Datasets for Face Recognition at a Distance (FRAD)

Brendan F. Klare, Ph.D.
Visual Analytics Lab @ Noblis

brendan.klare@noblis.org
Importance of Datasets

- Algorithm = $f$ (data)
  - Biometrics algorithms are a function of the data that they are trained on
  - More data $\rightarrow$ better algorithms

- In order to better understand biometric capabilities in operational settings, data is needed that best replicates the intended scenario

- As new problems and opportunities emerge that can leverage biometrics, it is imperative to have analogous datasets available to train and evaluate such algorithms
Understanding Face Recognition at a Distance

- Identifying persons at a distance is helpful in law enforcement, defense, and intelligence scenarios.
- Provides an understanding of what persons, and how many persons, are present at a location.

- Two primary operational paradigms:
  - Watch list identification
  - Re-identification

- Several challenges may exist:
  - Low resolution
  - Motion blur
  - Heterogeneous matching
  - Atmospheric noise
  - PIE variations
  - Sensors and illuminators

**Watch-list identification:**

- Distant image
- Match
- Watch-list image

**Re-identification:**

- Distant image
- Match
- Distant image
Heterogeneous Face Recognition

Frontal photograph exists for majority of the population. In many scenarios only alternate image modalities or can be captured.

Matching non-photograph face images (probe images) to large databases of frontal photographs (gallery images) is called heterogeneous face recognition (HFR).
Understand Infrared Spectrum

- FRAD often requires infrared capture due to nighttime conditions
- Active sensing bands (requires illumination):
  - Visible
  - NIR (Near infrared)
  - SWIR (short wave infrared)
- Passive sensing bands (does not require illumination):
  - MWIR (medium wave infrared)
  - LWIR (long wave infrared)

Image from:
Face Appearance Across the IR Spectrum

Visible  NIR  SWIR  MWIR

Wavelength (micrometers)

Visible  NIR  SWIR  MWIR

Infrared type

NIR = Near Infrared
SWIR = Short Wave Infrared
MWIR = Medium Wave Infrared (or thermal)
Long Distance Heterogeneous Face Database (LDHF)

- Face recognition database from Korea University containing both visible and infrared imagery from 100 subjects
- Each subject has the following imagery:
  - At a distance:
    - Images captured at 60m, 100m, and 150m outdoors
    - Both visible and near-infrared images collected
    - Both daytime and nighttime collection
  - Reference image:
    - Visible and near infrared image captures at 1m indoors (analogous to id card image)

WVU FRAD Data

- Datasets used in research by WVU [1] and partially collected by the West Virginia High Technology Consortium (WVHTC) Foundation contained at distance imagery

- Relevant datasets from [1]:
  - SWIR at a distance:
    - SWIR DB2 – 50 subjects with SWIR images collected at 50m and 106m with controlled pose variation and minor facial expression variations
    - SWIR DB3 – 16 subjects in an uncontrolled settings with distances ranging from 60 to 400 meters
  - NIR at a distance:
    - 103 subjects with NIR captured at 30, 60, 90, and 120 meters
    - One visible video per subject was collected

UCCS Large Scale Unconstrained Open Set Face Database

- Fully covert FRAD database, with images captured in public environments without subject cooperation
- The most covert dataset available to date
- Details:
  - Contains 6,337 visible images from 308 subjects
  - Images captured at a distance of roughly 100m
  - Dataset is ideal for re-identification scenario

UMD Remote Face Database

- At a distance images captured in an unconstrained environment from cooperating subjects
- Details:
  - Contains 2,106 visible face images from 17 subjects captured between 5m and 250m
  - Images labeled based on clarity, occlusion, and illumination

Pinellas County Sheriff’s Office (PCSO)

- PCSO has provided the research community many unprecedented datasets
- An FRAD dataset was collected previously collected to conduct accuracy experiments using at a distance images as a query against a database of 1.5 million controlled capture images
- Details:
  - 50 subjects total
  - For each subject the following images were collected:
    - 1m controlled capture “mugshot” style image
    - Indoor at a distance of 50, 75, and 100 yards
    - Outdoor at a distance of 50, 75, 100, 125, 150, and 200 yards
Thank you!