Face recognition algorithms: performance evaluation

Project Report

Marco Del Coco - Pierluigi Carcagnì

Institute of Applied Sciences and Intelligent systems
c/o Dhitech scarl Campus Universitario
via Monteroni sn, 73100 Lecce Italy

Institute of Applied Sciences and Intelligent systems
c/o Dhitech scarl Campus Universitario
via Monteroni sn, 73100 Lecce Italy

Title:
Project Title: Face recognition algorithms: performance evaluation

Theme:
Scientific Theme

Project Period:
September 2016

Participant(s):
Marco Del Coco
Pierluigi Carcagnì

Supervisor(s):
Prof. Cosimo Distante

Group:
U.O.S Lecce

Page Numbers: 10

Date of Completion:
November 11, 2016

Abstract:

This report aims at an accurately performance investigation on the running face recognition algorithm proposed by the CNR-ISASI for face recognition and two additional solutions at present under consideration. All the tests have been executed on three challenging datasets and curves relative to identification rate, as a function of rank, and false acceptance rate, have been reported and discussed.

1 Introduction

This report provides a validation of RECO3.26's latest face recognition technology. A team from the CNR-ISASI has performed the evaluation study. The results obtained demonstrate that RECO3.26's identification figures are correct, attaining excellent performance.

Automatic recognition and verification from face image are two hot topics both in research and industrial fields due to the increasing security requirements in public places. Recognition is the most challenging topic; it mainly consists in a One-to-Many association of subjects. Many solutions have been proposed among the years leading to results worth of note. Anyway, the recognition problem is intrinsically liable to a decrease of the accuracy when the dataset size grows. Indeed, given a subject under test (to be recognized) the operation is more challenging, bigger is the number of possible identities. Identification is, in some way, a simpler problem; it consists, given a pair of face images, in answering to the question: "Are the two images under test representing the same subject"? In this report the recognition problem will be detailed referring to the Reco3.26's algorithms. In the following, sections 2 will be aimed at the presentation of the generic processing scheme employed in the proposed system, the problem formulation, the definition of the employed performance metrics and finally at a detailed description of the employed testing procedure; in section 3 will be presented the employed datasets, algorithms and the experimental results; as a last step, in section 4 will be summed up the most relevant information.

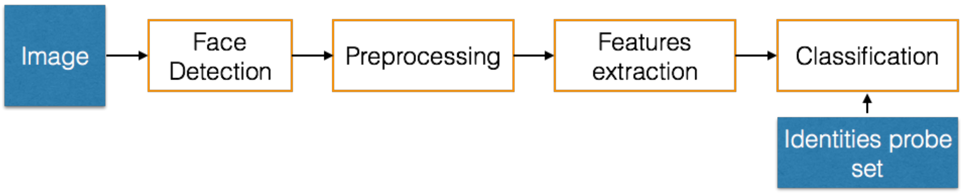
2 Evaluation procedure

This section aims at the description of the generic pipeline managing the recognition protocol and at the testing problem definition and description.

2.1 General Processing Scheme

Face recognition issue has been addressed by means of different approaches among the years. Each approach exploits different aspect and processing techniques, anyway it is possible to define a common scheme shared by all the solution. The scheme represented in figure 1 represents the key steps and elements in the face recognition process. The first step is devoted to the face detection in the running image. The detected face is then preprocessed in order to make it suitable to the constraints required by the next step. The obtained outcomes are the input of the *features extraction*, aimed at extracting a new representation of facial regions, depending on the selected approach. Finally the obtained data are compared with the probe set (a dataset of known identities) and the predicted identity is returned.

Figure 1: Generic processing scheme for face recognition



2.2 Problem formulation

In order to describe the testing procedure and the obtained results, as clear as possible, a problem formulation is mandatory. Let x_i be the i-th facial image, X the whole face image set whose criminality is N and containing L different identities with $L < N$. Moreover the j-th identity of the i-th facial image can be defined as $y_j = ID(x_i)$.

The recognition procedure consists in the assumption that both an images **probe set** and a **test set** are available. An image probe set is a set of images whose identities are known (enrolled) and one or more images are available for the same identity. On the other hand the set of image under test (virtually unknown) is defined as the **test set**. Finally, let T be the threshold used to define if the recognition score between two subjects (a test and an enrolled) is enough to consider the probe image a reasonable candidate.

