(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 March 2011 (31.03.2011)

PCT

(10) International Publication Number
WO 2011/037579 A1

(51) International Patent Classification:
G06K 9/00 (2006.01) *G06T 7/40* (2006.01)

(21) International Application Number:
PCT/US2009/058476

(22) International Filing Date:
25 September 2009 (25.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.** [US/US]; 11445 Compaq Center Dr. West, MC 060701, Houston, Texas 77070 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ZHANG, Wei** [CN/US]; Hewlett-Packard Co., 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US). **ZHANG, Tong** [US/US]; Hewlett-Packard Co., 1501 Page Mill Rd., Palo Alto, California 94304-1100 (US).

(74) Agents: **LEE, Denise** et al.; Intellectual Property Administration, 3404 E. Harmony Road, Mail Stop 35, Fort Collins, Colorado 80528 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

WO 2011/037579 A1

(54) Title: FACE RECOGNITION APPARATUS AND METHODS

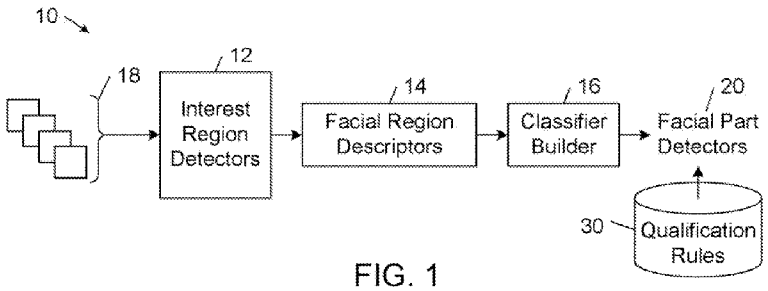


FIG. 1

(57) Abstract: Interest regions are detected in respective images (18) having face regions labeled with respective facial part labels. For each of the detected interest regions, a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region is determined. Ones of the facial part labels are assigned to respective ones of the facial region descriptor vectors. For each of the facial part labels, a respective facial part detector (20) that detects facial region descriptor vectors corresponding to the facial part label is built. The facial part detectors (20) are associated with rules (30) that qualify segmentation results of the facial part detectors (20) based on spatial relations between interest regions detected in images and the respective face part labels assigned to the facial part detectors (20). Faces in images are detected and recognized based on application of the facial part detectors (20) to images.

FACE RECOGNITION APPARATUS AND METHODS

BACKGROUND

[0001] Face recognition techniques oftentimes are used to locate, identify, or verify one or more persons appearing in images in an image collection. In a typical face recognition approach, faces are detected in the images; the detected faces are normalized; features are extracted from the normalized faces; and the identities of persons appearing in the images are identified or verified based on comparisons of the extracted features with features that were extracted from faces in one or more query images or reference images. Many automatic face recognition techniques can achieve modest recognition accuracy rates with respect to frontal images of faces that are accurately registered. When applied to other facial views (poses) and to poorly registered or poorly illuminated facial images, however, these techniques typically fail to achieve acceptable recognition accuracy rates.

[0002] What are needed are systems and methods that are capable of detecting and recognizing face images with wide variations in scale, pose, illumination, expression, and occlusion.

SUMMARY

[0003] In one aspect, the invention features a method in accordance with which interest regions are detected in respective images, which include respective face regions labeled with respective facial part labels. For each of the detected interest regions, a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region is determined. Ones of the facial part labels are assigned to respective ones of the facial region descriptor vectors determined for spatially corresponding ones of the face regions. For each of the facial part labels, a respective facial part detector that segments the facial region descriptor vectors that are assigned the facial part label from other ones of the facial region descriptor vectors is built. The facial part detectors are associated with rules that qualify segmentation results of the facial part detectors based on spatial relations between interest regions detected in images and the respective face part labels assigned to the facial part detectors.

[0004] In another aspect, the invention features a method in accordance with which interest regions are detected in an image. For each of the detected interest

regions, a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region is determined. A first set of the detected interest regions are labeled with respective face part labels based on application of respective facial part detectors to the facial region descriptor vectors. Each of the facial part detectors segments the facial region descriptor vectors into members and nonmembers of a class corresponding to a respective one of multiple facial part labels. A second set of the detected interest regions is ascertained. In this process, one or more of the labeled interest regions are pruned from the first set based on rules that impose conditions on spatial relations between the labeled interest regions.

[0005] The invention also features apparatus operable to implement the methods described above and computer-readable media storing computer-readable instructions causing a computer to implement the methods described above.

DESCRIPTION OF DRAWINGS

[0006] FIG. 1 is a block diagram of an embodiment of an image processing system.

[0007] FIG. 2 is a flow diagram of an embodiment of a method of building a face part detector.

[0008] FIG. 3A is a diagrammatic view of an exemplary set of face regions of an image labeled with respective face part labels in accordance with an embodiment of the invention.

[0009] FIG. 3B is a diagrammatic view of an exemplary set of face regions of an image labeled with respective face part labels in accordance with an embodiment of the invention.

[0010] FIG. 4 is a flow diagram of an embodiment of detecting face part regions in an image.

[0011] FIG. 5A is a diagrammatic view of an exemplary set of interest regions detected in an image.

[0012] FIG. 5B is a diagrammatic view of a subset of the interest regions detected in the image shown in FIG. 5A.

[0013] FIG. 6 is a flow diagram of an embodiment of a method of constructing a spatial pyramid representation of a face area in an image.

[0014] FIG. 7 is a diagrammatic view of a face area of an image partitioned into a set of different spatial bins in accordance with an embodiment of the invention.

[0015] FIG. 8 is a diagrammatic view of an embodiment of a process of matching a pair of images.

[0016] FIG. 9 is a diagrammatic view of an embodiment of an image processing system.

[0017] FIG. 10 is a block diagram of an embodiment of a computer system.

DETAILED DESCRIPTION

[0018] In the following description, like reference numbers are used to identify like elements. Furthermore, the drawings are intended to illustrate major features of exemplary embodiments in a diagrammatic manner. The drawings are not intended to depict every feature of actual embodiments nor relative dimensions of the depicted elements, and are not drawn to scale.

I. DEFINIITON OF TERMS

[0019] A "computer" is any machine, device, or apparatus that processes data according to computer-readable instructions that are stored on a computer-readable medium either temporarily or permanently. A "computer operating system" is a software component of a computer system that manages and coordinates the performance of tasks and the sharing of computing and hardware resources. A "software application" (also referred to as software, an application, computer software, a computer application, a program, and a computer program) is a set of instructions that a computer can interpret and execute to perform one or more specific tasks. A "data file" is a block of information that durably stores data for use by a software application.

[0020] As used herein, the term "includes" means includes but not limited to, the term "including" means including but not limited to. The term "based on" means based at least in part on. The term "ones" means multiple members of a specified group.

II. FIRST EXEMPLARY EMBODIMENT OF AN IMAGE PROCESSING SYSTEM

[0021] The embodiments that are described herein provide systems and methods that are capable of detecting and recognizing face images with wide variations in scale, pose, illumination, expression, and occlusion.

A. BUILDING A FACE RECOGNITION SYSTEM

[0022] FIG. 1 shows an embodiment of an image processing system 10 that includes interest region detectors 12, facial region descriptors 14, and a classifier builder (or inducer) 16. In operation, the image processing system 10 processes a set of training images 18 to produce a set of facial part detectors 20 that are capable of detecting facial parts in images.

[0023] FIG. 2 shows an embodiment of a method by which the image processing system 10 builds the facial part detectors 20.

[0024] In accordance with the method of FIG. 2, the image processing system 10 applies the interest region detectors 12 to the training images 18 in order to detect interest regions in the training images 18 (FIG. 2, block 22). Each of the training images 18 typically has one or more manually labeled face regions demarcating respective facial parts f_i appearing in the training images 18. In general, any of a wide variety of different interest region detectors may be used to detect interest regions in the training images 18. In some embodiments, the interest region detectors 12 are affine-invariant interest region detectors (e.g., Harris corner detectors, Hessian blob detectors, principal curvature based region detectors, and salient region detectors).

[0025] For each of the detected interest regions, the image processing system 10 applies the facial region descriptors 14 to the detected interest region in order to determine a respective facial region descriptor vector $\vec{V}_R = (d_1, \ldots, d_n)$ of facial region descriptor values characterizing the detected interest region (FIG. 2, block 24). In general, any of a wide variety of different local descriptors may be used to extract the facial region descriptor values, including distribution based descriptors, spatial-frequency based descriptors, differential descriptors, and generalized moment invariants. In some embodiments, the local descriptors 14 include a scale invariant feature transform (SIFT) descriptor and one or more textural descriptors (e.g., a local binary pattern (LBP) feature descriptor, and a Gabor feature descriptor).

[0026] The image processing system 10 assigns ones of the facial part labels in the training images 18 to respective ones of the facial region descriptor vectors that are determined for spatially corresponding ones of the face regions (FIG. 2, block 26). In this process, interest regions are assigned the labels that are associated with the face region that the interest regions overlap and each region descriptor vector $\vec{V}_R$ inherits the label assigned to the associated interest region. When the center of an

interest region is close to the boundaries of two manually labeled face regions or the interest region significantly overlaps two face regions, the interest region is assigned both facial part labels and the facial region descriptor vector associated with the interest region inherits both facial part labels.

[0027] For each of the facial part labels f_i, the classifier builder 16 builds (e.g., trains or induces) a respective one of the facial part detectors 20 that segments the facial region descriptor vectors $\vec{V}_R$ that are assigned the facial part label f_i from other ones of the facial region descriptor vectors $\vec{V}_R$ (FIG. 2, block 28). In this process, the facial region descriptor vectors $\vec{V}_R$ that are assigned the facial part label f_i are used as the positive training samples S_i^+, and the other facial region descriptor vectors are used as the negative training samples S_i^-. The facial part detector 20 for facial part label f_i is trained to discriminate S_i^+ from S_i^-.

[0028] The image processing system 10 associates the facial part detectors 20 with the qualification rules 30, which qualify segmentation results of the facial part detectors 20 based on spatial relations between interest regions detected in images and the respective face part labels assigned to the facial part detectors 20 (FIG. 2, block 32). As explained below, the qualification rules 30 typically are manually coded rules that describe favored and disfavored conditions on labeling of respective groups of interest regions with respective ones of the face part labels in terms of spatial relations between the interest regions in the groups. The segmentation results of the facial part detectors 20 are scored based the qualification rules 30, and segmentation results that have lower scores are more likely to be discarded.

