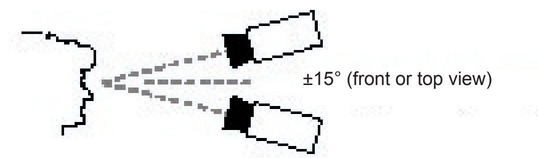
tech overview | Recognition Technology

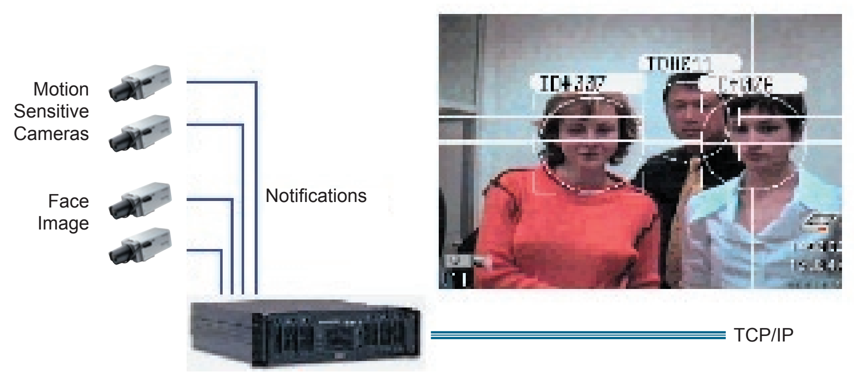
april 2005 . moscow. russia . iss research & development headquarters

INTELLIGENT SECURITY SYSTEMS

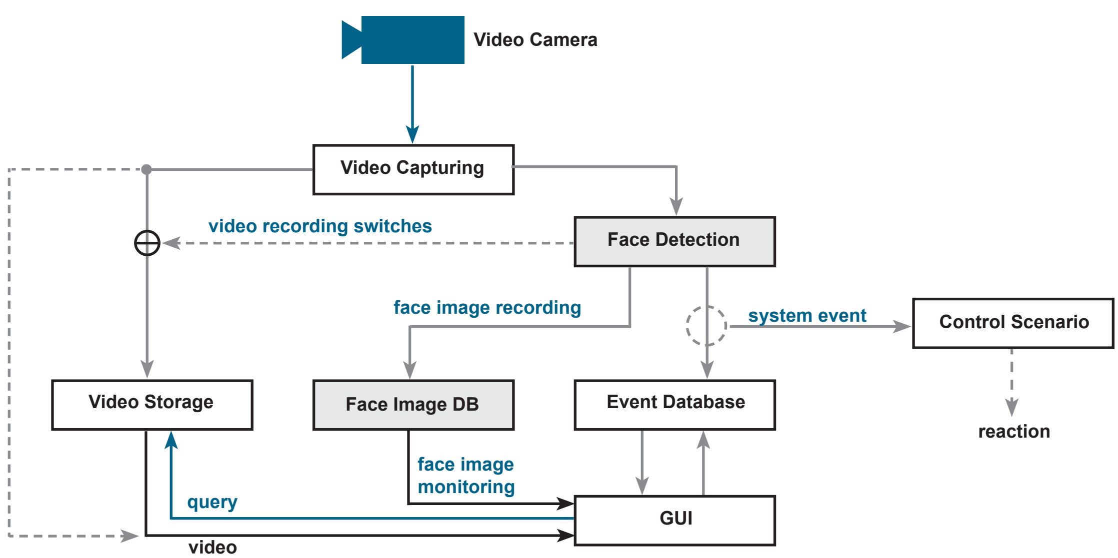
Pattern Recognition Technologies – Face Capture

ISS offers proprietary real-time algorithm of face image detection and capture. It does not require special cameras or a specific environment. Multiple faces in human stream may be detected, captured, recorded and delivered with further analysis, reporting and notification capabilities. The Face Capture is an application software for video surveillance, monitoring, law enforcement and other applications.





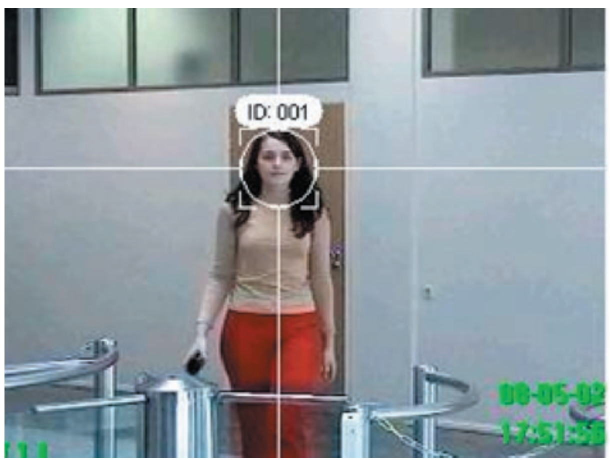
Individual facial patterns are recorded and stored in a digital photo database that can be viewed and used for different applications on-site or remotely. ISS developed several algorithms, supporting the real time processing of video data and image localization, determination of position of head and motion tracing for subsequent recognition. The algorithm might be divided into several steps.



