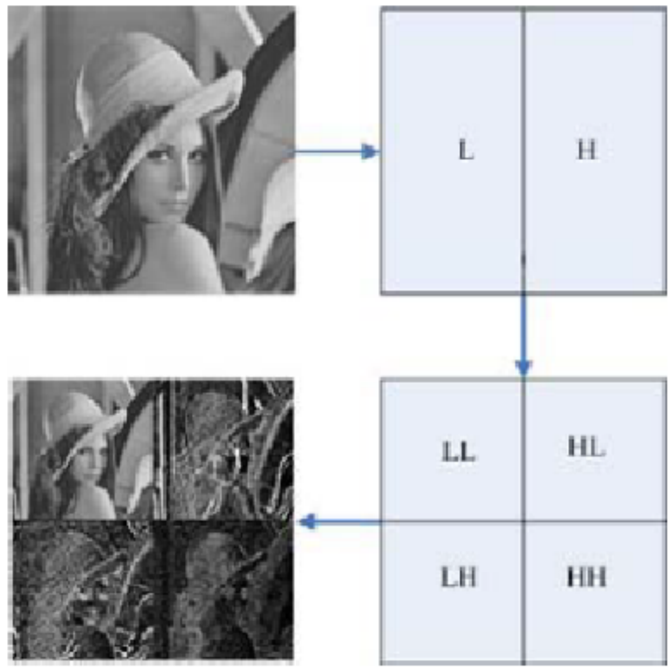
Different Levels of Fusion

- Data
- Feature
- Match score
- Decision
- Rank

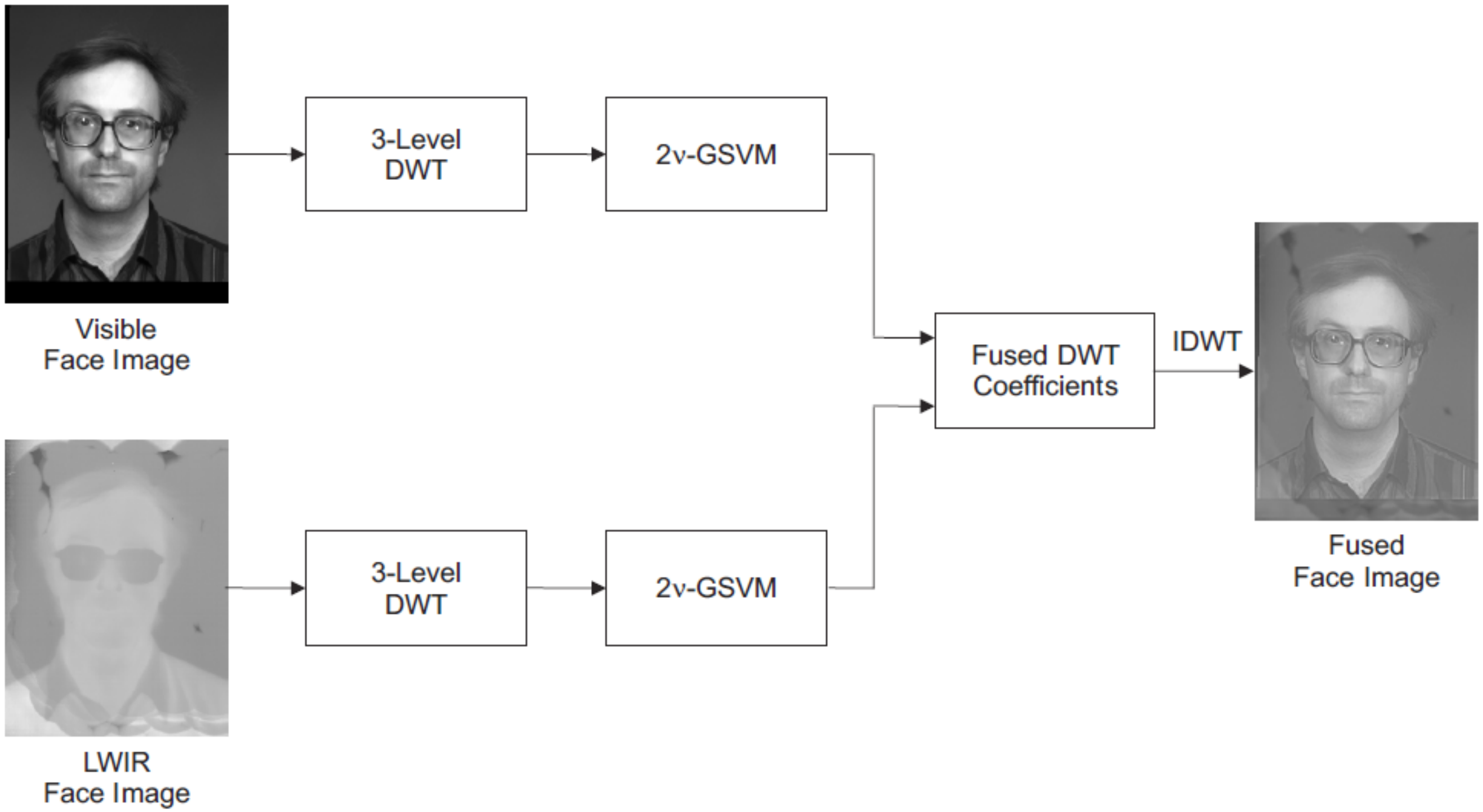


Sensor Level Fusion





Sensor Level Fusion

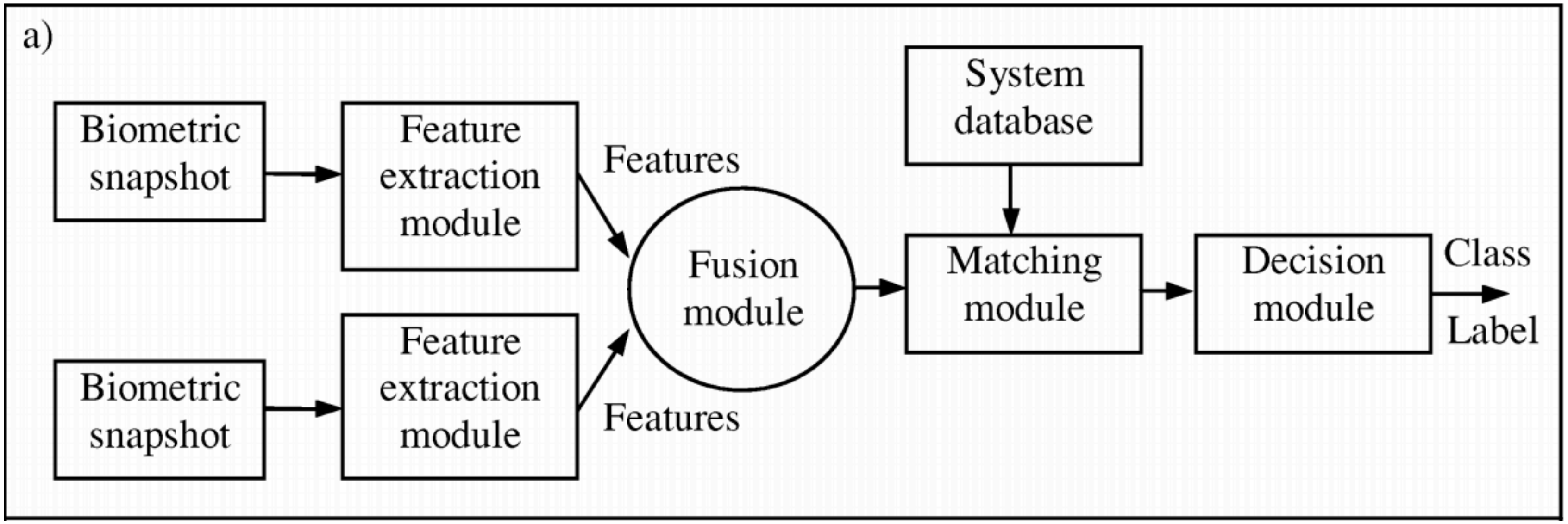


Sensor Level Fusion

- Face mosaicing: stitching multi-view images into one

R. Singh, M. Vatsa, A. Ross and A. Noore, A Mosaicing Scheme for Pose Invariant Face Recognition, *In IEEE Transactions on Systems, Man, and Cybernetics - B, Special Issue on Biometrics*, Vol. 37, No. 5, pp. 1212-1225, 2007.

