G06V

IMAGE OR VIDEO RECOGNITION OR UNDERSTANDING

Definition statement

This place covers:

Higher-level interpretation and recognition of images or videos, which includes pattern recognition, pattern learning and semantic interpretation as fundamental aspects. These aspects involve the detection, categorisation, identification, authentication of image or video patterns. For this purpose, image or video data are acquired and preprocessed. In the next step, distinctive features are extracted. Based on these features or representations derived from them, matching, clustering or classification is performed which may lead to one or several decisions, related confidence values (e.g. probabilities), classification or clustering labels. The aim is to find an explanation or to derive a specific meaning.

Pattern recognition or pattern learning in a specific, image or video-related context that includes:

- scene-related patterns and scene-specific elements – group G06V 20/00;
- character recognition or recognising digital ink; document-oriented image-based pattern recognition – group G06V 30/00;
- human-related, animal-related or biometric patterns in image or video data – group G06V 40/00.

Further details are given in the Definition statement of group G06V 10/00. Image or video recognition can be carried out by using electronic means (G06V 10/70) or by using optical means (G06V 10/88).

Typically, a pattern recognition system involves one or more of the following techniques:

Number of samples	Data entities (e.g. image objects) involved; Individual	Data entities (e.g. image objects) involved; Groups (classes)
One data sample	Authentication	Categorisation
Several data samples	Identification	Clustering

Relationships with other classification places

Pattern recognition techniques in general are classified in group G06F 18/00.

Some techniques of image or video understanding performed in the preprocessing step — which start with a bitmap image as an input and derive a non-bitmap representation from it — can also be encountered in general image analysis. If these techniques do not involve one of the functions of image or video pattern authentication, identification, categorisation or clustering, classification should be made only in the appropriate subgroups of subclass G06T.

Some examples of these techniques are: general methods for image segmentation, e.g. obtaining contiguous image regions with similar pixels, for position and size determination of an object without establishing its identity, for calculating the motion of an image region corresponding to an object irrespective as to the identity of the object, for camera calibration, etc.

Techniques based on coding, decoding, compressing or decompressing digital video signals using video object coding are classified in group H04N 19/20.

Velocity or trajectory determination systems or sense-of-movement determination systems using radar, sonar or lidar are classified in groups G01S 13/58, G01S 15/58, G01S 17/58, respectively. Radar, sonar or lidar systems specially adapted for mapping or imaging are classified in groups G01S 13/89, G01S 15/89, G01S 17/89.

General purpose image data processing, in particular image watermarking, is classified in group G06T 1/00, while selective content distribution, such as generation or processing of protective or descriptive data associated with content involving watermarking is covered by group H04N 21/8358. General purpose image data acquisition and related pre-processing using digital cameras, and processing used to control digital cameras is classified in group H04N 5/00. Play-back, editing or synchronising of a music score, including interpretation therefor, as well as transmission of a music score between systems of musical instruments for play-back, editing or synchronising is classified in subclass G10H.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Detecting, measuring and recording for medical diagnostic purposes	A61B 5/00
Identifications of persons in medical applications	A61B 5/117
Sorting of mail or documents using means for detection of the destination	B07C 3/10
Input arrangements for interaction between user and computer	G06F 3/01
Testing to determine the identity or genuineness of paper currency or similar valuable papers or for segregating those which are unacceptable, e.g. banknotes that are alien to a currency	G07D 7/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Programme-controlled manipulators	B25J 9/00
Optical viewing arrangements in vehicles	B60R 1/00
Photogrammetry or videogrammetry, e.g. stereogrammetry; Photographic surveying	G01C 11/00
Testing balance of machines or structures	G01M
Investigating or analysing materials by determining their chemical or physical properties	G01N
Radio direction-finding; Radio navigation; Determining distance or velocity by use of radio waves; Locating or presence-detecting by use of the reflection or reradiation of radio waves; Analogous arrangements using other waves	G01S
Geophysics	G01V
Optical elements, systems or apparatus	G02B
Photomechanical production of textured or patterned surfaces, e.g. for printing, for processing of semiconductor devices	G03F
Control or regulating systems in general	G05B
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements	G06F 3/00
Comparing digital values in methods or arrangements for processing data by operating upon the order or content of the data handled	G06F 7/02
Content-based image retrieval	G06F 16/50

Fourier, Walsh or analogous domain transformations in digital computers	G06F 17/14
Security arrangements for protecting computer systems against unauthorised activity	G06F 21/00
Authentication, i.e. establishing the identity or authorisation of security principals	G06F 21/30
Computer-aided design [CAD]	G06F 30/00
Handling natural language data	G06F 40/00
Methods or arrangements for sensing record carriers	G06K 7/00
Record carriers for use with machines and with at least a part designed to carry digital markings	G06K 19/00
Computer systems based on specific computational models	G06N
Data processing for business purposes, logistics, stock management	G06Q
General purpose image data processing, e.g. specific image analysis processor architectures or configurations	G06T 1/00
Geometric image transformation in the plane of the image, e.g. rotation of a whole image or part thereof	G06T 3/00
Image enhancement or restoration	G06T 5/00
Image analysis in general	G06T 7/00
Motion image analysis using feature-based methods	G06T 7/246
Image analysis using feature-based methods for determination of transform parameters for the alignment of images	G06T 7/33
Image analysis of texture	G06T 7/40
Image analysis for depth or shape recovery	G06T 7/50
Image analysis using feature-based methods for determining position and orientation of objects	G06T 7/73
Image analysis for determination of colour characteristics	G06T 7/90
Image coding	G06T 9/00
Image contour coding, e.g. using detection of edges	G06T 9/20
Two-dimensional [2D] image generation	G06T 11/00
Three-dimensional [3D] image rendering	G06T 15/00
Lighting effects in 3D image rendering	G06T 15/50
Three-dimensional [3D] modelling for computer graphics	G06T 17/00
Manipulating 3D models or images for computer graphics	G06T 19/00
Checking-devices for individual registration on entry or exit	G07C 9/00
Burglar, theft or intruder alarms using image scanning and comparing means	G08B 13/194
Traffic control systems for road vehicles	G08G 1/00
Labels, tag tickets or similar identification or indication means	G09F 3/00
Speech recognition	G10L 15/00
Speaker recognition	G10L 17/00
Bioinformatics	G16B
Chemoinformatics and computational material science	G16C

Healthcare informatics	G16H
Semiconductor devices	H01L
Arrangements for secret or secure communications; Network security protocols	H04L 9/00
Scanning, transmission or reproduction of documents, e.g. facsimile transmission	H04N 1/00
Studio circuitry for television systems	H04N 5/222
Closed circuit television systems	H04N 7/18
Methods or arrangements for coding, decoding, compressing or decompressing digital video signals using video object coding	H04N 19/20
Methods or arrangements for coding, decoding, compressing or decompressing digital video signals, region motion estimation for predictive coding	H04N 19/543

Special rules of classification

Pattern recognition or pattern learning techniques for image or video understanding involving feature extraction or matching, clustering or classification should be classified in groups G06V 10/40 or G06V 10/70 irrespective whether an application-related context provided by the groups G06V 20/00 - G06V 40/00 exists.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

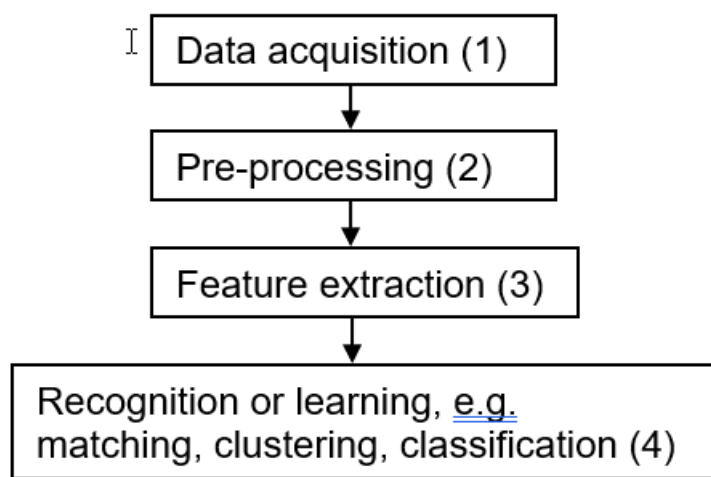| authentication	verifying the identity of a sample using a test of genuineness by undertaking a one-to-one comparison with the genuine (authentic) sample
categorisation	assigning a sample to a class according to certain distinguishing properties (or characteristics) of that class; it generally involves a one-to-many test in which one data sample is compared with the characteristics of several classes.
classification	assigning labels to patterns
clustering	grouping or separating samples in groups or classes according to their (dis)similarity or closeness. It generally involves many-to-many comparisons using a (dis)similarity measure or a distance function.
feature extraction	deriving descriptive or quantitative measures from data.
identification	in the context of collecting of samples, identification means selecting a particular sample having a (predefined) characteristic which distinguishes it from the others. Several samples are generally matched against the one to be identified in a many-to-one process.
image and video understanding	techniques for semantic interpretation, pattern recognition or pattern learning specifically applied to images and videos
pattern	data having characteristic regularity, or a representation derived from it, having some explanatory value or a meaning, e.g. an object depicted in an image

G06V 10/00

Arrangements for image or video recognition or understanding (character recognition in images or video G06V 30/10)

Definition statement

This place covers:

The functions performed at each step in the operation of an image or video recognition or understanding system.

These steps include:

Data acquisition (1)
↓
Pre-processing (2)
↓
Feature extraction (3)
↓
Recognition or learning, e.g. matching, clustering, classification (4)

Processing steps involved in a pattern recognition or understanding system

Classification of each of these steps may be made in groups as follows:
- G06V 10/10 – Image acquisition;
- G06V 10/20 – Image pre-processing;
- G06V 10/40 – Extraction of image or video features;
- G06V 10/70 – Arrangements for image recognition using pattern recognition or machine learning, e.g. matching, clustering or classification.

References

Limiting references

This place does not cover:

Character recognition in images or videos	G06V 30/10

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Image or video recognition or understanding of scene-related patterns and scene-specific elements	G06V 20/00
Image or video recognition or understanding of human-related, animal-related or biometric patterns in image or video data	G06V 40/00
Detecting, measuring and recording for medical diagnostic purposes	A61B 5/00
Identifications of persons in medical applications	A61B 5/117

Sorting of mail or documents using means for detecting the destination	B07C 3/10
Input arrangements for interaction between user and computer	G06F 3/01
Checking-devices for individual registration on entry or exit	G07C 9/00
Testing to determine the identity or genuineness of paper currency or similar valuable papers	G07D 7/00
Burglar, theft or intruder alarms using image scanning and comparing means	G08B 13/194
Traffic control systems for road vehicles	G08G 1/00
Scanning, transmission or reproduction of documents, e.g. facsimile transmission	H04N 1/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Optical viewing arrangements in vehicles	B60R 1/00
Investigating or analysing materials by determining their chemical or physical properties	G01N
Radio direction-finding; Radio navigation; Determining distance or velocity by use of radio waves; Locating or presence-detecting by use of the reflection or reradiation of radio waves; Analogous arrangements using other waves	G01S
Geophysics	G01V
Optical elements, systems or apparatus	G02B
Photomechanical production of textured or patterned surfaces, e.g. for printing, for processing of semiconductor devices	G03F
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements	G06F 3/00
Content-based image retrieval	G06F 16/50
Fourier, Walsh or analogous domain transformations	G06F 17/14
Security arrangements for protecting computer systems against unauthorised activity	G06F 21/00
User authentication in security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/31
Computer-aided design	G06F 30/00
Handling natural language data	G06F 40/00
Computer systems based on specific computational models	G06N
General purpose image data processing, e.g. specific image analysis processor architectures or configurations	G06T 1/00
Geometric image transformation in the plane of the image, e.g. rotation of a whole image or part thereof	G06T 3/00
Image enhancement or restoration	G06T 5/00
Image analysis in general	G06T 7/00
Motion image analysis using feature-based methods	G06T 7/246

Image analysis using feature-based methods for determination of transform parameters for the alignment of images	G06T 7/33
Image analysis of texture	G06T 7/40
Image analysis for depth or shape recovery	G06T 7/50
Image analysis using feature-based methods for determining position and orientation of objects	G06T 7/73
Image analysis for determination of colour characteristics	G06T 7/90
Image coding	G06T 9/00
Image contour coding, e.g. using detection of edges	G06T 9/20
Two-dimensional image generation	G06T 11/00
Three-dimensional [3D] image rendering	G06T 15/00
Lighting effects in 3D image rendering	G06T 15/50
Three-dimensional [3D] modelling for computer graphics	G06T 17/00
Manipulating 3D models or images for computer graphics	G06T 19/00
Bioinformatics	G16B
Chemoinformatics and computational material science	G16C
Healthcare informatics	G16H
Secret or secure communication	H04L 9/00
Studio circuitry for television systems	H04N 5/222
Closed circuit television systems	H04N 7/18
Methods or arrangements for coding, decoding, compressing or decompressing digital video signals using video object coding	H04N 19/20
Methods or arrangements for coding, decoding, compressing or decompressing digital video signals, region motion estimation for predictive coding	H04N 19/543

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

classification	assigning category labels to patterns
clustering	grouping or separating samples in groups or classes according to their (dis)similarity or closeness. It generally involves many-to-many comparisons using a (dis)similarity measure or a distance function.
feature extraction	deriving descriptive or quantitative measures from data
image and video understanding	techniques for semantic interpretation, pattern recognition or pattern learning specifically applied to images and videos
pattern	data having characteristic regularity, or a representation derived from it, having some explanatory value or a meaning, e.g. an object depicted in an image
pattern recognition	detection, categorisation, authentication and identification of patterns for explanatory purposes or to derive a certain meaning in images or video data, by acquiring, pre-processing or extracting distinctive features and matching, clustering or classifying these features or representations thereof

G06V 10/10

Image acquisition (document image scanning and transmission H04N 1/00; control of digital cameras H04N 23/60)

Definition statement

This place covers:

The process of acquiring still images or video sequences for the purpose of subsequently recognising patterns in the acquired images.

Image capturing arrangements which visually emphasise those features of the objects that are relevant to the pattern recognition process.

Optimising the image capturing conditions, such as correctly placing the object with respect to a camera, choosing the right moment for triggering the image sensor, or suitably setting the parameters of the image sensor.

Devices for image acquisition including sensors that generate a conventional two-dimensional image irrespective of its nature (e.g. grey level image, colour image, infrared image, etc.), a three-dimensional point cloud, a sequence of temporally-related images or a video.

Notes – other classification places

Constructional details of the image acquisition arrangements are covered by a hierarchy of subgroups branching from group G06V 10/12:

- Group G06V 10/14 covers the design of the optical path, including the light source (if any), the different optical elements such as lenses, prisms, mirrors, apertures/diaphragms, filters, the individual optical characteristics of these elements (e.g. refraction indices, focal lengths, chromatic aberrations or distortions) and their optical arrangement;

- Group G06V 10/141 covers the control of the illumination, e.g. strategies for activating additional light sources if the ambient lighting is insufficient for a reliable pattern recognition or if individual facets of an object are obstructed by shadows;

- Group G06V 10/143 covers processes or devices which emit or sense radiation in different parts of the electromagnetic spectrum (e.g. infrared light, the visual spectrum and ultraviolet light) so as to obtain a comprehensive set of sensor readings, which when combined facilitate an automated distinction of different kinds of objects. For example, an infrared image could be used for isolating living bodies from the background to analyse the presence of a living body in a second image modality, like an RGB image, the second image being aligned with the infrared image. The images captured in infrared could be used for night vision, e.g. detecting pedestrians or animals for collision avoidance. Sensors using multiple wavelengths are also typically used in remote sensing (e.g. when detecting different kinds of crops, forests, lakes, rivers or urban areas in multispectral or hyperspectral satellite imagery; see also group G06V 20/13);

- Group G06V 10/145 covers illumination arrangements which are specially adapted to increase the reliability of the pattern recognition process. For example, mitigating shadow artefacts, which are likely to deteriorate the pattern recognition process, by providing specially designed arrangements of light sources (light domes, softboxes, ring flashes, etc). The pattern recognition process can also be supported by means of a structured light projector, which projects specific patterns (e.g. stripes or fringe patterns) onto the object so as to augment the two-dimensional image data with three-dimensional information and for this purpose, additional optical elements such as gratings or filter masks may be added to the illumination system. These various special illumination arrangements are also commonly used for recognising patterns in microscopic imagery (see also group G06V 20/69);

- Group G06V 10/147 covers technical details of the image sensor, such as the sensor technology (photodiodes, CCD, CMOS, etc.), the size and the geometrical distribution of light receiving elements on the sensor surface, or the presence of additional optical elements on the sensor (e.g. micro-lenses, diaphragms, collimators or coded aperture masks).
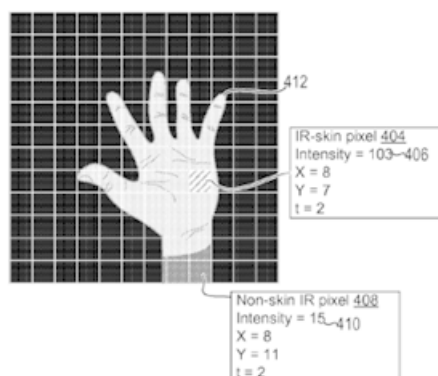
Illustrative examples of subject matter classified in this place:

1.
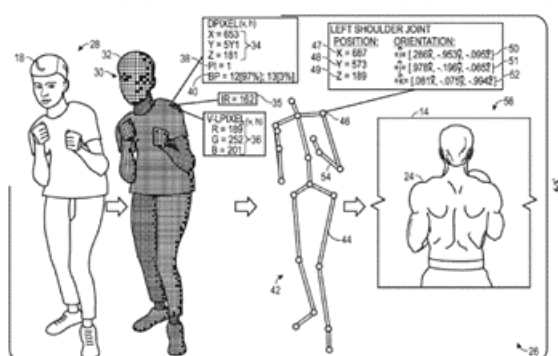
2.

Illumination by casting infrared (IR) light onto a person to highlight regions of a hand, to assist in gesture recognition.

Relationships with other classification places

CCTV and image transmission systems are classified in group H04N 7/00.

References

Limiting references

This place does not cover:

Image acquisition in photocopiers or fax machines	H04N 1/00
Controlling digital cameras	H04N 23/60

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Image acquisition arrangements specifically designed for optical character recognition	G06V 30/14
Image acquisition arrangements specifically designed for fingerprint or palmprint sensors	G06V 40/13

Image acquisition arrangements specifically designed for vascular sensors	G06V 40/145
Image acquisition arrangements specifically designed for taking pictures of the eye	G06V 40/19

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognising patterns in satellite imagery	G06V 20/13
Recognition of microscopic objects in scenes	G06V 20/69
Devices for illuminating a surgical field	A61B 90/30
Optical instruments for measuring contours or curvatures	G01B 11/24
Means for illuminating specimens in microscopes	G02B 21/06
Digital video cameras	H04N 23/00
Digital image sensors	H10F 39/12, H10F 39/15, H10F 39/18

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

CCD	charge-coupled device
CMOS	complementary metal-oxide-semiconductor
visible light	light as seen by the eye, typically in the range 400 – 750 nm
IR	infrared, wavelength longer than those of visible light, typically in the range 750 nm - 1 mm
LIDAR	light detection and ranging, optical range sensing method, which targets a laser at objects and generates a three-dimensional representation (a point cloud)
NIR	near-infrared, typically having wavelengths between 750 nm - 2,5 μm.
UAV	unmanned aerial vehicle, a drone
UV	ultraviolet light, wavelengths shorter than that of visible light, but longer than X-rays, typically having a range of 10 nm - 400 nm
X-rays	electromagnetic radiation in the range 10 pm – 10 nm

G06V 10/20

Image preprocessing

Definition statement

This place covers:

Any kind of processing of acquired image or video data before the steps of feature extraction and recognition; devices configured to perform this processing.

Processing to prepare an image for feature extraction.

Processing to enhance image quality with the intent to emphasise structures in the image, which inform the automated recognition of objects or categories of objects.

Processing to attenuate or discard elements of the image, which are unlikely to be useful for the pattern recognition process.

Processing converts image to a standard format suitable for feature extraction and pattern recognition routines.

Notes – other classification places

Specific aspects of pre-processing are covered by the subgroups of group G06V 10/20; they particularly relate to aspects such as:

- Processes or devices for identifying regions of the image, which should be subjected to the pattern recognition process, or which are likely to contain image information that is relevant for an object recognition task – covered by group G06V 10/22;
- Correcting wrongly oriented images (e.g. changing the orientation from an erroneous portrait mode to landscape mode), compensating for the pose change of the object by performing affine transformations (translation, scaling, homothety, similarity, reflection, rotation, shear mapping and compositions of them in any combination and sequence), or for correcting geometrical distortions induced by the image capturing – covered by group G06V 10/24;
- Determination of a bounding box containing the pattern of interest, processing within a region-of-interest [ROI] or volume-of-interest [VOI] to emphasise the pattern for recognition – covered by group G06V 10/25;
- Devices or processes for separating a candidate object from other, non-interesting image regions or the background; image segmentation to the extent that it is adapted to support a subsequent recognition step – covered by group G06V 10/26;
- Adjusting the bit depth, e.g. conversion to black-and-white images, and setting thresholds therefor, e.g. by analysis of the histogram of the image grey levels; Converting the image data to a predetermined numerical range, e.g. by scaling pixel values – covered by group G06V 10/28;
- Techniques for improving the signal-to-noise [SNR] ratio or denoising the image for the purpose of improving the recognition – covered by group G06V 10/30;
- Adjusting the size or the resolution of the image to a standard format, e.g. by scaling; adjusting the size of the detected object to a certain format – covered by group G06V 10/32;
- Smoothing or thinning to obtain an alternative, less complex representation of the pattern; applying morphological operators (e.g. morphological dilation, erosion, opening or closing) for filling in gaps or merging elements, with the aim of emphasising the structures relevant for recognition; skeleton extraction for characterising the shape of a pattern – covered by group G06V 10/34;
- Enhancing the contrast by convolving the image with a filter mask or by applying a non-linear operator to local image patches – covered by group G06V 10/36.

Illustrative example of subject matter classified in this place:
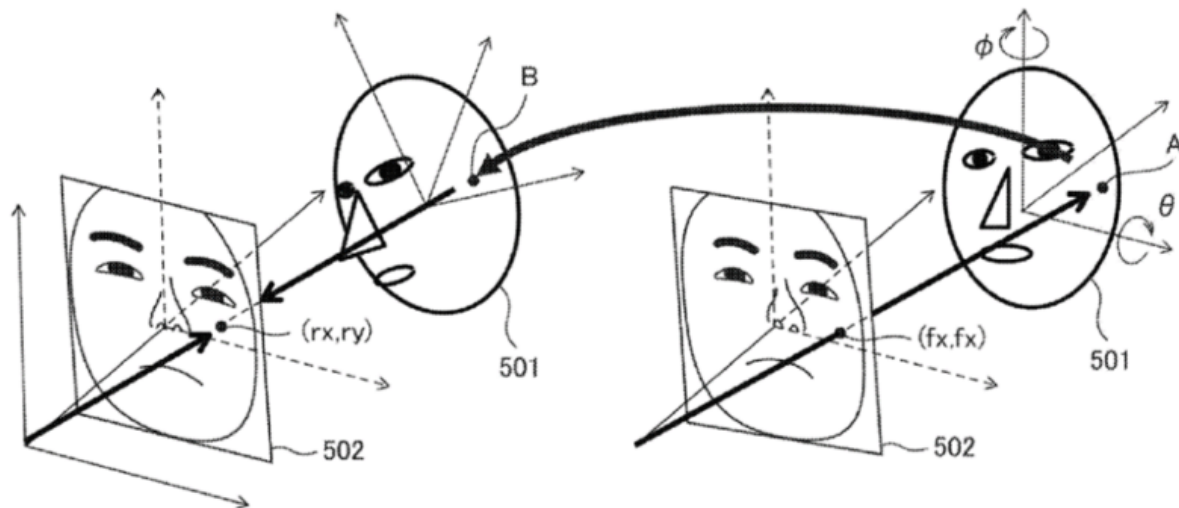
Alignment of the image of a face by affine transformations to obtain a pose-invariant image.

Relationships with other classification places

Different image pre-processing in general is covered in groups as follows:

- G06T 3/00 when geometric image transformations (e.g. image rotation) are involved;
- G06T 5/00 when image enhancement or restoration (e.g. denoising) is performed.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Recognising scenes; Scene-specific elements	G06V 20/00
Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition	G06V 30/00
Image or video recognition or understanding of human-related, animal-related or biometric patterns in image or video data	G06V 40/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Filter operations to reveal edges, corners or other image features, which are used to characterise objects	G06V 10/44
Image enhancement or restoration	G06T 5/00
Image segmentation	G06T 7/10
Morphological operators for image segmentation	G06T 7/155

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

DCT	discrete cosine transform
FFT	fast Fourier transform
FOV	field of view, the region of the environment that an image sensor observes
ROI	region of interest, an image patch that is likely to contain relevant information
skeletonisation	process of shrinking a shape to a connected sequence of lines, which are equidistant to the boundaries of the shape
SNR	signal-to-noise ratio
VOI	volume of interest, a cuboid that encloses three-dimensional data points that are likely to represent relevant information

G06V 10/22

by selection of a specific region containing or referencing a pattern; Locating or processing of specific regions to guide the detection or recognition

Definition statement

This place covers:

Guiding a pattern recognition process or device to a specific region of an image where the pattern recognition algorithm is to be applied, e.g. using fiducial markers.

The use of reference points in images, e.g. patterns having unique combinations of colours or other image properties, which make them useful for guiding a pattern recognition process.

Illustrative examples of subject matter classified in this place:

1.
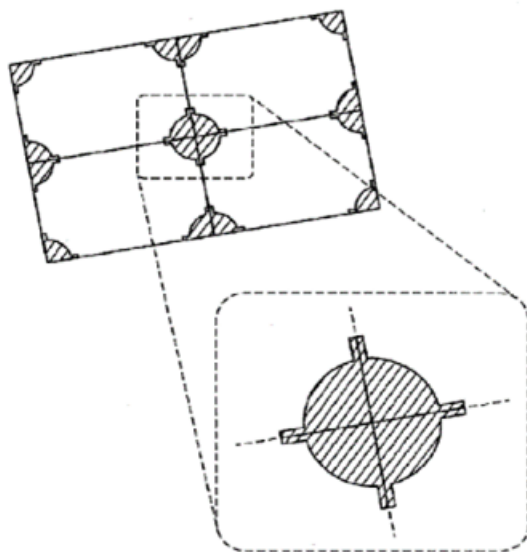
A fiducial marker placed in the centre of an object is used for its object detection and recognition.

2.

view 0   view 1   · · ·   view 7
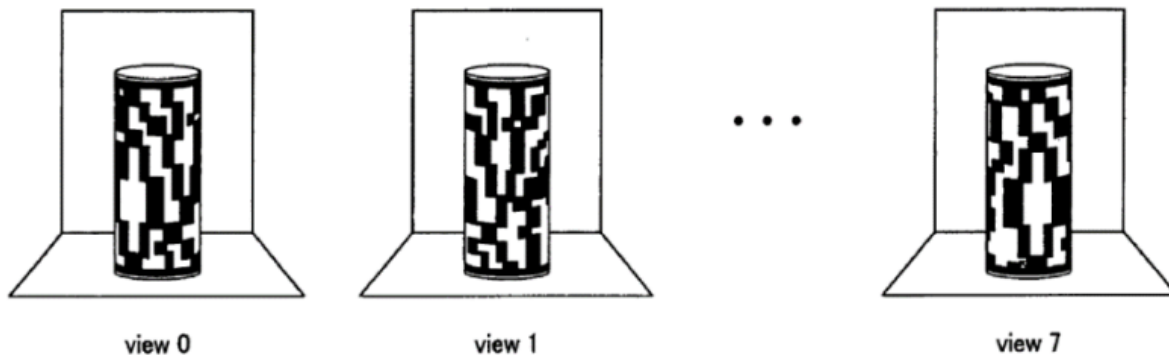
3.
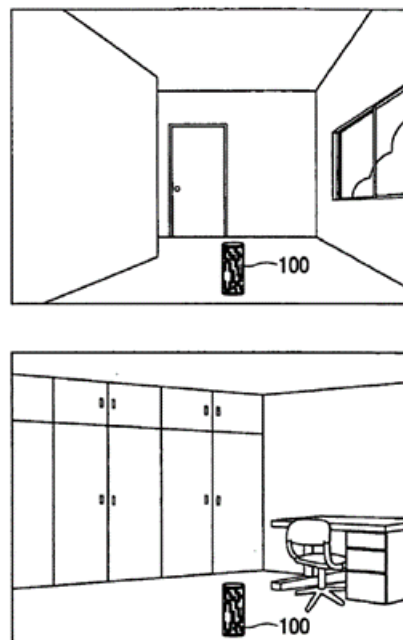
A pattern present on a marker gives additional information about the scene to be recognised.