Face localization in picture

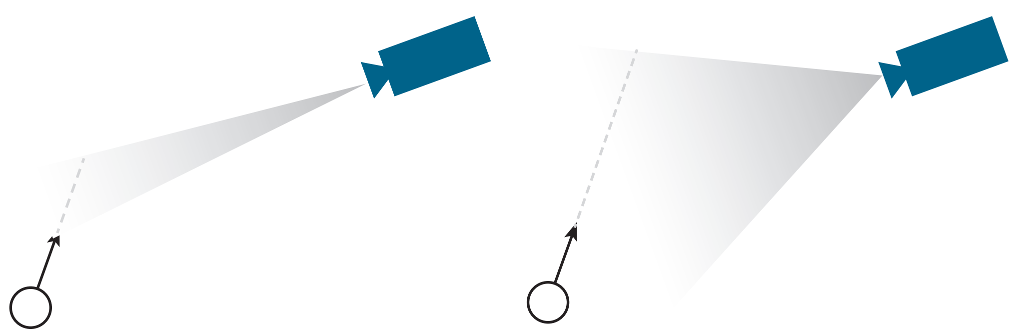
For face localization ISS developed the neural network-based classification algorithm, which scans original picture in several scales, valuates every area by key characteristics, and with set probability classifies the area as face or not. Key characteristics are selected via automatic analysis of quite large learning selection, which covers most part of probable conditions (such as changes of exterior, illumination, perspective, and so on).

The advantage of using neural networks for face detection is the feasibility of training a system to capture the complex class conditional density of face patterns. The neural network computes a face description by approximating the eigenvectors of the image's autocorrelation matrix. To detect the presence of a face in scene, the distance between an image region and the face space is computed for all location of image.



Head position determination

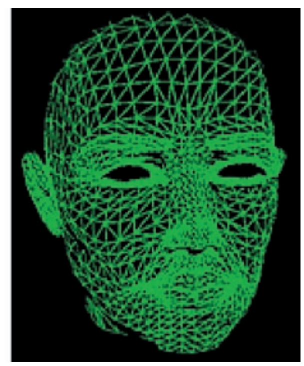
Determination of person's head position is another important step, which allows necessary corrections during further recognition. At this step the ISS-developed 3D head model is compared with the probable face image in the picture, thus allowing evaluation of such parameters, as rotation angle against X, Y, Z-axis, exact dimensions and visual displacement.



Tracing of face motion between frames

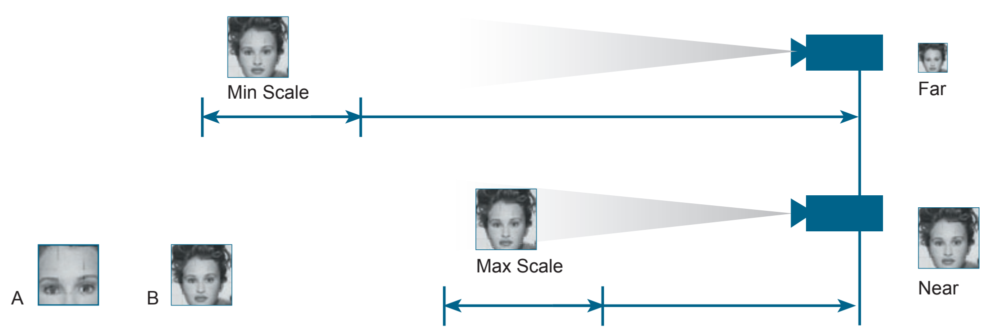
In identification of a person, moving in the viewing field of the camera, it is necessary to track movement of face between frames. Using several pictures of the same person in different perspectives the software selects best frame and stores it to database. In addition, processing of several images of the same person in different perspectives provides very high accuracy of recognition. At present ISS is developing algorithm of face comparison with database images, which will complete the range of video-based personal identification algorithms.

The Face Capture allows officials to document the movement of suspects linking individuals to a crime or significant event. An advanced search function and GUI can cut the time needed to evaluate a 24-hour time lapse observation video to less than 15 minutes. Facial images can also be exported in common formats or printed out on a standard PC printer.



The Face Capture can be used at airports, banks, casinos, public buildings, subways, factories, schools or in any other location where it makes sense to record the faces of visitors, with facilities for integration into existing CCTV systems.

The Face Capture GUI is very simple such that any operator can use all of its functions with just a minimal amount of training. The system is highly flexible, allowing images to be digitalized and recorded in either color or monochrome with a storage capacity typically exceeding 12 months of facial data recording. Face Capture screen simultaneously shows the live camera shot and the latest sequence of captured images.



Pattern Recognition Technologies – Face Recognition

This technology processes each image (8 bit grayscale) in the following way:

- Face localization: The image is analyzed to determine the position and size of one or more faces. (In all of the following steps, we assume that only a single face is found.)
- Eye localization: The positions of the eye centers in the face are determined.
- Image Quality Check: The quality of the face image is checked to see whether it is sufficient for the steps that follow.
- Normalization: The face is extracted from the image and is scaled and rotated such that an image of fixed size results, with the eye centers at fixed positions in that image.
- Preprocessing: The normalized image is preprocessed with standard techniques such as histogram equalization, intensity normalization, and others.
- Feature extraction: In the preprocessed image, features are extracted that are relevant for distinguishing one person from another.
- Construction of the reference set: During enrollment the facial features of (usually) several images of a person are extracted and combined into a reference set, also called the “biometric template”.
- Comparison: For verification, the set of extracted features is compared with the reference set of the person who the person in the image just processed is claimed to be; for identification, the feature set is compared to all stored reference sets, and the person with the largest comparison value is selected; in both cases the recognition is considered successful if the (largest) comparison value - which is interpreted as a similarity value - exceeds a certain threshold value.

