CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 96 (2016) 1571 – 1580
Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, 5-7 September 2016, York, United Kingdom

Face recognition in the wild

Wajdi Dhifli[a], Abdoulaye Baniré Diallo[a,*]

[a]*Department of Computer Science, University of Quebec At Montreal, PO box 8888, Downtown station, Montreal (Quebec), H3C 3P8, Canada*

Abstract

Face recognition is one of the most important tasks in pattern recognition and computer vision. The most conventional way to perform face recognition is to compare a set of facial features that are extracted from a source image or a video frame with a reference image database of known faces. Such a classification takes the form of a prediction within a closed-set of classes. However, a more realistic scenario that fits the ground truth of real-world face recognition applications is to consider the possibility of encountering faces that do not belong to any of the training classes, *i.e.*, an open-set classification. Such a constraint is very challenging to most existing face recognition systems since the latter are based on closed-set classification methods which always assign a training label to novel unknown instances even if they represent unseen faces that are not represented in the reference database. This results in a misclassification. In this paper, we introduce **F**ace **R**ecognition in the **W**ild (FRW), a novel face recognition system that allows (1) to efficiently recognize known faces from the reference database, and (2) to prevent misclassifying instances that represent unknown and unseen faces. FRW formulates this problem as a multi-class classification in an open-set context where the presence of instances from unknown classes is possible. Experimental results on the challenging Olivetti Faces benchmark dataset show the efficiency of our approach in open-set face recognition problems.
© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of KES International
Keywords: Face recognition, multi-class classification, open-set classification;

1. Introduction

Face recognition is one of the most important tasks in pattern recognition and computer vision as well as in many other related fields including machine learning and artificial intelligence. Most existing works on face recognition focus on the identification of the most relevant facial features for efficiently identifying and discriminating between the considered individuals. Such features are often mined through an extraction of explicit features (such as the size and shape of the eyes, those of the nose and the jaw [1], *etc*.), or latent features (such as the Eigenfaces [2,3], Fisherfaces [3,4], *etc*.). These features are used to construct a description feature vector for each face image. Then, the resulting description matrix is used to train a classifier which creates a classification model allowing to classify novel face images into one of the reference classes. Both parts of the framework (*i.e.*, the feature extraction step and the classification step) are fundamental for face recognition systems. Yet, in this work, we focus our attention on the second step of

* Corresponding author. Tel.: +1-234-567-8910 ; fax: +1-234-567-8910.
E-mail address: diallo.abdoulaye@uqam.ca

1877-0509 © 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.204

the framework *i.e*., the classification step. The most conventional classification scenario is to train a classifier on a set of instances of known classes, *i.e*. the training set, then to predict the class label of unknown instances within the same set of already seen classes[5,6,7,8]. Such a classification takes the form of a prediction within a *closed-set* of labels. However, in real-world face recognition applications, due to the growth of data collection, the training data could only represent a partial view of the domain and thus it may not contain training examples for all possible classes.

In such scenario, the classifier may be confronted, during prediction, to observations that do not belong to any of the training classes. This makes the target classes become an *open-set* of labels. In *open-set classification*, traditional closed-set classifiers will fail in the prediction for observations of unseen labels.

In applications where the user is interested in the identification of few classes from a large classification universe, the most conventional way is to fuse the set of uninteresting classes into one single large negative set which usually makes the dataset highly unbalanced. In this case, the classifier becomes overwhelmed by negative observations which hinders the discrimination of positive classes. Some attempts have emerged trying to remedy such situation, mainly based on the sampling of a subset of representatives from the negatives[9]. However, it is very difficult and somehow unfair to reduce all the negatives into a small summary that may not be sufficient to represent the whole set. A more appropriate transformation of such problem is an open-set classification where only positive classes are modeled in training and any observation that remarkably deviates from the distribution of known classes is rejected.

In face recognition applications, the system is only interested in recognizing a number of faces within an infinite set of possibilities, *i.e*. an open universe of classes.

In this paper, we introduce FRW (for **F**ace **R**ecognition in the **W**ild), a novel open-set multi-class classifier for face recognition in open-set contexts. For each class, FRW creates a minimum bounding hyper-sphere that encloses all of its instances. In such manner, it is able to distinguish between novel face instances that fit the distribution of a known class from those that diverge from it. FRW introduces a softening parameter for the adjustment of the minimum bounding hyper-spheres to add more generalization or specialization to the classification models. To appropriately evaluate open-set classification, we also propose a novel evaluation technique, namely *Leave-P-Class-Out-Cross-Validation*. We experimentally evaluate FRW on the challenging Olivetti Faces benchmark dataset. The obtained results show the efficiency of our approach in open-set face recognition problems.

