Not Afraid of the Dark

NIR-VIS Face Recognition via Cross-spectral Hallucination and Low-rank Embedding

José Lezama[2*], Qiang Qiu[1*], Guillermo Sapiro[1]

[1]Duke University
[2]Universidad de la República, Uruguay

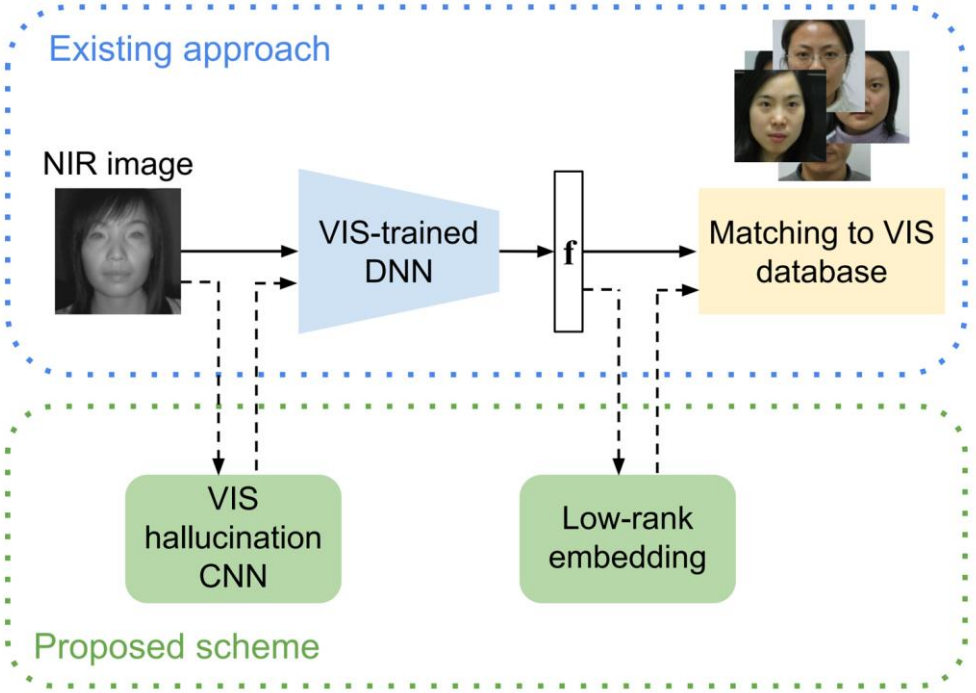
* Denotes equal contribution

Challenge

Apply deep learning for NIR-VIS face recognition:

- Limited near-infrared (NIR) training faces.
- Face galleries to be matched are mostly visible light (VIS).

Proposed Approach



No re-training of VIS-DNNs!

NIR-VIS Hallucination

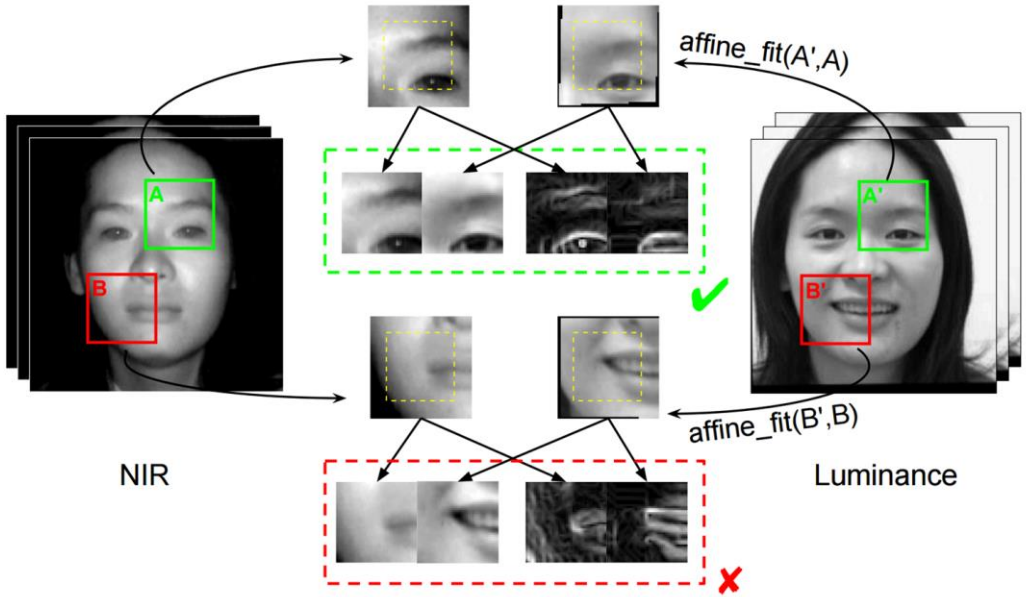
- Use a patch-based CNN to learn the mapping from NIR-VIS
- *Challenge:* No database of **aligned** NIR-VIS pairs
- *Approach:* Construct a database by mining aligned pairs from *CASIA NIR-VIS 2.0 Face Dataset*

