Datasets for Face Recognition at a Distance (FRAD)

Brendan F. Klare, Ph.D.

Visual Analytics Lab @ Noblis

brendan.klare@noblis.org

noblis

For the best of reasons

Importance of Datasets

- Algorithm = *f* (data)
  - Biometrics algorithms are a function of the data that they are trained on
  - More data → 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

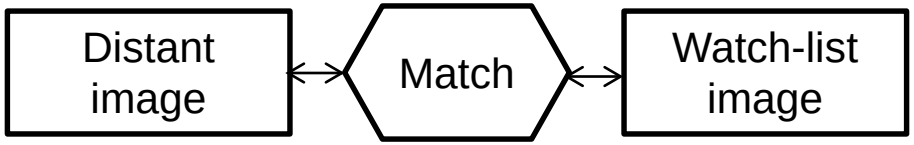
Understanding Face Recognition at a Distance

- 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

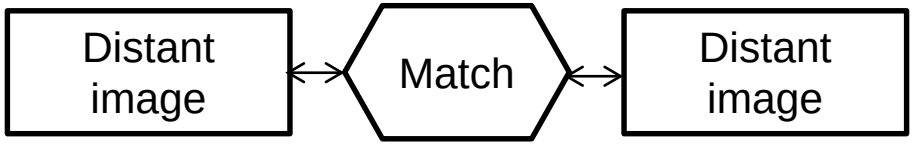
- Two primary operational paradigms:
  - Watch list identification
  - Re-identification

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

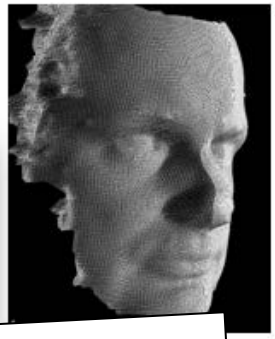
Watch-list identification:



Re-identification:



Heterogeneous Face Recognition

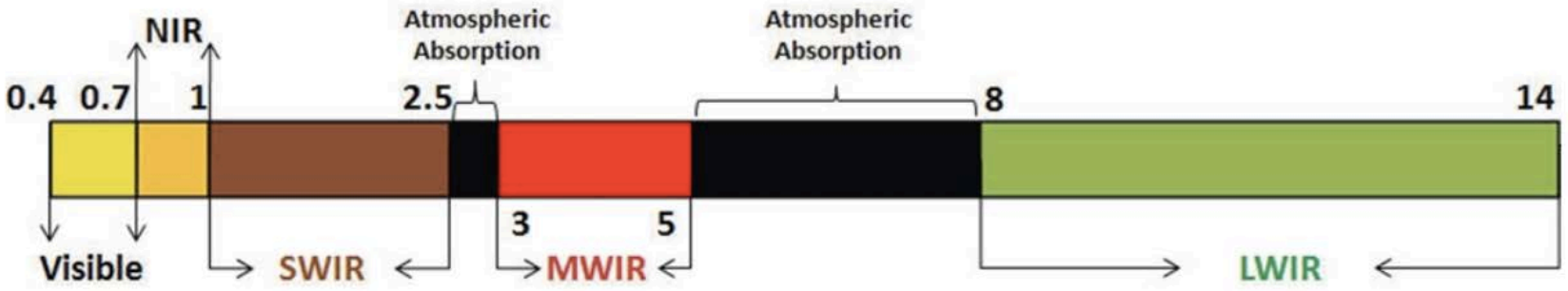


Watch-list identification in FRAD is a heterogeneous face recognition problem

Frontal photograph … nate image for majority of the p… …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)***

