Time Analysis of Pulse-based Face Anti-Spoofing in Visible and NIR

Javier Hernandez-Ortega, Julian Fierrez, Aythami Morales, and Pedro Tome

Biometrics and Data Pattern Analytics – BiDA Lab
Universidad Autónoma de Madrid, Spain
https://atvs.ii.uam.es/atvs/

CVPR 2018 Workshop on Biometrics, Salt Lake City, Utah, US.

BiDA Lab
Biometrics and Data Pattern Analytics

Escuela Politécnica Superior

UAM UNIVERSIDAD AUTONOMA DE MADRID

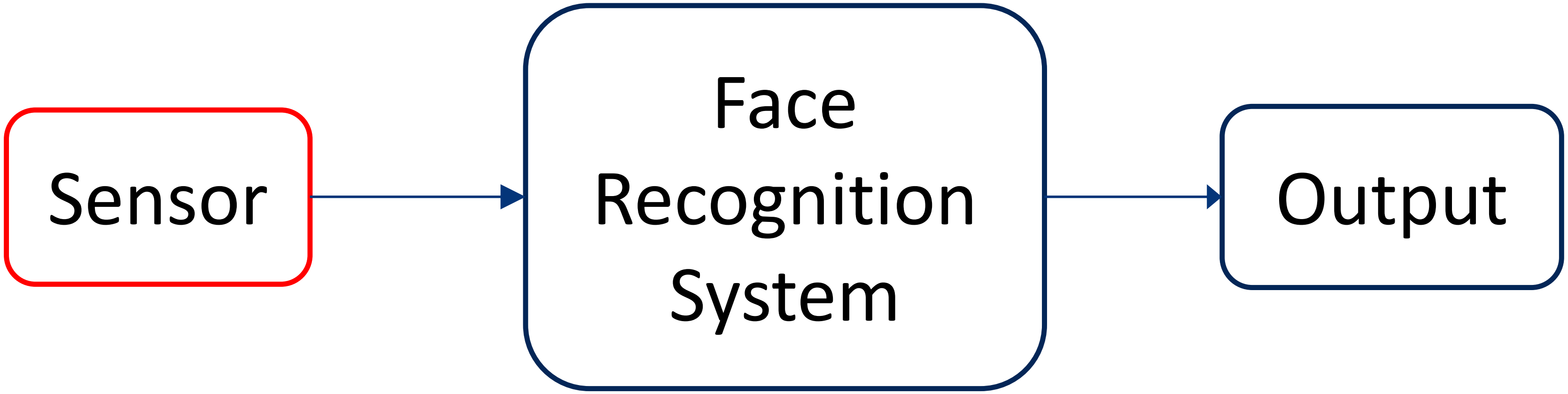
Scope

- Face Presentation Attack Detection.
- 3D Mask Attacks, photo-print attacks.
- Remote Photoplethysmography for detecting pulse.
- Performance in:
    - Short-time videos.
    - Variable scenarios over time.

Apple Face ID

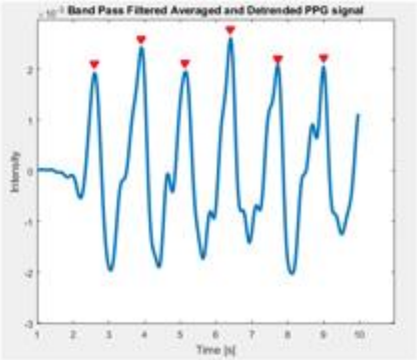
Presentation Attacks in Face Recognition

- Impersonate a genuine user.
- Artifacts.
- At sensor level.

