

## Method and System for Facial Recognition

The present invention relates to a method and a system for facial recognition.

5 US 2014/0241574 A1 discloses a method and an apparatus for the tracking and recognition of faces. Persons are identified through recognition of facial attributes in selected regions of the face and comparing the facial attributes to facial data stored in a database of known faces.

10 US 8,380,711 B2 describes a method and a system for determining a hierarchical ranking of facial attributes. Facial regions are estimated from face image data and in these facial regions attributes and/or features are determined. By vectorizing these attributes and features a ranking graph for facial recognition is constructed. The ranking graph represents a hierarchical ranking of the facial attributes. Thus, a person can be identified by their facial attributes with more efficiency.

15 US 2013/0129210 A1 discloses a recommendation system and a recommendation method based on the recognition of a face and style of a person. With the face recognition gender and age of the person are determined. The style recognition includes the recognition of color and pattern of the clothing of the person combined with information about the season, weather and  
20 time. The information of the face and style recognition is then used to generate a style recommendation for the person, concerning hair, make-up, products for the outfit, and the style in general.

25 US 7,236,615 B2 describes a method for face detection and pose estimation with energy-based models. The method enables a multi-view detector, which is able to detect faces in a variety of poses. Hereby variations in skin color, eye glasses, facial hair, lighting, scale and facial expressions and other facial attributes or face features, respectively, are effectively restrained.

30 US 2009/0087100 A1 discloses an apparatus for calculating the top of head position of a person in an image. This is done by using a high frequency analysis of the image to find areas of the person, which are provided with hair. By using this method, faces of persons in an image are

found and used as reference points in order to address issues with compositional balance in photograph editing.

5 CN 103679151 A describes a method for face clustering in an image or several images. The method is performed by transforming an RGB-image into a grey-scale-image for efficiency purposes and extracting Gabor and/or LBP (local binary patterns) characteristics from the grey-scale image. The images, which belong to one person, are clustered. Other attributes such as the background, illumination, different facial expressions, body postures, hair and hairstyles and head ornaments are effectively restrained.

10 US 2006/0140455 A1 pertains to a method for image recognition. The image recognition is based on face and non-face classifiers. The image sorting/retrieval process is described in the paragraphs. In this process the user may make certain selections of the classifiers to be used in this search or the search criteria may be predetermined through the configuration file and  
15 the search image step may thus be automated. After initiating the main retrieval process distances are calculated in the current classifier space between the feature vectors, the statistical mean and standard deviation is determined, and the determined distances are normalized using the mean and standard deviations. These normalized starter sets may now be combined in a decision fusion process, which generates a ranked output list of images. These may then be  
20 displayed by a UI module.

US 2013/0121584 A1 relates to a system and method for using contextual features to improve face recognition in digital images. The face recognition system of this method may determine both visual and non-visual features for each person for which a face is detected in a set of  
25 digital images. Visual features may be a variety of facial features that represent the characteristics of a face and/or contextual features which may be non-facial features that represent the characteristics of a person in a digital image. Non-visual contextual features may be determined from data that is associated with a digital image in which a face appears. Examples of non-visual features may include the time stamp, location (e.g. geotag), name and/or tags associated  
30 with the digital image. A face comparator may use a variety of different combinations of distance values to calculate a combined distance between a pair of faces. This variety comprises face only, clothes only, face plus clothes, face plus gender, face plus clothes plus gender, etc..

CN 100568262 C discloses a multi-camera human face recognition detection based on information fusion.

5 US 7,236,615 B2 describes a human face detection independently of their poses. Additionally, a pose estimation is disclosed.

C.P. Papageorgion, M. Oren, and T. Poggio; A general framework for object detection; Sixth International Conference on Computer Vision, pages 555-562, 1998 is one of the first publications in which Haar wavelets are described for real-time object detection.

10

From Paul Viola and Michael Jones; Rapid object detection using a boosted cascade of simple features, Mitsubishi Electric Research Laboratories, Inc., 2004 (TR-2004-043), Cambridge, Massachusetts, USA (accepted Conference on Computer Vision and Pattern Recognition, 2001), and Paul Viola and Michael Jones; Robust real-time object detection; International Journal of Computer Vision, 57(2): 137-154, 2002, a method for automatically recognizing faces in images is known, wherein Haar wavelets are used for detecting Haar-like features. This method uses a so-called "Integral Image", which is an intermediate representation of a certain image, wherein the sum of all pixels above and to the left with respect to the pixel, plus its own value, is assigned to each pixel of the integral image. By using these grid point assigned sums the sum of the pixels within any rectangle of the image in between four such integral pixels can be calculated very quickly. Therefore, Haar wavelets can be very quickly applied to the image. A learning algorithm is described which is based on AdaBoost, which selects a small number of critical Haar-like features from a larger set and yields extremely efficient classifiers. These classifiers can be combined to a classifier cascade which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions.