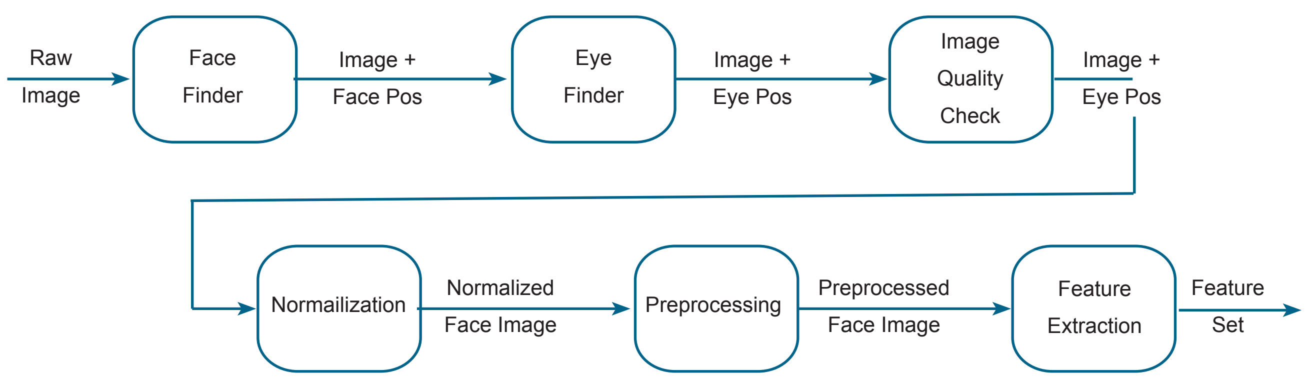


Figure 1 - Architecture : Feature Set Creation

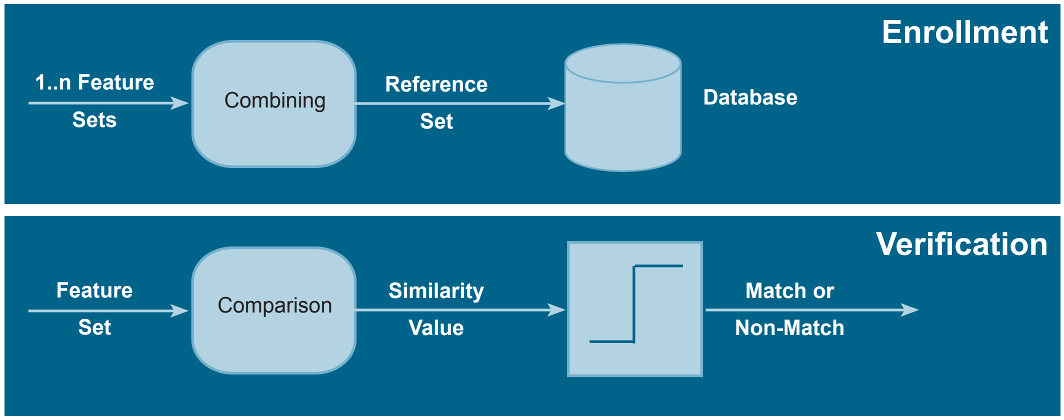


Figure 2 - Architecture : Enrollment & Verification

In addition, a “live check” is performed to ensure that the face in front of the camera is a real one and not just a picture. To this end, the changes in appearance occurring during movement of the face (rotations around the vertical axis in particular) are exploited. Due to the special 3D structure of a real face, those changes are quite different for a real face than for a photo. So when the user wants to pass the live check, he or she should briefly rotate his or her head back and forth.

In the following subsections, more details of the individual steps are given. An example image is used to illustrate the effect of each processing stage.

Figure 3 - Sample Image

Figure 4 - Eye locations found by algorithm

Face and eye localization

To locate the face, a so-called image pyramid is formed from the original image. An image pyramid is a set of copies of the original image at different scales, thus representing a set of different resolutions.

A mask is moved pixel-wise over each image in the pyramid, and at each position the image section under the mask is passed to a function that assesses the similarity of the image section to a face. If the similarity value is high enough, the presence of a face at that position and resolution is assumed. From that position and resolution, the position and size of the face in the original image can be calculated. From the position of the face, a first estimate of the eye positions can be derived. In a neighborhood around these estimated positions, a search for the exact eye positions is started. This search is very similar to the search for the face position, the main difference being that the resolution of the images in the pyramid is higher than the resolution at which the face was found before. The positions yielding the highest similarity values are taken as final estimates of the eye positions.

Image Quality Check

To be usable for the subsequent steps, the part of the image occupied by the face has to meet certain quality requirements; e.g., it should not be too noisy or blurred. The quality is measured by means of a set of functions that are applied to the image. If the quality is considered too low, the image is rejected.

Normalization and Preprocessing

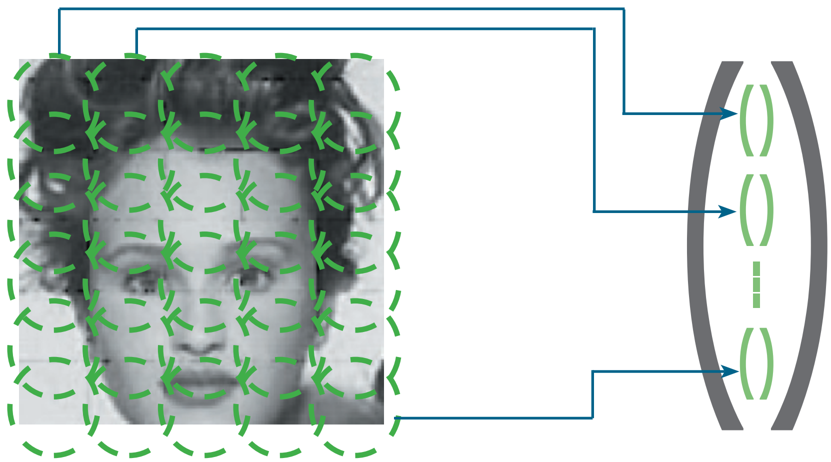
In the normalization step, the face is extracted, rotated and scaled such that the centers of the eyes lie at predefined positions. More precisely, they are positioned to lie on the same horizontal pixel row such that the midpoint of this row is aligned with the midpoint between the centers of the eyes.

The preprocessing step comprises, among other transformations, the elimination of very high and very low spatial frequencies and the normalization of contrast.

Figure 5 - After Normalization & Preprocessing

Feature Extraction, Reference Set Creation & Comparison

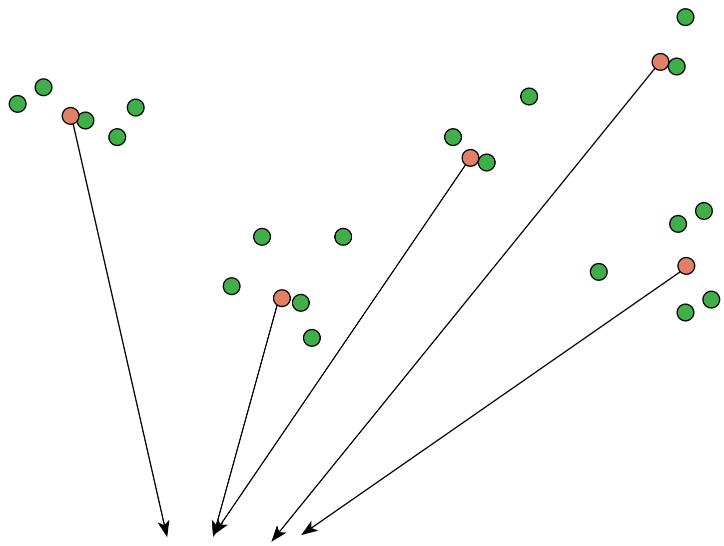
Feature extraction starts with local image transforms that are applied at fixed image locations. These transforms capture local information relevant for distinguishing people, e.g. the amplitudes at certain spatial frequencies in a local area. The results are collected in a vector.



A global transform is then applied to this vector. Using a large face-image database, the parameters of this transform were chosen to maximize the ratio of the inter-person variance to the intra-person variance in the space of the transformed vectors; i.e., the distances between vectors corresponding to images of different persons should be large compared to distances between vectors corresponding to images of the same person. The result of this transformation is another vector that represents the feature set of the processed face image.

Figure 7 – Global transformation, yielding the feature set of the face image

For the creation of the reference set, several images are usually taken of each person during enrollment in order to better cover the range of possible appearances of that person's face. The reference set generated for a person consists of up to five feature sets, which are the centers of clusters obtained through a clustering process on the feature sets created from those images.



Reference Set

Figure 8 – Combining cluster centers (red) into a reference set.
(Green dots are feature sets created from images)

The function that is used to compare a feature set with a reference set is simple and can be computed very fast. It makes identification a matter of seconds, even if a million reference sets have to be compared.

Annex 1 :
Facial Recognition Metrics

Introduction

Independent studies were conducted to determine the accuracy under multiple conditions for Face Capture & Recognition, including the ISS FCR subsystem. Testing was condutec jointly by several United States agencies, including the Defense Advanced Research Projects Agency, Department of Defense, and the National Institute of Standards and Technology. Evaluation consisted of the high computational intensity (HCInt) test and medium computational intensity (MCInt) test. The HCInt was designed to evaluate the performance of state-of-the-art systems on extremely challenging real-world problems. The MCInt was designed to provide an understanding of a participant's capability to perform face recognition tasks with several different formats of imagery (still and video) under varying conditions.

Analysis of the FCR role

The three primary face recognition tasks are :

Verification: "Am who I say I am?"
Identification: "Who am I?"
Watch list: "Are you looking for me?"

In a verification task, a person presents their biometric and an identity claim to a face recognition system. The system then compares the presented biometric with a stored biometric of the claimed identity. Based on the results of comparing the new and stored biometric, the system either accepts or rejects the claim. From an evaluation point of view, there are two types of system users. The first is a legitimate user. The second is a person who attempts to impersonate a legitimate user. Verification performance is characterized by two performance statistics. The two statistics characterize the success rate of the two types of users. The first is the rate at which legitimate users are granted access. This is the verification rate. The second is the rate at which imposters are granted access. This is the false accept rate. The ideal system would have a verification rate of 100% and a false accept rate of 0%. Unfortunately, such a system does not exist. In real-world systems, there is a trade-off between verification and false accept rates.

It is critically important to consider the false accept rates and verification rates together in order to understand the performance capabilities of a face recognition system. It is easy to build a system that always grants access to a subject. This system will have a 100% verification rate since access will always be granted in response to a legitimate user's request. Conversely, this system will also have a 100% false accept rate because it also grants access to imposters. The best system is one that balances the verification rate with a false accept rate in a manner consistent with operational needs. Examples of this trade-off can be viewed below.

Verification rates at different false accept rates for the ISS subsystem was as follows :

False Alarm Rate (%)	Verification Rate
0.1	82%
1.0	90%
10.0	96%

Environmental Conditions

Recognition from frontal images is an important capability of face recognition systems. The tests examined recognition performance using frontal images taken under varied conditions. The analysis included the results using the HCInt and MCInt data sets. In all cases, the database consists of images taken indoors under good lighting conditions.

Match original image with ...	Verification Rate
Same person taken indoors on the same day with the same illumination.	95%
Same person taken indoors on the same day with different illumination.	95%
Same person taken indoors on different days with the same illumination.	89%
Same person taken indoors on different days with different illumination.	90%
Same person taken indoors with an image taken outdoors.	54%

Watch List Performance

In the watch list task, a face recognition system must first detect if an individual is, or is not, on the watch list. If the individual is on the watch list, the system must then correctly identify the individual. The statistic for correctly detecting and identifying an individual on the watch list is called the detection and identification rate. In some instances, the system may incorrectly alarm on an individual that is not on the watch list. The rate at which an individual that is not on the watch list is incorrectly alarmed is called the false alarm rate. Typically, the watch list task is more difficult than the identification or verification tasks alone.

