Face Verification using Large Feature Sets and One Shot Similarity

Huimin Guo[1], William Robson Schwartz[2], Larry S. Davis[1]
[1]Department of Computer Science, University of Maryland, College Park, MD, 20740, USA
[2]Institute of Computing, University of Campinas, Campinas, SP, 13084-971, Brazil
hmguo@cs.umd.edu schwartz@ic.unicamp.br lsd@umiacs.umd.edu

Abstract

We present a method for face verification that combines Partial Least Squares (PLS) and the One-Shot similarity model[28]. First, a large feature set combining shape, texture and color information is used to describe a face. Then PLS is applied to reduce the dimensionality of the feature set with multi-channel feature weighting. This provides a discriminative facial descriptor. PLS regression is used to compute the similarity score of an image pair by One-Shot learning. Given two feature vector representing face images, the One-Shot algorithm learns discriminative models exclusively for the vectors being compared. A small set of unlabeled images, not containing images belonging to the people being compared, is used as a reference (negative) set. The approach is evaluated on the Labeled Face in the Wild (LFW) benchmark and shows very comparable results to the state-of-the-art methods (achieving 86.12% classification accuracy) while maintaining simplicity and good generalization ability.

1. Introduction

Face recognition research [30, 23] is driven by its variety of applications in areas such as public security, human computer interaction, and financial security. The two primary face recognition tasks are *identification* (*1:N matching* problem) and *verification* (*1:1 matching* problem). In the *identification* task, a probe image is matched against a set of labeled faces in a gallery set, and is identified as the person presenting the highest similarity score. In the *verification* task, given two face images, the goal is to decide whether these two images are of the same person or not. The method described in this paper addresses the verification task.

One of the most challenging problems in face recognition is with uncontrolled images. Recently, the Labeled Faces in the Wild (LFW)[8] dataset was released as a benchmark for the face verification (pair-matching) problem. The LFW images include considerable visual variations caused by, for example, lighting, pose, facial expression, partial occlusion, aging, scale, and misalignment. Figure 1 contains some examples of pairs of images from the same person that differ in lighting, pose, facial expression and partial occlusion.

The main challenges in face verification are the small-sample-size problem, since only a few (often a single) samples are available per subject, and large variations between the two images to be compared. Thus, one major issue in face verification is to find efficient and discriminative facial descriptors that can cope with problems caused by uncontrolled acquisition and are robust to the small-sample-size problem. The use of a rich descriptor set based on multiple feature channels has shown improvements over single feature channels for recognition [18, 22, 17]. In the method proposed by Schwartz et al. [17] to tackle the face identification task, these separate descriptors are concatenated into a feature vector. Then, feature weighting is performed using Partial Least Squares (PLS), which handles very high-dimensional data and works well when very few samples are available [26, 18]. In this work, we use PLS to discriminatively weight the large and rich feature set (facial descriptor).

Another important issue in face verification is learning an appropriate similarity measure. Most popular methods tailor the similarity measure to available training data by applying learning techniques [28]. In such methods, testing is performed using models (or similarity measures) learned beforehand. The other trend is to learn from one or very few training examples. Wolf et al. [29, 28] introduced One-Shot Similarity (OSS) which learns discriminative models exclusive to the vectors being compared, by using a set of background samples. In our work, we use this One-Shot framework to learn models for feature vectors representing face samples on-the-fly. The prediction scores are computed from Partial Least Squares Regression.

There are several advantages of our method: (1) It is 'unsupervised'. No labeled training set is needed, either

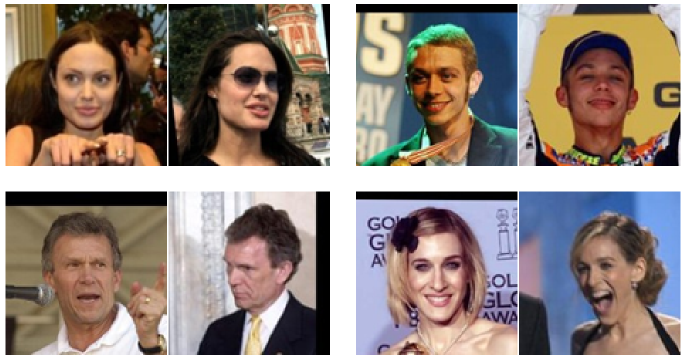


Figure 1. Examples of some images from the LFW dataset with variations in: Top row:(left) partial occlusion, (right)lighting; bottom row: (left) pose, (right) expression. Each corner shows a different subject. Note that each pair is from the same person.

matched/mismatched pair labels or identity information. All we need is a small unlabeled reference set. The discriminative models are learned online and exclusively for the pair being compared. (2) PLS has been shown, experimentally, to be robust to modest pose variations[19], expression, illumination, aging and other uncontrolled variations[17]. (3) There is almost no parameter tuning with PLS. The only parameter is the number of factors.

The rest of this paper is organized as follows. In Section 2, we provide a literature review on recent work on face verification, especially those evaluated on the LFW benchmark. In Section 3, we present our feature extraction procedure and review Partial Least Squares Regression. Then, we describe the proposed face verification approach: PLS One-Shot model. Finally, we present experimental results in Section 4 and draw conclusions in Section 5.

2. Related Work

There has been a significant amount of relevant works on face verification [22, 10, 5, 28, 25, 6, 2, 29, 13]. Here we briefly review those state-of-the-art methods that have been evaluated on the LFW benchmark. At the end of this section we also review some recent work using PLS in face identification[17, 19].

Some work focuses principally on face descriptors [13, 2, 25]. Pinto[13] et al. combined variants of intensity and V1-like models. The classification of face images was performed using large-scale multi kernel learning (MKL) associated with a support vector machine (SVM). In [2], an unsupervised learning-based encoding method was proposed to encode the micro-structures of a face with a single or a combination of multiple descriptors. In [25], Patterns of Oriented Edge Magnitudes (POEM) was introduced. The POEM feature is built by applying a self-similarity based structure on oriented magnitudes, calculated by accumulating a local histogram of gradient orientations over all pixels of image cells, centered on the pixel. Other works have employed metric learning[5, 11, 6] for learning similarity functions for verification. Guillaumin et al.[5] presented two methods for learning robust distance measures: (1) LDML: a logistic discriminant approach which learns the metric from a set of labeled image pairs and (2) MkNN: a nearest neighbour approach which computes the probability for two images belonging to the same class. In [6], a part based face representation (densely sampled overlapping image patches) is computed to enable elastic and partial matching. The distance metric is defined as each descriptor in one face is matched against its spatial neighborhood in the other face. [11] presented the Cosine Similarity Metric Learning as an alternative to Euclidean distance. The idea was to learn a transformation matrix by minimizing the cross-validation error with a gradient-based optimization algorithm.

Some of the best performing algorithms focus on the classifier design[10] and learning more discriminative models[29, 28, 21]. Kumar et al. [10] designed two methods: *attribute* classifiers, which are trained to recognize describable aspects of visual appearance, and *simile* classifiers, trained to recognize the similarity of faces, or regions of faces, with respect to specific reference people. Wolf et al.[29, 28, 21] introduced One-Shot Similarity [28] to learn discriminative models exclusive to the vectors being compared, by using a set of background samples. In [28], they used a random-patch based image representation with OSS as the similarity score and a SVM to classify. In [29], the OSS was extended to 'Two-Shot Similarity'(TSS). Also, the authors used the ranking of images most similar to a query image and employed these as a descriptor for that image. The best verification result was obtained by adding SVM based OSS and TSS to LDA.