15

20

25

P.I. Wilson, J. Fernandez; Facial Feature Detection Using Haar Classifiers, JCSC 21, 4 (April 2006), CCSC: South Central Conference, describes a further method for recognizing faces in an image by means of Haar-like features. The area of the image being analyzed for a facial feature is regionalized to a location with the highest probability of containing the feature. By regionalizing the detection area, false positives are eliminated and the speed of detection is increased due to the reduction of the area examined.

30

In Sebastian Schmitt, Real-Time Object Detection With Haar-Like Features, June 22, 2010, s-schmitt.de/ressourcen/haar\_like\_features.pdf several projects using Haar-like features for detecting objects in real-time are described. In these projects rotated Haar-like features are used. In order to compute rotated features as fast as the axis-aligned ones, a rotated summed area table (RSAT), which corresponds to a rotated integral image, is used.

In F. Abdat, C. Maaoui and A. Pruski (2010); Real Time Facial Feature Points Tracking with Pyramidal Lucas-Kanade Algorithm, Human-Robot Interaction, Daisuke Chugo (Ed.), ISBN: 978-953-307-051-3, InTech, Available from: <http://www.intechopen.com/books/human-robot-interaction/real-time-facial-feature-pointstracking-with-pyramidal-lucas-kanade-algorithm>, a method for facial expression tracking is known using Haar-like features. With this method selected facial feature points can be tracked in a video sequence.

N. Dalal and B. Triggs; Histograms of oriented gradients for human detection, [lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf](http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf), published in Computer Vision and Pattern Recognition, 2005, CVPR 2005, IEEE Computer Society Conference on June 25, 2005 (Volume 1), pages 886-893, vol. 1, ISSN 1063-6919, Print ISBN 0-7695-2372-2, publisher IEEE, describes a method for detecting humans in images using Histogram of Oriented Gradient (HOG) descriptors for discriminating humans in the images. This method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. This is implemented by dividing the image window into small spatial regions ("cells"), for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. The combined histogram entries form the representation. Tiling the detection window with a dense (in fact, an overlapping) grid of HOG descriptors and using the combined feature vector in a conventional SVM (Support Vector Machine) based window classifier results in a human detection chain.

The object of the present invention is to provide a system and method for facial recognition, which allows quick identification of a person with high reliability.

The object is solved by a method and a system according to the independent claims. Advantageous embodiments of the present invention are disclosed in the corresponding sub-claims.

A method for facial recognition comprises the steps of

5 a) reading an image which can show one or more persons,  
b) detecting whether said image shows at least one human face of a person, wherein the method is only continued if the image shows at least one face,  
c) analyzing the image for non-facial attributes of the person of this face,  
d) extracting facial attributes of this face from the image,

10 e) sorting and/or filtering face templates stored in a database by said non-facial-attributes,  
f) searching the sorted and/or filtered database for a face-template matching this face of the image.

By using non-facial attributes for sorting and/or filtering face-templates stored in a database

15 before searching for a match with a face of the image, the amount of face-templates to be matched can be greatly reduced so that the face-templates database can be searched very quickly or face-templates can be matched with a high accuracy.

The inventors of the present invention have realized that non-facial attributes of persons are

20 very specific. By using just a small number of non-facial attributes the face-templates stored in the database can be sorted and/or filtered very efficiently.

Most facial attributes of different faces are rather similar. All faces comprise two eyes, one nose, one mouth, and these elements are rather similarly arranged. Therefore, the corresponding attributes are mostly very similar. Only with the combination of a plurality of such facial attributes can different faces be discriminated. In contrast non-facial attributes are often

25 very specific to persons. E.g. clothing can show very specific patterns and/or colors and hair texture can be very specific. Therefore, a small number of non-facial attributes can be used for discarding a major portion of face-templates stored in the database, which do not have

30 corresponding non-facial attributes.

In other words, the non-facial attributes can be used for a highly efficient pre-selection of face-templates of the database. By using just a small number of non-facial attributes, such as skin color, clothing, hairstyle and eyewear, the relevant number of face-templates to be matched can be reduced to 0.5 % - 5 % of all face-templates stored in the database. Therefore, the search for a face-template matching the extracted face thumbnail can be drastically accelerated or can be carried out with very high accuracy.

This method enables large-scale real-time face recognition across multiple cameras, particularly on the same day.

The order of the steps c) and d) can be changed so that firstly the facial attributes and secondly the non-facial attributes are determined or these steps can be also combined to one single step for extracting both the facial attributes and the non-facial attributes.