It is typical that a search is conducted into an enrolled population of S identities (the **probe set**), and that the algorithm ranks all the scores returned by the search in the probe set in descending order. In practical applications, a human analyst might examine all S candidates or a subset made up by the fist R candidates, or only those with score greater than threshold, T .

Given the S returned candidates in a search, a shorter candidate list can be prepared by taking the top $R \leq S$ candidates and the threshold T can be applied. It is useful then to state accuracy in terms of rank R and threshold T. According with these definition the *false negative identification rate* (FNIR) can be defined as follows:

$$\texttt{FNIR}(N, R, T) = \frac{\begin{array}{c}\text{Num. mate searches with enrolled mate found}\\ \text{outside top R ranks or score below threshold, T}\end{array}}{\texttt{Num. mate searches attempted}} \tag{1}$$

whereas the *true positive identification rate* (TPIR) is defined as:

$$\texttt{TPIR}(N, R, T) = \frac{\begin{array}{c}\text{Num. mate searches with enrolled mate found}\\ \text{inside top R ranks or score above threshold, T}\end{array}}{\texttt{Num. mate searches attempted}} \tag{2}$$

.

This formulation does not distinguish between causes of misses (below the threshold or outside of top R).

In order to account the worst case, given a dataset of face images with labeled identities and accounting for the nomenclature presented above, the test procedure has been performed in the following way. For each face image x_i (such that the cardinality of its identity is greater than 1) in the dataset, x_i is taken as test set and $\overline{X} = \{X - x_i\}$ is the probe set; than the scores for x_i respect to the probe set is computed sorted in descending order.

Now, the identification rate can be easily expressed as a function of the rank. We would define the identification rate for a given rank R as the percentage of trial in which at least 1 occurrence x_z of the returned elements of the probe set fall inside the fists R identities such that $ID(x_z) == ID(x_i)$ with x_i the image under test.

A second metrics of interest is given by the TPIR expressed as a function of FNIR. More specifically, the thresholds for FNIR values between 0 and 1 have been computed and successively the TPIR values have been computed for each threshold (keeping the constraints $R = 1$).

3 Experimental setup and Results

This section is devoted to the presentation and discussion of experimental outcomes and to the employed set-up. First of all a brief description of the facial images datasets will be provided to the reader and successively the experimental results will be shown and discussed.

3.1 Dataset description

The experimental session exploited three different datasets in order to cover the wide scenario is possible. The Morph dataset, made up by frontal face images in a controlled environment, the Color Feret datset showing faces in a controlled environment but with poses variation and finally, the Labeled Faces in the Wild (LFW) dataset, a collection of unconstrained face collected from the web.

The Color Facial Recognition Technology (FERET) database NIST [1996] contains a total of 11338 facial images with a resolution of 512×768 pixel. They were collected by photographing 994 subjects at various angles, over the course of 15 sessions between 1993 and 1996. This database is mainly a color version of the original FERET database NIST [2001] which was released in 2001 and consisted of 14051 grayscale images. Some examples of pose variation among the same subject are reported in figure 2

The LFW data set contains more than 13,000 images of faces (most of them at 250×250 pixel) collected from the web. Each face is labeled with the name of the person pictured that we translated in a numeric identification number. 1680 of the people pictured have two or more distinct photos in the data set. Some examples of the unconstrained condition are reported in figure 3

Morph datset consists of face images of people of different gender, race and age, provided with complete annotation data. Most of the subject, present at least one replicated photo sample in different date (same person pictured in different years). The resolution of facial images is 200×240 or 400×480 pixels. The original dataset was built without taking into account the balancing of the cardinality of each class. It consists of 46645 male subjects and 8489 female subjects. Some examples of pose variation among the same subject and different ages are reported in figure 4

Figure 2: Some examples of face images and poses variation among the same subject in the Color Feret Dataset

(a) Frontal **(b)** Frontal **(c)** Rotated **(d)** Rotated

Figure 3: Some examples of face images and poses variation among the same subject in the LFW Dataset

(a) Rotated **(b)** Frontal **(c)** Rotated **(d)** Frontal

All the datasets present specific challenges. Morph is characterized by a high cardinality value, feret and LFW are challenging due to the many different poses from the frontal to the profile one; LFW, in particular, is made up by low quality image captured in unconstrained condition. It is deserved to highlight as some image could be missed by the detection step, depending on the specific adopted approach. Consequently, all the results have been computed accounting for the faces effectively detected in the detection step, whose cardinality is reported for every plot curve.