Schwartz et al.[17] proposed an approach to robust face identification based on a very large feature set using Partial Least Squares (PLS) to perform multi-channel feature weighting. Their results showed that PLS works well with only a single image per sample, in large galleries, and under different conditions, particulary when the data is acquired under uncontrolled conditions. More recently, in [19] the authors used PLS to linearly map images in differently modalities to a common linear subspace in which they are highly correlated. The work showed, in theory, that there exist linear projections of images taken in two modalities that map them to a space in which images of the same individual are very similar. In that work, PLS was shown to work well across modalities: high resolution vs. low resolution, pose variation, and in comparing images to sketches.

3. Proposed Method

3.1. Overview of the Framework

The pipeline of our PLS One-Shot Model based face verification approach is presented in Figure 2. First, a randomly selected set of images **A** (approximately 500 images from LFW) is set aside as background samples [29]. The images in this set are unlabeled and considered as 'negative' examples. It should not contain any images from individuals to be compared subsequently. All images in this set are aligned, cropped and their feature vectors are extracted and stored.

When a new pair of face images is presented, their two feature vectors are extracted and their similarity score is calculated by building PLS One-Shot Models using each of the vectors versus the set **A** and projecting the other vector to the other's model. The two projections provide the responses of the PLS regression models. The average of these two scores is used to measure the similarity of the image pair.

Finally, if the similarity score is above threshold, a match is declared; a non-match is declared, otherwise.

3.2. Feature Extraction

After cropping the faces, each sample is decomposed into overlapping blocks and a set of low-level feature descriptors is extracted from each block. Features used are: Gabor features, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG).

Gabor filters[22], which are spatially localized and selective to spatial orientations and scales, are comparable to the receptive fields of simple cells in the mammalian visual cortex. Due to their robustness to local distortions, Gabor features have been successfully applied to face recognition. Normally 5 scales and 8 orientations of Gabor filters are used. Convolving the face with each of these 40 Gabor filters generates the Gabor features. For each Gabor filter, one value is computed at each pixel position. In practice, the Gabor features are downsampled by a factor (e.g., 4).

Local Binary Patterns (LBP) [1] have been successfully applied in many applications. LBP characterizes the spatial structure of local image texture and is invariant to monotonic transformations of the pixel gray values.

Histogram of Oriented Gradients (HOG) captures edge or gradient structures that are characteristic of local shape [3]. Since the histograms are computed for regions of a given size, HOG is robust to some location variability of face parts. HOG is also invariant to rotations smaller than the orientation bin size. The 8-bin HOG is used.

After the feature extraction process is performed for all blocks inside a face, feature descriptors are concatenated creating a high-dimensional feature vector $\boldsymbol{v}$ to describe the face. Partial Least Squares will give higher weights to more discriminatory features when building each model.

3.3. Partial Least Squares Regression

Partial Least Squares (PLS) regression is a multivariate data analysis technique which can be used to relate several response ($\boldsymbol{y}$) variables to several predictor ($\boldsymbol{X}$) variables. The basic idea of PLS is to construct new predictor variables, latent variables, as linear combinations (weighed sum) of the original variables summarized in a matrix $\boldsymbol{X}$ of predictor variables (features) and a vector $\boldsymbol{y}$ of response variables. PLS can deal efficiently with a small number of examples and a large number of variables. Detailed descriptions of the PLS method can be found in [20, 26, 14, 4].

Let $\mathcal{X} \subset \mathbb{R}^m$ denote an m-dimensional feature space and let $\mathcal{Y} \subset \mathbb{R}$ be a 1-dimensional vector representing the response variable. Let the number of samples be n. PLS decomposes a zero-mean matrix $\boldsymbol{X}_{n\times m} \in \mathcal{X}$ and zero-mean vector $\boldsymbol{y}_{n\times 1} \in \mathcal{Y}$ into

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}$$
$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{q}^T + \boldsymbol{f}$$

where $\boldsymbol{T}$ and $\boldsymbol{U}$ are $n \times p$ matrices containing p extracted latent vectors, the $(m \times p)$ matrix $\boldsymbol{P}$ and the $(1 \times p)$ vector $\boldsymbol{q}$ represent the loadings. The $n \times m$ matrix $\boldsymbol{E}$ and the $n \times 1$ vector $\boldsymbol{f}$ are noise terms, or the residuals.