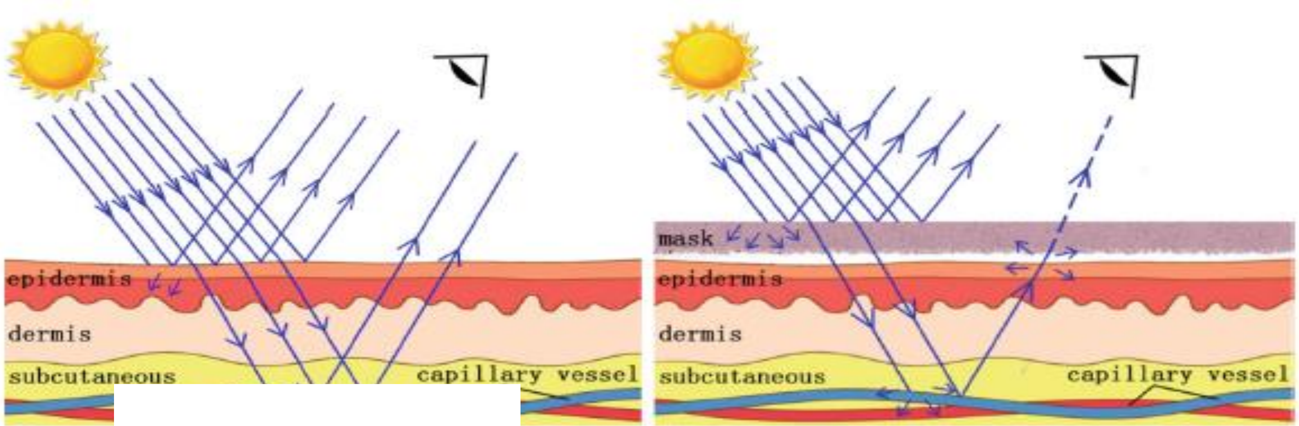


Remote Photoplethysmography (rPPG)

- Slight **changes in the skin color** at video recordings.



- 3D mask attack **highly different from a genuine pulse** [1].



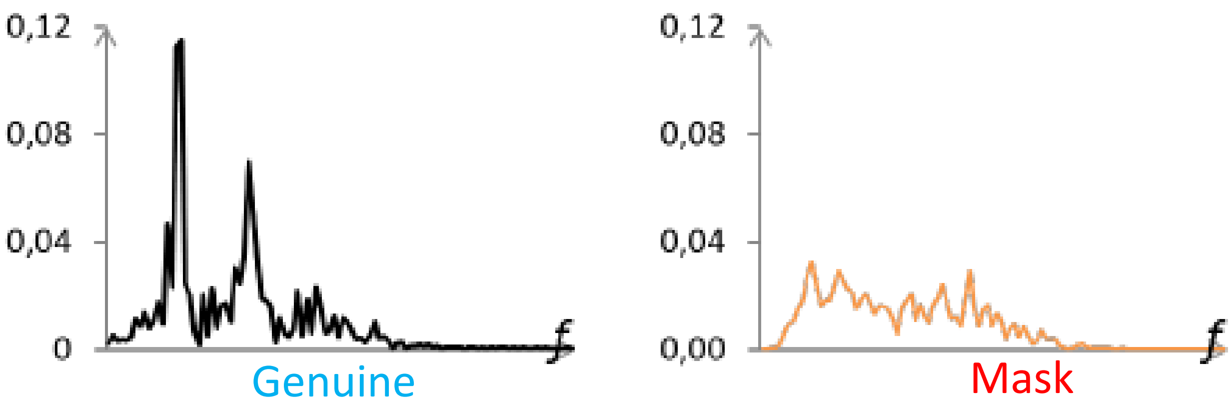
(b) Masked Face

[1] Liu et al.: 3D Mask face anti-spoofing with remote photoplethysmography. In European Conference on Computer Vision 2016 (pp. 85-100). Springer.

Reference Work

- **[2]: PAD with public dataset based on rPPG.**
  - Few public datasets.
  - Usually HR estimation, not PAD.
  - Relative short recordings.

- Exploiting **differences in HR spectrum**:



[2] Li et al.: Generalized face anti-spoofing by detecting pulse from face videos. In International Conference on Pattern Recognition (ICPR), 2016.

Databases

3DMAD: 3D Mask Attack Dataset [3]

