# Evaluating systems assessing face-image compliance with ICAO/ISO standards[1]

M. Ferrara, A. Franco, D. Maltoni

C.d.L. Scienze dell'Informazione - Università di Bologna, via Sacchi 3, 47023 Cesena, Italy
DEIS – Viale Risorgimento, 2 – 40126 Bologna, Italy.
E-mail: {ferrara, franco, maltoni}@csr.unibo.it

**Abstract.** This paper focuses on the requirements for face images to be used in Machine Readable Travel Documents, defined in the ISO/IEC 19794-5 standard. In particular an evaluation framework is proposed for testing software able to automatically verify the compliance of an image to the standard. The results obtained for thee commercial software are reported and compared.

## 1. Introduction

Face represents one of the most used biometric traits, for both computer automated and human assisted person identification. To allow interoperability among systems developed by different vendors and simplify the integration of biometric recognition in large-scale identification (e-passport, visas, etc.) a standard data format for digital face images is needed. In this context, the International Civil Aviation Organization (ICAO) started in 1980 a project focused on machine assisted biometric identity confirmation of persons. Initially three different biometric characteristics where identified for possible application in this context (face, fingerprint, iris), but finally face was selected as the most suited to the practicalities of travel document issuance, with fingerprint and/or iris available for choice by States for inclusion as complementary biometric technologies. Of course high quality, defect-free digital face images are needed to maximize both the human and computer assisted recognition accuracy. Starting from the ICAO work, in 2004 the International Standard Organization (ISO) defined a standard [3] for the digital face images to be used in the Machine Readable Travel Documents. The standard specifies a set of characteristics that the image has to comply, mainly related to the position of the face in the image and to the absence of defects (blurring, red eyes, face partially occluded by accessories, etc.) that would affect both the human and automatic recognition performance.

In view of the widespread adoption of the new standard, some vendors of biometric technologies started to develop and distribute software applications able to automatically verify the compliance of a face image to the ISO standard. However,

[1]This work was partially supported by CNIPA (Italian National Center for ICT in the Public Administration).

until now no independent and systematic evaluation of these algorithms have been done, and it is not clear if these systems can effectively assist or substitute humans in checking face-image compliance with the standards.

To the best of our knowledge one of the few experiments related to this issue has been carried out by the Federal Office for Information Security (BSI) in Germany, one of the first European countries to adopt the electronic passport; this evaluation [5], aimed at verifying the compliance of face images to the ISO/IEC 19794-5 standard [3], was performed on 3000 images from field applications, and was carried out mainly by manual inspection.

The aim of this paper is to define a testing protocol for the automatic evaluation of systems verifying compliance of face-images with ISO/IEC 19794-5 standard. Starting from the guidelines and the examples of compliant and non-compliant images provided in the ISO standard, a set of salient characteristics has been identified and encoded, a precise evaluation protocol has been defined, and a software framework has been developed to fully automate the test. We believe that the possibility of fully automating such evaluation is a crucial point since it allows to effortless repeat the test on new systems and new databases.

The paper is organized as follows: in section 2 the main ISO requirements are detailed, in section 3 the evaluation protocol and framework are introduced; section 4 presents the experiments carried out and finally in section 5 some concluding remarks are given.

## 2. The ISO/IEC 19794-5 standard and the tests defined

The ISO/IEC 19794-5 international standard [3] specifies a record format for storing, recording and transmitting the facial image information and defines scene constraints, photographic properties and digital image attributes of facial images.
Each requirement is specified for different face image types:

- *Full frontal*. Face image type that specifies frontal images with sufficient resolution for human examination as well as reliable computer face recognition. This type of image includes the full head with all hair in most cases, as well as neck and shoulders.
- *Token frontal*. Face image type that specifies frontal images with a specific geometric size and eye positioning based on the width and height of the image. This image type is suitable for minimizing the storage requirements and to simplify computer based recognition (the eyes are in a fixed position).

The requirements introduced by the ISO standard are organized in two categories: geometric and photographic requirements.

The *geometric requirements* are related to the position of the face and of its main components (eyes, nose, etc.) within the digital image. In Fig. 1 the geometric characteristics of the digital image used to specify the requirements for the full frontal format are shown. The following basic elements are considered in the definition of the requirements:

– A: image width, B: image height;
– AA: imaginary vertical line positioned at the center of the image;
– BB: vertical distance from the bottom edge of the image to an imaginary horizontal line passing through the center of the eyes;
– CC: head width defined as the horizontal distance between the midpoints of two imaginary vertical lines; each imaginary line is drawn between the upper and lower lobes of each ear and shall be positioned where the external ear connects the head;
– DD: head height defined as the vertical distance between the base of the chin and the crown.