[0029] In some embodiments, the image processing system 10 additionally segments the facial region descriptor vectors that are determined for all the training images 18 into respective clusters. Each of the clusters consists of a respective subset of the facial region descriptor vectors and is labeled with a respective unique cluster label. In general, the facial region descriptor vectors may be segmented (or quantized) into clusters using any of a wide variety of vector quantization methods. In some embodiments, the facial region descriptor vectors are segmented as follows. After extracting a large number of facial region descriptor vectors from a set of training images 18, k-means or hierarchical clustering is used to group these vectors into M clusters (types or classes), where M has a specified integer value. The center (e.g., the centroid) of each cluster is called a "visual word", and a list of the cluster centers forms

a "visual codebook," which is used to spatially match pairs of images, as described below. Each cluster is associated with a respective unique cluster label that constitutes the visual word. In the spatial matching process, each facial region descriptor vector that is determined for a pair of images (or image areas) to be matched is "quantized" by labeling it with the most similar (closest) visual word, and only the facial region descriptor vectors that are labeled with the same visual word are considered to be matches.

[0030] FIGS. 3A and 3B show examples of training images 33, 35. Each of the training images 33, 35 has one or more manually labeled rectangular face part regions 34, 36, 38, 40, 42, 44 demarcating respective facial parts (e.g., eyes, mouth, nose, etc.) appearing in the training images 33, 35. Each of the face part regions 34-44 is associated with a respective face part label (e.g., "eye" and "mouth"). The detected elliptical interest regions 46-74 are assigned the face part labels that are associated with the face part regions 34-44 with respect to which they have significant spatial overlap. For example, in the exemplary embodiment shown in FIG. 3A, the interest regions 46, 48, and 50 are assigned the face part label (e.g., "left eye") that is associated with face part region 34; the interest regions 52, 54, and 56 are assigned the face part label (e.g., "right eye") that is associated with face part region 36; and the interest regions 51, 53, and 55 are assigned the face part label (e.g., "mouth") that is associated with face part region 38. In the exemplary embodiment shown in FIG. 3B, the interest regions 58 and 60 are assigned the face part label (e.g., "left eye") that is associated with face part region 40; the interest regions 62, 64, and 66 are assigned the face part label (e.g., "right eye") that is associated with face part region 42; and the interest regions 68, 70, 72, and 74 are assigned the face part label (e.g., "mouth") that is associated with face part region 44.

[0031] In some embodiments, the image processing system 10 includes a face detector that provides a preliminary estimate of the location, size, and pose of the faces appearing in the training images 18. In general, the face detector may use any type of face detection process that determines the presence and location of each face in the training images 18. Exemplary face detection methods include but are not limited to feature-based face detection methods, template-matching face detection methods, neural-network-based face detection methods, and image-based face detection methods that train machine systems on a collection of labeled face samples. An

exemplary feature-based face detection approach is described in Viola and Jones, "Robust Real-Time Object Detection," Second International Workshop of Statistical and Computation theories of Vision – Modeling, Learning, Computing, and Sampling, Vancouver, Canada (July 13, 2001). An exemplary neural-network-based face detection method is described in Rowley et al., "Neural Network-Based Face Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1 (January 1998).

[0032] The face detector outputs one or more face region parameter values, including the locations of the face areas, the sizes (i.e., the dimensions) of the face areas, and the rough poses (orientations) of the face areas. In the exemplary embodiments shown in FIGS. 3A and 3B, the face areas are demarcated by respective elliptical boundaries 80, 82 that define the locations, sizes, and poses of the face areas appearing in the images 33, 35. The poses of the face areas are given by the orientation of the major and minor axes of the ellipses, which are usually obtained by locally refining the originally detected circular or rectangular face areas.

[0033] The image processing system 10 normalizes the locations and sizes (or scales) of the detected interest regions based on the face region parameter values so that the qualification rules 30 can be applied to the segmentation results of the facial part detectors 20. For example, the qualification rules 30 typically describe conditions on labeling of respective groups of interest regions with respective ones of the face part labels in terms of spatial relations between the interest regions in the groups. In some embodiments, the spatial relations model the relative angle and distance between face parts or the distance between face parts and the centroid of the face. The qualification rules 30 typically describe the most likely spatial relations between the major face parts, such as eyes, nose, mouth, cheeks. One exemplary qualification rule promotes segmentation results in which, on a normalized face, the right eye is most likely to be found displaced from the left eye along a line at a 0° angle (horizontal) at a distance of half the face area width. Another exemplary qualification rule reduces the likelihood of segmentation results in which a labeled eye region overlaps with a labeled mouth region.

B. RECOGNIZING FACES IN IMAGES

[0034] The image processing system 10 uses the facial part detectors 20 and the qualification rules in the process of recognizing faces in images.

[0035] FIG. 4 shows an embodiment by which the image processing system 10 detects face parts in an image.

[0036] In accordance with the embodiment of FIG. 4, the image processing system 10 detects interest regions in the image (FIG. 4, block 90). In this process, the image processing system 10 applies the interest region detectors 12 to the image in order to detect interest regions in the image. FIG. 5A shows an exemplary set of elliptical interest regions 89 that are detected in an image 91.

[0037] For each of the detected interest regions, the image processing system 10 determines a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region (FIG. 4, block 92). In this process, the image processing system 10 applies the facial region descriptors 14 to each of the detected interest regions in order to determine a respective facial region descriptor vector $\vec{V}_k = (d_1, \ldots, d_n)$ of facial region descriptor values characterizing the detected interest region.

[0038] The image processing system 10 labels a first set of the detected interest regions with respective face part labels based on application of respective ones of the facial part detectors 20 to the facial region descriptor vectors (FIG. 4, block 94). Each of the facial part detectors 20 segments the facial region descriptor vectors into members and nonmembers of a class corresponding to a respective one of the facial part labels that are associated with the facial part detectors 20. The classification decision is soft with a prediction confidence value. An exemplary classifier with real-valued confidence value is Support Vector Machine described in Christopher, J. C. B. "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, volume 2(2), pages 121-167 (1998).

[0039] The image processing system 10 ascertains a second set of the detected interest regions (FIG. 4, block 96). In this process, the image processing system 10 prunes one or more of the labeled interest regions from the first set based on the qualification rules 30, which impose conditions on spatial relations between the labeled interest regions.

[0040] In some embodiments, the image processing system 10 applies a robust matching algorithm to the first set of classified facial region descriptor vectors in order to further prune and refine facial region descriptor vectors based on the classification of the interest regions corresponding to the labeled facial region descriptor

vectors. The matching algorithm is an extension of a Hough Transform process that incorporates the face-specific domain knowledge encoded in the qualification rules 30. In this process, each instantiation of a group of the facial region descriptor vectors at the corresponding detected interest regions vote for a possible location, scale and pose of the face area. The confidence of voting is decided by two measures: (a) confidence values associated with the classification results produced by the facial part detectors; and (b) the consistency of the spatial configuration of the classified facial region descriptor vectors with the qualification rules 30. For example, a facial region descriptor vector labeled as a mouth is not likely to be collinear with a pair of facial region descriptor vectors labeled as eyes, thus, the vote for this group of labeled facial region descriptor vectors will have near-zero confidence no matter how confident the detectors are.

[0041] The image processing system 10 obtains a final estimation of the location, scale and pose of the face area based on the spatial locations of the group of labeled facial region descriptor vectors that have the dominant vote. In this process, the image processing system 10 determines the location, scale and pose of the face area based on a face area model that takes as inputs the spatial locations particular ones of the labeled facial region descriptor vectors (e.g., the locations of the centroids of facial region descriptor vectors respectively classified as a left eye, a right eye, a mouth, lips, a cheek, and/or a nose). In this process, the image processing system 10 aligns (or registers) the face area so that the person's face can be recognized. For each detected face area, the image processing system 10 aligns the extracted features in relation to a respective face area demarcated by a face area boundary that encompasses some or all portions of the detected face area. In some embodiments, the face area boundary corresponds to an ellipse that includes the eyes, nose, mouth but not the entire forehead or chin or top of head of a detected face. Other embodiments may use face area boundaries of different shapes (e.g., rectangular).

[0042] The image processing system 10 further prunes the classification of the facial region descriptor vectors based on the final estimation of the location, scale and pose of the face area. In this process, the image processing system 10 discards any of the labeled facial region descriptor vectors that are inconsistent with a model of the locations of face parts in a normalized face area that corresponds to the final estimate of the face area. For example, the image processing system 10 discards

interest regions that are labeled as eyes that are located in the lower half of the normalized face area. If no face part label is assigned to a facial region descriptor vector after the pruning process, that facial region descriptor vector is designated as being "missing." In this way, the detection process can handle the recognition of occluded faces. The output of the pruning process includes "cleaned" facial region descriptor vectors that are associated with interest regions that are aligned (e.g., labeled consistently) with corresponding face parts in the image, and parameters that define the final estimated location, scale, and pose of the face area. FIG. 5B shows the cleaned set of elliptical interest regions 89 that are detected in the image 91 and a face area boundary 98 that demarcates the final estimated location, scale, and pose of the face area. The final estimation of the location, scale and pose of the face area is expected to be much more accurate than the original area detected by the face detectors.

[0043] FIG. 6 shows an embodiment of a method by which the image processing system 10 constructs from the cleaned facial region descriptor vectors and the final estimate of the face area a spatial pyramid that represents a face area that is detected in an image.

[0044] In accordance with the method of FIG. 6, the image processing system 10 segments (or quantizes) the facial region descriptor vectors into respective ones of the predetermined face region descriptor vector cluster classes (FIG. 6, block 100). As explained above, each of these clusters is associated with a respective unique cluster label. The segmentation process is based on the respective distances between the facial region descriptor vectors and the facial region descriptor vector cluster classes. In general, a wide variety of vector difference measures may be used to determine the distances between the facial region descriptor vectors and the cluster classes. In some embodiments, the distances correspond to a vector norm (e.g., the L2-norm) between the facial region descriptor vectors and the centroids of the facial region descriptor vectors in the clusters. Each of the facial region descriptor vectors is segmented into the closest (i.e., shortest distance) one of the cluster classes.