The ISS FCR detection and identification rate is 77% using a watch list of 25 people. Increasing the size watch list to 3,000 people, decreases the detection and identification rate to 56%. In general, systems achieve better performance for a smaller watch list. If the impetus of the watch list application is to detect and identify the "most wanted" individuals, the watch list should be kept as small as possible. Increasing the size of the watch list reduces the probability that an individual on the watch list is correctly detected and identified when presented to the system.

Watch List Size	Verification Rate
25	77%
50	76%
100	75%
300	70%
400	65%
800	63%
1600	59%
3000	56%

Affects of Age, Sex, and Time Delay

The test examined how sex, age, and time-delay affect recognition. Males are easier to recognize than females. Identification rates for males were 6% to 9% points higher than that of females. Older people are easier to identify than younger people. For example, 18 to 22 year olds the average identification rate was 62%, and for 38 to 42 year olds was 74%. For every ten years increase in age, on average performance increases approximately 5% points.

Conclusion

The ISS FCR proved in independent tests to be the top performer in alost very sub-test. Even with this impressive performance, certain basic axioms can be derived for the usage of Face Capture and Recognition systems :

- Given reasonable controlled indoor lighting, the ISS FCR achieves 90% verification at a 1% false accept rate.
- The use of morphable models can significantly improve non-frontal face recognition.
- Watch list performance decreases as a function of size - performance using smaller watch lists is better than performance using larger watch lists.
- In face recognition applications, accommodations should be made for demographic information since characteristics such as age or sex can significantly affect performance.
- Males are easier to recognize than females.
- Younger people are harder to recognize than older people.
- Indoor FCR performance is substantially better than outdoor.
- Face recognition performance decreases approximately linearly with elapsed time database and new images.
- Face recognition systems do not appear to be sensitive to normal indoor lighting changes.
- Three-dimensional morphable models substantially improve the ability to recognize non-frontal faces.

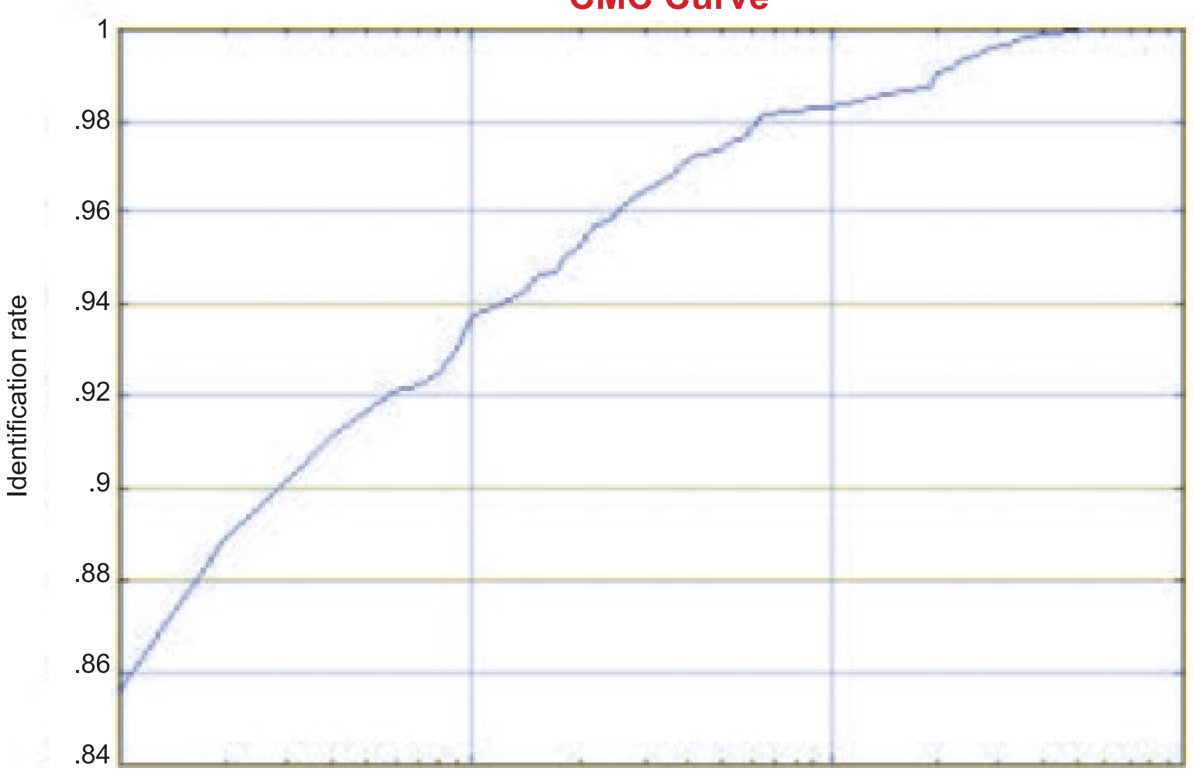
Annex 2 :

A8T4 Algorithm Performance

FERET “Duplicate I” Identification Test

This test was one of the tests used to measure the performance of various face recognition algorithms, both in the FERET tests themselves and in the Facial Recognition Vendor Test 2000. The «Duplicate I» test involves the following subsets of the FERET database: a gallery including 1196 images of 1183 persons (for a few persons, two images are enrolled in the gallery) and a probe set of 722 images of 242 persons. The main characteristic of this test is that the probe images and the corresponding gallery image(s) were taken in different