2. Related Work

Very few works have addressed open-set classification in the literature. Scheirer et al.[10] presented a formalization of open-set classification and showed its importance in real-world applications including face recognition. The authors discussed the bias related to the evaluation of learning approaches on popular datasets. They showed how recognition accuracies are inflated in closed-set scenarios, leading to an over-estimated confidence in the evaluated approaches[10,11]. In binary closed-set classification, SVM defines a hyper-plane that best separates between two classes. Scheirer et al. proposed an SVM based open-set multi-class classifier termed one-vs.-set SVM[10], which defines an additional hyper-plane for each class such that the latter becomes delimited by two hyper-planes in feature space. A testing instance is then classified as of one training class or as of an unknown class, depending on its projection in feature space. Although this strategy delimits each training class from two sides, the class "acceptance-space" is left unlimited within the region between the hyper-planes and no additional separators are provided to prevent misclassifying unknown instances that are within the same class hyper-planes bound but far away from the distribution of its training instances in feature space.

Semi-supervised classification[12] have addressed open-set classification to some extent, where part of the dataset is unlabeled and the goal is to label as many as possible of the unlabeled data then to use them in training in order to enhance the classification performance. Unlabeled instances are labeled based on their distances from the distribution of labeled data where far instances can be rejected by all classes. Although this somehow resembles an open-set classification, in this context the acceptance/rejection is performed in training and the goal is to optimize the classification performance by minimizing the loss on the training set. However, in open-set classification, the acceptance/rejection is performed in prediction on testing data to classify them as of one training-class or as of an unknown one.

Another important approach for open-set problems is one-class classification. The most known technique is one-class SVM[7] where the latter is trained on one positive class and the aim is to define a contour that encloses it from the rest of the classification universe. Any instance that is projected outside of the defined class boundary is considered as negative. One-class classification is mainly used in outlier and novelty detection. It can be used for open-set face recognition. However, it is limited to single class classification and cannot be directly used in multi-class classification.

One-vs.-one and one-vs.-rest[8] are popular techniques for multi-class classification. One-vs.-one constructs a model for each pair of classes and test examples are evaluated against all the constructed models. A voting scheme is applied and the predicted label is the one with the highest number of votes. One-vs.-rest creates a single classifier per class, with the examples of that class as positives and all the other examples as negatives. In prediction, all classifiers are applied on the test example and the predicted label is the one with the highest confidence score. It is possible to consider one-vs.-rest for open-set classification by iteratively using each class as the positive training set, and all the remaining (known) classes as the rest of the classification universe. However, in open-set classification, the rest of the classification universe is (theoretically) unlimited and thus the one-vs.-rest classifier will suffer a negative set bias.

Based on Landgrebe et *al.*[13] and Tax et *al.*[14], it is possible to build a simple open-set multi-class classifier using a combination of a one-class classifier and a multi-class classifier. In the first step, all training classes are fused into a single large super-class and the one-class classifier is trained on the entire super-class. In this setting, the one-class classifier will directly reject and label as "unknown" all testing instances that do not fit the distribution of all known training classes. In the second step, the multi-class classifier is trained on the original training classes and it is used to classify instances that were not rejected by the one-class classifier.

3. Open-set Face Recognition

3.1. Preliminaries and Problem Definition

Let $\mathcal{D}$ be a training set of n instances and $\mathcal{L}$ be the set of possible labels in $\mathcal{D}$, $\mathcal{D} = \{(x_1, l_1), ..., (x_n, l_n)\}$ where $l_i \in \mathcal{L}$ and x_i is defined by a vector in d-dimensional space, $\forall i \in [1, n]$. In open-set classification, the classifier should be able to assign to a test instance x a label l_x that is known $l_x \in \mathcal{L}$ or that is *unknown*, *i.e.*, $l_x \in \mathcal{L} \cup \{\text{"}unknown\text{"}\}$. In this setting, it is necessary to define a boundary envelop for each class in order to make it distinguishable from other unknown possibilities. The definition of such boundary is hard and delicate as the delimited class-space should reflect the class distribution by enclosing as many as possible of its instances while keeping outside as many as possible of the rest of instances. Indeed, this can be seen as an optimization problem of the classification error that considers a trade-off between generalization and specialization. As a possible solution, we define the minimum bounding hyper-sphere $\mathcal{M}$ as the smallest hyper-sphere that circumscribes all instances of a considered class. For a class $\mathcal{D}_l \subseteq \mathcal{D}$ of label $l \in \mathcal{L}$, the hyper-sphere $\mathcal{M}_l$ represents the class model that resembles the distribution of $\mathcal{D}_l$ instances. Each class model $\mathcal{M}_l$ ($\forall l \in \mathcal{L}$) is defined as:

