FaceVACS Algorithms White Paper

Feb 7 2012

1. Introduction

All products using FaceVACS technology employ the same algorithms to process facial data. This text describes the key features of these algorithms and the processing sequence applied to data samples provided to the facial recognition engines. ’Data samples’ within this context describes facial data. While until recently facial data has mainly been presented as intensity images, upcoming 3D sensors allow to acquire shape information, too. Starting with FaceVACS-SDK 4.0 Cognitec provides algorithms utilizing shape information, too. For this reason, the more general term ’data sample’ will be used to describe facial information comprising mandatory intensity image (shortly: image) and optional shape data.

2. Intensity Image Processing

Intensity Image Processing Sequence

Images are processed as as follows:

- Face localization: The image is analyzed to determine the position and size of one or more faces. (In all of the following steps it is assumed that only one face is found.)
- Eye localization: The positions of the centers of the eye within the face are determined.
- Normalization: The face is extracted from the image and is scaled and rotated in such a way that the result is an image of fixed size, with the centers of the eye at fixed positions within 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 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 recognition is considered successful if the (largest) score value exceeds a certain threshold value.

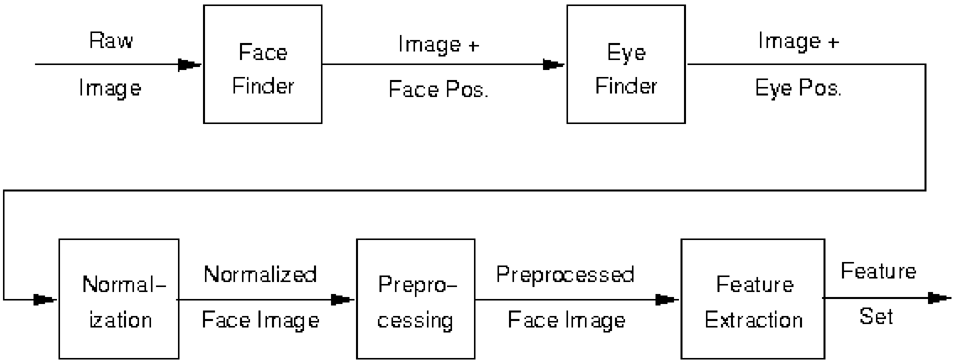


Figure 1. FaceVACS architecture: Feature set creation

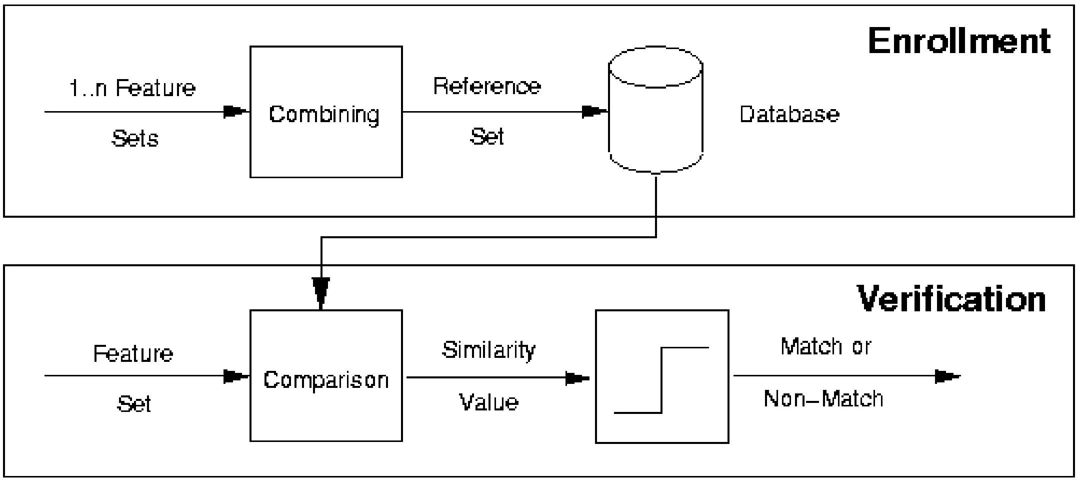


Figure 2. FaceVACS architecture: enrollment and verification

In addition, FaceVACS has a "live check" facility to ensure that the face in front of the camera is a real one and not just a photograph. 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 very different for a real face compared to the changes in a photo. So if the user wants to pass the live check, he or she should briefly rotate their head back and forth. Another way to provide the 3D structure information is to utilize 2 or more cameras providing different views at the face.

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.