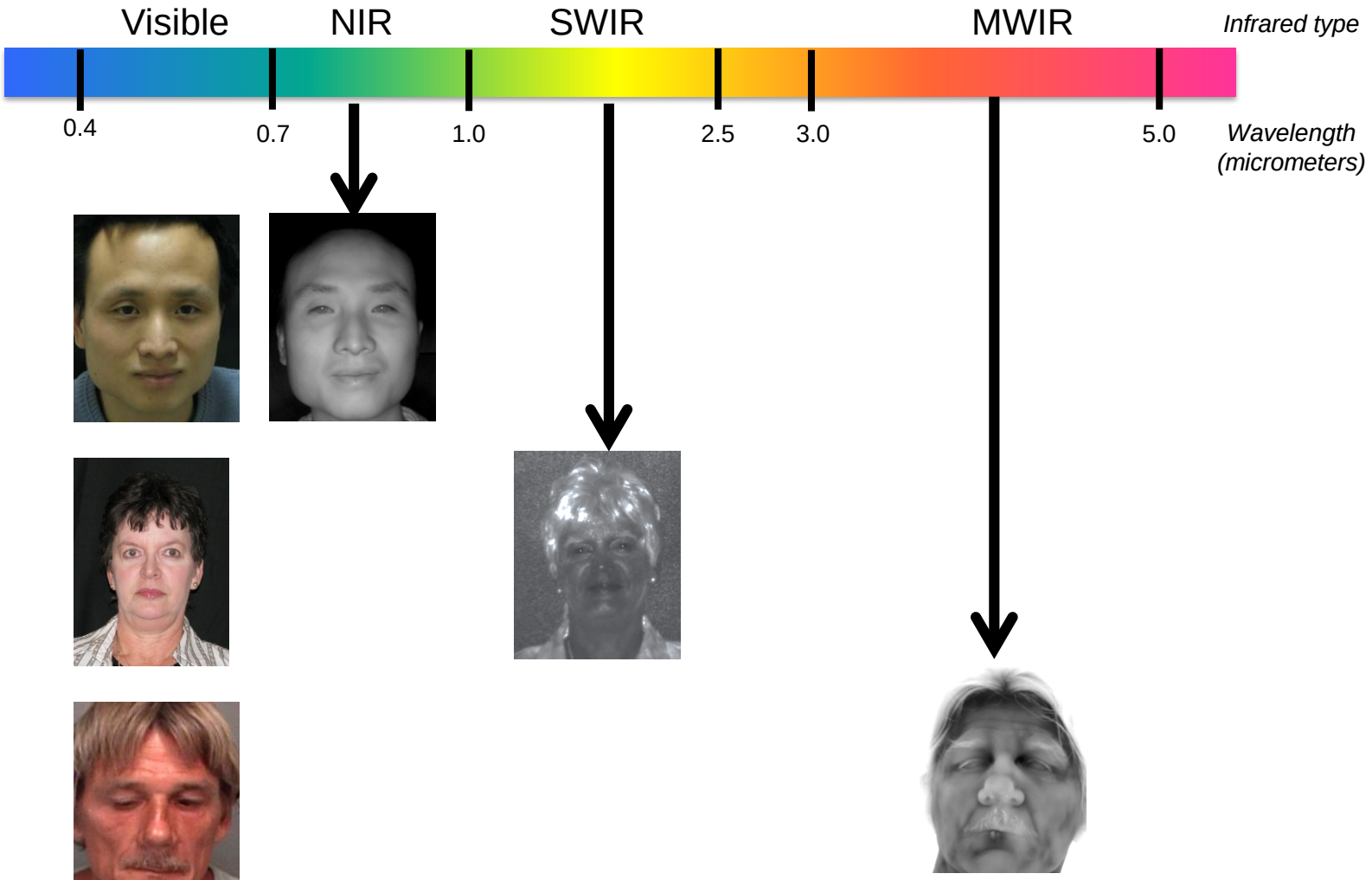
Understand Infrared Spectrum



- 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:
T. Bourlai, B. Cukic, Multi-Spectral Face Recognition: Identification of People in Difficult Environments, in Proc. ISI, 2012, pp. 196-201.

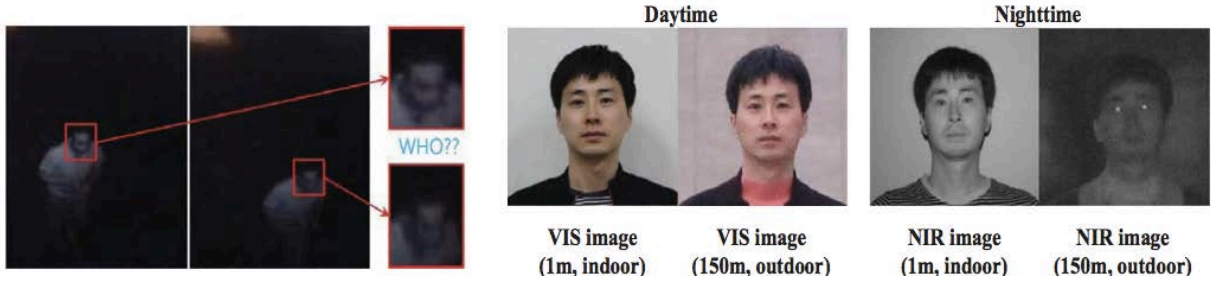
Face Appearance Across the IR Spectrum



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)



D. Kang, H. Han, A. K. Jain, and S. W. Lee, "Nighttime Face Recognition at Large Standoff: Cross-Distance and Cross-Spectral Matching", Pattern Recognition, Vol. 47, No. 12, pp. 3750-3766, December 2014.
http://biolab.korea.ac.kr/database/index.html

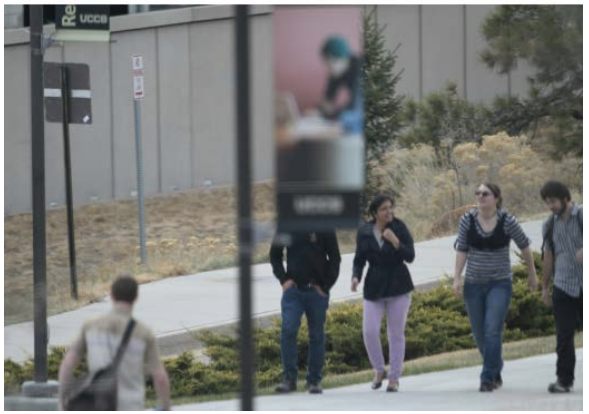
WVU FRAD Data

- 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

T. Bourlai, B. Cukic, Multi-Spectral Face Recognition: Identification of People in Difficult Environments, in Proc. ISI, 2012, pp. 196-201.

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



Sapkota, Archana, and Terrance E. Boult. "Large scale unconstrained open set face database." Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on. IEEE, 2013.

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

Ni, Jie, and Rama Chellappa. "Evaluation of state-of-the-art algorithms for remote face recognition." Image Processing (ICIP), 2010 17th IEEE International Conference on. IEEE, 2010.

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!