Similarity Scores based on Background Samples

Lior Wolf[1*], Tal Hassner[2], and Yaniv Taigman[1,3]

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

Abstract. Evaluating the similarity of images and their descriptors by employing discriminative learners has proven itself to be an effective face recognition paradigm. In this paper we show how "background samples", that is, examples which do not belong to any of the classes being learned, may provide a significant performance boost to such face recognition systems. In particular, we make the following contributions. First, we define and evaluate the "Two-Shot Similarity" (TSS) score as an extension to the recently proposed "One-Shot Similarity" (OSS) measure. Both these measures utilize background samples to facilitate better recognition rates. Second, we examine the ranking of images most similar to a query image and employ these as a descriptor for that image. Finally, we provide results underscoring the importance of proper face alignment in automatic face recognition systems. These contributions in concert allow us to obtain a success rate of 86.83% on the Labeled Faces in the Wild (LFW) benchmark, outperforming current state-of-the-art results.

1 Introduction

In a learning framework, we define *background samples* as samples that do not belong to the classes being learned. Collecting such samples is often easy as they do not require labeling. For example, in a face identification scenario, these samples could be a face set of individuals not among those which the system is being trained to recognize. Besides being easy to collect, we believe such examples may provide valuable information about which images may be considered "the same" and which may not. In this paper we present similarity measures designed to exploit such background samples. These measures are then shown to outperform state-of-the-art techniques on standard image similarity tests.

Why would background samples be useful for defining similarity functions? The sample vectors are embedded in a vector space in which various metrics can be employed. In order to know which metric is most suitable for the similarity task at hand, the underlying structure of the manifold on which the samples reside needs to be analyzed. Supervised learning can sometimes be used, but may require extra labeling information. On the other hand, background samples without additional information directly answer questions such as "is this sample closer to that one than to a typical example from the background set?" (One-Shot); "are these two examples well separated from the background sample set?"

(Two-shot); and "do these two samples have similar sets of neighboring samples in the background set?" (ranking similarity).

As a benchmark for testing our methods, we use the Labeled Faces in the Wild (LFW) database [1]. It offers a unique collection of annotated faces captured from news articles on the web. It can be used to estimate face recognition performance on faces detected automatically in web images, and may serve as a reasonable benchmark for photo album applications. The dataset is published with a specific benchmark, which focuses on the face recognition task of *pair matching*. In this task, given two face images, the goal is to decide whether the two pictures are of the same individual. This is a binary classification problem, with two possible outcomes: "same" or "not-same".

The best results currently reported on the LFW benchmark were obtained by [2] using the One-Shot Similarity measure. Our tests here on the same benchmark indicate that exploiting background samples yields improved performance.

The rest of the paper is organized as follows. Section 2 reviews related work. In Section 3 we describe the novel Two-Shot similarity measure and its efficient computation. Using image ranking as an additional image descriptor is proposed in Section 4. Section 5 discusses the importance of accurate face alignment for recognition. We present our results in Section 6 and finally conclude in Section 7.

2   Related Work

The literature on similarity functions is extensive. Some similarity measures proposed in the past have been hand crafted (e.g., [3, 4]). Alternatively, a growing number of authors have proposed tailoring similarity measures to available training data by applying learning techniques (e.g., [5–9]). In all these methods testing is performed using models (or similarity measures) learned beforehand.

Recently [10, 11], the One-Shot Similarity (OSS) score was introduced as an alternative approach which utilizes background samples. The OSS draws its motivation from the growing number of so called "One-Shot Learning" techniques; that is, methods which learn from one or few training examples (see for example [12, 13]). Unlike previous methods for computing similarities, the OSS score of two signals is computed by training a discriminative model *exclusive* to the two signals being compared, by using a set of background samples. It was consequently shown to be instrumental in obtaining state-of-the-art results on the Labeled Faces in the Wild (LFW) image pair-matching challenge [1].

Employing background samples differs from semi-supervised learning [14] and from transductive learning [15] since in both cases the unlabeled samples belong to the set of training classes. It differs from flavors of transfer learning that use unlabeled samples [16], since they use separate supervised learning tasks in order to benefit from the unlabeled set.

Although learning with background samples can be seen as belonging to the group of techniques called "learning with side-information", it differs from existing methods in the literature known to us. In particular, some of the previous

contributions, e.g., [17–19], require having training samples with the same identity. Other side-information contributions, e.g., [20] assume that the variability in the side information differs from that in the relevant data.

3   The Two-Shot Similarity Score

We begin our description of the TSS measure by reviewing the OSS [10, 11]. Given two vectors **I** and **J** their OSS score is computed by considering a training set of background sample vectors **A**. This set contains examples of items not belonging in the same class as neither **I** nor **J**, but are otherwise unlabeled. A measure of the similarity of **I** and **J** is then obtained as follows. First, a discriminative model is learned with **I** as a single positive example, and **A** as a set of negative examples. This model is then used to classify the vector, **J**, and obtain a confidence score. The nature of this score depends on the classifier used. Using linear SVM, for example, this score may be the signed distance of **J** from the hyperplane separating **I** and **A**. A second such score is then obtained by repeating the same process with the roles of **I** and **J** switched. The final OSS score is the average of these two scores.

The Two-Shot similarity score is obtained in a single step by modifying the process described above. Again, we consider the same auxiliary set of negative examples **A**. This time, however, we train a single discriminative model using both **I** and **J** as positive examples, and the set **A** as a set of negative examples. The Two-Shot score is then defined as a measure of how well this model discriminates the two sets. Again, the particular definition of this score depends on the underlying classifier used. Using the SVM classifier, for example, this can simply be the width of the margin between the two sets. In the following sections we provide detailed analysis of this new similarity score.

3.1   Background-Sample Based Similarities with LDA

The OSS and TSS scores are actually meta-similarities which can be fitted to work with almost any discriminative learning algorithm. In our experiments, we focused on the Fisher Discriminant Analysis (FDA or LDA) [21, 22] as the underlying classifier. Similarities based on LDA can be efficiently computed by exploiting the fact that the set **A** of negative samples is used repeatedly, and that the positive class, which contains just one or two elements, contributes either nothing or a rank-one matrix to the within class covariance matrix.

We focus on binary LDA, which is relevant to this work. Let $p_i \in \mathbb{R}^d, i = 1, 2, ..., m_1$ be a set of positive training examples, and let $n_i \in \mathbb{R}^d, i = 1, 2, ..., m_2$ be a set of negative training examples. Let μ be the average of all points and μ_p (resp. μ_n) be the average of the positive (negative) training set. Two matrices are then considered [23], S_B measuring the covariance of the class centers, and S_W, which is the sum of the covariance matrices of each class. The LDA algorithm computes a projection v which maximizes the quotient:

$$v = \arg\max_v \frac{v^\top S_B v}{v^\top S_W v} \tag{1}$$

In the two class case, v is easily determined as:

$$v = \frac{S_W^+ (\mu_p - \mu_n)}{\|S_W^+ (\mu_p - \mu_n)\|} \tag{2}$$

Note that we use the pseudo-inverse S_W^+ instead of the inverse S_W^{-1} in order to deal with cases where the within-class covariance matrix is not full rank. This is equivalent to requiring in Eq. 1 that v be spanned by the training vectors.

Once v has been computed, the classification of a new sample $x \in \mathbb{R}^d$ is given by the sign of $v^\top x - v_0$, where v_0 is the bias term (see below).

LDA-based One-Shot Similarity. The LDA-based OSS score and its computation was recently analyzed in [11]. By exploiting the fact that the positive set contains a single sample and the negative set is fixed, it was shown that the LDA-based OSS between samples **I** and **J** given the auxiliary set **A** becomes:

$$\frac{(I - \mu_A)^\top S_W^+ (J - \frac{I + \mu_A}{2})}{\|S_W^+ (I - \mu_A)\|} + \frac{(J - \mu_A)^\top S_W^+ (I - \frac{J + \mu_A}{2})}{\|S_W^+ (J - \mu_A)\|} \tag{3}$$

The overall complexity for the OSS per pair is thus $O(d^2)$ once the (pseudo) inverse S_W has been computed. In addition, if similarities are computed for the same point repeatedly, one can factor the positive definite $S_W^+ = HH^\top$ and pre-multiply this point by the factor H.

LDA-based Two-Shot Similarity. In the two-shot case, **I** and **J** serve as the positive class, while the set **A** of background samples is used repeatedly as the negative class. In contrast to the One-Shot case, the within class covariance matrix S_W changes from one similarity computation to another.

In order to be robust to the size of the background set and for simplicity, we balance the positive and the negative classes and define the within-class convenience matrix as $S_W = \frac{1}{2} S_A + \frac{1}{2} S_{IJ}$, where $S_A = \frac{1}{|A|} \sum_{x \in A} (x - \mu_A)(x - \mu_A)^\top$, and $S_{IJ} = \frac{1}{2}((I - \frac{(I+J)}{2})(I - \frac{(I+J)}{2})^\top + (J - \frac{(I+J)}{2})(J - \frac{(I+J)}{2})^\top) = \frac{1}{4}(I - J)(I - J)^\top$

Since S_{IJ} is a rank-one matrix, the inverse of S_W can be computed by updating the inverse of S_A with accordance to the Sherman-Morrison formula as:

$$\frac{1}{2} S_W^{-1} = S_A^{-1} - \frac{S_A^{-1}(I - J)(I - J)^\top S_A^{-1}}{4 + (I - J)^\top S_A^{-1}(I - J)} \tag{4}$$

If S_W is not full rank, a similar formula can be applied to update the pseudoinverse, based on rank-one updates [24] of the Cholesky factor or SVD of S_A. The details are omitted. Note that the matrix S_W^{-1} need not be computed explicitly. Let $\nu = (I + J)/2 - \mu_A$. From equation 2, v can be computed up to scale as:

$$S_A^{-1} \nu - \frac{S_A^{-1}(I - J)(I - J)^\top (S_A^{-1} \nu)}{4 + (I - J)^\top S_A^{-1}(I - J)}$$

The TSS itself measures the separability of the two classes, i.e., the distance between the centers of the two classes in the direction of v. Thus, once the covariance matrix of the background samples is inverted, computing the TSS requires $O(d^2)$ operations. If points I_i are used repeatedly, $S_A^{-1}I_i$ can be pre-computed, and future TSS computations become $O(d)$.

4   Ranking Based Background Similarity

The idea of representing an image by a set of similarities to other images or to prelearned classifiers is well known [25]. Bart and Ullman [26] have proposed to use it for learning a novel class from one example. We have tried using a vector of similarities to the background samples as a face descriptor. Specifically, we generated for image I and for image J vectors of similarities by comparing I or J to each image in A. The resulting vectors produce much worse classification results than the original similarity between I and J.

Instead, we consider a retrieval system in which images I or J are used to retrieve similar images from the set A, and examine the order in which the images are retrieved. In other words, image I (or J) produces an order on the elements of A from the most similar to the least similar.

To compare two such orders, we can employ the non-parametric technique of computing the correlation between the rank vectors. Each image (I or J) is represented by a vector which contains the ranking of each image in the set A from 1 (most similar image) to $|A|$ (least similar image). The correlation between the two rank vectors measures the similarity between the two permutations.

In our experiments, we have found that it is better to focus on the most similar images. We propose the following statistical test. For each of the two samples I and J we compute the rank vectors r_I and r_J as before. Let π_I (π_J) be the order of images in A according to their similarity to I (J). We then compute the similarity s as the sum of the ranking by one image to the first 100 images in the order of the second image: $s(I, J) = -\sum_{k=1}^{100} r_I(\pi_J(k)) + r_J(\pi_I(k))$. (higher values mean more similar examples). We are yet to conduct a full test for the value of the parameter; currently 100 seems to produce good results.

5   Face Alignment

In order to produce an aligned version of the LFW images, we automatically processed them using a commercial face alignment system. The alignment system is based on localization of fiducial points. An affine transformation that brings those feature points to fixed locations is applied to the image. In order to train the feature detectors a set of labeled face images with manually marked points were collected. These images do not intersect the LFW set in images or in identity.

Our experiments, reported in Section 6, show that this alignment method significantly improves the performance of all tested methods. To illustrate the importance of this improved alignment, we tested the performance of the system

Fig. 1. Two images aligned using the Funneling technique of [28] (on the left) and the fiducial-point based alignment system. Misalignments on the left hand pair are visible when comparing the positions of the mouth and the eyes to the markers. These misalignments are all but removed in the right hand pair.
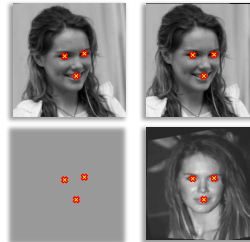
designed by Nowak and Jurie [27] on our own aligned version of the LFW image set. Originally, the LFW images were aligned using the "Funneling" technique of [28]. On this funneled set, Nowak and Jurie obtained a recognition rate of 0.7393. The same method obtains a recognition score of 0.7912 on our aligned set. Note that this performance boost was gained even tough the method of [27] has build-in mechanisms to deal with misalignments.

Figure 1 presents an example of one image pair aligned using both the original Funneling technique of [28] and our own alignment method. The improved alignment of both the eyes and the mouth is evident by comparing their positions to the markers. It is important to note that while the Funneling technique requires no additional training (it is an unsupervised technique), feature-point based alignment techniques, including the method employed by the authors of [29], rely on the existence of a training set of images with marked fiducial points.

6   Experiments

We test the effect of the various contributions on the 10 folds of view 2 of the LFW dataset [1]. Similarly to previous contributions, we employ "image-restricted training". This benchmark consists of $6,000$ pairs, half marked "same" and half not, and is divided into 10 equally sized sets. The benchmark test is repeated 10 times, each time using one set for testing and nine others for training. The goal is to predict which of the test pairs match using only the training data.

We used one of the nine training splits for the background set A and the other eight for classifier training. The background split contains $1,200$ images. The subjects in these images do not appear in the test set, as the LFW benchmark is constructed to have subjects in the different splits mutually exclusive [1].

6.1   The contribution of alignment

Our first set of experiments repeats the experiments of [10] while introducing the automatic alignment. Note that we did not make any attempt to verify the alignment. If the alignment fails for any reason, we still use the resulting image.

The results are described in Table 2. We use the same descriptors of [10] with the addition of a SIFT descriptor: the LBP descriptor [30], two variants called Three-patch and Four-patch LBP (TPLBP and FPLBP) [10], the C1 image descriptor [31], and SIFT [32]. The parameters of all descriptors were copied from [10]. To compute the SIFT descriptor, we subdivide the image into a grid

Table 1. Mean (± standard error) scores on the LFW, Image-Restricted Training benchmark ("view 2") using Euclidean similarities.

Image Descriptor	Original images		Funneled		Alignment	
	Euclidian	SQRT	Euclidian	SQRT	Euclidian	SQRT
LBP	0.6649	0.6616	0.6767	0.6782	0.6824	0.6790
Gabor (C1)	0.6665	0.6654	0.6293	0.6287	0.6849	0.6841
TPLBP	0.6713	0.6678	0.6875	0.6890	0.6926	0.6897
FPLBP	0.6627	0.6572	0.6865	0.6820	0.6818	0.6746
Above combined	0.7107 ± 0.0045		0.7062 ± 0.0046		0.7450 ± 0.0068	
SIFT	0.6617	0.6672	0.6795	0.6870	0.6912	0.6986
All combined	0.7223 ± 0.0092		0.7193 ± 0.0049		0.7521 ± 0.0055	

of 7x7, and compute a 128D SIFT descriptor for each one of the 49 patches. All descriptors are then concatenated to a single vector. Compared to the LBP variants, the SIFT descriptor is less sensitive to misalignment, however, it is easily misled by sharp edges caused by glasses or illumination.

We use either the descriptor vectors or their square roots (i.e., the Hellinger distance). In the latter case, instead of using the descriptor vector $g(I)$ we use $\sqrt{g(I)}$. The 10 descriptor/mode scores in the table are obtained by training SVM on $4,800$ (8 sets) 1D vectors containing the similarity scores. The "Combined" classification is based on learning and classifying the 8D/10D vectors which are the concatenations of the eight/ten 1D vectors (including or excluding SIFT). The results are reported in Table 2. The contributions of adding SIFT and of performing a proper alignment are clearly seen.

6.2   The contribution of one-shot

Next, we examine the performance on the one-shot measure in Table 2. The descriptors used are the same as above. Here again we use either the original descriptor vectors, or their square roots. The "Combined" classification is based on learning and classifying the 8D/10D vectors which are the concatenations of the eight/ten 1D One-Shot similarites. Results are reported without SIFT (to allow comparison to [10]) and with SIFT. The "Hybrid" results contain all direct (Euclidean) similarities above and the One-Shot similarities. Note the gap in performance compared to the funneled no-sift hybrid previously reported.

6.3   The contribution of two-shot

The two-shot similarity adds another layer of information. By itself, it is not very discriminative. For the aligned images, all 10 (5 descriptors and using or not using square root) two-shot similarities provide a combined score of 0.6593 ± 0.0076, which is lower than the corresponding figure of 0.8207 for the One-Shot Similarities and the 0.7521 for the baseline similarities.

However, in combination with the baseline similarities and the One-Shot Similarities, the Two-Shot Similarities boost performance considerably. Adding those similarities to the mix increases the performance in the aligned images from 0.8398 ± 0.0035 to 0.8513 ± 0.0037.

Table 2. Mean (± standard error) scores on the LFW, Image-Restricted Training benchmark ("view 2") using OSS.

Image Descriptor	Original images		Funneled		Alignment	
	OSS	OSS SQRT	OSS	OSS SQRT	OSS	OSS SQRT
LBP	0.7292	0.7390	0.7343	0.7463	0.7663	0.7820
Gabor (C1)	0.7066	0.7097	0.7112	0.7157	0.7396	0.7437
TPLBP	0.7099	0.7164	0.7163	0.7226	0.7453	0.7514
FPLBP	0.7092	0.7112	0.7175	0.7145	0.7466	0.7436
Above OSS Comb.	0.7582± 0.0067		0.7653 ± 0.0054		0.8002 ± 0.0018	
Above Hybrid	0.7752 ± 0.0063		0.7847 ± 0.0051		0.8255 ± 0.0031	
SIFT	0.7126	0.7199	0.7202	0.7257	0.7576	0.7597
All OSS Combined	0.7673 ± 0.0039		0.7779 ± 0.0072		0.8207 ± 0.0041	
All Hybrid	0.7782 ± 0.0036		0.7895 ± 0.0053		0.8398 ± 0.0035	

6.4   The contribution of the ranking descriptor

The ranking based similarities obtained by the proposed score, which considers the ranking by one example of the first 100 images closest to the other example. It is slightly more effective than Two-Shot Similarity above, and the score obtained by combining all 10 rank similarities using SVM is 0.6918 ± 0.0062. As mentioned in Sec. 4, using other forms of representation by similarity are not better.

Similar to the the Two-Shot Similarity above, the contribution of the ranking descriptor by adding it to the other descriptors. A hybrid descriptor which contains 10 original distances, 10 One-Shot distances, 10 Two-Shot distances, and 10 ranking based distances produces a result of 0.8557 ± 0.0048, which is much higher than the current record of 0.7935 ± 0.0055 [33].

6.5   Combining background similarities beyond LDA

The One-Shot and Two-Shot similarities are frameworks that can be applied with LDA as above or with other classifiers. Applying it with SVM instead of LDA gives very similar results. However, a considerable boost in performance is obtained when adding SVM based OSS and TSS to those of LDA. Adding those 20 additional dimensions results in a performance of 0.8297 ± 0.0037 for the funneled images and 0.8683 ± 0.0034 for the aligned images.

The ROC curves of the final combined result, as well as the results of previous work is presented in Figure 2. As can be seen, the present result is considerably better than previous method. This is especially so in the low-false-positive region, which is the crucial region for most applications.

7   Conclusions

We follow the Detection-Alignment-Recognition pipeline devised in [1] for the study of face recognition in unconstrained environments. For alignment, we demonstrate the significance of proper localization by improving upon results obtained on already aligned images. For representation, we augment the set of descriptors by adding SIFT. For similarity we study three frameworks for employing background samples, shifting focus from one-shot to two examples to
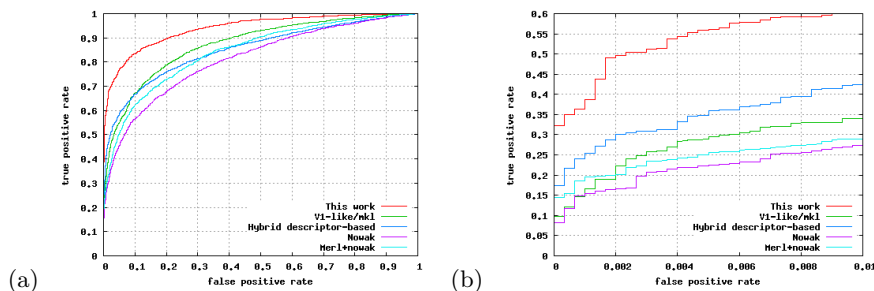
Fig. 2. ROC curves for View 2 of the LFW data set. Each point on the curve represents the average over the 10 folds of (false positive rate, true positive rate) for a fixed threshold. (a) Full ROC curve. (b) A zoom-in onto the low false positive region. The proposed method is compared to scores reported in http://vis-www.cs.umass.edu/lfw/results.html. These methods include the combined nowak+Merl system [29], the Nowak method [27], the hybrid method of [10], and the recent V1-like/mkl method of [33].

many examples. This form of side information has not gained considerable attention previously, and we demonstrate its effectiveness. The obtained leap in performance is impressive given the law of diminishing returns and the amount of work invested by various groups on the LFW benchmark.