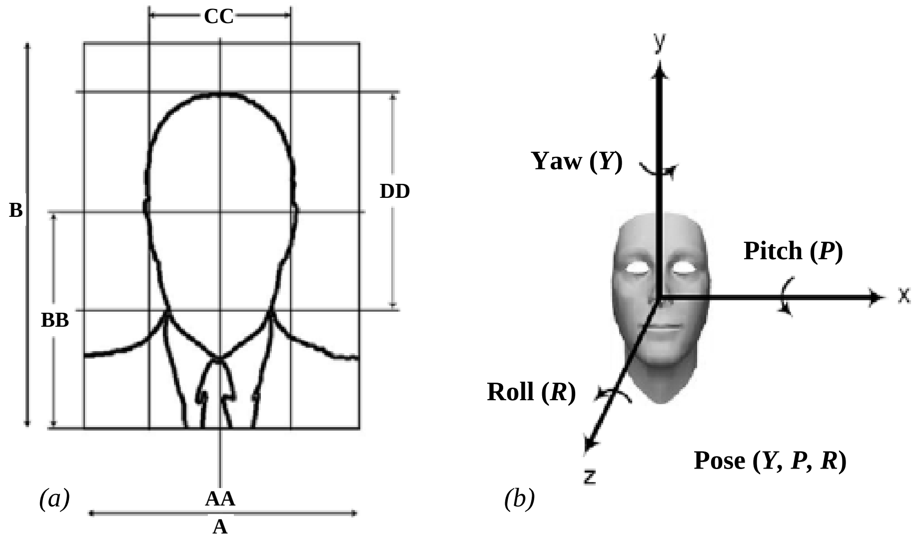


**Fig. 1** Geometric characteristics of the Full Frontal Face Image *(a)* and definition of the pose angles with respect to the frontal view of the subject *(b)*.

The *photographic requirements* refer to characteristics of the face (e.g. expression, mouth open) and of the image (e.g. focus, contrast, natural skin tone). Starting from the guidelines and the examples of acceptable/unacceptable images provided in [3], we defined a set of tests (see Table 1).

The token face image format inherits the requirements of the frontal face image type [3], does not require to comply with the geometric constraints of full frontal images (see tests 3..7 in Table 1), but enforces other geometric constraints related to the eyes position and the image size proportion (see Table 2).

**Table 1.** Tests defined to evaluate systems for ISO compliance check. The last column of the table (Section) denotes the section of [3] from which the test was derived.

| N° | Description of the test | Section |
|---|---|---|
| **Feature extraction accuracy tests** | | |
| 1 | Eye Location Accuracy | |
| 2 | Face Location Accuracy (other points) | |
| **Geometric tests (Full Frontal Image Format)** | | |
| 3 | Eye Distance (min 90 pixels) | 8.4.1 |
| 4 | Relative Vertical Position (0.5B<=BB<=0.7B) | 8.3.3 |
| 5 | Relative Horizontal Position (no tolerances) | 8.3.2 |
| 6 | Head Image Width Ratio (0.5A<=CC<=0.71A) | 8.3.4 |
| 7 | Head Image Height Ratio (0.7B<=DD<=0.8B) | 8.3.5 |
| **Photographic and pose-specific tests** | | |
| 8 | Blurring | 7.3.3 |
| 9 | Looking Away | 7.2.3 |
| 10 | Ink Marked/Creased | A3.2.3 |
| 11 | Unnatural Skin Tone | 7.3.4 |
| 12 | Too Dark/Light | 7.3.2 |
| 13 | Washed Out | 7.4.2.1 |
| 14 | Pixelation | A3.2.3 |
| 15 | Hair Across Eyes | A3.2.3 |
| 16 | Eyes Closed | 7.2.3 |
| 17 | Varied Background | A2.4 |
| 18 | Roll/Pitch/Yaw Greater 5 | 7.2.2 |
| 19 | Flash Reflection on Skin | 7.2.10 |
| 20 | Red Eyes | 7.3.4 |
| 21 | Shadows Behind Head | A3.2.3 |
| 22 | Shadows Across Face | 7.2.7 |
| 23 | Dark Tinted Lenses | 7.2.11 |
| 24 | Flash Reflection on Lenses | 7.2.11 |
| 25 | Frames too Heavy | A4.3 |
| 26 | Frame Covering Eyes | 7.2.3 |
| 27 | Hat/Cap | A3.2.3 |
| 28 | Veil over Face | A3.2.3 |
| 29 | Mouth Open | 7.2.3 |
| 30 | Presence of Other Faces or Toys too Close to Face | A3.2.3 |