[0045] The image processing system 10 assigns to each of the facial region descriptor vectors the cluster label that is associated with the facial region descriptor vector cluster class into which the facial region descriptor vector was segmented (FIG. 6, block 102).

[0046] At multiple levels of resolution, the image processing system 10 subdivides the face area into different spatial bins (FIG. 6, block 104). In some embodiments, the image processing system 10 subdivides the face area into log-polar spatial bins. FIG. 7 shows an exemplary embodiment of image 91 in which the face region, which is demarcated by the face region boundary 98, is divided into a set of log-polar bins at four different resolution levels, each corresponding to a different set of the elliptical boundaries 98, 106, 108, 110. In other embodiments, the image processing system 10 subdivides the face area into rectangular spatial bins.

[0047] For each of the levels of resolution, the image processing system 10 tallies respective counts of instances of the cluster labels in each spatial bin to produce a spatial pyramid representing the face area in the given image (FIG. 6, block 112). In other words, for each cluster label, the image processing system 10 counts the facial region descriptor vectors that fall in each spatial bin to produce a respective spatial pyramid histogram.

[0048] The image processing system 10 is operable to recognize a person's face in the given image based on comparisons of the spatial pyramid with one or more predetermined spatial pyramids generated from one or more known images containing the person's face. In this process, the image processing system constructs a pyramid match kernel that corresponds to a weighted sum of histogram intersections between the spatial pyramid representation of the face in the given image and the spatial pyramid determined for another image. A histogram match occurs when facial descriptor vectors of the same cluster class (i.e., have the same cluster label) are located in the same spatial bin. The weight that is applied to the histogram intersections typically increases with increasing resolution level (i.e., decreasing spatial bin size). In some embodiments, the image processing system 10 compares the spatial pyramids using a pyramid match kernel of the type described in S. Lazebnik, C. Schmid, J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," IEEE Conference on Computer Vision and Pattern Recognition 2006.

[0049] FIG. 8 shows an embodiment of a process by which the image processing system 10 matches two face areas 98, 114 that appear in a pair of images 91, 35. The image processing system 10 subdivides the face areas 98, 114 into different spatial bins as described above in connection with block 104 of FIG. 6. Next,

the image processing system 10 determines spatial pyramid representations 116, 118 of the face areas 98, 35 as described above in connection with block 112 of FIG. 6. The image processing system 10 calculates a pyramid match kernel 120 from the weighted sum of intersections between the spatial pyramid representations 116, 118. The calculated value of the pyramid match kernel 120 corresponds to measure 122 of similarity between the faces areas 98, 114. In some embodiments, the image processing system 10 determines whether or not a pair of face areas match (i.e., are images of the same person) by applying a threshold to the similarity measure 122 and declares a match when the similarity measure 122 exceeds the threshold (FIG. 8, block 124).

III. SECOND EXEMPLARY EMBODIMENT OF AN IMAGE PROCESSING SYSTEM

[0050] FIG. 9 shows an embodiment 130 of the image processing system 10 that includes the interest region detectors 12, the facial region detectors 14, and the classifier builder 16. The image processing system 130 additionally includes auxiliary region detectors 132, and an optional second classifier builder 136

[0051] In operation, the image processing system 130 processes the training images 18 to produce the facial part detectors 20 that are capable of detecting facial parts in images as described above in connection with the image processing system 10. The image processing system 130 also applies the auxiliary region descriptors to the detected interest regions to determine a set of auxiliary region descriptor vectors 132 and builds the set of auxiliary region detectors 136 from the auxiliary region descriptor vectors. The process of applying the auxiliary region descriptors 132 and building the auxiliary part detectors 136 is essentially the same as the process by which the image processing system 10 applies the facial region descriptors 14 and builds the facial part detectors 20; the primary difference being the nature of the auxiliary region descriptors 132, which are tailored to represent patterns typically found in contextual regions, such as eyebrows, ears, forehead, chin, and neck, which do not tend to change much over time and different occasions.

[0052] In these embodiments, the image processing system 130 applies the interest region detectors 12 to the training images 18 in order to detect interest regions in the training images 18 (see FIG. 2, block 22). Each of the training images 18 typically has one or more manually labeled face regions demarcating respective facial parts f_i appearing in the training images 18 and one or more manually labeled auxiliary regions

demarcating respective auxiliary parts a_i appearing in the training images 18. In general, any of a wide variety of different interest region detectors may be used to detect interest regions in the training images 18. In some embodiments, the interest region detectors 12 are affine-invariant interest region detectors (e.g., Harris corner detectors, Hessian blob detectors, principal curvature based region detectors, and salient region detectors).

[0053] For each of the detected interest regions, the image processing system 130 applies the facial region descriptors 14 to the detected interest region in order to determine a respective facial region descriptor vector $\vec{V}_{FR} = (d_1, \ldots, d_n)$ of facial region descriptor values characterizing the detected interest region (see FIG. 2, block 24). The image processing system 130 also applies the auxiliary (or contextual) region descriptors 14 to each of the detected interest region in order to determine a respective auxiliary region descriptor vector $\vec{V}_{AR} = (c_1, \ldots, c_n)$ of auxiliary region descriptor values characterizing the detected interest region. In general, any of a wide variety of different local descriptors may be used to extract the facial region descriptor values and the auxiliary region descriptor values, including distribution based descriptors, spatial-frequency based descriptors, differential descriptors, and generalized moment invariants. In some embodiments, the auxiliary and facial descriptors 132, 14 include a scale invariant feature transform (SIFT) descriptor and one or more textural descriptors (e.g., a local binary pattern (LBP) feature descriptor, and a Gabor feature descriptor). The auxiliary descriptors also include shape-based descriptors. An exemplary type of shape-based descriptor is a shape context descriptor that describes a distribution over relative positions of the coordinates on an auxiliary region shape using a coarse histogram of the coordinates of the points on the shape relative to a given point on the shape. Addition details of the shape context descriptor are described in Belongie, S., Malik, J. and Puzicha, J., "Shape matching and object recognition using shape contexts," In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 24(4), pages 509–522 (2002).

[0054] The image processing system 130 assigns ones of the facial part labels in the training images 18 to respective ones of the facial region descriptor vectors that are determined for spatially corresponding ones of the face regions (see FIG. 2, block 26). The image processing system 130 also assigns ones of the auxiliary part labels in the training images 18 to respective ones of the auxiliary region descriptor

vectors that are determined for spatially corresponding ones of the auxiliary regions. In this process, interest regions are assigned the labels that are associated with the auxiliary region that the interest regions overlap and each auxiliary region descriptor vector $\vec{V}_{AR}$ inherits the label assigned to the associated interest region. When the center of an interest region is close to the boundaries of two manually labeled auxiliary regions or the interest region significantly overlaps two auxiliary regions, the interest region is assigned both auxiliary part labels and the auxiliary region descriptor vector associated with the interest region inherits both auxiliary part labels.

[0055] For each of the facial part labels f_i, the classifier builder 16 builds (e.g., trains or induces) a respective one of the facial part detectors 20 that segments the facial region descriptor vectors $\vec{V}_{FR}$ that are assigned the facial part label f_i from other ones of the facial region descriptor vectors $\vec{V}_{FR}$ (see FIG. 2, block 28). For each of the auxiliary part labels a_i, the classifier builder 134 builds (e.g., trains or induces) a respective one of the auxiliary part detectors 136 that segments the auxiliary region descriptor vectors $\vec{V}_{AR}$ that are assigned the auxiliary part label a_i from other ones of the auxiliary region descriptor vectors $\vec{V}_{AR}$. In this process, the auxiliary region descriptor vectors $\vec{V}_{AR}$ that are assigned the auxiliary part label a_i are used as the positive training samples T_i^+, and the other auxiliary region descriptor vectors are used as the negative training samples T_i^-. The auxiliary part detector 136 for auxiliary part label a_i is trained to discriminate T_i^+ from T_i^-.

[0056] The image processing system 130 associates the facial part detectors 20 with the qualification rules 30, which qualify segmentation results of the facial part detectors 20 based on spatial relations between interest regions detected in images and the respective face part labels assigned to the facial part detectors 20 (see FIG. 2, block 32). The image processing system 130 also associates the auxiliary part detectors 136 with auxiliary part qualification rules 138, which qualify segmentation results of the auxiliary part detectors 136 based on spatial relations between interest regions detected in images and the respective auxiliary part labels assigned to the auxiliary part detectors 136. The auxiliary part qualification rules 138 typically are manually coded rules that describe favored and disfavored conditions on labeling of respective groups of interest regions with respective ones of the auxiliary part labels in terms of spatial relations between the interest regions in the groups. The segmentation results of the auxiliary

part detectors 136 are scored based the auxiliary part qualification rules 138, and segmentation results that have lower scores are more likely to be discarded in a manner analogous to the process described above in connection with the face part qualification rules 30.

[0057] In some embodiments, the image processing system 130 additionally segments the auxiliary region descriptor vectors that are determined for all the training images 18 into respective clusters. Each of the clusters consists of a respective subset of the auxiliary region descriptor vectors and is labeled with a respective unique cluster label. In general, the auxiliary region descriptor vectors may be segmented (or quantized) into clusters using any of a wide variety of vector quantization methods. In some embodiments, the auxiliary region descriptor vectors are segmented as follows. After extracting a large number of auxiliary region descriptor vectors from a set of training images 18, k-means or hierarchical clustering is used to group these vectors into K clusters (types or classes), where K has a specified integer value. The center (e.g., the centroid) of each cluster is called a "visual word", and a list of the cluster centers forms a "visual codebook", which is used to spatially matching pairs of images, as described above. Each cluster is associated with a respective unique cluster label that constitutes the visual word. In the spatial matching process, each auxiliary region descriptor vector that is determined for a pair of images (or image areas) to be matched is "quantized" by labeling it with the most similar (closest) visual word, and only the auxiliary region descriptor vectors that are labeled with the same visual word are considered to be matches in the spatial pyramid matching process described above.

[0058] The image processing system 130 seamlessly integrates the auxiliary part detectors 136 and the auxiliary part qualification rules 138 into the face recognition process described above in connection with the image processing system 10. The integrated face recognition process uses the auxiliary part detectors 136 to classify auxiliary region descriptor vectors that are determined for each image, prunes the set of auxiliary region descriptor vectors using the auxiliary part qualification rules 138, performs vector quantization on the cleaned set of auxiliary region descriptor vectors to build a visual codebook of auxiliary regions, and performs spatial pyramid matching on the visual codebook representation of the auxiliary region descriptor vectors in respective ways that are directly analogous to the corresponding ways described above

in which the image processing system 10 recognizes faces using the facial part detectors 20 and the qualification rules 30.