Relationships with other classification places

Determination of position or orientation of image objects using fiducial markers is covered by group G06T 7/70.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Devices for tracking or guiding surgical instruments	A61B 34/20
Fiducial marks and measuring scales in optical systems	G02B 27/32
Marks applied to semiconductor devices	H01L 23/544

Attention is drawn to the following places, which may be of interest for search:

Aligning, centring, orientation detection or correction of the image	G06V 10/24
Image pre-processing for image or video recognition or understanding involving the determination of region of interest [ROI] or volume of interest [VOI]	G06V 10/25
Image analysis for determining position or orientation of objects or cameras	G06T 7/70

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

AR	augmented reality
ARTag	fiducial marker system based on ARToolKit
ARToolKit	open-source software library for augmented reality
FFT	fast Fourier transform
fiducial marker	an image element which is explicitly designed for serving as a visual landmark point. A fiducial marker can be as simple as a set of lines forming crosshairs or a rectangle, but it can also be a more elaborate pattern such as an augmented reality tag, which additionally conveys information encoded as a two-dimensional barcode. Fiducial markers generally provide information about the position and, often, the orientation or the three-dimensional arrangement of objects in images. Additionally, they can comprise unique identifiers to support the recognition process. Fiducial markers are designed for being easily distinguishable from other image elements; therefore, they commonly have sharp image contrasts (e.g. by limiting their colours to black and white), and they are often designed to generate sharp peaks in the frequency space, allowing them to be easily recognisable by a two-dimensional Fourier transform. Commonly known fiducial markers are those defined by the augmented reality toolkit (ARToolKit).

G06V 10/24

Aligning, centring, orientation detection or correction of the image

Definition statement

This place covers:

Methods or arrangements for aligning or centring the image pattern so that it meets the requirements for successfully recognising it; for example, adjusting the camera's field of view such that a face, a person or another object of interest is located at the centre of the image.

Adjusting the field of view such that the object is entirely visible without any parts of the object extending beyond the boundaries of the image
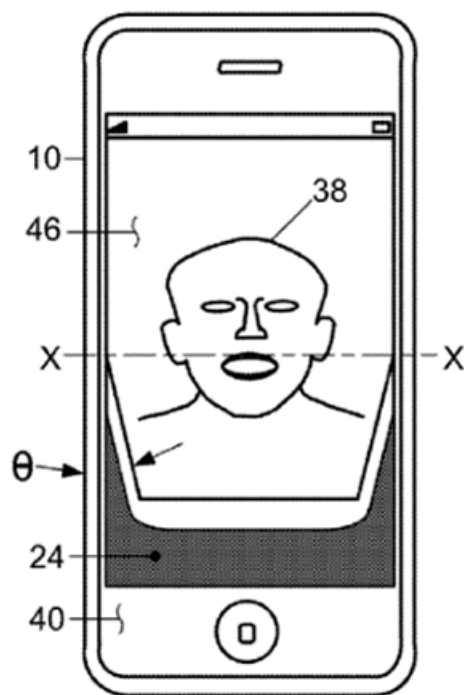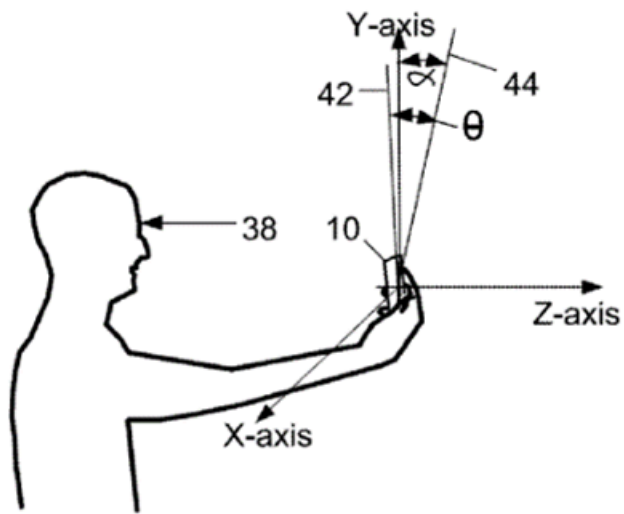
Correcting the image alignment by changing from landscape to portrait mode.

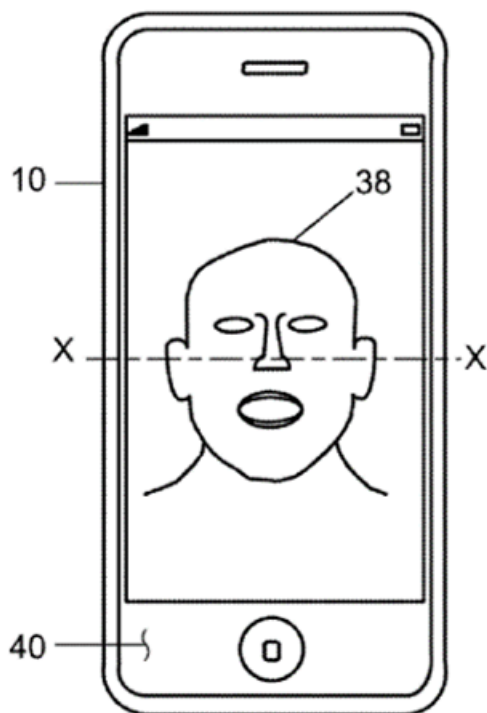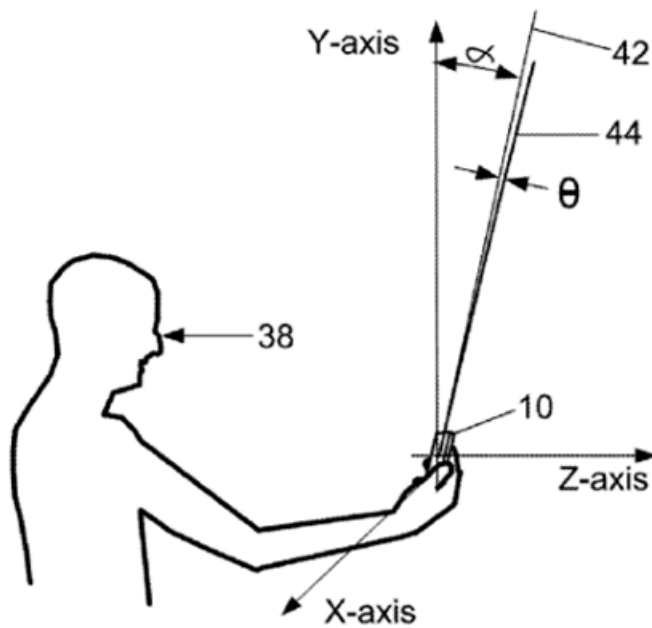Detecting or correcting images that were flipped upside-down or left-right.

Compensating for image skew.

Illustrative examples of subject matter classified in this place:

1A.

1B.

Compensation for the tilt angle of a face captured by a mobile phone by aligning the image.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image pre-processing by selection of a specific region containing or referencing a pattern; Locating or processing of specific regions to guide the detection or recognition	G06V 10/22
Image pre-processing for image or video recognition or understanding involving the determination; Determining of region of interest [ROI] or volume of interest [VOI]	G06V 10/25

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

FOV	field of view, the region of the environment that an image sensor observes

G06V 10/25

Determination of region of interest [ROI] or a volume of interest [VOI]

Definition statement

This place covers:

Methods or arrangements for identifying regions in two-dimensional images, or volumes in three-dimensional point cloud data sets, which contain information relevant for recognition.

Identifying regions or volumes of interest in an image, point cloud or distance map which are likely to lead to successful object recognition.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

A region or volume of interest [RoI or VoI] could include, for example, a human face (in case of a CCTV system), a vehicle or a pedestrian (in case of a camera-based traffic monitoring system), an obstacle on the road (in case of an advanced driver assistance system) or an item on a conveyor belt (in case of an industrial automation system).

The determination of a region or volume of interest is in essence a task of object detection, that is to say detecting the presence of a particular kind of object in images and localising the object(s).

It is the necessity of localising an object and, in particular, of describing the position and the spatial extent of the object (e.g. by outputting a bounding box around it) that distinguishes "object detection" algorithms from "object recognition" algorithms. This is because an "object detection" algorithm will merely assess whether a given visual object exists at a given image location. It may automatically generate a bounding box (e.g. around weeds in a field of vegetables) without solving the problem of "object classification" (e.g. analysing an image of a weed to determine its species and to output its botanical name).

Algorithms for detecting ROIs or VOIs in video sequences typically use frame differencing or more advanced optical flow methods for detecting moving objects.Algorithms that determine a region or volume of interest [ROI or VOI] may use visual cues to establish the location of a boundary box, e.g. by evaluating features such as colour distributions or local textures.

The determination of a region or volume of interest may be facilitated by using special illumination, such as casting light in a specific direction where an object is to be expected in autonomous driving, or by treating the images of specimens with special staining, as is the case in classification of objects in microscopic imagery.

More recently developed algorithms use neural networks [NN] which integrate object detection and recognition. An example is the region-based convolutional neural network [R-CNN] which uses segmentation algorithms for splitting the image into individual segments to find candidate ROIs, followed by inputting each ROI to a classifier for subsequent object recognition.

Other solutions, such as the you only look once [YOLO], region-proposal networks [RPN] or single shot detector [SSD] networks integrate the ROI detection into the actual object recognition step.

Illustrative example of subject matter classified in this place:
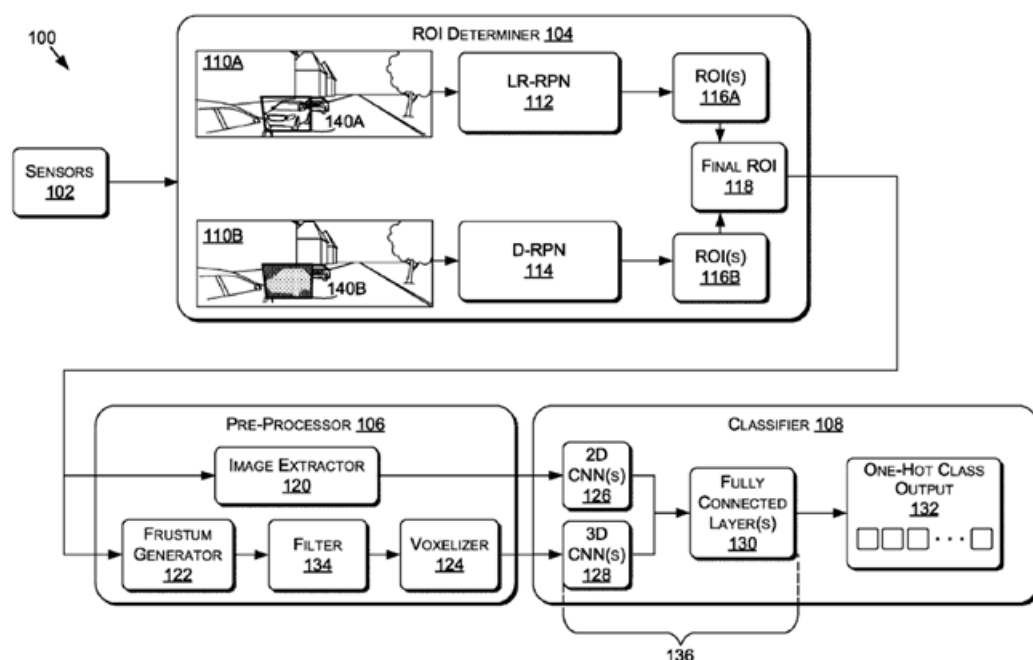
Using a mixed architecture based on region-proposal convolutional networks [R-CNN or RPN] to define a region of interest [ROI] and classifying it by another mixed convolutional neural network [CNN] using 2D and 3D information.

Relationships with other classification places

Determination of a ROI for character recognition is classified in group G06V 30/146.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Devices for radiation diagnosis	A61B 6/00
Diagnostic systems using ultrasound, sound or infrasound	A61B 8/00
Computer-aided diagnosis systems	G16H 50/20

Region-based segmentation image analysis	G06T 7/11

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

AOI	area of interest, synonym for ROI
FOV	field of view, the region of the environment that an image sensor observes
R-CNN	convolutional neural network using a region proposal algorithm for object detection (variants: fast R-CNN, faster R-CNN, cascade R-CNN)
ROI	region of interest, an image region that is likely to contain relevant information concerning an object to be detected and recognised
RPN	region proposal network, an artificial neural network architecture which defines a ROI
SSD	single shot (multibox) detector, a neural network for object detection
VOI	volume of interest, a cuboid that encloses three-dimensional data points that are likely to represent relevant information concerning an object to be detected and recognised
YOLO	you only look once, an artificial neural network used for object detection (comes in various versions: YOLO v2, YOLO v3, etc.).

G06V 10/26

Segmentation of patterns in the image field; Cutting or merging of image elements to establish the pattern region, e.g. clustering-based techniques; Detection of occlusion

Definition statement

This place covers:

Methods and arrangements for segmenting patterns in images or video frames, e.g. segmentation algorithms. Note: segmentation algorithms divide images or video frames into distinct regions, so that boundaries between neighbouring regions coincide with changes of some image properties.

Segmentation algorithms which operate directly on the image by considering the pixel values and their neighbourhood relationships, e.g. mathematical-morphology based algorithms, such as region growing, watershed methods and level-set methods.

Segmentation algorithms which generate a hierarchy of segmentations by starting with a coarse segmentation, which includes only few segments, and successively refine this coarse segmentation by splitting (possibly recursively) the coarse image segments into finer segments (coarse-to-fine approaches).

Graph-cut algorithms such as normalised cuts or min-cut which use graph-based clustering algorithms for image segmentation.

Region growing algorithms which start with few seed points and iteratively expand these into larger regions until some optimality criterion is fulfilled.

The use of classifiers for foreground-background separation. Note: classifiers calculate a score function which expresses a probability (or belief) that a given region of the image is a foreground object or part of the background. The image is then segmented based on these score values.

Deep learning models, in particular different encoder-decoder architectures based on convolutional neural networks [CNN's], applied to semantic image segmentation (a task which requires not only splitting the image into regions, but also consistently assigning labels to image object categories, e.g. "sky", "trees", "road").

Detection of occlusion. Note: Sometimes an object (e.g. a trunk of a tree) partly occludes another object, e.g. a dog behind the tree, which may cause the other object to be split into multiple disjoint segments; occlusion detection algorithms deal with such situations so as to join semantically linked segments into a single segment.

Other algorithms (e.g. some active contour models) which start from an initial image region, which is large enough to surely enclose an object in the image, and they iteratively shrink this region until its boundary is tightly aligned with the contour of the object.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Segmentation algorithms divide images or video frames into distinct regions, so that boundaries between neighbouring regions coincide with changes of some image properties.

Segmentation algorithms may determine regions of homogeneous texture, regions having characteristic colours, regions enclosing individual objects, etc.

Some segmentation algorithms are in essence clustering algorithms. They disregard the spatial arrangement of pixels in the image and compute clusters in a feature space (e.g. by running the k-means algorithm on all colour values in an image). They then group spatially-connected pixels belonging to the same cluster into a region (a "segment").

Illustrative examples of subject matter classified in this place:

1.
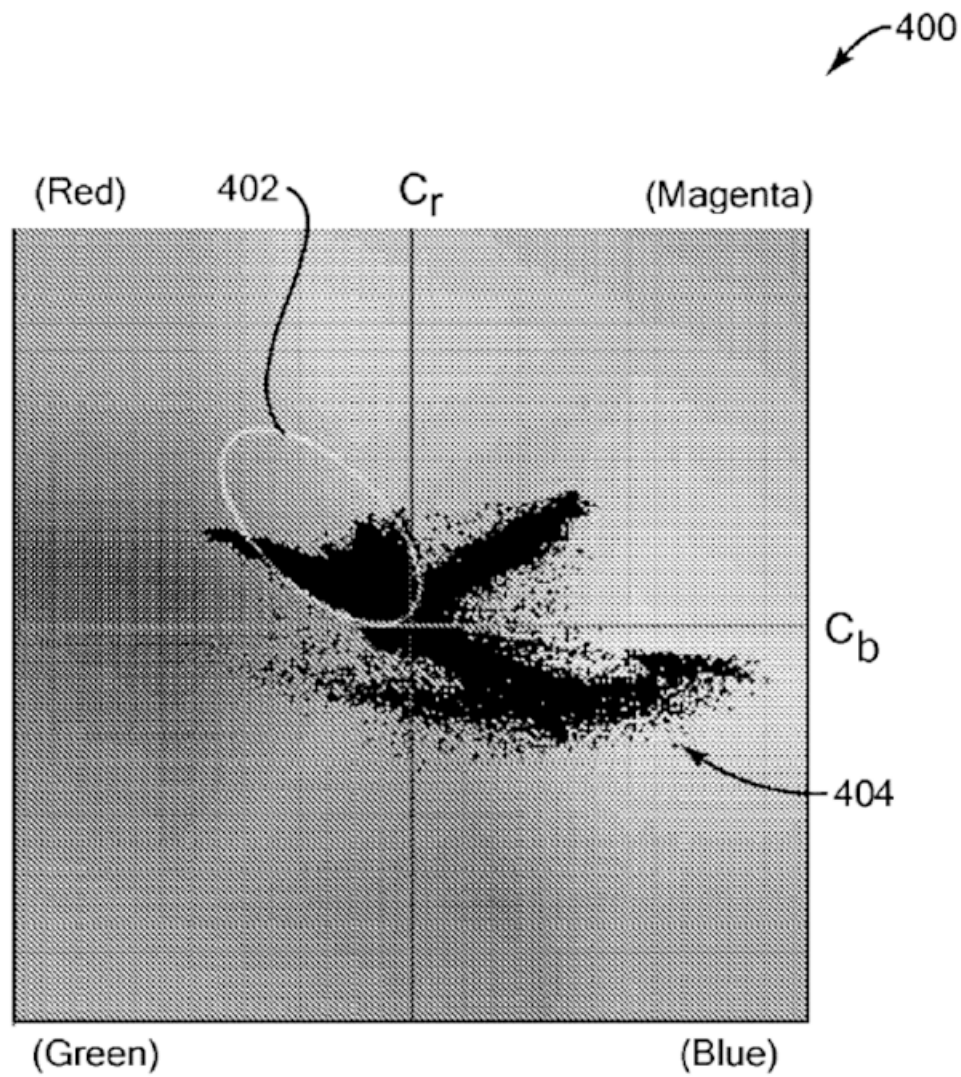
Colour segmentation of a skin region of a face using the clustering in a colour space.

2.

Example of a scene frequently encountered in autonomous driving and its semantic segmentation map with regions such as "road", "sky", "trees", etc.

Relationships with other classification places

Variational methods used for object recognition, such as active contour models [ACM, or "snakes"], active shape models [ASM] or active appearance models [AMM] are classified in group G06V 10/74.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Clustering algorithms for image or video recognition or understanding	G06V 10/762
Image segmentation in general	G06T 7/10
Region-based segmentation image analysis	G06T 7/11
Edge-based image segmentation in general	G06T 7/12
Motion-based image segmentation in general	G06T 7/215

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

BSD	Berkeley segmentation data set, a collection of manually segmented images
K-Means	clustering algorithm
NCUTS	normalised cuts, a graph-based segmentation algorithm
PASCAL VOC	collection of image data sets for evaluating the performance of computer vision algorithms; it includes a dedicated data set for evaluating segmentation algorithms.

G06V 10/28

Quantising the image, e.g. histogram thresholding for discrimination between background and foreground patterns

Definition statement

This place covers:

Methods and arrangements for quantising the image with the effect that the number of possible pixel values does not exceed a predetermined maximum number.

Note: In the limit, this quantisation generates a binary two-tone (black-and-white) image, e.g. an image in which foreground objects appear white and any objects in the background appear black. Quantisation to other numbers of pixel values is also possible.

Quantisation algorithms which calculate a histogram of the grey value distribution and then use one or more thresholds to divide the grey values into different ranges, and then map grey values in the same range to the same target value.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Image quantisation refers to a technique in which the number of possible pixel values is set so that it does not exceed a predetermined maximum number. The source image may be an analogue image, which is quantised for being storable in digital form, or a digital image, which is quantised to a smaller bit depth.

Quantisation may be uniform or non-uniform. The subsequent encoding might adapt the number of bits to encode the representation, using more bits to represent those ranges of grey values which are considered to be particularly relevant for the subsequent image recognition step.

Colour or grey value quantisation can cause artefacts in the resulting image, such as apparent edges or quantisation boundaries which did not exist in the original image (e.g. colour banding). These artefacts can be mitigated by dithering techniques (e.g. by using the Floyd-Steinberg algorithm).
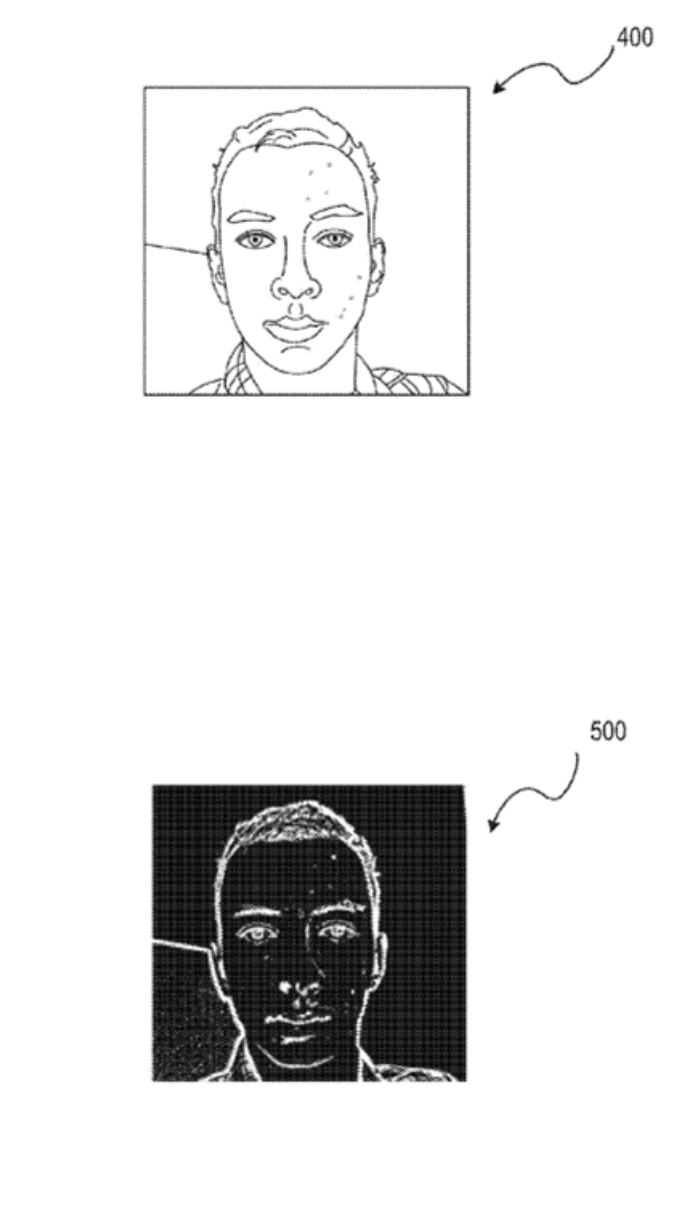
Quantisation may be performed globally (using the histogram of the whole image) or locally (using statistics of local image patches).

Illustrative examples of subject matter classified in this place:

1.


```
300 ─╮
      ╰→    ( START )
                │
                ▼
   ┌─────────────────────────────────────────────┐
   │ IDENTIFYING AN IMAGE CAPTURE MODE FROM A SET  │
   │ OF IMAGE CAPTURE MODES                        │
   │                   302                         │
   └─────────────────────────────────────────────┘
                │
                ▼
   ┌─────────────────────────────────────────────┐
   │ IDENTIFYING A SINGLE CHANNEL ASSOCIATED WITH  │
   │ THE IMAGE CAPTURE MODE                        │
   │                   304                         │
   └─────────────────────────────────────────────┘
                │
                ▼
   ┌─────────────────────────────────────────────┐
   │ ACCESSING A SET OF IMAGES IN A VIDEO STREAM   │
   │                   310                         │
   └─────────────────────────────────────────────┘
                │
                ▼
   ┌─────────────────────────────────────────────┐
   │ CONVERTING ONE OR MORE IMAGES OF THE SET OF   │
   │ IMAGES TO A SET OF SINGLE CHANNEL IMAGES      │
   │                   320                         │
   └─────────────────────────────────────────────┘
                │
                ▼
   ┌─────────────────────────────────────────────┐
   │ GENERATING A SET OF APPROXIMATION IMAGES FROM │
   │ THE SET OF SINGLE CHANNEL IMAGES              │
   │                   330                         │
   └─────────────────────────────────────────────┘
                │
                ▼
   ┌─────────────────────────────────────────────┐
   │ GENERATING A SET OF BINARIZED IMAGES BY       │
   │ THRESHOLDING THE SET OF APPROXIMATION IMAGES  │
   │                   340                         │
   └─────────────────────────────────────────────┘
                │
                ▼
              ( END )
```

2.

Detection of faces in colour images by creating a single-channel image, e.g. a greyscale image, and subsequent binarisation by thresholding.

Relationships with other classification places

Image enhancement or restoration by the use of histograms is covered by group G06T 5/40.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image coding	G06T 9/00
Circuits or arrangements for halftone screening	H04N 1/52
Systems for transmitting or storing colour picture signals	H04N 1/64
Quantisation for adaptive video coding	H04N 19/124

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

bit depth	number of bits, which is available for indicating the grey level or colour of an individual pixel

G06V 10/30

Noise filtering

Definition statement

This place covers:

Techniques for noise removal or filtering such as thresholding in the frequency domain (e.g. after a Fourier or wavelet transform), edge-preserving smoothing techniques such as anisotropic diffusion (also called Perona-Malik diffusion) or deep learning approaches to image denoising, e.g. using convolutional neural networks [CNN's].

Linear smoothing filters (e.g. for convolving the original image with a low-pass filter such as a Gaussian kernel matrix or applying a Wiener filter) and non-linear filtering such as median filtering or bilateral filtering (see also group G06V 10/36), when applied for the purpose of noise removal.

Noise estimation techniques based on a reference image, wherein the reference image may be:
  • a previously captured image which was obtained with the same camera set-up;
  • a previously captured image which was obtained with an optical system of higher quality, potentially downscaled or otherwise converted to match the expected performance parameters of a lower-quality system;
  • artificially generated patterns, obtained, e.g. by blurring or smoothing the original image or by means of computer graphics techniques (e.g. rendered from a 3D model of an object).

Estimation of noise parameters based on different noise models, e.g. additive white Gaussian noise, speckle noise, etc.

Detection of blur or defocusing of the image pattern.

Illustrative examples of subject matter classified in this place:

1.

Face image denoising

2.

