Paper link:

Discriminative Invariant Kernel Features:

A Bells-and-Whistles-Free Approach to Unsupervised Face Recognition and Pose Estimation

Dipan K. Pal, Felix Juefei Xu, and Marios Savvides

dipanp@cmu.edu, felixu@cmu.edu, msavvid@ri.cmu.edu

IEEE 2016 Conference on Computer Vision and Pattern Recognition

CVPR 2016

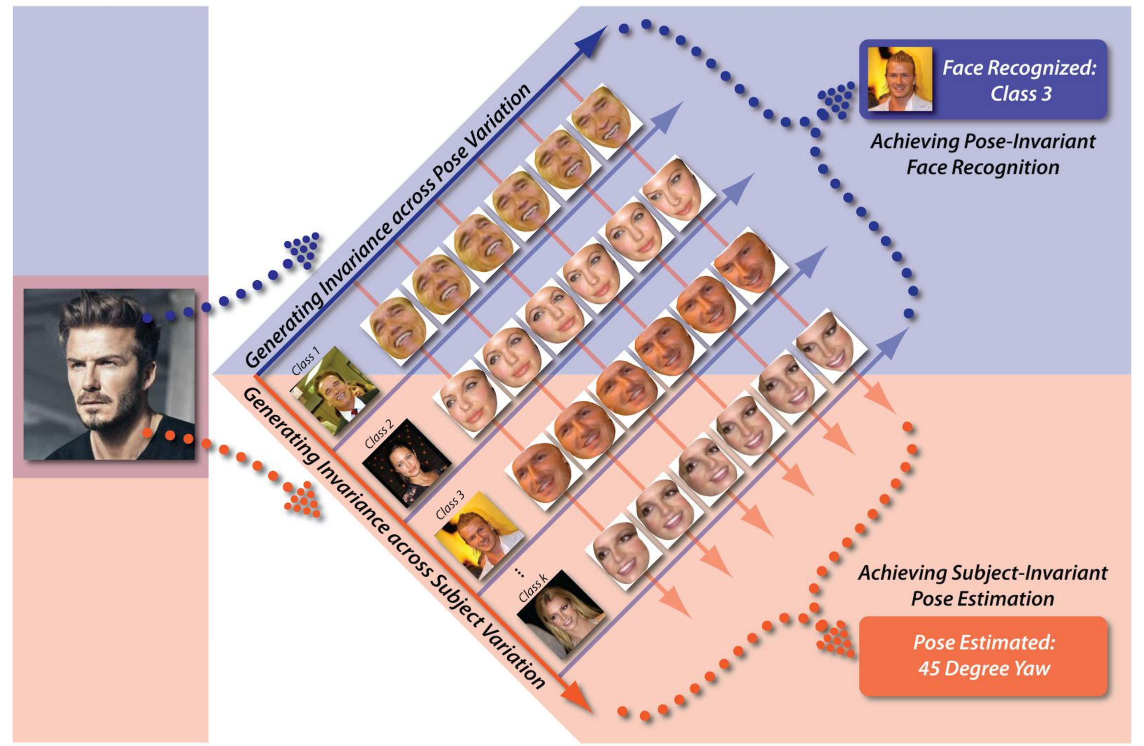
The Problem

Two complementary tasks: To perform two complementary tasks simultaneously using a single unsupervised feature extractor.

Who is this subject?

What is the subject's pose?

Landmark-free: The paper focuses on dense landmark-free (only two eye center locations) face recognition and pose estimation.
Also extends to a completely landmark-free approach which is also alignment free.