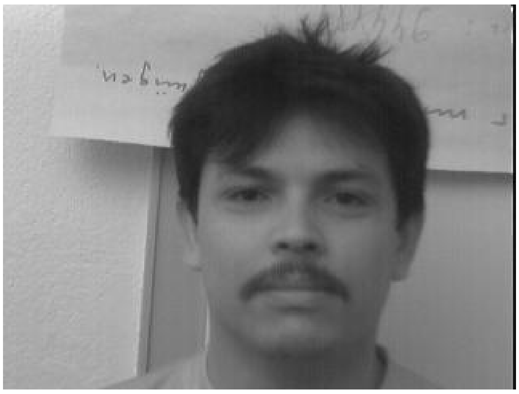


Figure 3. Example image

2.1. 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 from one pixel to the next 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 score 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. Within this estimated positions and its neighborhood, 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 previously found. The positions yielding the highest score values are taken as final estimates of the eye positions.

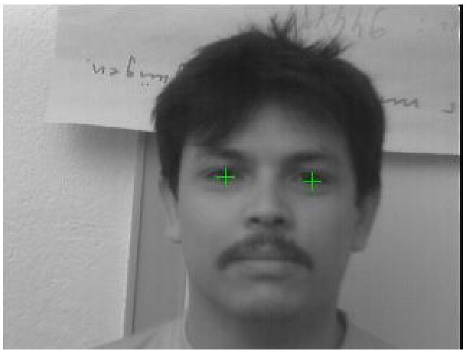


Figure 4. Eye locations found by the algorithm

2.2. Normalization and Preprocessing

In the normalization step, the face is extracted, rotated and scaled in a way that the centers of the eyes lie at predefined positions. More precisely, they are positioned to lie on the same horizontal pixel row so that the mid-point of this row is aligned with the mid-point between the centers of the eyes.

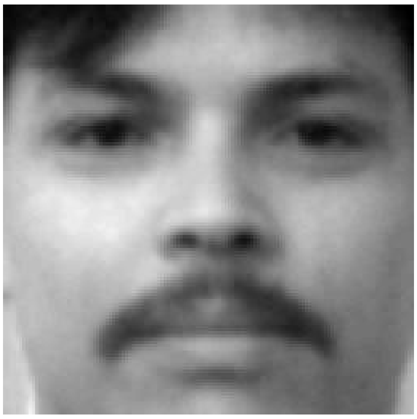


Figure 5. After normalization

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

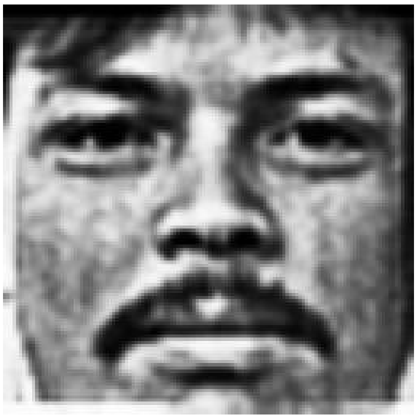


Figure 6. After preprocessing

2.3. Feature Extraction and Reference Set Creation and 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.

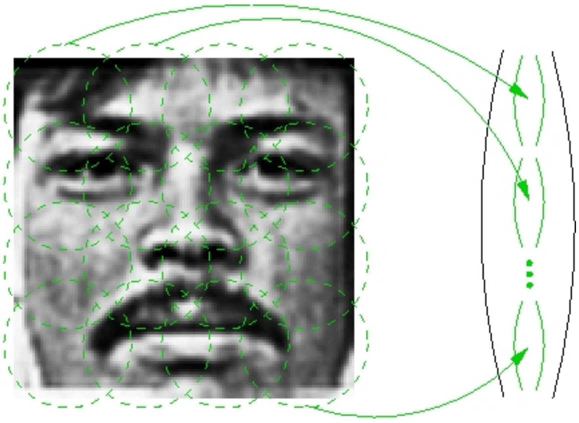


Figure 7. Extracting local features

A global transform is then applied to this vector. Using a large face-image database, the parameters of this transform are 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 transform is another vector that represents the feature set of the processed face image.

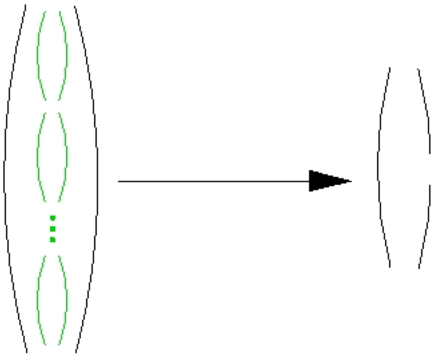


Figure 8. Global transform, yielding the feature set of the face image

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 in the feature sets created from those images.