The non-facial attributes can comprise the

- color of skin, particularly the color of neck,
- hairstyle comprising e.g. shape of hair, length of hair, color of hair, texture of hair,
- style of clothing comprising e.g. color of clothing, texture of clothing, pattern of clothing, presence of collar,
- body form comprising e.g. shape of neck, shape of shoulder,
- presence of eyewear,
- color of eyewear.

For the purposes of facial recognition, some of these non-facial attributes, such as style of clothing and hairstyle are only valid for a short time, such as for a day. Other non-facial attributes, such as shape of neck, color of neck, shape of shoulder, usually remain stable for a long time. Therefore, it can be helpful to assign a timestamp to non-facial attributes to mark the time of taking the image or the time of extracting the non-facial attributes from the image. When the face-templates stored in the database are sorted and/or filtered by non-facial attributes the timestamp of the non-facial attributes can be combined with a weight according to a mean validity duration of the respective non-facial attributes.

A face thumbnail can be picked out from the image before carrying out step d). The face thumbnail is preferably determined in step b). This face thumbnail has the size of the face dimensions. The facial attributes are extracted from this face thumbnail. The search for a matching face-template according to step f) is carried out on the face thumbnail.

5

An attribute thumbnail can be extracted from the image which contains the face thumbnail and which is larger than the face thumbnail. Thus, the attribute thumbnail shows additional parts of a person besides his/her face. These parts should particularly comprise the hair, chest, the neck and/or the shoulders of the person. The size of the attribute thumbnail is preferably 2 to 4 times larger than the face thumbnail in order to contain the non-facial attributes. The size of the attribute thumbnail is preferably no larger than 2, 3, or 4 times the face thumbnail, because such a region is large enough to capture surrounding attributes yet does not have much of a chance of capturing nearby persons and background which are distractions.

15 The detection of a human face according to step b) is carried out by performing a classification method on the image by means of a wavelet transformation. The wavelet transformation preferably uses 2-dimensional Haar wavelets for the detection of Haar-like features. This classification method can be based on the above-mentioned methods for detecting objects in images by means of Haar-like features (Paul Viola and Michael Jones; Rapid object detection using a boosted cascade of simple features – P.I. Wilson, J. Fernandez; Facial Feature Detection Using Haar Classifiers – Sebastian Schmitt; Real-Time Object Detection With Haar-Like Features). Therefore, these documents are incorporated in its entirety.

25 The non-facial attributes relating to one of shapes are determined by an object detection method or by an edge detection method. The preferred object detection method is histograms of gradients. But there are further suitable edge detection methods, such as Canny edge detector, Canny-Deriche edge detector, differential edge detection, Sobel operator, Prewitt operator and Roberts cross operator.

30 Non-facial attributes relating to colors are determined by a color detection method. The preferred color detection method is color histograms.

The non-facial attributes relating to textures or patterns can be determined by a texture classification method, such as local binary patterns (LBP) or Gabor filters.

5 The extracting of facial attributes from the image or the face thumbnail, respectively, is carried out by a texture classification method, such as local binary patterns or Gabor filters.

The non-facial attributes of the image taken can form a non-facial vector. Each face-template of the database comprises a corresponding non-facial vector of non-facial attributes. The filtering according to step e) is carried out by selecting all face-templates of the database having a non-facial vector of non-facial attributes being less distanced from the non-facial vector of the image taken than a predetermined threshold distance.

10

Such a non-facial vector can also be used for sorting of the face-templates stored in the database according to step e) in that the face-templates of the database are sorted according to the distance of the non-facial vectors from the non-facial vector of the image taken.

15

By determining the distance of the non-facial vectors of the face-templates from the non-facial vector of the image taken, individual non-facial attributes can be weighted. The weight of the individual non-facial attributes can correspond to a tolerance with which the values of the respective non-facial attributes are determined. E.g. clothing comprising only a single color, which can be determined very clearly, has a higher weight for the attribute "color of clothing" than clothing having a pattern comprising many different small segments of different colors. The weight can also be applied in combination with the above-mentioned timestamp. The weight of a certain non-attribute corresponds to the attribute stability. The non-facial attributes relating to clothing do not usually have a stability lasting longer than one day. Therefore, the weight will be significantly reduced after a duration of more than one day. Attributes relating to the hair color, hair texture or hair shape of the person are usually more stable so that these non-facial attributes have a weight function which does not decrease as much over time as the non-facial attributes relating to clothing. The non-facial attributes relating to the shape of the neck or the shape of the shoulder are usually very stable and therefore, these non-facial attributes have a constant time weight.