Depth: 3 | Depth: 256 | Depth: 256 | Depth: 512 | Depth: 1024 | Depth: 512 | Depth: 256 | Depth: 256 | Depth: 128
Width: 128 | Width: 64 | Width: 32 | Width: 16 | Width: 8 | Width: 16 | Width: 32 | Width: 64 | Width: 128
Height: 160 | Height: 80 | Height: 40 | Height: 20 | Height: 10 | Height: 20 | Height: 40 | Height: 80 | Height: 160
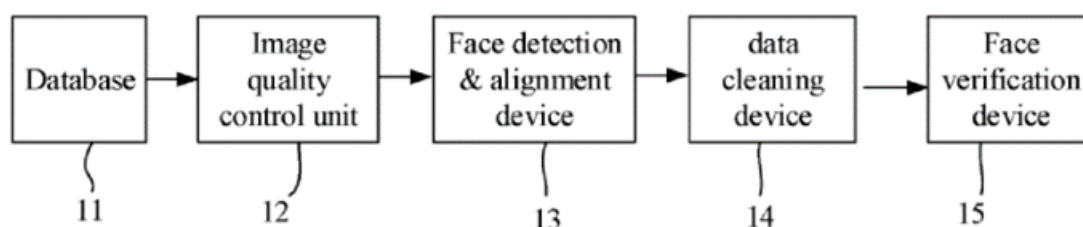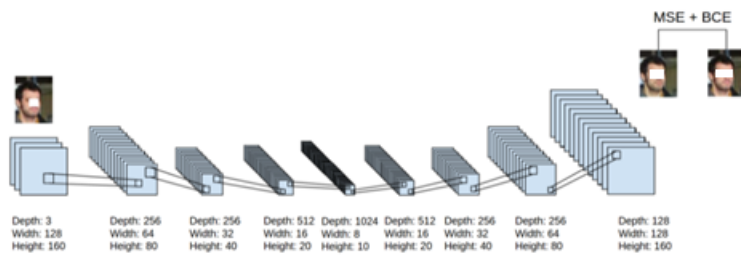
Face denoising using an autoencoder convolutional neural network architecture (above), followed by face recognition using a discriminator architecture (below).

3.
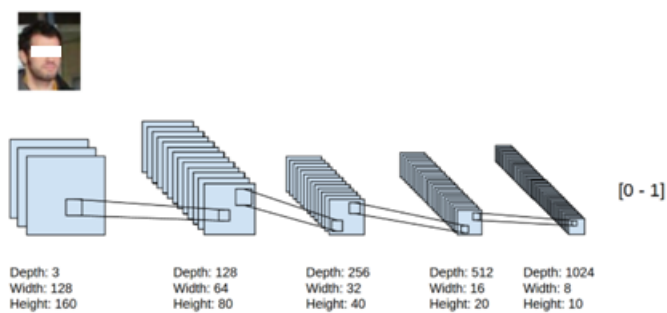
Depth: 3 | Depth: 128 | Depth: 256 | Depth: 512 | Depth: 1024
Width: 128 | Width: 64 | Width: 32 | Width: 16 | Width: 8
Height: 160 | Height: 80 | Height: 40 | Height: 20 | Height: 10

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Aligning, centring, orientation detection or correction for image or video recognition or understanding	G06V 10/24
Segmentation of patterns in the image field	G06V 10/26
Local image operators for image or video recognition or understanding, e.g. median filtering	G06V 10/36
Enhancement or restoration for general image processing	G06T 5/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

DCT	discrete cosine transform
FFT	fast Fourier transform
PDF	probability density function
SNR	signal to noise ratio

G06V 10/32

Normalisation of the pattern dimensions

Definition statement

This place covers:

Processes and devices for bringing image or video data to a standard format, so that it may be compared with reference data (e.g. with images in an image database or gallery images serving as reference templates).

Normalisation or standardisation of the size of images, e.g. by cropping, by reducing the image size via downscaling or sub-sampling, or by enlarging images via up-scaling and interpolation.

Note – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Normalisation can involve adjustments to guarantee that all objects to be recognised have similar size or appearance (e.g. by rescaling facial images so that they are centred and that the area of the face covers a predetermined fraction of the image, or by only selecting frontal images).

Illustrative example of subject matter classified in this place:
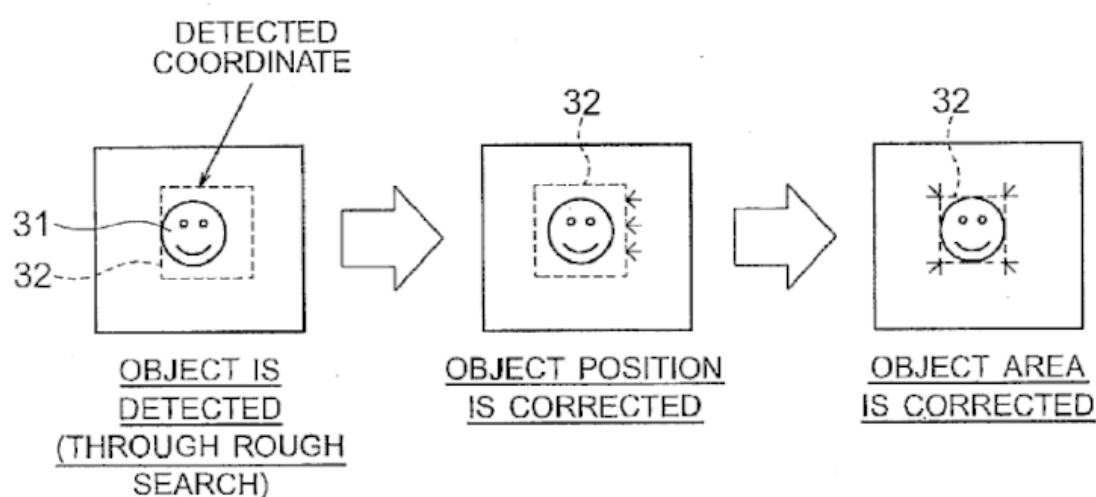
Correction of the region detected for the face image by cropping around the face region.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image enhancement or restoration in general, e.g. dynamic range modification	G06T 5/00

G06V 10/34

Smoothing or thinning of the pattern; Morphological operations; Skeletonisation

Definition statement

This place covers:

Techniques for binary and grayscale morphological analysis of image patterns. These techniques include:

- Basic morphological operators: erosion, dilatation, opening, closing, watershed analysis, etc.;
- Detection of patterns or arrangements of pixels in a binary or a greyscale image, e.g. by using the hit-or-miss transform;
- Finding the outline or contour of a foreground object by morphological processing (morphological edge detection, watershed processing, etc.);
- Finding the skeleton of a foreground object, e.g. by thinning, medial axis transformation, contour-based erosion, etc.;
- Distance transformation.

Illustrative example of subject matter classified in this place:
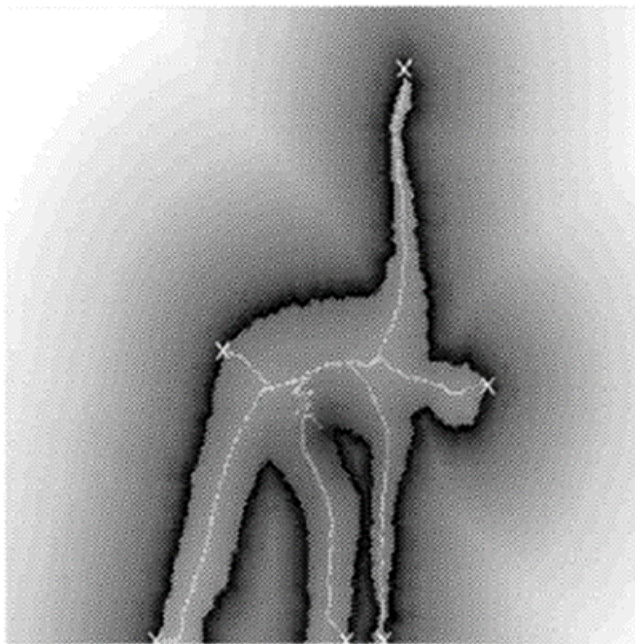
Extraction of the skeleton representation of a human body by applying morphological operations.

Restoration for general image processing	G06T 5/00
Erosion or dilatation, e.g. thinning	G06T 5/30

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

skeletonisation	process of shrinking a shape to a connected sequence of lines, which are equidistant to the boundaries of the shape

G06V 10/36

Applying a local operator, i.e. means to operate on image points situated in the vicinity of a given point; Non-linear local filtering operations, e.g. median filtering

Definition statement

This place covers:

Image or video pre-processing techniques which examine a local neighbourhood around a pixel and assign a value to the pixel, which is a function of the values (e.g. colour values or luminance values) of the pixels in this local neighbourhood.

The application of local operators in the spatial domain (e.g. by convolving the image with a predefined kernel matrix) or in the frequency domain (e.g. by calculating the Fourier transform and performing a point-wise multiplication in the frequency domain).

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

1. Usually the local neighbourhood is defined as a rectangular region of pixels with the pixel of interest placed at its centre, e.g. a 3*3 pixels neighbourhood or a 5*5 pixels neighbourhood; neighbourhoods having other shapes are also possible, but they are less common.

2. Local operators include:
- Linear operators, e.g. convolutions with low-pass filter matrices (such as a Gaussian kernel or a boxcar function; and convolutions with high-pass filter matrices for sharpening the image, or convolutions with spatial band-pass filters (such as the difference of Gaussians filter);
- Non-linear operators, e.g. median filters and more complex operators such as those for evaluating local luminance differences in order to detect sparkle points which are significantly brighter than their immediate surroundings;
- Non-linear operators, e.g. the Sobel operator and the Marr Hildreth operator are also frequently used for emphasising object boundaries or elongated structures in images;
- Differential operators such as the Laplace operator and filter matrices for calculating image gradients.

Notes – other classification places

Use of low-pass filter matrices for noise removal – group G06V 10/30.

Use of median filters for noise removal – group G06V 10/30.

Use of the Sobel operator and the Marr Hildreth operator for edge detection – group G06V 10/44.

Illustrative example of subject matter classified in this place:
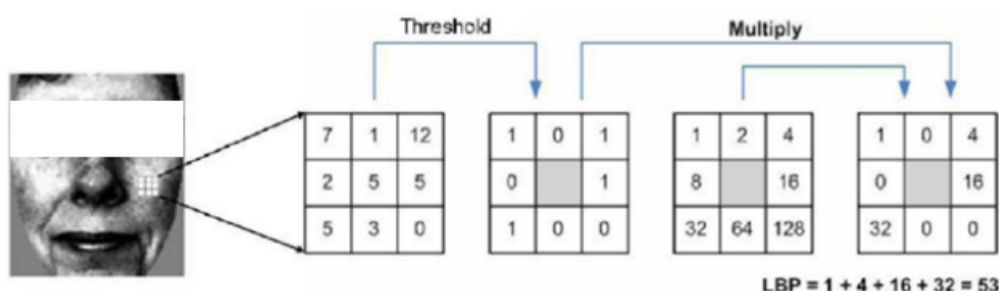
Analysis of local image patches of a face image using a local operator and encoding the representation for subsequent face recognition.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Noise removal for image or video recognition or understanding	G06V 10/30
Detecting edges or corners for image or video recognition or understanding	G06V 10/44
Extracting features from image blocks	G06V 10/50
Local operators for general image enhancement	G06T 5/20

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

BPF	band-pass filter
DCT	discrete cosine transform
DoG	difference of Gaussians
DWT	discrete wavelet transform
FFT	fast Fourier transform
HPF	high-pass filter
Kernel	filter kernel, a matrix which an image is convolved with
LPF	low-pass filter

G06V 10/40

Extraction of image or video features

Definition statement

This place covers:

Methods and arrangements for extracting visual features which are subsequently input to an object recognition algorithm.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Formerly, the choice of suitable feature extraction algorithms was a crucial design choice in the art of pattern recognition algorithms. It had a strong influence of the overall performance. With the advent of deep learning, particularly in convolutional neural networks, the need for the hand-picked design of dedicated feature extraction algorithms has decreased to some extent, because the inner layers of the neural networks are trained to automatically find suitable features from the training data.

Notes – other classification places

Subgroups of group G06V 10/40 focus on specific kinds of feature extraction techniques. These include:

- Features which describe characteristics of the entire image or an entire object (group G06V 10/42);

Note: Global feature extraction techniques often involve domain transformations, such as frequency domain transformation. The global descriptors contain numerical data, such as vectors or matrices, but they can also represent the image or object in an abstract form as a string of symbols from a predetermined alphabet, which are integrated using a grammar (covered by group G06V 10/424).

- Graph structures having vertices and edges (e.g. directed attributed graphs or trees) are another way of representing patterns in images; the vertices of such graph structures represent qualitative or quantitative feature measurements; the edges represent relations between them (covered by group G06V 10/426);
- Local features (covered by group G06V 10/44) build representations of the local image content. Examples of local features include luminance values or colour characteristics, potentially from more than three colour channels, local edges, corners, gradients and texture. Edges can be extracted by convolutions with specially designed filter masks (e.g. Prewitt, Sobel) or by convolutions with a numerical filter, e.g. wavelet filters (Haar, Daubechies), or by difference of Gaussians, Laplacian of Gaussians, Gabor filters etc. Local features such as edges and corners, which can be extracted by applying a pre-defined image operator, are also referred to as low-level features to distinguish them from features such as objects or events, which are extracted using a machine learning algorithm;
- Higher-level features, obtained e.g. by detecting silhouettes of shapes and describing them, e.g. using a chain code, by a Fourier expansion of the contour, by curvature scale-space analysis or by sampling points along object boundaries and quantifying their relative locations;
- Algorithms for evaluating the saliency of local image regions; selecting salient points as key points (covered by group G06V 10/46);
- For the purpose of feature extraction, techniques for converting image or video data to a different parameter space, e.g. using a Hough transform for detecting linear structures in images, or performing a conversion from the spatial domain to the frequency domain or vice versa (group G06V 10/48);
- Techniques for combining individual low-level features into feature vectors by first calculating local statistics of low-level image features in a block of pixels and subsequently generating histograms or deriving other statistical measures in a local neighbourhood (group G06V 10/50);
- Multi-scale feature extraction algorithms for analysing image or video data at different resolutions; scale space analysis, e.g. wavelet decompositions (group G06V 10/52);
- Techniques for describing textures, such as convolution with Gabor wavelets, grey-level co-occurrence matrices or edge histograms (group G06V 10/54);
- Descriptors which capture colour properties of the image, such as colour histograms, possibly after conversion to a suitable colour space (group G06V 10/56);
- Descriptors which are specially designed for more than three colour channels, in particular for hyperspectral images which contain sensor readings in a multitude of different wavelengths not limited to the visual spectrum (group G06V 10/58);
- Descriptors obtained by integrating information about the imaging conditions, such as the position, the orientation and the spectral properties of light sources, diffuse or specular reflections at object surfaces, etc. (group G06V 10/60);
- Temporal descriptors derived from object movements, e.g. optical flow (group G06V 10/62).

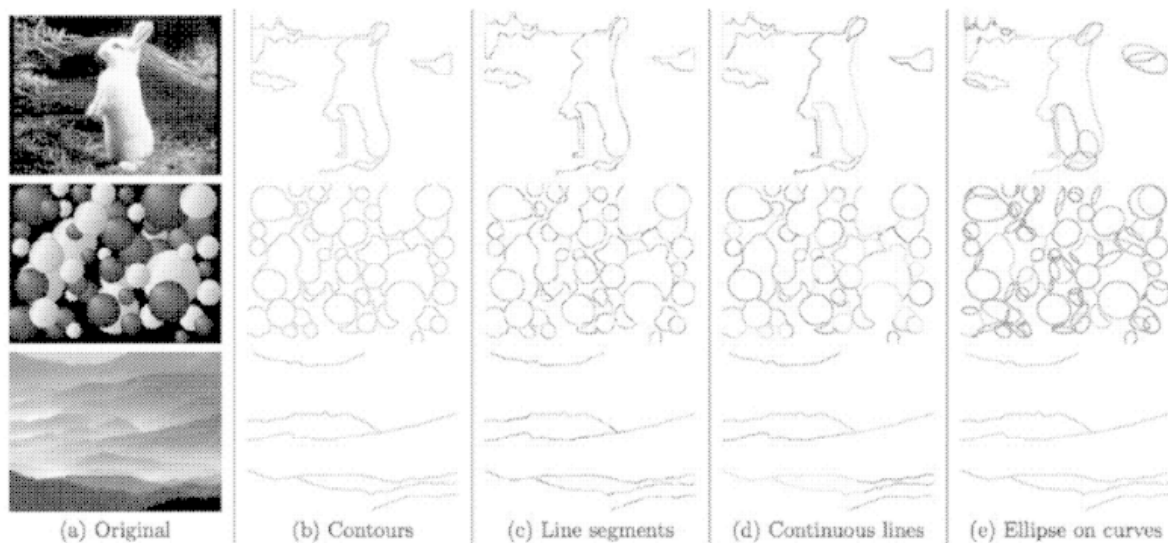Illustrative examples of subject matter classified in this place:

1.
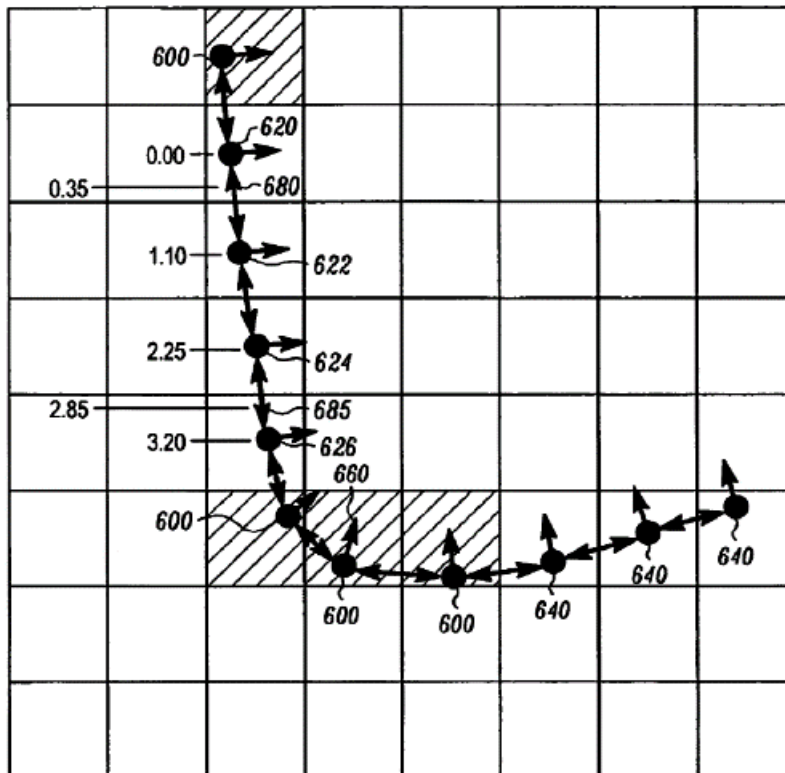
2.
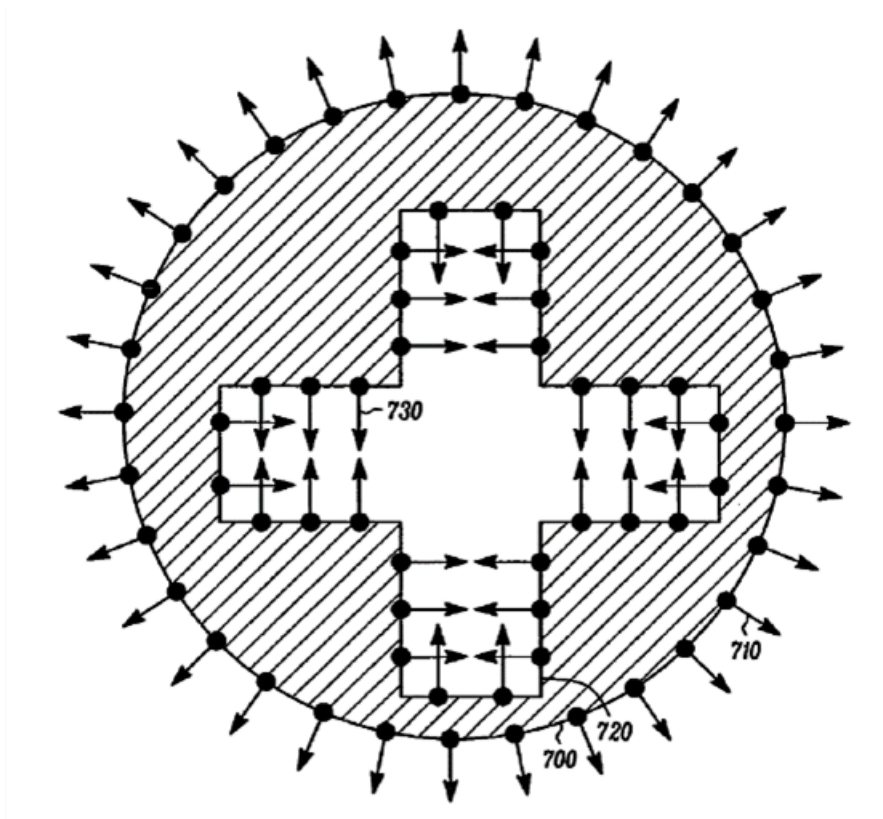
Quantifying local image properties, in particular the local gradient, using a local probe.

3.

(a) Original   (b) Contours   (c) Line segments   (d) Continuous lines   (e) Ellipse on curves

Different types of features used for object recognition, e.g. contours, line segments, continuous lines.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Recognition of scene and scene-specific elements	G06V 20/00
Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition	G06V 30/00
Image or video recognition or understanding of human-related, animal-related or biometric patterns in image or video data	G06V 40/00
Recognition of fingerprints or palmprints	G06V 40/12
Recognition of vascular patterns	G06V 40/14
Recognition of human faces, e.g. facial parts, sketches or expressions within images or video data	G06V 40/16
Recognition of eye characteristics within image or video data, e.g. of the iris	G06V 40/18

Informative references

Attention is drawn to the following places, which may be of interest for search:

Spectrometry, measurement of colour	G01J 3/46
Image analysis using feature-based methods, in particular, for determination of transform parameters for the alignment of images	G06T 7/33
Image analysis for depth or shape recovery	G06T 7/50
Image contour coding, e.g. using detection of edges	G06T 9/20

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

BoW	bag of words, a model originally developed for natural language processing; when applied to images, it represents an image by a histogram of visual words, each visual word representing a specific part of the feature space.
edge	region in the image, at which the image exhibits a strong luminance gradient
GLCM	grey-level co-occurrence matrix
HOG	histogram of oriented gradients, a feature descriptor described by N. Dalal and B. Triggs
SIFT	scale-invariant feature transform, a feature detection algorithm
SURF	speeded up robust features, a feature descriptor

G06V 10/46

Descriptors for shape, contour or point-related descriptors, e.g. scale invariant feature transform [SIFT] or bags of words [BoW]; Salient regional features (colour feature extraction G06V 10/56)

Definition statement

This place covers:

Feature extraction techniques in which additional (invariant) information is calculated from certain image regions or patches or at certain points, which are visually more relevant in the process of comparison or matching.

Feature extraction techniques in which information from multiple local image patches can be combined into a joint descriptor by using an approach called "bag of features" (from its origin in text document matching), "bag of visual features" or "bag of visual words".

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

1. The image regions referred to in this place are called "salient regions", and the points are called "keypoints", "interest points" or "salient points". The information assigned to these regions or points is referred to as a local descriptor due to the inherent aspect of locality in the image analysis.

A local descriptor aims to be invariant to transformations of the depicted image object (e.g., invariant to affine transforms, object deformations or changes in image capturing conditions such as contrast or scene illumination, etc.).

A local descriptor may capture image characteristics across different scales for reliably detecting objects at different sizes, distances or resolutions. Typical descriptors of this kind include:
  • Blob detectors (e.g. SIFT, SURF);
  • Region detectors (e.g. MSER, SuperPixels).

At a salient point, the pixels in its immediate neighbourhood have visual characteristics, which are different from those of the vast majority of the other pixels. The visual appearance of patches around a salient point is, therefore, somewhat unique; this uniqueness increases the chance of finding a similar patch in other images showing the same object.

Generally, salient points can be expected to be located at boundaries of objects and at other image regions having a strong contrast.

2. A "bag of visual words" is a histogram, which indicates the frequencies of patches with particular visual properties; these visual properties are expressed by a codebook, which is commonly obtained by clustering a collection of typical feature descriptors (e.g. SIFT features) in the feature space; each bin of the histogram corresponds to one specific cluster in the codebook.

The process of generating a bag of features typically involves:

A training phase comprising:
 • Extracting local features (e.g. SIFT) from a set of training images;
 • Clustering these features into visual words (e.g. with k-means).

And an operating phase comprising:
 • Extracting local features from a target image;
 • Associating each feature with its closest visual word;
 • Building a histogram of visual words over the whole image and match them with templates using a statistical distance (e.g. Mahalanobis distance).

Illustrative example of subject matter classified in this place:
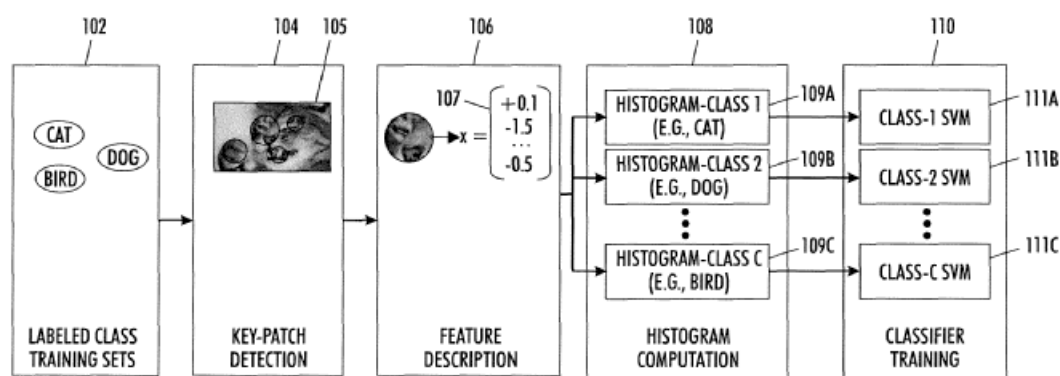
Defining key-patches for different object classes from a training set, computing features from them and using a set of support vector machine [SVM] classifiers to detect those objects in new images.

References

Limiting references

This place does not cover:

Colour feature extraction	G06V 10/56

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding involving the determination of a region or volume of interest [ROI, VOI]	G06V 10/25
Global feature extraction, global invariant features (e.g. GIST)	G06V 10/42
Local feature extraction; Extracting of specific shape primitives, e.g. corners, intersections; Computing saliency maps with interactions such as reinforcement or inhibition	G06V 10/44

Local feature extraction, descriptors computed by performing operations within image blocks (e.g. HOG, LBP)	G06V 10/50
Organisation of the matching process; Coarse-fine approaches, e.g. multi-scale approaches; using context analysis; Selection of dictionaries	G06V 10/75
Obtaining sets of training patterns, e.g. bagging	G06V 10/774
Extracting salient feature points for character recognition	G06V 30/18
Image retrieval systems using metadata	G06F 16/583

Special rules of classification

The present group does not cover biologically inspired approaches of feature extraction based on modelling the receptive fields of visual neurons, such as Gabor filters, and convolutional neural networks [CNN].

The use of neural networks for image or video pattern recognition or understanding is classified in group G06V 10/82.

When a document presents details on a sampling technique and a clustering technique (bagging), then it should also be classified in group G06V 10/774.

Classical "bag of words" techniques remove most image localisation information (geometry).