$$\mathcal{M}_l = (c_l, r_l), \forall l \in \mathcal{L}, r_l > 0 \tag{1}$$

where c_l is the center of $\mathcal{M}_l$ hyper-sphere (the class mean $\overline{x}$):

$$c_l = \overline{x}, \forall x_i \in \mathcal{D}_l, \forall l \in \mathcal{L} \tag{2}$$

and r_l is the radius of $\mathcal{M}_l$ hyper-sphere, *i.e.*, the distance between c_l and the most divergent instance from $\mathcal{D}_l$:

$$r_l = \max\left(\Delta(x_i, c_l)\right), \forall x_i \in \mathcal{D}_l, \forall l \in \mathcal{L} \tag{3}$$

where Δ is a function returning the distance between c_l and x_i with respect to a distance measure.

3.2. The Training Process

Algorithm 1 describes the training phase in FRW. Given a training set $\mathcal{D}$ and a training label-set $\mathcal{L}$ over $\mathcal{D}$, we create a model $\mathcal{M}_l$ for each class $l \in \mathcal{L}$ that is composed of the class minimum bounding hyper-sphere center c_l and radius r_l as defined in the Equations 1, 2 and 3.

3.3. Acceptance of Instances

In open-set classification, the classifier should be able to discriminate between instances of the different known classes and to reject those of unknown classes. This would allow our approach to reject images of faces that are not

Algorithm 1: FRW: The training process

Data: $\mathcal{D}$: training set, $\mathcal{L}$: training labels
Result: $\mathcal{M}$: set of class models
begin
  $\mathcal{M} \leftarrow \emptyset$
  foreach *($l \in \mathcal{L}$)* **do**
    $c_l \leftarrow$ `Centroid` (D_l)
    $r_l \leftarrow$ `Boundary` (D_l)
    $\mathcal{M}_l \leftarrow (c_l, r_l)$
    $\mathcal{M} \leftarrow \mathcal{M} \cup \mathcal{M}_l$

represented in the training set as well those of objects that do not represent faces at all. Therefore, we define a score of acceptance of an instance by a class depending on the distance of its projection from the class boundary.

Definition 1. *(Acceptance Score) The acceptance score, denoted by ϕ, for an instance x by a class of label $l \in \mathcal{L}$, is defined as follows:*

$$\phi(x, l) = 1 - \frac{\Delta(x, c_l)}{r_l} \tag{4}$$

where Δ is an appropriate distance measure, c_l is the center of the class of label l, and r_l is its radius.

The acceptance score is defined in $]-\infty, 1]$ (*i.e.* $\phi \in \mathbb{R}_{\leq 1}$). It allows to decide whether an instance is accepted or rejected by a class. The score is interpreted as follows:

- $\phi(x, l) \in [0, 1]$: the query instance x is accepted by the class l:
  - $\phi(x, l) \in]0, 1[$: x is inside the minimum bounding hyper-sphere of l,
  - $\phi(x, l) = 1$: x is in the class center, *i.e.*, $x = c_l$,
  - $\phi(x, l) = 0$: x is on the class boundary, *i.e.*, $\Delta(x, c_l) = r_l$.
- $\phi(x, l) < 0$: x is out of the class boundary (rejected).

FRW tries to minimize the classification error (Err) that can be formulated as:

$$Err = \sum_{\forall l \in \mathcal{L}} \sum_{\forall x_i \in \mathcal{D}} \psi(x, l) \tag{5}$$

where $\psi \in [0, 1]$ is a binary function that is defined as follows:

$$\psi(x, l) = \begin{cases} 1, & \text{if } \phi(x, l) \in [0, 1] \text{ and } x \in \mathcal{D}_l, \text{ or} \\ & \text{if } \phi(x, l) < 0 \text{ and } x \notin \mathcal{D}_l \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

3.4. The Classification Process

3.4.1. Filtering Prediction Candidate labels

Based on the acceptance score, it is possible, for a given query instance x, to filter a subset of candidate labels $\mathcal{L}_x \subseteq \mathcal{L}$. The latter is the subset of remaining possible candidates, such that if $\mathcal{L}_x \neq \emptyset$, then the predicted label l_x is an element of it, $l_x \in \mathcal{L}_x$. The general algorithm of filtering of the candidate labels is described in Algorithm 2. It starts with an empty set of candidate labels. Given the set of training class models, it tests whether the query instance x is accepted or rejected by each training class according to Definition 1. Indeed, it rejects all the class labels where the query instance do not fit the class distribution, *i.e.* when x lies outside of the class boundary. Only the subset of labels $\mathcal{L}_x$ where x is accepted is retained as the set of candidate labels for prediction.