**Table 2.** Geometric tests for the token image type. The last column of the table (Section) denotes the section of [3] from which the test was derived.

| Geometric tests (Token Frontal Image Format) | Section |
|---|---|
| Image Width W (min 240 pixels) | 9.2.4 |
| Image Height (= W / 0.75) | 9.2.3 |
| Y Coordinate of Eyes (=0.6 * W) | 9.2.3 |
| X Coordinate of First (right) Eye (=(0.375 * W) – 1) | 9.2.3 |
| X Coordinate of Second (left) Eye (=(0.625 * W) – 1) | 9.2.3 |
| Width from Eye to Eye (inclusive) (=0.25 * W) | 9.2.3 |

## 3. The software framework

A software framework has been developed to evaluate and analyze the performance of algorithms provided in the form of SDK (Software Development Kit). The framework offers the following functionalities.

– *Manual image labeling*. It allows to load a database of images and, for each of them, to manually:
  - label by point and click the main facial features such as eye centers, center of mouth, nostrils, etc.;
  - specify the compliance of the image with respect to the characteristics underlying the tests 8..30 in Table 1. Labels are tri-state values (compliant, non-compliant and *dummy*). A *dummy* label is assigned when the human expert is not confident enough whether the image is compliant or not.

– *Artificial dataset generation*. Most of the images used for the tests belong to face databases available to the scientific community; for some of the tests it is very difficult to find in these databases a sufficient number of non-compliant images. The framework offers a tool to generate artificial images non-compliant with respect to a particular characteristic by applying some image processing to "real" compliant images. In the current version of the framework the following transformations are available: blurring, brightness and contrast adjustment, pixelation, addition of red eyes. Each transformation is characterized by a specific set of parameters that can be tuned to control the effect of the operation on the real image (see Fig. 2 for some example).

– *Automatic SDK testing*. In order to interface the framework with different SDKs a simple interface protocol based on a command-line executable has been defined and provided to the SDK vendors. The executable evaluates the compliance of a single image and provides in output a compliance degree (in the range 0..100) for each of the characteristics underlying the tests in Table 1. The results obtained can be analyzed and compared on the basis of several performance indicators among which EER, FAR/FRR curves, DET graphs. Any new SDK, in order to be tested, simply needs to comply with the defined testing protocol.

## 4. Experiments

Three commercial SDKs, whose names cannot be disclosed (here referred to as A, B and C), have been evaluated in our experiments; for each of them the compliance of each image in the dataset has been measured with respect to the characteristics 1, 8..30 underlying the cases reported in Table 1; the geometric tests 2..7 are not included in this study, because of the non-uniform way the different SDKs provide in output details about the location of internal face-feature. This part of the evaluation will be done in a successive study.

Analogously to a biometric verification systems the SDKs here evaluated can make two types of errors: declaring compliant with respect to a given characteristic an image that is non-complaint (False Acceptance) and declaring non-compliant and image that is compliant (False Rejection). Images labeled as dummy for a given characteristic are excluded from the corresponding test.

According to this protocol, the results are reported for each characteristic in terms of EER and rejection rate. A rejection occurs when either the SDK is not able to process an image or the image is processed but the SDK is not able to evaluate the specific characteristic. According to the best practices the rejection is here included in the calculation of EER [4]: this is implicitly done by, assuming that a 0 compliance degree (for the given characteristic) is returned in case of rejection. This choice is aimed at discouraging the software to reject the most uncertain cases thus improving the performance over processed images.

### 4.1 The database

The dataset used is the publicly available AR Face Database [1] containing 2000 images of different subjects. Unfortunately some of the images are defective or not available, so that finally 504 images from 126 subjects have been selected. The images, whose original size is 768×576 pixels, have been cropped to 480 (w) × 640 (h). The database contains about four images of each subject: one with natural lighting and expression, two with evident facial expressions (smile and angry) and one with a strong lateral illumination.

