ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
ONLINE 10 | 15 January 2021

IAPR
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY

January 11th, 2020

An Adaptive Video-to-Video Face Identification System Based on Self-Training

Eric Lopez-Lopez
Carlos V. Regueiro
Xosé M. Pardo

UNIVERSIDADE DA CORUÑA

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Face Identification in Video-surveillance

Key challenges

- Individual cooperation is costly
- Low-quality video-frames
- Dynamic context
- Labelled data is scarce

Face Identification in Video-surveillance

Key challenges

- Individual cooperation is costly
- Low-quality video-frames
- Dynamic context
- Labelled data is scarce

How can they be addressed?

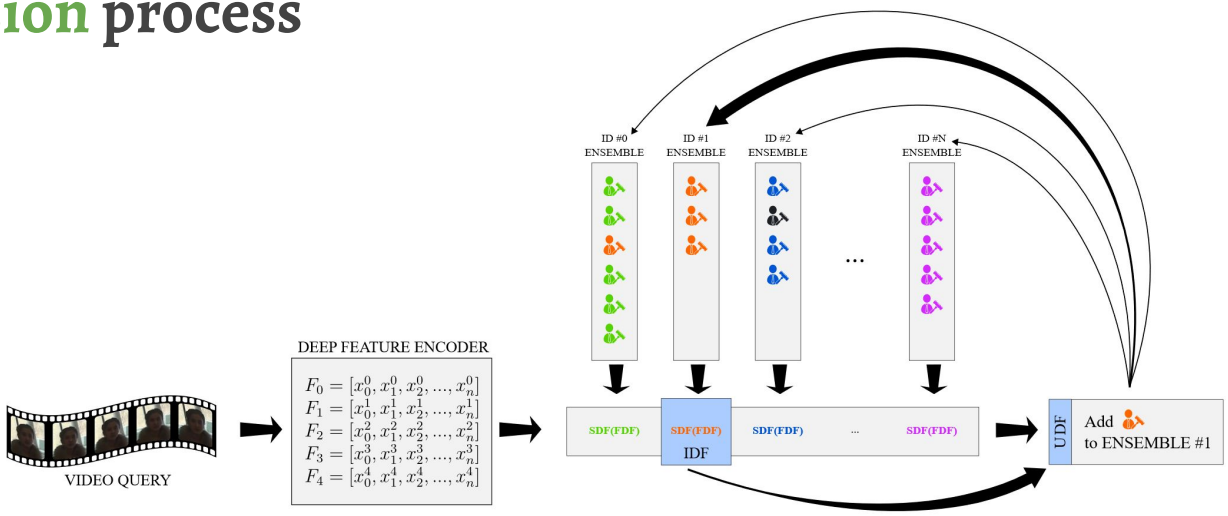
- Video-to-video face identification
- Adaptation to target domain data
- Adaptation over time
- Ability to use non-labelled data

Proposed system: *Dynamic Ensembles of SVM*

Basic Pillars

- Deep feature encoders as a basis
- One ensemble composed by SVM classifiers is assigned to each individual
- Ensembles are initialised with a few frames (5)
- Self-training approach: system's predictions are pseudo-labels
- Temporal coherence within each sequence
- Encouraging diversity when adding new classifiers in update

Decision process



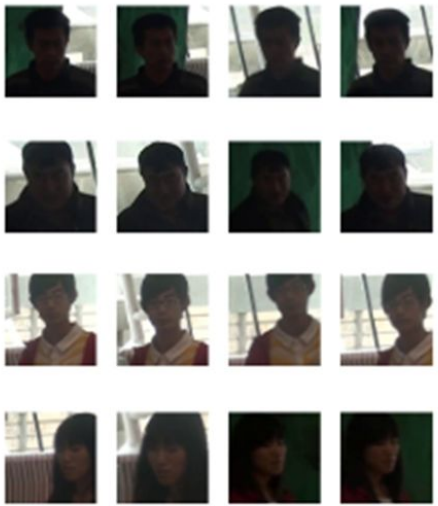
Decision rules

- Frame Decision Function (FDF) — *Assigns scores to frames (median)*
- Sequence Decision Function (SDF) — *Assigns scores to sequences (median)*
- Identification Decision Function (IDF) — *Assigns Identities to sequences (best)*
- Update Decision Function (UDF) — *How to update? (worst scoring frames)*