When local features are matched directly from one image to another without involving a bagging technique (and thereby retaining geometric information), e.g. when triplets of features are matched using a geometric transformation with a RANSAC algorithm, then the document should also be classified in group G06V 10/75.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

BOF	bag of features, see BOW
BOVF	bag of visual features, see BOVF
BOVW	bag of visual words, see BOW
BOW	bag of words, a model originally developed for natural language processing; when applied to images, it represents an image by a histogram of visual words, each visual word representing a specific part of the feature space.
MSER	maximally stable extremal regions, a technique used for blob detection
RANSAC	random sample consensus, a popular regression algorithm
SIFT	scale-invariant feature transform
superpixels	sets of pixels obtained by partitioning a digital image for saliency assessment
SURF	speeded up robust features

G06V 10/48

by mapping characteristic values of the pattern into a parameter space, e.g. Hough transformation

Definition statement

This place covers:

Techniques that map the image space into a parameter space using a transformation, such as the Hough transform.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The object of the transformations classified in this place is to allow better interpretation and increase the separability between the pattern classes. Each dimension of the parameter space may be linked to a specific feature parameter of an object, e.g. its distance from the origin of the image coordinate system and its orientation. The function which performs the mapping to the parameter space may be invertible, i.e. the original representation could be recovered from the representation in the parameter space.

In case of the Hough transform, the parameter space is partitioned into individual bins, which form a so-called accumulator array (a two-dimensional histogram). A voting process maps features in images to individual bins of the accumulator array to finally determine the most probably parameter configuration by retrieving the bin, which has received the maximum bin count.

The generalised Hough transform can be applied for recognising arbitrary shapes, e.g. analytic curves such as lines and circles, or binary or grey-value pattern templates.

Other examples are the generalised Radon transform, the Trace transform, etc.

Illustrative examples of subject matter classified in this place:

1.
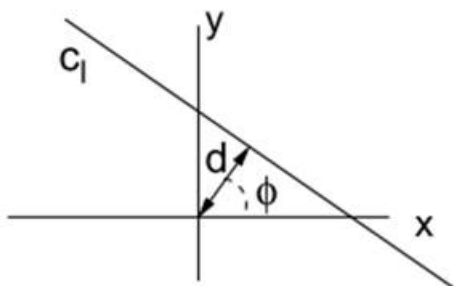
A line in the plane is described by the parameters "d" and "ϕ" (distance to the origin and angle).

2A.

2B.

Rendering of Transform Results
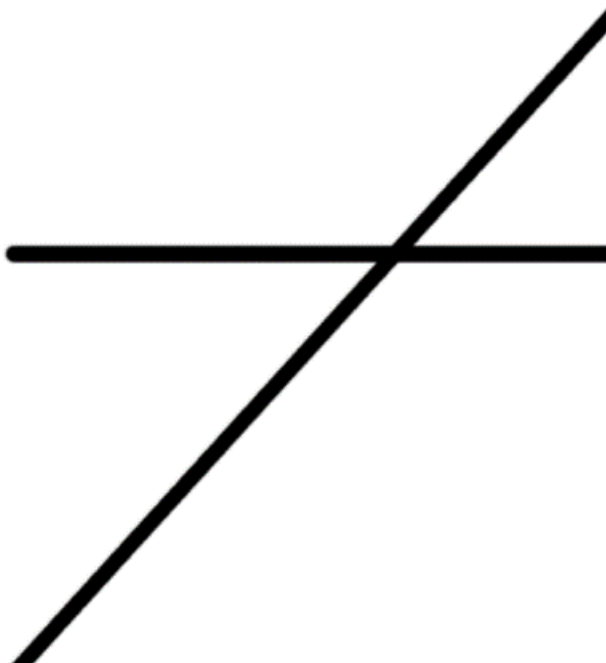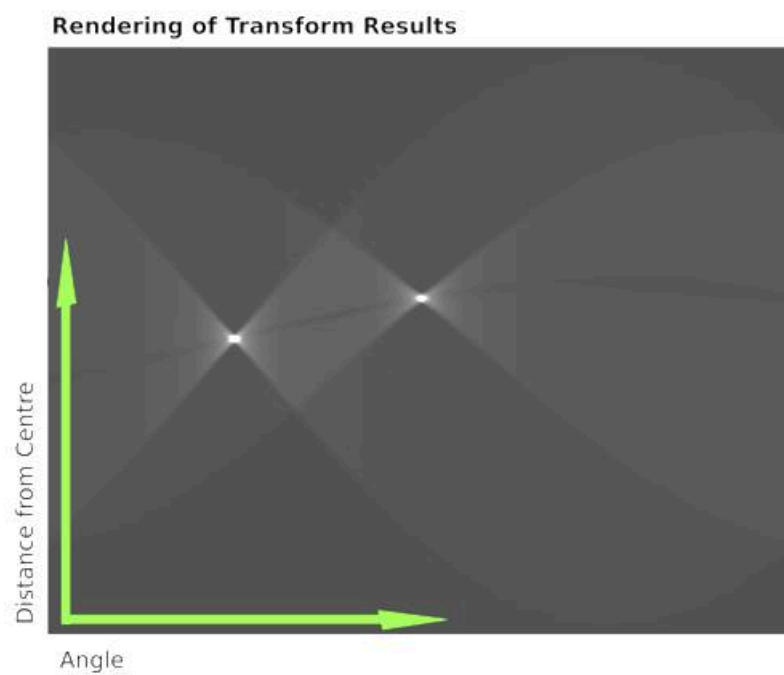
The two lines in the input image (fig. 2A) are mapped by the Hough transform in the parameter space (d,ϕ), and the representation leads to two distinct corresponding bright spots (fig. 2B).

3.

Hough Transform Example

(a) Original Image   (b) Binary Gradient Image

(c) Hough Space   (d) Clusters of Edges
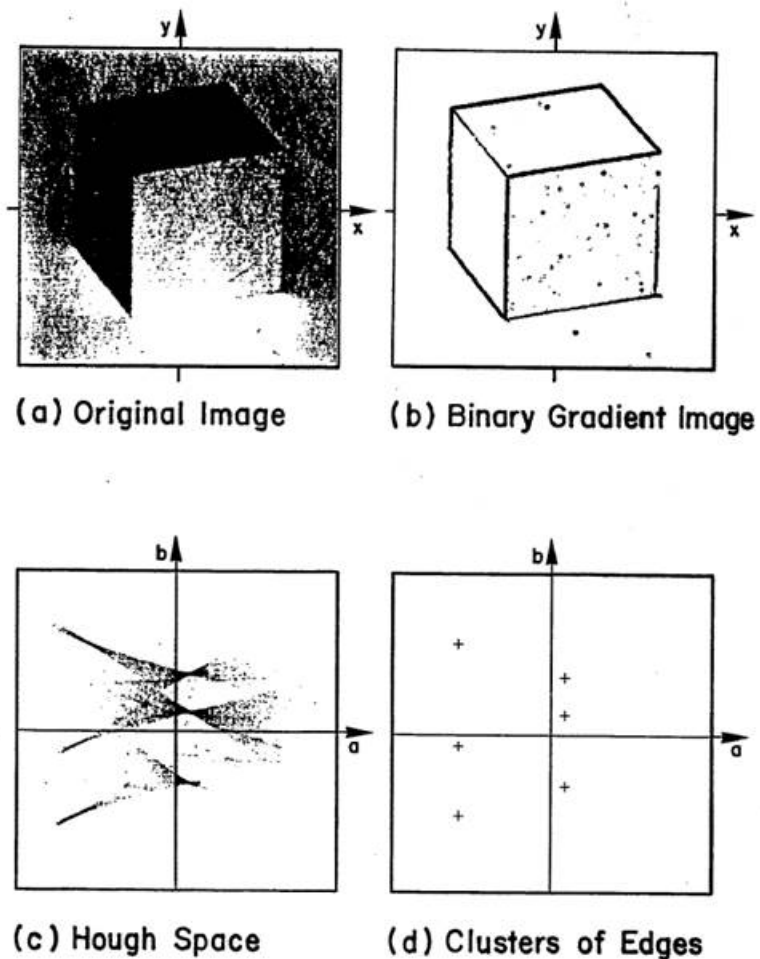
Detection of the visible edges of a cube as points in the Hough parameter space.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern	G06V 10/42
Local feature extraction by analysis of parts of the pattern	G06V 10/44
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Image analysis in general	G06T 7/00

Special rules of classification

Global feature extraction for image or video recognition or understanding is classified in group
G06V 10/42.

Fourier-transform based representations, scale-space representations or wavelet-based representations have a different aim than improving the discriminability in the representation space. The Fourier transform is usually chosen for its geometric invariance properties in the Fourier space (e.g. translation invariance), while the scale-space and wavelet-based representations aim at capturing the variability of the pattern at multiple representation scales. For this reason, the latter two representations are classified as global feature extraction (group G06V 10/42) and, respectively, local feature extraction by scale-space analysis (group G06V 10/52).

G06V 10/50

by performing operations within image blocks; by using histograms, e.g. histogram of oriented gradients [HoG]; by summing image-intensity values; Projection analysis

Definition statement

This place covers:

Feature extraction techniques that perform operations within image blocks or by using histograms.

Summation of image intensity values and projection along an axis, e.g. by binning the values into a histogram, to arrive at a more compact feature representation.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The processing classified in this group might involve:

Block-based arithmetic or logical operations (including non-linear operators such as "max", "min", etc.);

Histograms of various measurements computed on a block-basis, e.g. histogram of oriented gradients [HOG];

Quantification of local geometric arrangements of features by block-based analysis, e.g. local binary patterns [LBP].

The blocks need not necessarily be arranged in a form of a grid. They can overlap or can be arranged in different geometrical patterns.

Frequently used local feature descriptors which are classified in this group include:

- Histogram of oriented gradients [HOG];

- Edge oriented histogram [EOH];

- Local binary pattern [LBP] and its refinements:
  • Local Gabor binary pattern [LGBP];
  • Local edge pattern [LEP];
  • Heat kernel local binary pattern [HKLBP];
  • Oriented local binary pattern [OLBP];
  • Elliptical binary patterns [EBP];
  • Local ternary Patterns [LTP];
  • Probabilistic LBP [PLBP];
  • Elongated quinary patterns [EQP];
  • Thee-patch local binary patterns [TPLBP], four-patch local binary patterns [FPLBP];
  • Local line binary patterns, etc.;

- Shape context;

- Gradient location and orientation histogram [GLOH];

- Local energy-based shape histogram [LESH];

- Oriented histogram of flows [OHF];

- Binary robust independent elementary features [BRIEF];

- Spin image.

Illustrative examples of subject matter classified in this place:

1.
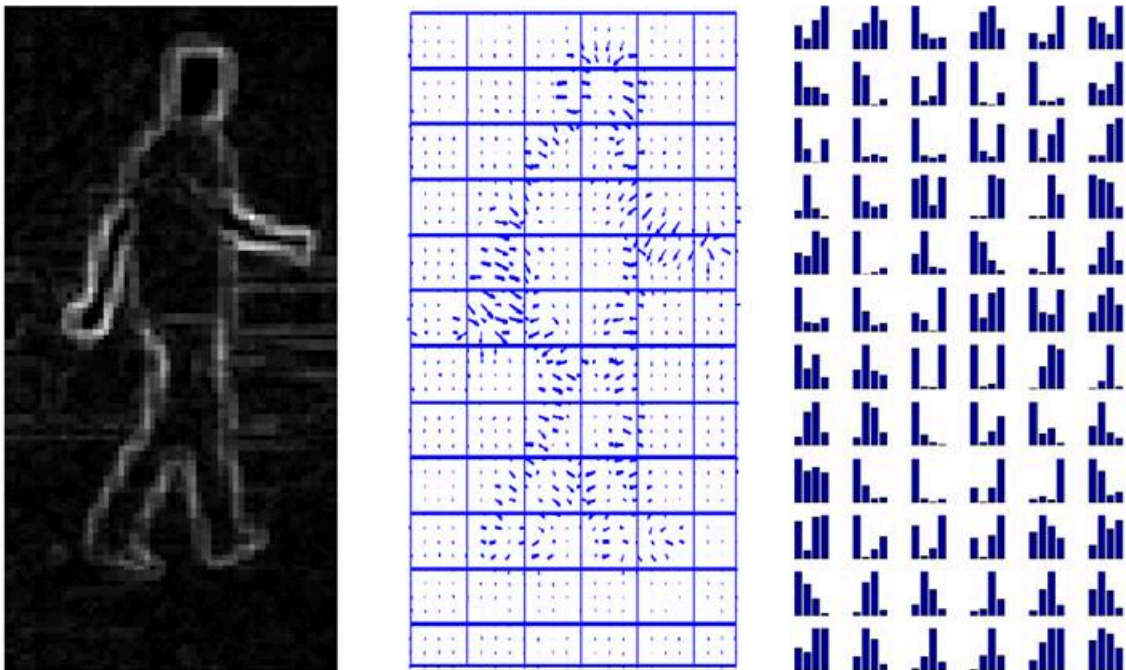
The local oriented histograms of the gradients or HOG descriptor.

2.

(a)   (b)
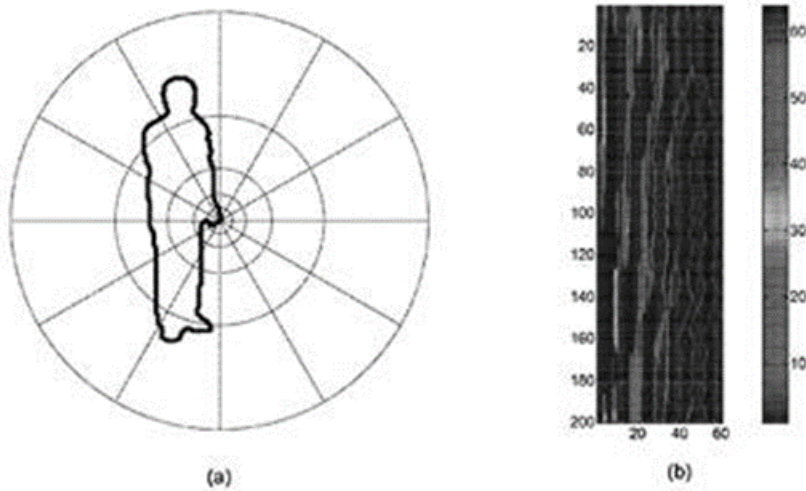
The "shape context", a representation which performs binning of the contours of the shape in a circular-like pattern.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern	G06V 10/42
Local feature extraction by analysis of parts of the pattern	G06V 10/44
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Image analysis in general	G06T 7/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

BRIEF	binary robust independent elementary features
EOH	edge oriented histogram
GLOH	gradient location and orientation histogram
HOG	histogram of oriented gradients
LBP	local binary pattern
LESH	local energy-based shape histogram
OHF	oriented histogram of flows
OLBP	oriented local binary pattern

G06V 10/52

Scale-space analysis, e.g. wavelet analysis (multi-scale boundary representations G06V 10/42)

Definition statement

This place covers:

Scale-space representations which allow analysis of the image or video at multiple scales.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

One primary goal of scale space methods is to achieve scale-invariance, i.e. being able to detect and recognise objects regardless of their size in the image. The scale is usually selected by convolving the image with a parametric "size function", also called a kernel. After the convolution, which typically blurs the fine-scale structures to a certain degree and which is often followed by a suitable sub-sampling of the blurred image, the actual feature extraction can take place at the selected scale.

A very common example of a kernel is the Gaussian kernel:

$$g(x, y; t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}$$

Given the input image f, the scale-space representation is obtained by convolving it with the Gaussian kernel: $L(\cdot, \cdot; t) = g(\cdot, \cdot; t) * f(\cdot, \cdot)$, where t is the scale of analysis.

Scale space approaches can also use Gaussian derivatives, Laplacians of Gaussians, difference of Gaussians (DoG's), Gabor functions, wavelets (in continuous or discrete form, e.g. Haar, Daubechies).

Other alternatives to constructing a scale space which do not use a kernel exist, for instance, by applying to the image a diffusion equation $\partial_t L = \frac{1}{2} \nabla^2 L$, starting with the initial condition $L(x, y; 0) = f(x, y)$. In more general terms, these techniques analyse the differential intrinsic structure of the image in order to construct scale-space representations.

Techniques based on morphological scale-space construct representations at different scales using mathematical morphology methods, e.g. erosion, dilation, opening, closing.

Some other techniques construct multi-scale temporal representations based on the analysis of optical flow for feature extraction in video.

Multi-resolution methods implicitly provide representations at multiple scales; such methods are also classified in the present group in as far as they concern image or video feature extraction.

Illustrative examples of subject matter classified in this place:

1.
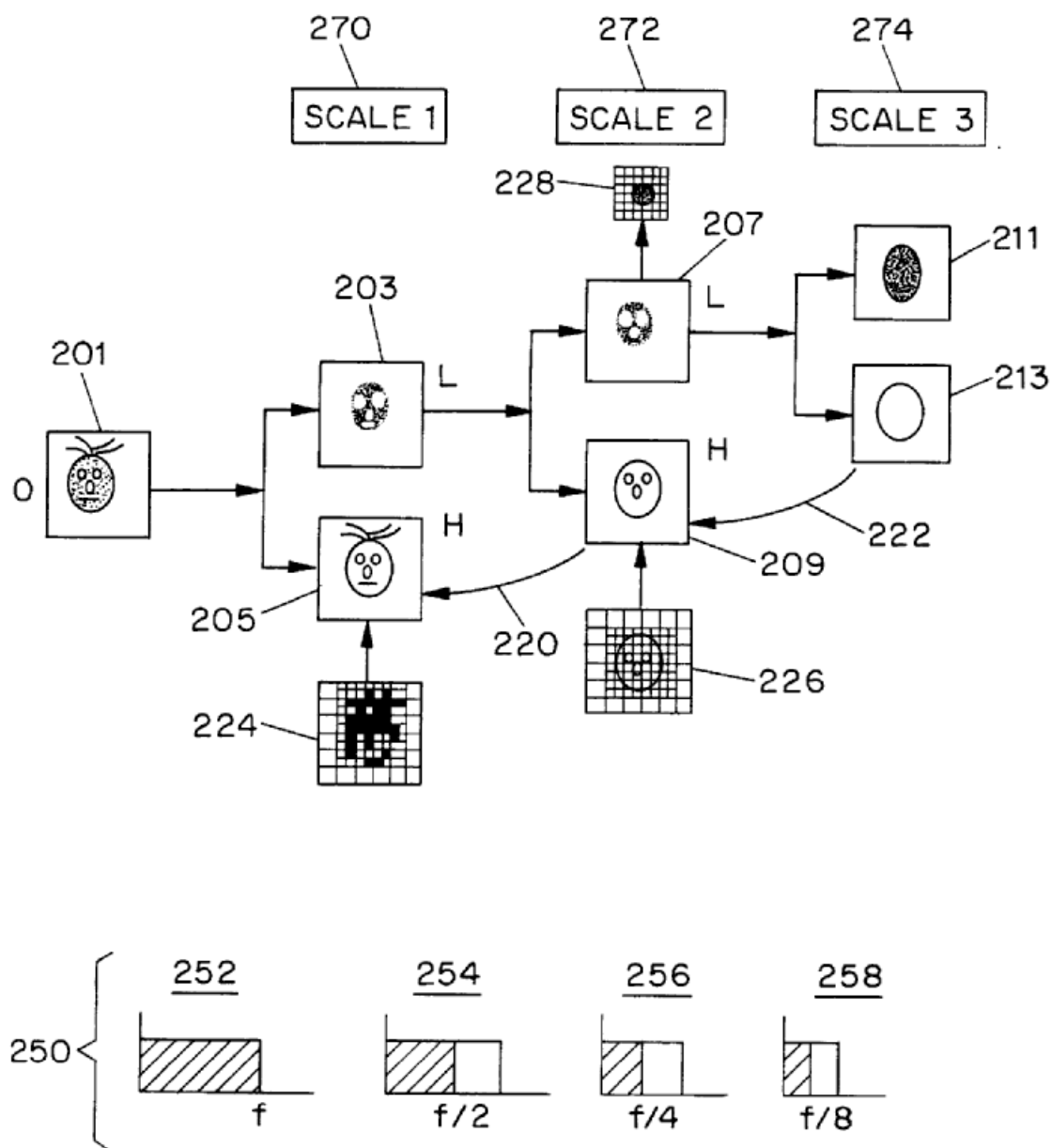
2.
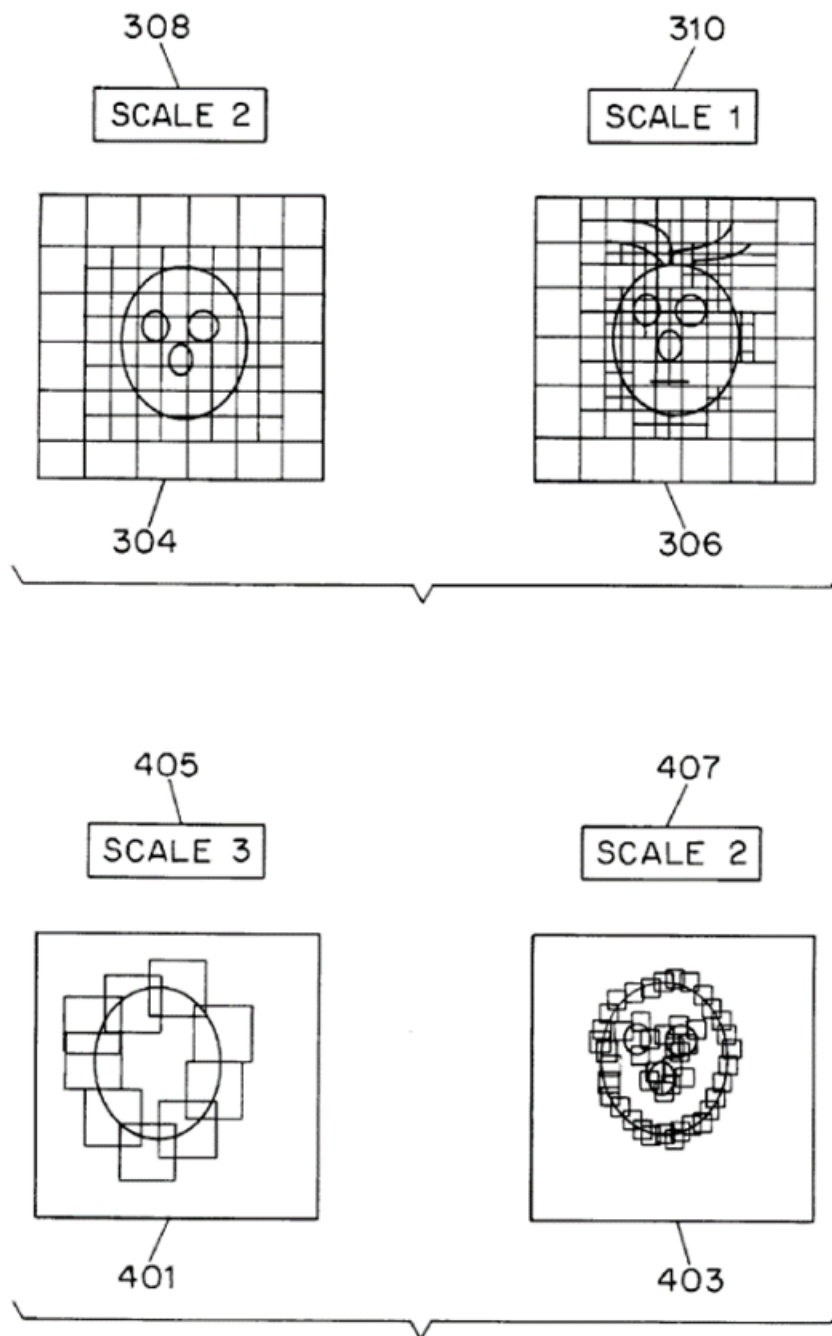
Wavelets applied at different scales for the extraction of facial features.

References

Limiting references

This place does not cover:

Multi-scale boundary representations	G06V 10/42

Attention is drawn to the following places, which may be of interest for search:

Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Image analysis in general	G06T 7/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

CWT	continuous wavelet transform
DoG	difference of Gaussians
DWT	discrete wavelet transform
LoG	Laplacian of Gaussian
Haar wavelets	family of wavelets constructed from rescaled square-shaped functions
steerable filter	class of orientation-selective convolution kernels used for feature extraction that can be expressed via a linear combination of a small set of rotated versions of themselves. As an example, the oriented first derivative of a 2D Gaussian is a steerable filter

G06V 10/54

relating to texture

Definition statement

This place covers:

Texture feature extraction for image or video recognition or understanding, either by identifying the boundaries of texture regions, or by analysing the content of the regions themselves.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Examples of algorithms used for feature extraction include:
- Statistical approaches which characterise the texture by local statistical measures such as "edgeness" (local variation of the image gradient), co-occurrence matrices and Haralick features, Laws texture energy, local histogram-based measures, autocorrelation, power spectrum, etc.;
- Structural approaches based on primitives, morphological operations or representations derived from them, or graph-based methods in which image quantities (e.g. pixels or local patches) are represented as graph nodes and are clustered together using graph-based clustering algorithms (e.g. graph-cuts) to identify texture regions;
- Model-based approaches such as auto-regressive models, fractal models, random fields, texton model;
- Transform methods such as Fourier (spectral) analysis, Gabor filters, wavelets, curvelet transform.

Illustrative example of subject matter classified in this place:

Texture feature extraction allows identification of an animal (zebra) in natural images.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern	G06V 10/42
Local feature extraction by analysis of the parts of the pattern, e.g. by detecting edges, contours, loops, corners, intersections; Connectivity analysis, e.g. connected component analysis	G06V 10/44
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Colour feature extraction	G06V 10/56
Feature extraction related to illumination properties	G06V 10/60
Pattern recognition or image understanding, using clustering	G06V 10/762
Analysis of texture in general	G06T 7/40

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

GLCH	grey-level co-occurrence histogram (synonym of GLCM)
GLCM	grey-level co-occurrence matrix (Haralick invariant texture features)
Texton	basic component of an image that may be recognised visually before the entire image is recognised, and that repeats itself to generate a texture region

G06V 10/56

relating to colour

Definition statement

This place covers:

Colour feature extraction for image or video recognition or understanding.

Colour feature extraction based on colour invariance.

Colour feature extraction based on colour descriptors.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

1. Colour invariance or, conversely, compensation of colour variations, is important for increasing the robustness in image matching or object recognition. Colour variations are often caused by changing lighting conditions (e.g. the colour of an object typically looks different under ambient light or when the object is being illuminated by an incandescent light bulb). They can also be caused by other factors (e.g. sun-tanned skin has a different colour than pale skin).

2. Colour descriptors associate colour information with various image structures such as points, contours or blobs/regions. Colour descriptors (e.g. colour histograms, average colour values etc.) are frequently used in image recognition or image understanding. Typical applications include feature detection based on a model of skin colour, traffic sign detection based on colour information, colour image object detection or video analysis for finding objects with a special colour (e.g. nudity detection).

Illustrative examples of subject matter classified in this place:

1.
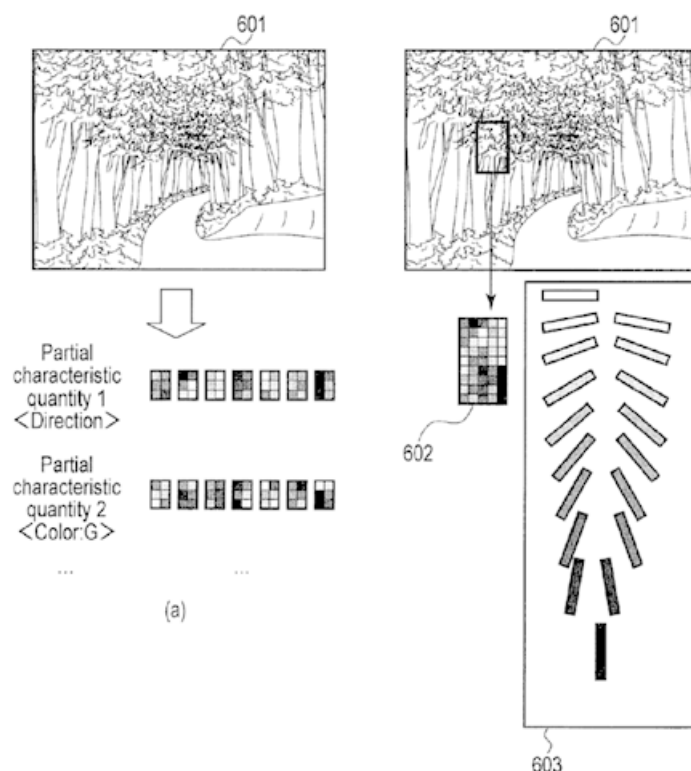
Colour histograms used to detect vegetation in natural scenes.

2.
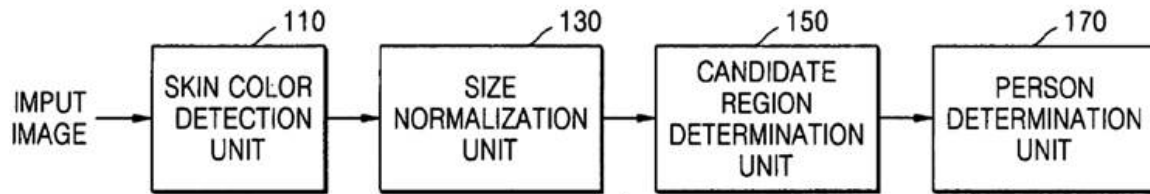
Detection of a person based on his/her skin colour.

3.
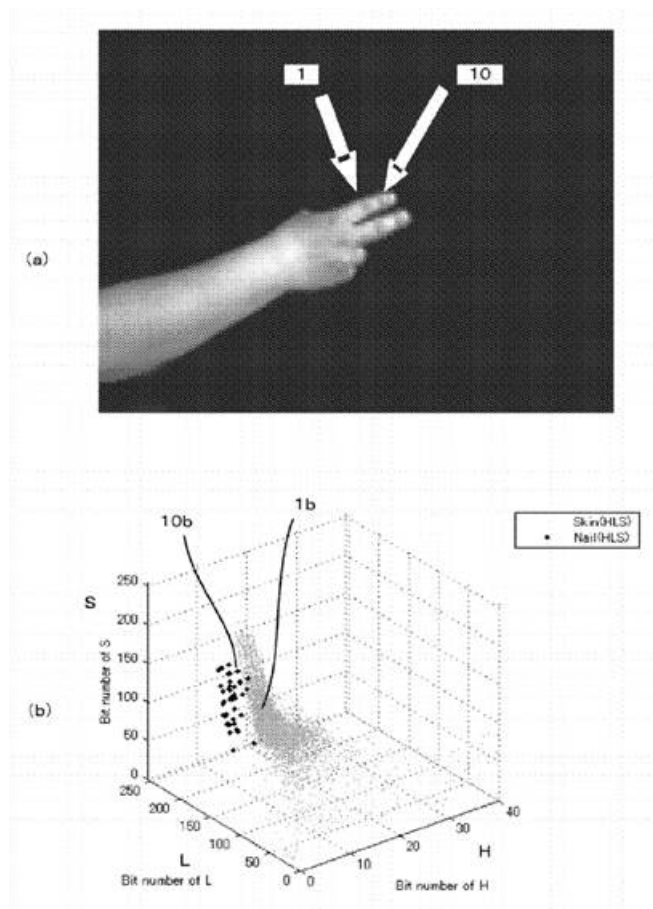
Discrimination between skin image regions and nail image regions by clustering in a three-dimensional colour space.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern	G06V 10/42
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Local feature extraction by performing operations within image blocks or by using histograms	G06V 10/50

Image analysis for determination of colour characteristics	G06T 7/90
Colour picture communication systems	H04N 1/46

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

CIELAB, L*a*b*	colour space representation using a lightness value L*, a value a* on a red-green axis and a value b* on a blue yellow axis; these axes reflect human perception
CMYK	colour space representation using cyan, magenta, yellow and black
HSB	colour space representation using separate channels for hue, saturation and brightness (also called HSV)
HSL	colour space representation using separate channels for hue, saturation and lightness
HSV	colour space representation using separate channels for hue, saturation and value (also called HSB)
RGB	colour space representation using red, green and blue colour channels
YCbCr	colour space representation using separate channels for a luminance component Y, a blue-difference component Cb, and a red-difference component Cr, respectively
YUV	colour space representation using separate channels for a luminance component Y and two chrominance components U and V

G06V 10/58

relating to hyperspectral data

Definition statement

This place covers:

Techniques for feature extraction in hyperspectral image data.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The goal of feature extraction in hyperspectral imaging is to obtain a representation of the relevant features captured by the spectral content of a scene, with the purpose of finding relevant objects and identifying materials. The data can be visualised as a 3D cube, also called a hyperspectral cube, where 2D images corresponding to different spectral wavelengths are superposed. Typical examples of applications are in astronomy, microscopy and satellite image analysis.

Depending on the number of spectral bands, one often distinguishes between multispectral imaging (e.g. 3 to 15 bands) and hyperspectral imaging (often several hundred spectral bands). Group G06V 10/58 encompasses both alternatives.

Illustrative example of subject matter classified in this place:
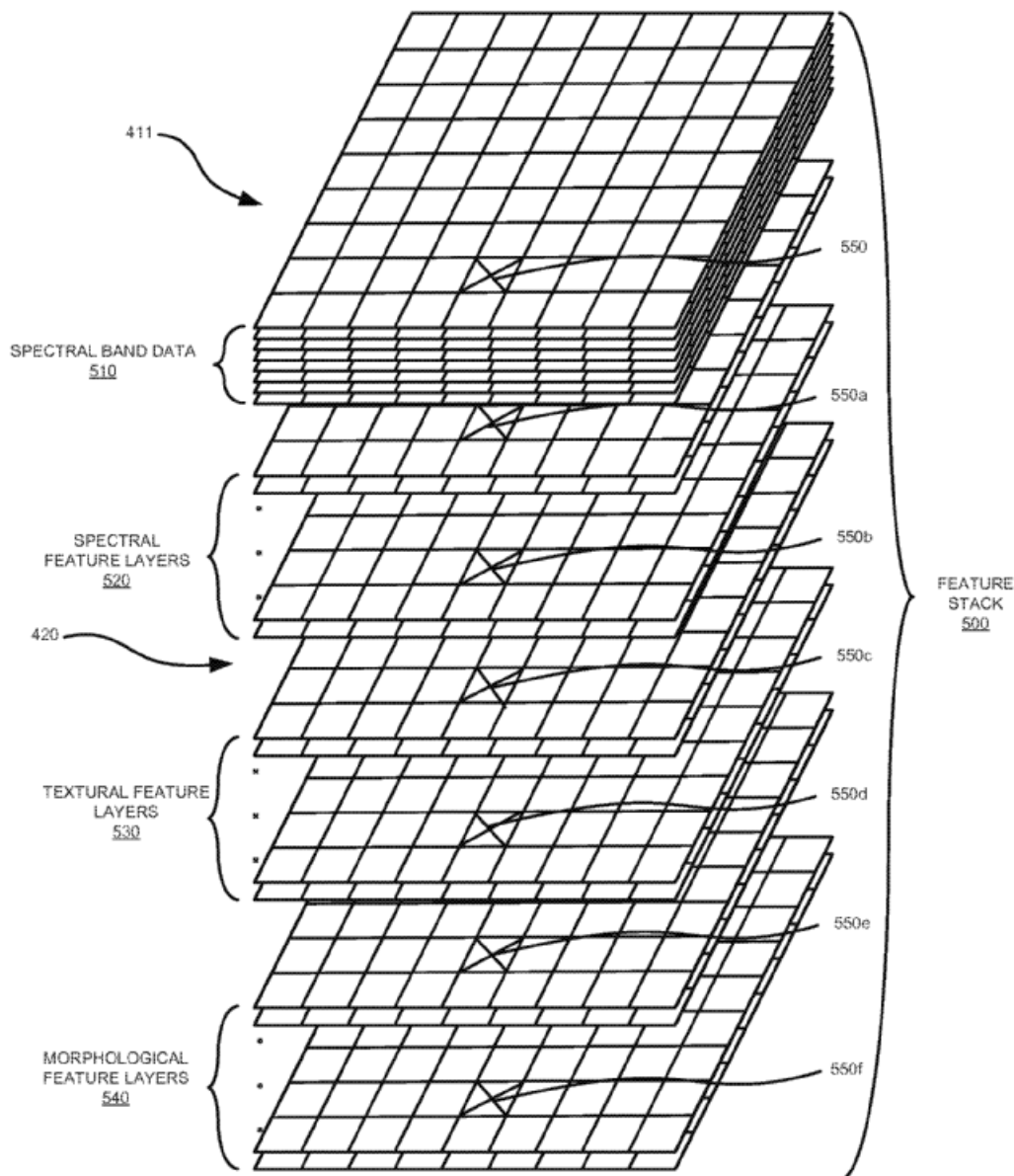
Example of a 3D representation containing hyperspectral features or representations derived from them.

Relationships with other classification places

Feature extraction in the visible spectrum using colour representations is not regarded as pertaining to this group and it is provided under colour feature extraction – group G06V 10/56.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Scenes; Scene-specific elements, terrestrial scenes taken from satellites	G06V 20/13
Scenes; Scene-specific elements, microscopic objects	G06V 20/69

| Geographic models | G06T 17/05 |

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern	G06V 10/42
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Local feature extraction by performing operations within image blocks or by using histograms	G06V 10/50
Feature extraction related to colour	G06V 10/56
Geographic information databases	G06F 16/29

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

| hyperspectral images | images in which one continuous spectrum is measured for each pixel. Generally, the spectral resolution is given in nanometres or wave numbers. |

G06V 10/60

relating to illumination properties, e.g. using a reflectance or lighting model

Definition statement

This place covers:

Techniques in which a model of illumination or reflectance of the image object is relevant for performing feature extraction or for object detection/recognition.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Information relating to the object in terms of its surface or geometry and other information relating to the scene (e.g. camera and illumination sources) may be used to compensate for, or to eliminate, the effect of changes in illumination.

When involved in the process of image or video recognition or understanding, the techniques covered include:
- analysing the scene and changing the acquisition to eliminate for undesired illumination conditions (e.g. reflections, albedo);
- evaluating the amount of illumination in the scene, adapting the processing according to this amount;
- estimating the position of the illumination source(s);
- estimating or modelling other properties of the illumination source;
- estimating the amount of front or back light;
- illumination invariant representations for object recognition (e.g. using local/global transforms);
- computing an illumination or reflectance map of the image scene and taking this map into account in object detection/recognition, e.g. "de-lighting" or "re-lighting" techniques;
- extracting features in presence of shadows, estimating shadows;
- representing objects using shape-illumination manifolds.

Illustrative examples of subject matter classified in this place:

1.
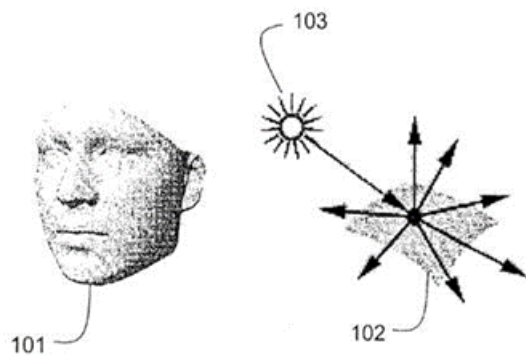
Modelling the reflection of human skin by considering its light scattering properties.
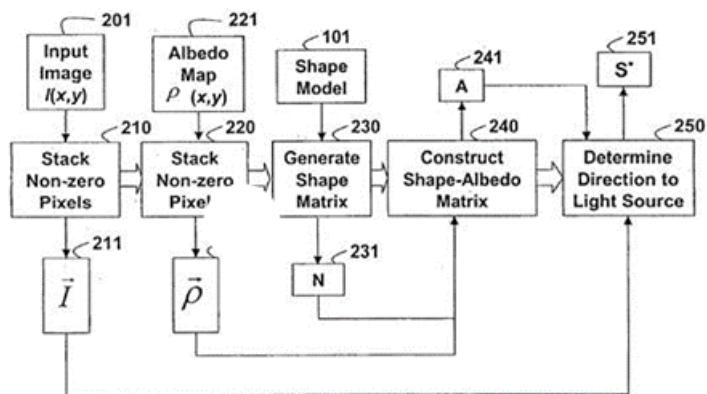
2.

Light source direction determination by modelling the albedo and the shape of an object.

3.
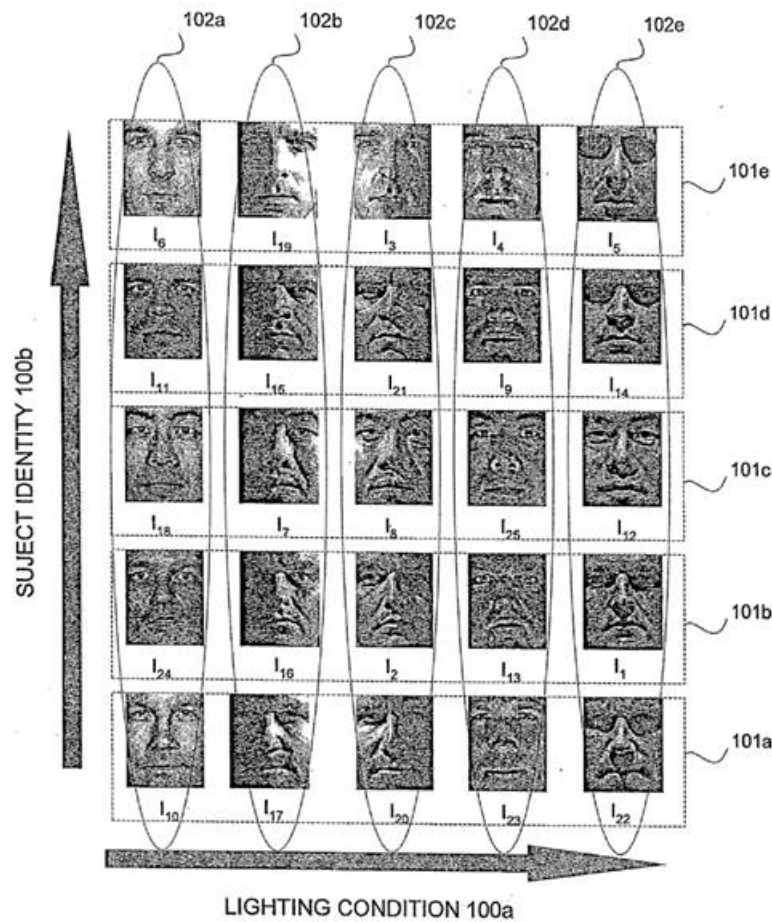
Person identification in different illumination (lighting) conditions by grouping the images pertaining to a certain illumination condition.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern	G06V 10/42
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Local feature extraction by performing operations within image blocks or by using histograms	G06V 10/50
Feature extraction related to colour	G06V 10/56
Image analysis for depth or shape recovery	G06T 7/50
Image analysis for determining position or orientation of objects or cameras	G06T 7/70
Colour picture communication systems	H04N 1/46

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

Albedo	the proportion of the incident light or radiation that is reflected by a surface
diffuse reflection	reflection having the property that incident light rays are scattered in many different directions
illumination cone	representation of a set of all possible images of a convex Lambertian surface created by varying the strength and direction of an arbitrary number of light sources at infinity
Lambertian model	model according to which the radiant intensity or luminous intensity observed from an ideal diffusely reflecting surface or ideal diffuse radiator is directly proportional to the cosine of the angle θ between the direction of the incident light and the surface normal ($I = I_0 \cos(\theta)$)
reflectance	effectiveness of a surface in reflecting radiant energy, component of the response of the electronic structure of the material to the electromagnetic field of light, and is in general a function of the frequency, or wavelength, of the light, its polarisation, and the angle of incidence
specular reflection	mirror-like reflection
spherical harmonics	special functions defined on the surface of a sphere, generally used to model the reflectance properties of a 3D surface

G06V 10/62

relating to a temporal dimension, e.g. time-based feature extraction; Pattern tracking

Definition statement

This place covers:

Techniques involving time-related feature extraction and pattern tracking for image or video recognition or understanding. Such techniques include:

- generative methods, such as kernel-based tracking [KT], Kalman filtering [KF], particle filtering [PF];
- discriminative tracking methods, such as joint probability data association filtering (JPDAF), multiple-hypothesis tracking [MHT], flow network framework [FNF].

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

1. Tracking may be implemented using a single camera or a system with multiple cameras, with possibly overlapping field of views [FOV].

2. In time-related feature extraction and pattern tracking, the features extracted from the video can be low-level (e.g. pixel colours, gradient, motion cues), mid-level (e.g. edges, corners, interest points, regions, etc.) or high-level (e.g. geometrical arrangements of parts of an object). The tracking often involves the foreground-background segmentation or background modelling in order to focus only on the objects of interest and reduce the overall complexity. Target representations are models of the objects of interest which rely on visual cues such as shape, texture, colour. There are rigid models (e.g. regions or volumes of interest), articulated models (e.g. kinematic chains) or deformable models (e.g. fluid models, point-distributions, appearance models).

An inherent problem during tracking is that of localisation, which is usually solved:
- using single-hypothesis localisation in which only one track candidate estimate is evaluated over time, e.g. gradient-based trackers such as Kanade-Lucas-Tomasi [KLT], mean-shift [MS] tracker, Bayes tracker, Kalman filtering; or
- a multiple-hypothesis localisation where multiple tracks are evaluated simultaneously, e.g. grid sampling, particle filter, hybrid methods such as hybrid particle mean shift tracker.

Models employed during tracking include graphical models (e.g. Markov models), graph-matching based tracking, camera-link model [CLM] or statistical models such as maximum a-posteriori estimation (MAP).

Problems frequently occurring are that of context modelling (e.g. changes in background, clutter, duration of the tracking events), or in the case of a multiple camera system, that of re-identification, i.e. detection of the same object in the field of view of these cameras.

Neural networks have been more recently applied to the problem of tracking, examples of architectures include: generic object tracking using regression networks [GOTURN], multi-domain network [MDNet], long short-term memory [LSTM] networks, recurrent you only look once [ROLO] networks.

Illustrative example of subject matter classified in this place:
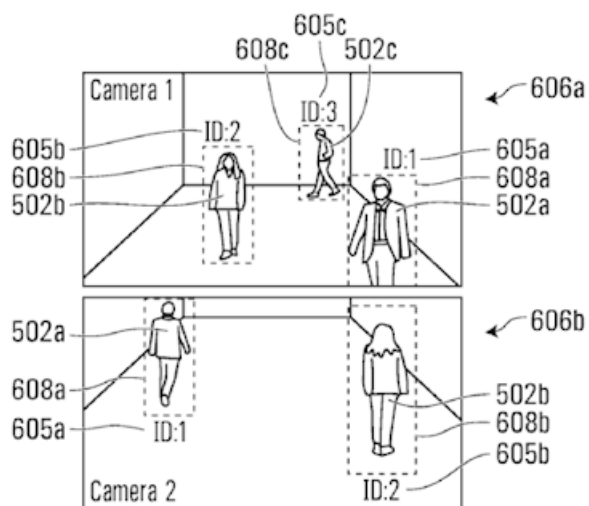
Tracking, person re-identification in a multiple camera system.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding involving the determination of a region or volume of interest [ROI, VOI]	G06V 10/25
Global feature extraction by analysis of the whole pattern	G06V 10/42
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Local feature extraction by performing operations within image blocks or by using histograms	G06V 10/50
Feature extraction related to texture	G06V 10/54
Feature extraction related to colour	G06V 10/56

Pattern recognition or machine learning for image or video recognition or understanding using probabilistic graphical models	G06V 10/84
Analysis of motion in images	G06T 7/20

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

CLM	camera link model
FOV	field of view
GM	graph matching
KF	Kalman filter
KT	kernel tracking
MAP	maximum a-posteriori estimation
MHT	multiple hypothesis tracking
PF	particle filtering

G06V 10/70

using pattern recognition or machine learning (optical pattern recognition or electronic computations therefor G06V 10/88)

Definition statement

This place covers:

Methods and arrangements for pattern recognition or machine learning in image or video data.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Pattern recognition algorithms try to identify or discover regularities in data (such as a collection of representative features derived from images) through the use of computer algorithms. These regularities are used to take actions such as classifying the data into different categories. Modern approaches include the use of techniques from machine learning for this purpose.

Pattern recognition and machine learning algorithms can operate in a supervised fashion, an unsupervised fashion or in hybrid forms (e.g. semi-supervised). Supervised methods require not only exemplary feature patterns for training the model, but also a-priori knowledge in the form of associated class labels that indicate a respective category or class. Using labelled inputs and outputs, the accuracy can be measured and the method can adapt/learn over time. In contrast, unsupervised learning may discover hidden patterns in data without the need for human intervention, with the goal of, e.g. clustering unlabelled data sets.

Notes – other classification places

Specific aspects of the pattern recognition or machine learning in the recognition or understanding of images or video are classified in subgroups as follows:
- Preparation of data items for being fed into a pattern recognition or machine learning algorithm (e.g. complementing missing data, statistical pre-processing, discarding feature vectors which have been identified as outliers), is classified in group G06V 10/72;
- Pattern matching based on a measure of (dis)similarity, e.g. template matching, is classified in group G06V 10/74. The definition of suitable criteria (e.g. similarity thresholds) for deciding whether a match is successful or not is also classified in group G06V 10/74;

- Clustering algorithms are classified in group G06V 10/762;
- Classification algorithms are classified in group G06V 10/764;
- Regression algorithms are classified in group G06V 10/766;
- The processing of image or video features in feature spaces is classified in group G06V 10/77; also classified there are techniques of data integration or data reduction, e.g. principal component analysis [PCA], independent component analysis [ICA], self-organising maps [SOM] or blind source separation. Feature selection methods which pick the most informative vectors/dimensions of high-dimensional feature vectors during model training, and which disregard the others, are classified in group G06V 10/771;
- Generating sets of training patterns and bootstrap methods (e.g. bagging, boosting) are classified in group G06V 10/774;
- Validation and performance evaluation of the methods of pattern recognition and machine learning are classified in group G06V 10/776;
- Active pattern learning, e.g. online learning of features, is classified in group G06V 10/778;
- Fusion, i.e. combining data from various sources, is classified in group G06V 10/80;
- Artificial neural networks [ANN] are classified in group G06V 10/82;
- Graphical models, e.g. Markov models or Bayesian networks, are classified in group G06V 10/84;
- Syntactic or structural representations and graph matching are classified in group G06V 10/86.

References

Limiting references

This place does not cover:

Pattern recognition performed by an arrangement of optical devices rather than by machine learning	G06V 10/88

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Scenes; Scene-specific elements	G06V 20/00
Character recognition	G06V 30/00
Image or video recognition or understanding of human-related, animal-related or biometric patterns in image or video data	G06V 40/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Neural network models per se, not specially adapted to a particular data modality	G06N 3/02
Genetic algorithms per se, not specially adapted to a particular data modality	G06N 3/12
Machine learning in general	G06N 20/00
Identification of individual speakers or sound sources by multimodal pattern matching	G10L 17/10

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

AE	auto-encoder network
AlexNet	CNN designed by Alex Krizhevsky et al.
Backprop	backpropagation, an algorithm for adjusting the weights of an artificial neural network
BERT	bidirectional encoder representations from transformers, a transformer based artificial neural network
C4.5	an algorithm for learning decision trees
CART	classification and regression trees
CNN	convolutional neural network, an artificial neural network that includes convolutional layers
CPD	coherent point drift, an algorithm for matching point clouds
DAG	directed acyclic graph
DBSCAN	density-based spatial clustering of applications with noise, a non-parametric clustering algorithm which does not require specifying the number of clusters in advance
DNN	deep neural network
EMD	earth mover's distance/Wasserstein metric
FCL	fully connected layer of an artificial neural network
FCNN	fully convolutional neural network
GAN	generative adversarial network
GMM	Gaussian mixture model
GoogLeNet	deep convolutional neural network
ICA	independent component analysis
ICP	iterative closest point, an algorithm for matching point clouds
ID3	iterative Dichotomiser 3, an algorithm for learning decision trees
Inception	convolutional neural network which concatenates several filters of different sizes at the same level of the network
IoU	intersection over union, a measure for quantifying the accuracy of an object detection algorithm
KDE	kernel density estimation, an algorithm for estimating the probability density function of a random variable
kernel	function which expresses an inner product of two inputs in another feature space
KLT	Karhunen-Loève transform
K-Means	data clustering algorithm
KNN	K-nearest neighbour: a classification algorithm which, for a given data sample, chooses the k most similar samples from a training set, retrieves their respective class labels, and assigns a class label to the data sample by majority decision. Variant - 1NN, which is KNN for k=1
LASSO	least absolute shrinkage and selection operator

LDA	linear discriminant analysis
LeNet	early CNN that firstly demonstrated the performance of CNNs on handwritten character recognition
LSTM	long short-term memory, a recurrent neural network
LVQ	learning vector quantisation
MDS	multi-dimensional scaling
MLP	multi-layer perceptron
MRF	Markov random field
MS COCO	annotated image data set
overfitting	trained model suffers from overfitting if it performs well on the training data, but generalises poorly on new test data
PASCAL VOC	collection of data sets for object detection
PCA	principal component analysis
PDF	probability density function
Perceptron	simple feed-forward neural network
RANSAC	random sample consensus, a popular regression algorithm
RBF	radial basis function
Res-Net	residual neural network, an artificial neural network having shortcuts / skip connections between different layers
R-CNN	convolutional neural network using a region proposal algorithm for object detection (variants: fast R-CNN, faster R-CNN, cascade R-CNN)
ROC	receiver-operating characteristics
RPM	robust point matching, an algorithm for matching point clouds
RVM	relevance vector machine
SOM	self-organising maps, an algorithm for generating a low-dimensional representation of data while preserving the topological structure of the data
SSD	single shot (multibox) detector, a neural network for object detection
SVD	singular value decomposition
SVM	support vector machine
test data	data set different from the training data, used for testing the performance of a trained model
training data	data set used for adjusting the parameters of the model during training
transformer	deep learning model that uses attention to give different weights to individual parts of the input data
U-Net	neural network having a specific layer structure
validation data	data set used for testing the performance of the model during training
YOLO	you only look once, an artificial neural network used for object detection (comes in various versions: YOLO v2, YOLO v3, etc.)

G06V 10/72

Data preparation, e.g. statistical preprocessing of image or video features

Definition statement

This place covers:

Techniques that handle data quality issues such as data accuracy (obtaining the correct data entries), data completeness and data consistency (data written to a database must be valid according to all defined rules) in the context of image or video recognition or understanding.

Examples of techniques classified here include:

- data cleaning, e.g. by filling in missing values, smoothing noisy data, identifying or removing outliers, resolving inconsistencies, etc.;
- compensating for missing data by supplying alternative default values;
- eliminating unreliable samples/outliers;
- data reduction, i.e. building a reduced representation of a data set through a reduction technique (e.g. PCA) or a numerosity reduction technique such as data aggregation;
- data normalisation, so that all attributes have an equal weight, e.g. min-max normalisation, z-scores, normalisation by decimal scaling, etc.