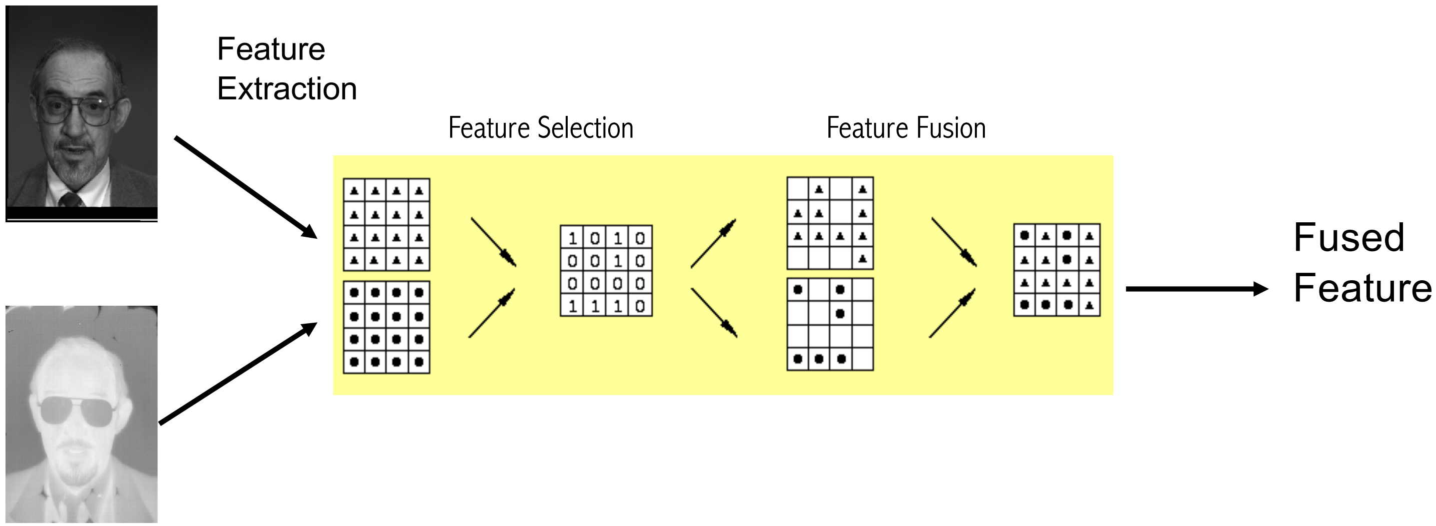
Fusion at the feature extraction level



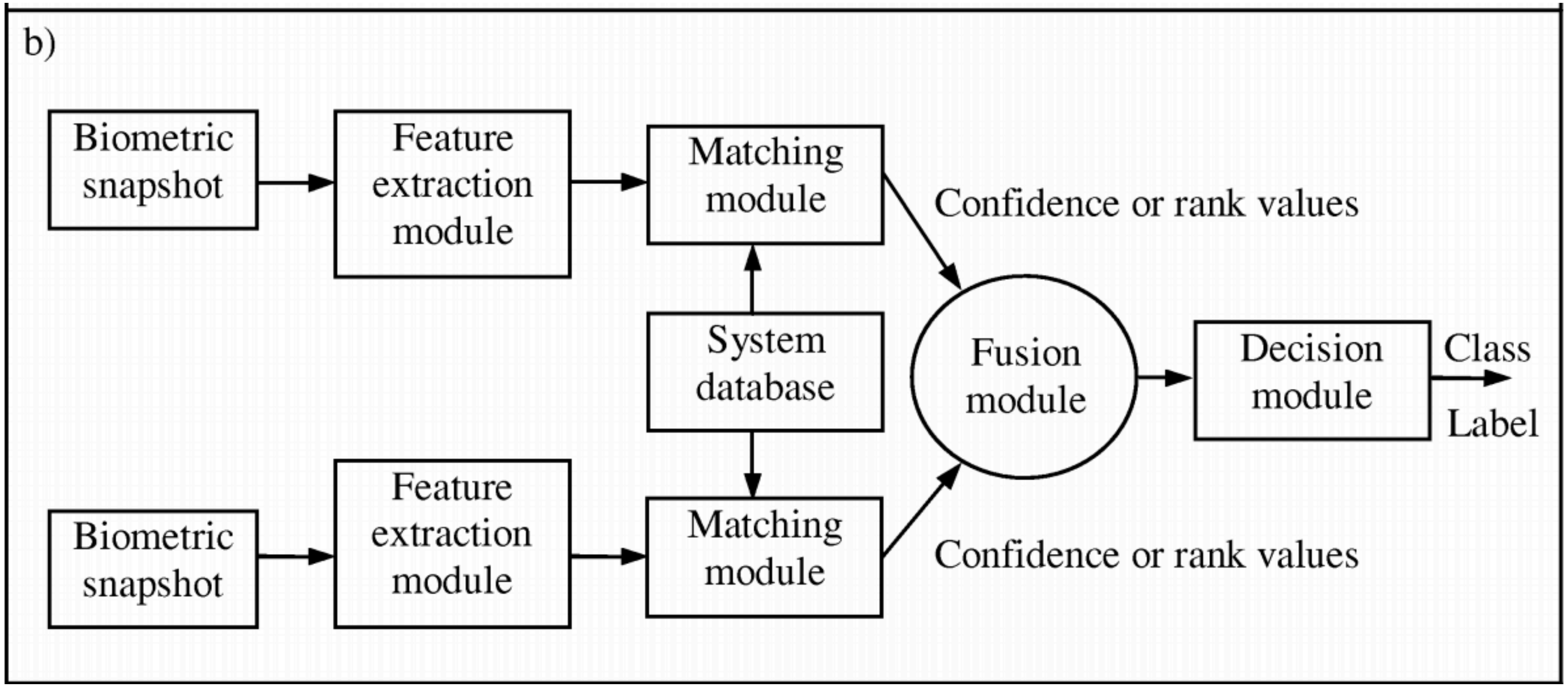
Feature Level Fusion

- Simplest: Concatenation
- Issues: ???
- How can we do better concatenation?

Feature Level Fusion



Fusion at matching score (confidence or rank) level



Match Score Level Fusion

- Compared to image level and feature level fusion – do we have any benefits of fusing at match score level?