IV. EXEMPLARY OPERATING ENVIRONMENT

[0059] Each of the training images 18 (see FIG. 1) may correspond to any type of image, including an original image (e.g., a video keyframe, a still image, or a scanned image) that was captured by an image sensor (e.g., a digital video camera, a digital still image camera, or an optical scanner) or a processed (e.g., sub-sampled, filtered, reformatted, enhanced or otherwise modified) version of such an original image.

[0060] Embodiments of the image processing systems 10 (including image processing system 130) may be implemented by one or more discrete modules (or data processing components) that are not limited to any particular hardware, firmware, or software configuration. In the illustrated embodiments, these modules may be implemented in any computing or data processing environment, including in digital electronic circuitry (e.g., an application-specific integrated circuit, such as a digital signal processor (DSP)) or in computer hardware, firmware, device driver, or software. In some embodiments, the functionalities of the modules are combined into a single data processing component. In some embodiments, the respective functionalities of each of one or more of the modules are performed by a respective set of multiple data processing components.

[0061] The modules of the image processing systems 10, 130 may be co-located on a single apparatus or they may be distributed across multiple apparatus; if distributed across multiple apparatus, these modules and the display 24 may communicate with each other over local wired or wireless connections, or they may communicate over global network connections (e.g., communications over the Internet).

[0062] In some implementations, process instructions (e.g., machine-readable code, such as computer software) for implementing the methods that are executed by the embodiments of the image processing systems 10, 130, as well as the data they generate, are stored in one or more machine-readable media. Storage devices suitable for tangibly embodying these instructions and data include all forms of non-volatile computer-readable memory, including, for example, semiconductor memory devices, such as EPROM, EEPROM, and flash memory devices, magnetic disks such as internal hard disks and removable hard disks, magneto-optical disks, DVD-ROM/RAM, and CD-ROM/RAM.

[0063] In general, embodiments of the image processing systems 10, 130 may be implemented in any one of a wide variety of electronic devices, including desktop computers, workstation computers, and server computers.

[0064] FIG. 10 shows an embodiment of a computer system 140 that can implement any of the embodiments of the image processing system 10 (including image processing system 130) that are described herein. The computer system 140 includes a processing unit 142 (CPU), a system memory 144, and a system bus 146 that couples processing unit 142 to the various components of the computer system 140. The processing unit 142 typically includes one or more processors, each of which may be in the form of any one of various commercially available processors. The system memory 144 typically includes a read only memory (ROM) that stores a basic input/output system (BIOS) that contains start-up routines for the computer system 140 and a random access memory (RAM). The system bus 146 may be a memory bus, a peripheral bus or a local bus, and may be compatible with any of a variety of bus protocols, including PCI, VESA, Microchannel, ISA, and EISA. The computer system 140 also includes a persistent storage memory 148 (e.g., a hard drive, a floppy drive, a CD ROM drive, magnetic tape drives, flash memory devices, and digital video disks) that is connected to the system bus 146 and contains one or more computer-readable media disks that provide non-volatile or persistent storage for data, data structures and computer-executable instructions.

[0065] A user may interact (e.g., enter commands or data) with the computer 140 using one or more input devices 150 (e.g., a keyboard, a computer mouse, a microphone, joystick, and touch pad). Information may be presented through a user interface that is displayed to a user on the display 151 (implemented by, e.g., a display monitor), which is controlled by a display controller 154 (implemented by, e.g., a video graphics card). The computer system 140 also typically includes peripheral output devices, such as speakers and a printer. One or more remote computers may be connected to the computer system 140 through a network interface card (NIC) 156.

[0066] As shown in FIG. 10, the system memory 144 also stores the image processing system 10, a graphics driver 158, and processing information 160 that includes input data, processing data, and output data. In some embodiments, the image processing system 10 interfaces with the graphics driver 158 (e.g., via a DirectX® component of a Microsoft Windows® operating system) to present a user

interface on the display 151for managing and controlling the operation of the image processing system 10.

V. CONCLUSION

[0067] The embodiments that are described herein provide systems and methods that are capable of detecting and recognizing face images with wide variations in scale, pose, illumination, expression, and occlusion.

[0068] Other embodiments are within the scope of the claims.

CLAIMS

1. A method, comprising:

detecting interest regions in respective images (18), wherein the images (18) comprise respective face regions labeled with respective facial part labels;

for each of the detected interest regions, determining a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region;

assigning ones of the facial part labels to respective ones of the facial region descriptor vectors determined for spatially corresponding ones of the face regions;

for each of the facial part labels, building a respective facial part detector (20) that segments the facial region descriptor vectors that are assigned the facial part label from other ones of the facial region descriptor vectors; and

associating the facial part detectors (20) with rules (30) that qualify segmentation results of the facial part detectors (20) based on spatial relations between interest regions detected in images and the respective face part labels assigned to the facial part detectors (20);

wherein the determining, the assigning, the building, and the associating are performed by a computer (140).

2. The method of claim 1, wherein at least one of the rules (30) describes a condition on labeling of a given group of interest regions with respective ones of the face part labels in terms of a spatial relation between the interest regions in the given group.

3. The method of claim 1, wherein the images (18) comprise respective auxiliary regions that are outside the face regions and are labeled with respective auxiliary part labels, and further comprising:

for each of the detected interest regions, determining a respective auxiliary region descriptor vector of region descriptor values characterizing the detected interest region;

assigning ones of the auxiliary part labels to respective ones of the auxiliary region descriptor vectors determined for spatially corresponding ones of the auxiliary regions;

for each of the auxiliary part labels, building a respective auxiliary part detector (136) that segments the auxiliary region descriptor vectors (136) that are assigned the auxiliary part label from other ones of the auxiliary region descriptor vectors (136); and

associating the auxiliary part detectors (136) with rules (138) that qualify segmentation results of the auxiliary part detectors (136) based on spatial relations between interest regions detected in images and the respective auxiliary part labels assigned to the auxiliary part detectors (136).

4. The method of claim 3, further comprising:

labeling interest regions detected in a given image with respective ones of the face part labels and the auxiliary part labels based on application of the facial part detectors (20) to respective facial region descriptor vectors determined for the labeled interest regions and further based on application of the auxiliary part detectors (136) to respective auxiliary region descriptor vectors determined for the interest regions;

ascertaining a face area (98, 114) in the given image (91, 35) based on the labeled interest regions;

at multiple levels of resolution, subdividing the face area (98, 114) into different spatial bins;

for each of the levels of resolution, tallying respective counts of instances of the face part labels in each spatial bin; and

constructing from the tallied counts a spatial pyramid representation (116, 118) of the face area (98, 114) in the given image (91, 35).

5. The method of claim 1, wherein the determining comprises: applying facial region descriptors (14) to the detected interest regions to produce a first set of facial region descriptor vectors of facial region descriptor values characterizing the detected interest regions; and segmenting the first set of facial region descriptor vectors into clusters, wherein each of the clusters consists of a respective subset of the first set of facial region descriptor vectors and is labeled with a respective unique cluster label.

6. A method, comprising:

detecting interest regions (89) in an image (91);

for each of the detected interest regions (89), determining a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region (89);

labeling a first set of the detected interest regions (89) with respective face part labels based on application of respective facial part detectors (20) to the facial region descriptor vectors, wherein each of the facial part detectors (20) segments the facial region descriptor vectors into members and nonmembers of a class corresponding to a respective one of multiple face part labels; and

ascertaining a second set of the detected interest regions, wherein the ascertaining comprises pruning one or more of the labeled interest regions from the first set based on rules (30) that impose conditions on spatial relations between the labeled interest regions;

wherein the detecting, the determining, the labeling, and the ascertaining are performed by a computer (140).

7. The method of claim 6, wherein at least one of the rules (30) describes a condition on the labeling of a given group of interest regions (89) with respective ones of the face part labels in terms of a spatial relation between the interest regions (89) in the group.

8. The method of claim 7, further comprising identifying respective groups of the labeled interest regions (89) that satisfy the rules (30), and determining parameter values specifying location, scale, and pose defining a face area (98) in the image (91) based on locations of the labeled interest regions (89) in the identified groups.

9. The method of claim 8, further comprising segmenting the facial region descriptor vectors into respective predetermined face region descriptor vector cluster classes based on respective distances between the facial region descriptor vectors and the facial region descriptor vector cluster classes, wherein each of the facial region descriptor vector cluster classes is associated with a respective unique cluster label, and each of the facial region descriptor vectors is assigned the cluster label associated with the facial region descriptor vector cluster class into which the facial region descriptor vector was segmented.

10. The method of claim 9, further comprising:

at multiple levels of resolution, subdividing the face area (98) into different spatial bins; and

for each of the levels of resolution, tallying respective counts of instances of the unique cluster labels in each spatial bin to produce a spatial pyramid (116) representing the face area (98) in the given image (91).

11. The method of claim 10, further comprising recognizing a person's face in the image (89) based on comparisons of the spatial pyramid (116) with one or more predetermined spatial pyramids (118) generated from other images (35).

12. The method of claim 6, further comprising:

for each of the detected interest regions (89), determining a respective auxiliary region descriptor vector of auxiliary region descriptor values characterizing the detected interest region (89);

labeling a third set of the detected interest regions (89) with respective auxiliary part labels based on application of respective auxiliary part detectors (136) to the auxiliary region descriptor vectors, wherein each of the auxiliary part detectors (136) segments the auxiliary region descriptor vectors into members and nonmembers of a class corresponding to a respective one of the auxiliary part labels;

ascertaining a fourth set of the detected interest regions (89), wherein the ascertaining of the fourth set comprises pruning one or more of the labeled interest regions from the third set based on rules (138) that impose conditions on spatial relations between the labeled interest regions in the third set.

13. Apparatus, comprising:

a computer-readable medium (144, 148) storing computer-readable instructions; and

a processor (142) coupled to the computer-readable medium (144, 148), operable to execute the instructions, and based at least in part on the execution of the instructions operable to perform operations comprising

detecting interest regions in respective images (18), wherein the images (18) comprise respective face regions labeled with respective facial part labels,

for each of the detected interest regions, determining a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region,

assigning ones of the facial part labels to respective ones of the facial region descriptor vectors determined for spatially corresponding ones of the face regions,

for each of the facial part labels, building a respective facial part detector (20) that segments the facial region descriptor vectors that are assigned the facial part label from other ones of the facial region descriptor vectors, and

associating the facial part detectors (20) with rules (30) that qualify segmentation results of the facial part detectors based on spatial relations between interest regions detected in images and the respective face part labels assigned to the facial part detectors.

