CBSA ASFC

PROTECTION ▪ SERVICE ▪ INTEGRITY   PROTECTION ▪ SERVICE ▪ INTÉGRITÉ

Evaluation of Real-Time Face Recognition Technologies for Video Surveillance Applications

Dmitry Gorodnichy and Eric Granger

Science and Engineering Directorate
Video Surveillance and Biometrics Section

Outline

- **PROVIT –** evaluate state-of-the-art commercial technologies and academic systems for FRiVS:
  - public data sets for medium- to large-scale evaluation
  - experimental protocols for different still-to-video and video-to-video surveillance applications, e.g.,
    - screening of faces according to their resemblance to a wanted list
    - matching a face across several video feeds
    - fusion of face recognition from different cameras while tracking a person
  - performance measures: transaction-based (P-R curve) and subject-based (biometric menagerie) analysis

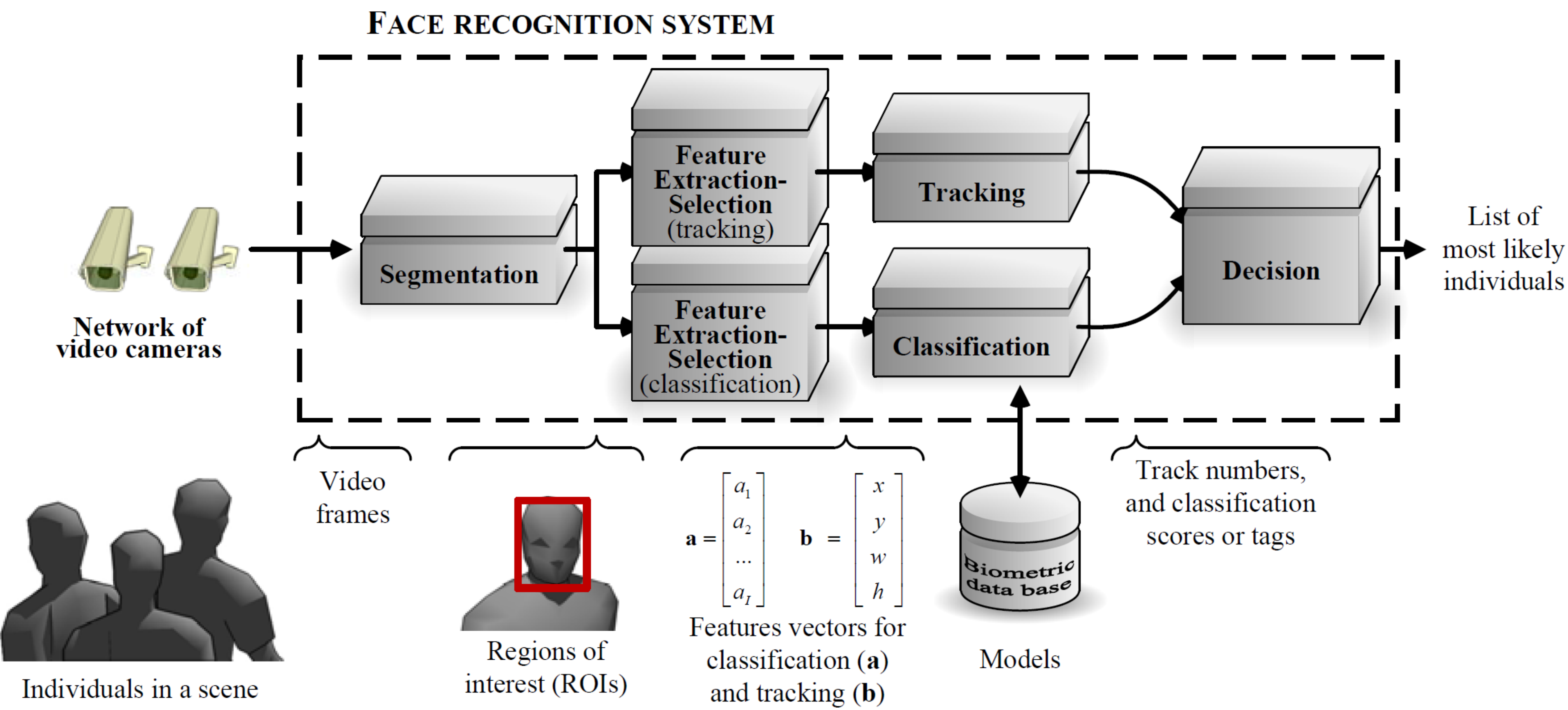
Outline

1. **Background – Face Recognition in Video Surveillance**
   - objectives and challenges
   - where biometrics meets video surveillance
   - academic and commercial solutions

2. **Evaluation of Systems for FRiVS**
   - publicly-available data sets and lab mock-up
   - specialized performance metrics and protocols
   - CBET: multi-order score analysis, threshold-validated analysis