Algorithm 2: FRW: The label filtering process

Data: $\mathcal{M}$: set of class models, x: test instance
Result: $\mathcal{L}_x$: retained candidate labels
begin
  $\mathcal{L}_x \leftarrow \emptyset$
  foreach *(class model* $\mathcal{M}_l \in \mathcal{M}$*)* **do**
    if $\phi(x, l) \geq 0$ **then**
      $\mathcal{L}_x \leftarrow \mathcal{L}_x \cup l$

3.4.2. Handling Class Overlapping

It is possible to obtain a disjoint set of minimum bounding hyper-spheres in the case where training classes are perfectly separable. In such case, if a query instance x is circumscribed by a hyper-sphere then x takes the latter's class label otherwise x is considered as of an unknown class. However, in real-world cases the hyper-spheres may overlap mainly in the presence of high inter-class similarity. In fact, the overlapping space between classes resembles a local closed-set classification within an open-set classification context. In this case, we train a local closed-set classifier only on the overlapping classes then we use it for only classifying query instances that are within the overlapping space, *i.e.*, instances that are accepted by multiple classes in Algorithm 2, $|\mathcal{L}_x| > 1$.

3.4.3. The Classification Process

Algorithm 3 describes the classification process of FRW. The first step in the prediction is the filtering of candidate labels according to Algorithm 2. If the retained set of candidate labels is an empty set $\mathcal{F}_x^L = \emptyset$, then the query instance x does not fit the distribution of any of the training classes. In this case, the predicted label l_x is set to *"unknown"*. However, if $|\mathcal{F}_x^L| = 1$, then x is only accepted by one training class. In such case, the predicted label is that single filtered possibility $l_x \leftarrow \mathcal{F}_x^L$. In the case where $|\mathcal{F}_x^L| > 1$, x shares similarities with more than one class and its feature vector is projected within the overlapping area between the hyper-spheres of the retained class labels. Since this situation presents a typical closed-set classification, we locally train a closed-set classifier $\mathcal{E}$ only on the retained classes of $\mathcal{F}_x^L$, then we use it to predict the class label l_x of x such that $l_x \leftarrow \mathcal{E}(x)$ and $l_x \in \mathcal{F}_x^L$.

Algorithm 3: FRW: The classification process

Data: $\mathcal{M}$: set of class models, $\mathcal{E}$: local closed-set classifier, x: test instance
Result: l_x: predicted label for x
begin
  $\mathcal{F}_x^L \leftarrow \texttt{FilterLabels}(\mathcal{M}, x)$
  if $\mathcal{F}_x^L = \emptyset$ **then**
    $l_x \leftarrow$ "unknown"
  else
    if $|\mathcal{F}_x^L| = 1$ **then**
      $l_x \leftarrow \mathcal{F}_x^L$
    else
      Train($\mathcal{E}, \mathcal{F}_x^L$)
      $l_x \leftarrow \mathcal{E}(x)$

3.5. Softening Class Boundaries

In order to add flexibility to the classification models, we introduce a softening parameter $\delta \in \mathbb{R}$ that allows to perform a distortion of the class boundary. Indeed, it allows to add more generalization or specialization to the classification models as a trade-off between sensitivity (recall) and specificity. A positive softening extends the radius of the minimum bounding hyper-sphere allowing to add more generalization to the class model. Extending the class boundary may help in detecting test instances that are from the same class but slightly deviate from the training instances. In contrast, a negative softening shrinks the radius of the hyper-sphere which adds more specialization to the class model. Shrinking the class boundary may help in rejecting instances that do not belong to the class but that are within the class hyper-sphere near to the class boundary. In addition, it can be used to alleviate or remove overlapping between classes. If the softening is performed, the value of δ has to be carefully chosen as an over-generalization engenders many false positives. Whereas an over-specialization makes the model under-fit the class and it will only represent a small portion of the class instances.

Definition 2. *(Soft Acceptance Score) The softening can be introduced in the acceptance score. We define the soft acceptance score (φ) as follows:*

$$\varphi(x, l, \delta) = 1 - \frac{\Delta(x, c_l)}{r_l + \delta} \tag{7}$$

where δ is the softening parameter, Δ is an appropriate distance measure, c_l is the center of the class of label l, and r_l is its radius.

Similarly to ϕ, φ is defined in $]-\infty, 1]$ and is interpreted in the same way.

It is worth noting that softening can also be introduced in the training step (instead of φ) during the definition of class boundaries such that line 5 in Algorithm 1 becomes $r_l \leftarrow \texttt{Boundary}\ (D_l) + \delta$. According to Equations 5 and 7, the optimal δ value, denoted δ^*, should be the one that minimizes the classification error as follows:

$$\delta^* = \operatorname*{argmin}_{\delta} Err = \operatorname*{argmin}_{\delta} \sum_{\forall l \in \mathcal{L}} \sum_{\forall x_i \in \mathcal{D}} \Psi(x, l, \delta) \tag{8}$$

where Ψ is defined similarly to Equation 6 but based on φ.