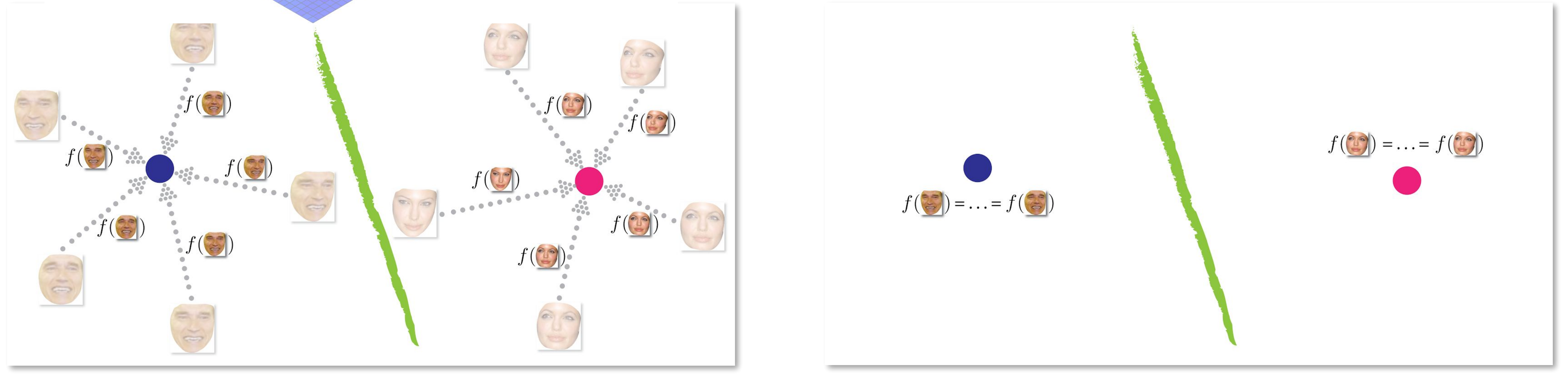
Face recognition sub-feature

Pose estimation sub-feature

The Approach

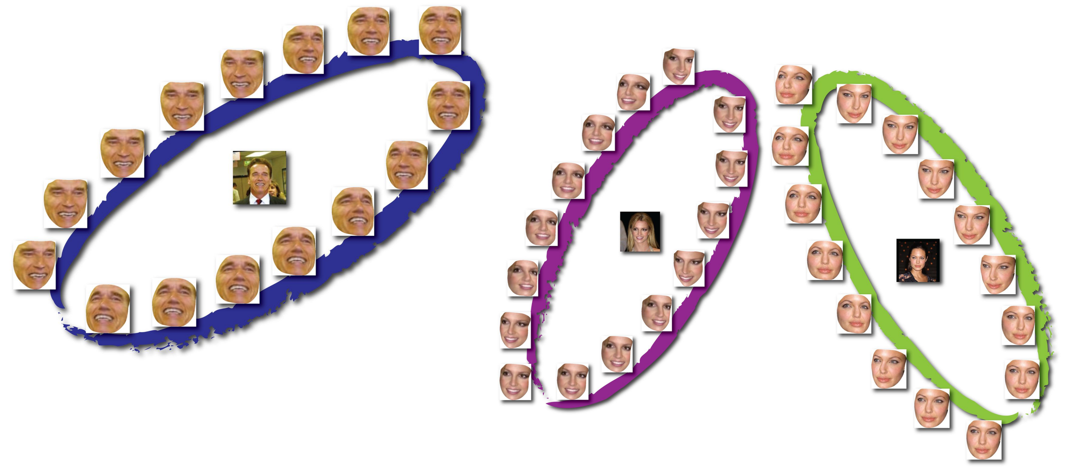
Discriminative Invariant Feature: We extract a single highly discriminative provably group invariant non-linear feature for both tasks from raw pixels.

Invariance to Transformations: Nuisance transformations groups such as the translation, rotation group, increase complexity of the learning problem. Invariance to such transformations can drastically reduce complexity.



The Approach

Linear Invariant Features: Previous work [1] builds linear invariant that are implicitly (but not explicitly) discriminative. When a group of transformations act on an object, they create an orbit.