Patch Mining

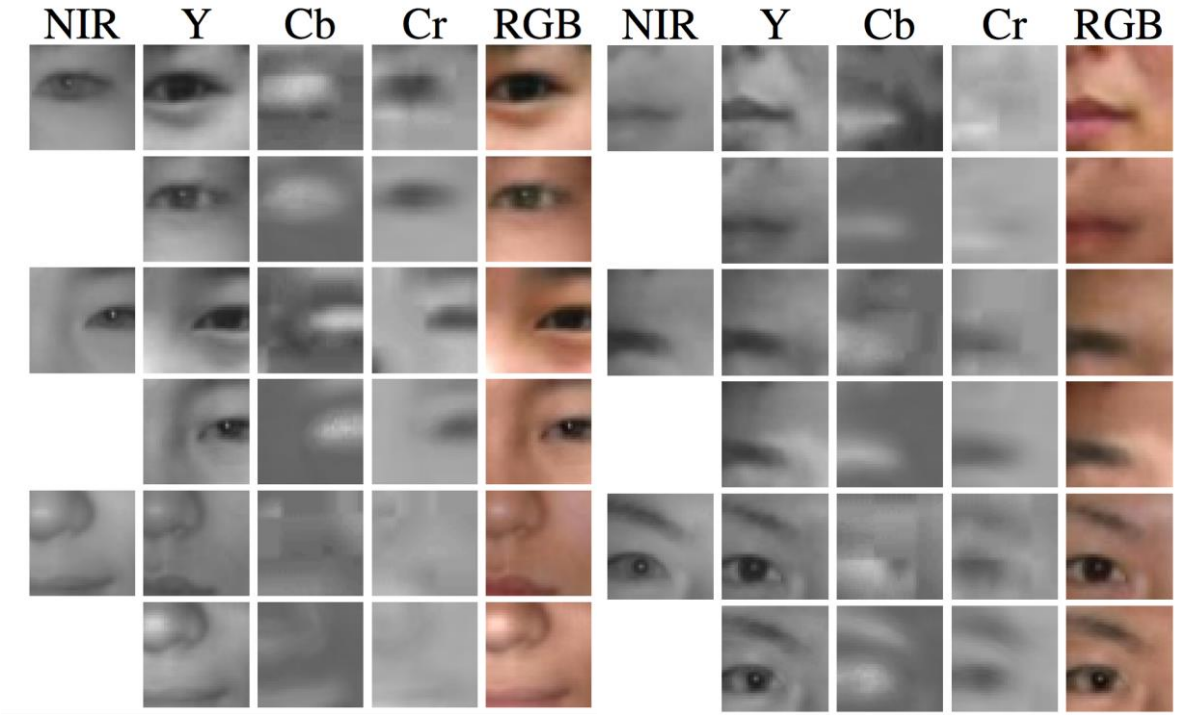
Over pairs of NIR/Luminance images, extract pairs of patches that are highly correlated after an affine transforms.