Lemma 1. *Given a classification scenario $S_{\mathcal{D}}$, a performance evaluation technique P, and a closed-set classifier $\mathcal{X}$:*

$$\forall S_{\mathcal{D}}, P(\text{FRW-}\mathcal{X}, S_{\mathcal{D}}) \geq P(\mathcal{X}, S_{\mathcal{D}})$$

Proof 1. *In the worst case, the optimal softening value δ of* FRW-$\mathcal{X}$ *will be very high until the training models completely overlap resembling a typical closed-set classification. In this case, the evaluation instances will be classified using only the local closed-set classifier $\mathcal{X}$. Consequently, P(*FRW-$\mathcal{X}$*, $S_{\mathcal{D}}$) = P($\mathcal{X}$, $S_{\mathcal{D}}$).*

4. Experimental Evaluation

Evaluating an open-set multi-class face recognition technique requires defining proper measures and protocols.

4.1. How Open is Your Open-set Face Recognition?

We propose *Openness* as a measure to quantify the openness of an open-set face recognition scenario ($S_{\mathcal{D}}$).

Definition 3. *(Openness) It measures the ratio of labels that are unseen in training but encountered in prediction to all the labels of the dataset $\mathcal{D}$. Openness is defined as follows:*

$$openness(S_{\mathcal{D}}) = \frac{|UnseenLabels|}{|\mathcal{L}|} \tag{9}$$

Openness is defined in $\mathbb{R}^+$. An *openness* value of 0 means that it is a closed-set classification scenario, otherwise it is an open-set classification. Theoretically, the value of *openness* can be even $+\infty$ which means an infinite set of possibilities (due to the presence of both *known unkown* labels and *unknown unkown* ones). However, in practical cases, the number of test labels can usually be delimited (*known unkowns*). In our experiments, *openness* $\in [0, 1[$ where the open-set face recognition will be simulated from a benchmark dataset were all possible labels are known, *i.e.*, $|\mathcal{L}| = |TrainingLabels| + |UnseenLabels|$. An *openness* of 1 means that $|TrainingLabels| = 0$. This corresponds to a clustering context which is out of the scope of this work.

4.2. Evaluation Technique

To appropriately evaluate the open-set face recognition, careful experimental procedures need to be designed. Conventional evaluation techniques (including hold out, cross validation, random sampling, and their variants) are not suitable for open-set face recognition. They were originally designed for closed-set classification and hence they do not present sufficient restrictions on the labels to simulate an open-set face recognition evaluation. We propose *Leave-P-Class-Out-CrossValidation* as a novel evaluation technique for open-set classification. It allows to simulate an open-set classification that better resembles real-world face recognition applications where we do not have knowledge of all the possible prediction classes. The general procedure of *Leave-P-Class-Out-CrossValidation* is described in Algorithm 4. First, all possible combinations C of $\mathcal{P}$ labels from $\mathcal{L}$ are computed. In each iteration, one possible combination *comb* is randomly chosen from C without replacement. All instances of a label l_{comb}, $\forall l_{comb} \in comb$, are temporarily discarded from the dataset $\mathcal{D}$ to be directly added to the test set. These instances are referred to as the *Leave-out-instances*. All labels in *comb* are unseen in training but encountered in testing which simulates an open-set scenario. A $\mathcal{N}$-fold-cross-validation is performed on the remaining instances ($\mathcal{D}\backslash$*Leave-out-instances*) where in each cross-validation the *Leave-out-instances* are directly added to the test set. The evaluation is repeated until a maximum number of iterations α is reached or no more combination is possible.

Algorithm 4: Leave-P-Class-Out-CrossValidation

Data: $\mathcal{D}$: the classification dataset, $\mathcal{L}$: the set of labels of $\mathcal{D}$, α: the maximum number of iterations, $\mathcal{P}$: the number of labels to leave out in each iteration, $\mathcal{N}$: the number of cross-validation folds, $\mathcal{E}$: the open-set classifier

Result: $Scores$: the classification scores

begin
- $C \leftarrow$ All possible combinations of $\mathcal{P}$ labels form $\mathcal{L}$
- **while** $(\alpha > 0)$ *and* $(C \neq \emptyset)$ **do**
  - Randomly chose a combination *comb* from C
  - *Leave-out-instances*$\leftarrow$ all instances of $\mathcal{D}_{l_{comb}} \mid \forall l_{comb} \in comb$, $comb \subseteq \mathcal{L}$
  - **foreach** *TrainingSet, TestSet* $\in$ $\mathcal{N}$*-CrossValidation(*$\mathcal{D}\backslash$*Leave-out-instances)* **do**
    - Train($\mathcal{E}$, *TrainingSet*)
    - *TestSet* $\leftarrow$ *TestSet* $\cup$ *Leave-out-instances*
    - *PredictedLabels* $\leftarrow$ Predict($\mathcal{E}$, *TestSet*)
    - Scores $\leftarrow$ Scores $\cup$ Statistics(*PredictedLabels*)
  - $C \leftarrow C\backslash comb$
  - $\alpha \leftarrow \alpha - 1$