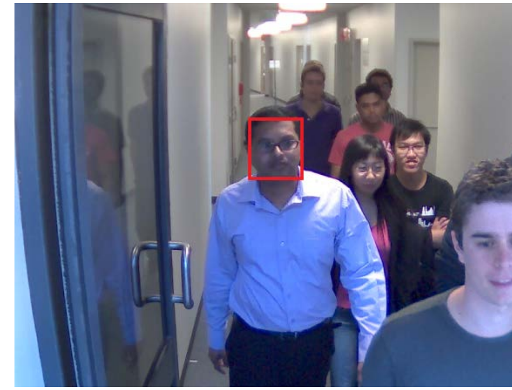
3. **TRL-based evaluation**
   - Issues with conventional performance evaluations
   - Integrating FR into operational CCTV environment
   - PROVE-IT (FRiV) methodology & results
   - Preliminary TRL assessment

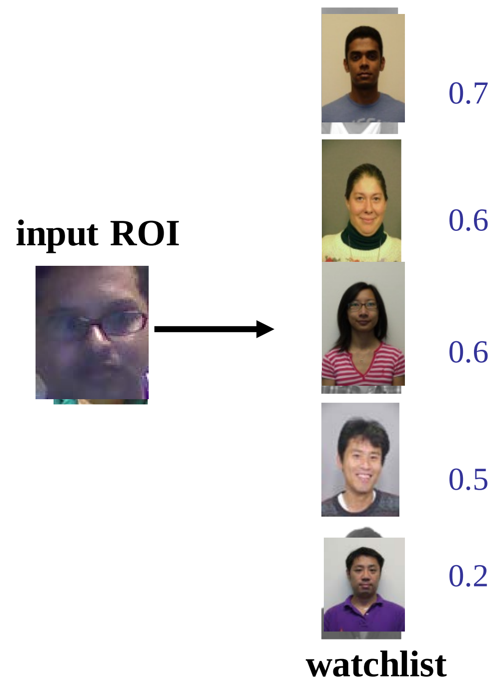
A Generic System for FR in Video



1) Face Recognition in Video Surveillance

- **Enhanced screening and situation analysis across a network of surveillance cameras**
  - automatically recognize and track individuals within dense and moving crowds, as found at major events and airports
  - determine if faces captured in video streams correspond to individuals of interest populating a restrained list of individuals





CBSA ASFC

Problem statement

- **ROIs extracted from video frames (probes) are matched against facial model of individual of interest**

- **Still-to-video recognition**: facial model of each individual consists of 1+ templates extracted from a gallery of stills

  Typical CBSA application: watchlist-based surveillance

- **Video-to-video**: facial model of each individual are extracted from videos

  Typical CBSA application: operator captures an individual of interest in a video stream and the system tracks him over a network of cameras

1) Challenges of FR in Video Surveillance

CBSA ASFC

- **Environments are complex and change over time due to:**
  - low quality and resolution of video frames
  - limited control of acquisition conditions – variation in poses, expressions, illumination, cooperation of individuals, occlusion…
  - inter- and intra-class variability and noise in the feature space
  - ageing and variation of interaction between sensor and individual
  - facial models are often poor representatives of real faces
  - highly skewed data distributions: very few positives (from individuals of interest) w.r.t. negative samples (from open world)

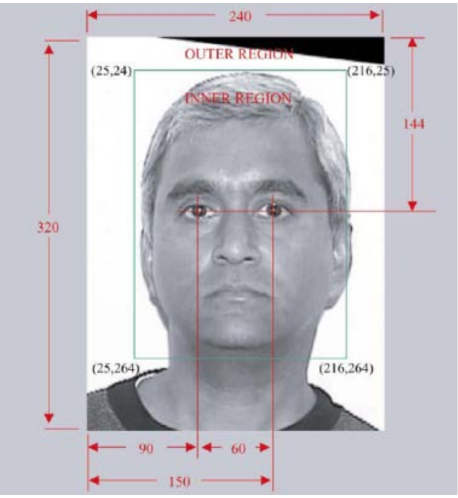
- **Computational resources** – video surveillance networks are comprised of a growing number of IP-based cameras
  - transmit or archive massive quantities of data
  - memory requirements: storage and retrieval of facial models
  - processing time: matching ROIs against facial models

1) Face Recognition in Video Surveillance

CBSA ASFC

Biometric Setup

- **Faces captured in *controlled environment* (as in e-Gates with e-Passport) are much easier to recognize**
- **Still images captures from these environments may provide:**
  - canonical face model adopted by ICAO’02 for passport-type documents
  - high resolution (60 pixels between eyes)
  - well positioned face (front-faced, eye-level) without occlusion (eye-glasses, scarf)
  - neutral facial expression
  - high quality:
    - no motion, blur, compression artifacts, etc
    - in focus
    - best possible illumination



1) Challenges of FR in Video Surveillance

CBSA ASFC

Taxonomy of Surveillance Setups

Type 0: Cooperative Biometric setup (access control, eGate)

Type 1: semi-constrained setup
- primary inspection lane (PIL)

Type 2: unconstrained free-flow, one-at-time
- port of entry / chokepoint entry

Type 3: unconstrained free-flow, many-at-time
- airport

Type 4: Outdoor (no lighting or structural constraints)