Experiments and Results

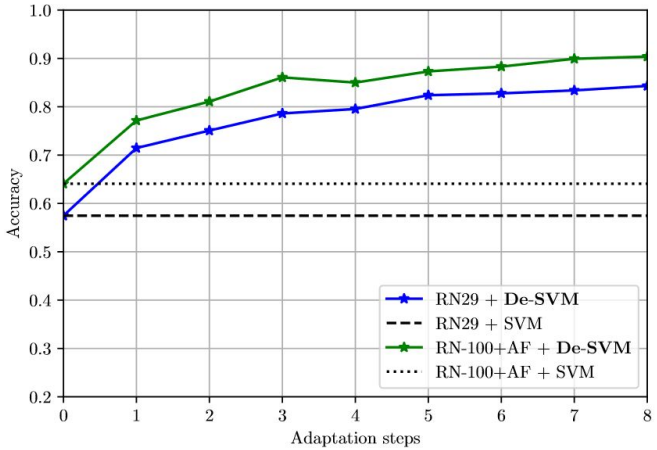
Experimental methodology

- COX Face database
  - 1000 individuals
  - 3 non-overlapping cameras
  - Each camera sequence is divided into 3 sub-sequence = 8 adaptation sub-sequences + 1 testing sub-sequence
- Deep Feature Encoders:
  - ResNet Dlib features (RN29)
  - ResNet50-ArcFace features (RN50-AF)
- Accuracy performance after each adaptation step
- Repeated for 100 and 700 identities, in closed-set scenario



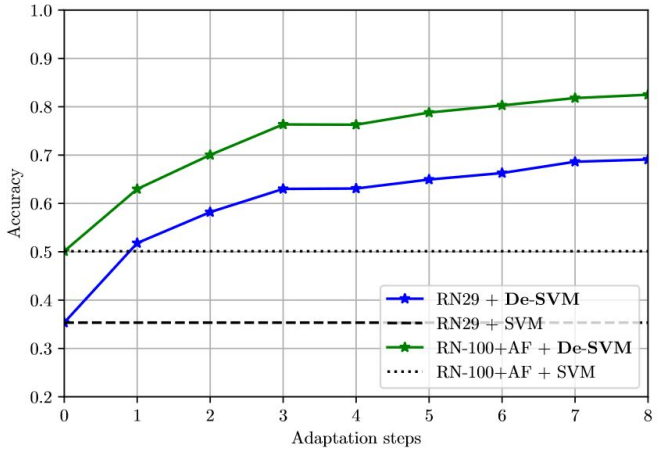
Results

100 identities



	Accuracy
RN29+SVM	57.5±4.1%
RN50-AF+SVM	64±11%
RN29+De-SVM (After Adapt.)	84.3±5.0%
RN50-AF+De-SVM (After Adapt.)	90.3±6.1%

700 identities



	Accuracy
RN29+SVM	35.34±0.60%
RN50-AF+SVM	50.11±0.99%
RN29+De-SVM (After Adapt.)	69.09±0.19%
RN50-AF+De-SVM (After Adapt.)	82.49±0.63%

Conclusions

- The type of target domain data is crucial for certain applications
- Self-training is useful to achieve unsupervised incremental learning
- De-SVM adaptation provides for up to 25% accuracy

Conclusions

- The type of target domain data is crucial for certain applications
- Self-training is useful to achieve unsupervised incremental learning
- De-SVM adaptation provides for up to 25% accuracy

Future work

- Mistaken updates corrections
- Extend to the more general open-set setting

An Adaptive Video-to-Video Face Identification System Based on Self-Training

Eric Lopez-Lopez

Thank you so much for your attention.

UNIVERSIDADE DA CORUÑA

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA