Face Recognition with semi-supervised learning and Multiple Classifiers

NEAMAT EL GAYAR*, SHABAN A. SHABAN† SAYED HAMDY†
†Institute of Statistical Studies and Research
*Faculty of Computers and Information
Cairo University
5 Ahmed Zewel St., 12613 Orman, Giza
EGYPT
*

Abstract: - Face recognition using labeled and unlabelled data has received considerable amount of interest in the past years. In the same time, multiple classifier systems (MCS) have been widely successful in various pattern recognition applications such as face recognition. MCS have been very recently investigated in the context of semi-supervised learning. Very few attention has been devoted to verifying the usefulness of the newly developed semi-supervised MCS models for face recognition. In this work we attempt to access and compare the performance of several semi-supervised MCS training algorithms when applied to the face recognition problem. Experiments on a data set of face images are presented. Our experiments use non-homogenous classifier ensemble, majority voting rule and compare between a three semi-supervised learning models: the *self-trained single classifier* model, the *ensemble driven* model and a newly proposed *modified co-training* model. Experimental results reveal that the investigated semi-supervised models are successful in the exploitation of unlabelled data to enhance the classifier performance and their combined output. The proposed semi-supervised learning model has shown a significant improvement of the classification accuracy compared to existing models.

Key-Words: - Semi-Supervised learning, Multiple Classifier System, Classifier ensembles, Face recognition, Majority vote, Learning using labeled and unlabelled data.

1 Introduction

Over the past years, there has been an emerging interest in face recognition as a result of the availability of advanced technologies and a growing requirement for automatic human identification in numerous applications. Examples of such applications are: security, access control to buildings, identification of criminals and human computer interfaces. Face recognition is a difficult task because of the inherent variability of the image formation process in terms of image quality, photometry, geometry, occlusion, change and disguise. A recent survey on face recognition discusses these challenges in some detail [1].

Among the challenges that can also be present in face recognition problems is the availability of only a small amount of labeled data (which is often costly to obtain) along with a large pool of unlabeled data. Only a few work has addressed using unlabeled data in face recognition to improve classifier accuracy when only a small set of labeled examples is available [2-4].

Multiple classifier Systems, also called classifier ensembles, have been shown very useful for improving performance in numerous pattern recognition problems ranging from handwriting character recognition to speech recognition[3]. They also have been used efficiently for face recognition purposes [6-12].

Recently there has been a strong direction to develop semi-supervised learning algorithms for MCS that are able to exploit both labeled and unlabelled data [13]. The demand on such models is increasing in important pattern recognition problems such as multi-sensor remote sensing and multimodal biometrics. Recent research has attempted to extend the single-classifier versions of semi-supervised learning techniques [14] to semi-supervised MCS models [15-18]. In [15] a method is proposed to use the combined output of classifier ensembles working on different views of the input sample to label unlabelled patterns and use them to retrain the original classifiers. In [16] Zhou presents *democratic co-learning* in which multiple algorithms instead of multiple views enable learners to label data for each other. The work in [17] suggests two extensions of the co-training algorithms over MCS that use different views of the patterns. Finally in [18], Roli investigates the use of co-training for a generic MCS whose classifiers can be created with different classification techniques.

Almost no work has been done on investigating recently developed models of semi-supervised multiple classifier systems (MCS) for face recognition.

It is the purpose of this study to experimentally compare several semi-supervised MCS learning techniques when applied to face recognition in a heterogeneous ensemble of classifiers, using the same input features. In particular we compare between the *single classifier self-training*, the *ensemble driven training* and suggest a new technique for semi-supervised learning in MCS that combines between co-training[13-14] and self-supervised learning[12,15].

Experiments were conducted on face images from a benchmark data set. The eigenfaces technique was used to reduce the dimensionality of the image space. The eigenfaces technique [19] can help us to deal with multidimensionality because it reduces the dimension of the image space to a small set of characteristics called eigenfaces, making the calculations manageable and with minimal information loss.

The experiments were performed using *k-nearest-neighbor*, *Fuzzy k-nearest-neighbor* and *PCA-based recognition* algorithms as base classifiers. We experiment with different sizes of the initially labeled training set and use the majority rule for classifier combination.

The paper is organized as follows: Section 2 briefly surveys the used semi-supervised models; Section 3 presents the data under investigation and describes the experiments conducted; Section 4 summarizes and discusses experimental results; finally, some conclusions and directions for future work are presented in Section 5.

2 Semi-Supervised MCS for Face recognition

In this section we describe three different techniques for semi-supervised MCS that can be applied to the face recognition problem, namely *single classifier self-training*, *ensemble driven training* and *modified co-training*. We assume K classifiers representing K learning algorithms; hence given a set L of labeled data and a set U of unlabelled data (which is usually much larger than L), it is required to use both data sets to design the K classifiers.

The main idea of using unlabeled data is to improve classifier accuracy when only a small set of labeled examples is available. The basic steps to achieve this are summarized as follows: First, each learning algorithm is trained with the set of labeled examples (L). Then each classifier is used to assign pseudo labels to a randomly chosen subset of the unlabeled examples, U which we refer to as U'. Some of the newly labeled examples of U' are chosen –according to a certain criterion- to be appended to the training set L. The new training set is used to retrain the classifiers so as to increase their accuracy. This process is iterated usually until most data is labeled. As follows we outline 3 different methods to complete this procedure.

2.1 Single Classifier Self-training

Fig. 1 describes basic steps for the *Single classifier self-training* algorithm for MCS. In the self-training procedure, as outlined above, the K classifiers are initially trained using the labeled data set L. Each classifier is then used to assign pseudo class labels to a subset of unlabelled examples in U. The pseudo labeled data; that is labeled with high confidence by each classifier is then added to increase its own training set. Each classifier is then retrained with the new augmented data set. This process is repeated for a given number of times. It should be noted that here classifiers work and learn independently; only they start with the same initial conditions (same labeled data set). Classifiers can be fused after training is complete.

- ***Given:***
  - L, a set of crisp labeled training examples x of M different classes.
  - K different classification models: $(CL_1, CL_2, \ldots CL_{i\ldots}, CL_K)$
  - U, a set of unlabeled example
- ***Create a pool U' of examples by choosing u examples at random from U***
- ***For each classifier*** CL_i ***let*** $L_i = L$
- ***Loop for N iteration:***

1. For each classifier $\boldsymbol{CL_i}$
   a. (Re)train $\boldsymbol{CL_i}$ with L_i
   b. Use $\boldsymbol{CL_i}$ to label all examples from U'
   c. Select n most confident examples per class $(n_{i1}, n_{i2}, \ldots n_{ij\ldots}, n_{iM})$
   d. Add these self-labeled examples to the training set L_i, for classifier $\boldsymbol{CL_i}$
   e. Randomly choose examples from U to refill U' equal to the number of examples added to L_i

- **Combine classifiers outputs with fusing rule**.

Figure 1 Single Classifier Self-training

2.2 Ensemble Driven training

Fig. 2, summarizes the *ensemble driven* semi-supervised learning algorithm. Again, *K* classifier are trained on the initial, small, labeled data set *L*. Each classifier is then applied to the u unlabelled samples in *U'*. Classifiers are combined using majority vote, if the majority of the classifiers (i.e. more than half) agree on a certain class then the pattern u is labeled according to this class and added to the set L; otherwise it is discarded. All classifiers are then retrained with the new set *L*. This method is similar to the self-supervised learning model described in [15], but uses different classification algorithms (i.e non-homogenous ensembles) instead of different views. In the majority vote, if no majority candidate is present, then the sample is rejected; also not added to L.

- ***Given:***
- *L*, a set of crisp labeled training examples x of *M* different classes.
- K different classification models: $(CL_1,\ CL_2, \ldots CL_{i\ldots},\ CL_K)$
- *U*, a set of unlabeled example
- ***Create a pool U' of examples by choosing u examples at random from U***
- ***Loop for N iteration:***

1. Use L to train K classifier
2. Allow each $\boldsymbol{CL_i}$ to label all examples from *U'*
3. For each unlabelled sample in U'; Combine classifiers outputs using majority vote.
4. Add only to L the labeled patterns from U' on which the K classifiers have agreed with majority.
5. Randomly choose examples from U to replenish U' equal to the number of examples added to L.

- ***Combine classifiers outputs with fusing rule.***

Figure 2 Ensemble Driven Training

2.3 Modified Co-training

Fig. 3 introduces a new semi-supervised learning algorithm, the *modified co-training* algorithm. The difference to the ensemble driven training explained above is that each classifier labels *U'* and sorts patterns in each class according to their confidence. For an unlabelled pattern to be added to *L* it has to be ranked among the most confident patterns for the same class for the majority of the classifiers. This is different than the co-training algorithms introduced in [17] and [18] in that they put more strict conditions on the choice of the unlabelled patterns to be added to *L*. The *modified co-training* algorithm can be considered a combination between co-training [16,17] and self-supervised learning ensembles[15,18].

- ***Given:***
- *L*, a set of crisp labeled training examples x of *M* different classes.
- K different classification models: $(CL_1,\ CL_2, \ldots CL_{i\ldots},\ CL_K)$
- *U*, a set of unlabeled example
- ***Create a pool U' of examples by choosing u examples at random from U***
- ***Loop for N iteration:***

1. Train each classifier with L
2. Allow each $\boldsymbol{CL_i}$ to label all examples from *U'*
3. Select n most confident examples per class for each $\boldsymbol{CL_i}$ $(n_{i1},\ n_{i2}, \ldots n_{ij\ldots},\ n_{iM})$
4. Add these self-labeled examples to L only if the K classifier agree upon them in majority
5. Randomly choose examples from U to refill U' equal to the number of examples added to L.

- ***Combine classifiers outputs with fusing rule.***

Figure 3 Modified Co-training

3 Data and Experiments

Experiments were conducted on the *UMIST Face database* from the University of Manchester Institute of Science and Technology [19]. >From this database we used images of 20 people with 25 to 55 images per class. Figure 6 shows one example of one class that exists in the *UMIST Face database*. Due to that the images had different sizes we processed them to obtain images of 110 × 110 pixels in BMP and PGM format. The data set for each class was partitioned into a labeled data set *L*, an unlabelled data set *U* and a test set *T*.

We used principal component analysis (PCA) for dimensionality reduction as described in [2]. The PCA finds the vectors which best account for the distribution of face images within the entire image space.. With this technique the calculations are greatly reduced from the order of the number of pixels in the images *N2* to the order of the number of images in the training set *N*, and the calculations become quite manageable.

Figure 4

In our experiments we used 3 non-homogenous classifiers in the ensemble: a *k-nearest neighbor classifier* (KNN), a *fuzzy k-nearest neighbor classifier* (FKNN) and simple PCA-based *classifier* (PCA).

KNN classifiers belong to the family of instance-based learning algorithms and are popular for their simplicity to use and implementation, robustness to noisy data and their wide applicability in a lot of appealing applications [20]. FKNN classifiers on the other hand, have the advantage over the traditional (crisp) K-Nearest Neighbor algorithms that they can take into account the ambiguous nature of the neighbors of a pattern to be classified and can assign a membership that represents the *strength* or *confidence* with which the current pattern belongs to a particular class. We implement a FKNN model as described in [21]. As for the PCA classifier we compare the input image (after performing a PCA operation on it) with the mean image of each class and the input face is consider to belong to the class with minimum Euclidean distance [12].

Experiments were conducted to determine whether the semi-supervised techniques described in section 2 can successfully use unlabeled data to improve classifier performance on the face images, opposed to the simple supervised training model. We used variable initial number of labeled examples/class (3,5 and 7) to study the effect of the size of the initially labeled data on the performance of the compared techniques. We divided the rest of available data into 2/3 for the unlabelled data set *U* and 1/3 for the test data set *T*. We used 3 non-homogeneous classifiers: KNN, FKNN and the PCA described above that were initially trained on the same labeled data set (*L*). For each number of labeled samples we repeated our experiments 10 times and the results presented are the average over 10 random choices of the initial labeled data set, *L*, the initial unlabelled data set, *U* and the test data set, *T*. Detailed results are presented and discussed in the following section.

4 Results

Table 1, summarizes the results of the experiments conducted. It compares the accuracy of each single classifier after learning with the labeled data set *L* through a purely *supervised* manner (*Sup*) to the single classifier performance when the classifiers are retrained with semi-supervised models: the *single classifier self training* (SS-single), the *ensemble driven training* (*SS-ED*) and the *modified co-training* (*SS-COT*) as described previously in Fig. 1., Fig 2. and Fig. 3, respectively. As can be seen, Table 3 is divided into 3 sub-tables; presenting the results when the size of the initial labeled data set *L* consists of 3, 5 and 7 samples/class . The last row in each sub-table lists the results of fusing the 3 classifiers in the ensemble; i.e the KNN, the FKNN and the PCA. As mentioned before we use the majority vote rule, that takes the majority of the classifiers decisions into account. If no clear majority agreement exists on a certain class (in our case if no two classifier agree on the same class), then a certain pattern is *refused to be decided on* (i.e rejected) . In the last row in each sub-table the rejection rate is listed under the dotted line.

3 lab.	*Sup.*	*SS (Single)*	*SS-ED*	*SS-COT*
PCA	42%	43%	62%	65%
KNN	40%	46%	51%	67%
FKNN	41%	47%	53%	69%
Voting	47.6%	58%	61.23%	85%
	1%	0.01%	0.07%	5%
5 lab.	*Sup.*	*SS (Single)*	*SS-ED*	*SS-COT*
PCA	63%	69%	78%	79%
KNN	*60%*	*66%*	*70%*	*76%*
FKNN	62%	67%	77%	77%
Voting	74%	83%	84.94%	92%
	1%	0.1%	0.06%	2%
7 lab.	*Sup.*	*SS (Single)*	*SS-ED*	*SS-COT*
PCA	*77%*	*86%*	*82%*	*88%*
KNN	*66%*	*79%*	*81%*	*81%*
FKNN	*72%*	*81%*	*80%*	*85%*
Voting	*79%*	*90%*	*92.36%*	*97%*
	0%	*0%*	*0.04%*	*1%*

Table 1 Summary of Results

Note the results presented are using K=3 for the KNN, FKNN. Experimentation with other values of K did not yield to further improvements.

As for the PCA we used 60 eigenvectors, containing about 90% of the information. For the PCA we use the augmented training set resulting from the semi-supervised cycle to update the class template similar to method described in [4]. As mentioned before, in our implementation template used to represent each class is simply the mean of the faces. More sophisticated methods could be used [22].

Choosing the most confident pattern depends on the base classifier used in the ensemble. For the KNN and the FKNN the most confident pattern is considered to be the pattern having maximum probability for all elements belonging to that specific class. The confidence for the PCA based classification is simply a measure of the closeness to the class template.

The most confident patterns of U' were sorted in each class and only the top five of the most confident patterns in each class were considered to be added to the labeled data set.

From the results summarized in Table 1 it can be easily seen that all semi-supervised models cause an improvement over the supervised models for all cases investigated; indicating their ability to effectively use information provided by the unlabelled data set to increase accuracy. The improvement can be seen on the level of the single classifiers (KNN, FKNN and PCA) and on the level of the ensemble, i.e combined output. Comparing the three investigated semi-supervised learning algorithm it can be seen that proposed *modified co-training* algorithm significantly improves the performance and outperforms the *single classifier self-training* model and the *ensemble driven model*. Although the rejection rate of the *modified co-training* is highest it still can be regarded as to reduce misclassified patterns compared to other methods.

5 Conclusions and Future Work

We have presented a new ensemble method for face recognition using labeled and unlabeled data. The study compares the proposed model to two recently existing semi-supervised MCS models. Experiments on a data set of face images indicate that all investigated semi-supervised models can improve both single classifier accuracy and the accuracy of the ensemble that is composed from the classifier; if only trained with a small initial data set. We use non-homogenous ensembles of a k-nearest classifier, a fuzzy k-nearest classifier and a PCA-based classifier. The combination of the classifiers is performed by a majority voting rule to alleviate the difficulty of combining classifiers that give different types of output.

Future work includes extending the experiments to other image databases and using other learning algorithms and fusing rules. We also intend to investigate other applications such as multi-sensor remote sensing and multimodal biometrics.