2

3

1

Survey of Academic Solutions

Author	Description	Recognition	Set	Tracking	Applications
Beveridge	CSU Ellastic Graph Bunch Matching	Still-to-video, local	Both	No	watch list screening
Zhou 2003	Simultaneous Face Tracking and Recognition	Still-to-video, video-to-video, hollistic	Closed	Yes	access control
Ekenel 2007	Local Appearence-Based Face Models	Video-to-video, hollistic	Open	No	access controll
Stallenkamp 2008	Local Appearence-Based Face Models	Video-to-video, hollistic	Open	No	watch list screening
Kamgar-Parsi 2011	Face Morphing to Boost Training Data	Still-to-video, local	Open	No	watch list screening
Li 2005	TCM-kNN	Still-to-still, hollistic	Open	No	watch list screening
Connolly 2010	**Evolving ensembles using Dynamic PSO**	**Video-to-video, holistic**	**Closed**	**No**	**access control**
Pagano 2011	**Adaptive Ensemble of Detectors**	**Video-to-video, hollistic**	**Open**	**Yes**	**watch list screening**

1) Face Recognition in Video Surveillance

CBSA ASFC

Survey of Commercial Technologies

Technology	Vendor	Type	Track	Approach	Applications
Verilook Surveillance SDK	Neurotechnology	SDK	Multiple	Still-to-video, video-to-video	Face anotation, watch list screening, enrollment from video, multi-modal biometrics
FaceR	Animetrics	SDK	No	Still-to-still	Watch list screening, enrollment from video
FaceIT SDK	L1	SDK	No	Still-to-still	Watch list screening, multi-modal biometrics
PittPatt SDK	Google*	SDK	Multiple	Still-to-video, video-to-video	Face anotation, watch list screening, enrollment from video
FaceVACS	Cognitec	SDK	Multiple	Still-to-video, video-to-video	Face anotation, watch list screening, enrollment from video
Acsys FRS SDK	Acsys	SDK	Multiple	Video-to-video	Face anotation, watch list screening, enrollment from video
SureMatch 3D	Genex	App	No	Still-to-still	Watch list screening
Notiface II	FACE-TEK	App	No	Still-to-still	Watch list screening
Face First	Face First	App	No	Still-to-video	Watch list screening

2) Evaluation of Systems for FRiVS

CBSA ASFC

Public Data Sets for for medium- to large-scale evaluation

DATASET	TARGET APPLICATIONS
CMU MOBO: [GRO01] Carnegie Mellon University Motion of Bodies	subjects performing different walking patterns on a treadmill
CMU FIA: [GOH05] Carnegie Mellon University Faces in Action	subjects mimicking passport checkpoint at airport
Chokepoint [WON11]	video-surveillance subjects walking though portals
MOBIO: [MCC10] EC FP7 Mobile Biometry	m-modal unconstrained authentication on mobile device
ND-Q0-Flip: [BAR11] Notre-Dame Crowd Data	detection of questionable observers that appear often in crowd videos
NIST-MBGC: [PHI09] National Institute of Standards and Technology - Multiple Biometric Grand Challenge	m-modal verification of subjects walking through portal or access control checkpoint (still- and video-to-video)
NRC-IIT: [GOR05] National Research Council – Institute for Information Technology	user identification for secured computer login
XM2VTS: [MAT03] Multi-Modal Verification for Teleservices and Security Applications	multi-modal verification for tele-service and security

CBSA ASFC

Data sets for FRiV - summary

- **datasets have been characterized according to:**
  - **demographics:** distribution of individuals per session and in the entire dataset;
  - **complexity in scene:** the systematic variations of illumination, motion, occlusion, expression and/or pose for some target application;
  - **capture properties**: the number and type of cameras, duration of video sequences, frame rate and resolution.

CBSA ASFC

CMU – FIA (mono-modal, 1 face)

- **PIL**: subjects mimicking passport checkpoint at airport

CBSA ASFC

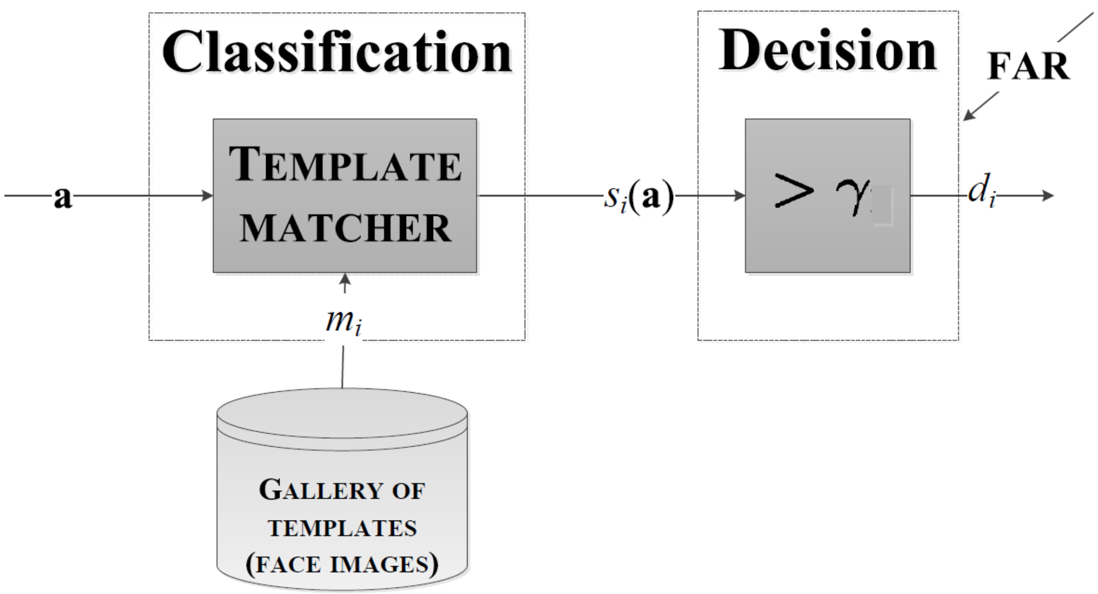
Chokepoint (mono-modal, 1 to 24 faces)

- **CATSA checkpoint**: subjects walking though portals

CBSA ASFC

Performance metrics

- **Fundamental task under evaluation:**
  - independent, user-specific detection of an individual of interest among a restrained cohort of individuals
  - data from a restrained cohort ≠ universal world model



Performance Metrics

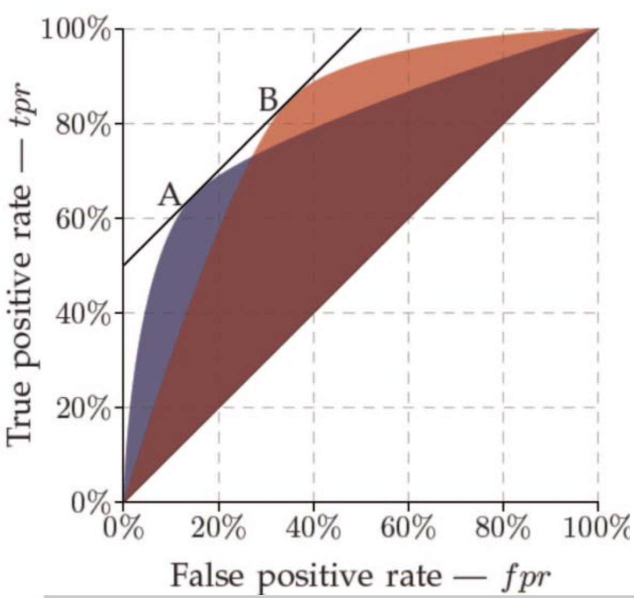
- **Open-set FR problem with imbalanced class distributions (few positive samples from a restrained cohort)**
  - precision-recall space, and F-scores for transaction-based analysis
- **Complex environment and uncertainty of facial models**
  - quality of acquired ROIs and tracks
  - test for confidence or significance on quality estimates
- **Performance varies across a population of individuals, and some individuals are harder to recognize**
  - menageries – statistical tests to characterize individual
- **Growing complexity of surveillance networks**
  - analysis of time and memory complexity

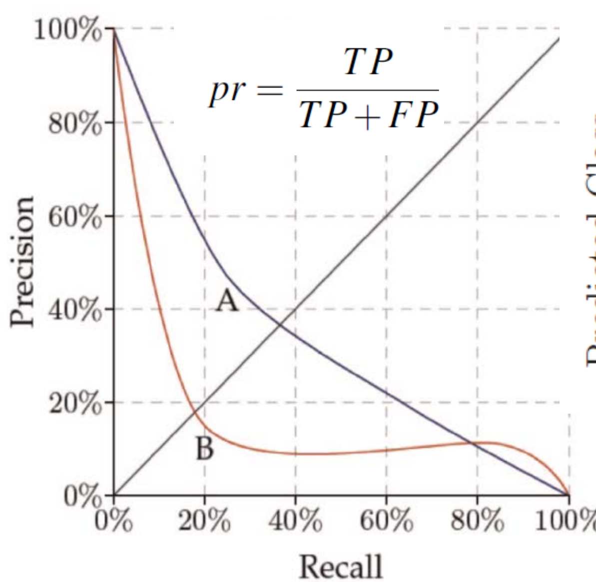
2) Evaluation of Systems for FRiVS

CBSA ASFC

Transaction-Based Analysis

- **Evaluation of detectors** – count correct and incorrect decisions over a test set, and express performance trade-offs using a curve or scalar metric
  - **Traditional:** ROC or DET curves (accuracy, AUC, pAUC)
  - **Dependent on class distributions and miss-classification costs:** precision-recall (F-score), ROC isometrics, cost curves and others





True Class

Predicted Class	p	n
$\hat{p}$	True Positive (TP) Correct detection	False Positive (FP) Type I error
$\hat{n}$	False Negative (FN) Type II error	True Negative (TN) Correct rejection
	$P = TP + FN$	$N = FP + TN$