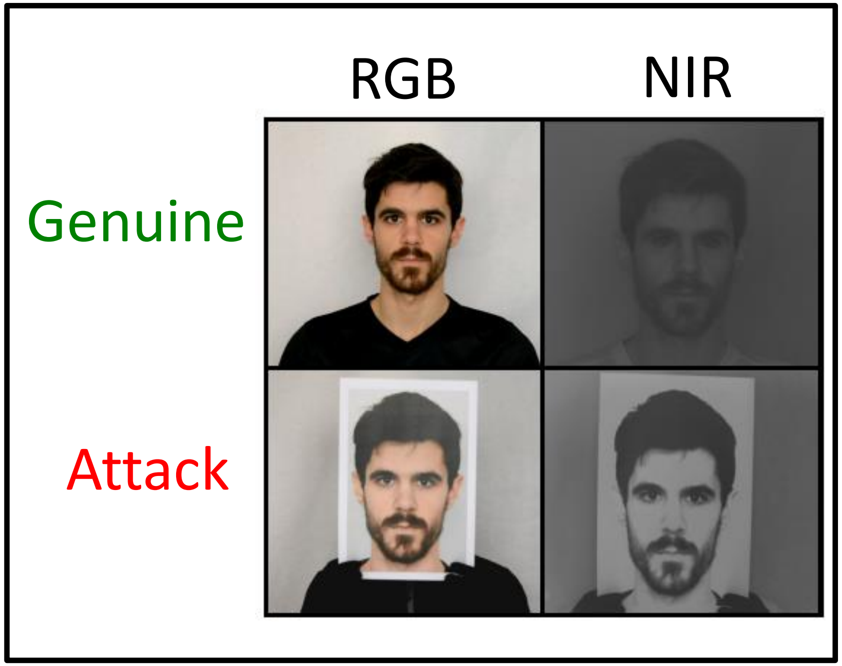
- **3D Hard resin masks.**
- **10 sec. RGB videos.**
- 17 users.
- 3 sessions/user:
  - 2 real access
  - 1 mask attack
- 5 videos/session.
- 640x480 pixels
- 30 fps.

Extracted from [9].

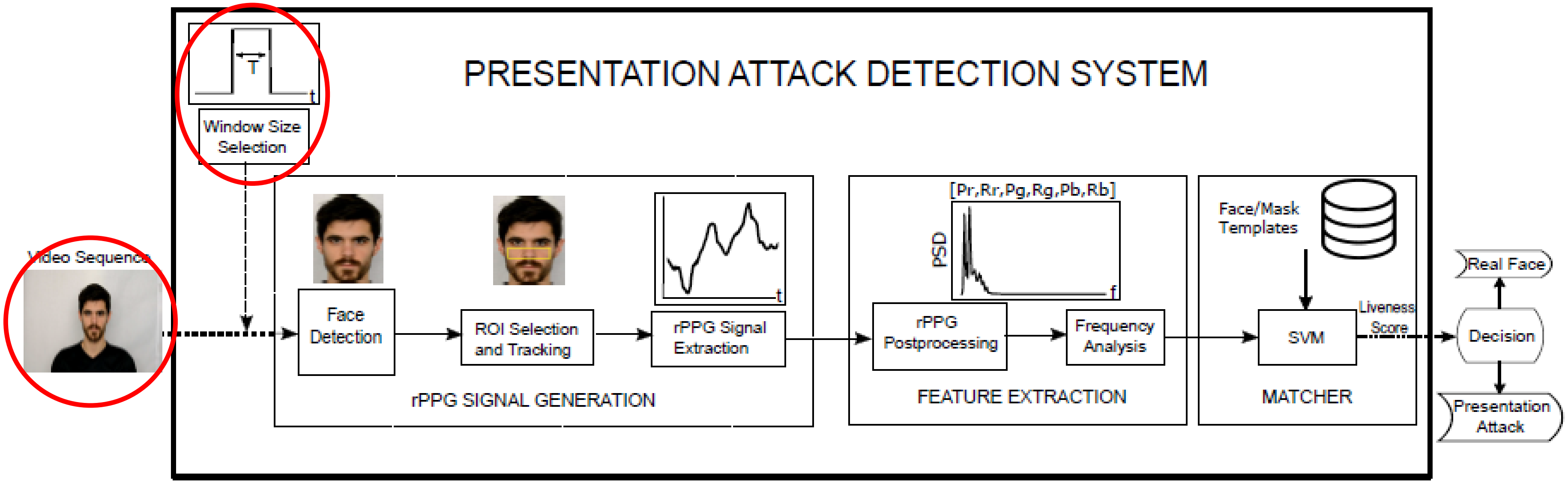
[3] Erdogmus and S. Marcel. Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect. In IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems, 2013.

BiDA HR: BiDA Heart-Rate database

- **Supplementary**.
- RGB videos:
  - 1920×1080 pixels. 25 fps
- NIR videos:
  - 1032×778 pixels. 30fps
- **60 seconds.**
- **Photo print attacks.**
- 10 users, 3 sessions/user:
  - 2 real access (rest & accelerated pulse)
  - 1 photo print attack.