4.3. Evaluation Measures

The natural way to evaluate classification is to use the accuracy measure which refers to the amount of correctly classified instances from the evaluation set. In multi-class classification the accuracy is averaged over all classes of the dataset. However, in open-set face recognition, the negative set can extremely outnumber the positive set which inflates the classification results causing an over-estimation of the performance of the classifier. Moreover, the number of testing classes is (at least theoretically) undefined. F-measure (also so-called f-score), which is the harmonic mean of precision and recall, represents a good alternative for open-set face recognition. Formally, it is defined as:

$$F\text{-}measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

Where Precision is $\frac{TP}{TP+FN}$ and Recall is $\frac{TP}{TP+FP}$. We use the weighted version of f-measure as the evaluation metric for our experiments. F-measure is computed for each label, then the results are averaged, weighted by the support of each label which makes it account for label imbalance.

4.4. Experimental Protocol and Settings

In order to guarantee an equal participation of all used attributes in the classification, we apply a min-max normalization on each attribute independently such that no attribute will dominate in the prediction ($x_{normalized} = \frac{x-min}{max-min}$, where x is an attribute value, min and max are the minimum and maximum values of the attribute vector). In each experiment, we use FRW to classify a considered dataset in a simulated open-set face recognition using the Leave-P-Class-Out-CrossValidation evaluation. The maximum number of iterations α is set to 10 and the number of cross validations in each iteration is 5. We evaluate the classification performance in terms of weighted f-measure using incremental values of *openness*. We compare our classification results with the gold standard multi-class classification strategy One-vs.-Rest [8] using a linear SVM as the baseline classifier (OvR-SVM), and with the open-set multi-class classifier One-vs.-Set SVM (OvS-SVM) proposed by Scheirer et al. [10]. OvS-SVM is used with the default parameters as requested by the authors, where the generalization/specialization of the hyper-planes are performed automatically through an iterative greedy optimization of the classification risk. We also build a two-step open-set multi-class classifier, termed OCSVM+OvR-SVM, based on Landgrebe et *al.* [13] and Tax et *al.* [14] as discussed in related work (Section2). In the first step of OCSVM+OvR-SVM, a one-class SVM (OCSVM) with an RBF kernel is trained on the entire training instances considered as a single super-class. For OCSVM, instances that deviate from distribution of the super-class are rejected and thus they are labeled as "unknown". Otherwise, the instance is passed to the OvR-SVM for classification where the latter is trained on the original training classes using a linear SVM. For FRW, we use the same closed-set classifier as OvR-SVM, OvS-SVM, and OCSVM+OvR-SVM (*i.e*. SVM with a linear kernel). We show results of FRW-SVM using a fixed softening value δ=-0.3 (*i.e*. -30% expressed in terms of class radius). We also show results of our approach (denoted shortly H-FRW-SVM for Hyper FRW-SVM) using the optimal δ value for each openness where δ^* is obtained through a greedy search within a range of [-0.5, 0.5] with a step size of 0.1. The used distance measure for our approach is the euclidean distance. It is worth noting that FRW is not limited to SVM but it can integrate other closed-set classification algorithms as well. In contrast, OvS-SVM is restricted to the SVM framework. Thus, we use SVM for FRW, OvR-SVM and OCSVM+OvR-SVM for consistency.

4.5. Evaluation Dataset

We evaluate FRW-SVM on a challenging benchmark face recognition dataset namely the Olivetti faces dataset from AT&T Laboratories Cambridge[1]. This dataset consists of a set of 400 pictures, 10 pictures each of 40 individuals. The pictures were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling, *etc*.) and facial details (with/without glasses, *etc*.). Each picture is of a size of 64x64 resulting in a feature vector of 4096 values of gray levels. The task is to identify the identity of the pictured individuals. With so many classes and only 10 examples per class, the classification of this dataset is very challenging. Transforming this dataset into an open-set face recognition task makes the task even more challenging.

