

To Frontalize or Not To Frontalize: Do We Really Need Elaborate Pre-processing To Improve Face Recognition?

S. Banerjee¹ and J. Brogan¹, J. Krizaj², A. Bharati¹, B. RichardWebster¹, V. Struc², P. Flynn¹, and W. Scheirer¹

¹University of Notre Dame, USA, ²University of Ljubljana, Slovenia

(Code: https://github.com/joelb92/ND_Frontalization_Project/Release)



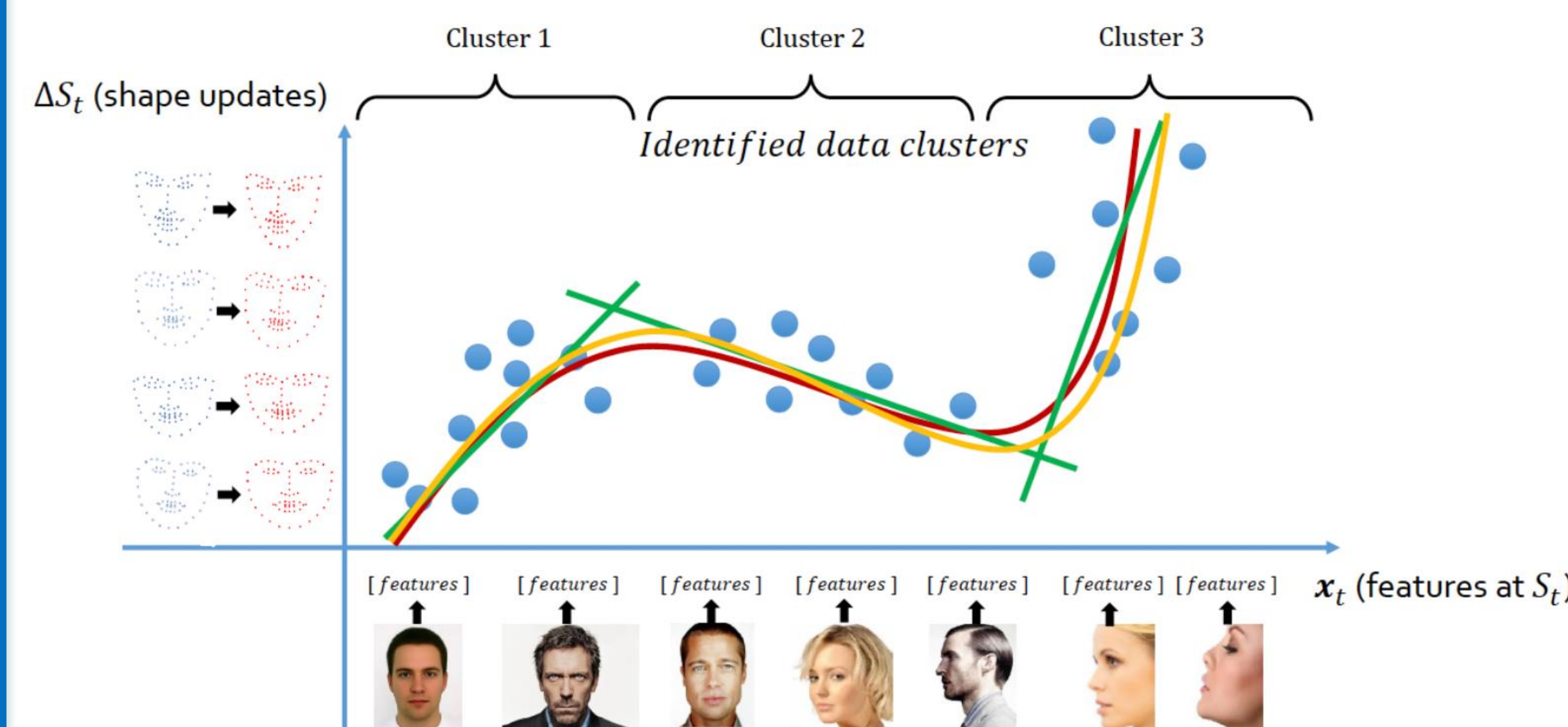
A. Motivation

- What should be the orientation of face images meant for CNN training?
- Should they be frontalized or is 2D alignment sufficient?
- What mode of frontalization is ideal?
- Different methods yield different results, which one is optimal?