Illustrative examples of subject matter classified in this place:

1.
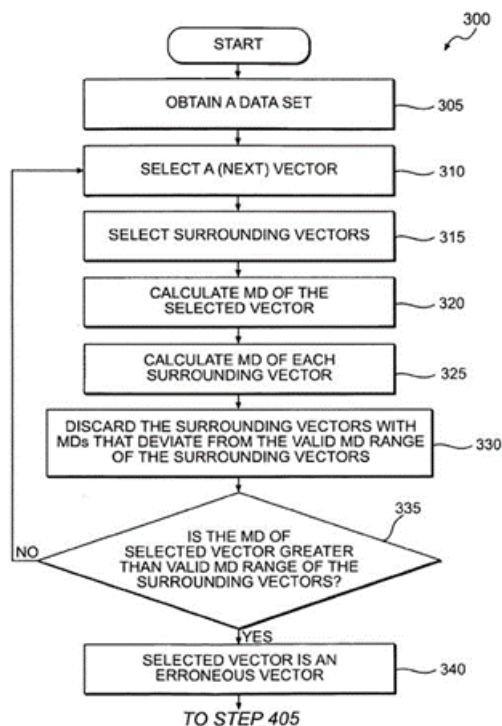
2.
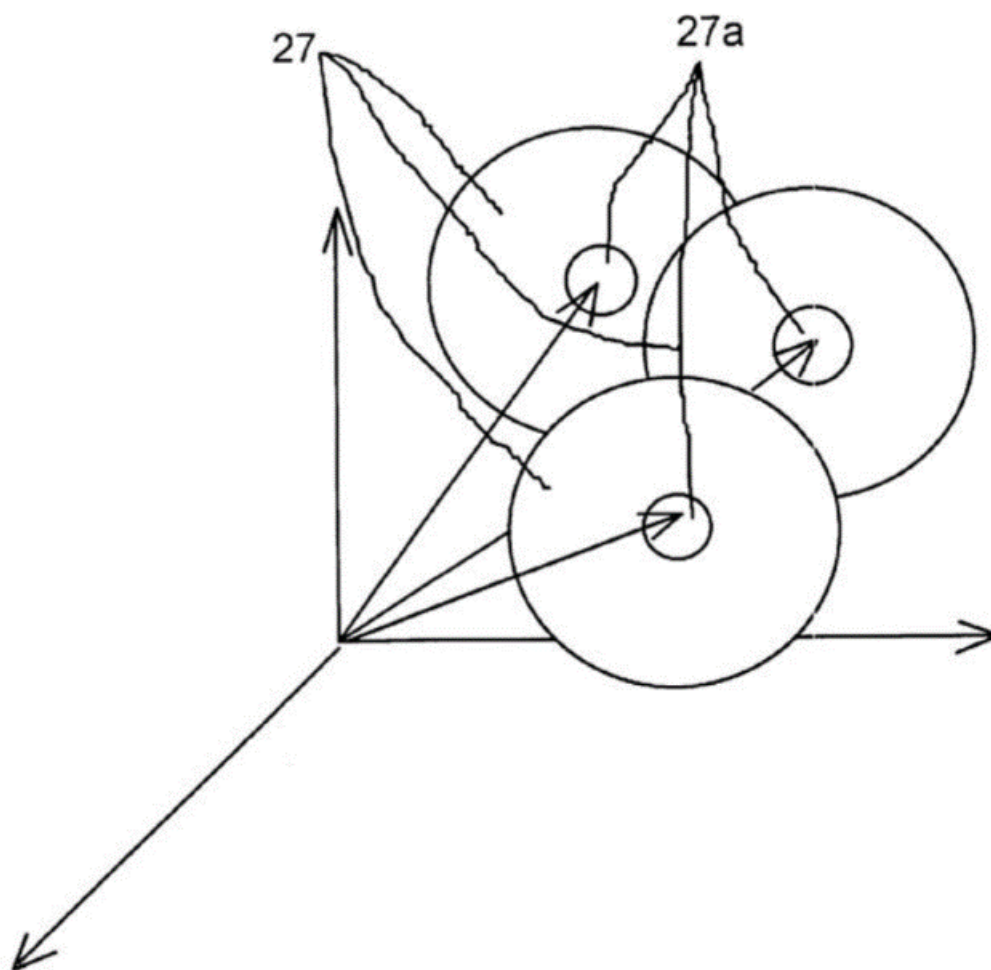
Selection of vectors in a multi-dimensional space by considering the median of their subsets and discarding those above a certain distance range from the median.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, processing image or video features in feature spaces	G06V 10/77

G06V 10/74

Image or video pattern matching; Proximity measures in feature spaces

Definition statement

This place covers:

Matching, which involves comparison of pixels values, combinations thereof or features derived from them, in which one entity is considered as a template pattern and the other is the input pattern (template matching). The matching process may involve shifting, deforming or transforming patterns to accommodate distortions or positional errors.

Histogram-based matching, wherein a histogram can be regarded as a quantised representation of the grey-level probability distribution function of pixels into intervals, called bins. Other statistical measures may be used for matching include probabilities, confidence intervals, etc.

Variational techniques such as active contour models [ACM, or "snakes"], active shapes models [ASM] or active appearance models [AAM] in which a contour or a shape of the object is obtained by iterative matching.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Aside pixels, other types of entities that may be involved in matching processes include lines, edges, object contours, object shapes, corners, key-points and statistical measures computed in a defined image neighbourhood.

The matching may be performed in a different representation space than the image space, e.g. using an eigenspace representation of the image object, using shape manifolds, using a Hough transform, using a Fourier transform etc., which implies applying a transformation from the image to this representation space prior to matching. The transformation is usually chosen due to the invariant properties sought by the matching process (e.g. Fourier transformation offers invariance to translation of the pattern in the image).

The proximity measures used during matching may include classical distances, such as Euclidian distances, or more involved distances, divergences or other measures between probability distribution functions or other statistical representations (e.g. mean, standard deviation, moments, kurtosis, Chi-square distance), for instance:
- Kullback-Leibler divergence;
- Mutual information;
- Bhattacharyya distance;
- Hamming distance;
- Earth mover, Wasserstein distance;
- Chi-square distance;
- Hellinger distance.

Notes – other classification places

Group G06V 10/75 covers more detailed aspects of the matching process, such as:
- its organisation: e.g. sequential, parallel, an initial matching with a small set of patterns each representing an entire set of patterns can be followed by a subsequent matching with all patterns in the most relevant sub-set; or matching in a randomised order, or in a predetermined order of relevance;
- precision-related aspects, e.g. rough matching with a large set of templates can be followed by a more elaborate matching with a few candidate matches; coarse-to-fine approaches at different scales of analysis, e.g. starting with a rough image resolution and then refining it to more precise resolutions;

- organisation of templates in dictionaries according to their properties in order to speed-up the matching process;
- matching using context, i.e. by taking into account secondary aspects not necessarily related to the intrinsic properties of the pattern, e.g. its proximity to other patterns, co-occurrences, etc.

Illustrative examples of subject matter classified in this place:

1.
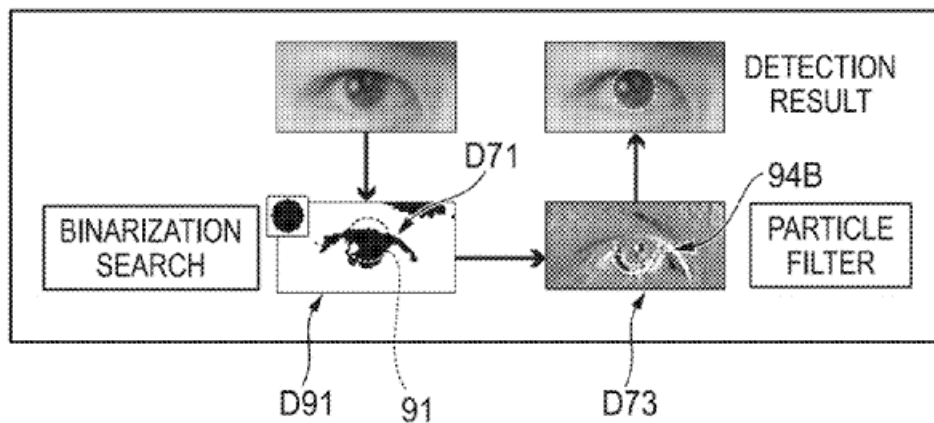
2.
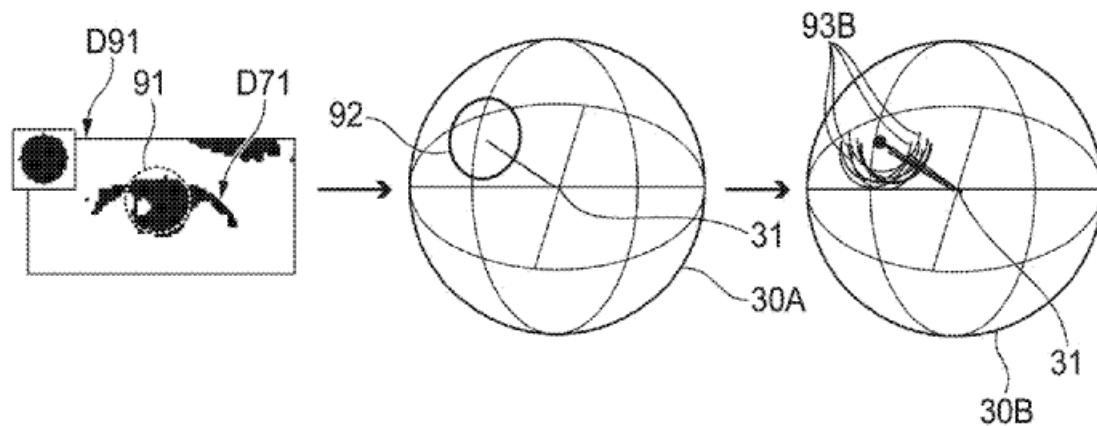
Eye detection by matching a circle/ellipse to the iris using a 2D projection onto a 3D representation of the eye.

3.
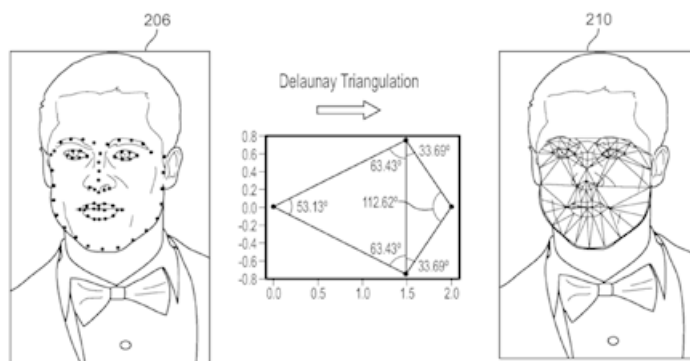
Fitting an active appearance model [AAM] to the face using key points detected for prominent facial features.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern	G06V 10/42
Descriptors for shape, contour or point-related descriptors, e.g. SIFT	G06V 10/46
Local feature extraction by performing operations within image blocks or by using histograms	G06V 10/50
Feature extraction related to texture	G06V 10/54
Feature extraction related to colour	G06V 10/56

G06V 10/762

using clustering, e.g. of similar faces in social networks

Definition statement

This place covers:

Techniques of grouping patterns together in order to reveal a certain structure or a meaning in images or video.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The object of techniques classified here is to identify groups of similar entities and to assign entities to a group (cluster) according to a measure of their similarity.

Separability is determined by measuring the similarity or dissimilarity. Such techniques are usually performed in a high-dimensional feature space constructed by extracting features from the image or video, but can also be performed in the original domain, e.g. in the image domain in the case of image segmentation.

Regarding the grouping of patterns, any pattern may belong exclusively to a single cluster (hard clustering) or it may belong simultaneously to more than one cluster up to a certain degree (fuzzy clustering) according to a similarity (or proximity) measure. In addition, depending on the clustering

method used, proximity may be defined (a) between vectors, (b) between a vector and a set of vectors (or a cluster), and (c) between sets of vectors (or different clusters).

Examples of proximity measures are: dissimilarity measures (based on l_1, l_2, and l_∞ norms), similarity measures (inner product, cosine, Pearson's correlation coefficient, Tanimoto distance, etc.).

Clustering algorithms include:

a) clustering based on statistical measures (which mainly employ numerical data) which adopt a cost function J related to possible groupings which is subject to a global or local optimisation criterion, and return a clustering that optimises J. Examples of such algorithms are:
- Hard clustering algorithms, where a vector belongs exclusively to a specific cluster, e.g. k-means, k-medoids, Linde-Buzo-Gray, ISODATA, DBSCAN, Neural Gas;
- Fuzzy clustering algorithms, where a vector belongs to a specific cluster up to a certain degree, e.g. fuzzy c-means, adaptive fuzzy C-shells [AFCS], fuzzy C quadric shells [FCQS], modified fuzzy C quadric shells [MFCQS];
- Probabilistic clustering algorithms, which follow Bayesian classification arguments and in which each vector is assigned to the cluster according to a probabilistic set-up, e.g. expectation maximisation [EM], Gaussian mixture model [GMM], mean-shift;
- b) Graph-based clustering algorithms, e.g. minimum spanning tree [MST] clustering, clustering based on directed trees, spectral clustering, graph-cut optimisation;
- c) Competitive learning algorithms for clustering, in which a set of representatives is selected and the goal is to move all representatives to regions of a vector space that are "dense" in terms of other vectors. Examples are leaky learning algorithms, self-organising maps [SOM], learning vector quantisation [LVQ].

Hierarchical clustering is a popular technique in the class of graph-based clustering, with its agglomerative or divisive variants. Various criteria can be used for determining the groupings, such as those based on matrix theory involving dissimilarity matrices.

Algorithms included in this scheme are:
- Single link algorithm;
- Complete link algorithm;
- Weighted pair group method average [WPGMA];
- Unweighted pair group method average [UPGMA];
- Weighted pair group method centroid [WPGMC];
- Ward or minimum variance algorithm.

Illustrative examples of subject matter classified in this place:

1.
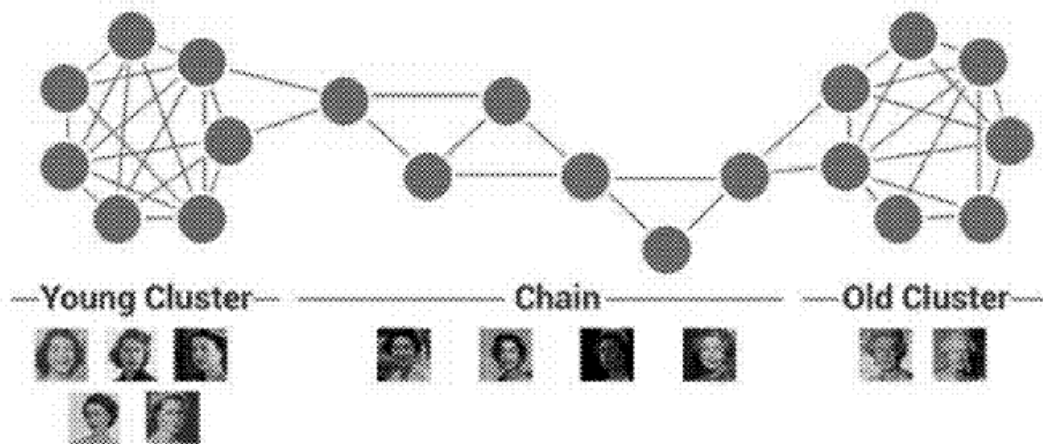
2.
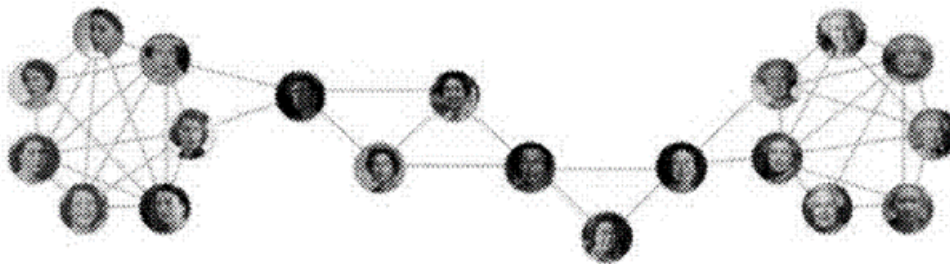
Clustering face images to detect affinity between persons using a graph-based clustering algorithm.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, processing image features in feature spaces	G06V 10/77
Pattern recognition or machine learning, fusion	G06V 10/80
Information retrieval of still images; Clustering; Classification	G06F 16/55
Information retrieval of video data; Clustering; Classification	G06F 16/75
Image analysis; Segmentation; Edge detection	G06T 7/10

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

AFC	adaptive fuzzy clustering
alternating cluster Estimation [ACE]	when a partitioning with a specific shape is to be obtained, the user can define membership functions U(V, X) and prototype functions V(U, X). The clustering will be estimated as follows: $$\begin{aligned}&\text{input } X = \{x_1, \ldots, x_n\} \in \mathcal{R}^p\\&\text{initialize } V = \{v_1, \ldots, v_c\} \in \mathcal{R}^p\\\\&\text{repeat}\\&\quad \text{store } V' = V\\&\quad \text{compute } U(V, X)\\&\quad \text{compute } V(U, X)\\&\text{until } \|V - V'\| < \varepsilon\\&\text{output } U, V\end{aligned}$$
AO	alternative optimisation
CCM	compatible cluster merging
clustering by graph partitioning	a weighted graph is partitioned into disjoint subgraphs by removing a set of edges (cut). The basic objective function is to minimise the size of the cut, which is calculated as the sum of the weights of all edges belonging to the cut.
compatible cluster merging [CCM]compatible cluster merging	starts with a sufficiently large number of clusters, and successively reduces the number by merging similar (compatible) clusters with respect to some criteria such as: $$\lvert \phi_{in}^T \phi_{jn} \rvert \quad \geq \quad k_1, \quad k_1 \text{ close to 1}$$ where: $\{\phi_{i1}, \ldots, \phi_{in}\}$ the set of eigenvectors of the ith cluster.
DBSCAN	density-based spatial clustering of applications with noise, a non-parametric clustering algorithm which does not require specifying the number of clusters in advance.
FACE	Fast-ACE
FCQS (Fuzzy C-quadric shells)FCQSFuzzy C-quadric shells	in case of quadric shaped clusters, FCQS can be employed for recovering them. For the estimation of the clusters the following cost function is minimised: $$J_a(\theta, U) = \sum_{i=1}^{N} \sum_{j=1}^{m} u_{ij}^q d_a^2(x_i, \theta_j) = \sum_{i=1}^{N} \sum_{j=1}^{m} u_{ij}^q p_j^T M_i p_j$$
FCSS	fuzzy C-spherical shells
FCV	fuzzy C-varieties
FHV	fuzzy hyper volume

fuzzy c-means clustering	Choose a number of clusters.Assign randomly to each point coefficients for being in the clusters using the formula. $w_{ij} = \dfrac{1}{\sum_{k=1}^{c} \left(\frac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\|\mathbf{x}_i - \mathbf{c}_k\|} \right)^{\frac{2}{m-1}}}$. Repeat until the algorithm has converged:Compute the centroid for each cluster, using the formula; $c_k = \dfrac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}$. For each point, compute its coefficients of being in the clusters.
Gustafson-Kessel [GK]	the GK algorithm associates each cluster with the cluster centre and its covariance. The main feature of GK clustering is the local adaptation of distance matrix in order to identify ellipsoidal clusters. The objective function of GK is: where: $$J_m = \sum_{i-1}^{n} \sum_{i-1}^{k} \mu_{ij}^m d_{ij}^2$$ $$d_{ij}^2 = (x_j - z_i) A_i (x_j - z_i)^T$$
HCM	hard c-Means

K-means clustering	■ Choose arbitrary initial estimates $\boldsymbol{\theta}_j(0)$ for the $\boldsymbol{\theta}_j$'s, $j = 1, \dots, m$.
	 ■ Repeat
	   ● For $i = 1$ to N
	     ○ Determine the closest representative, say $\boldsymbol{\theta}_j$, for $\boldsymbol{x}_i$.
	     ○ Set $b(i) = j$.
	   ● End {For}
	   ● For $j = 1$ to m
	     ○ Parameter updating: Determine $\boldsymbol{\theta}_j$ as the mean of the vectors $\boldsymbol{x}_i \in X$ with $b(i) = j$.
	   ● End {For}.
	 ■ Until no change in $\boldsymbol{\theta}_j$'s occurs between two successive iterations.

KNN	K-nearest neighbour; a classification algorithm which, for a given data sample, chooses the k most similar samples from a training set, retrieves their respective class labels, and assigns a class label to the data sample by majority decision; variant: 1NN, which is KNN for k=1.
LVQ	learning vector quantisation
partitioning around medoids [PAM] – the most common realisation of k-medoid type algorithms partitioning around medoids PAM	1. Initialise: randomly select k of the n data points as the medoids. 2. Associate each data point to the closest medoid. ("closest" here usually in a Euclidean/Manhattan distance sense). 3. For each medoid m.- For each non-medoid data point x. *Swap m and x and compute the total cost of the configuration. 4. Select the configuration with the lowest cost. 5. Repeat steps 2 to 4 until there is no change in the medoid.

G06V 10/764

using classification, e.g. of video objects

Definition statement

This place covers:

Classification of images or videos to identify the category or set of categories (classes) to which a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

Novelty detection (e.g. classification of "unseen" observations), anomaly detection or outlier detection.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Individual observations may be analysed into a set of quantifiable properties, known as explanatory variables or features. These properties may be categorical, ordinal, integer-valued, real-valued, etc. Other classifiers perform a class assignment by comparing current observations to previous observations by means of a similarity or distance function.

A classifier can be parametric or non-parametric depending on the type of model adopted for the observations.

Classification algorithms include those:
- based on the distance between a decision surface and training patterns, e.g. support vector machines [SVM];
- based on the distance between the pattern to be recognised and a reference, where the reference can be a prototype, a centroid of samples of the same class or the closest patterns from the same class or different classes, e.g. nearest-neighbour classification;
- based on a parametric, probabilistic model, where the model uses the Neyman-Pearson lemma, likelihood ratios, receiver operating characteristics [ROC], plotting the false acceptance rate [FAR] versus the false rejection rate [FRR], Bayesian classification, etc.;
- based on a graph-like or tree-like model, e.g. decision trees, random forests, etc. Examples are the classification and regression trees [CART], ID3 [Iterative Dichotomiser 3], C4.5, etc.

The decision surface of the classifier may be a linear classifier or a non-linear classifier. Linear classifiers model the boundaries between different classes in the feature space as hyperplanes. Non-linear classifiers use, e.g. quadratic, polynomial or hyperbolic functions instead.

Illustrative examples of subject matter classified in this place:

1.

Persons   Non-persons

2.

Region of Ambiguous Data
225
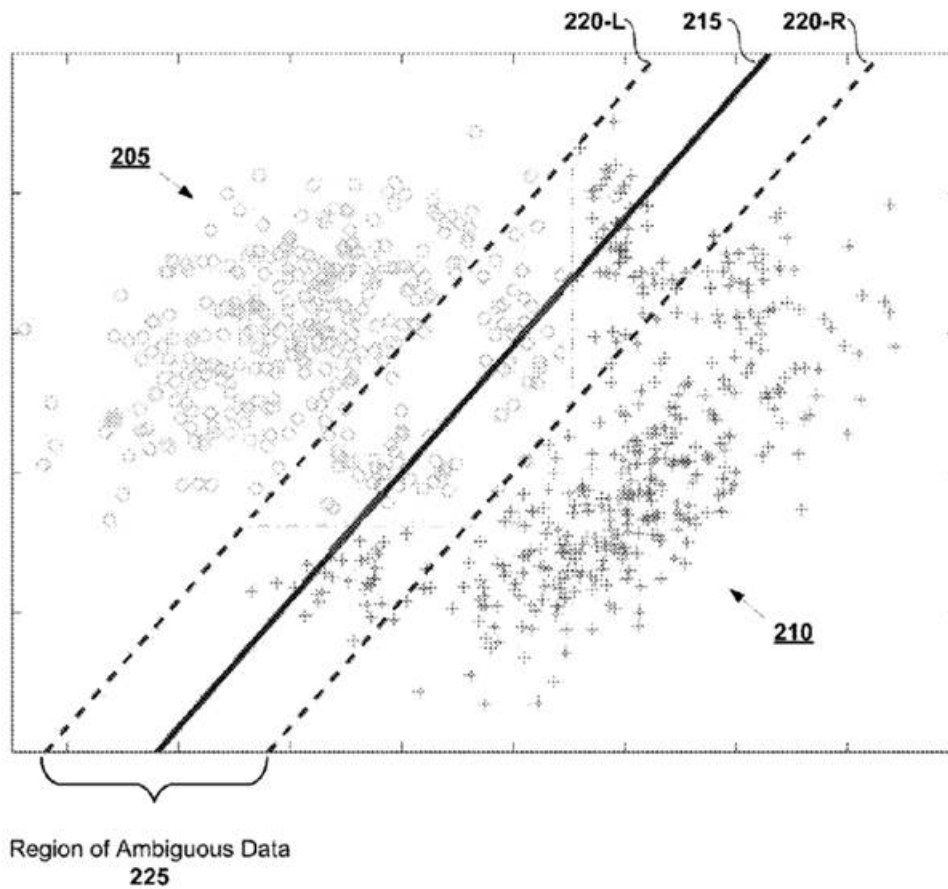
A linear support vector machine classifier which attempts to define a linear boundary between two classes (205, 210) of feature vectors originating from images containing "persons" and "non-persons", such as to separate them into two different classes.

3.
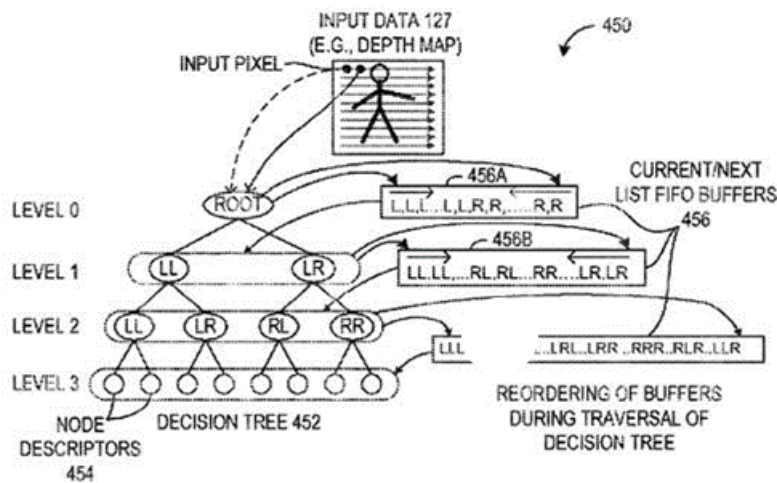
Decision tree classifying objects in the image data using an efficient hardware implementation with FIFO buffers.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, processing image features in feature spaces	G06V 10/77
Pattern recognition or machine learning, fusion	G06V 10/80
Information retrieval of still images; Clustering; Classification	G06F 16/55
Information retrieval of video data; Clustering; Classification	G06F 16/75
Image analysis; Segmentation; Edge detection	G06T 7/10

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

C4.5	classification algorithm using a decision tree
CART	classification and regression Tree
FAR	false acceptance rate
FRR	false rejection rate
Gini impurity	measure of how often a randomly chosen element from the set would be incorrectly labelled if it was randomly labelled according to the distribution of labels in the subset; usually used at the level of the nodes of tree-based classifiers.
ID3	iterative Dichotomiser 3, a precursor of C4.5
ROC	receiver operating characteristics

G06V 10/766

using regression, e.g. by projecting features on hyperplanes

Definition statement

This place covers:

Techniques for image or video recognition or understanding using regression.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The term "regression" refers to statistical techniques for estimating the relationships between a dependent variable (often called the "outcome" or "response" variable) and one or more independent variables (often called "predictors", "covariates" or "explanatory variables"), where the variables model the underlying image or video data.

Common forms of regression are:
- Linear regression - the model specification is that the dependent variables are a linear combination of the parameters (but need not be linear in the independent variables). The goal is to find a line (or a more complex linear combination) that most closely fits the data according to a specific mathematical criterion (e.g. by minimising the least-mean-squares criterion). For example, the

method of ordinary least squares computes the unique line (or hyperplane) that minimises the sum of squared differences between the true data and that line (or hyperplane);
- Non-linear regression, e.g. polynomial, binomial, binary, logistic, multinomial logistic, etc.