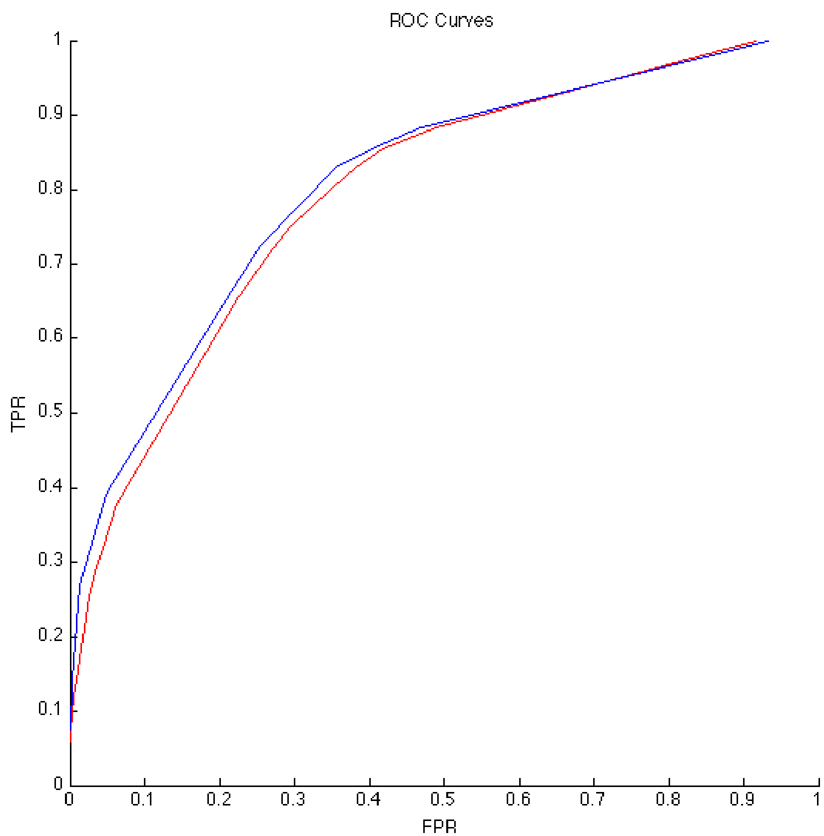
$$F_\beta = (\beta^2 + 1)\frac{tpr \cdot pr}{\beta^2 \cdot pr + tpr}$$

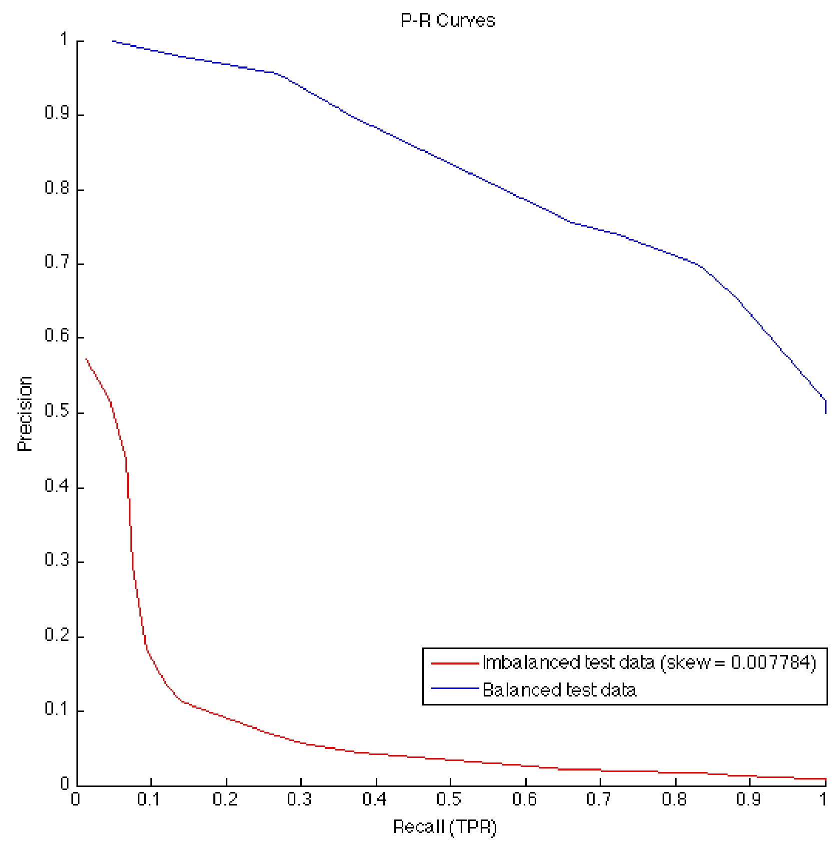
2) Evaluation of Systems for FRiVS

CBSA ASFC

Transaction-Based Analysis

- Results on FIA data





CBSA ASFC

Subject-Based Analysis

- **Doddington's zoo** – performance is assessed with different types of individuals in mind
  - performance of face recognition systems may vary drastically from one individual to the next
  - an analysis of these individuals and their common properties can:
    - expose fundamental weaknesses in a biometric system
    - schemes for user-specific thresholds, score normalization, and fusion