14. The apparatus of claim 13, wherein at least one of the rules (30) describes a condition on labeling of a given group of interest regions with respective ones of the face part labels in terms of a spatial relation between the interest regions in the given group.

15. The apparatus of claim 13, wherein in the determining the processor (142) is operable to perform operations comprising: applying facial region descriptors to the detected interest regions to produce a first set of facial region descriptor vectors of facial region descriptor values characterizing the detected interest regions; and segmenting the first set of facial region descriptor vectors into clusters, wherein each of the clusters consists of a respective subset of the first set of facial region descriptor vectors and is labeled with a respective unique cluster label.

16. At least one computer-readable medium (144, 148) having computer-readable program code embodied therein, the computer-readable program code adapted to be executed by a computer (140) to implement a method comprising:

detecting interest regions in respective images (18), wherein the images (18) comprise respective face regions labeled with respective facial part labels;

for each of the detected interest regions, determining a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region;

assigning ones of the facial part labels to respective ones of the facial region descriptor vectors determined for spatially corresponding ones of the face regions;

for each of the facial part labels, building a respective facial part detector (20) that segments the facial region descriptor vectors that are assigned the facial part label from other ones of the facial region descriptor vectors; and

associating the facial part detectors (20) with rules (30) that qualify segmentation results of the facial part detectors (20) based on spatial relations between interest regions detected in images and the respective face part labels assigned to the facial part detectors (20).

17. The at least one computer-readable medium of claim 16, wherein at least one of the rules (30) describes a condition on labeling of a given group of interest regions with respective ones of the face part labels in terms of a spatial relation between the interest regions in the given group.

18. The at least one computer-readable medium of claim 16, wherein the determining comprises: applying facial region descriptors to the detected interest regions to produce a first set of facial region descriptor vectors of facial region descriptor values characterizing the detected interest regions; and segmenting the first set of facial region descriptor vectors into clusters, wherein each of the clusters consists of a respective subset of the first set of facial region descriptor vectors and is labeled with a respective unique cluster label.

19. Apparatus, comprising:

a computer-readable medium (144, 148) storing computer-readable instructions; and

a processor (142) coupled to the computer-readable medium (144, 148), operable to execute the instructions, and based at least in part on the execution of the instructions operable to perform operations comprising

detecting interest regions (89) in an image (91);

for each of the detected interest regions (89), determining a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region;

labeling a first set of the detected interest regions (89) with respective face part labels based on application of respective facial part detectors (20) to the facial region descriptor vectors, wherein each of the facial part detectors (20) segments the facial region descriptor vectors into members and nonmembers of a class corresponding to a respective one of multiple face part labels; and

ascertaining a second set of the detected interest regions (89), wherein the ascertaining comprises pruning one or more of the labeled interest regions (89) from the first set based on rules (30) that impose conditions on spatial relations between the labeled interest regions (89).

20. At least one computer-readable medium (144, 148) having computer-readable program code embodied therein, the computer-readable program code adapted to be executed by a computer (142) to implement a method comprising:

detecting interest regions (89) in an image (91);

for each of the detected interest regions (89), determining a respective facial region descriptor vector of facial region descriptor values characterizing the detected interest region;

labeling a first set of the detected interest regions (89) with respective face part labels based on application of respective facial part detectors (20) to the facial region descriptor vectors, wherein each of the facial part detectors (20) segments the facial region descriptor vectors into members and nonmembers of a class corresponding to a respective one of multiple face part labels; and

ascertaining a second set of the detected interest regions (89), wherein the ascertaining comprises pruning one or more of the labeled interest regions (89) from the first set based on rules (30) that impose conditions on spatial relations between the labeled interest regions (89).

5

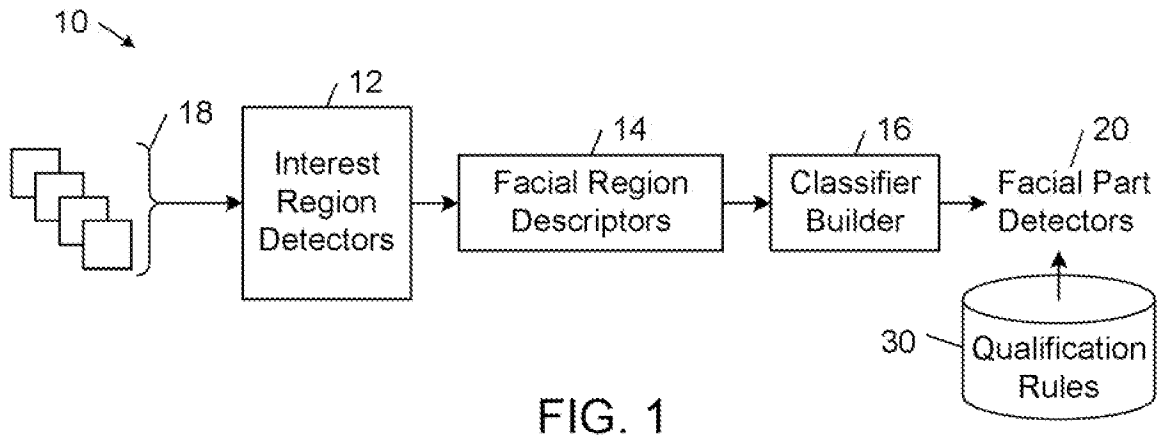

10
18
12
Interest Region Detectors
14
Facial Region Descriptors
16
Classifier Builder
20
Facial Part Detectors
30
Qualification Rules

FIG. 1

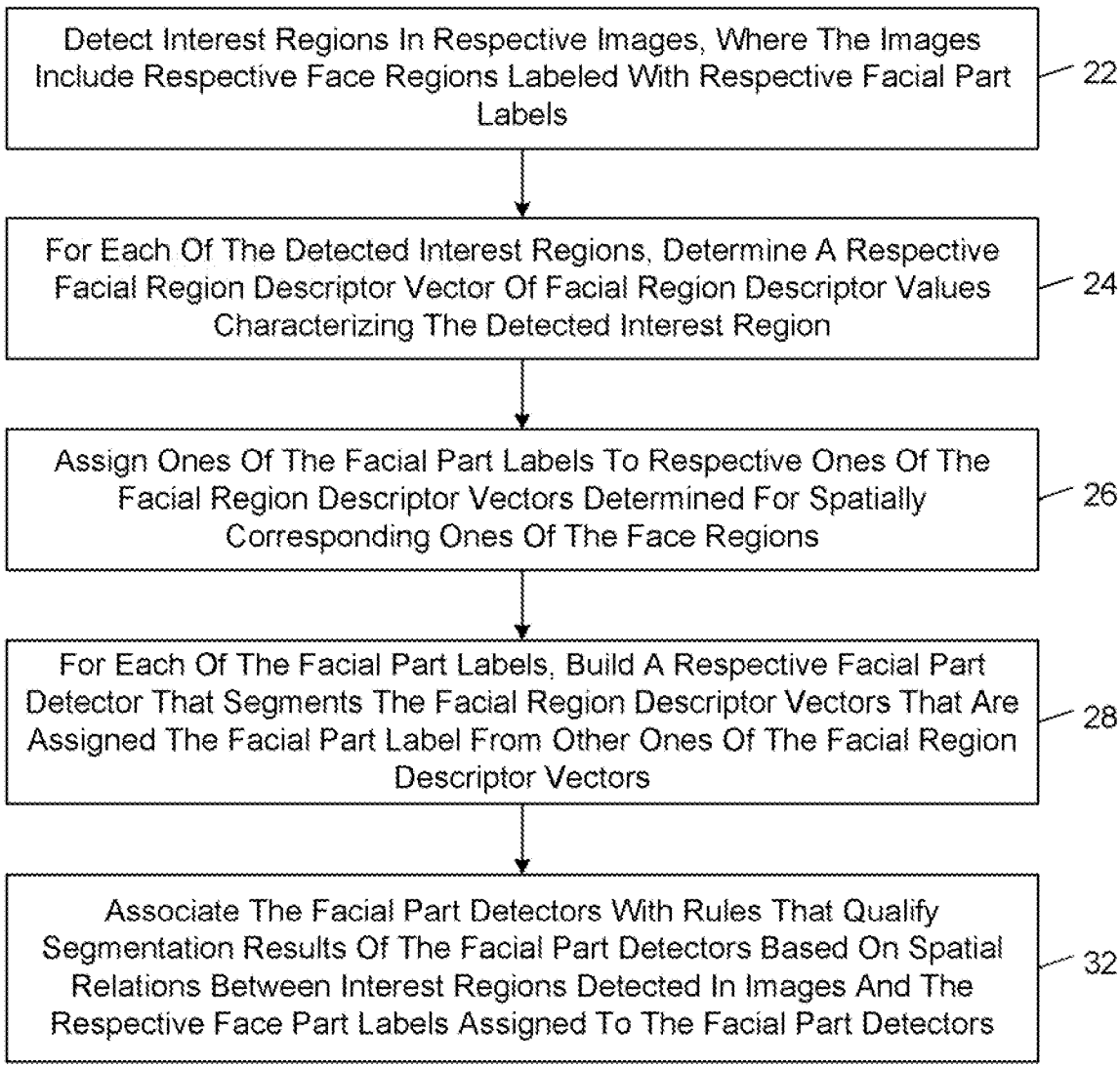

Detect Interest Regions In Respective Images, Where The Images Include Respective Face Regions Labeled With Respective Facial Part Labels
22
For Each Of The Detected Interest Regions, Determine A Respective Facial Region Descriptor Vector Of Facial Region Descriptor Values Characterizing The Detected Interest Region
24
Assign Ones Of The Facial Part Labels To Respective Ones Of The Facial Region Descriptor Vectors Determined For Spatially Corresponding Ones Of The Face Regions
26
For Each Of The Facial Part Labels, Build A Respective Facial Part Detector That Segments The Facial Region Descriptor Vectors That Are Assigned The Facial Part Label From Other Ones Of The Facial Region Descriptor Vectors
28
Associate The Facial Part Detectors With Rules That Qualify Segmentation Results Of The Facial Part Detectors Based On Spatial Relations Between Interest Regions Detected In Images And The Respective Face Part Labels Assigned To The Facial Part Detectors
32