CMC Curve



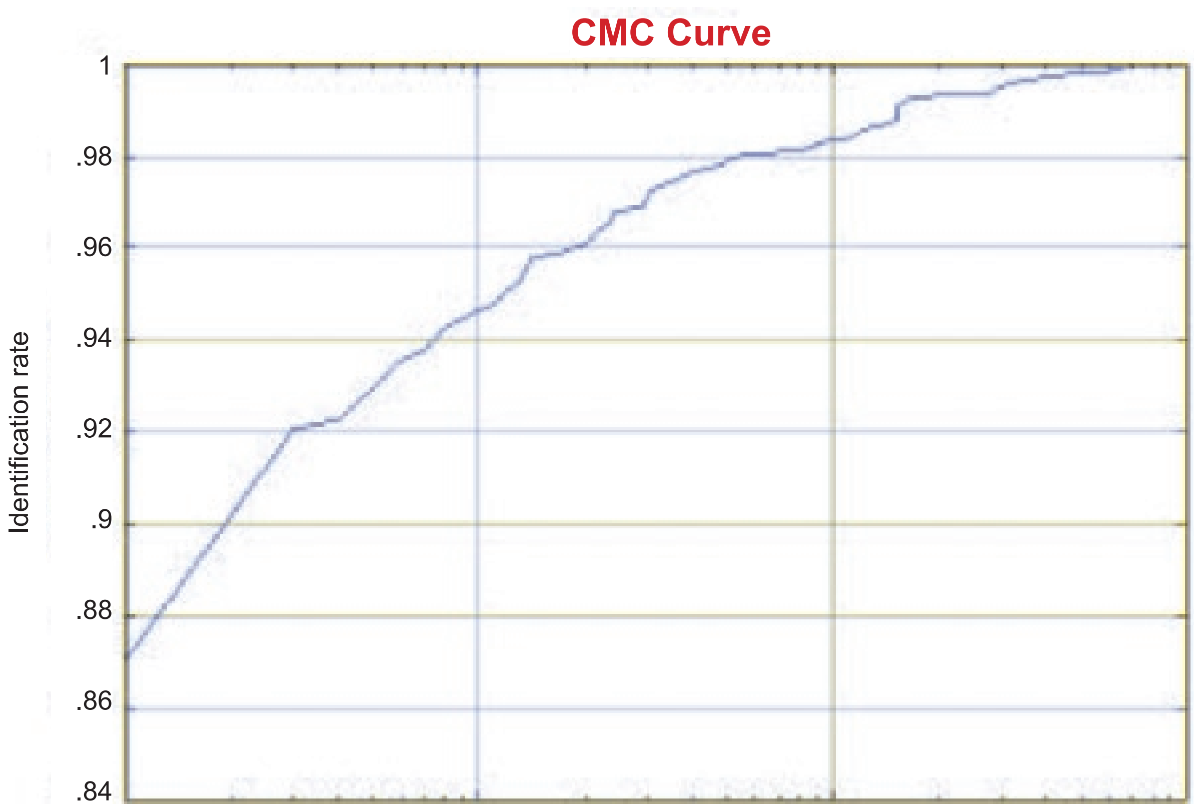
For more information on this test, see the articles listed below.

The research in this paper uses the FERET database of facial images collected under the FERET program [6].

References:

1. «The Facial Recognition Technology (FERET) Database», see Annex I
2. P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, «The FERET evaluation methodology for face recognition algorithms», IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, October 2000.
3. P. J. Phillips, P. J. Rauss, and S. Z. Der, «FERET (Face Recognition Technology) Recognition Algorithm Development and Test Results», October 1996. Army Research Lab technical report 995.
4. P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, «The FERET Evaluation Methodology for Face Recognition Algorithms».
5. S. Rizvi, P. J. Phillips and H. Moon, «The FERET Verification Testing Protocol for Face Recognition Algorithms».
6. P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing J, Vol. 16, No. 5, pp 295-306, 1998.

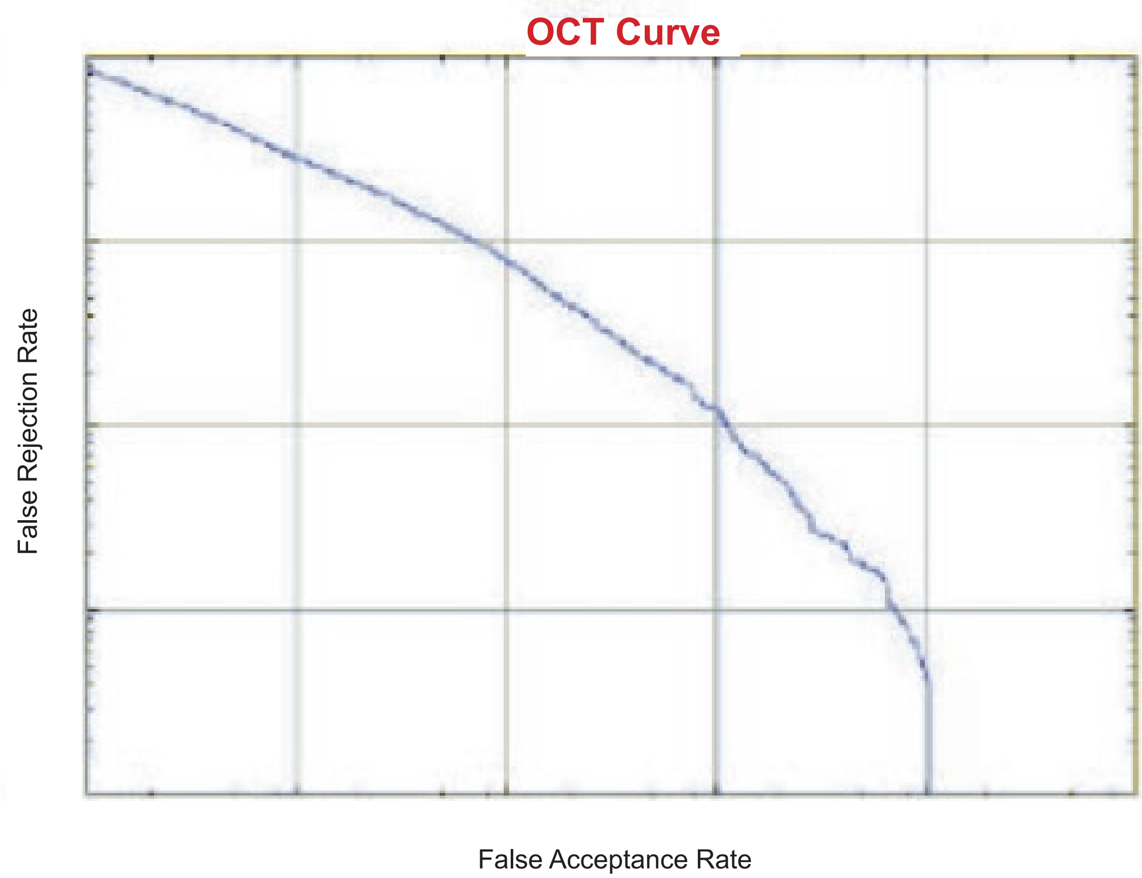
sessions of the FERET image collection process. All images show the face in a frontal or nearly frontal pose. Here is a quote from the FERET DB documentation on this test: «The Duplicate I probe set holds 722 images whose matches were taken between 0 and 1031 days after the match. The median is 72 days and the mean is 251 days.» Between gallery and probe set, the images of a person display differences in various aspects such as illumination, facial expression, wearing of glasses, pose (slightly), age (up to 34 months), or resolution of the face within the image. The demographic characteristics of the image subsets seem not to be disclosed. Visual inspection of the images revealed that they cover an age range of at least between 20 and 60 years and that various ethnicities occur in the following proportions: 9% Black (i.e. Afro-American and native Australian), 12% Asian (i.e. East and Southeast Asian) and 79% White (i.e. all other ethnicities).