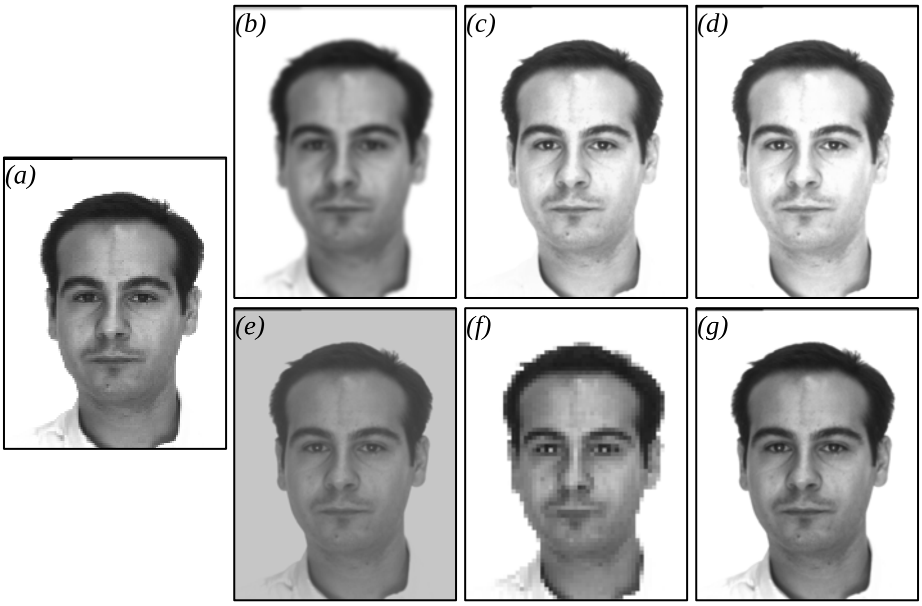


**Fig. 2.** An example of application of each image operation. *(a)* Original image; *(b)* blurred; *(c)* unnatural skin tone; *(d)* too dark/light; *(e)* washed out; *(f)* pixelation; *(g)* red eyes.

The presence of images with varying expression and lighting allows to verify the ability of the various SDKs to evaluate the compliance with respect to some of the characteristics given in section 2. Unfortunately the original dataset contains no (or a few) images non-compliant with respect to some of the requirements identified. In order to carry out a more precise evaluation of all the characteristics, some additional "artificial" datasets have been generated by applying specific digital image operations (see Fig. 2) that cannot be described here in detail for lack of space. In particular, derived datasets have been generated for blurring (1008 images), unnatural skin tone (748), too dark/light (735), washed out (1008), pixelation (1008) and red eyes (1000). The images in these datasets are equally distributed between compliant and non-compliant.

### 4.2 Experimental results

The results of the evaluation carried out are reported in this section. As to the geometric requirements, the eye localization accuracy of the SDKs is shown in Fig. 3. The columns refer to increasing intervals of localization errors (in pixels). The column "Correct" includes all the cases where the maximum error (among the single errors for the two eyes) is lower than 6 pixels. On average, in the images used for testing, the distance between the two eyes is 125 pixels. The result for SDK B is not reported since it does not output the eye position. The two SDKs achieve a very good localization accuracy, even in the presence of difficult cases; A is the most accurate.