The PLS method, which in its classical form is based on the nonlinear iterative partial least squares (NIPALS) algorithm [26], constructs a set of weight vectors, stored in the weight (projection) matrix $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p)$, such that

$$[cov(\boldsymbol{t}_i, \boldsymbol{u}_i)]^2 = \max_{|\boldsymbol{w}_i|=1} [cov(\boldsymbol{X}\boldsymbol{w}_i, \boldsymbol{y})]^2 \quad (1)$$

where $\boldsymbol{t}_i$ is the i-th column of matrix $\boldsymbol{T}$, $\boldsymbol{u}_i$ the i-th column of matrix $\boldsymbol{U}$. $cov(\boldsymbol{t}_i, \boldsymbol{u}_i) = \boldsymbol{t}_i{}^T\boldsymbol{u}_i/n$ denotes the sample covariance between latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$. After extracting the latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$, the matrix $\boldsymbol{X}$ and vector $\boldsymbol{y}$ are deflated by subtracting their rank-one approximations based on $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$. This process is repeated until the desired number (p) of latent vectors have been extracted.

In other words, partial least squares regression shrinks the predictor matrix by sequentially extracting orthogonal

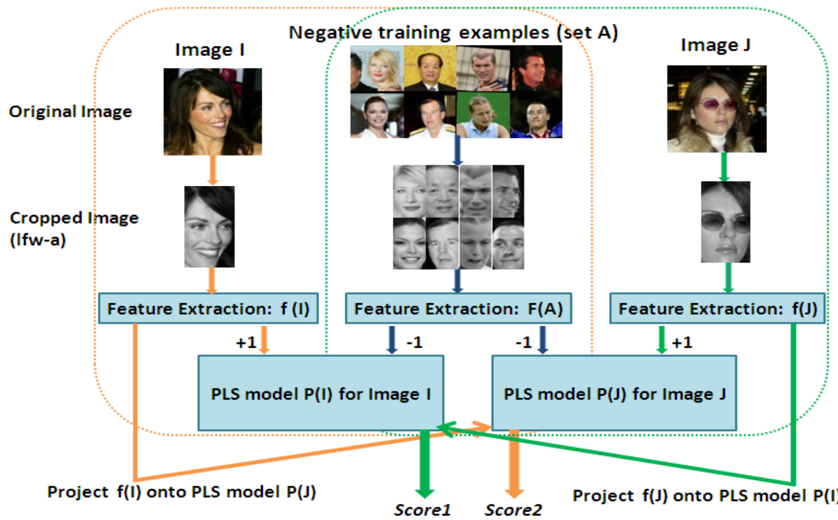


Figure 2. Our processing pipeline for face verification using the PLS One-Shot Model.

components which at the same time summarize the explanatory variables and allow modeling and predicting the response variables. Finally, it provides a classical regression equation, in which the dependent (response) variable $\boldsymbol{y}$ is estimated as a linear combination of a set of predictor variables (features) matrix $\boldsymbol{X}$.

The regression model is given by [14]

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{f} \tag{2}$$

where $\boldsymbol{B}_{m\times 1}$ is the matrix of regression coefficients, Algebraic manipulations yield

$$\boldsymbol{B} = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{T}^T\boldsymbol{y}. \tag{3}$$

The regression response, y_v, for a new observed feature vector $\boldsymbol{v}$ is obtained by

$$y_v = \overline{y} + \boldsymbol{B}^T\boldsymbol{v} \tag{4}$$

where $\overline{y}$ is the sample mean of $\boldsymbol{y}$.