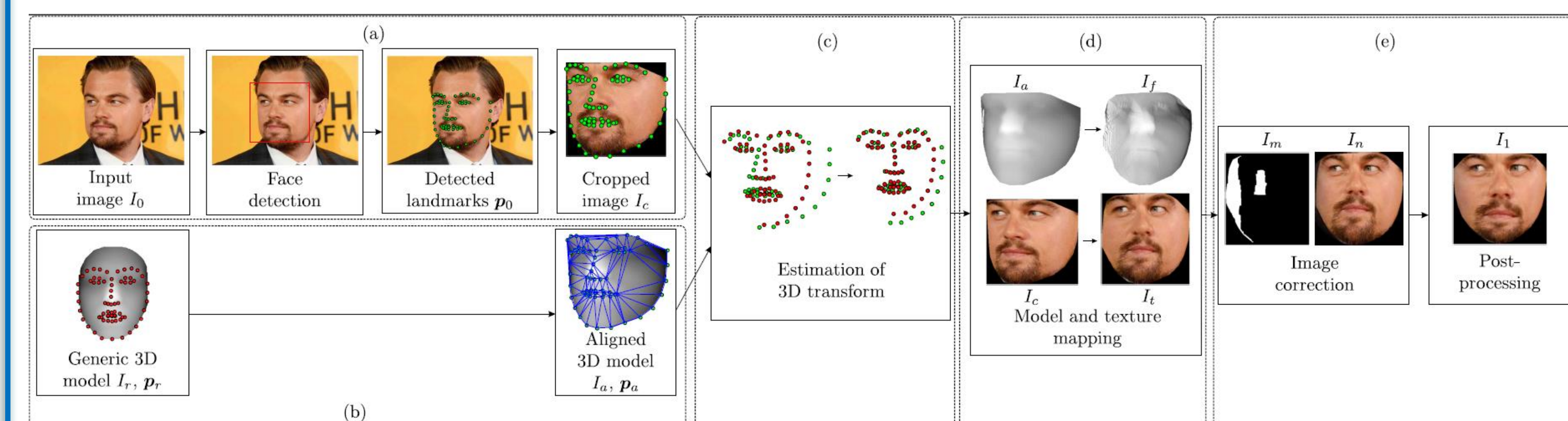
B. Landmarking (CMR)

- Start with initial face shape (landmark points).
- Using a mixture of regressors, keep updating landmarks till convergence.
- CMR is fast due to simultaneous update and accurate due to regressor set.