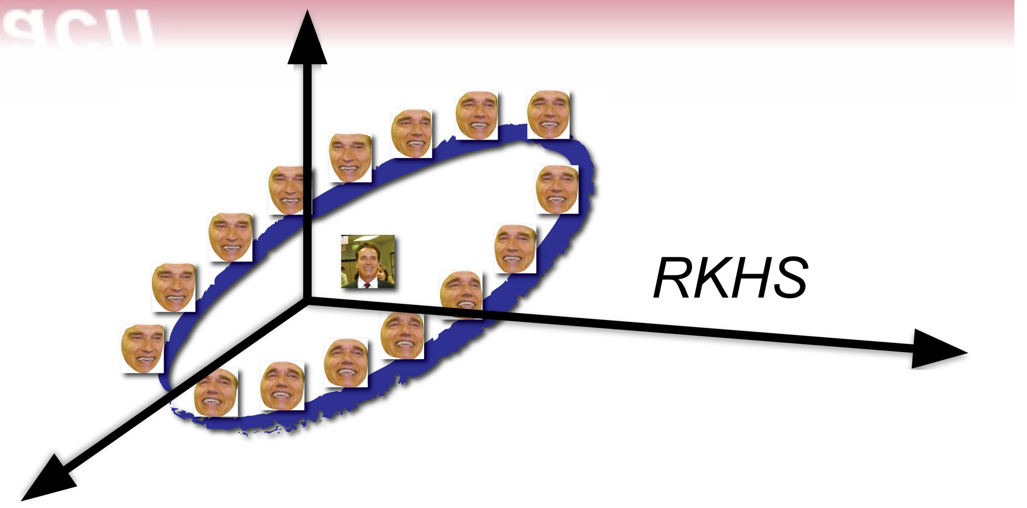
The orbit is unique to the object, and is an invariant to the transformation group

is invariant

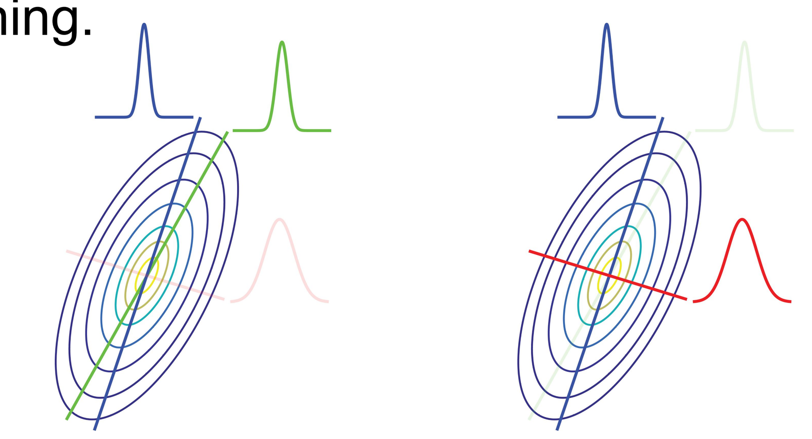
Hence any measure of the orbit is an invariant implicitly discriminative feature.

is invariant

Non-linear Discriminative Invariance: To improve discrimination, we can compute invariant features in the RKHS. We show the discriminative non-linear templates form a group in the RKHS, leading to Discriminative Invariant Kernel Features.



To characterize the orbit, previously simply sampled templates were used. Explicit discrimination provides better matching.



Sampled templates

Discriminatively learned templates

The learnt templates still form a group of transformed templates, hence invariance theory holds.

Definition 3.1 (*Unitary Kernel*). We define a kernel $k(x,y) = \langle \phi(x), \phi(y) \rangle$ to be a unitary kernel if, for a unitary group G, the mapping $\phi(x) : \mathcal{X} \rightarrow \mathbb{H}$ satisfies $\langle \phi(gx), \phi(gy) \rangle = \langle \phi(x), \phi(y) \rangle \; \forall g \in \mathcal{G}, \forall x, y \in \mathcal{X}$.