We obtain a dataset of ~1 Million **aligned** NIR-VIS patches of 40x40 pixels

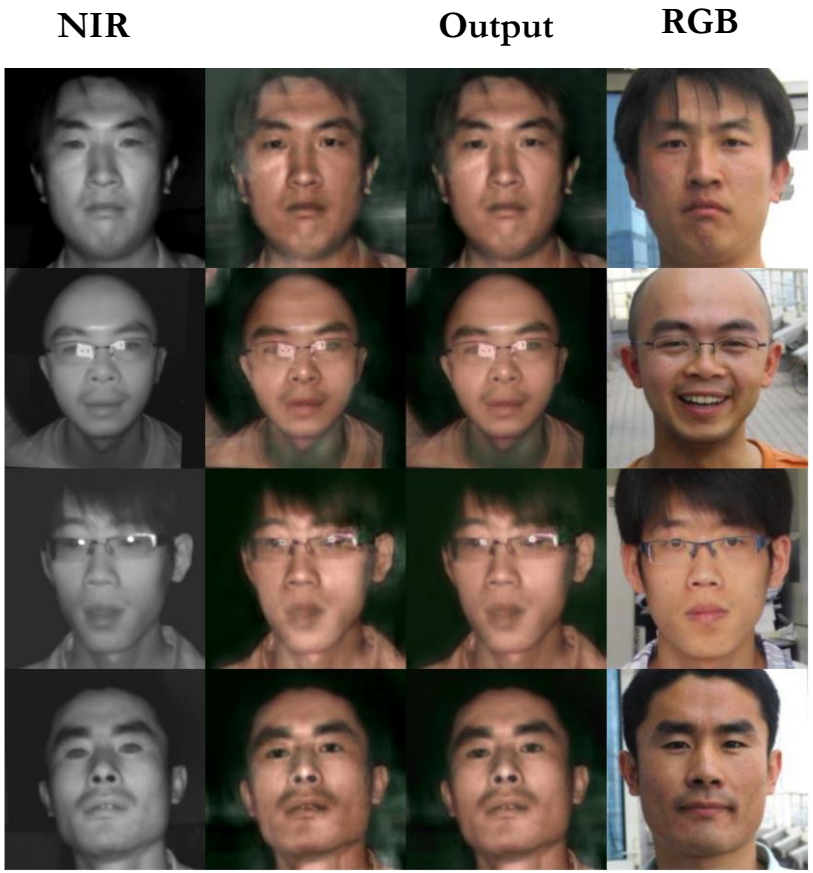
Using this dataset we train a NIR-VIS hallucination CNN



NIR-VIS Hallucination Results - Patches



NIR-VIS Hallucination Results - Images



Cross-modal Low-rank Embedding

Apply low-rank embedding to the output features of the VIS-DNN:

Group together features of the same subjects across modalities

Separate features of different subjects across modalities

$$\sum_{c=1}^{C} ||\mathbf{TY}_c||_* - ||\mathbf{TY}||_*$$

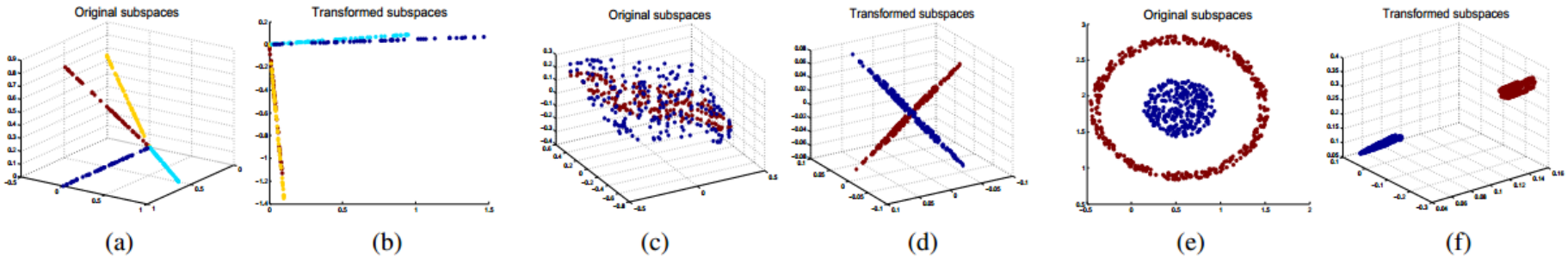
Qiang Qiu, Guillermo Sapiro, "Learning Transformations for Clustering and Classification", Journal of Machine Learning Research (JMLR), 16(Feb):187−225, 2015.