Figure 4: Some examples of face images and poses variation among the same subject in the Morph Dataset

(a) Age 67 **(b)** Age 68 **(c)** Age 69 **(d)** Age 69

3.2 Results

The figures presented in this report reflect performance of the most recent face RECO3.26 recognition engines; the former based on appearance and the latter involve a Convolutional Neural Network (CNN), respectively:

LA : *Local Appearance Algorithm* employs appearance information evaluated on specific face zones plus a whole face image analysis.

CNNA : *Convolutional Neural Network Algorithm* employs a features analysis based on the most recent convolutional neural network analysis.

It is worth noting as LA and CNN algorithms employ different detection strategy.

3.2.1 Results on Feret

Feret datset represents a good compromise between the challenge of pose variation and the constraints of a controlled environment situation. LA recognition step detects 5451 images, whereas, the CNN one is capable to detect a total amount of 7592 images, probably due to its higher generalization capabilities.

The first plot (figure 5) reports the recognition percentage against the rank. The running algorithm (LAA) shows good performance and a good capability to manage the challenge of pose variation. The CNNA solution, anyway, shows better results due to its generalization capability in detection and recognition. It is important to underline that CNNA is currently under investigation on different dataset sand under different condition and that, consequently, they can not be considered mature.

The second plot (figure 6) reports the TPIR against FNIR for rank=1. The curves are close to be planar meaning an high independence on the threshold problem.

3.2.2 Results on LFW

LFW datset is characterized by a completely in-the-wild environment that represent the most challenging aspect in the recognition goal.

LA recognition step detects 12717 images, whereas, the CNN one is capable to detect a total amount of 13175 images.

Figure 7 reports the recognition percentage against the rank. It is clear as, compared the Feret dataset, the challenges introduced by the pose variation and low resolution affect the recognition capabilities pushing down all the curves. Anyway, the discussed difficulties bring out the best of CNNA approach showing an increased gap among the LA approach.

The second plot (figure 8) reports the TPIR against FNIR for rank=1. Also in this case the curves are close to be planar.

3.2.3 Results on Morph

Last comparison is done on the Morph that is the most constrained dataset accounting only frontal face. Also in this case the LA algorithms is outperformed by the CNNA ones in terms of number of recognized images (54062 and 55022 respectively). Anyway, surprising, the Rank vs Identification rate plot (figure 8) shows an inversion in performance on the low rank values where the CNNA solution is outperformed by the LA one that, probably, designed on the frontal face environment, take advantage to work in its comfort context. The criticism of identification rate under the case of rank= 1 for CNNA algorithm make the change in FPIR vs TPIR more significant as highlighted in figure 10

4 Conclusion

In this report the performance of face recognition algorithms proposed by the CNR-ISASI have been investigated. All the test have been conducted on three challenging datasets made up by face image collection of frontal and non frontal views. The running algorithm exhibits good performance compatible with the application field. The CNNS solution shows interesting results capable to outperforms the running one; anyway a deeply investigation is necessary.

5 Images

Figure 5: Rank vs identification rate (Feret)

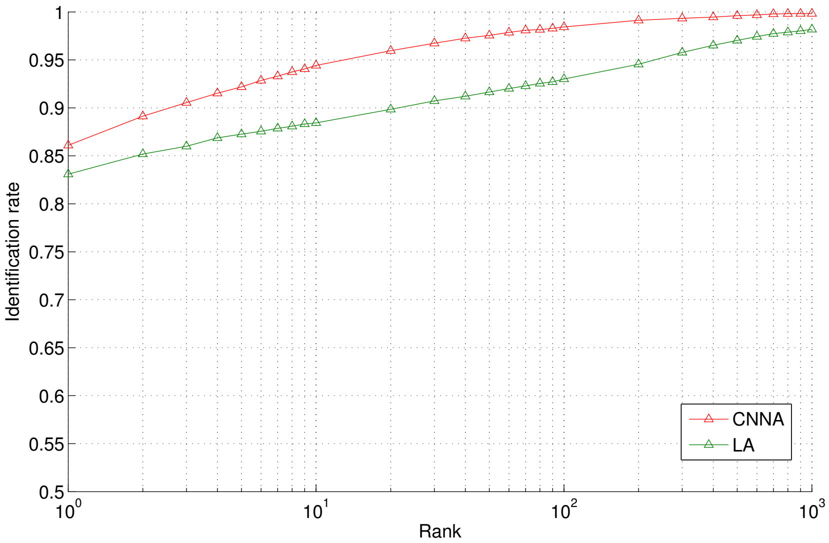


Figure 6: TPIR vs FNIR (Feret)