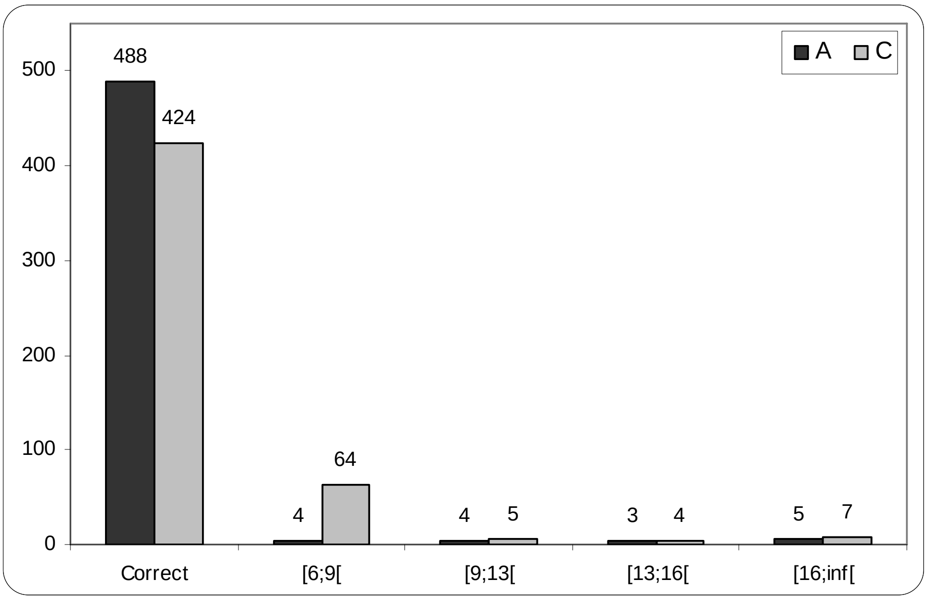


**Fig. 3.** Eye localization accuracy.

The results obtained by the three SDKs on tests 8..30 are reported in Table 3 where the EER and rejection rate are given. The rejection rate is in most cases quite low, but it is worth noting that this value for SDK A is noticeable for some characteristics (e.g. hair across eyes). For a further comparison of the three SDKs, the results in terms of EER shown in Table 3 are summarized in Fig. 4 where the EER distribution for the three SDKs is reported. Five EER intervals have been defined and each bar of the graph represents the number of tests that a given SDK is able to manage with an accuracy value included in the related range.

**Table 3.** EER and Rejection Rate of the three SDKs evaluated.

| Characteristic | A | | B | | C | |
|---|---|---|---|---|---|---|
| | **EER** | **Rej.** | **EER** | **Rej.** | **EER** | **Rej.** |
| 8 Blurred | **1.88%** | 3.47% | 2.48% | 0.30% | 65.87% | 0.60% |
| 9 Looking Away | 1.79% | 0.20% | - | - | **1.19%** | 0.40% |
| 10 Ink Marked/Creased | - | - | - | - | - | - |
| 11 Unnatural Skin Tone | 7.09% | 0.00% | 50.00% | 0.13% | **2.67%** | 0.40% |
| 12 Too Dark/Light | - | - | **25.15%** | 0.14% | 25.17% | 0.54% |
| 13 Washed Out | - | - | 23.11% | 0.99% | **0.79%** | 1.98% |
| 14 Pixelation | - | - | **1.39%** | 0.50% | - | - |
| 15 Hair Across Eyes | **50.00%** | 94.44% | - | - | - | - |
| 16 Eyes Closed | **12.11%** | 2.90% | - | - | 22.59% | 0.41% |
| 17 Varied Background | **17.91%** | 0.24% | 48.87% | 0.72% | 46.86% | 0.48% |
| 18 Roll/Pitch/Yaw Greater 5 | - | - | **13.96%** | 0.60% | 43.72% | 0.40% |
| 19 Flash Reflection on Skin | **0.51%** | 0.20% | 49.38% | 0.60% | - | - |
| 20 Red Eyes | 4.86% | 0.60% | 50.00% | 0.99% | **3.70%** | 1.10% |
| 21 Shadows Behind Head | - | - | - | - | - | - |
| 22 Shadows Across Face | **28.94%** | 2.78% | - | - | 34.77% | 0.40% |
| 23 Dark Tinted Lenses | - | - | - | - | **25.00%** | 0.40% |
| 24 Flash Reflection on Lenses | - | - | - | - | **22.77%** | 0.42% |
| 25 Frames too Heavy | - | - | - | - | - | - |
| 26 Frame Covering Eyes | 50.00% | 93.82% | - | - | **16.18%** | 0.20% |
| 27 Hat/Cap | - | - | - | - | - | - |
| 28 Veil over Face | - | - | - | - | / | 0.40% |
| 29 Mouth Open | **5.88%** | 23.72% | - | - | 14.64% | 0.20% |
| 30 Objects too Close to Face | - | - | - | - | - | - |

- the SDK does not support the test for this characteristic
/ the EER is not calculated since the dataset does not contain non-compliant images
The bolded values indicate the best performance for each characteristic.
The grayed rows correspond to characteristics evaluated mainly on compliant images. For these characteristics additional tests on extended datasets are needed.

It is worth noting that the number of characteristics evaluated by the three SDKs is different: in particular, A verifies 11 requirements, reaching in most cases a good accuracy; B evaluates only 9 requirements and the results obtained are mostly unsatisfactory; finally, C deals with 14 requirements and the accuracy is quite variable and strictly dependent on the specific requirement.