2) Evaluation of Systems for FRiVS

CBSA ASFC

Transaction-Based Analysis

- **Results on FIA data: pAUC(10%) and F_1-measure (Pagano at al. IEEE WCCI 2012)**

Architecture	Individuals									
	2	23	58	106	147	151	176	188	190	209
Global	0.37 ± 0.038	0.58 ± 0.095	0.68 ± 0.12	0.94 ± 0.036	0.42 ± 0.13	0.71 ± 0.11	0.73 ± 0.05	0.81 ± 0.076	0.53 ± 0.065	0.9 ± 0.068
Modular	0.35 ± 0.04	0.64 ± 0.15	0.85 ± 0.04	0.84 ± 0.075	0.69 ± 0.13	0.85 ± 0.035	0.61 ± 0.054	0.84 ± 0.06	0.66 ± 0.096	0.92 ± 0.054
Modular w. EoDs	0.45 ± 0.036	0.72 ± 0.094	0.89 ± 0.035	0.9 ± 0.069	0.82 ± 0.11	0.91 ± 0.043	0.75 ± 0.054	0.88 ± 0.049	0.7 ± 0.062	0.97 ± 0.024

TABLE I

AVERAGE PAUC ACCURACY FOR 10 INDIVIDUAL IN INTEREST.

CBSA ASFC

Comprehensive Biometrics Evaluation Toolkit (CDET)

- **Developed by CBSA-S&E for evaluation of biometric systems for border control for:**
  1. 1-to-*N* entry control applications, e.g., to investigate the risks of having non-confident matches in iris systems
  2. 1-to-*M* screening applications, e.g., to evaluate stand-off and iFR technologies

- **Integrates best practices and recommendations, such as:**
  - all ISO-SC 37 / NIST metrics
  - Multi-order score analysis
  - subject-based analysis
  - “Non-confident” match analysis for fully-automated systems
  - Threshold-validated ranking analysis
  - Case studies (iris, face, voice)

3. TRL-Based Evaluation

CBSA ASFC

Technology Readiness: Preliminary assessment

FRiV applications	**Type 0 (eGate)**	**Type 1**	**Type 2**	**Type 3**
Face Tracking (in consecutive frames)	√	?	?	–
Face Detection	√	√	√	?
Face Grouping / Tagging (across multiple feeds)	?	?	?	–
Instant “Watch List” Screening	?	?	–	–
Forensic examination from video (off-line)	√	?	?	–
Expression analysis	√	?	?	–
Face + Voice + Iris	√	?	-	–
Video-to-video face matching	√	?	?	–
Soft biometrics (e.g., height)	?	?	?	–