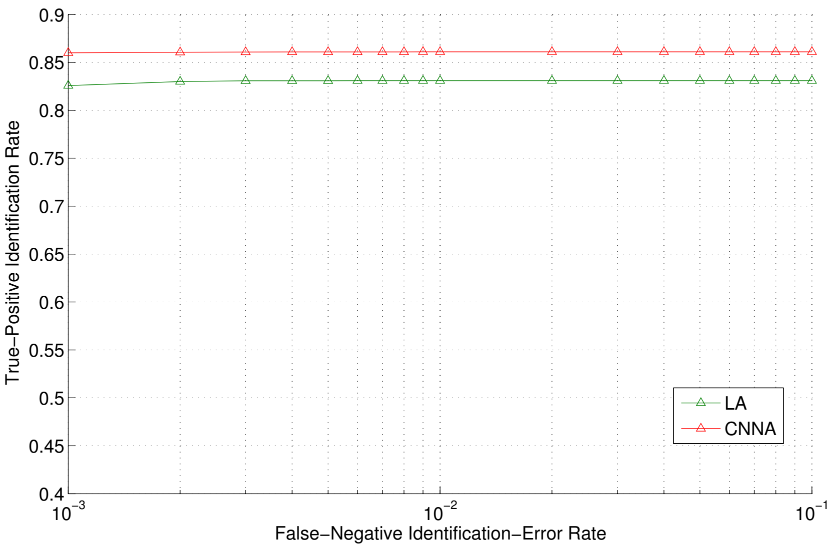


Figure 7: Rank vs identification rate (LFW)

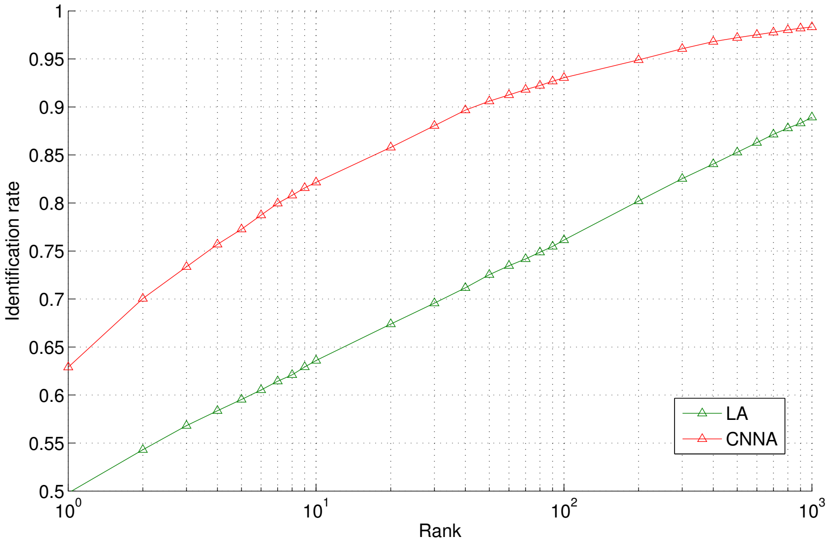


Figure 8: TPIR vs FNIR (LFW)