20

25

30

The searching according to step f) can be carried out by sorting the selected face-templates or by sorting a limited number of sorted face-templates having a distance of the non-facial vectors from the non-facial vector of the image taken below a certain threshold value, wherein sorting is further carried out on the basis of the facial attributes. The facial attributes preferably form a face vector so that sorting can be carried out on the basis of the distance between the face vector of the image taken with respect to the face vectors of the stored face-templates. The sorting can be carried out by multi-dimensional indexing.

Multiple cameras can be used for taking a plurality of images, wherein facial recognition is carried out for each image. This method can be used for tracking individual persons in a certain time frame. In dependence of the time frame, non-facial attributes have to be selected. For a time frame of one day, all the above-mentioned non-facial attributes are suitable. In cases where the time frame runs longer than one day, non-facial attributes having stronger time stability are selected. This method is also suitable for monitoring or tracking individual persons in a large crowd of people. This is very advantageous for monitoring the audience of sports events with the purpose of identifying offenders such as hooligans.

This method for facial recognition can also be used for determining customer behavior for evaluating e. g. advertising measures or product displays. This method can also be used for recognizing customer acceptance of service and support centers.

This method can particularly be used for tracking and counting people in sales sectors and public spaces, especially in combination with the multiple-camera system.

The images that are processed by the method according to the present invention which can contain faces of persons can be captured by means of one or more cameras. These images can also stem from a database containing a plurality of images showing faces.

The present invention also relates to a system for facial recognition comprising at least one camera for taking images and a control unit connected to the at least one camera. The control unit is embodied for facial recognition according to the method described above.

The system preferably comprises a plurality of cameras e. g. at least five cameras, preferably at least ten cameras, and more preferably at least hundred cameras. The cameras can be placed in certain closed areas. The cameras can also be distributed in unconnected areas such as railway stations, airports, for tracking the movement of individual persons.

5

The invention is explained in more detail by means of the enclosed drawings which show in:

Figure 1        schematically, a system for facial recognition in a block diagram,

10    Figure 2        a method for facial recognition in a flow chart,

Figure 3        a statistical data collection software in a block diagram,

Figure 4a       a simple Haar-like feature set,

15

Figure 4b       an extended Haar-like feature set,

Figure 5        a first and second Haar-like feature selected by an AdaBoost algorithm.

20    Fig. 1 shows an embodiment of a system 1 for facial recognition according to the present invention which is designed for monitoring the use of shopping paths 2 in a shopping center 3.

The shopping path 2 extends between an entrance 4 and an exit 5 of the shopping center 3. The shopping path 2 comprises bifurcations with several bifurcated sections 6.

A customer can pass one or more of these bifurcated sections 6 on his way from the entrance 4 to the exit 5. A customer selects one or more of these bifurcated sections 6 in dependence of his needs, the products and promotions which are displayed in the bifurcated sections 6. The customer behavior is mainly influenced by the arrangement of the products and the pro-  
5 motions. Therefore, statistical data showing which places along the shopping paths 2 are attractive for customers to display certain products or promotions are very helpful for a shopping center manager.

The system 1 for facial recognition allows for the collection of this kind of statistical data.

10

The system 1 comprises a central control unit 7 having a processor unit 8 and a storage media 9 for storing a database. The processor unit 8 comprises a CPU, RAM (Random Access Memory) and ROM (Read Only Memory).

15

Several cameras 10 are provided which are connected by means of datalines 11 with the central control unit 7. In the present embodiment the cameras 10 are still image cameras. Basically, it is also possible to use video cameras or a combination of still image cameras and video cameras.

20

The cameras 10 can also be arranged in a remote place such as a parking lot of the shopping center, and be connected to the central control unit 7 via the internet 25.

The cameras 10 are digital cameras for generating electronically readable image files. These image files are transmitted to the central control unit 7. On the central control unit 7 a software 12 is stored and executable for collecting statistical data, wherein a facial recognition on  
25 the basis of the images delivered by the cameras 10 is carried out automatically.

25

The statistical data collection software 12 comprises several software modules (Fig. 3). A change detection module 13 is provided for detecting whether an incoming image comprises  
30 a change with respect to the previous image of the same camera. If an image is identical to the previous image it does not have to be analyzed and can be discarded.

30

A human detection module 14 is provided for detecting whether the image shows at least one human.

A face detection module 15 is designed for detecting one or more faces in the image. If the face detection module 15 detects a face then it extracts a face thumbnail and an attribute thumbnail. The face thumbnail is a rectangular section of the image showing the face from the forehead to the chin of the face. The attribute thumbnail is a section of the image which encloses the corresponding face thumbnail and a margin around the face thumbnail which shows at least the hair, neck and shoulder of the person relating to this face.