Proposed system



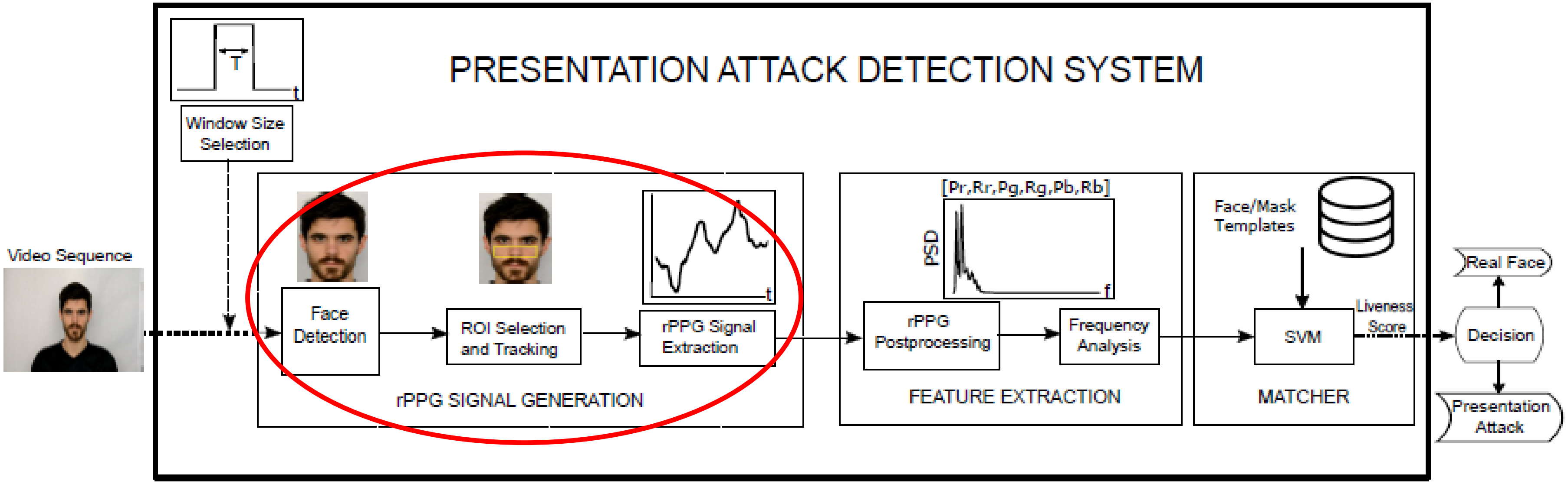
Input: facial video sequence.

Output: real face or presentation attack.

Preprocessing:

- Extracting short video sequencies
- Variable duration T
- Rectangular window.
- No overlapping.