Match Score Fusion

- Simplest score fusion:
  - C = A + B

- Weighted Sum score fusion
  - C = w1*A + w2*B

- What are the problems with this? What all do you have to consider before you apply this equation?

Match Score Fusion

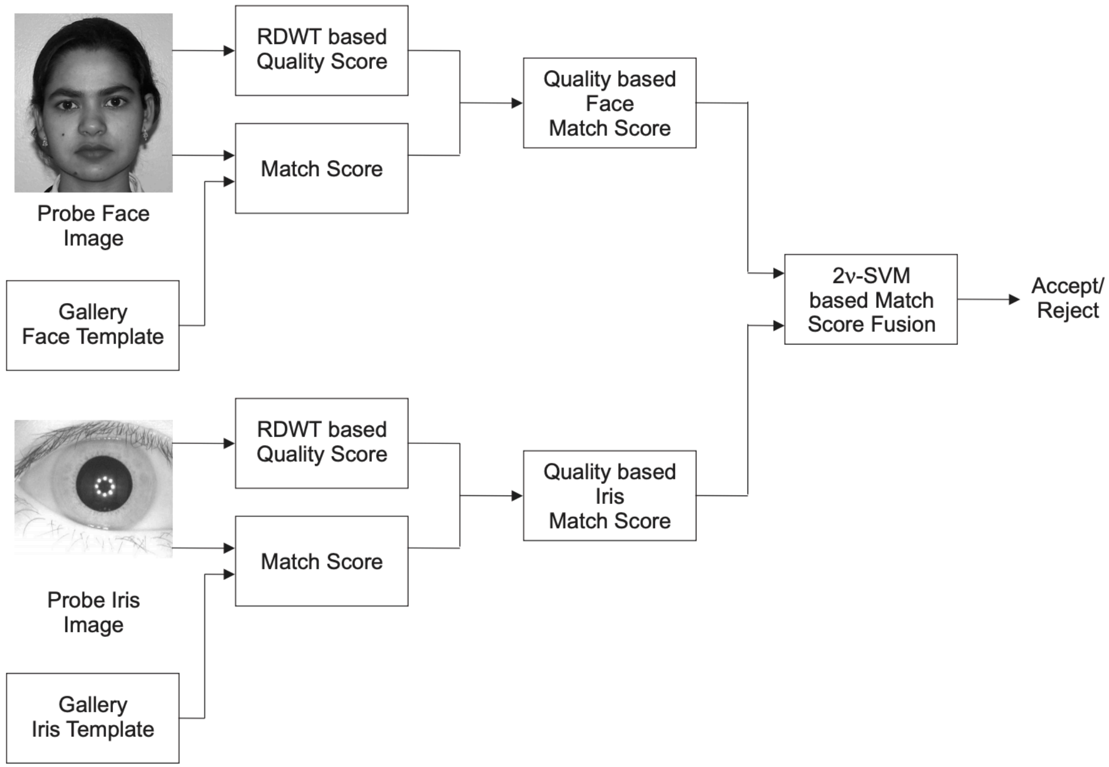
- Normalized sum rule

- A(0, 1) and B(0, 1) and they both have to be either distance score or similarity score