Illustrative example of subject matter classified in this place:
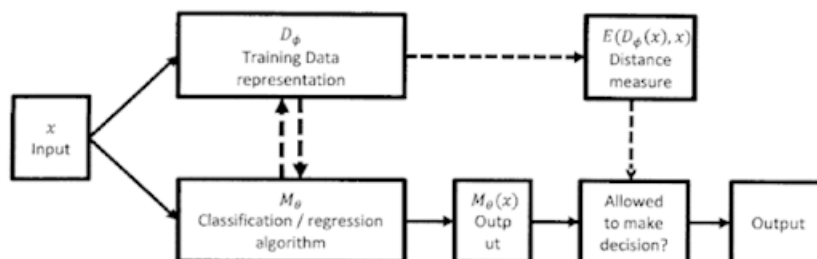
Example of adaptive regression analysis for classification.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, processing image features in feature spaces	G06V 10/77
Pattern recognition or machine learning, fusion	G06V 10/80
Digital computing; Complex mathematical operations	G06F 17/10

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

LMS	least mean squares
RANSAC	RANdom SAmple Consensus – an iterative algorithm for fitting a linear mathematical model such as a line or a plane through a set of points by eliminating the influence of outliers

G06V 10/77

Processing image or video features in feature spaces; using data integration or data reduction, e.g. principal component analysis [PCA] or independent component analysis [ICA] or self-organising maps [SOM]; Blind source separation

Definition statement

This place covers:

Techniques which deal with the problem of reducing the dimensionality of a representation of features in high-dimensional feature spaces.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The problem of reducing the dimensionality of a representation of features in high-dimensional feature spaces is sometimes referred to as "the curse of dimensionality". Generally, having to consider too many features increases the requirements regarding processing power and memory capacity; moreover, available data samples may be too sparsely distributed in a high-dimensional feature space for reliably recognising patterns and the number of training samples, which are necessary to obtain a good estimate of the actual data distribution, increases exponentially. The distances between randomly chosen pairs of training samples can be expected to exhibit little differences, causing a nearest neighbour search to become unreliable the more dimensions the feature space has.

Different types of analysis can be considered:
- based on a discrimination criterion, e.g. discriminant analysis such as linear discriminant analysis [LDA];
- based on evaluating a naturality criterion, non-negative matrix factorisation;
- based on an approximation criterion, e.g. principal component analysis [PCA];
- based on a separation criterion, e.g. independent component analysis [ICA];
- measuring the statistical independence, e.g. mutual information;
- decorrelating the data in the feature space;
- enforcing sparsity or performing a domain transformation, or evaluating a sparsity criterion, e.g. representations with an overcomplete basis;
- based on topology preservation, e.g. multidimensional scaling, self-organising maps [SOM].

Another way of dealing with the problem is to integrate or reduce data by deriving representatives through clustering.

A further concept covered by this group is the blind source separation [BSS] which involves estimating individual source components from mixtures of multiple sources, e.g. blended images that are obtained by superimposing one image onto another, or an image that is deteriorated by a noise process.

The reduction of the representation by principal component analysis [PCA] has been extensively applied in various application-related contexts, one example being face recognition ("Eigenfaces").

Notes – other classification places

Various subgroups cover further aspects relating to processing features in high-dimensional feature spaces.

In particular, group G06V 10/771 covers techniques relating to feature selection, e.g. selecting those features which are the most representative from a multi-dimensional feature space. Well-known ways to carry out feature selection are:
- by ranking or filtering the set of features, e.g. using a statistical measure such as variance or cross-correlation;
- by evaluating different subsets according to an optimisation criterion such as class separability in forward selection or backward elimination;
- using evolutionary computational techniques, such as genetic algorithms.

Group G06V 10/772 covers techniques for determining representative reference patterns, e.g. by averaging or distorting patterns, or for generating dictionaries, i.e. sets of templates which are usually organised efficiently for different purposes, such as matching.

Group G06V 10/776 covers techniques for validation and performance evaluation. They usually involve considerations of partitioning the available data into a training set to be used for training a classification model, and a validation set used to assess the validity of the classification and to evaluate its performance.

Group G06V 10/778 covers techniques for active pattern learning, e.g. online learning.

Group G06V 10/80 covers techniques for fusion, i.e. combining data from various sources at the sensor level, pre-processing, feature extraction or classification level, mainly to improve the performance of a pattern recognition system for images or video.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, fusion	G06V 10/80
Information retrieval of still images; Clustering; Classification	G06F 16/55
Information retrieval of video data; Clustering; Classification	G06F 16/75

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

BSS	blind source separation
eigenface	name given to a set of eigenvectors obtained by principal component analysis when used in face recognition.
ICA	independent component analysis
LDA	linear discriminant analysis
MDS	multidimensional scaling
PCA	principal component analysis
SOM	self-organising map

G06V 10/776

Validation; Performance evaluation

Definition statement

This place covers:

Techniques for validation and performance evaluation of algorithms for image or video recognition or understanding.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Validation and performance evaluation of algorithms for image or video recognition or understanding normally involve:
- a training data set which is a set of examples used to fit the parameters of a pattern recognition or machine learning model;
- a validation data set which provides an unbiased evaluation of the model fit on the training data set (while optionally tuning the model's parameters); and

- a test data set used to provide an unbiased evaluation of a final model. If the data in the test data set has never been used in training (for example in cross-validation), the test data set is also called a holdout data set.

Common classification metrics to evaluate the models are true positive rate [TPR] or sensitivity, false positives rate [FPR] or fall-out, true negatives rate [TNR] or specificity, false negative rate [FNR] or miss rate, receiver operating characteristic [ROC] curves (TP rate divided by the FP rate), z-score, accuracy, precision (or positive predictive value), recall, negative predictive value, intersection over union [IoU], the Jaccard index (also referred to as Tanimoto index), etc. Other metrics are also possible, for instance regression metrics, explained variance, validation curves, detection error trade-off etc. In case of decision-tree learning, the compactness of a cluster, the purity of a cluster in terms of class labels, the minimum distance of samples from the class boundary, a calculated likelihood score, etc.

In order to get more stable results and use all valuable data for training, a data set can be repeatedly split into several training and validation data sets. This strategy is known as cross-validation.

The performance can be measured automatically, e.g. by a stochastic process such as when using bootstrapping, or by a human operator in the case of relevance feedback.

Illustrative examples of subject matter classified in this place:

1.
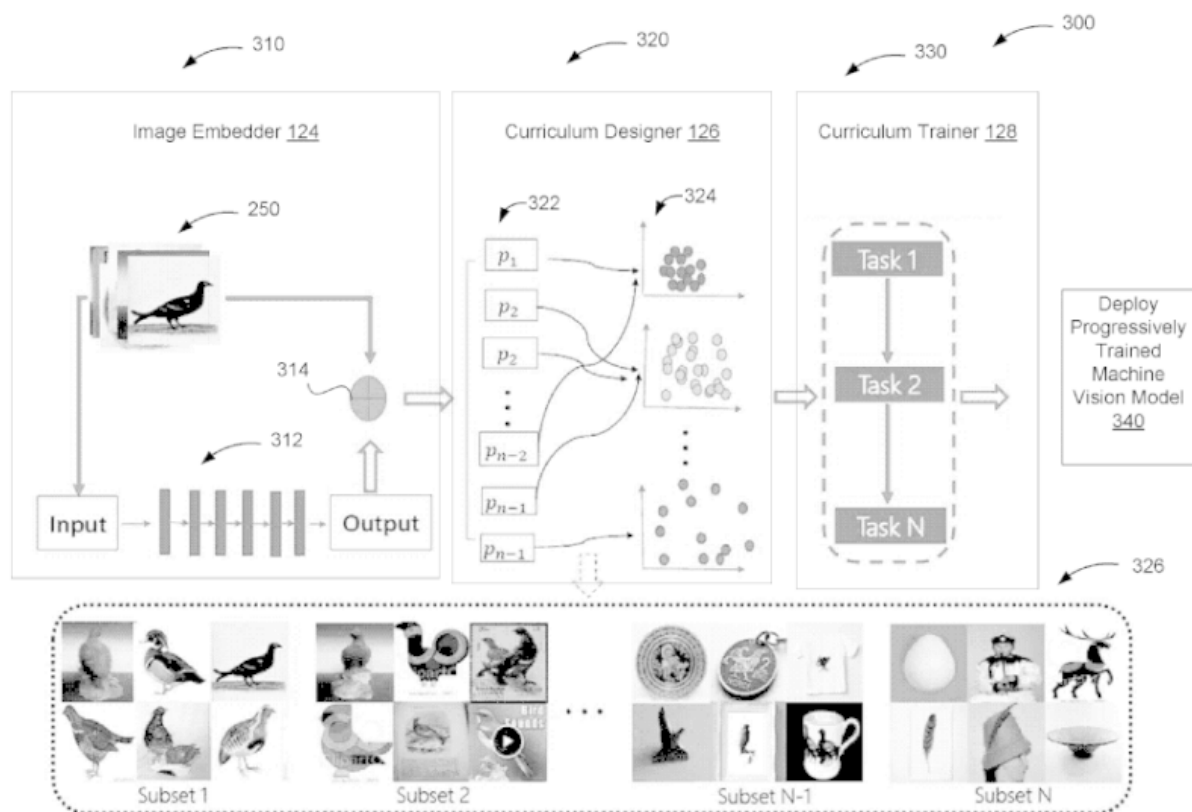
2.
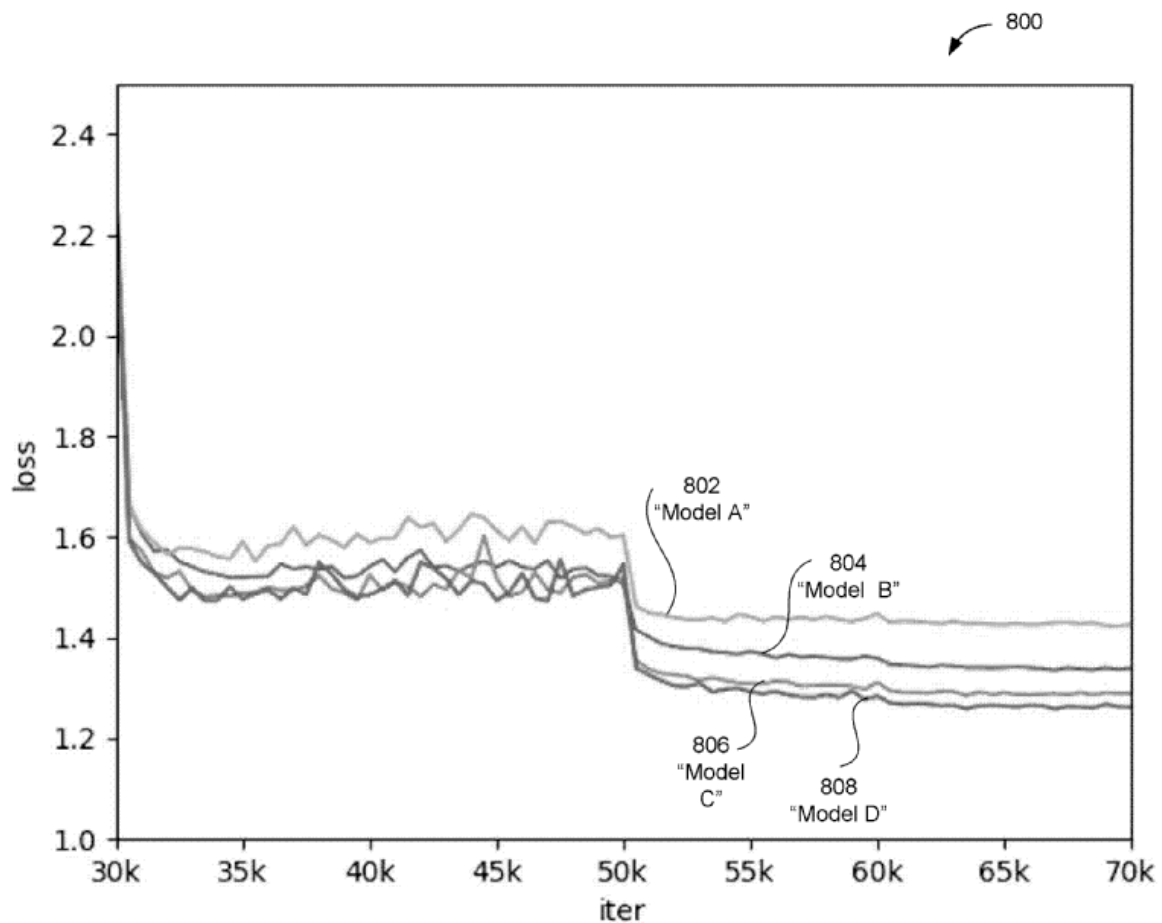
Example of an iterative "loss function" calculation for four different recognition models trained with different subsets of images, which is indicative of the performance of the classification of each model.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, processing image features in feature spaces	G06V 10/77
Pattern recognition or machine learning, fusion	G06V 10/80
Digital computing; Complex mathematical operations	G06F 17/10

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

FNR	false negative rate or miss rate
FPR	false positive rate or fall-out rate

IoU	intersection over union
ROC	receiver operating characteristic
TNR	true negative rate or specificity
TPR	true positive rate or sensitivity

G06V 10/778

Active pattern-learning, e.g. online learning of image or video features

Definition statement

This place covers:

Techniques for active pattern learning for image or video recognition or understanding which dynamically adapt a learning algorithm (e.g. a neural network), either by interactively querying a supervisor (user) or from some other information source, like a teacher module, to classify or learn new data. Examples of techniques classified here include:

- Membership query synthesis, where the learner generates its own instance from an underlying data set. For example, if the data set is pictures of humans and animals, the learner could send a clipped image of a leg to the teacher and query if this appendage belongs to an animal or human;
- Pool-based sampling, where instances are drawn from the entire data pool and assigned a confidence score, a measurement of how well the learner "understands" the data. The system then selects the instances for which it is the least confident and queries the teacher for the labels;
- Stream-based selective sampling, where unlabelled data samples are examined one at a time with the machine evaluating the informativeness of each item against its query parameters. The learner decides for itself whether to assign a label or query the teacher for each sample.

Illustrative example of subject matter classified in this place:
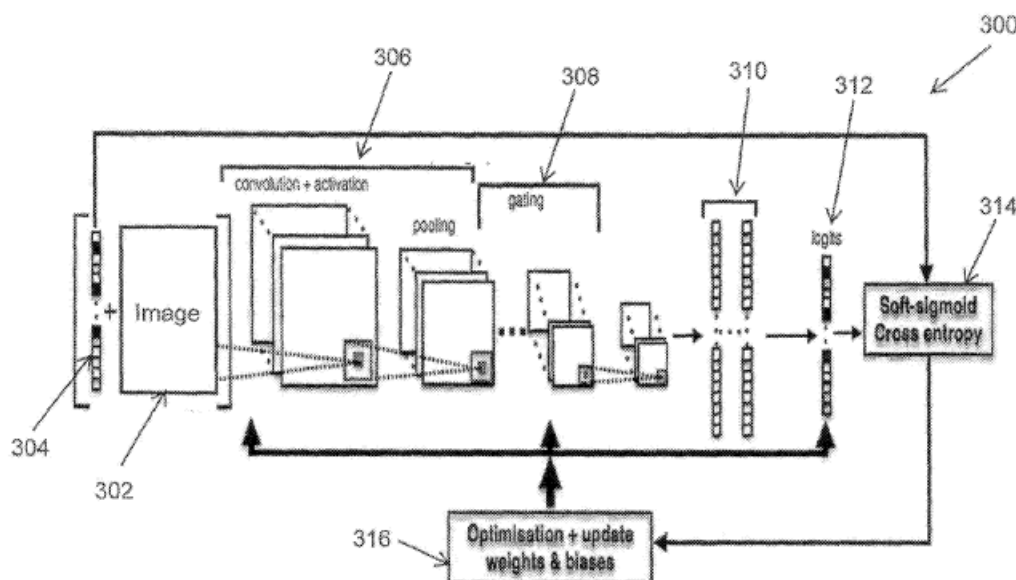
Active learning of weights and biases at different stages of a convolutional neural network for image classification.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, processing image features in feature spaces	G06V 10/77
Pattern recognition or machine learning, fusion	G06V 10/80
Digital computing; Complex mathematical operations	G06F 17/10
Machine learning in general	G06N 20/00

G06V 10/80

Fusion, i.e. combining data from various sources at the sensor level, preprocessing level, feature extraction level or classification level (multimodal speaker identification or verification G10L 17/10)

Definition statement

This place covers:

Combining the information from several sources in order to form a unified representation for image or video recognition or understanding.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

A simple fusion process combines raw data from several sensors or several sensor modalities (e.g. fusing spatial and temporal data). Besides fusing the raw sensor data, it is also possible to first process the sensor data to extract features and then combine the extracted features into a joint feature vector. Alternatively, it is possible to fuse classification results, e.g. inputting the features from different sensor modalities to separate classifiers, receiving respective classification scores from each classifier, and combining the individual scores into a final classification result.

Examples are probabilistic fusion, statistic fusion, fuzzy reasoning fusion, fusion based on evidence and belief theory, e.g. Dempster-Shafer, fusion by voting.

Fusion can also be applied at different stages of a recognition system for different purposes, e.g. for dimensionality reduction, computing robustness, improving precision and certainty in the classification decisions, etc.

Illustrative examples of subject matter classified in this place:

1.
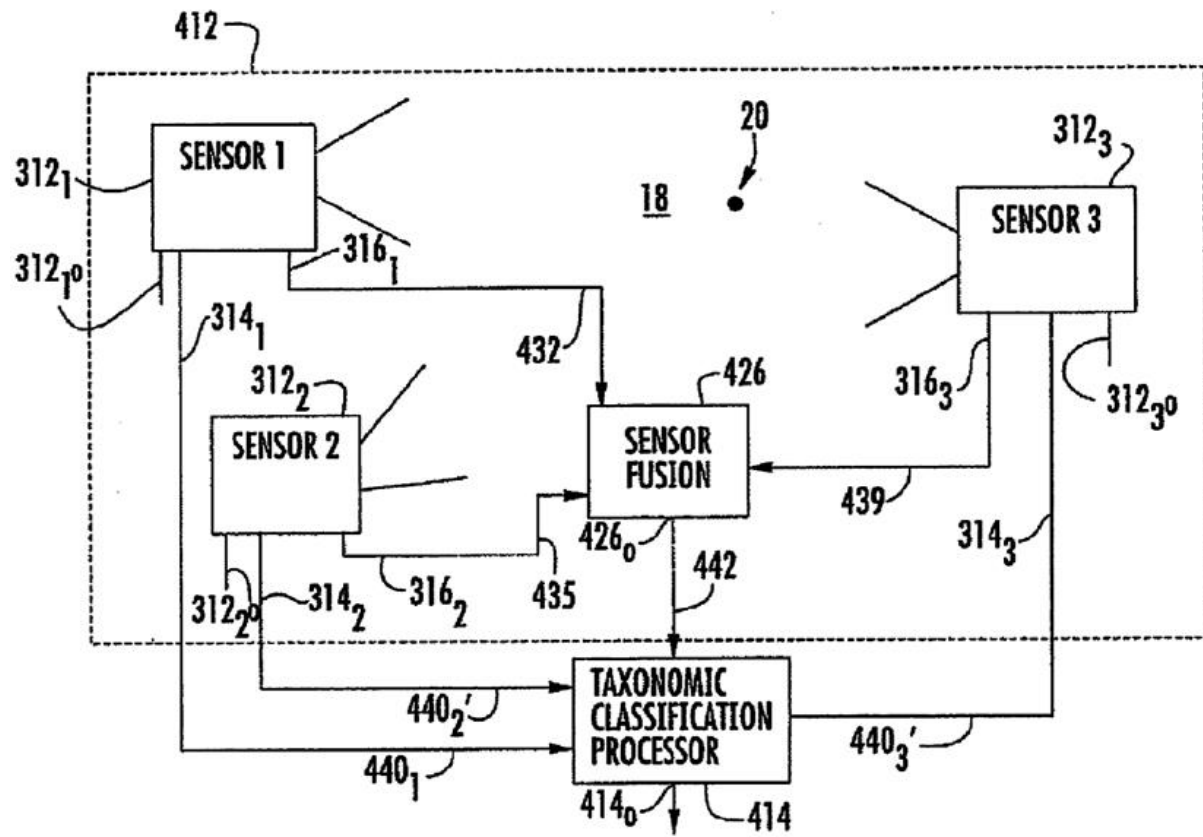
Sensor-level fusion followed by classification.

2.
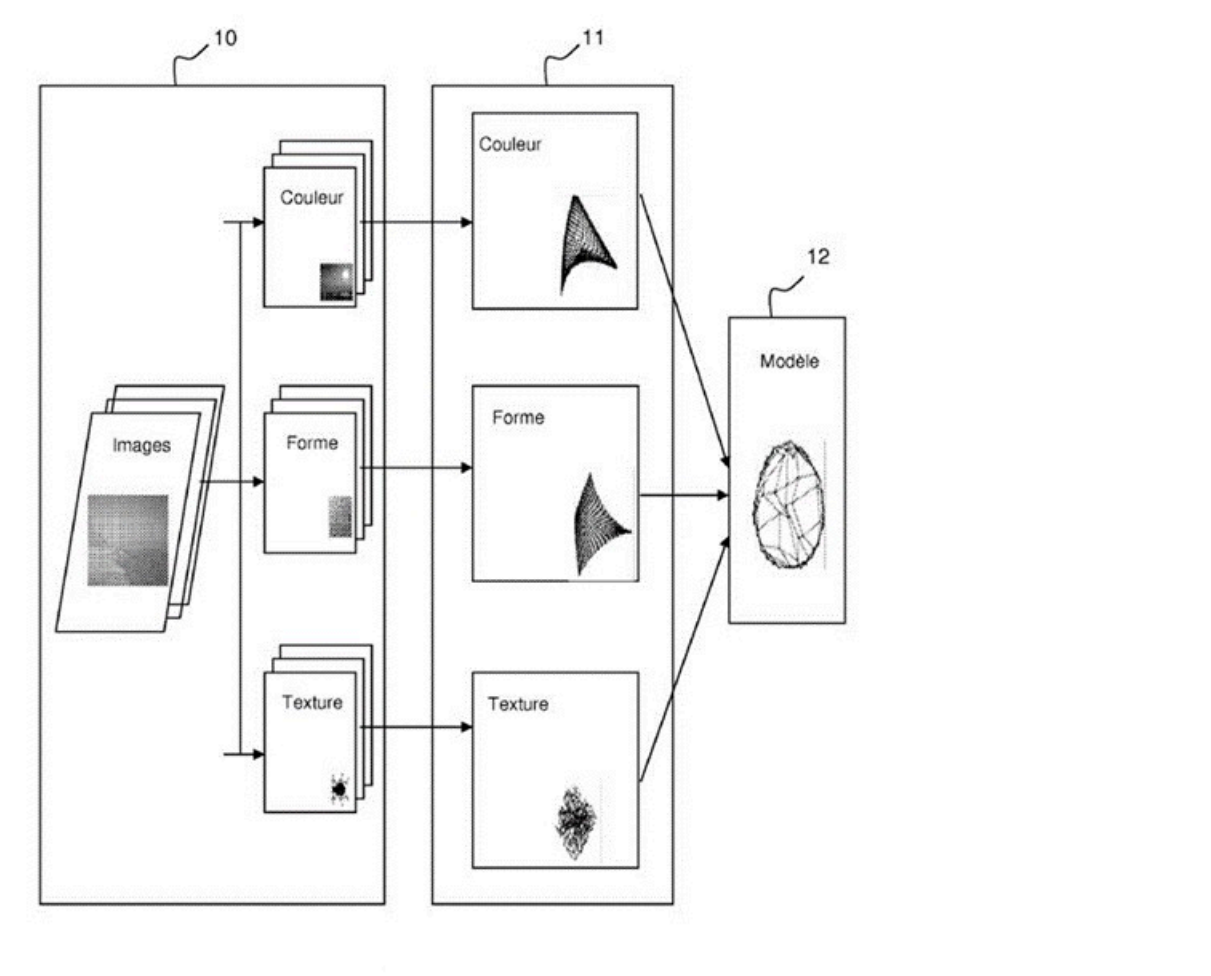
Feature-level fusion by combining colour, shape and texture representations.

References

Limiting references

This place does not cover:

Multimodal speaker identification or verification	G10L 17/10

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

Dempster-Shafer	general framework for reasoning with uncertainty which combines evidence from different sources and arrives at a degree of belief (represented by a mathematical object called belief function) that takes into account all the available evidence.

G06V 10/82

using neural networks

Definition statement

This place covers:

Neural networks [NN] specially adapted for image or video recognition or understanding, in particular specific architectures and specific learning tasks for this purpose.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Examples of architectures are:
- Attention based neural networks such as transformer architectures;
- Autoencoders consisting of encoder and decoder blocks, where the output has the same form as the input, i.e. input and outputs are both images for example;
- Convolutional neural networks consisting of repetitive convolutional and pooling layers;
- Pyramidal or multi-scale neural network, mostly of the convolutional type that process either differently scaled input images, have convolutional kernels of varying sizes, and/or contain skip connections from lower-level layers to higher level layers or the output layer;
- Recurrent neural networks, where the input data is sequential by nature. That is either the pixels of the input image are processed sequentially, or a plurality of image frames such as in videos are processed sequentially. Long-short-term-memory [LSTM] and gated recurrent units [GRU] are some specific examples of recurrent neural networks;
- Region proposal networks, where the main task is not only to correctly classify objects in an input image but also provide an indication where a specific object has been found. Example architectures are R-CNN and YOLO;
- Residual neural networks [ResNet] containing skip connections or shortcuts to jump over some layers;
- Siamese neural networks, that work on input pairs and consist of two identical neural networks that process each pair of the input and then merges the output to provide a judgement about the input pair, such as if they are belonging to the same class or not.

Examples of learning tasks are:
- Adversarial learning such as in generative adversarial networks (GANs);
- Meta learning;
- Metric learning, learning a distant metric between two input objects mostly done with a Siamese neural network;
- Reinforcement learning, learning how to take optimal actions for performing a task, e.g. deep reinforcement learning for robotics, self-driving vehicles etc.;
- Representation or feature learning, learning representations or features from raw input, mostly done with some form of encoder-decoder architecture or simply by using intermediate representations of a classification network;

• Transfer or multitask learning, reusing a network trained on task A for task B or jointly training a neural network on multiple tasks.

Illustrative examples of subject matter classified in this place:

1A.
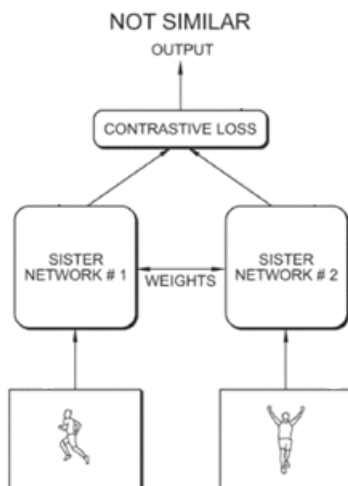
1B.
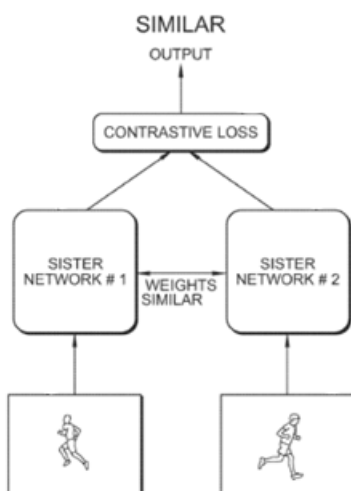
Siamese network showing not similar (left) and similar (right) input pairs.

2.

Deep Recurrent Embedded Clustering (DREC)

Video Frame at 1080p with multiple people — 400

Part Affinity Field — 302

402

$L=KL(P||Q)$

Q P

404

Action Recognition Classifier — 408

406

Control of Autonomous Agents — 310
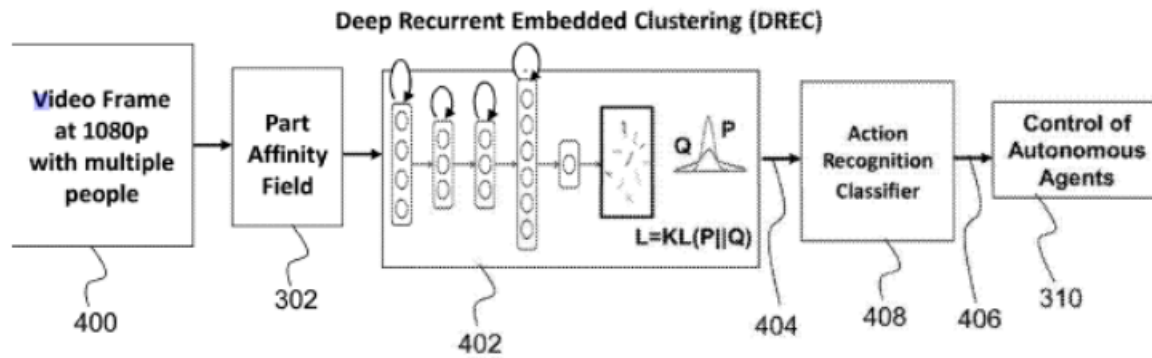
Recurrent neural network for action recognition.