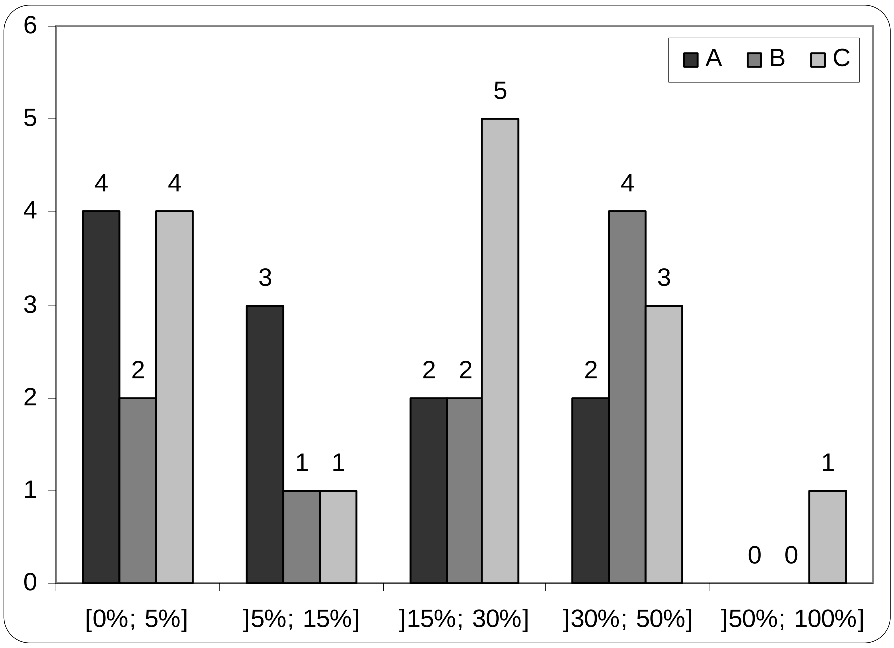


**Fig. 4.** Distribution of the three SDKs accuracy in five EER intervals. The x-axis reports the EER ranges, and the y-axis indicates the number of tests on which a SDK reaches an EER included in that range.

## 5. Conclusions

This work addresses the problem of evaluating the accuracy of automatic software for ISO/IEC 19794-5 compliance check. To this purpose, a testing protocol and an evaluation framework have been developed.

The results show that the three SDKs evaluated are able to accurately check only some characteristics while achieving unsatisfactory performance for others. An analysis of the results in Table 3 and Fig. 4 show that some requirements (e.g. blurred, unnatural skin tone, washed out) are easily verifiable by an automatic software. On the other hand, characteristics like hair across eyes or frame covering eyes are difficult to be automatically evaluated, and a human expert inspection is recommended. Finally, characteristics such as looking away, too dark/light and mouth open are not classified accurately by the three SDKs, but a deeper analysis of the problem and the availability of training images would certainly allow to significantly improve the performance. As future work the dataset will be extended by including new samples of non-compliant images with respect to all the grayed characteristics in Table 3. It is our intention to make a new database (labeled and partitioned into training and test sets) available to the scientific community to allow the comparison with other SDKs, the improvement of existing techniques and development of new algorithms.

## Bibliography

[1] A.M. Martinez, R. Benavente, “The AR Face Database”, *CVC Technical Report #24*, June 1998.

[2] BSI-TR 03104, “Technical Guideline for production, data acquisition, quality testing and transmission for passports - Annex 1 - Quality requirements for the acquisition and transmission of the facial image data as biometric feature for electronic identification documents”, 2007.

[3] ISO International Standard ISO/IEC JTC 1/SC 37 N506, “Text of FCD 19794-5, Biometric Data Interchange Formats – Part 5: Face Image Data”, 2004. Available at http://isotc.iso.org

[4] R. Cappelli, D. Maio, D. Maltoni, J.L. Wayman and A.K. Jain, “Performance Evaluation of Fingerprint Verification Systems", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.28, no.1, pp.3-18, 2006.

[5] U. Seidel, “Experiences with facial image standard compliance- Do we need to relax ISO 19794-5 tolerances?”, *E-Passport Interoperability Test Event*, available at http://www.interoptest-berlin.de/reading_material.htm.