Proposed system

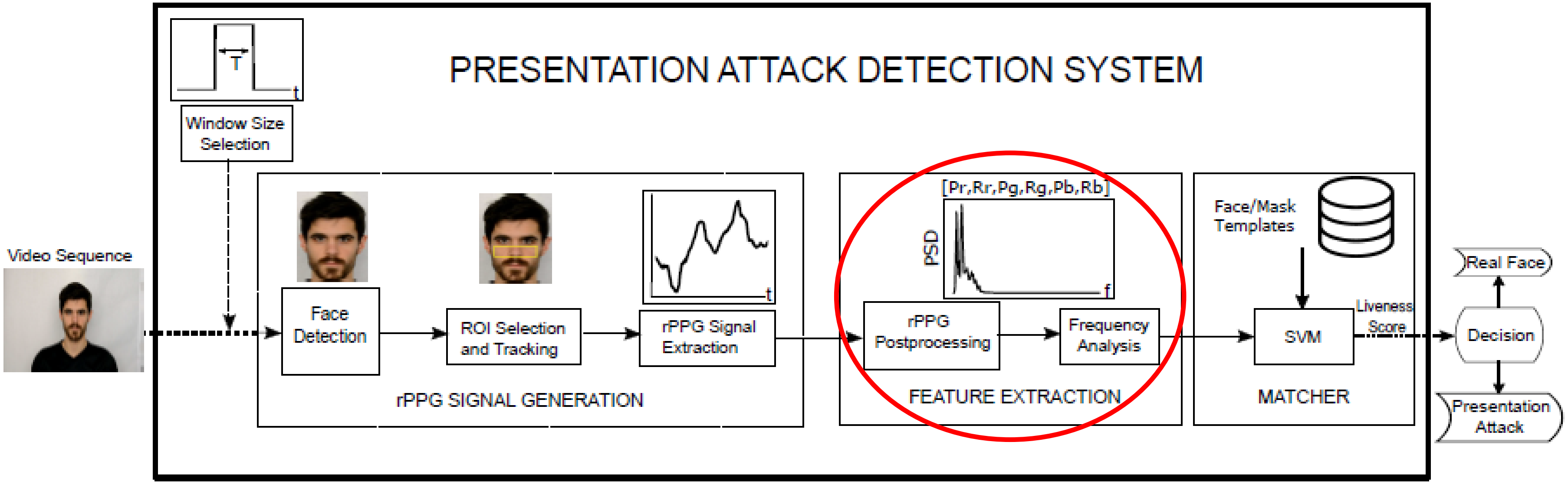


- **rPPG Signal Generation:**

  1) Face detection → Viola & Jones
  2) ROI selection. → Cheeks and Nose
  3) Raw rPPG signal extraction. → Avg. pixel values.

Quick → low-latency study.

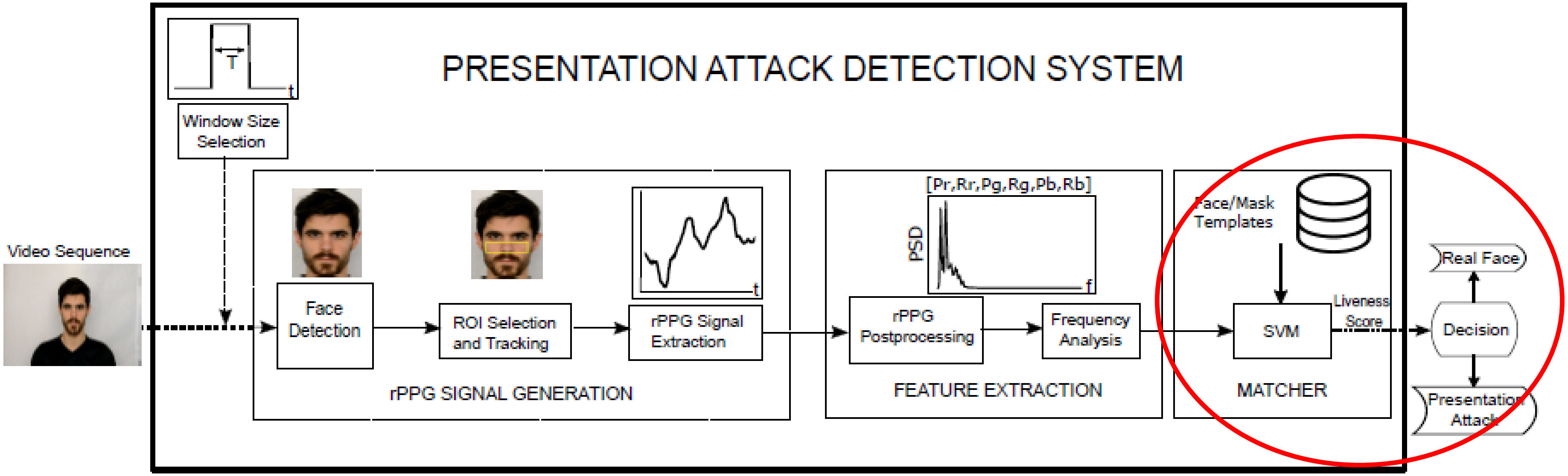
Proposed system



- **Feature Extraction:**

  1) rPPG postprocessing: **filtering.**
  2) Feature extraction from **power spectral density** (PSD) distribution:
     P: maximum power response.
     R: ratio of P to the total power in the 0.6 - 4 Hz frequency range.

Proposed system



- **Match Score Generation:**

  1) Classifier: **SVM** with Linear Kernel.

     1 score for each temporal window.

Experimental protocol

- Based on Li et al. [2]
- Feature vector: size 6 for RGB→[P_r,R_r,P_g,R_g,P_b,R_b]
- *T*: from 1 to **10 seconds**
- SVM: linear kernels, cost parameter C = 1000.
- **Leave-One-Out Cross-Validation protocol**.
  - 1 user for testing → EER.
  - Remaining users for training the classifier.
  - Repeat for all users → **Results are averaged**

- BiDA HR: NIR videos feature vector size 2 (only one channel).
  - *T*: from 1 to **60 seconds**

Results

- Averaged EER [%] on 3DMAD and BiDA HR **RGB:**

T [s]	1	2	3	5	7	10
3DMAD	42.8	45.0	37.8	33.1	**25**	22.1
BiDA HR	46.9	45.7	46.5	42.1	34.1	40.1

Better results with 3DMAD:

- Frame rate more relevant than resolution.