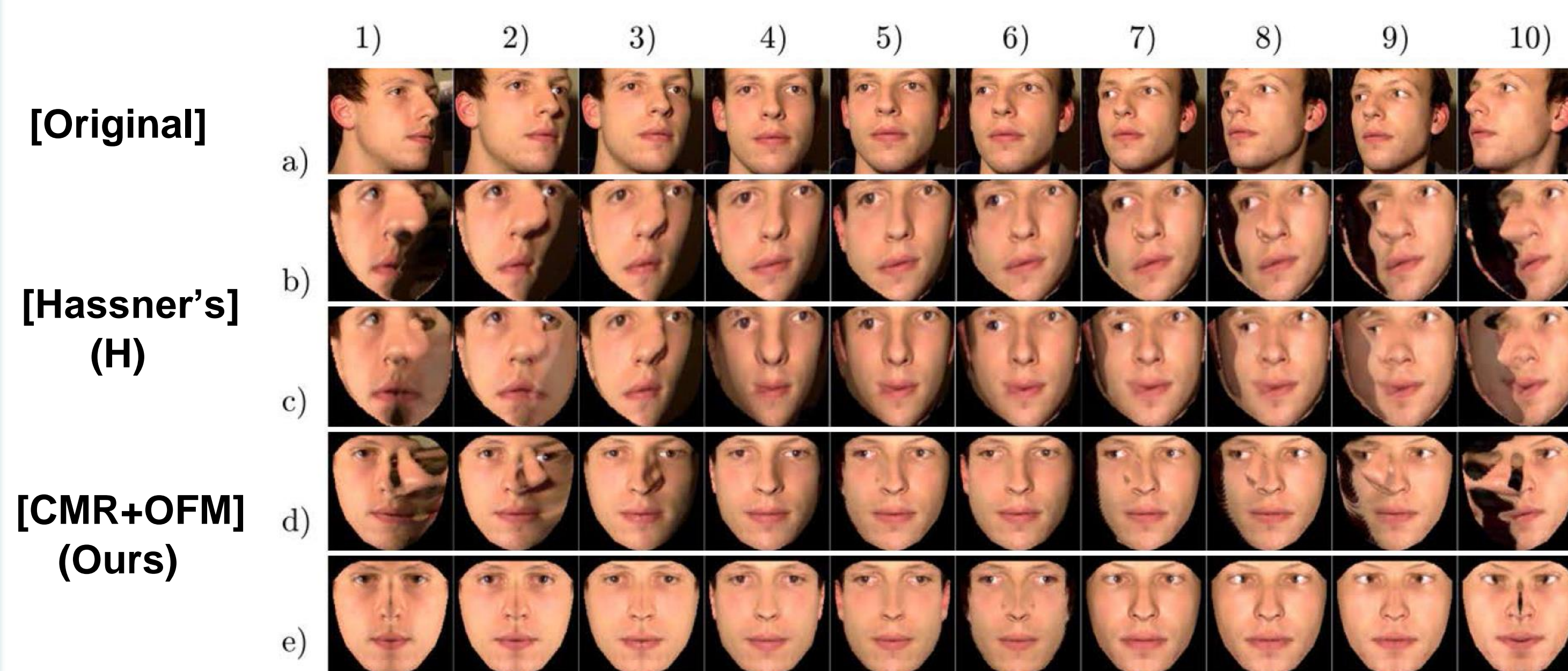


C. Frontalization (OFM)

- Detect 68 landmark points using CMR by regressing initial landmarks.
- Align generic 3D model to input face by piecewise warping.
- Find matrix T to transform aligned 3D model to generic 3D model.
- Using T, map aligned 3D model to generic 3D model and texture of input face using 2D to 3D correspondences.
- Distorted pixels replaced from opposite side of the face.



D. Qualitative Results



E. Data Yield

Pre-processing method	CMR & H	KS & H	ZR & H	CMR & OFM	KS & OFM	ZR & OFM	2D alignment (not frontalized)
CASIA images (yield)	252,294 (83.13%)	255,571 (84.22%)	261,951 (86.31%)	252,222 (83.11%)	266,269 (87.74%)	254,381 (83.82%)	268,455 (88.45%)
PaSC videos (yield)	2,691 (96.03%)	2,510 (89.57%)	2,497 (89.11%)	2,604 (92.93%)	2,476 (88.36%)	2,508 (89.51%)	2,726 (97.28%)

- Considerable data loss when any frontalization is used, compared to a simple 2D alignment.