Cross-modal Low-rank Embedding

Apply low-rank embedding to the output features of the VIS-DNN:

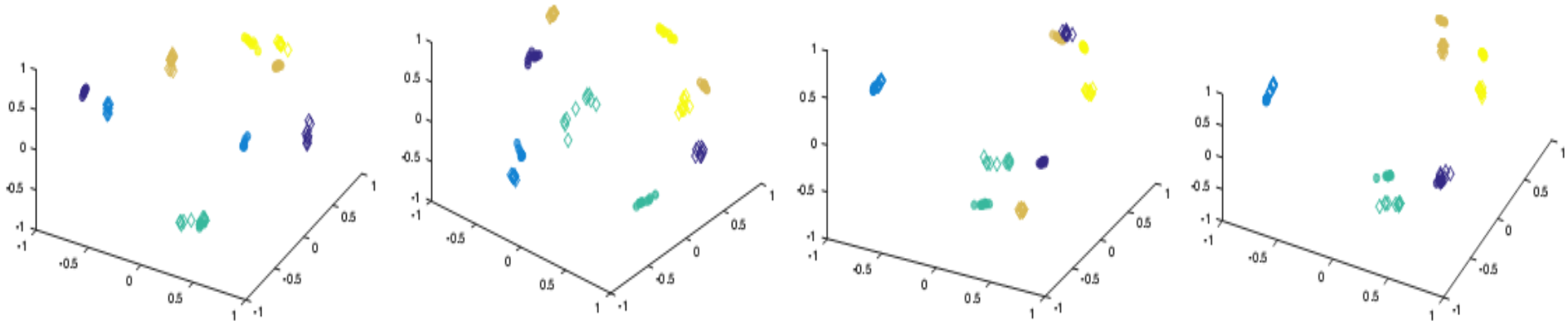
Group together features of the same subjects across modalities

Separate features of different subjects across modalities



Qiang Qiu, Guillermo Sapiro, "Learning Transformations for Clustering and Classification", Journal of Machine Learning Research (JMLR), 16(Feb):187−225, 2015.

Cross-modal Low-rank Embedding



(a) No embedding (b) Pairwise embedding (c) Triplet embedding (d) Low-rank embedding

5 subjects. Filled circle: VIS samples. Empty diamond: NIR samples.

Results

	Accuracy (%)
Jin et al. [15]	75.70 ± 2.50
Juefei-Xu et al. [16]	78.46 ± 1.67
Lu et al. [25]	81.80 ± 2.30
Saxena et al. [30]	85.90 ± 0.90
Yi et al. [37]	86.16 ± 0.98
Liu et al. [24]	**95.74 ± 0.52**
VGG-S	57.53 ± 2.31
VGG-face	66.97 ± 1.62
COTS	79.29 ± 1.54
VGG-S + Triplet	67.13 ± 3.01
VGG-face + Triplet	75.96 ± 2.90
COTS + Triplet	84.91 ± 3.32
VGG-S + Low-rank	82.07 ± 1.27
VGG-face + Low-rank	80.69 ± 1.02
COTS + Low-rank	*89.59 ± 0.89*

Results on CASIA NIR-VIS 2.0 dataset (not cross-modal hallucination). [24] performs DNN fine-tuning.

	Accuracy (%)
VGG-S	75.04
VGG-S + Hallucination	80.65
VGG-S + Low-rank	89.88
VGG-S + Hallucination + Low-rank	95.72
VGG-face	72.54
VGG-face + Hallucination	83.10
VGG-face + Low-rank	82.26
VGG-face + Hallucination + Low-rank	91.01
COTS	83.84
COTS + Hallucination	93.02
COTS + Low-rank	91.83
COTS + Hallucination + Low-rank	**96.41**

Results on CASIA NIR-VIS 2.0
NIR-VIS hallucination and low-rank embedding separately improve recognition of a black-box DNN, and more so when used in combination

Conclusions

Adapted pre-trained state-of-the-art VIS DNN to generate discriminative features for both VIS and NIR faces, *without retraining the DNN*

Cross-spectral hallucination performs a conversion of the NIR image into the VIS spectrum

Low-rank embedding restores low-rank structure for cross-spectral features

Significant improvement in cross-spectral face recognition with the proposed approach

New direction in the intersection of transfer learning and joint embedding