The face detection module 15 uses the technique for object detection in images based on so-called Haar-like features. Haar-like features represent meta features that are not explicitly present in the pixel intensities of an image. In general, a Haar-like feature encodes differences in average intensities of sub-regions within an image. The simplest feature set consists of a quadratic area that contains two or four rectangular sub-areas of the same size (Fig. 4a). These Haar-like features are applied to the image in that the sum of pixel values in the sub-areas is calculated wherein an intensity difference is determined of the white sub-areas on one side and the hatched sub-areas on the other side according to Fig. 4a. This difference represents the feature value.

The features can be scaled in their size to obtain feature information on different magnitudes.

An extended feature set is shown in Fig. 4b comprising edge features, line features and center-surround features. Some of the Haar-like features are rotated by 45°.

To be able to calculate the feature value in real time an image is converted into a so-called Integral Image or Summed Area Table (SAT). Such a Summed Area Table has the same size as the original image, wherein to each pixel the sum of all pixels to the left of and above the original image is assigned. Once the Summed Area Table is computed, pixel intensities within any axis-aligned rectangular sub-region of the original image can be summed up efficiently by only calculating a sum of four values.

In order to compute the rotated features as fast as the axis-aligned ones, a Rotated Summed Area Table (RSAT) is used. In such a Rotated Summed Area Table to each pixel the sum of the pixels in the original image is assigned, wherein the pixels in the original image are arranged in a rectangular area having edges inclined by  $45^\circ$  and wherein the pixel, to which the sum is assigned, forms the right most corner of the rectangular area.

To further improve the computation speed the Haar-like features are preferably applied in cascades for classifying a sub-window 16 of the image 17 which is to be analyzed with respect to the presence of a face. The Haar-like features are used to classify the sub-windows 19 and are therefore called Haar-classifiers when applied to the image.

The feature value of each Haar-classifier is compared with a feature-weight, wherein the Haar-classifier is true or false, if the feature value is larger or smaller than the feature weight or vice-versa. In a cascade of Haar-classifiers a sub-window 19 is rejected, if one Haar-classifier is false, and the calculation of the cascade is terminated and a further sub-window 19 can be analyzed by means of the cascade of Haar-classifiers.

For detecting human facial features, such as a mouth, eyes, and nose, it is necessary that Haar-classifier cascades are to be trained. A number of machine learning approaches can be used to learn the Haar-classifiers. The preferred algorithm is the AdaBoost learning procedure. Alternative learning procedures are e.g. a feature selection based on feature variance, a feature selection process based on the Winnow exponential perceptron learning rule or learning procedures using neuronal networks or support vector machines.

Fig. 5 shows the first and second Haar-like feature selected by the AdaBoost method. The two Haar features are shown in the top row and then overlaid on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and the region across the cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the mouth. This example is taken from Paul Viola et al. as discussed above.

With this face detection module 15 a plurality of sub-windows 19 can be analyzed quickly, wherein sub-windows of different sizes and different locations in the image are analyzed. Sub-windows, which show background only, are usually discarded by the first or at least by the second Haar-classifier.

5

If a face is detected then the corresponding sub-window forms a face thumbnail. An attribute thumbnail is generated on the basis of the face thumbnail, wherein the attribute thumbnail comprises the face thumbnail and a certain margin around the face thumbnail. Preferably, the attribute thumbnail is twice to four times as large as the face thumbnail.

10

A non-facial attribute extraction module 20 is provided for extracting non-facial attributes of a person shown in the image, wherein these non-facial attributes do not comprise features of the face of this person. These non-facial attributes comprise one or more of the following attributes: color of skin, shape of hair, color of hair, texture of hair, color of clothing, texture of clothing, pattern of clothing, shape of neck, color of neck, shape of shoulder, presence of eyewear, color of eyewear, hairstyle and/or presence of collar.

15

The non-facial attributes relating to shapes are determined by an object detection method or an edge detection method. In a preferred embodiment histograms of gradients are used as the object detection method for extracting shape related attributes. In N. Dalal et al.; Histograms of oriented gradients for human detection, as discussed above, a histogram of gradients method is disclosed which can be used for extracting shape relating attributes. Therefore, this document is incorporated in its entirety.

20

Non-facial attributes defining a certain color in a certain segment of the image are determined by a color detection method. In the present embodiment, a color histogram is used as color detection method according to which the frequency of pixels of certain colors in the segment is determined.

25

The non-facial attributes relating to a texture or a pattern are determined by texture classification methods. The texture classification method of the preferred embodiment is Local Binary Patterns (LBP).

30

A facial attribute extraction module 21 is provided for extracting features relating to the detected face. This facial attribute extracting module can copy the Haar-like features determined by the face detection module 15 and store them as facial attributes. Additionally, or alternatively, further facial attributes can be extracted by means of e. g. a texture classification method such as Local Binary Patterns.