Acknowledgement

The first author of this paper was partially supported by the DFG (German Research Society) grants SCHW 623/3-2 and SCHW 623/4-2.

References:

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, Face Recognition: A Literature Survey, *ACM Computing Surveys*, 2003, pp. 399-458.

[2] C. Martinez and O. Fuentes, Face recognition using unlabeled data, *Journal of Computer Science Research*, Vol. 7. no. 2, 2003, pp. 123-129.

[3] M.F. Balcan et al., Person Identification in Webcam Images: An Application to Semi-Supervised Learning, *ICML2005 Workshop on Learning with Partially Classified Training Data, Bonn, Germany, 7 August 2005*

[4] F. Roli and G. Luca Marcialis:, Semi-supervised PCA-Based Face Recognition Using Self-training. SSPR/SPR 2006, *Lecture Notes in Computer Science*, Springer Verlag, Vol. 4109, 2006, pp. 560-568

[5] Multiple Classifier Systems, Springer Verlag, *Lecture Notes in Computer Science*, Vols. 1857 (2000), 2096 (2001), 2364 (2002), 2709 (2003), 3077 (2004), 3541 (2005).

[6] S. Gutta , J. Huang , B. Takacs and H. Wechsler, Face Recognition Using Ensembles of Networks, *International Conference on Pattern Recognition (ICPR)*, 1999

[7] S. Ben-Yacoub, Y. Abdeljaoued and E. Mayoraz, Fusion of face and speech data for person identity verification, *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, September 1999, pp. 1065–1074,.

[8] G.-Dong Guo and Hong-Jiang Zhang, Boosting for fast face recognition, In *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems* , 2001, pp. 96–100.

[9] J. Czyz, J. Kittler and L. Vandendorpe, Combining face verification experts, In *16th International Conference on Pattern Recognition (ICPR),* Vol. 2, , August 2002, pages 28–31.

[10] Jie Zhou and David Zhang, Face recognition by combining several algorithms, In *16th International Conference on Pattern Recognition (ICPR),* Vol. 3, August 2002, pages 497–500.

[11] A. Lemieux and M. Parizeau, Flexible multi-classifier architecture for face recognition systems, *Vision Interface*, 2003.

[12] X. Lu, Y. Wangy and A. K. Jain, Combining classifiers for face recognition, *Proceedings of the International Conference on Multimedia and Expo,* ICME '03, 2003 *n*

[13] F. Roli, Semi-Supervised Multiple Classifier Systems: Background and Research Directions, *Multiple Classifier Systems*, Springer Verlag , Vol. 3541, 2005.

[14] X. Zhu, Semi-supervised learning literature survey, Technical report, Computer Sciences TR 1530, Univ. Wisconsis, Madison, USA, Jan. 2006.

[15] N. El Gayar, An Experimental Study of a Self-Supervised Classifier Ensemble, *International Journal of Information Technology*, Vol. 1, No. 1, 2004.

[16] Y. Zhou and S. Goldman, Democratic Co-learning, *16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004*), 15-17 November 2004, Boca Raton, FL, USA, pp. 594-602.

[17] M. Solyman, N. El Gayar, A Co-training Approach for Semi-Supervised Multiple Classifiers, *ICGST-AIML Journal, Vol. 6, Special Issue on Multiple Classifier Systems*, May 2006, pp 9-16.

[18] L. Didaci and F. Roli, Using Co-training and Self-training in Semi-supervised Multiple Classifier Systems, *Lecture Notes in Computer Science*, Springer Verlag, Vol. 4109, 2006, pp. 522-530

[19] D. Graham and N. Allinson, Character-rising Virtual Eigensignatures for Face Recognition. In *Face Recognition: From Theory to Applications*. Springer-Verlag, 1998.

[20] T. M. Mitchell. *Machine Learning.* McGraw-Hill, 1997.

[21]. J. M. Keller, M. R. Gray and J. A. Givens, A fuzzy K-Nearest Neighbor Algorithm. *IEEE Trans. on SMC,* Vol. SMC-15, No.4, 1985.

[22] U. Uludag, A. Ross and A.K. Jain, Biometric template selection and update: a case study in fingerprints, Pattern Recognition, Vol 3, No. 7, July 2004, pp.1533-1542.