- Space is averaged.

Lower EER with more data:

-For small T values random behaviour.

Results

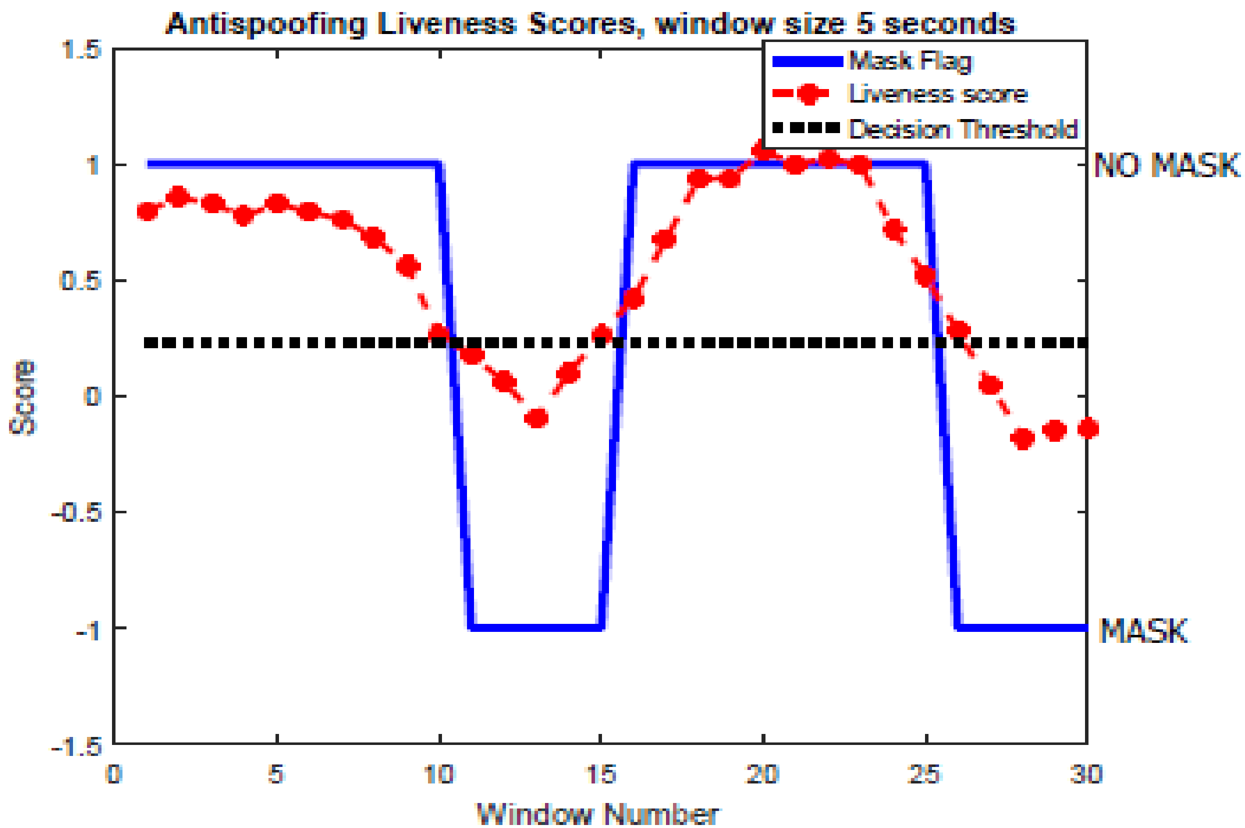
- Averaged EER [%] on BiDA HR database:

T [s]	1	2	5	10	20	30	40	50	60
RGB	46.9	45.7	42.1	40.1	**40.0**	40.0	36.6	30.0	25.0
NIR	42.4	41.7	38.4	30.9	30.0	16.6	5.0	0.0	0.0

- Longer video sequences (up to 60 secs.)
- Much better results with NIR:
  - Higher frame rate, similar resolution.
  - Better hardware quality → Less noise added.
  - Robust to external illumination.