A template pre-selection module 22 is designed for selecting face-templates of faces stored in the database in the storage media 9 on the basis of the non-facial attributes. The database in the storage media 9 comprises data sets to a plurality of face-templates. Each data set comprises at least one non-facial vector comprising non-facial attributes and at least one face vector comprising facial attributes of the corresponding face. Preferably, the data set also comprises the face thumbnail and/or the attribute thumbnail of the corresponding face and/or a data stamp or a time stamp.

The template pre-selection module 22 comprises a filter and/or a sorting algorithm for filtering and/or sorting the templates of the database on the basis of the non-facial attributes. This is carried out by calculating a distance between a non-facial vector of the face detected in the actual image by the face detection module 15 and the non-facial vectors of the face-templates of the database.

The face-templates are either sorted according to the computed distance or filtered according to this distance. If the face-templates are sorted, a certain number of face-templates is selected which have the smallest distance. This number can range from 10 – 10,000 and is preferably no smaller than 100 and particularly no smaller than 200 and is preferably no larger than 2,000 and particularly no larger than 1,000 or 500. The number of selected face-templates typically lies in the range of 0.5% to 5% of the unselected face-templates.

If a filter is used for selecting the face-templates, only those face-templates are selected which have a distance below a certain threshold distance. With both alternatives the number of face-templates, which have to be considered further, is drastically reduced. Preferably, the face-template pre-selection module 22 is adjusted in that no more than 10 % and particularly

no more than 5 % and preferably no more than 2 % of the face-templates of the database are selected for further processing.

5 The face-template pre-selection module can also be embodied for discarding face-templates showing a certain non-facial attribute. In shopping centers the staff often have to wear certain clothes. Attributes that relate to such a kind of clothing can be used to discard the face-templates relating to the staff of the shopping center, because only the customers, but not the staff, shall be monitored.

10 A matching module 23 is provided for searching for the best match of a face-template of the database with the detected face of the actual image.

15 The search for the best match is carried out on the basis of the facial attributes and particularly by means of the face vector of the face detected in the actual image and the face vectors of the face-templates. The best match is the face-template that has the smallest distance between its face vector and the corresponding face vector of the face thumbnail. The search is preferably carried out by multi-dimensional indexing. If there is no match below a predetermined threshold distance then the result is "no match".

20 A statistical analyzing module 24 uses the detected faces for a statistical analysis and can combine this information with additional information, such as the time, when the corresponding picture was taken, or the location of the person in the picture or the location of the camera.

25 In the following a method for collecting statistical data in the shopping center 3 by means of the system explained above for facial recognition is explained (see flow chart of Fig. 2). This method begins with step S1.

30 In step S2 an image is taken with one of the cameras 10. The cameras 10 can be embodied in such a way that they take images at regular intervals. These intervals can be e.g. between

0.1 s to 10 s. The cameras 10 can also be coupled to a proximity sensor so that a human being in front of the camera is detected by the proximity sensor. The proximity sensor triggers the capture of an image.

- 5 When an image is taken preferably a data stamp is generated and coupled to the image. The data stamp can comprise the time, when the image is taken, and/or a description of the location which is shown in the image. The description of the location can be coordinates or a speaking term, such as "Entrance of shopping center".
- 10 The camera 10 transmits the image via the dataline 11 to the central control unit 7.

The incoming image is checked by the change detection module 13, if there are any changes in the image with respect to the last image taken with the same camera 10 (step S3). If there is no change in the image then the image is discarded, because the same image has already  
15 been analyzed before. If no person is in front of a certain camera 10 in the shopping center 3 then the camera takes several identical images in a row. It is clear that it does not make any sense to analyze the same image in detail again and again.

If in step S3 it is determined that there is no change in the image then the program flow goes  
20 back to step S2. If in step S3 a change in the image is detected it is checked whether a human being is shown in the image (step S4). The typical contour of human beings can be readily detected by means of histograms of oriented gradients. If there is no human being shown in the image then the program flow goes back to step S2. If in step S4 a human being is detected then, preferably, also the number of the human beings in the image is determined and  
25 stored.

The face detection module 15 analyzes and detects a face in the image (step S5) by means of the above-described Haar-like features. The face thumbnail and the attribute thumbnail are also generated in this step.

30 The non-facial attributes extraction module 20 extracts non-facial attributes. In the present embodiment the persons are only detected during their stay in the shopping center which

lasts for a maximum of several hours. Therefore, it is appropriate to use non-facial attributes which are very significant, but which do remain valid over a longer period of time. Such non-facial attributes are e.g. all attributes relating to clothing and/or hairstyle. It is very unlikely that anyone will change his/her clothing or his/her hairstyle during his/her stay in the shopping center. In other applications it can be appropriate to select different non-facial attributes. The non-facial attributes are extracted from the attribute thumbnail.