FIG. 2

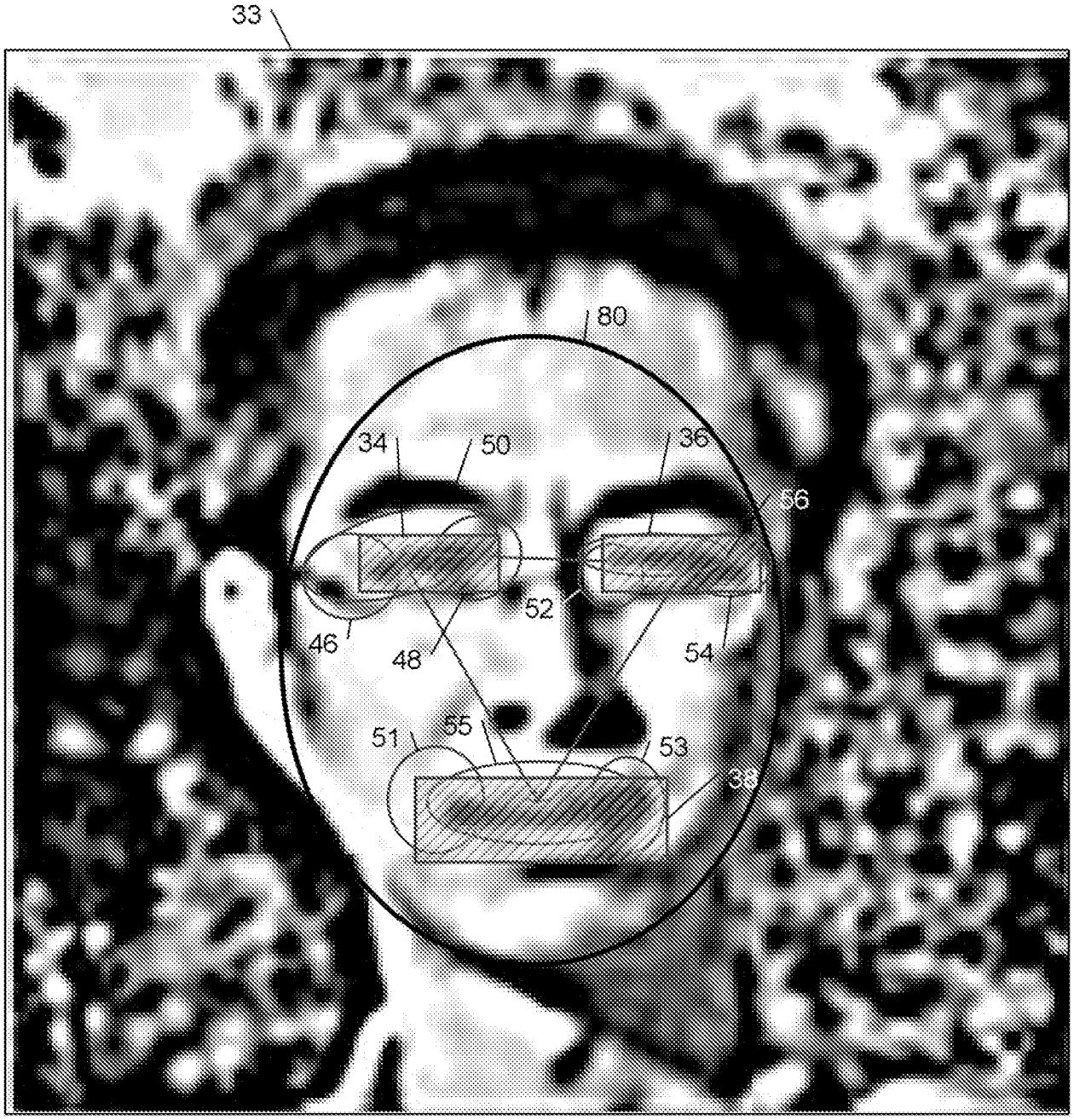


FIG. 3A



FIG. 3B

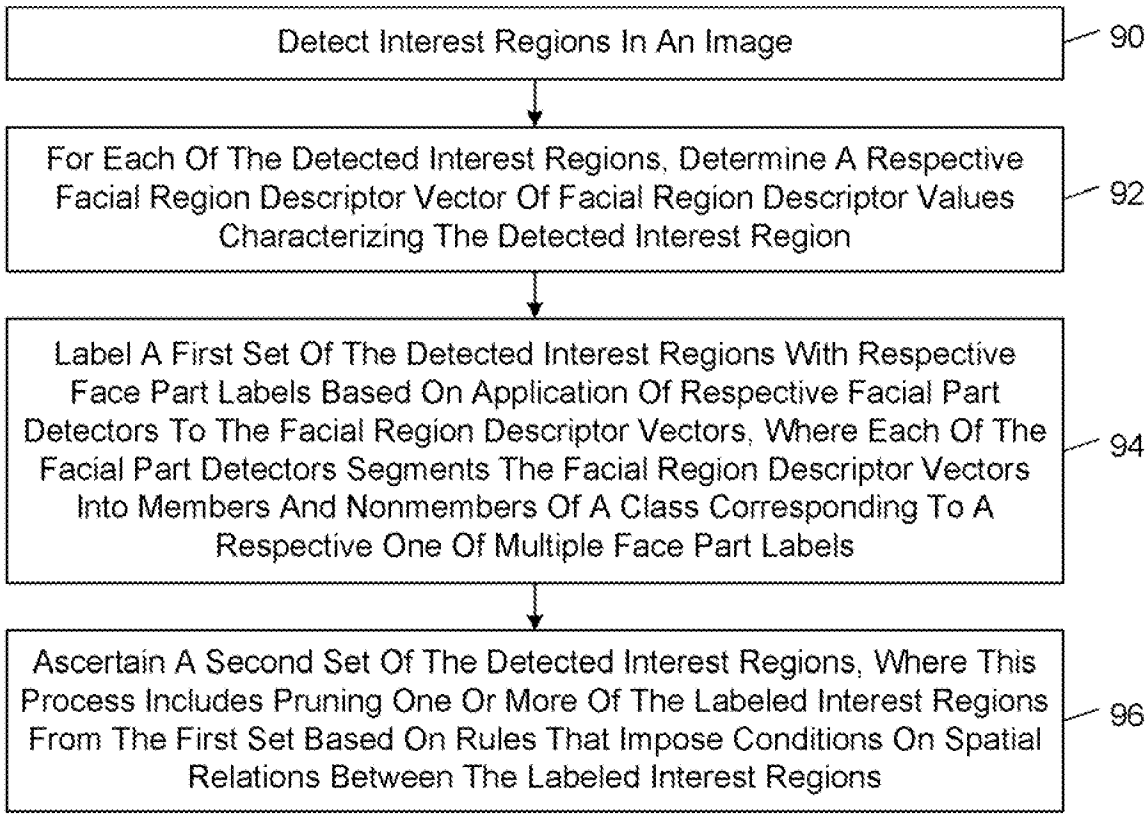


FIG. 4

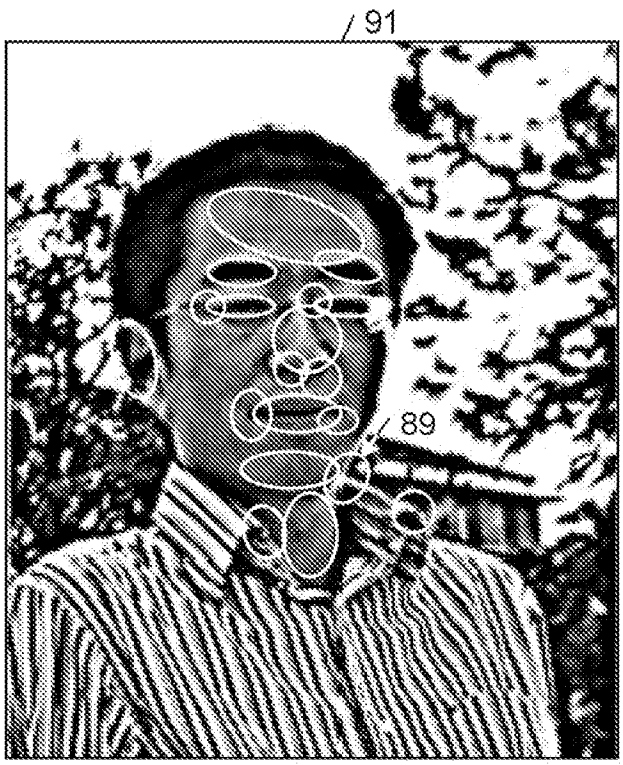


FIG. 5A

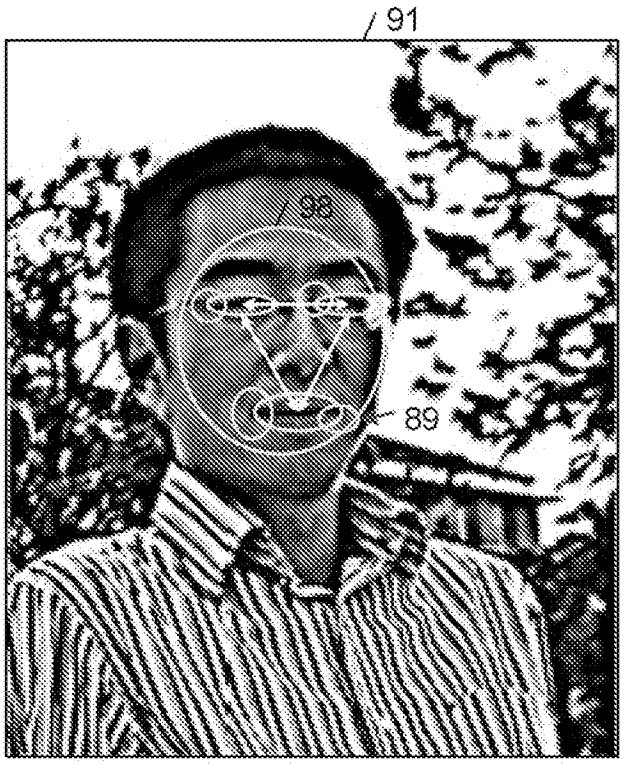


FIG. 5B

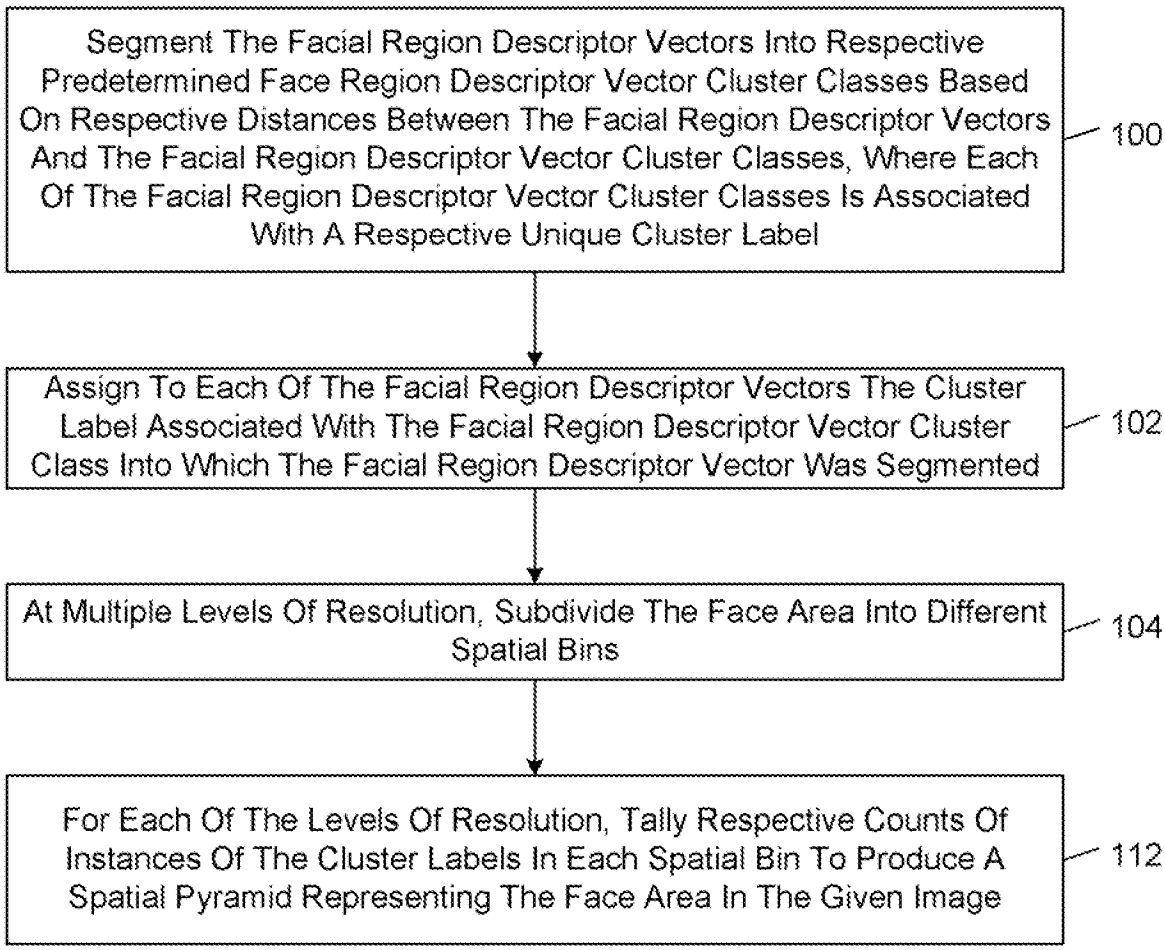

Segment The Facial Region Descriptor Vectors Into Respective Predetermined Face Region Descriptor Vector Cluster Classes Based On Respective Distances Between The Facial Region Descriptor Vectors And The Facial Region Descriptor Vector Cluster Classes, Where Each Of The Facial Region Descriptor Vector Cluster Classes Is Associated With A Respective Unique Cluster Label
100
Assign To Each Of The Facial Region Descriptor Vectors The Cluster Label Associated With The Facial Region Descriptor Vector Cluster Class Into Which The Facial Region Descriptor Vector Was Segmented
102
At Multiple Levels Of Resolution, Subdivide The Face Area Into Different Spatial Bins
104
For Each Of The Levels Of Resolution, Tally Respective Counts Of Instances Of The Cluster Labels In Each Spatial Bin To Produce A Spatial Pyramid Representing The Face Area In The Given Image
112