Theorem 3.2 (DIKF filters form a set of transformed templates in the kernel space under a group). *Given a group G of unitary transformation elements g with $|G| = N$, if $k(x,y) = \langle \phi(x), \phi(y) \rangle$* i.e. *$k$ is a unitary kernel, and $\{\mathbf{X}_n \mid \mathbf{X}_n = g_n(\mathbf{X}), g_n \in G\}$ are a set of pre-whitened matrices acted upon by G, then the set of DIKF filters*

$$\mathcal{T}_k = \left\{ \Phi(\mathbf{t}_{kn}) = \Phi(\mathbf{X}_n) \big(\Phi(\mathbf{X}_n) \cdot \Phi(\mathbf{X}_n) \big)^{-1} \mathbf{u}_k \mid \forall n \right\}$$

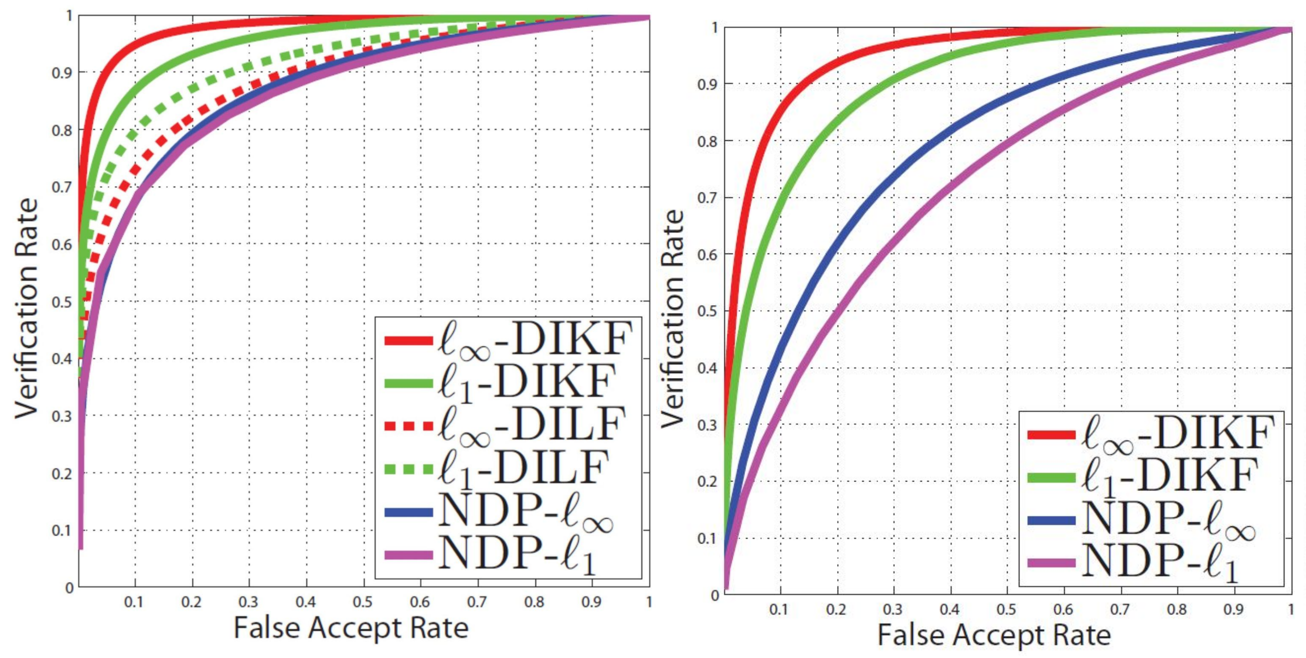
is a set of transformed templates under a group.