4.6. Results and Discussion

Figure 1 shows f-measure results of FRW-SVM using different δ values in a simulated open-set face recognition of *openness*=0.5, meaning that only 20 classes are seen in training and all the 40 classes are encountered in prediction. The obtained results are compared with those of SVM. We notice that SVM performance is very poor compared to that of FRW-SVM. Indeed, SVM assigns one training label to all test instances of the 20 unknown classes leading to a misclassification. FRW-SVM highly outperforms SVM in terms of f-measure by more than 70% order of magnitude in the best case. Indeed, even with no softening (δ=0), FRW-SVM was able to outperform SVM by more than 60% order of magnitude. However, with higher values of δ, the performance of FRW-SVM leans toward that of SVM. This is due to the effect of over-generalization since the bounding hyper-spheres become progressively larger with higher δ values until they completely overlap. In this setting, no rejection will be performed and only the local closed-set classifier (*i.e*., SVM) will be used to classify all instances. With lower δ values, the hyper-spheres become tighter adding more specialization to the class models. This allows FRW-SVM to better reject instances that do not resemble the overall

[1] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

distribution of training classes. However, the value of δ should be carefully specified since an over-specialization leads to a high distortion of the models making them incapable of covering the variance of training classes. The value of δ^* is the one that guarantees the highest f-measure representing the best trade-off between generalization and specialization for the classification scenario.

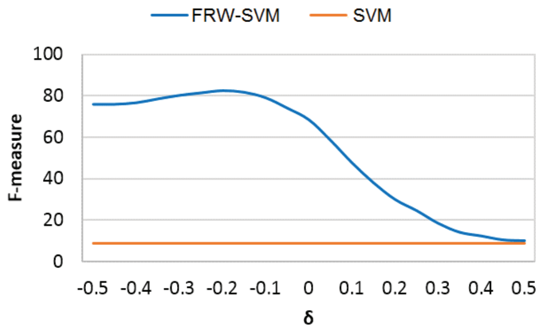
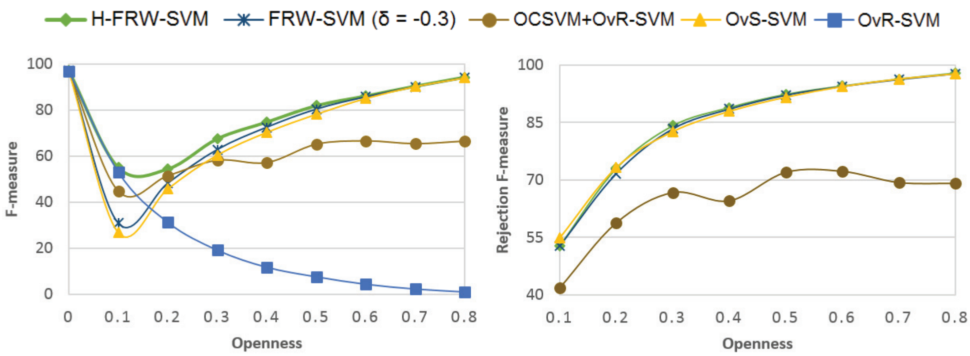


Fig. 1: F-measure performance of FRW-SVM and OvR-SVM in open-set classification of the Olivetti faces dataset with *openness* of 0.5 and using different δ (softening) values.

Figure 2 shows the classification performance in terms of f-measure (Figure 2a) and rejection f-measure (Figure 2b) for H-FRW-SVM (using δ^* in each iteration), FRW-SVM (with a fixed δ=-0.3), OvS-SVM, OCSVM+OvR-SVM, and OvR-SVM using different *openness* values. The value of *openness* ranges from 0 to 0.8 corresponding to a number of held-out classes (P) from 0 to 32 with a step size of 4. As shown in the figure, FRW-SVM handles higher values of *openness* better than all the other approaches. Indeed, even at an extreme *openness* value of 0.8 corresponding to only 8 training classes and 40 testing classes that contain 32 classes that were unseen in training, FRW-SVM was able to classify known as well as unknown class instances with high f-measure of almost 95%. H-FRW-SVM outperformed all the other approaches in open-set classification cases. H-FRW-SVM and FRW-SVM (δ=-0.3) gave close results for open-set classification cases except for *openness*=0.1 where H-FRW-SVM performed better. This can be explained by the fact that in that case more generalization was needed whereas FRW-SVM (δ=-0.3) performed a specialization of -0.3. This conclusion is supported by the f-measure result of the closed-set classifier OvR-SVM in that case, where it outperformed FRW-SVM (δ=-0.3) with no rejection at all. Even though H-FRW-SVM and OvS-SVM used the same closed-set classifier (SVM) and gave very similar results in terms of rejection f-measure, H-FRW-SVM outperformed OvS-SVM in terms of classification f-measure in all open-set classification cases. This is due to the difference between the class representation models used in each approach. Our approach encapsulates each class with a minimum bounding hyper-sphere that isolates it from the rest of the classification universe from all sides. However, OvS-SVM defines two hyper-planes for each class that delimit the latter from only two sides in feature space. In this setting, the class "acceptance space" is left unlimited within the region between the hyper-planes as discussed in Section 2. The classification technique of our approach is more efficient in such classification scenarios with high inter-class overlapping. OvS-SVM outperformed OCSVM+OvR-SVM in most open-set face recognition cases of the Olivetti faces dataset. This is due to the high inter-class overlapping that prevents the one class classifier OCSVM from efficiently isolating the training classes (when considered as one super-class) from the overlapping unknown classes. This is clearly illustrated in the rejection f-measure results in Figure 2b where OCSVM+OvR-SVM scored less than the other open-set classification methods.