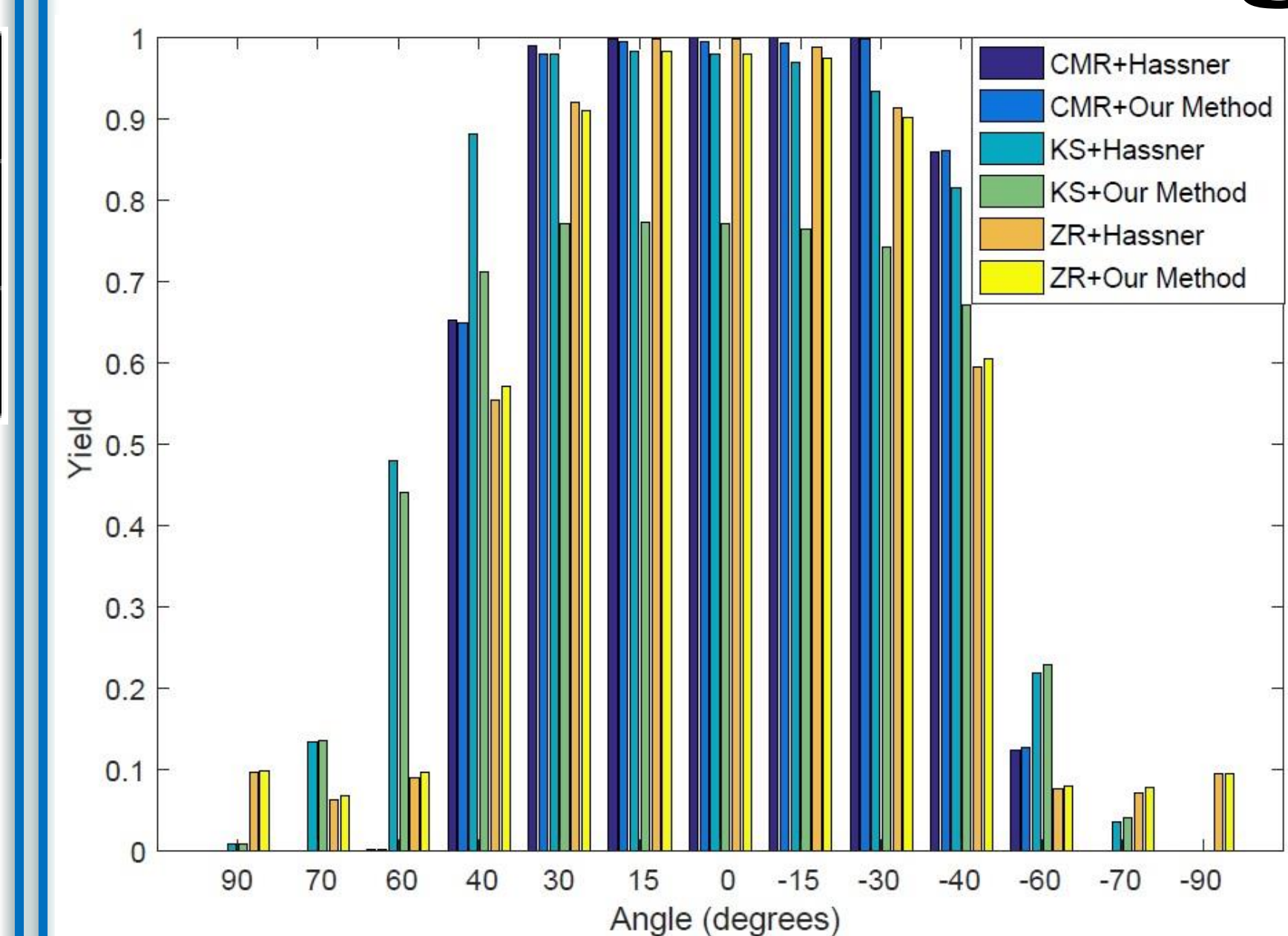
Network model – fine-tune VGG-Face (Parkhi, BMVC 2015).

Training Data – CASIA-WebFace subset (303K images); Testing Data - PaSC.

H – Hassner, CVPR 2015.

KS – Kazemi and Sullivan, CVPR 2014; ZR – Zhu and Ramanan, CVPR 2014.