This test was run on a collection of typical passport images (scanned passport photographs) of 1000 persons (2000 images with 2 images per person). The collection is split evenly in a gallery and a probe set such that each set contains exactly one image per person. In almost all cases, the two images of a person display differences in one or more of the following aspects: illumination, facial expression, wearing of glasses, pose (slightly), age (up to 2 years), resolution of face in image The image collection covers an age range from below 10 years (toddlers) up to over 70 years as well as various ethnicities such as 2.2% Black people (i.e. Afro-American and native Australian), 7.7% Asian (i.e. East and Southeast Asian), and 90.1% White (i.e. all other ethnicities, here mainly Caucasian and Indian). 61% of the persons are female.

Verification test on images from a access control scenario

This test uses enrollment and verification images collected with an actual installation for physical access control. The gallery contains 6324 enrollment images of 527 different persons, 12 images per person. For each person in the gallery, a single biometric template is created from the 12 images. The probe set contains 47008 verification images, both of the 527 persons in the gallery and of 2800 further persons not in the gallery (i.e. true impostors). The gallery and probe images of the same person differ in one or more of the following aspects: illumination, facial expression, wearing of glasses, pose, age (up to 1 year), resolution of face in image. The image collection covers an age range from 20 to 60 years as well as various ethnicities such as 0.1% Black (i.e. Afro-American and native Australian), 10.8% Asian (i.e. East and Southeast Asian) and 89.2% White (i.e. all other ethnicities).



Annex 3 :
The Facial Recognition Technology (FERET) Database

Introduction

The FERET program ran from 1993 through 1997. Sponsored by the Department of Defense's Counter-drug Technology Development Program through the Defense Advanced Research Products Agency (DARPA), its primary mission was to develop automatic face recognition capabilities that could be employed to assist security, intelligence and law enforcement personnel in the performance of their duties.

The FERET image corpus was assembled to support government monitored testing and evaluation of face recognition algorithms using standardized tests and procedures. The final corpus, presented here, consists of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles.

Nomenclature

The naming convention for the FERET imagery in this distribution is of the form nnnnnxxfffq_yymmdd.ext where:

1. nnnnn is a five digit integer that uniquely identifies the subject
2. xx is a two lowercase character string that indicates the kind of imagery:

Two letter code	Pose Angle (degrees)	Description	Number in Database	Number of Subjects
Fa	0 = frontal	Regular facial expression	1762	1010
Fb	0	Alternative facial expression	1518	1009
ba	0	Frontal "b" series	200	200
bj	0	Alternative expression to ba	200	200
bk	0	Different illumination to ba	200	200
bb	+60	Subject faces to his left which is the photographer's right	200	200
bc	+40		200	200
bd	+25		200	200
be	+15		200	200

bf	-15	Subject faces to his right which is the photographer's left	200	200
bg	-25		200	200
Bh	-40		200	200
bi	-60		200	200
ql	-22.5	Quarter left and right	763	508
qr	+22.5		763	508
hl	-67.5	Half left and right	1246	904
hr	+67.5		1298	939
pl	-90	Profile left and right	1318	974
pr	+90		1342	980
Ra	+45	Random images. See note below. Positive angles indicate subject faces to photographer's right	322	264
Rb	+10		322	264
Rc	-10		613	429
Rd	-45		292	238
Re	-80		292	238

Notes:

1. fa indicates a regular frontal image
2. fb indicates an alternative frontal image, taken seconds after the corresponding fa
3. ba is a frontal images which is entirely analogous to the fa series
4. bj is an alternative frontal image, corresponding to a ba image, and analogous to the fb image
5. bk is also a frontal image corresponding to ba, but taken under different lighting
6. bb through bi is a series of images taken with the express intention of investigating pose angle effects (see below). Specfically, bf - bi are symmetric analogues of bb - be.
7. ra through re are «random» orientations. Their precise angle is unknown. It appears that the pose angles are random but consistent. The pose angles in the table were derived by manual measurement of inter-eye distances in the image, and in their corresponding frontal image.
8. fff is a set of three binary (zero or one) single character flags. In order these denote:
    a. Indicates whether the image is releasable for publication. The flag has fallen into disuse t.
    b. Image is histogram adjusted if this flag is 1
    c. Indicates whether the image was captured using ASA 200 or 400 film, 0 implies 200.
9. q is a modifier that is not always present. When it is, the meanings are as follows:
    a. Glasses worn. Note that this flag is a sufficient condition only, images of subjects wearing glasses do not necessarily carry this flag. Some

retroactive re-truthing of such images to fix this problem is warranted. See also „c“ below.