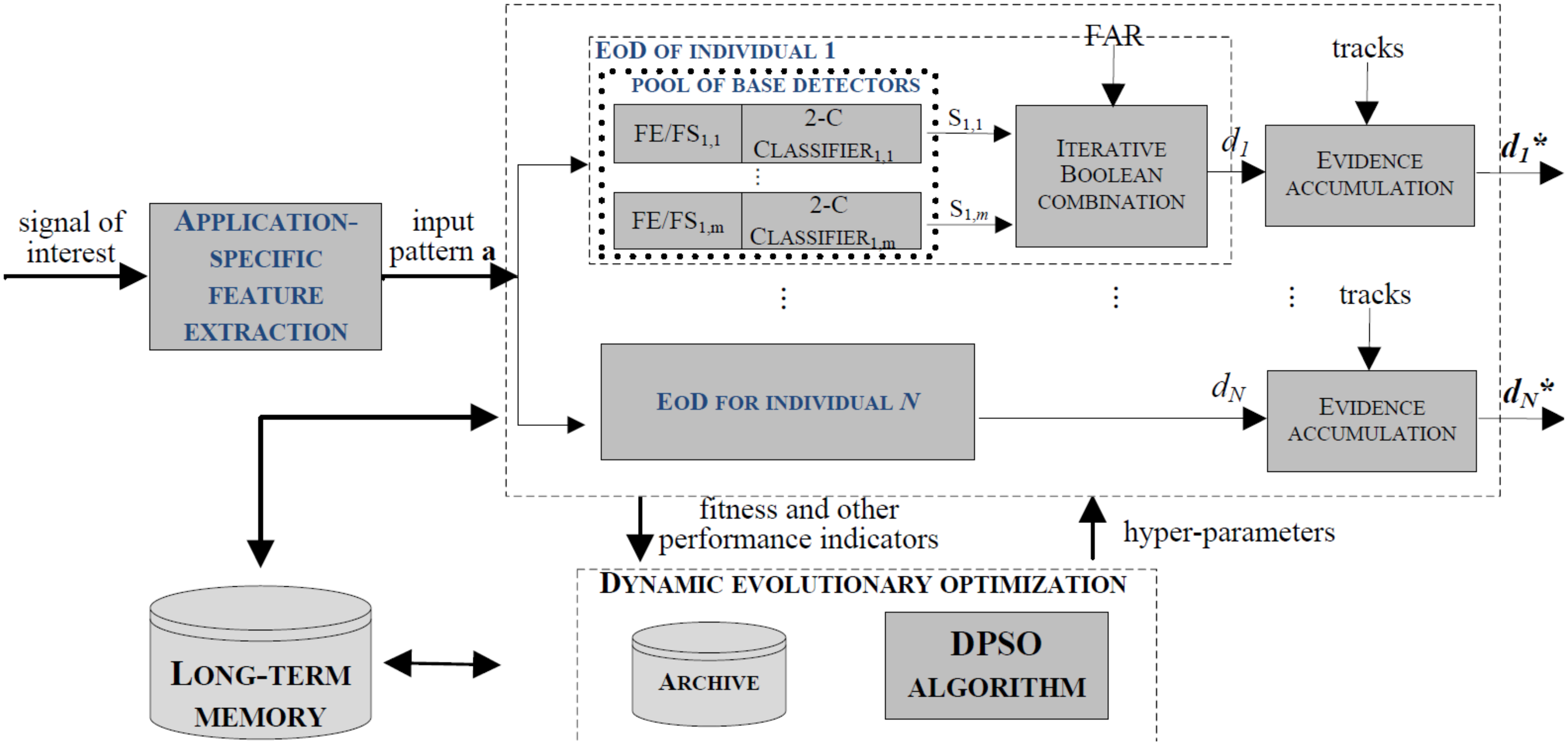
Conclusions

- Current COTS and academic systems can be found useful for some FRiSV applications
- Post-processing and pre-processing (including Video Analytics) are critical for their success
- Potential for new video-based (eg spatio-temporal recognition) techniques, as opposed to status-quo still-image-based.
- There is no all-inclusive evaluation methodology for FRiVS
  - conventional metric can be misleading
  - for operational agency, TRL-based evaluation should prevail
- Ultimate metric - satisfaction of the end-user (border officer)!

Backup Slides

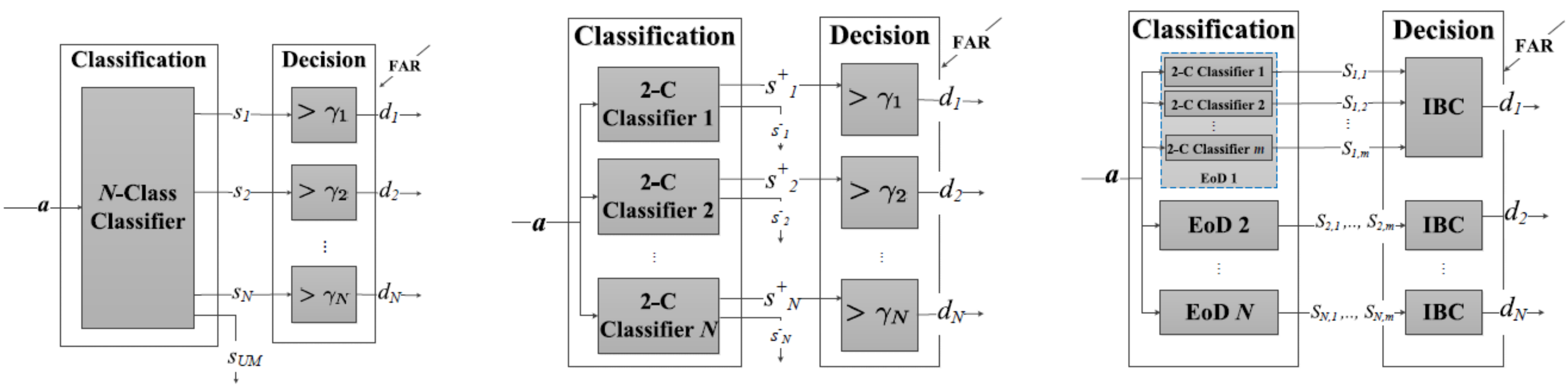
Adaptive Multi-Classifier Systems

(Pagano *et al.*, IEEE WCCI 2012)



Adaptive Multi-Classifier Systems
(Pagano *et al.*, IEEE WCCI 2012)

- **Classification and decision architectures for open-set FR**



(a) monolithic architecture with UM (b) modular architecture (c) modular architecture with EoDs

Adaptive Multi-Classifier Systems

(Pagano *et al*., submitted to IEEE WCCI 2012)

Architecture	Individuals									
	2	23	58	106	147	151	176	188	190	209
Global	0.37 ± 0.038	0.58 ± 0.095	0.68 ± 0.12	0.94 ± 0.036	0.42 ± 0.13	0.71 ± 0.11	0.73 ± 0.05	0.81 ± 0.076	0.53 ± 0.065	0.9 ± 0.068
Modular	0.35 ± 0.04	0.64 ± 0.15	0.85 ± 0.04	0.84 ± 0.075	0.69 ± 0.13	0.85 ± 0.035	0.61 ± 0.054	0.84 ± 0.06	0.66 ± 0.096	0.92 ± 0.054
Modular w. EoDs	0.45 ± 0.036	0.72 ± 0.094	0.89 ± 0.035	0.9 ± 0.069	0.82 ± 0.11	0.91 ± 0.043	0.75 ± 0.054	0.88 ± 0.049	0.7 ± 0.062	0.97 ± 0.024

TABLE I

AVERAGE PAUC ACCURACY FOR 10 INDIVIDUAL IN INTEREST.