Results

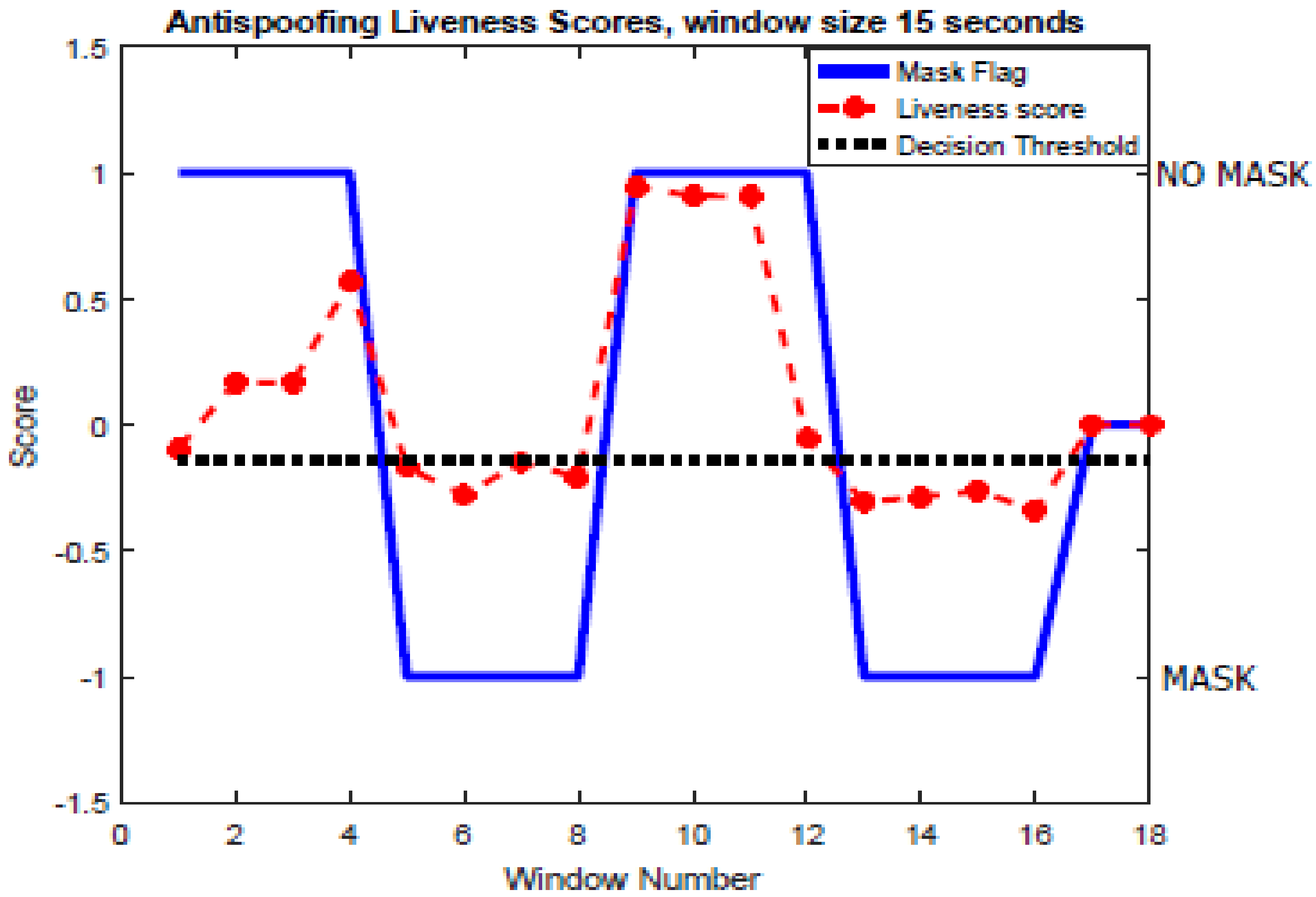
- Score evolution over time (1):



(a) Scores from 3DMAD RGB videos.