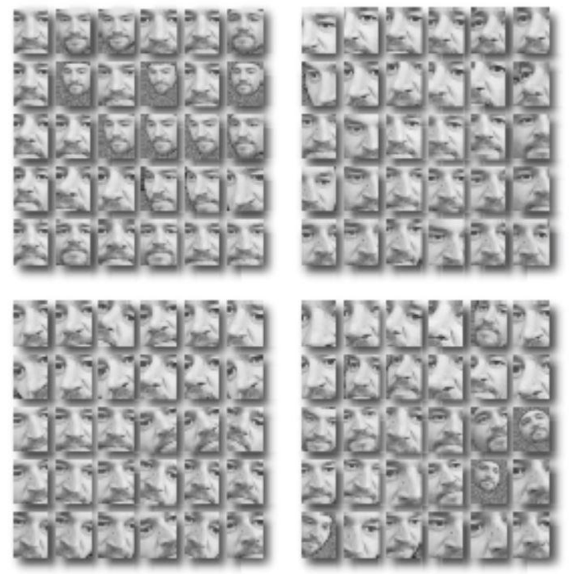
The Experiments

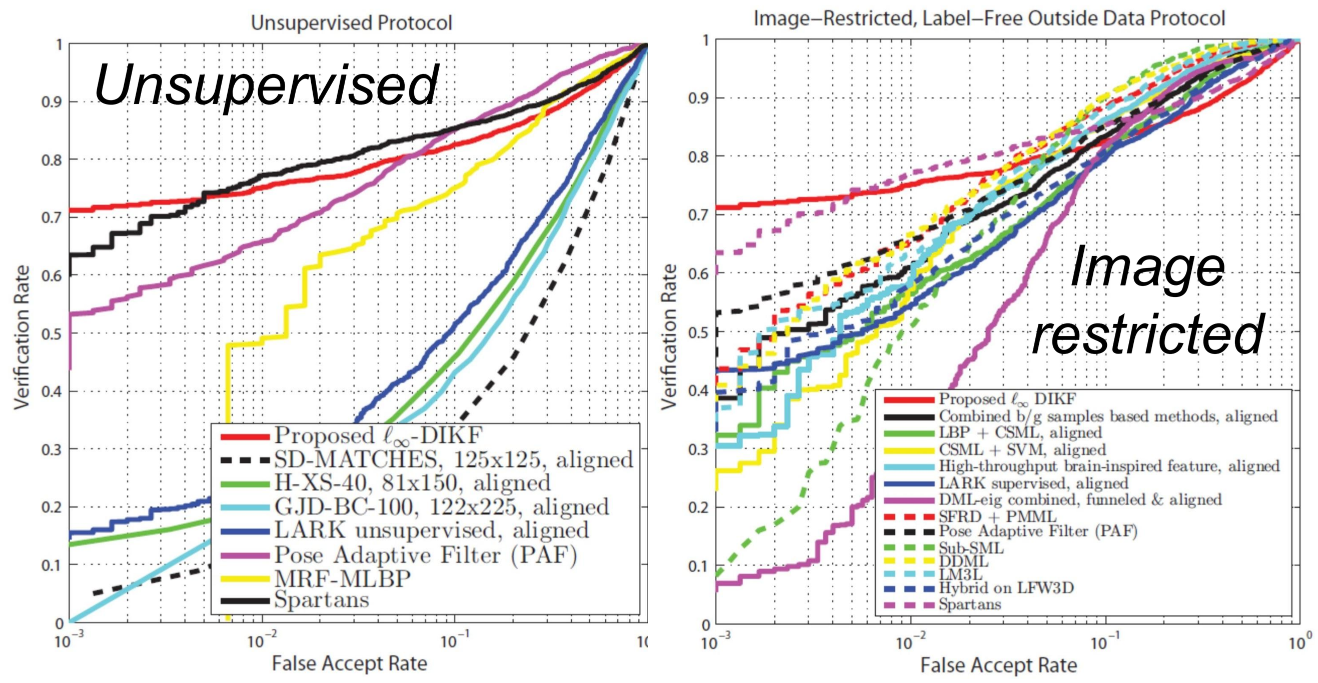
(1) ***Face recognition (153,000 semi-synthetic image dataset)***: 1000 subjects with 153 poses each. Images rendered from a 3D model with real texture. We compare DIKF against sampled templates (NDP) and discriminative linear templates (DILF).

(2) ***Face recognition (LFW)***: Max-pooled DIKF (in red) matches state-of-the-art results on two LFW protocols, despite being simpler than competing methods and working on raw pixels.

(3) ***Pose estimation***: 15 poses (-40 to 40 yaw and -20 to 20 pitch, step of 20). Train on the 250 subjects and test on the 1500 images of the remaining 100 subjects.







Paper link:

Carnegie Mellon University

CyLab Biometrics Center

[1] F. Anselmi, J. Z. Leibo, L. Rosasco, J. Mutch, A. Tacchetti, and T. Poggio. Magic materials: a theory of deep hierarchical architectures for learning sensory representations. MIT, CBCL paper, 2013.