FIG. 6

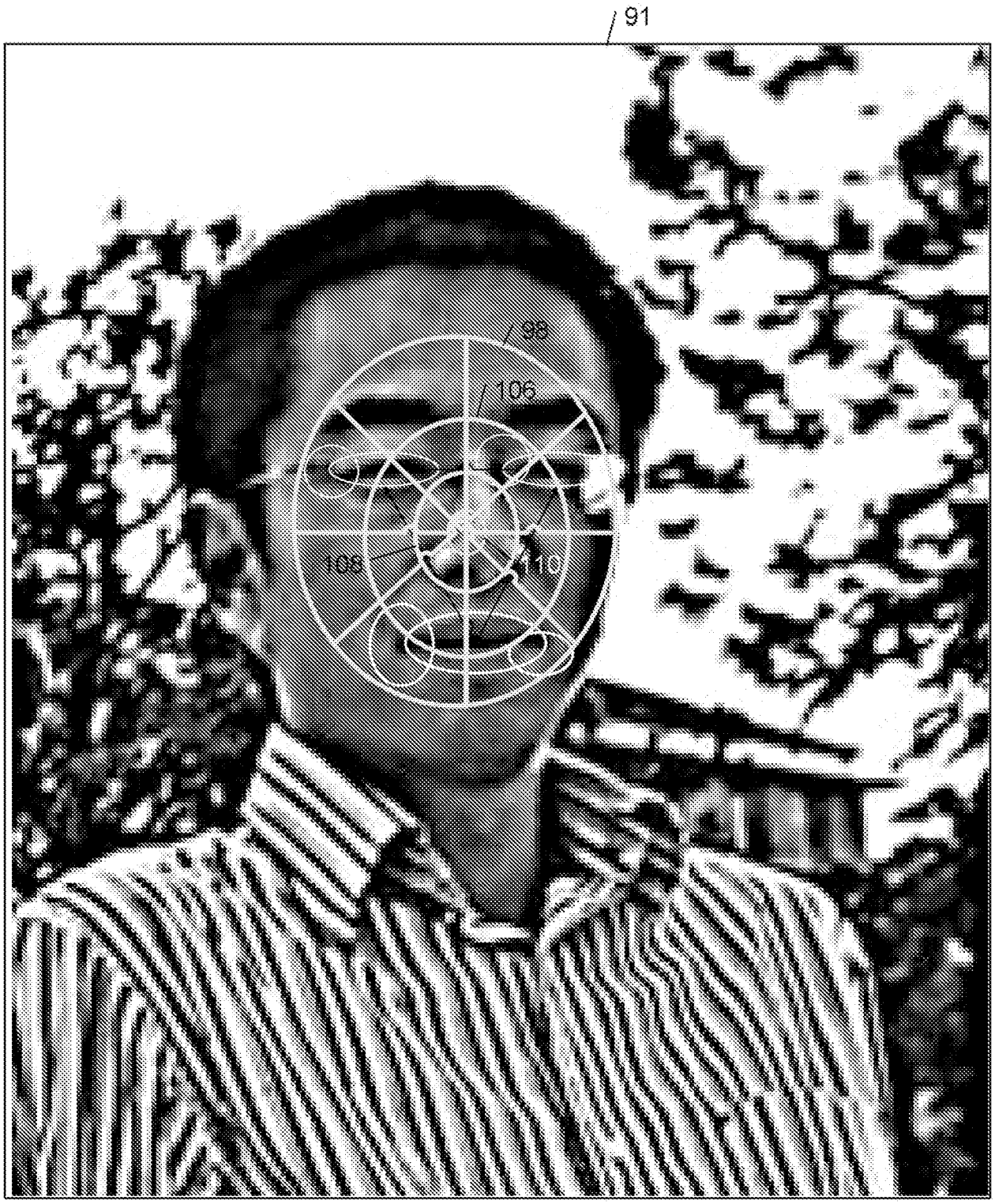


FIG. 7

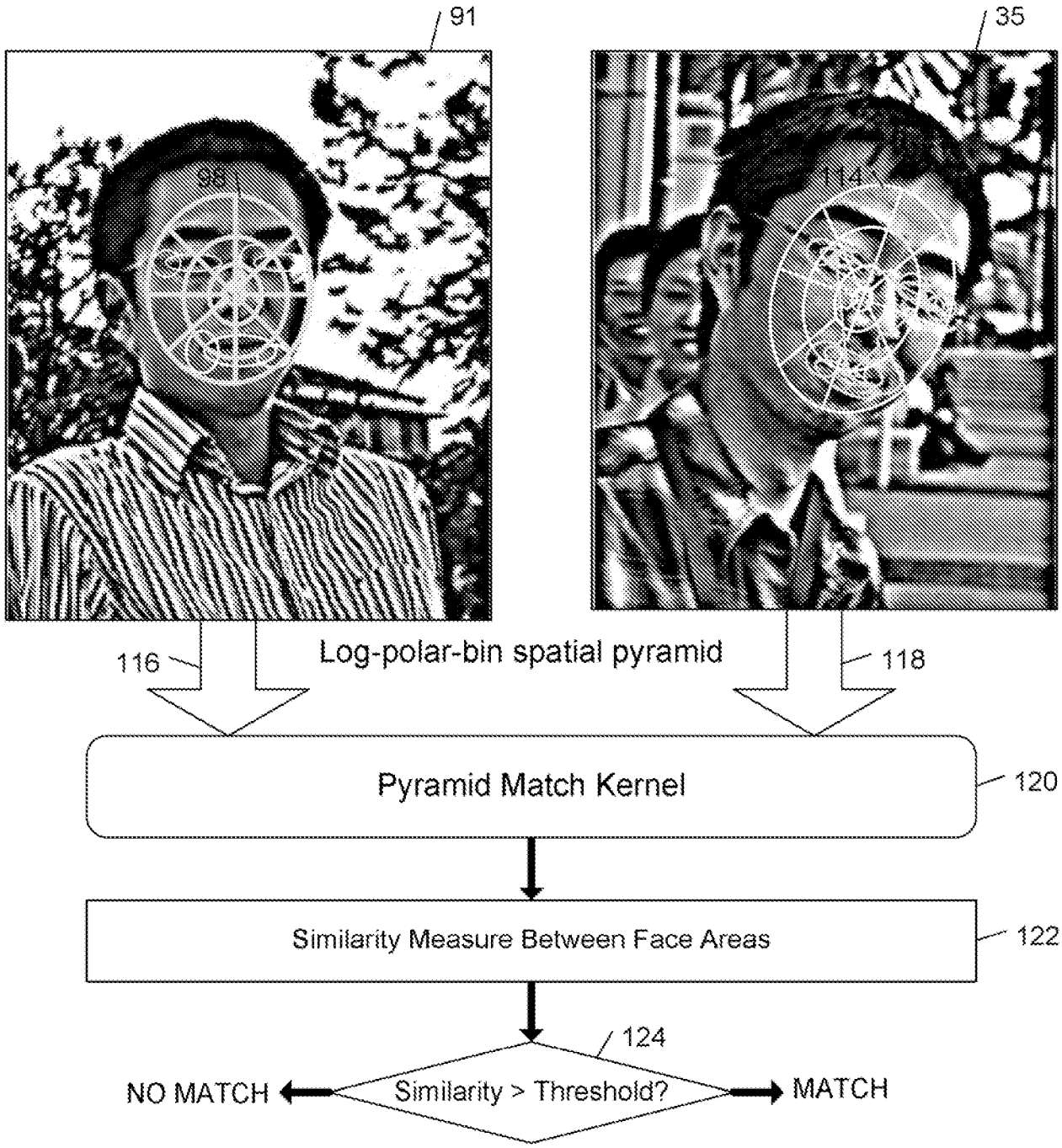

91
35
96
114
116
Log-polar-bin spatial pyramid
118
Pyramid Match Kernel
120
Similarity Measure Between Face Areas
122
124
NO MATCH
Similarity > Threshold?
MATCH

FIG. 8

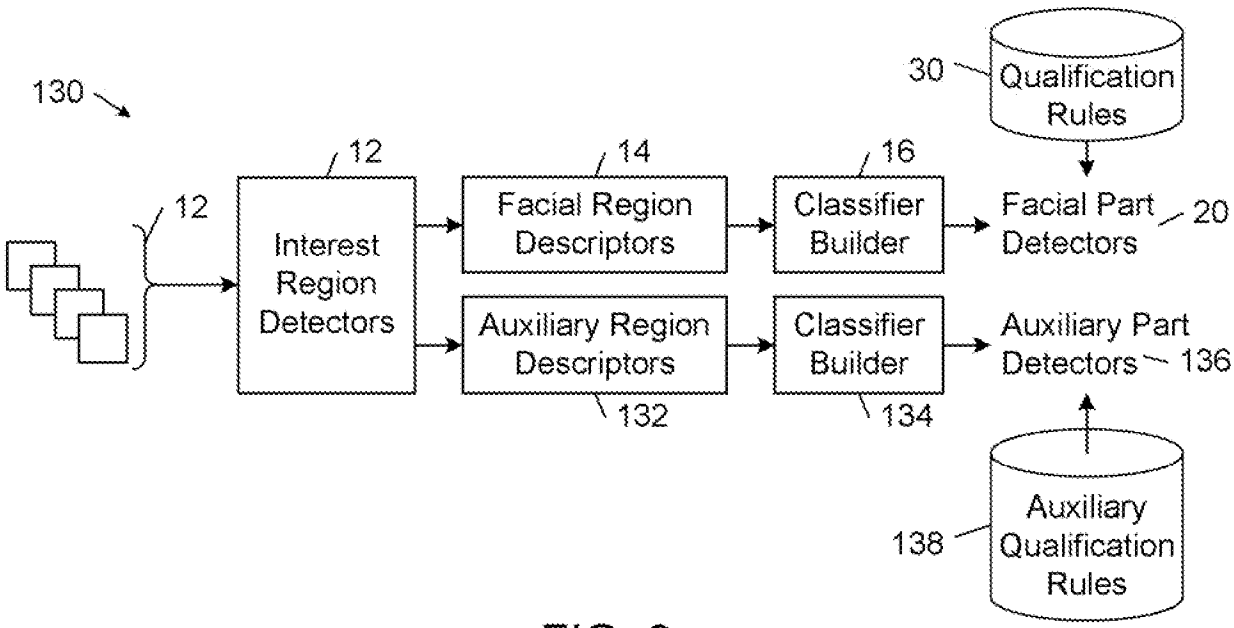

130
30
Qualification Rules
12
12
14
16
Interest Region Detectors
Facial Region Descriptors
Classifier Builder
Facial Part Detectors
20
Auxiliary Region Descriptors
Classifier Builder
Auxiliary Part Detectors
136
132
134
Auxiliary Qualification Rules
138

FIG. 9

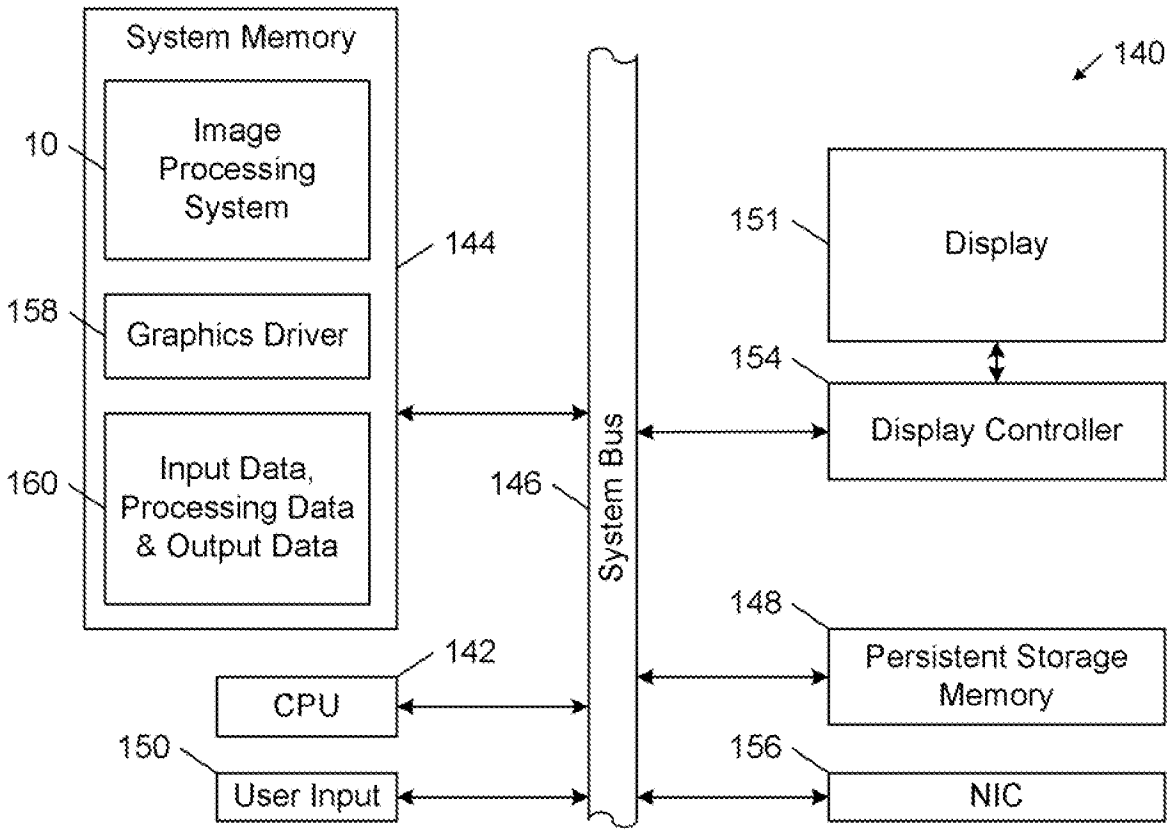

System Memory
140
10
Image Processing System
144
151
Display
158
Graphics Driver
154
Display Controller
160
Input Data, Processing Data & Output Data
146
System Bus
148
Persistent Storage Memory
142
CPU
150
User Input
156
NIC

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2009/058476

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i, G06T 7/40(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K 9/00; G06T 1/00; G06T 7/00; G06T 7/20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: face, detection, recognition, feature, label, relation, descriptor, vector

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2007-087345 A (CANON INC) 05 April 2007 See Abstract; claims 1,7-9; figures 2,8,20-24	1-20