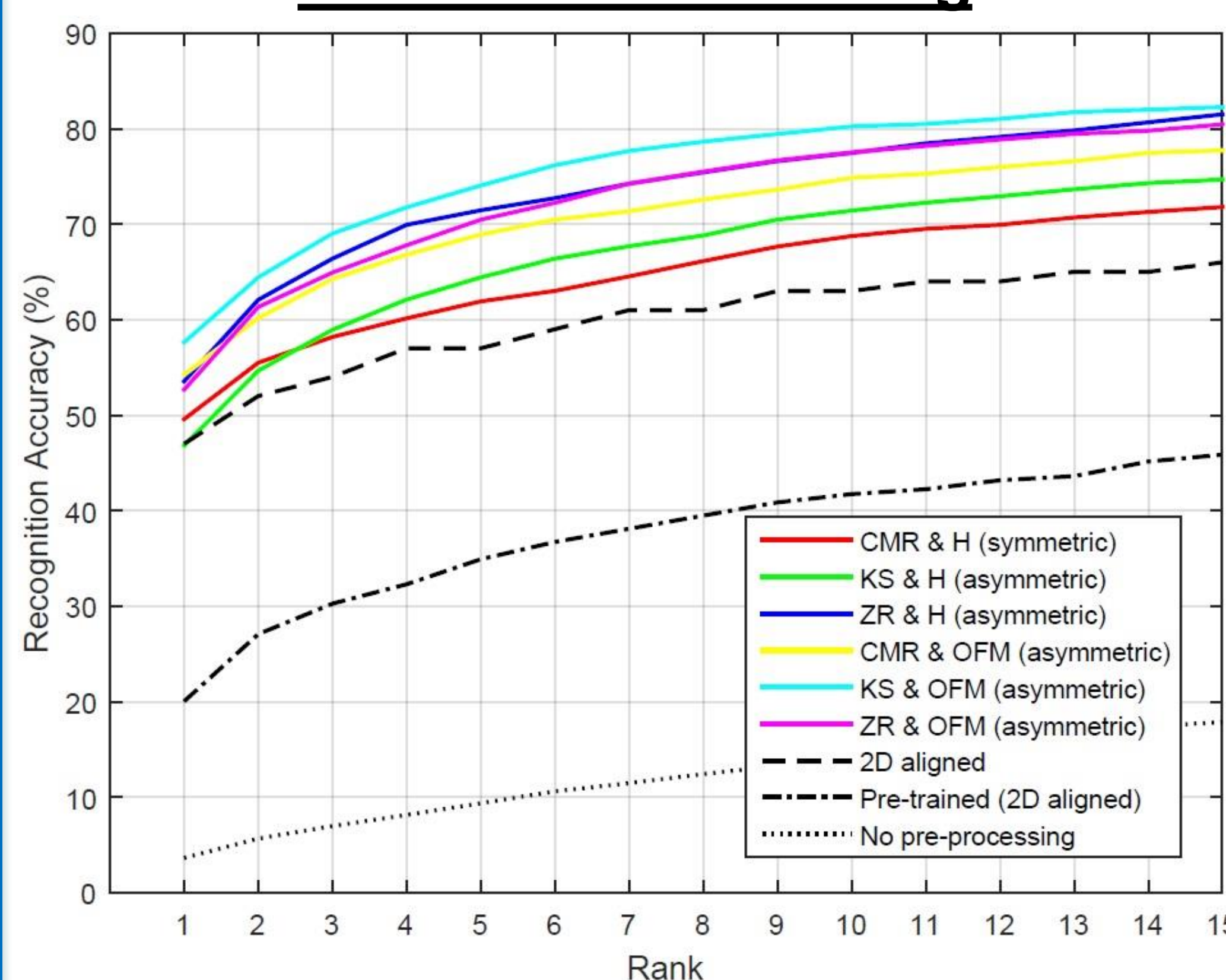
F. Pose Range



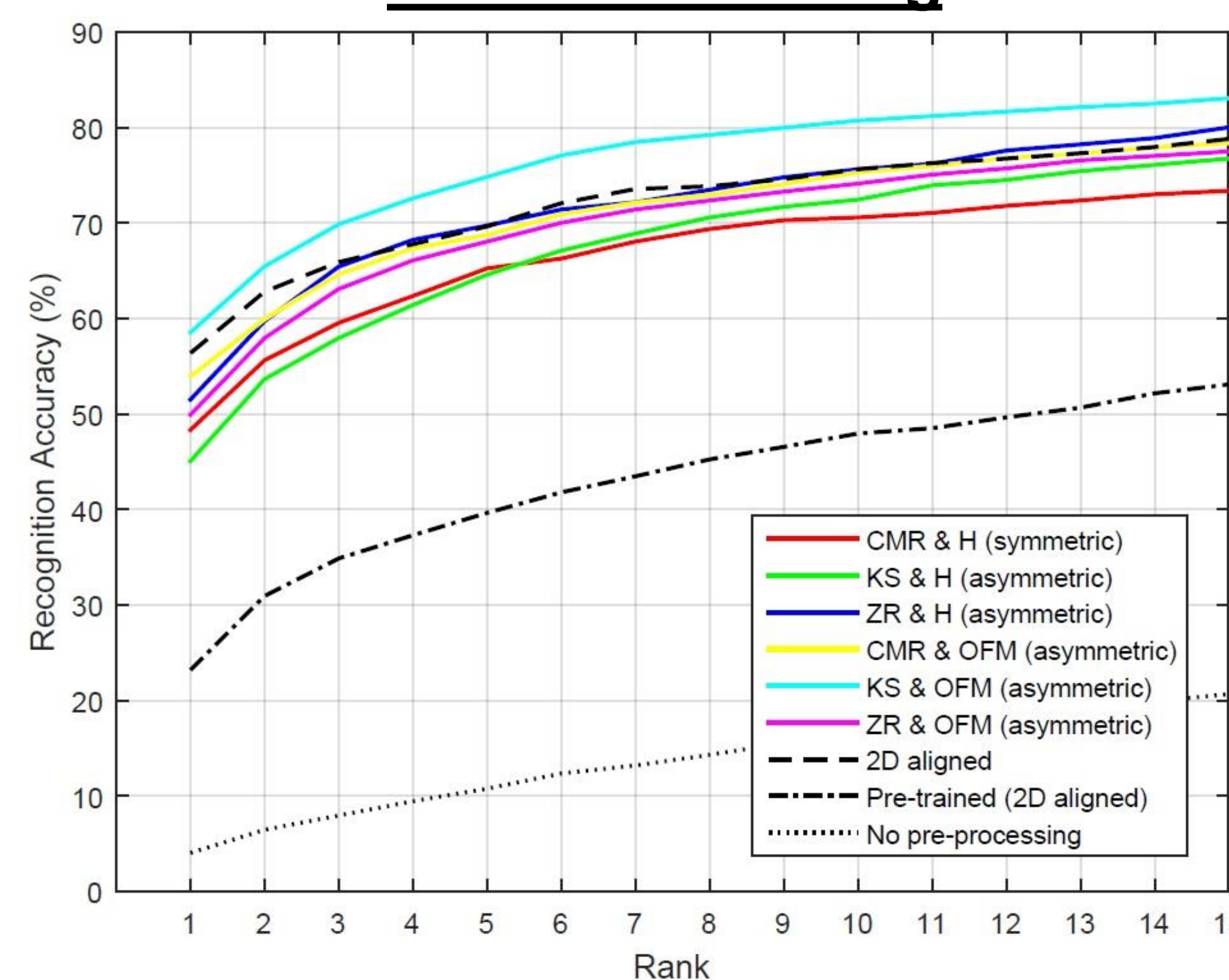
- CMR performs best in 0-40 yaw range.
- OFM and H has similar yield in 0-40 range.
- KS outperforms other landmarks when yaw > 40.