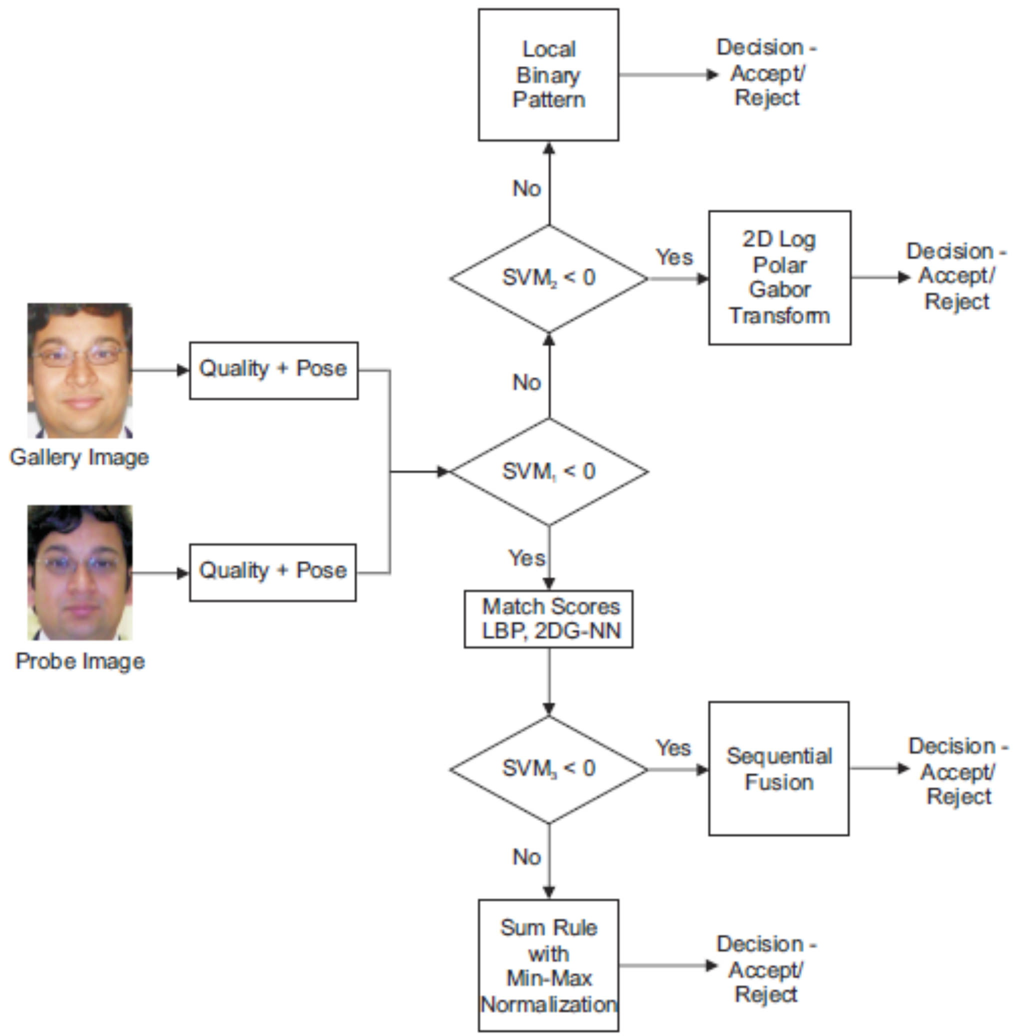
- $C = (A + B)/2$

Match Score Fusion

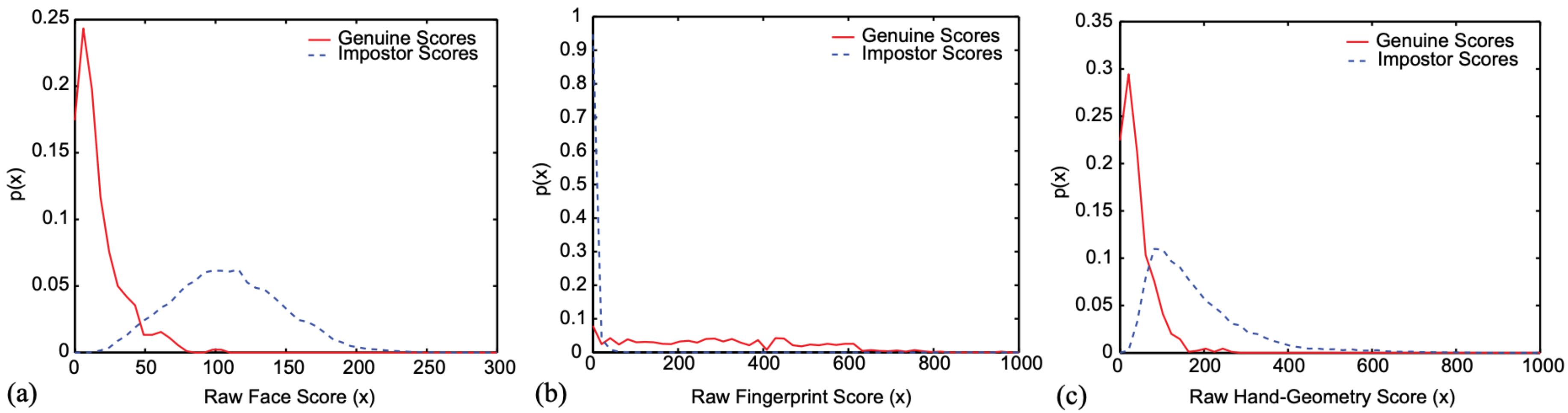
- SVM Fusion – Another powerful and widely used match score fusion rule
- How can we perform SVM based match score fusion?



Dynamic Selection



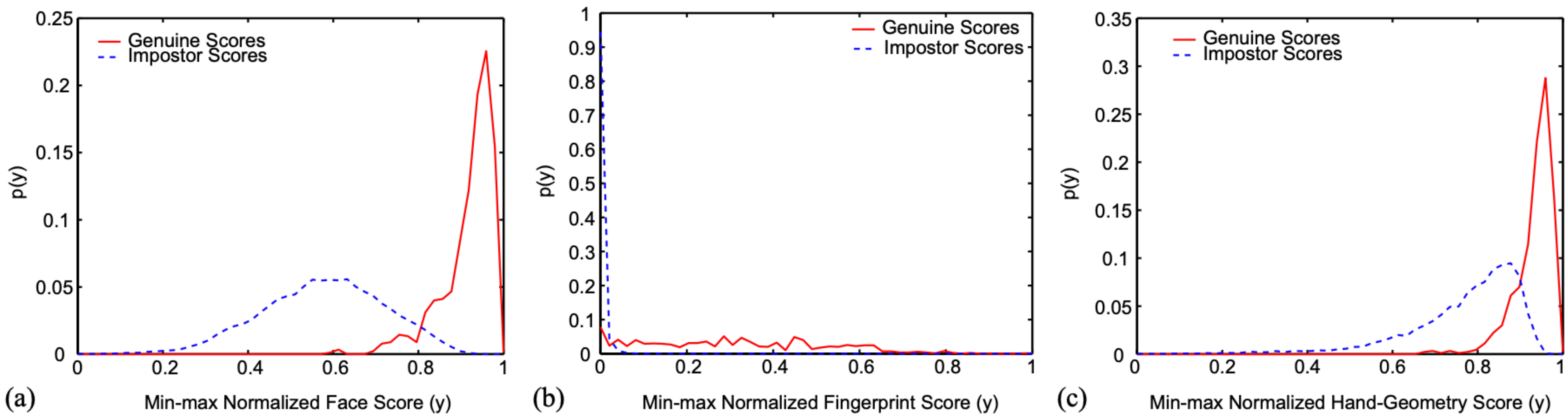
Score Distributions



Score normalization in multimodal biometric systems
A Jain, K Nandakumar, A Ross
Pattern recognition 38 (12), 2270-2285