Results

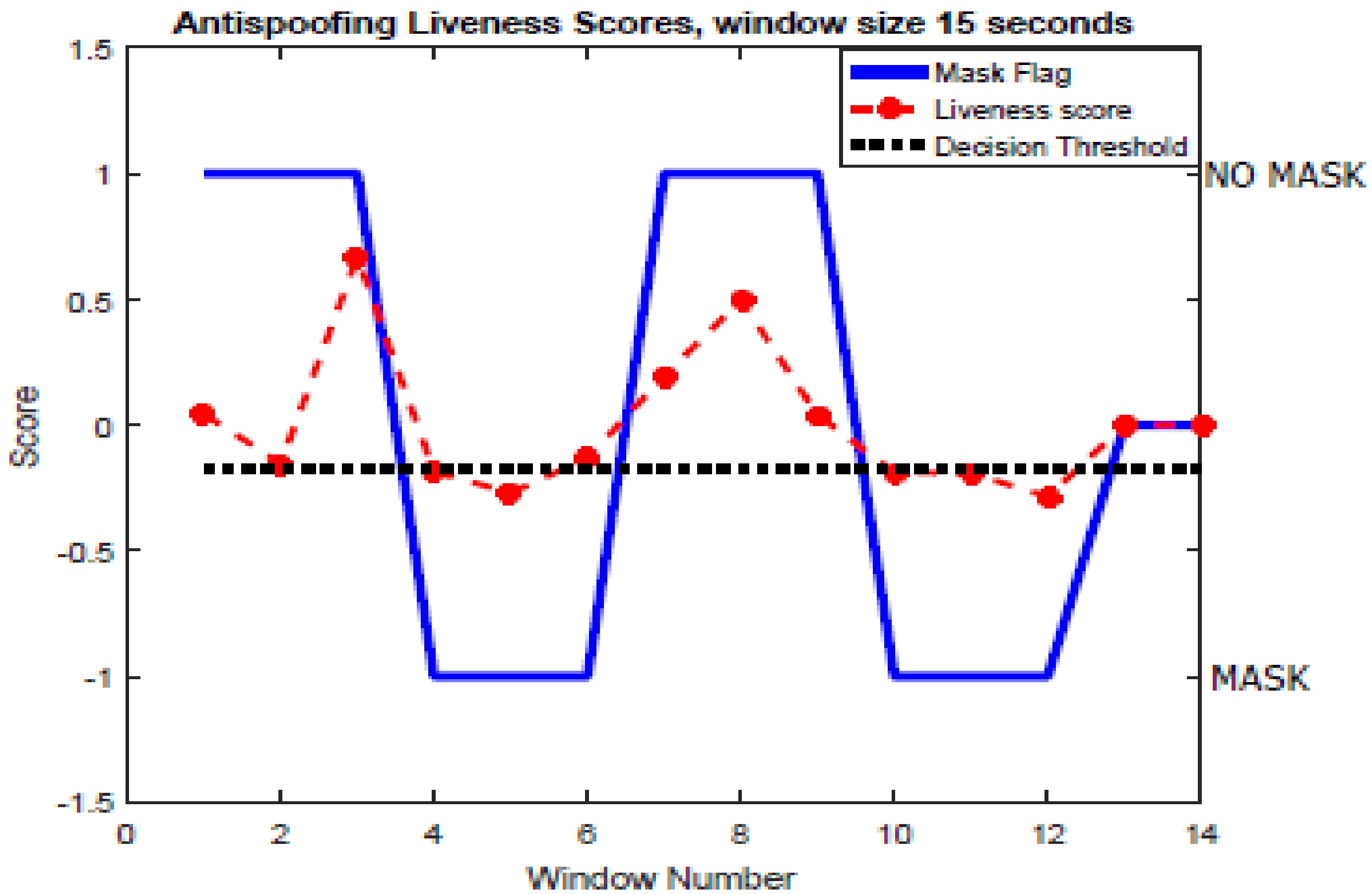
- Score evolution over time (2):



(b) Scores from HR RGB videos.

Results

- Score evolution over time (3):



(c) Scores from HR NIR videos.

Conclusions

- **Short-time PAD:** reacts to quick temporal changes in the attacking scenario.

- **Minimum length** for robust individual PAD score.

- **Frame rate** more relevant than resolution (with limits).

- **Better with NIR.**
  - More robust to external variations.

Future Work

1) **Improving the baseline** system for short videos.
   - More robust individual scores.

2) **Temporal integration** of scores for continuous PAD.

3) **Study impact** of spatial and temporal resolution.

4) **Larger database**:
   - # users, diff. artifacts and sensors.
   - Challenging scenarios: **mobile.**