Facial attributes are extracted by the face feature extracting module 21 from the face thumbnail (step S7). The facial attributes can be extracted either by just copying face features which were already determined in step S5, e.g. Haar-like features, or by applying a certain extraction routine to the face thumbnail.

The face-templates of the database are pre-selected by the face-template pre-selection module 22 by means of the extracted non-facial attributes (step S8). By this pre-selection only a small number of the face-templates stored in the database is selected.

These selected face-templates are then used to search for a match between the face thumbnail generated in step S5 and one of these face-templates in the database (step S9).

If no match can be found in step S9 then the program flow goes to step S10. In step S10 a new data set is added to the database relating to the detected face of the actually captured image. This data set comprises at least the corresponding face vector and the corresponding attribute vector. Preferably, this data set comprises also the face thumbnail and/or the attribute thumbnail. This data set can also include the data stamp, which was generated in step S2 comprising the time and/or the place when and where the image was taken.

In step S11 either the matching face-template found in step S9 or the new face-template stored in step S10 in the database are processed in a statistical analysis. In the present case it is analyzed which person uses which bifurcated section 6 of the shopping path 2. Furthermore, it can be analyzed how long the person stays in a certain bifurcated section 6 of the shopping path 2. This information can also be correlated to the products which are actually

bought by this person. The products bought by a certain person can be determined by detecting the corresponding person at the Point Of Sale (POS), wherein this information is correlated to the data that are registered at the cash register.

- 5 In step S12 it is checked whether a further human being was detected in the actual image. If this is the case then the program flow goes back to step S5 for detecting the next face. Otherwise, the program flow proceeds to step S13, in which it is checked, whether a further image is received by the central control unit 7. Then the program flow goes back to step S3. Otherwise the method is finished with step S14.

10

The above-described method is an example for collecting data in a shopping center. In this example, the face information revealed by the facial recognition process is used for statistical analysis. This kind of facial recognition process can also be used for other applications. With this facial recognition process, e.g. a crowd of people can be monitored, wherein individual people in the crowd can be easily tracked by means of the non-facial attributes. This can be used for monitoring the audience of a sports event, which could be disturbed by offenders, such as hooligans. This process can simultaneously analyze images of a plurality of cameras or images showing a plurality of faces. Once a person is registered in the database, the same person can be found in real time, even if he changes his position and has his image taken by different cameras. If a certain offender is identified and detected in a sports stadium, where it is difficult to isolate the offender, then this offender can easily be detained at a train station or any other public place which is monitored by a camera so long as this camera is linked to the system for facial recognition.

- 25 In the embodiment described above the number of human beings is detected in step S4 and the faces are detected in step S5. These two steps can also be combined into one step, wherein a face detection is also used for detecting faces or people, respectively, and for counting the people shown in the image.

- 30 Furthermore, the order of the steps S6 and S7 can be changed. It is also possible to combine the steps S5 and S7 into one single step, wherein, by detecting the faces simultaneously, the

face features are extracted. This is particularly suitable if Haar-like features are used as face features.

5 This method and system can also be used for monitoring security relevant areas, such as banks. This method allows the recognition people who are approach areas relevant to security several times during one day.

10 This method and system is also useful for analyzing a service process in a service center, where it can be detected easily how long a certain customer has to stay in the service center and which spots in the service center are addressed by a certain customer.

15 The basic principle of the present invention is to consider a small number of non-facial attributes for carrying out a pre-selection of templates stored in a database. Due to the high information content of the non-facial attributes it is possible to select a small number of potential relevant face-templates very quickly with a high reliability. Therefore, the face-templates ("faces") can be found very quickly and with a high accuracy. This system and method are particularly advantageous for monitoring people during a limited time period, such as one to five hours, or one to five days, or during one month. The non-facial attributes have to be selected according to the period during which the people shall be monitored.

20 By carrying out the pre-selection of the templates in step S8 on the basis of the non-facial attributes the distance between the corresponding non-facial vectors is calculated. By calculating this distance it is also possible to use time-dependent weights for each attribute, because there are attributes, which are more likely to be changed, and other attributes which are stable. Furthermore, it is possible to weight the attributes according to a tolerance calculated or  
25 estimated when the value of the corresponding attribute is determined. The smaller the tolerance is, the larger is the corresponding weight of the attribute.