5. Conclusion

In this paper, we addressed a fundamental problem in pattern recognition and computer vision namely face recognition. We discussed the bias related to the classification strategy in most existing approaches and we showed that a more realistic scenario that better fits real-world face recognition applications is to transform the problem into an open-set classification. In open-set classification, it is possible to encounter during in prediction, instances that belong to classes that were unseen in training. In this setting, it is necessary to define a decision boundary for each class



(a) (b)

Fig. 2: F-measure (a) and rejection f-measure (b) results of H-FRW-SVM, FRW-SVM (δ=-0.3), OvS-SVM, OCSVM+OvR-SVM and OvR-SVM in open-set classification of the Olivetti faces dataset with different *openness* values.

that envelops the class instances and resembles its distribution. In many real-world applications where the closed-world hypothesis does not hold, it is important to detect such unknown instances and raise the attention of experts to address them separately, preventing a misclassification. We introduced FRW, an approach for open-set face recognition. FRWencapsulates each class with a minimum bounding hyper-sphere that resembles the class distribution by enclosing as many as possible of its instances. In such manner, our method is able to distinguish instances that resemble previously seen classes from those that are of unknown ones. FRWpresents a high flexibility through a softening parameter that allows extending or shrinking class boundaries to add more generalization or specialization to the classification models. Experimental results on a challenging benchmark dataset show the efficiency of FRWin open-set face recognition compared to gold standard approaches from the literature. An evaluation procedure was also introduced to adequately evaluate open-set face recognition .

An interesting future work is to propose a representation model for non spherical like shaped classes in order to avoid the risk of over-generalization in empty regions of the hyper-sphere. Furthermore, another direction is to develop an estimation method for a fast discovery of δ^*, avoiding a greedy search across all possibilities.

References

1. Kanade, T.. Picture processing system by computer complex and recognition of human faces. In: *doctoral dissertation, Kyoto University*. 1973, .
2. Turk, M.A., Pentland, A.P.. Face recognition using eigenfaces. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; CVPR '91. 1991, p. 586–591.
3. Belhumeur, P.N., Hespanha, J.a.P., Kriegman, D.J.. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997;**19**(7):711–720.
4. Cheng, Y.Q., Liu, K., Yang, J., Zhuang, Y.M., Gu, N.C.. Human face recognition method based on the statistical model of small sample size. *Proc SPIE* 1992;**1607**:85–95.
5. Quinlan, J.R.. Simplifying decision trees. *International Journal of Man-Machine Studies* 1987;**27**(3):221–234.
6. Cortes, C., Vapnik, V.. Support-vector networks. *Machine Learning* 1995;**20**(3):273–297.
7. Tax, D.. *One-class classification: Concept learning in the absence of counter-examples*. Ph.D. thesis; Technische Universiteit Delft; 2001.
8. Rocha, A., Goldenstein, S.. Multiclass from binary: Expanding one-versus-all, one-versus-one and ecoc-based approaches. *IEEE Transactions on Neural Networks and Learning Systems* 2014;**25**(2):289–302.
9. Raskutti, B., Kowalczyk, A.. Extreme re-balancing for svms: A case study. *ACM SIGKDD Explorations Newsletter* 2004;**6**(1):60–69.
10. Scheirer, W., Rocha, A., Sapkota, A., Boult, T.. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2013;**35**(7):1757–1772.
11. Torralba, A., Efros, A.A.. Unbiased look at dataset bias. In: *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*. IEEE Computer Society; 2011, p. 1521–1528.
12. Zhu, X., Goldberg, A.B., Brachman, R., Dietterich, T.. *Introduction to Semi-Supervised Learning*. Morgan and Claypool Publishers; 2009. ISBN 1598295470, 9781598295474.
13. Landgrebe, T., Paclík, P., Tax, D.M.J., Duin, R.P.W.. Optimising two-stage recognition systems. In: *Proceedings of the 6th International Conference on Multiple Classifier Systems*; MCS'05. Springer-Verlag; 2005, p. 206–215.
14. Tax, D.M.J., Duin, R.P.W.. Growing a multi-class classifier with a reject option. *Pattern Recognition Letters* 2008;**29**(10):1565–1570.