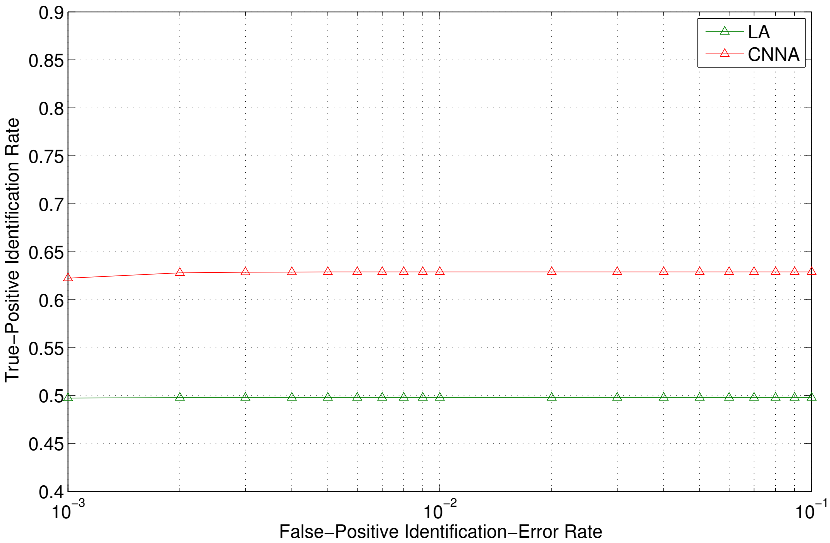


Figure 9: Rank vs identification rate (Morph)

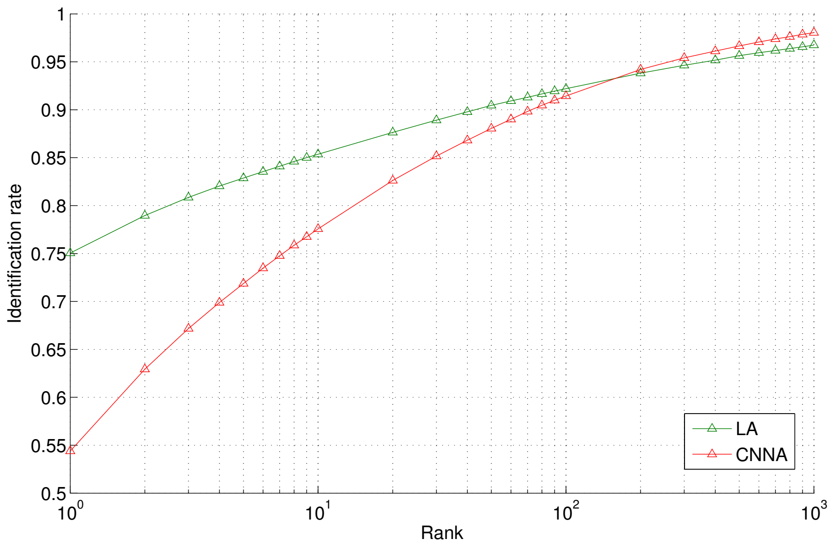
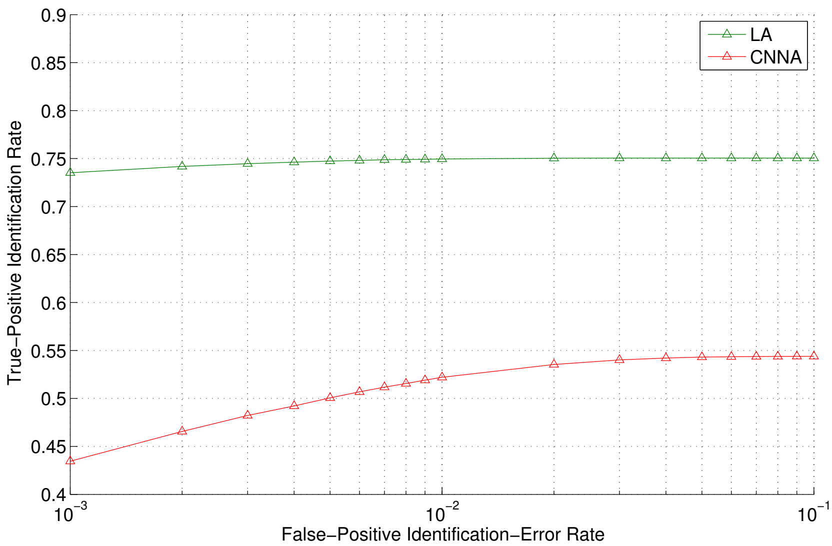


Figure 10: TPIR vs FNIR (Morph)



Bibliography

NIST (1996). Color feret (facial recognition tecnology). http://www.nist.gov/humanid/colorferet.

NIST (2001). Gray scale feret (facial recognition tecnology). http://www.nist.gov/humanid/eret.