### List of reference numbers

|    |    |                                        |
|----|----|----------------------------------------|
|    | 1  | system                                 |
| 5  | 2  | shopping path                          |
|    | 3  | shopping center                        |
|    | 4  | entrance                               |
|    | 5  | exit                                   |
|    | 6  | bifurcated section                     |
| 10 | 7  | central control unit                   |
|    | 8  | processor unit                         |
|    | 9  | storage media                          |
|    | 10 | camera                                 |
|    | 11 | dataline                               |
| 15 | 12 | statistical data collection software   |
|    | 13 | change detection module                |
|    | 14 | human detection module                 |
|    | 15 | face detection module                  |
|    | 16 | Haar-feature                           |
| 20 | 17 | sub-image                              |
|    | 18 | image                                  |
|    | 19 | sub-window                             |
|    | 20 | non-facial attribute extraction module |
|    | 21 | face feature extraction module         |
| 25 | 22 | template pre-selection module          |
|    | 23 | matching module                        |
|    | 24 | statistical analyzing module           |
|    | 25 | internet                               |

## Claims

1. A method for facial recognition comprises the steps of
- a) reading an image which can show one or more persons,
- 5 b) detecting whether said image shows at least one human face of a person, wherein the method is only continued if the image shows at least one face,
- c) analyzing the image for non-facial attributes of the person of this face,
- d) extracting facial attributes of this face from the image,
- e) sorting and/or filtering face templates stored in a database by said non-facial-attributes,
- 10 f) searching the sorted and/or filtered database for a face-template matching this face of the image;
- wherein the non-facial attributes comprise one or more of the following attributes:
- shape of hair
  - color of hair
  - 15 - texture of hair
  - color of clothing
  - texture of clothing
  - shape of neck
  - color of neck
  - 20 - shape of shoulder
  - presence of eyewear
  - presence of collar.
2. Method according to claim 1,
- 25 characterized in that
- a face thumbnail having a size of the face dimensions is picked out from the image before carrying out step d) and said facial attributes are extracted from the face thumbnail, wherein searching for a face-template according to step f) is carried out by matching the facial attributes of the extracted thumbnail.
- 30
3. Method according to claim 1 or 2,
- characterized in that

the detecting of a human face according to step b) is carried out by performing a classification method on the image by means of a wavelet transformation.

4. Method according to claim 3,

5 characterized in that

the wavelet transformation uses 2-dimensional Haar-like wavelets for the detection of Haar-like features.

5. Method according to claim 1,

10 characterized in that

non-facial attributes relating shapes are determined by an edge detection method (histograms of gradients) and/or

non-facial attributes relating colors are determined by a color detection method (color histograms) and/or

15 non-facial attributes relating textures are determined by a texture classification method (local binary patterns).

6. Method according to one of the claims 1 to 5,

characterized in that

20 extracting face features from the image is carried out by a texture classification method (local binary patterns).

7. Method according to one of the claims 1 to 6,

characterized in that

25 the non-facial attributes of said image taken form a non-facial vector and each template of the database comprises a corresponding non-facial vector of non-facial attributes, wherein the filtering according to step e) is carried out by selecting all face templates of said database having a non-facial vector of non-facial attributes being less distant from the non-facial vector of the image taken than a predetermined threshold distance.

30

8. Method according to one of the claims 1 to 7,

characterized in that

the non-facial attributes of said taken image forms a non-facial-vector and each face-template of the database comprises a corresponding non-facial vector of non-facial attributes, wherein the sorting according to step e) is carried out by sorting the face-templates of said database according to the distance of their non-facial vectors from the non-facial vector of the image taken.

9. Method according to claim 7 or 8,  
characterized in that  
the distance of the non-facial vectors of the templates from the non-facial vector of the image taken is determined by weighting the individual non-facial attributes.

10. Method according to one of the claims 1 to 9,  
characterized in that  
in step b) a face thumbnail is determined showing the face, wherein this face thumbnail is used for extracting face features, and an attribute thumbnail which contains the face thumbnail and which is larger than the face thumbnail is used for analyzing the non-facial attributes.

11. Method according to one of the claims 1 to 10,  
characterized in that  
the searching according to step f) is carried out by sorting the selected face templates or by sorting a limited number of sorted face templates having a distance between their non-facial vectors and the non-facial vector of the image taken that lies below a certain threshold, wherein the sorting is carried out on the basis of face features.

12. Method according to claim 11,  
characterized in that  
the sorting is carried out by multi-dimensional indexing.

13. Method according to one of the claims 1 to 11,  
characterized in that  
multiple cameras are used for taking a plurality of images, wherein for each image the facial recognition is carried out.

14. System for facial recognition comprising

- at least one camera for taking images,
- a control unit connected to the at least one camera, wherein a server is embodied for carrying out the method according to one of the claims 1 to 13.