The face verification problem is considered as a two-class problem: same and not-same. Thus the sample's class labels $\boldsymbol{y}$ can be set as +1 or -1. We use the PLS regression algorithm as a feature weighing tool. The regression response y_v of a PLS model is used to measure the similarity of a given image pair. Therefore, no extra classifier is needed.

3.4. Face Verification with PLS One-Shot Model

To decide whether the images of two faces **I** and **J** are of the same individual or not, traditional methods for face verification use a large training set to learn models (or similarity measures) beforehand, and then employ this model to calculate the similarity of images **I** and **J**. In contrast, we learn the models for the images to be matched on-the-fly using the PLS One-Shot Model (PLS + OSS). The key idea behind the OSS[28] is to use a set **A** of *negative* training examples not containing images belonging to the people being compared. The details of OSS are described in [28].

To perform face verification, we use the given training set as the fixed *negative* set (or reference set) **A**, see Figure 2. When a new pair of face images **I** and **J** is to be matched, they are cropped to the same size and the features are extracted. Then, a discriminative PLS Regression model is learned by taking **A** as the negative samples (with labels -1) and **I** to be the single positive sample (with label +1). Then, **J** is evaluated by this model to obtain a response (similarity score). This score gives a measure of how likely **J** shares the same label as **I** or belongs to the negative set (which means **J** might be very different from **I**). Symmetrically, we switch the roles of **I** and **J** and execute the same procedure. The final similarity score for this pair is the average of the two scores.

4. Experimental Results

In this section, we evaluate our PLS One-Shot model based face verification on the LFW.

4.1. LFW Dataset

The Labeled Faces in the Wild (LFW)[8] dataset contains 13,233 face images labelled by the identity of the person. The faces show large variation in pose, expression, lighting, occlusion, and aging. There are three versions of the LFW dataset available: original, funneled and aligned. Wolf et al.[29] showed that the aligned version (lfw-a) is better than the funneled version in dealing with misalignment. Therefore, we use the lfw-a n in all of our experiments.

The dataset comes with a division into 10 fully independent splits (folds) that can be used for cross validation [9]. Using only the given image pairs in the training set is referred to as the **image-restricted** paradigm; in this case, it is known whether an image pair belongs to the same person or not, while identity information is not used at all. The **unrestricted** paradigm refers to training methods that can use all available data, including the identity of the people in the images. Additionally, there is an **unsupervised** paradigm when there is no supervised information, such as in the form of same/not-same labels used. As an example of the unsupervised paradigm, in[15], the authors randomly selected 100 images from LFW for the Borda-count method that was used together with the Gabor descriptor. The 100 images were used simply as a reference set; their pair or identity information was not used.

In our evaluation, while performing each independent fold, we randomly choose 500 images from the training set (other 9 splits, 5400 image pairs) without using their pair information. The number 500 is chosen because experiments with several datasets show sufficiently good performance when the 'negative' set contains 300 to 1000 images. Then, these images are fixed as the 'negative' set (background samples) for this fold. According to the protocol, the 10 splits are mutually exclusive with respect to subject identities. Below, we present results using both the unsupervised and the image restricted paradigms.

4.2. Preprocessing

In our evaluation we consider three different crop regions: (1) centered face region with hair (2) centered face region without mouth (3) centered face region without mouth and hair. Figure 3 shows the three different croppings: 80×148 (crop1), 80×110 (crop2), 80×64 (crop3). There are pros and cons for each different cropping: removing some parts like the mouth or hair region could help alleviate effects due to expression and hat occlusion, while mouth/chin and hair style might also include some informative features. We tried these croppings and fused the three scores in a simple way (rely more on full region than the other two partial regions, fusion gives about 1% improvement from 'crop1 only', see Table 1):

$$finalscore = score(crop1) + score(crop2)/2 + score(crop3)/2$$

For illumination normalization, our experiments indicated that the un-normalized images and images filtered by Difference of Gaussian give similar results with PLS One-Shot model. We report the results with DoG, as in [6, 2]. Since there is significant pose variation within LFW, we additionally use the flipped (mirror) image. For example, when comparing image pair **I** and **J**, we also compare **I** and the flipped image of **J**. Then the average of the two scores is taken as the final similarity score. We will show this simple flip step improves performance.