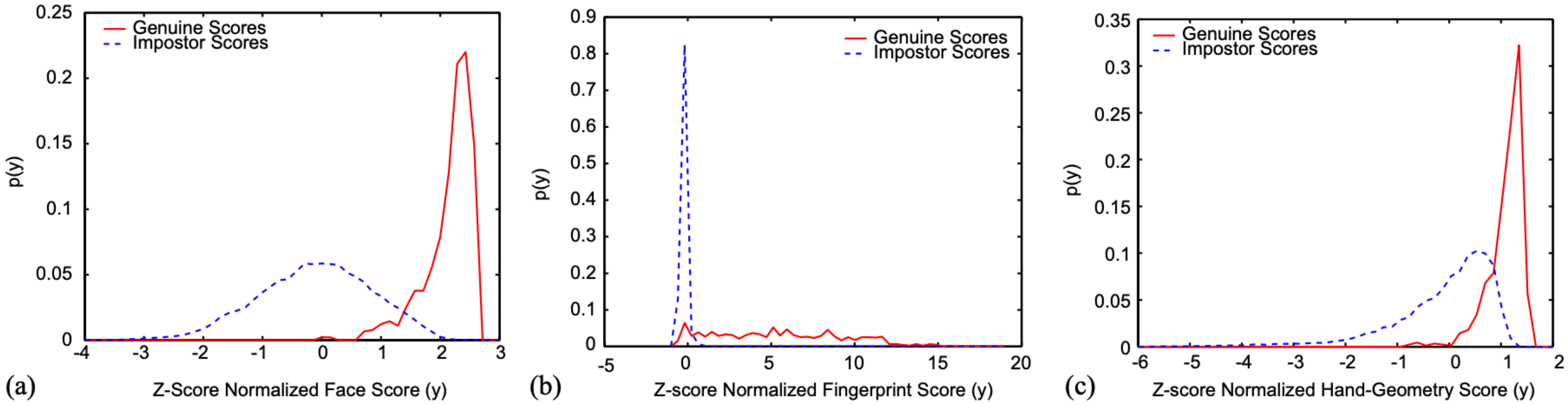
Min-max Normalization

$$s'_k = \frac{s_k - \min}{\max - \min}$$



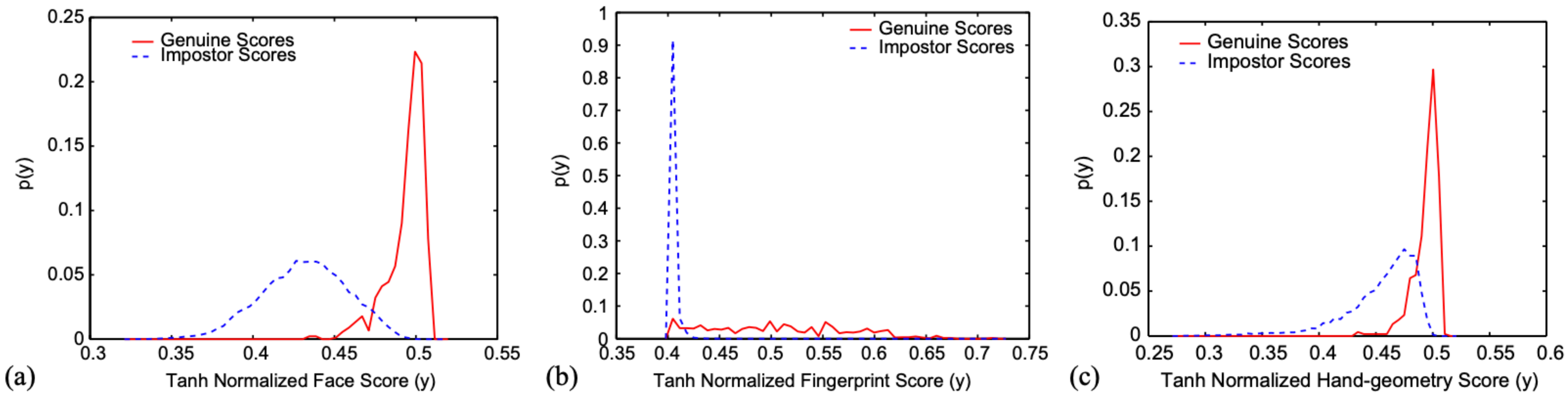
Z-score Normalization

$$s'_k = \frac{s_k - \mu}{\sigma}$$

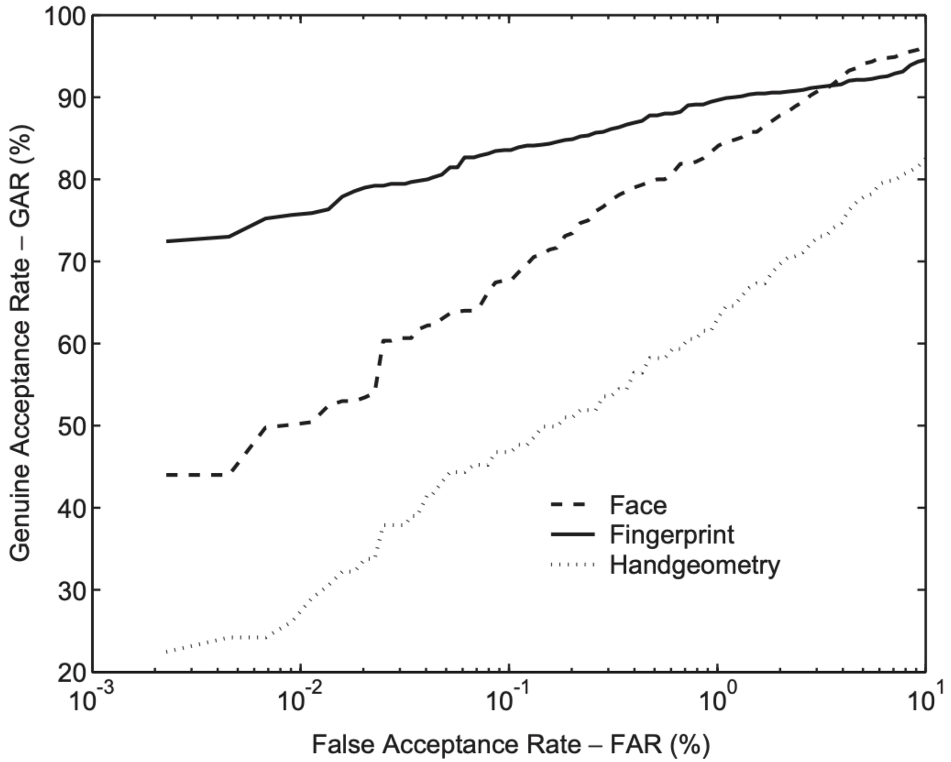


Tan-h Estimators

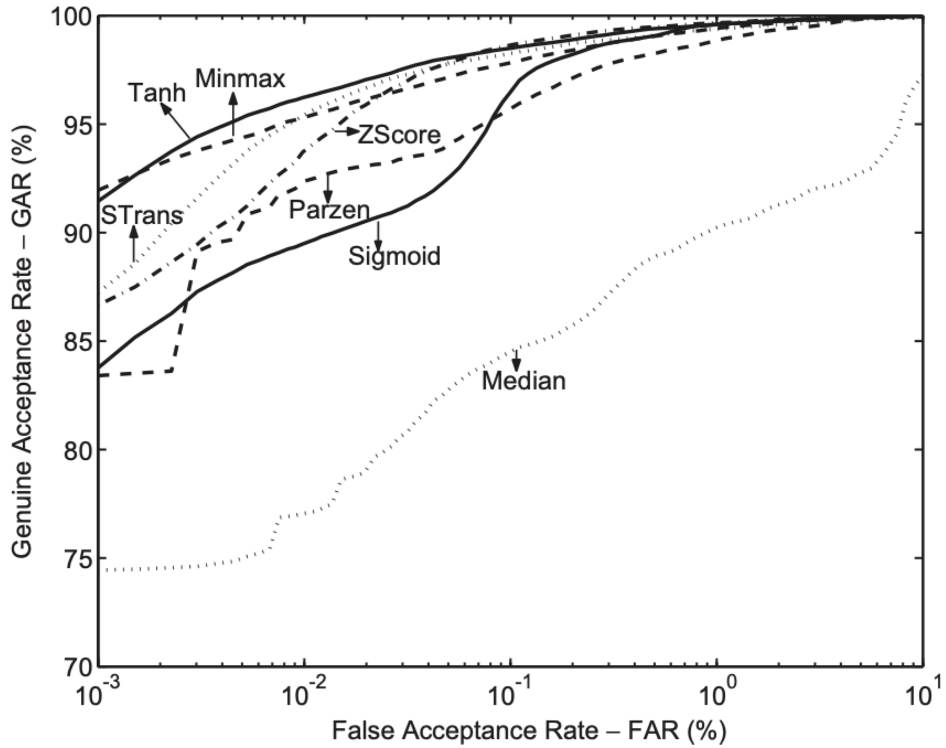