b. Duplicate with different hair length.
c. Glasses worn and different hair length
d. Electronically scaled (resized) and histogram adjusted.
e. Clothing has been electronically retouched.
f. Image brightness has been reduced by 40%
g. Image brightness has been reduced by 80%
h. Image size has been reduced by 10%, with white border replacement
i. Image size has been reduced by 20%, with white border replacement
j. Image size has been reduced by 30%, with white border replacement

Note that the modifications d through j are the result of applying various off-line operations to real images in the database; the “parent” image is that image without the “q” modifier present at all.

10. The three fields are the date that the picture was taken in year, month, day format.
11. The filename extension is .tif.

Standard Testing Subsets

Empirical testing of pattern recognition algorithms is predicated on standardized data sets. Previously publications using FERET images have reported performance on the images described below. The concepts of gallery and probe sets apply: Each probe image1[1] is matched against those in a gallery, and the ranked matches can be analyzed to produce recognition performance measures such a cumulative match score for identification, and receiver operating characteristic for verification applications.

Users are strongly encouraged to utilize the specific gallery and probe sets included in this FERET distribution. The reports detailed in the previous footnotes quote performance measures for contemporary algorithms on exactly these standardized test sets. These partitions of the corpus defined by the *.names files in the directories underneath the to-level folder partitions/.

The standard subsets have two distinct provenances. The first was employed originally in the FERET tests themselves, and were used again in the Face Recognition Vendor Test 2000 (FRVT2000, see footnote Erreur ! Signet non défini.). The second series were designed to evaluate the sensitivity of algorithms to head direction (pose angle) as part of the FRVT 2000 tests.

[1] In the FERET protocol a probe set contains one or more images from a set of individuals. Each person will have exactly one match in the gallery. The gallery may contain images from other individuals who are not in the probe population.

In the tables below the parenthesized numbers indicate the number of images in the respective sets.

FERET Tests September 1996

These tests employed frontal images gathered between 1993 and 1996. The image sets are in the partitions/by_previously_reported/feret/ directory.

Evaluation Task	*Recognized Names*	*Gallery (1196)*	*Probe Set*
Aging of subjects	Duplicate I or T1	gallery.names	probe_dup_1_*.names (722)
Aging of subjects	Duplicate II or T2	gallery.names	probe_dup_2_*.names (234)
Facial Expression	fafb	gallery.names	probe_fafb_*.names (1195)
Illumination	fafc	gallery.names	probe_fafc_*.names (194)

All the above tests used a single gallery containing 1196 images1[2]. The Duplicate I probe images were obtained anywhere between one minute and 1031 days after2[3] their respective gallery matches. The harder Duplicate II probe images re a strict subset of the Duplicate I images; they are those taken only at least 18 months3[4] after their gallery entries. For assessment of the effect of facial expression two probe sets have been used. There is usually only a few seconds between the capture of the gallery-probe pairs.

[2] The tables in the published FERET reports in the previous footnote indicate that the gallery for the Duplicate II test held 864 images. This is erroneous; the correct number is 1196.

[3] The Duplicate I probe set holds 722 images whose matches were taken between 0 and 1031 days after the match. The median is 72 days and the mean is 251 days.

[4] The Duplicate II probe set contains 234 images from subjects whose gallery match was taken between 540 and 1031 days beforehand. The median is 569 and the mean is 627 days. Thus the Duplicate II probe images were taken at least 18 months after their gallery match.

[5] The FRVT 2000 tests evaluated leading commercial face recognition vendors. They were sponsored by the DoD Counterdrug Technology Development Program Office, the National Institute of Justice, and the Defense Advanced Research Projects Agency, and were administered between May and June 2000.

[6] To
direction other than toward the camera with eyes a horizontal plane. The other degrees of freedom in which the head rotates about horizontal axes: pitch (looking up or down) and roll (tipping head on it's side) are not represented in the FERET database.

FRVT 2000 Tests May 2000

The following four tests were conduced as part of the Facial Recognition Vendor Test 20004[5] and used a single frontal gallery, and four non-frontal probe sets taken at increasing azimuthal5[6] angles. The intention of these tests was to quantify the effect of non-frontal image capture on recognition performance. The images are named in the partitions/by_previously_reported/frvt2000/ directory.

Evaluation Task	Recognized Names	Gallery (1196)	Probe Set
Aging of subjects	Duplicate I or T1	gallery.names	probe_dup_1_*.names (722)
Aging of subjects	Duplicate II or T2	gallery.names	probe_dup_2_*.names (234)
Facial Expression	fafb	gallery.names	probe_fafb_*.names (1195)
Illumination	fafc	gallery.names	probe_fafc_*.names (194)

The FRVT 2000 evaluation repeated the tests using the FERET Duplicate I and II imagery, used the working names T1 and T2 The table and the accompanying figure show the relationships between the sets defined in the preceeding sections.

FERET Subset 1		Feret Subset 2	Cardinality
FERET Gallery	$\cap$	Training CD	270
Duplicate I	$\cap$	Training CD	184
Duplicate II	$\cap$	Training CD	0
Probe set fafb	$\cap$	Training CD	270
Probe set fafc	$\cap$	Training CD	0
Probe set fafb	$\cap$	Probe set fafc	0
Probe set fafb	$\cap$	Duplicate I	0
Duplicate II	$\subset$	Duplicate I	234

notes