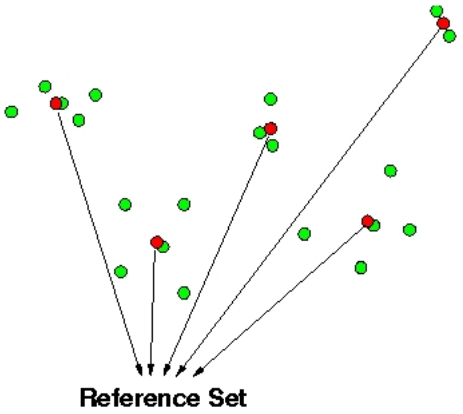


Figure 9. Combining cluster centers (red) into a reference set. (Green dots are feature sets created from images.)

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.

3. Combined Shape and Intensity Image Processing

Data Sample Processing Sequence

Regarding the intensity image part, data samples containing both intensity image and shape information are processed in the same way as described above. The eye locations obtained in this stage are important for subsequent shape data processing.

The data sample comparison is based on a fusion step which merges the results of the intensity and the shape recognition substeps.

The entire processing sequence for data samples runs as follows:

- Intensity image processing as described above
- Shape data preprocessing: Depending on the sensor type and the acquisition conditions. shape data as delivered by 3D sensors is frequently noisy and incomplete. Before shape data can be passed to pattern recognition steps, it has to be preprocessed and smoothed in some way.
- Normalization: Similar to what is done with intensity images, the face shapes are scaled to some standard size and aligned in space in order to minimize variations due to translation, rotation and scaling.
- Feature extraction: From the preprocessed and normalized shape, facial shape features are extracted that are relevant to describe similarities or differences between faces.
- Construction of the sample reference set: During enrollment both the intensity and the shape based facial features of one or more samples of a person are extracted and combined into a reference set, also called the "biometric template".
- Comparison: To compare a feature set with a reference set, a score is determined considering both their intensity and shape subfeatures.

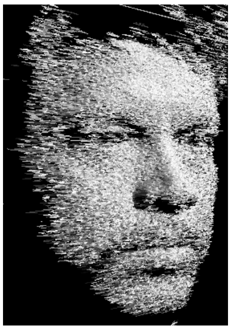
  Score computation for reference sets from samples

  1. Subscore computation from intensity based features
  2. Subscore computation from shape based features
  3. Final score computation based on a fusion algorithm

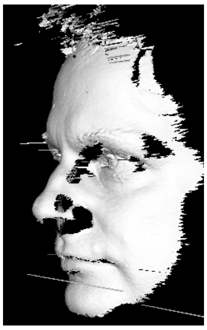
3.1. Shape Data Preprocessing

Data as provided by 3D sensors usually contains noise, data gaps and outliers, i.e. small groups of vertex positions far distant from the face shape. Also, depending on the 3D sensor principle, there can be even large parts of the face shape missing if that part of the face is occluded.

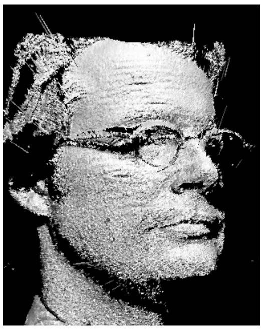
Figures below show examples for all of these flaws.



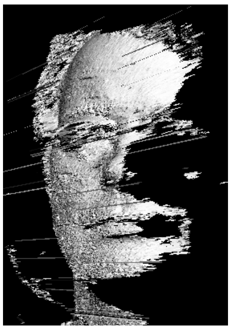
Noisy shape data from sensor



Sensor image, while smoother than the left one, contains data gaps and outliers



Non-frontal views result in occlusions and large holes in the face shape. This view of the face shape hides the gaps...



but when rotating the shape, the missing shape portions become obvious.

Of all the manifold algorithm qualified to cope with these problems only those whose time consumption is compatible with face recognition in real-world scenarios can be employed.

The main steps required to obtain shape data suitable for shape featuere extraction out of the original sensor data are:

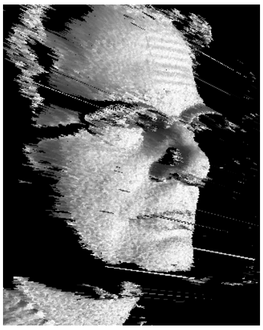
1. Outlier Removal
2. Noise Reduction
3. Gap Filling

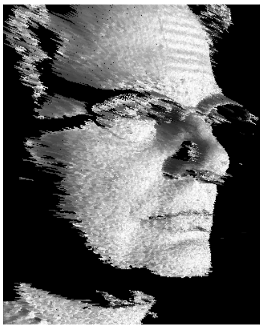
Outlier Removal

Outlier removal is the first step of the preprocessing sequence, since outliers in sensor data can heavily disturb subsequent smoothing operations on shape data.

The problem with outlier removal is to detect what is an ′outlier′ and what is ′normal′ data. Since sensor data can contain gaps and leaps, a naive definition like ′an outlier is what is not a smooth continuation of the face surface′ will fail in many cases. One approach to make this distinction is to compute local statistics of the face surface and to eliminate all vertices which are too distant from the local average.