A	JP 2007-034723 A (GLORY LTD) 08 February 2007 See Abstract; claims 1,6-7; figures 1,4	1-20
A	JP 2007-265367 A (FUJIFILM CORP) 11 October 2007 See Abstract; claims 1-9; figures 1,4	1-20
A	JP 2007-065766 A (SONY CORP) 15 March 2007 See Abstract; claims 1; figures 1-4	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 MAY 2010 (31.05.2010)	**31 MAY 2010 (31.05.2010)**

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Jang, Gi Jeong

Telephone No. 82-42-481-5648

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2009/058476

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
JP 2007-087345 A	05.04.2007	CN 1940965 A	04.04.2007
		CN 1940965 C0	04.04.2007
		EP 1768058 A2	28.03.2007
		JP 2007-087346 A	05.04.2007
		KR 10-0886407 B1	02.03.2009
		US 2007-0122036 A1	31.05.2007
JP 2007-034723 A	08.02.2007	CA 2616460 A1	01.02.2007
		CN 101228552 A	23.07.2008
		CN 101228552 C0	23.07.2008
		EP 1909228 A1	09.04.2008
		EP 1909228 A4	01.07.2009
		JP 04-410732 B2	20.11.2009
		KR 10-2008-0031031 A	07.04.2008
		US 2009-0041357 A1	12.02.2009
		WO 2007-013529 A1	01.02.2007
JP 2007-265367 A	11.10.2007	US 2007-0230797 A1	04.10.2007
JP 2007-065766 A	15.03.2007	CN 1924897 A	07.03.2007
		CN 1924897 C0	07.03.2007
		EP 1760635 A1	07.03.2007
		EP 1760635 B1	27.01.2010
		KR 10-2007-0026080 A	08.03.2007
		US 2007-0047775 A1	01.03.2007