G. Quantitative Results

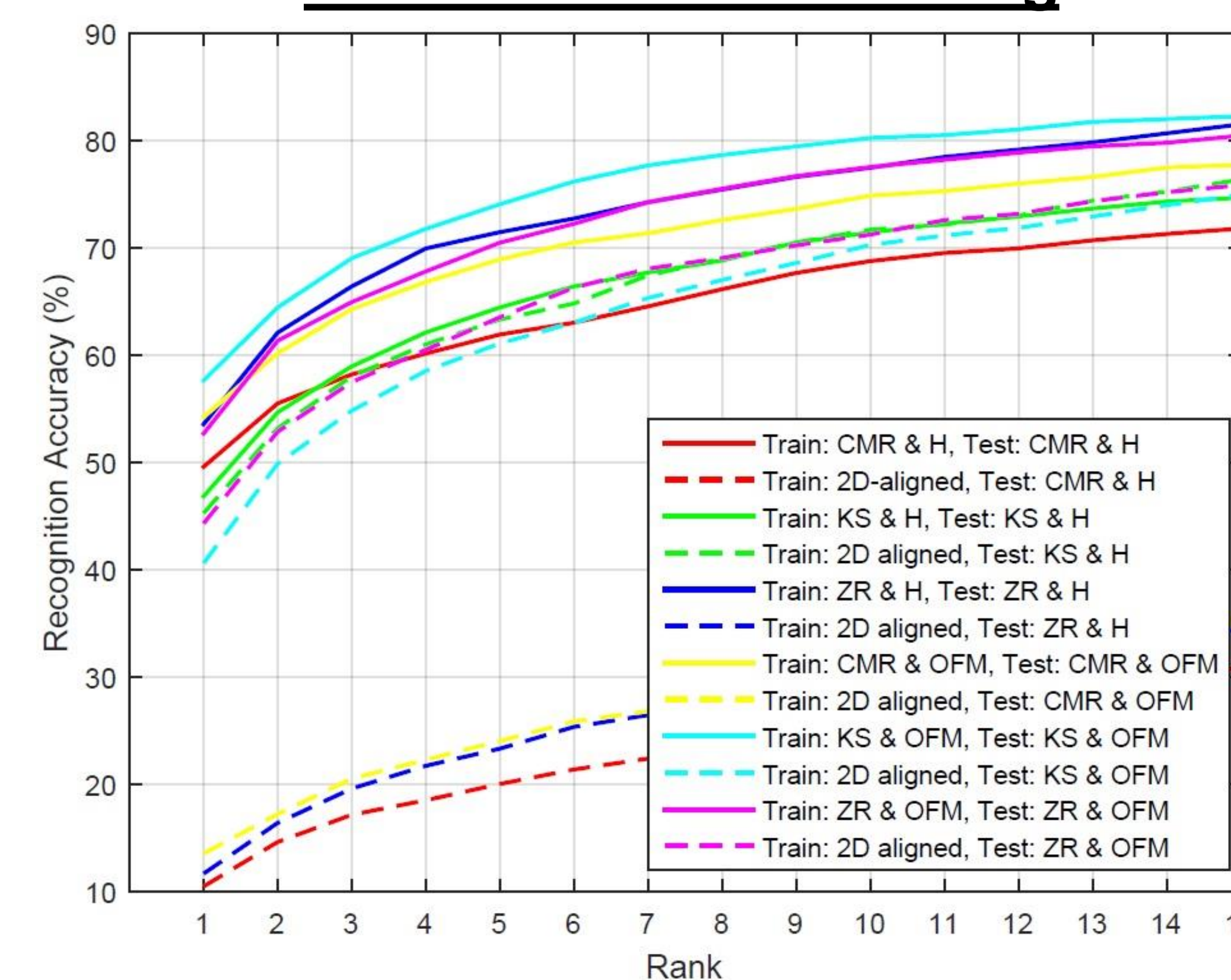
I. Non-uniform Testing



II. Uniform Testing



III. Cross-domain Testing



H. Takeaways

- Our proposed frontalization method (OFM), which dynamically adapts local areas of the 3D reference model to the given input face, provides better performance improvements than Hassner's (H) for PaSC video recognition.
- Due to the significant loss of data (specifically for extreme pose or occlusion) face image frontalization introduces, its performance benefit must be weighed against a computationally cheaper 2D alignment.
- Both the training and testing data must be pre-processed under consistent methods to realize any performance benefit out of frontalization.
- Asymmetric frontalization provides slightly superior performance for face recognition.
- Training with millions of face images makes a network more robust (pre-trained VGG-Face ~ 25-30% accuracy).
- The network trained with non-preprocessed images always performed poorly (~5%). A simple 2D alignment helps!