$$s'_k = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{s_k - \mu_{GH}}{\sigma_{GH}}\right)\right) + 1\right\},$$



ROC with Score Fusion

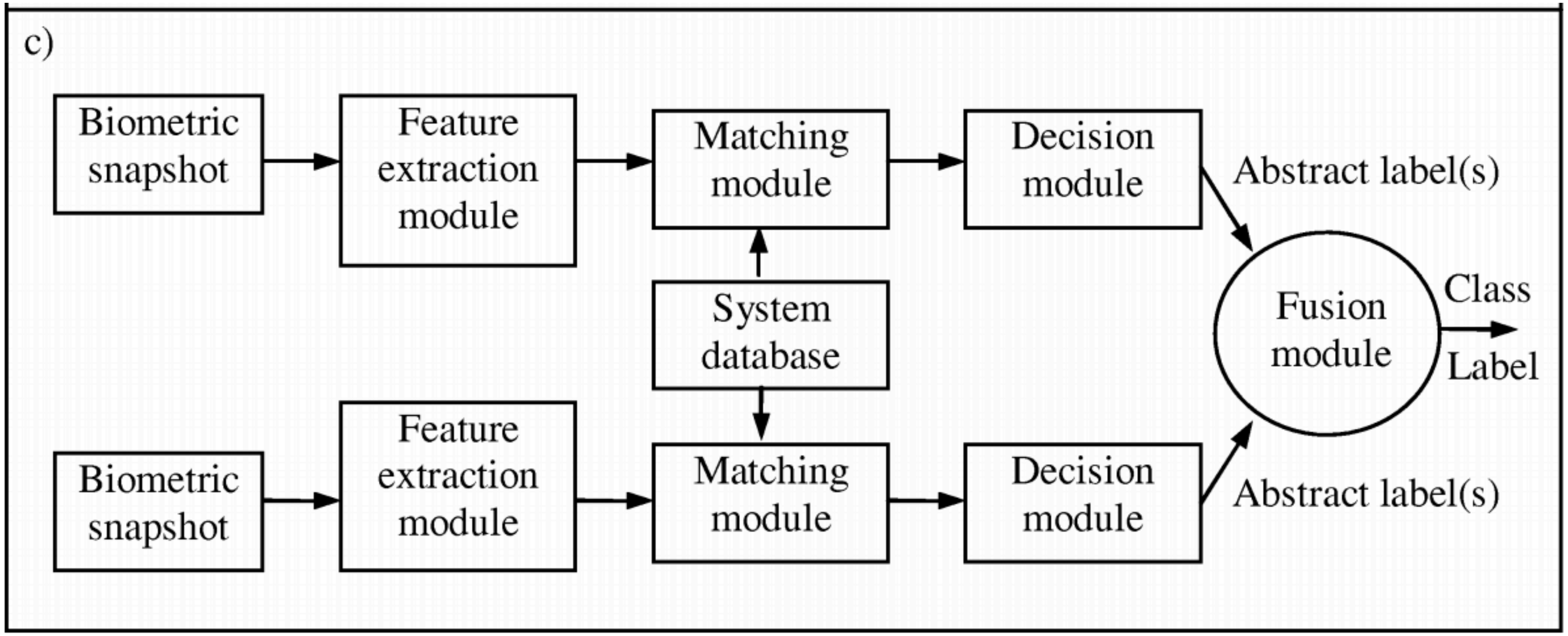


Individual Modalities



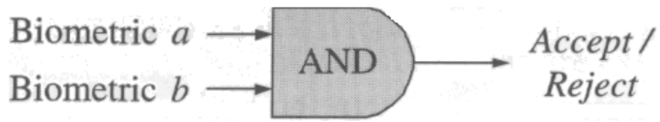
Score Fusion

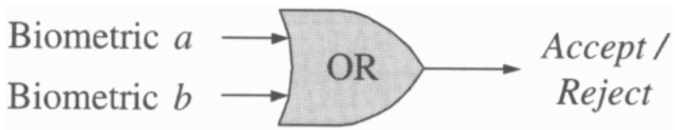
Fusion at decision (abstract label) level



