Multiple One-Shots for Utilizing Class Label Information

Yaniv Taigman[1,3]
yaniv@face.com
Lior Wolf[1]
wolf@cs.tau.ac.il
Tal Hassner[2]
hassner@openu.ac.il

[1] The Blavatnik School of Computer Science, Tel-Aviv University, Israel
[2] Computer Science Division, The Open University of Israel
[3] face.com Tel-Aviv, Israel

The One-Shot Similarity (OSS) kernel [3, 4] has recently been introduced as a means of boosting the performance of face recognition systems. Given two vectors, their One-Shot Similarity score (Fig. 1) reflects the likelihood of each vector belonging to the same class as the other vector and not in a class defined by a fixed set of "negative" examples. In this paper we explore how the One-Shot Similarity may nevertheless benefit from the availability of such labels. (a) we present a system utilizing identity and pose information to improve facial image pair-matching performance using multiple One-Shot scores; (b) we show how separating pose and identity may lead to better face recognition rates in unconstrained, "wild" facial images; (c) we explore how far we can get using a single descriptor with different similarity tests as opposed to the popular multiple descriptor approaches; and (d) we demonstrate the benefit of learned metrics for improved One-Shot performance.

We test the performance of our system on the challenging Labeled Faces in the Wild [2] unrestricted benchmark. Using the label information we split the set $\mathbf{A}$ of examples to n sets, $\mathbf{A_i} \subset \mathbf{A}, i = 1..n$, each one containing examples from a single class. The OSS is then computed multiple times, where each time only one subset $\mathbf{A_i}$ is used.

```
function One-Shot-Similarity(I, J,  A)
    Model1 = train(I, A)
    Score1 = classify(J, Model1)

    Model2 = train(J, A)
    Score2 = classify(I, Model2)

    return ½(Score1+Score2)
```

Figure 1: One-Shot score for the vectors **I** and **J** given a set **A** of negatives.

The rational for the split is as follows. The set $\mathbf{A}$ contains variability due to a multitude of factors including pose, identity and expression. During the computation of the (regular) OSS one tries to judge whether $\mathbf{J}$ is more likely to belong to the set containing just the point $\mathbf{I}$ or to the set $\mathbf{A}$. $\mathbf{I}$ contains one person captured at one pose under a particular viewing condition. The classifier trained to distinguish between the two sets can distinguish based on any factor, not necessarily based on the identity of the person. Fig. 2 demonstrates the various OSS scores for pairs of similar/non-similar identities with similar and non similar poses.

Our system consists of the following steps (Fig. 3). We start by aligning the image pair and producing descriptor representations for each of the images. We next employ Information Theoretic Metric Learning (ITML) [1] to provide global alignment of the feature vector coordinates systems. Multiple OSS scores are then computed using different sets of negatives. Finally, the vector of similarity scores is fed to a binary, linear, SVM classifier for the final pair-matching decision.

The ROC curve for our method, using only a single image descriptor, is shown in Fig. 4 (accuracy of 0.8517 ± 0.0061 S_E). In addition, we show results obtained by combining 16 different descriptors and Multiple OSS scores, each trained separately using this same method, into one vector of $16D$. This combination yields 0.8950 ± 0.0051 S_E, which is, to date, the state-of-the-art result for the LFW unrestricted benchmark.

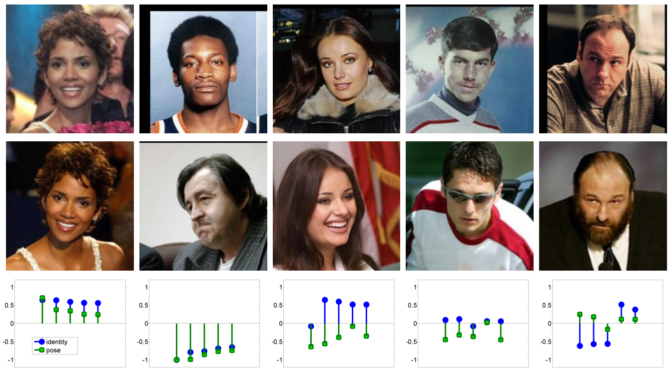

Figure 2: Each group contains two images and 10 sample multiple OSS scores. Identity based multiple OSS scores are plotted with circle markers and pose based are with squares. As can be seen the value of each type of OSS score is a good indication of the type of similarity between the images of the pair. From left to right: (1) Same person, same pose. (2) Different persons and pose. (3) Same person, different pose. (4) Different persons, same pose. (5) Same person and pose, however, a mode of variability not modeled in the system is present.

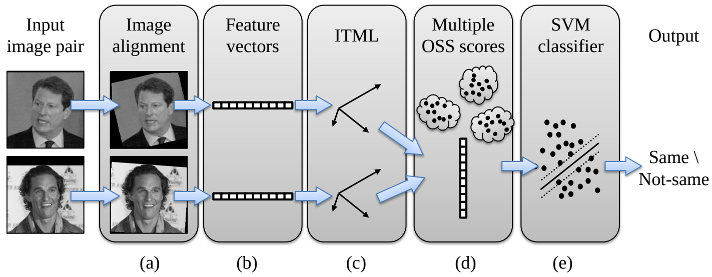

Figure 3: A schematic description of our system. Please see text for more details.

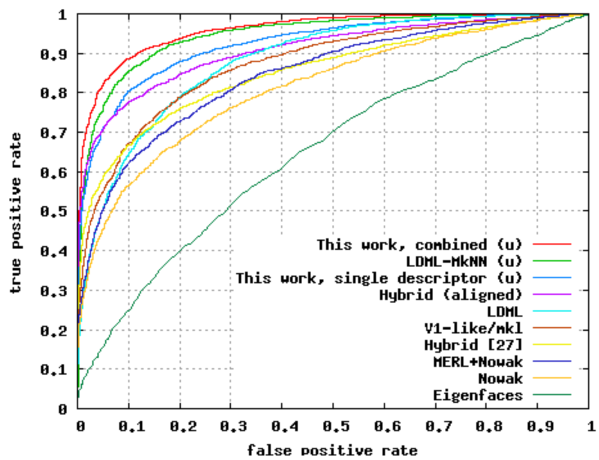

Figure 4: ROC curves. (u) indicates results for the unrestricted setting

[1] J.V. Davis, B. Kulis, P. Jain, S. Sra, and I.S. Dhillon. Information-theoretic metric learning. In *International Conference on Machine Learning (ICML)*, pages 209–216, June 2007.

[2] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, Technical Report 07-49, October, 2007, 2007.

[3] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Faces in Real-Life Images Workshop in ECCV*, 2008.

[4] L. Wolf, T. Hassner, and Y. Taigman. The one-shot similarity kernel. In *International Conference on Computer Vision (ICCV)*, September 2009.