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
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.

**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.

**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.
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.

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.