Decision Level Fusion - Boolean Combination

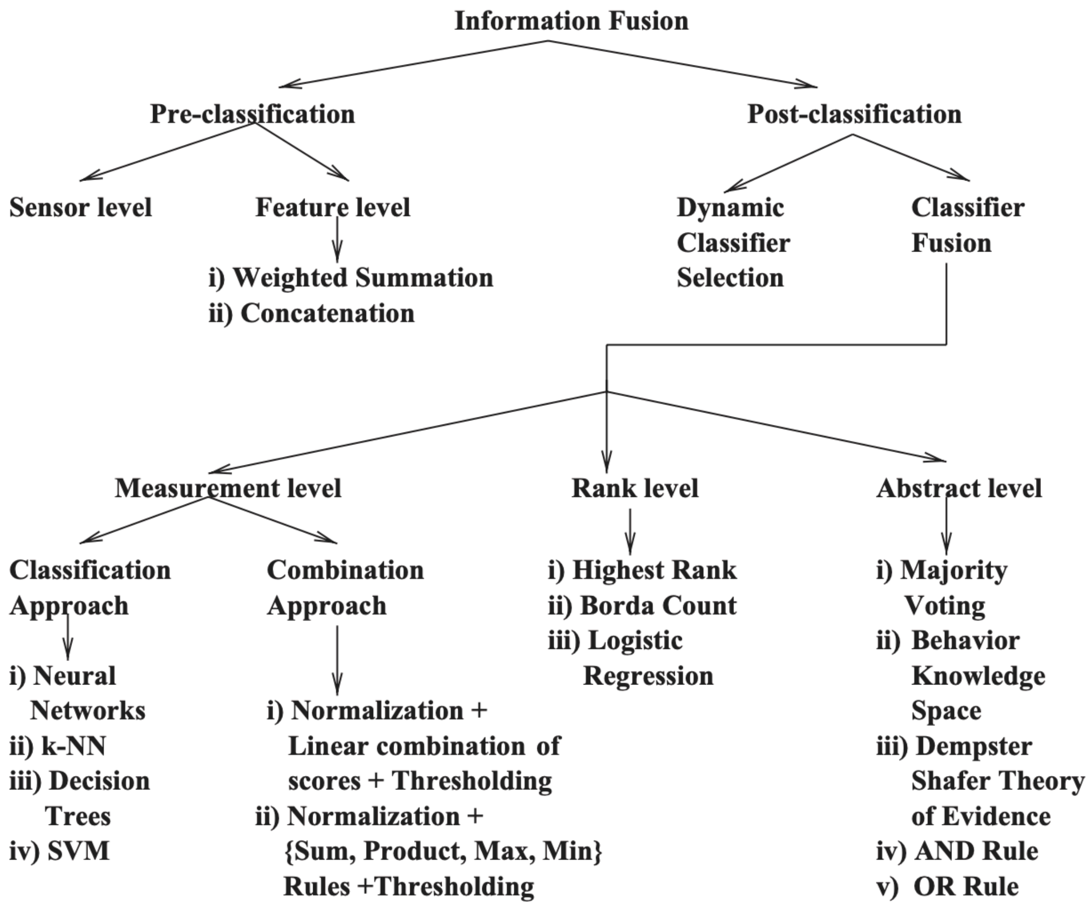
- The **AND** rule and the **OR** rule:
  - the most prevalent rules for multiple biometrics combination in practical systems.





  - Used in both **Identification** and **Verification** protocol.

Before Deep Learning Era…



Multimodal Deep Learning

- Major challenges in multimodal deep learning

Representation

Alignment

Fusion

Translation

Co-Learning

Multimodal Deep Learning

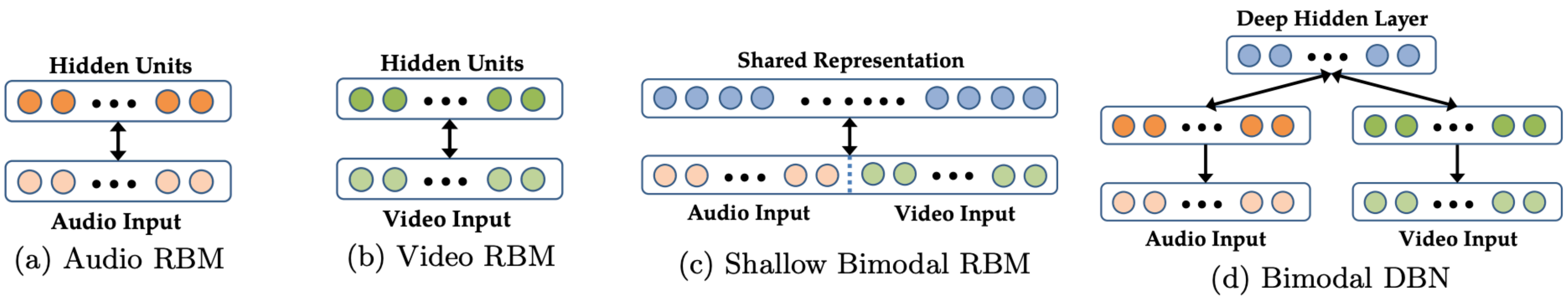
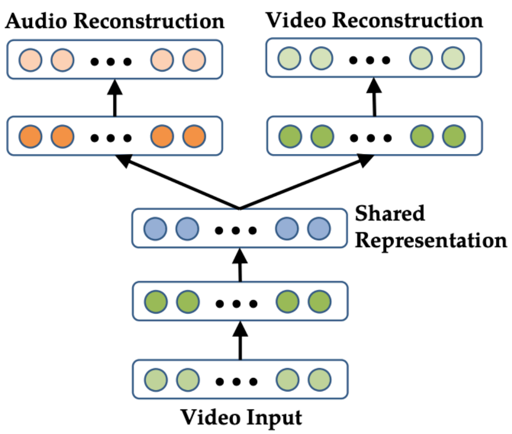