4.3. Experimental Setup

For HOG features, we use block sizes of 16×16 and 8×8 with strides of 4 and 4 pixels, respectively. For LBP features, we use block size of 16×16 with strides of 8 pixels. The Gabor features have 5 scales and 8 orientations, down sampled by a factor of 4. The PLS factor (number of latent vectors p) is set to 11.

In prior work on the LFW benchmark, algorithms are typically evaluated by ROC curves and the classification accuracy (true positive rate) at the Equal Error Rate (EER). EER is the location on the ROC curve where the false positive rate and false negative rate are equal. We report our results in the form of ROC curves and the estimated mean classification accuracy and the standard error of the mean for the 10 cross-validation folds in View 2 of the dataset.

4.4. Comparison with the State-of-the-art Methods

As stated previously, we only uses a very small number of images from the training set as a reference set. No pair label or identity is used. Thus, we compare our method using the unsupervised paradigm. PLS One-Shot method outperforms other methods using the unsupervised paradigm by a large margin. At the same time, its performance is comparable to the best results using the image-restricted paradigm whose methods use pair information.

Comparison with Unsupervised paradigm. Table 1 shows the classification accuracy (at EER) of our method in comparison with other methods using the unsupervised paradigm. Figure 4(a) presents the ROC curve of our approach (pink line), along with the ROC curves of previous methods. As can be seen, PLS One-Shot outperforms the other methods by a very large margin. Similar to these methods, only a very small set from the LFW is used as a reference set. No other supervised information is used. Our PLS One-Shot model based face verification approach is simple and effective for this challenging real-world dataset.

Comparison with Image-Restricted paradigm. Since many state-of-the-art methods use the pair information and report their results using the Image-Restricted paradigm, we compare our results with them too. Table 2 shows the classification accuracy of our method in comparison with those methods with the Image-Restricted paradigm. Figure 4(b) contains the ROC curve of our approach (blue line), along with the ROC curves of previous methods with the Image-Restricted paradigm.

The results show that our approach is comparable with the state-of-the-art methods on the LFW benchmark (we achieved 86.12% classification accuracy). On the LFW

Figure 3. Examples of face images with different croppings. Left to right: 80×148 (crop1), 80×110 (crop2), 80×64 (crop3).

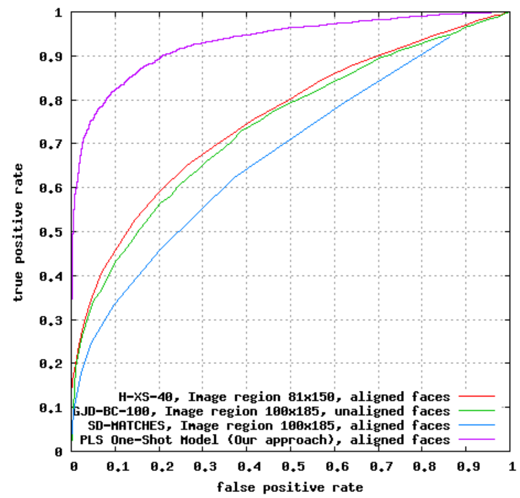
Method	Accuracy
SD-MATCHES, aligned [15]	0.6410±0.0042
H-XS-40, aligned [15]	0.6945±0.0048
GJD-BC-100, aligned [15]	0.6847±0.0065
Our method (PLS + OSS, crop1 only)	**0.8418±0.0052**
Our method (PLS + OSS)	**0.8533±0.0038**
Our method (PLS + OSS, flip)	**0.8612±0.0047**

Table 1. Mean (± standard error) classification accuracy on the LFW dataset, Unsupervised Training benchmark using the PLS One-Shot Model, and the same model except the addition of the flipped image idea. The 'crop1 only' gives the result of the main cropping.

