Tom-vs-Pete Classifiers and Identity-Preserving Alignment for Face Verification

Thomas Berg
tberg@cs.columbia.edu

Peter N. Belhumeur
belhumeur@cs.columbia.edu

Columbia University
New York, NY

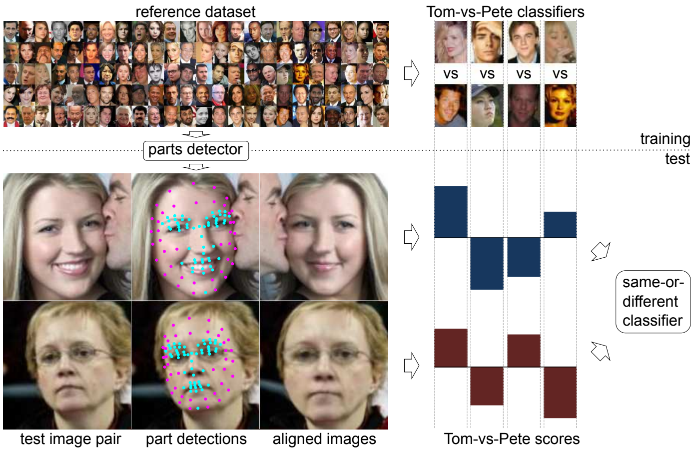

Figure 1: Overview. A reference set of images labeled with parts and identities is used to train a parts detector and a large number of binary "Tom-vs-Pete" classifiers. Given a test image pair, parts are detected and used to perform an "identity-preserving" alignment. The Tom-vs-Pete classifiers are run on the aligned images, and the results are passed to a same-or-different classifier to produce a decision.

In face verification, we are given two face images and must determine whether they are the same person. In this paper, we present a method for face verification that uses a reference dataset of images of other (non-test) people in two novel ways. First, we use part labels on the faces in the training set to perform an *identity-preserving alignment* that reduces differences due to pose and expression, but preserves identity-related differences such as nose width and lip thickness. Second, we train a large set of linear *Tom-vs-Pete classifiers* that are likely to be able to find differences between almost any two people. The outputs of these first-stage classifiers are used as features for a second-stage same-vs-different classifier that makes the verification decision. An overview of the process is shown in Figure 1.

The reference dataset consists of 20,639 images of 120 people, downloaded from the internet. Half the images are from the PubFig [3] "development set," with the rest collected online. In addition to the identity labels, each face is labeled with the locations of 95 parts, including 55 "inner" parts at well-defined points such as the corners of the eyes and mouth, and 40 less well-defined "outer" points around the boundary of the face.

Identity-preserving Alignment

Our alignment procedure is based on a set of part locations on the face. Given a test image, we first find the fifty-five inner parts using the detector of [1]. We then look to the reference dataset, and for each of the 120 reference people, find the image whose inner part locations are closest to the detected parts. This gives a set of 120 images of different people with nearly the same pose and expression. We take the avearage of all ninety-

(a) (b) (c) (d)

Figure 2: Alignment. (a) Original images. (b) Global affine alignment. (c) Alignment using detected parts. (d) Alignment using generic parts.

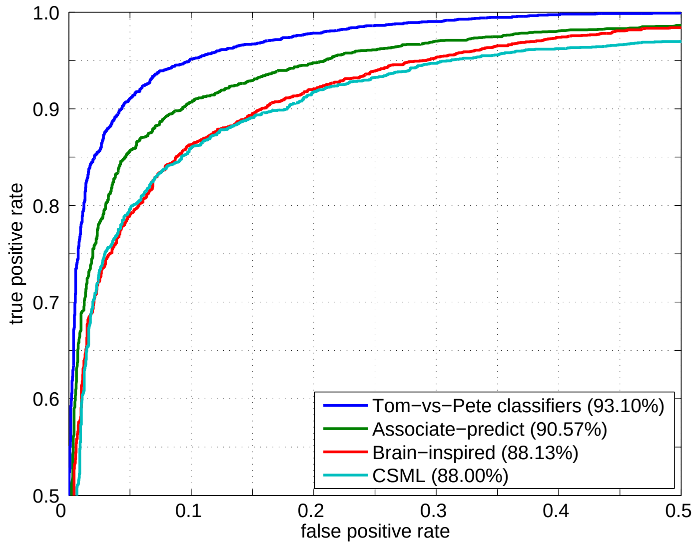

Figure 3: Our results on LFW, compared with the best previously published results [4, 5, 6].

five part locations on these images to be the "generic part" locations for this image – where the parts would be on an average face with the pose and expression of the test face.

Each part has a canonical location, where it occurs in an average, frontal face with neutral expression. To align the image, we perform a piecewise affine warp that takes the generic parts to the canonical part locations. While this sort of alignment using the original detected points can produce a "too-aligned" image in which important, identity-corelated differences like nose width have been removed, we demonstrate in the paper that using the generic parts preserves these differences. This effect is visible in Figure 2, where (c) appears anonymized relative to (b) or (d).

Tom-vs-Pete Classifiers and Verification

Each first-stage classifier is a linear SVM trained on SIFT features from some small region of the face to separate two people. We define eleven such regions, allowing us to train $11 \cdot \binom{120}{2} = 78,540$ classifiers. We call them "Tom-vs-Pete" classifiers to emphasize that each is trained on just two individuals. Intuitively, these classifiers represent a large number of ways in which two people can differ. We describe an adaboost-based heuristic to find a subset of these classifiers that will complement each other and will generalize well to other people, and use this subset to extract features for the second-stage classifier as shown in Figure 1.

We evaluate our system on the Labeled Faces in the Wild (LFW) [2], image-restricted benchmark, obtaining a mean accuracy of $93.10\% \pm 1.35\%$, a 26.86% reduction in error rate relative to the best previously published result [6]. Results are shown in Figure 3.

[1] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *CVPR*, 2011.

[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. Technical report, UMass-Amherst, 2007.

[3] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009.

[4] H. V. Nguyen and L. Bai. Cosine similarity metric learning for face verification. In *ACCV*, 2011.

[5] N. Pinto and D. D. Cox. Beyond Simple Features: A Large-Scale Feature Search Approach to Unconstrained Face Recognition. In *Face and Gesture (FG)*, 2011.

[6] Q. Yin, X. Tang, and J. Sun. An associate-predict model for face recognition. In *CVPR*, 2011.