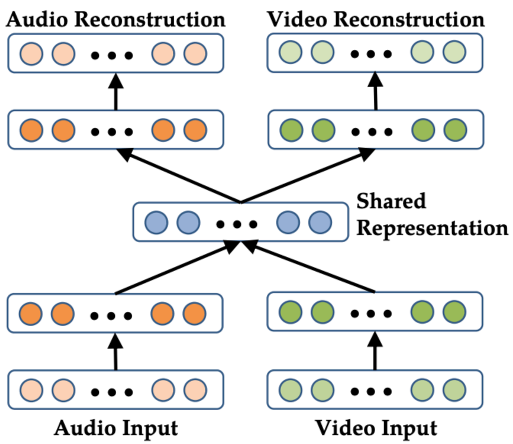
Figure 2: RBM Pretraining Models. We train RBMs for (a) audio and (b) video separately as a baseline. The shallow model (c) is limited and we find that this model is unable to capture correlations across the modalities. The bimodal deep belief network (DBN) model (d) is trained in a greedy layer-wise fashion by first training models (a) & (b). We later "unroll" the deep model (d) to train the deep autoencoder models presented in Figure 3.

https://people.csail.mit.edu/khosla/papers/icml2011_ngiam.pdf

Multimodal Deep Learning



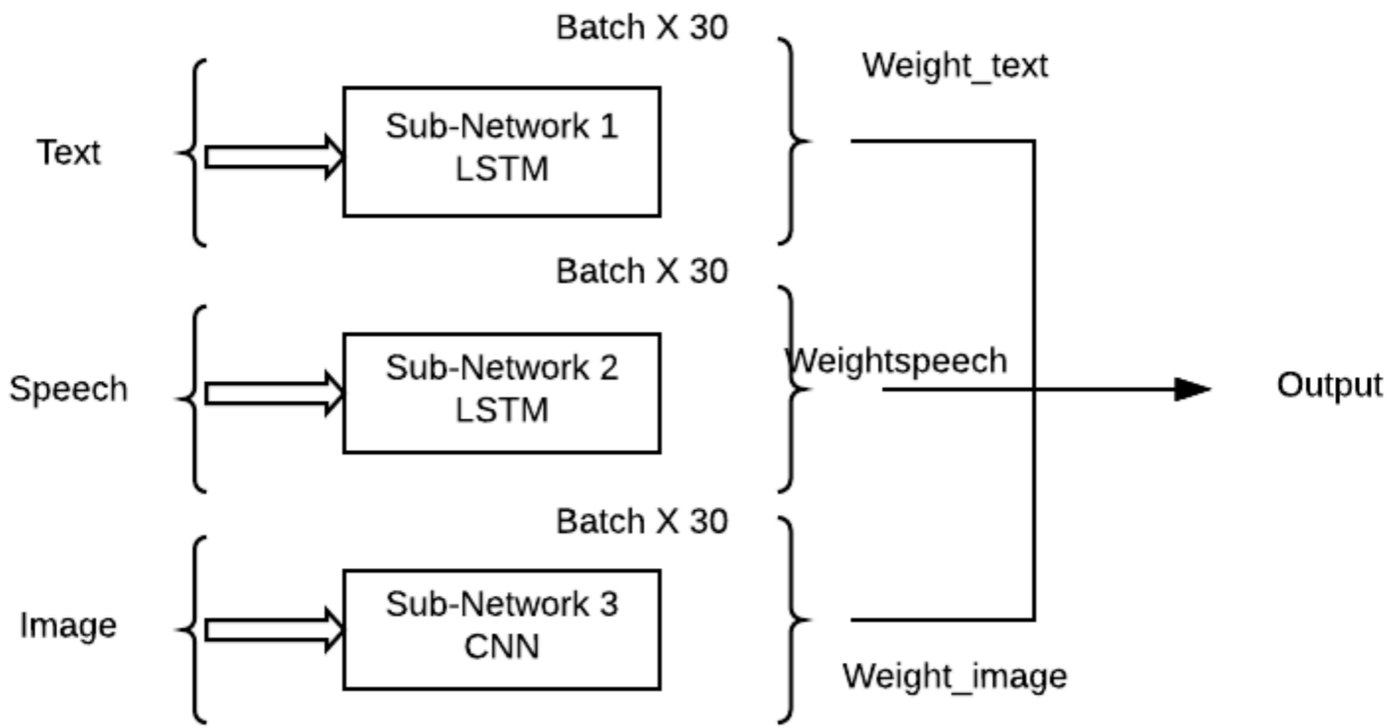
(a) Video-Only Deep Autoencoder



(b) Bimodal Deep Autoencoder

Figure 3: Deep Autoencoder Models. A "video-only" model is shown in (a) where the model learns to reconstruct both modalities given only video as the input. A similar model can be drawn for the "audio-only" setting. We train the (b) bimodal deep autoencoder in a denoising fashion, using an augmented dataset with examples that require the network to reconstruct both modalities given only one. Both models are pre-trained using sparse RBMs (Figure 2d). Since we use a sigmoid transfer function in the deep network, we can initialize the network using the conditional probability distributions $p(\mathbf{h}|\mathbf{v})$ and $p(\mathbf{v}|\mathbf{h})$ of the learned RBM.

Multimodal Deep Learning



https://towardsdatascience.com/multimodal-deep-learning-ce7d1d994f4

Other Resources on Deep Learning

- https://sites.google.com/site/multiml2016cvpr/
- Multimodal Machine Learning: Integrating Language, Vision and Speech by Louis-Philippe Morency, Tadas Baltrušaitis https://www.cs.cmu.edu/~morency/MMML-Tutorial-ACL2017.pdf
- https://telecombcn-dl.github.io/2019-mmm-tutorial/
- https://www.mitpressjournals.org/doi/full/10.1162/neco_a_01273

- https://paperswithcode.com/task/multimodal-deep-learning