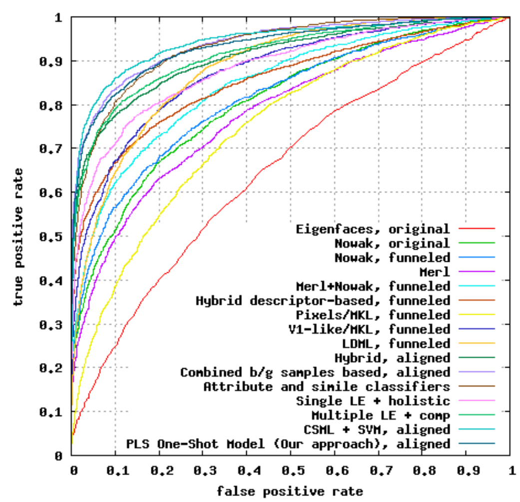
Method	Accuracy
Eigenfaces, original[24]	0.6002±0.0079
Nowak, funneled [12]	0.7393±0.0049
MERL+Nowak, funneled [7]	0.7618 ±0.0058
Hybrid Descriptor, funneled [27]	0.7847±0.0051
Multi-Region Histograms [16]	0.7295±0.0055
V1-like/MKL [13]	0.7935±0.0055
LDML, funneled [5]	0.7927±0.0060
SVM + OSS [28]	0.7637±0.0065
POEM, aligned [25]	0.7542 ±0.0071
Hybrid, aligned [21]	0.8398±0.0035
Combined b/g samples based [29]	0.8683±0.0034
Attribute and Simile classifiers [10]	0.8529± 0.0123
Single LE + holistic, aligned [2]	0.8122±0.0053
Multiple LE + comp, aligned [2]	0.8445±0.0046
CSML + SVM, aligned [11]	0.8800 ±0.0037
Our method (PLS + OSS)	**0.8533±0.0038**
Our method (PLS + OSS, flip)	**0.8612±0.0047**

Table 2. Mean (± standard error) classification accuracy on the LFW dataset, compared to Image-Restricted Training benchmark using the PLS One-Shot Model.

benchmark, Wolf's work[29] and the recently published CSML[11] have the best performance. Wolf's model has several layers and requires a large amount of training data. CSML uses the View 2 training data intensively to conduct their cross-validation error minimization based metric learning. Kumar[10] shows excellent result, marginally lower than ours. However, Kumar's work requires train-



(a)



(b)

Figure 4. ROC curves for View 2 of the LFW dataset. Each point on the curve represents the score over the 10 folds (of false positive rate, true positive rate) for a fixed threshold. (a)Unsupervised paradiam (b) Image-Restricted paradiam. The plots are gernerated from results reported on http://vis-www.cs.umass.edu.lfw.results.html.

ing high-level classifiers requiring a huge volume of images outside of the LFW dataset. The LE [2] method in the component-level relies on facial feature point detectors that have been trained with supervision. Overall, our approach achieves competitive accuracy without using any label information or local feature identification. Thus it could be easily generalized to other datasets.

5. Conclusions and Discussion

We proposed a robust face verification approach based on PLS One-Shot model. This model is versatile - it performs multi-channel feature weighting on a rich feature set and the PLS regression response can be used efficiently to construct a similarity measure. The One-Shot learning builds discriminative models online exclusively to the pair being compared. A small set of unlabeled images used as the reference (negative) set is all that is needed. The approach was evaluated on the LFW benchmark and showed very comparable results to the state-of-the-art methods (image-restricted setting). When compared with other methods using the unsupervised setting, the proposed method outperformed them by a large margin.