Adaptive Multi-Classifier Systems

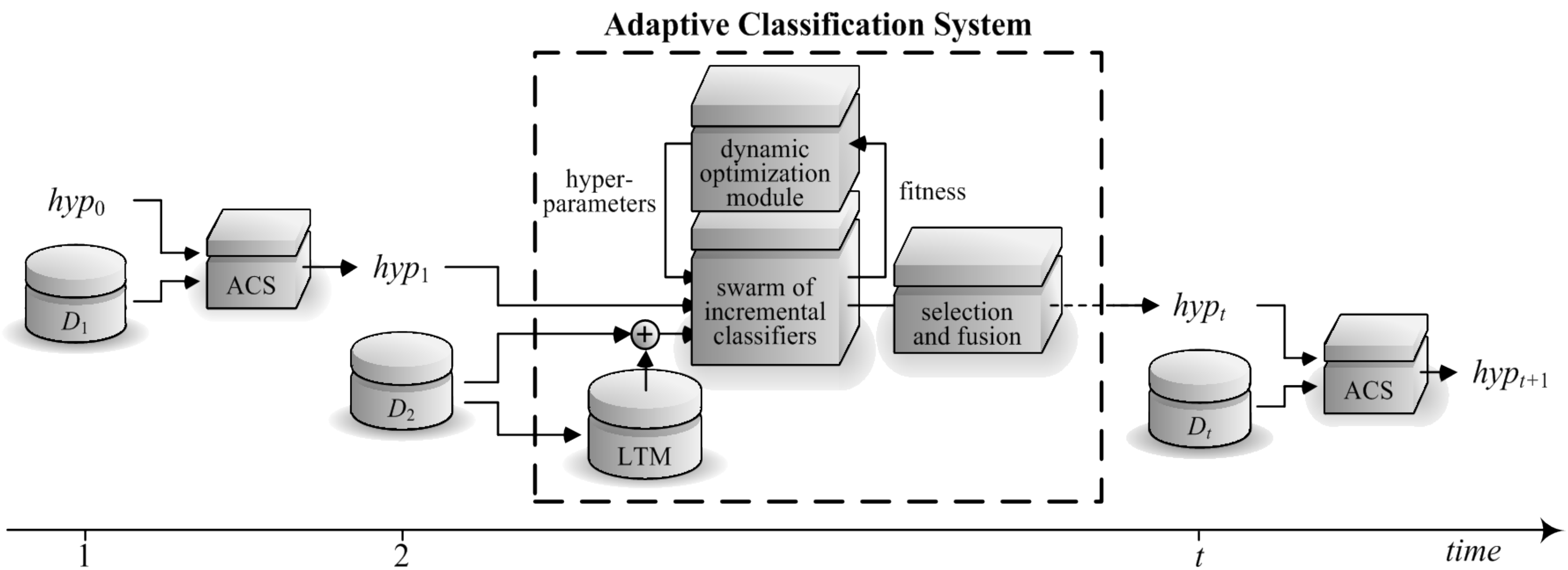
(de la Torre *et al.*, IEEE WCCI 2012)

▶ Incremental learning of new data using L&C

Classifier	Compression	AUC	pAUC-10
		D_1	
k-NN	1.00±0.00	0.9127±0.020	0.4798±0.069
PFAM_{batch}	8.63±1.13	0.9499±0.022	0.7223±0.068
PFAM_{inc}	8.63±1.13	0.9499±0.022	0.7223±0.068
Learn++	**10.43±1.25**	0.8352±0.079	0.5477±0.098
L&C	9.00±1.19	**0.9523±0.024**	**0.7496±0.067**
		D_3	
k-NN	1.00±0.00	0.9398±0.016	0.5933±0.068
PFAM_{batch}	10.52±1.52	0.9512±0.018	0.7125±0.074
PFAM_{inc}	**11.82±1.83**	0.9382±0.022	0.6502±0.074
Learn++	9.46±0.91	0.8422±0.068	0.5080±0.097
L&C	8.22±1.10	**0.9649±0.018**	**0.7957±0.062**
		D_5	
k-NN	1.00±0.00	0.9496±0.013	0.6442±0.060
PFAM_{batch}	11.76±2.24	0.9589±0.014	0.7403±0.065
PFAM_{inc}	**16.07±3.00**	0.9164±0.032	0.5965±0.085
Learn++	9.34±0.89	0.8174±0.069	0.4280±0.089
L&C	7.32±0.79	**0.9732±0.013**	**0.8240±0.058**

Adaptive Multi-Classifier Systems (Connolly *et al.* PR2011)

- **Framework** – a 'swarm' of incremental classifiers, a dynamic optimization module and a LTM:



Adaptive Multi-Classifier Systems

- **Adaptive Fusion:** Incremental Boolean Combination