Example: Outlier removal based on local statistics



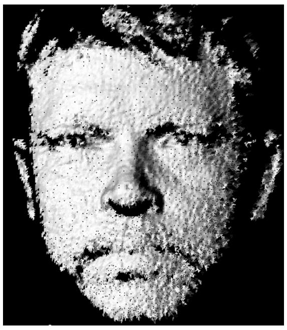


Noise Reduction

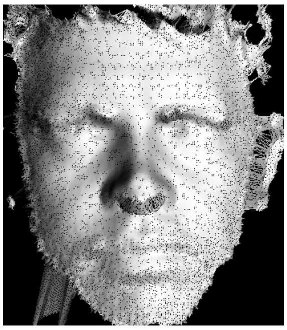
Generalisations of well-known 2D image processing operations like rank order and mean filtering to shape data often yield satisfactory results.

In addition to data smoothing, rank order filters also contribute to outlier removal to some extent.

See below for some examples demonstrating shape noise reduction by median and mean filtering



Original noisy data

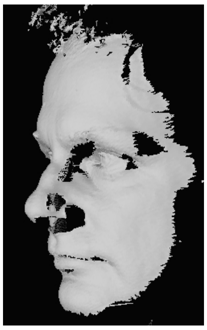


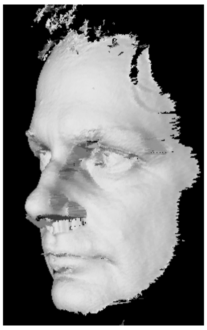
Result after combined median and mean filtering (Note: The mesh-like structure in these images is a rendering artifact)

Gap Filling

Missing data can be reconstructed e.g. based on local surface approximations. While reconstruction of nearly plane surface patches is mostly appropriate, applicability of such methods in regions with high curvature is limited.

Gap Filling Example





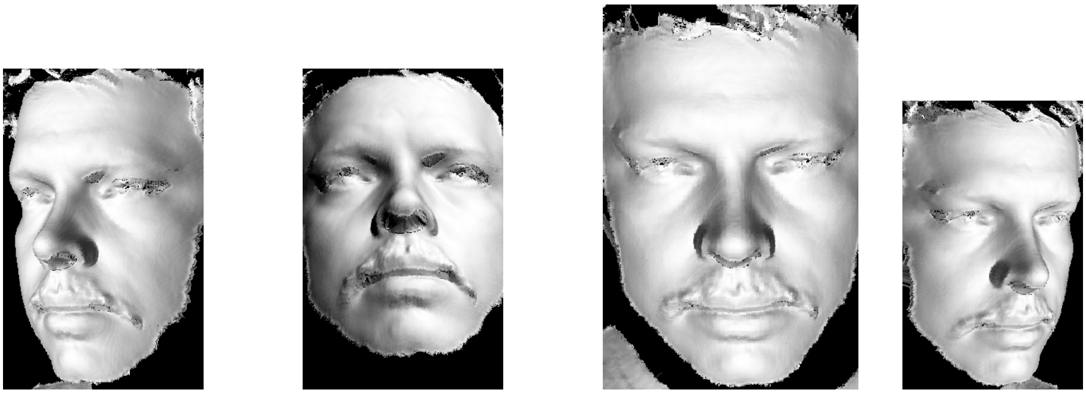
3.2. Normalization

Identity information contained in the facial shape is 'intrinsic' to this shape, i.e. is not affected by translating and rotating the shape.

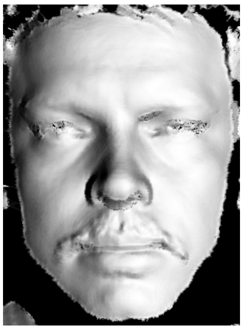
On the other hand, shape data as delivered by the sensor can have an arbitrary orientation in space. To eliminate differences between shape data sets merely due to their spatial orientation, a normalization step is applied after preprocessing where the faces are centered and aligned to frontal view.

Normalization Step Example

Different views of a face retrieved from sensor:



Standard view after normalization:



3.3. Feature Extraction

Feature extraction from shape data is a process similar to that applied to intensity data. On a set of fixed spatial locations defined relative to the eye positions, shape descriptors are retrieved, which are collected into a vector.

As with intensity image data, this vector is transformed by a global transform into a representation which optimally discriminates face shapes of different persons.

3.4. Fusion

The fusion of the intensity image and shape image processing is performed at score level, that is, the score obtained from comparing intensity feature sets is combined with the score obtained from comparing shape feature sets, resulting in a single score. The fusion function takes into account the different degrees of reliability with which an intensity or shape score reflects the probability of the two respective images showing the same person.