Acknowledgments

This work was partially supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). All statements of fact, opinion or conclusions contained herein should not be construed as representing the official views or policies of IARPA and the ODNI. This research was also partially supported by FAPESP grant 2010/10618-3.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *ECCV'04*, 35, pages 469–481, 2004.

[2] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *CVPR'10*, pages 2707–2714, 2010.

[3] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR'05*, pages 886–893, 2005.

[4] L. Elden. Partial least-squares vs. Lanczos bidiagonalization–I: analysis of a projection method for multiple regression. *Computational Statistics & Data Analysis*, 46(1):11 – 31, 2004.

[5] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In *ICCV'09*, pages 498 –505, 2009.

[6] G. Hua and A. Akbarzadeh. A Robust Elastic and Partial Matching Metric for Face Recognition. In *International Conference on Computer Vision (ICCV)*, 2009.

[7] G. B. Huang, M. J. Jones, and E. Learned Miller. LFW Results Using a Combined Nowak Plus MERL Recognizer. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille France, 2008.

[8] 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 07-49, University of Massachusetts, Amherst, October 2007.

[9] 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 07-49, University of Massachusetts, Amherst, October 2007.

[10] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV'09*, 2009.

[11] H. V. Nguyen and L. Bai. Cosine Similarity Metric Learning for Face Verification. In *ACCV*. LNCS, 2010.

[12] E. Nowak. Learning visual similarity measures for comparing never seen objects. In *Proc. IEEE CVPR*, 2007.

[13] N. Pinto, J. DiCarlo, and D. Cox. How far can you get with a modern face recognition test set using only simple features? In *CVPR'09*, volume 0, pages 2591–2598, Los Alamitos, CA, USA, 2009.

[14] R. Rosipal and N. Kramer. Overview and recent advances in partial least squares. *Lecture Notes in Computer Science*, 3940:34–51, 2006.

[15] J. Ruiz-del Solar, R. Verschae, and M. Correa. Recognition of faces in unconstrained environments: a comparative study. *EURASIP J. Adv. Signal Process*, 2009:1:1–1:19, January 2009.

[16] C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. In *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, pages 199–208, Berlin, Heidelberg, 2009. Springer-Verlag.

[17] W. R. Schwartz, H. Guo, and L. Davis. A robust and scalable approach to face identification. In *ECCV'10*, volume 5, pages 476–489, Sept. 2010.

[18] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human detection using partial least squares analysis. In *ICCV'09*, 2009.

[19] A. Sharma and D. Jacobs. Pls based multi-modal face recognition. In *CVPR'11*, 2011.

[20] K. Strimmer, A. laure Boulesteix, and A. laure Boulesteix. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. In *Briefings in Bioinformatics*, pages 32–44, 2007.

[21] Y. Taigman, L. Wolf, and T. Hassner. Multiple one-shots for utilizing class label information. In *British Machine Vision Conference (BMVC)*, 2009.

[22] X. Tan and B. Triggs. Fusing Gabor and LBP feature sets for kernel-based face recognition. In *AMFG'07*, pages 235–249, 2007.

[23] A. Tolba, A. El-Baz, and A. El-Harby. Face recognition: A literature review. *International Journal of Signal Processing*, 2:88–103, 2006.

[24] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591. IEEE Comput. Sco. Press, 1991.

[25] N.-S. Vu and A. Caplier. Face recognition with patterns of oriented edge magnitudes. In *Proceedings of the 11th European conference on Computer vision: Part I*, ECCV'10, pages 313–326, 2010.

[26] H. Wold. Partial least squares. In S. Kotz and N. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 6, pages 581–591. Wiley, New York, 1985.

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

[28] L. Wolf, T. Hassner, and Y. Taigman. The one-shot similarity kernel. In *ICCV'09*, Sept. 2009.

[29] L. Wolf, T. Hassner, and Y. Taigman. Similarity scores based on background samples. In *ACCV'09*, 2009.

[30] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.