Acknowledgments

References

1. Huang, G., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. UMASS, TR 07-49 (2007)
2. Taigman, Y., Wolf, L., Hassner, T.: Multiple one-shots for utilizing class label information. In: BMVC. (2009)
3. Belongie, S., Malik, J., Puzicha, J.: Shape context: A new descriptor for shape matching and object recognition. In: NIPS. (2001)
4. Zhang, H., Berg, A., Maire, M., Malik, J.: Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In: CVPR. (2006)
5. Bilenko, M., Basu, S., Mooney, R.: Integrating constraints and metric learning in semi-supervised clustering. In: ICML. (2004)
6. Cristianini, N., Kandola, J., Elisseeff, A., Shawe-Taylor, J.: On kernel-target alignment. In: NIPS. (2002)
7. Shental, N., Hertz, T., Weinshall, D., Pavel, M.: Adjustment learning and relevant component analysis. In: ECCV. (2002)
8. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. NIPS (2006)
9. Xing, E., Ng, A.Y., Jordan, M., Russell, S.: Distance metric learning, with application to clustering with side-information. In: NIPS. (2003)
10. Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: Faces in Real-Life Images Workshop in ECCV. (2008)

11. Wolf, L., Hassner, T., Taigman, Y.: The one-shot similarity kernel. In: ICCV. (2009)
12. Fei-Fei, L., Fergus, R., Perona, P.: One-shot learning of object categories. PAMI **28**(4) (2006) 594–611
13. Fink, M.: Object classification from a single example utilizing class relevance pseudo-metrics. In: NIPS. (2004)
14. Chapelle, O., Scholkopf, B., Zien, A.: Semi-Supervised Learning. MIT Press, Cambridge, MA (2006)
15. Joachims, T.: Transductive learning via spectral graph partitioning. In: International Conference on Machine Learning (ICML). (2003) 290–297
16. Quattoni, A., Collins, M., Darrell, T.: Transfer learning for image classification with sparse prototype representations. CVPR (June 2008)
17. Xing, E., Ng, A., Jordan, M., Russell, S.: Distance metric learning with application to clustering with side-information. In: NIPS, Cambridge, MA (2003)
18. Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D.: Learning distance functions using equivalence relations. In: ICML. (2003)
19. Liu, W., Hoi, S., Liu, J.: Output Regularized Metric Learning with Side Information. In: ECCV. (2008) 358–371
20. Chechik, G., Tishby, N.: Extracting relevant structures with side information. In: NIPS. (2002) 857–864
21. Fisher, R.: The use of multiple measurements in taxonomic problems. Annals Eugenics **7** (1936) 179–188
22. Hastie, T., , Tibshirani, R., Friedman, J.: The elements of statistical learning. Springer (2001)
23. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification, 2nd ed. Wiley (2001)
24. Brand, M.: Fast low-rank modifications of the thin singular value decomposition. Linear Algebra and its Applications **415**(1) (2006) 20 – 30
25. Edelman, S.: Representation and recognition in vision. MIT Press, Cambridge, MA, USA (1999)
26. Bart, E., Ullman, S.: Single-example learning of novel classes using representation by similarity. In: British Machine Vision Conference. (2005)
27. Nowak, E., Jurie, F.: Learning visual similarity measures for comparing never seen objects. In: CVPR. (June 2007)
28. Huang, G., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: IEEE International Conference on Computer Vision. (2007)
29. Huang, G., Jones, M., Learned-Miller, E.: Lfw results using a combined nowak plus merl recognizer. In: Faces in Real-Life Images Workshop in ECCV. (2008)
30. Ojala, T., Pietikainen, M., Harwood, D.: A comparative-study of texture measures with classification based on feature distributions. Pattern Recognition **29**(1) (1996)
31. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nature Neuroscience **2**(11) (1999) 1019–1025
32. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2) (November 2004) 91–110
33. Pinto, N., DiCarlo, J., Cox, D.: How far can you get with a modern face recognition test set using only simple features? In: CVPR. (2009)