3.

classifier

RoI pooling

135 — proposals

120 — feature maps

Region Proposal Network (130)

110 — "base network" conv layers

101 — image
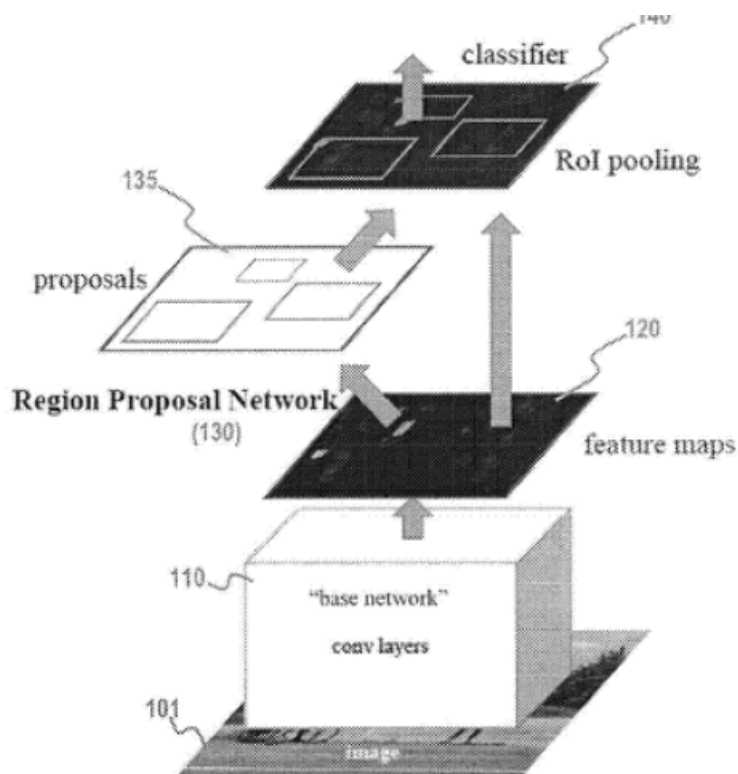
Region proposal neural network for region of interest (ROI) detection.

4.
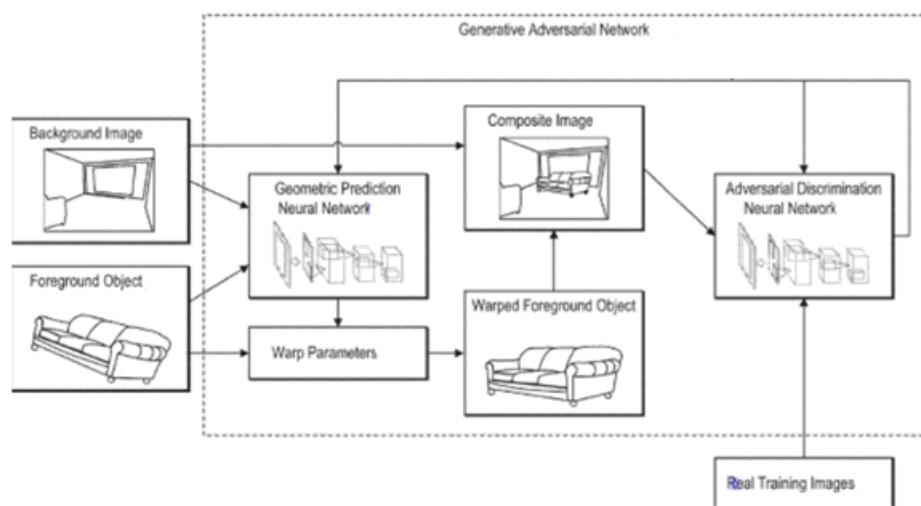
Adversarial learning with a generative adversarial neural network for object recognition on different backgrounds.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Feature extraction related to a temporal dimension; Pattern tracking	G06V 10/62
Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, fusion	G06V 10/80
Information retrieval of video data; Clustering; Classification	G06F 16/75
Computer systems based on biological models using neural networks	G06N 3/02
Computer systems using knowledge-based models; Inference methods	G06N 5/04
Machine learning	G06N 20/00
Motion image analysis	G06T 7/20

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

AE	auto-encoder network
AlexNet	CNN designed by Alex Krizhevsky et al.
Backprop	backpropagation, an algorithm for computing the gradient of the weights of an artificial neural network
BERT	bidirectional encoder representations from transformers, a transformer based artificial neural network

CNN	convolutional neural network, an artificial neural network that includes convolutional layers
DNN	deep neural network
FCL	fully connected layer of an artificial neural network
FCNN	fully convolutional neural network
GAN	generative adversarial network
GoogLeNet	deep convolutional neural network
Inception	convolutional neural network which concatenates several filters of different sizes at the same level of the network
LeNet	early CNN that firstly demonstrated the performance of CNNs on handwritten character recognition
LSTM	long short-term memory, a recurrent neural network
MLP	multi-layer perceptron
MS COCO	annotated image data set
Perceptron	simple feed-forward neural network
RBF	radial basis function
R-CNN	convolutional neural network using a region proposal algorithm for object detection (variants: fast R-CNN, faster R-CNN, cascade R-CNN)
Res-Net	residual neural network, an artificial neural network having shortcuts / skip connections between different layers
SOM	self-organising maps, an algorithm for generating a low-dimensional representation of data while preserving the topological structure of the data
SSD	single shot (multibox) detector, a neural network for object detection
U-Net	neural network having a specific layer structure
YOLO	you only look once, an artificial neural network used for object detection (comes in various versions: YOLO v2, YOLO v3, etc.)

G06V 10/84

using probabilistic graphical models from image or video features, e.g. Markov models or Bayesian networks

Definition statement

This place covers:

Graphical models for image or video recognition or understanding, with states modelled as nodes in a graph and transitions between states as graph edges, and where it is assumed that the future state of a system depends on the present state.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Examples of graphical models include:

- probabilistic models such as state machines, Bayesian networks, dynamic Bayesian networks, tree-structured models, probabilistic latent semantic analysis (PLSA), conditional random fields,

     Markov models and variations, e.g. hidden Markov models, Markov random fields, partially observable Markov models, Markov decision processes, or variable length Markov models;

- inference using graphical models, e.g. by the junction tree algorithm, factor graphs, belief propagation, message passing, Gibbs sampling, variational inference, Monte Carlo inference, Markov chains;
- learning using graphical models, e.g. by expectation maximisation, latent variable methods, Baum-Welch algorithm, Viterbi training, forward-backward propagation, Monte Carlo methods;
- learning the graphical structure of the model itself.

Applications include learning spatial context for object detection, learning spatio-temporal events for activity recognition, gesture recognition, video segmentation and understanding, etc.

Illustrative examples of subject matter classified in this place:

1.
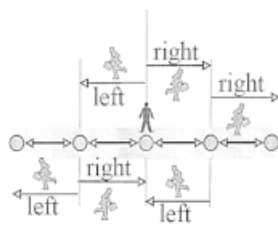
FIG. 1A
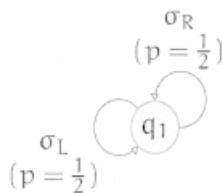
2.
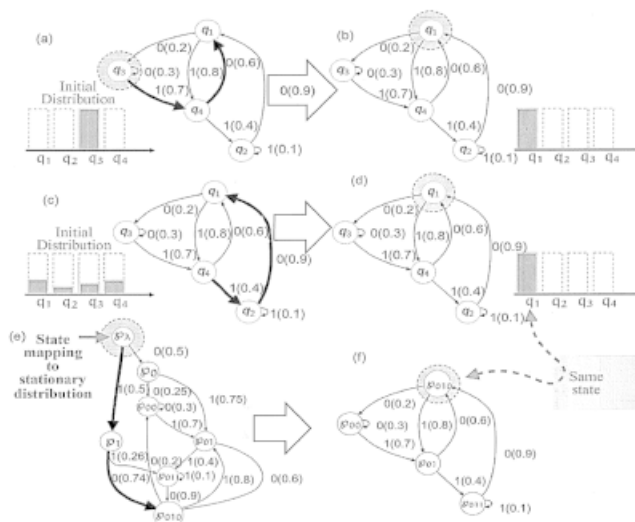
Human activity recognition using a hidden Markov model [HMM].

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Feature extraction related to a temporal dimension; Pattern tracking	G06V 10/62
Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Pattern recognition or machine learning, using regression	G06V 10/766
Pattern recognition or machine learning, fusion	G06V 10/80
Information retrieval of video data; Clustering; Classification	G06F 16/75
Motion image analysis	G06T 7/20
Speaker identification and verification; Hidden Markov models [HMM]	G10L 17/16

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

EM	expectation maximisation, iterative method to find (local) maximum likelihood or maximum a posteriori [MAP] estimates of parameters in statistical models, where the model depends on unobserved latent variables.
HMM	hidden Markov model, statistical Markov model in which the system being modelled is assumed to be a Markov process with unobservable ("hidden") states.
PLSA	probabilistic latent semantic analysis, a representation model in which the probability of co-occurrence of data is modelled as a mixture of conditionally independent multinomial distributions.

G06V 10/86

using syntactic or structural representations of the image or video pattern, e.g. symbolic string recognition; using graph matching

Definition statement

This place covers:

Methods and arrangements which use syntactic or structural representations of the image or video patterns for recognition or understanding where objects can be represented by a variable-cardinality set of symbolic, nominal features.

Syntactic pattern recognition which represents structures by means of strings of symbols and formal language analysis algorithms, such as parsing with grammars.Recognition based on graph matching to find relations between patterns.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

These methods allow the representation of structures by taking into account interrelationships between patterns or their attributes.

Illustrative examples of subject matter classified in this place:

1.
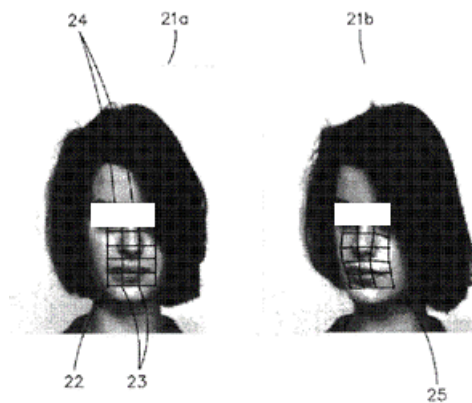
2.

Face recognition by elastic bunch graph matching.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition or machine learning, using clustering	G06V 10/762
Pattern recognition or machine learning, using classification	G06V 10/764
Complex mathematical operations	G06F 17/10
Handling natural language data	G06F 40/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

Hopcroft-Karp algorithm	graph matching algorithm that takes as input a bipartite graph and produces as output a maximum cardinality matching – a set of as many edges as possible with the property that no two edges share an endpoint.

G06V 10/88

Image or video recognition using optical means, e.g. reference filters, holographic masks, frequency domain filters or spatial domain filters

Definition statement

This place covers:

Optical devices which are specially adapted for recognising patterns; methods making use of these devices.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Detection and recognition of an object usually involves optical correlation between an input (optical) image and either a reference optical mask containing the object of interest, or an optical image of the reference object. The convolution is efficiently performed as superposition of these images (dot products) in the Fourier domain, the latter representation being obtained by a lens or a system of lenses.

At least one element of a processing chain for recognising patterns in image and video data may be an optical hardware component, e.g. a ring-wedge detector, or an optical correlator to determine the similarity between two patterns. The effect of the optical element may be present in both spatial domain and in the frequency domain.

Typically, the optical elements used in these approaches are: specially designed filter masks, spatial light modulators [SLM], holographic masks, phase-only filters, acousto-optic cells, waveguides, polarisers, etc.

Illustrative example of subject matter classified in this place:
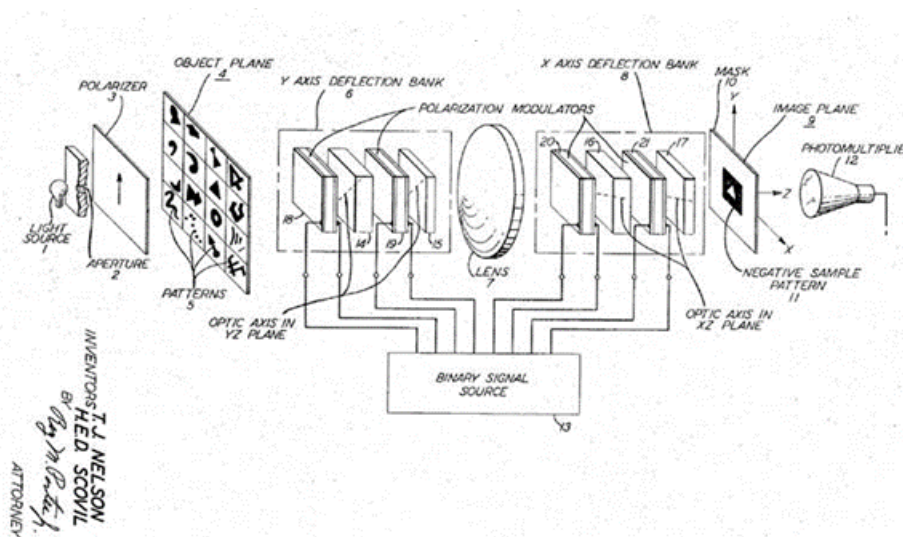
Example of an optical correlator.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Arrangements for image or video recognition or understanding using pattern recognition or machine learning	G06V 10/70

Optical elements per se	G02B
Diffraction optics, systems using spatial filters	G02B 27/46
Spatial light modulators per se	G02F
Optical or electro-optical devices for carrying out mathematical operations	G06E 3/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

SLM	spatial light modulator

G06V 10/94

Hardware or software architectures specially adapted for image or video understanding

Definition statement

This place covers:

Hardware solutions (e.g. individual electronic circuits or networks of interacting electronic devices) or software architectures (e.g. data structures or software libraries), which are specially adapted for pattern recognition or image or video understanding.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

An example of special adaptation of software might include software arranged to perform a sequence of mathematical operations, which can run particularly efficiently on a particular graphical processing unit [GPU].

An example of special adaptation of hardware might be a processor which is designed to perform operations, that are particularly relevant for pattern recognition (e.g. convolutions), in a power-efficient manner. Another example might be a hardware interface, which makes it possible to communicate an extracted visual pattern very efficiently (in terms of speed or bandwidth) to a server for further processing.

Illustrative example of subject matter classified in this place:
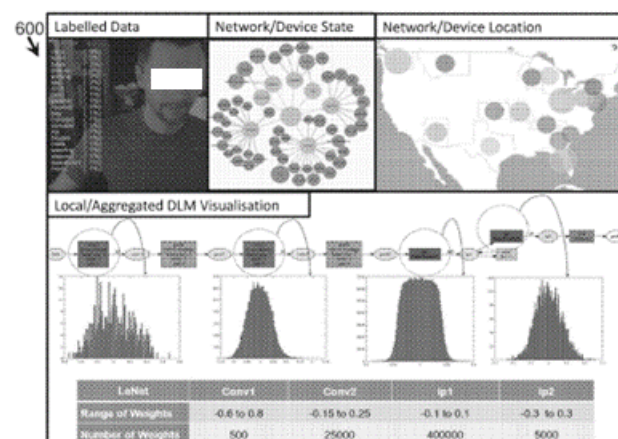
Distributed pattern recognition system in which different resources are placed at different geographical locations.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Optical devices for pattern recognition	G06V 10/88
Sensors specially adapted for fingerprint or palmprint recognition	G06V 40/13
Sensors specially adapted for recognising vascular patterns	G06V 40/145
Sensors specially adapted for eye recognition	G06V 40/19

Informative references

Attention is drawn to the following places, which may be of interest for search:

Processor architectures for image data processing	G06T 1/20
Hardware or software architectures for video coding	H04N 19/42

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

Core	CPU cores in an individual physical CPU
CPU	central processing unit
DSP	digital signal processor
edge device	device that provides an entry point to a digital communication network, such as routers, switches, etc.
GPU	graphical processing unit
LAN	local area network
WAN	wide area network

G06V 10/96

Management of image or video recognition tasks

Definition statement

This place covers:

Processes and devices which control the execution of pattern recognition algorithms.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The control may cause the result of a pattern recognition algorithm to be available within a predetermined time frame, e.g. by prioritising pattern recognition tasks over other tasks, by delegating tasks to processors which are currently idle, by lowering the frame rate (e.g. discarding every other image frame), or by altering the resolution of the image.

The control may also take the urgency of pattern recognition tasks into account; for example, an autonomous vehicle could prioritise tracking a detected pedestrian over tracking other more distant objects.

The control may also be adaptive to the available processing power or to the available bandwidth, e.g. by selecting simpler and less accurate algorithms when being executed on a mobile device, by dynamically altering between batch processing and real-time processing depending on predetermined criteria, or by skipping dispensable pre-processing steps.

Illustrative example of subject matter classified in this place:
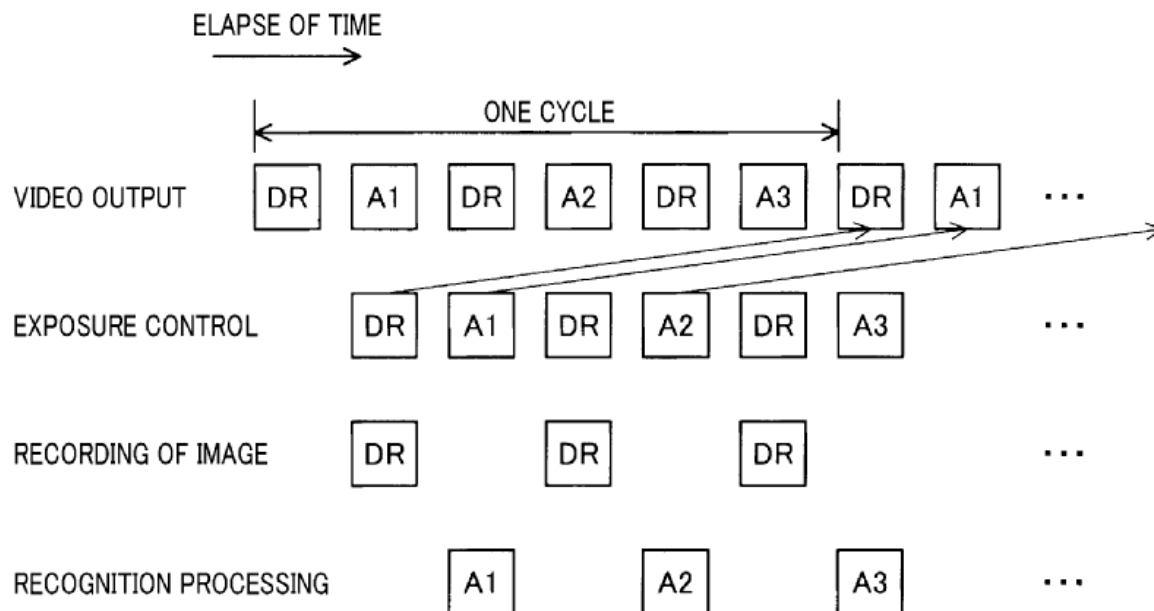
Task scheduling to accommodate different steps, e.g. image recording, exposure control, recognition and video output.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Recognition of scenes or scene-specific elements	G06V 20/00
Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition	G06V 30/00
Image or video recognition or understanding of human-related, animal-related or biometric patterns in image or video data	G06V 40/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image or video recognition or understanding, algorithms using pattern recognition or machine learning	G06V 10/70
Allocating computer resources to programs	G06F 9/50
Data processing for complex mathematical operations	G06F 17/10

G06V 10/98

Detection or correction of errors, e.g. by rescanning the pattern or by human intervention; Evaluation of the quality of the acquired patterns

Definition statement

This place covers:

Methods and arrangements for detecting or correcting errors in an acquired pattern.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Regarding error correction in acquired images, errors may be detected or corrected automatically, or with the help of an operator. The process may also be semi-automatic; for example, it may involve the displaying of an image on a graphical user interface in order to prompt human intervention, if the quality of the image is insufficient for successful recognition.

Detecting errors, in particular, may comprise evaluating the quality of given image or video data in order to assess its suitability for analysis by an automated pattern recognition process. Typical quality criteria are sharpness/blurriness, resolution, contrast and brightness.

More advanced quality assessment algorithms check the image for objects that are only partly visible (e.g. due to occlusions or because parts of the object have moved outside the field of view). These algorithms may also detect the presence of clutter or shadows, and check whether the position and orientation of the object are as expected, or they determine whether the image complies with quality standards of a particular technical application (e.g. the visibility of the eyes in case of biometric authentication).

If the quality of the image or video data is considered insufficient, the process or device may attempt to improve the quality by capturing a further image, potentially by changing parameter settings of the image capturing process (e.g. by switching on active infrared illumination when the captured image is found to be too dark) or by providing the user with instructions on how to re-capture the image.

Illustrative examples of subject matter classified in this place:

1A.
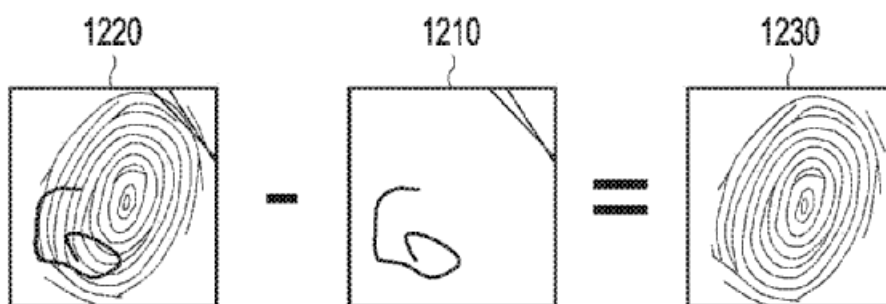
1B.
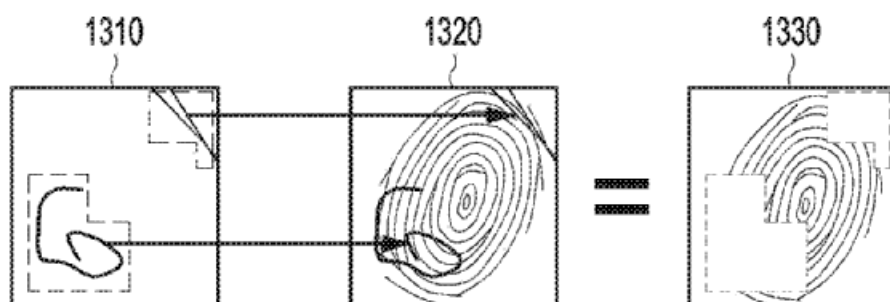
The quality of the acquisition of a fingerprint image is influenced by a hair present on the sensor. A stand-alone image of the hair is either removed (left) from the acquired image or the regions containing the hair are discarded in the subsequent analysis (right).

2.
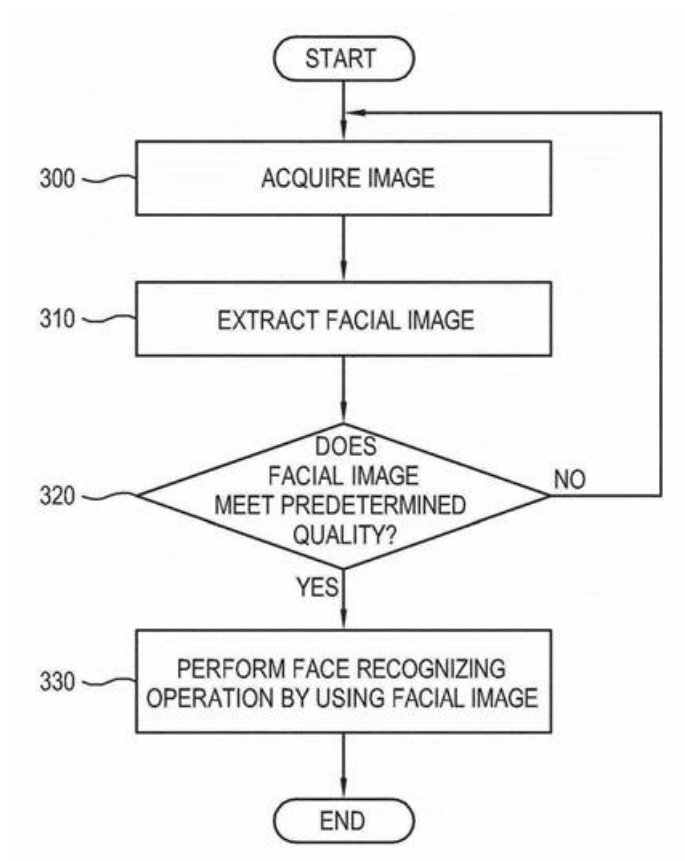
Flowchart according to which face recognition is performed only when the acquired image fulfils a predetermined quality standard.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Recognition of scenes or scene-specific elements	G06V 20/00
Recognition of human or animal bodies within image or video data	G06V 40/10
Recognition of fingerprints or palmprints	G06V 40/12
Recognition of vascular patterns	G06V 40/14
Recognition of human faces, e.g. facial parts, sketches or expressions within images or video data	G06V 40/16
Recognition of eye characteristics within image or video data, e.g. of the iris	G06V 40/18
Maintenance of biometric data or enrolment thereof	G06V 40/50
Multimodal biometrics	G06V 40/70

Informative references

Attention is drawn to the following places, which may be of interest for search:

Validation or performance evaluation for pattern recognition	G06V 10/776
Investigating the presence of flaws or contamination by the use of optical means	G01N 21/88
Investigating the presence of flaws in materials by the use of thermal means	G01N 25/72
Image analysis	G06T 7/00
Arrangements for detecting and preventing errors in the information received	H04L 1/00
Details of television systems	H04N 5/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

quality	quality within the meaning of the present group is a property of the acquired pattern insofar as it has an effect on the accuracy or performance of the pattern recognition process

G06V 20/00

Scenes; Scene-specific elements (control of digital cameras H04N 23/60)

Definition statement

This place covers:

Scene-specific image or video recognition or understanding according to the category of scene that is perceived by the observer or the scene-specific processing performed.

Examples of different categories of scenes are underwater scenes, terrestrial scenes, augmented reality scenes, albums, collections, shared content such as social networks photos or video, videos

such as a film of a TV broadcasting. The context of an image or video includes scenes under surveillance, traffic scenes, scenes exterior to a vehicle, scenes on the interior of a vehicle. Various types of objects can be analysed, such as three-dimensional objects, microscopic objects, food, trinkets, scene text, etc. Examples of scene-specific processing includes semantic and syntactic analysis and classifying the scene content.

References

Limiting references

This place does not cover:

Devices for controlling television cameras, e.g. remote control	H04N 23/60

Informative references

Attention is drawn to the following places, which may be of interest for search:

Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition	G06V 30/00
Recognition of biometric, human-related or animal-related patterns in image or video data	G06V 40/00
Measuring arrangements characterised by the use of optical means	G01B 11/00
Systems using the reflection or reradiation of acoustic waves, e.g. sonar systems	G01S 15/00
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements	G06F 3/00
Image analysis	G06T 7/00
Burglar, theft or intruder alarms	G08B 13/00
Selective content distribution, e.g. interactive television, Video on Demand [VoD]	H04N 21/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

scene	visual representation of the world or of some elements of it, as captured by a sensor or generated by a computer

G06V 20/05

Underwater scenes

Definition statement

This place covers:

Detection, identification and recognition of objects specifically adapted to underwater scenes.

Categorising underwater objects.

Detection, identification and recognition of underwater structures, such as oil or gas pipes.

Detection, identification and recognition of objects or animals located on the sea floor.

Adapting the recognition according to the underwater conditions, e.g. light scattering or absorption, artifacts, blurring, non-uniform lighting, etc.

Recognising underwater objects in the context of simultaneous localisation and mapping [SLAM].

Illustrative example of subject matter classified in this place:
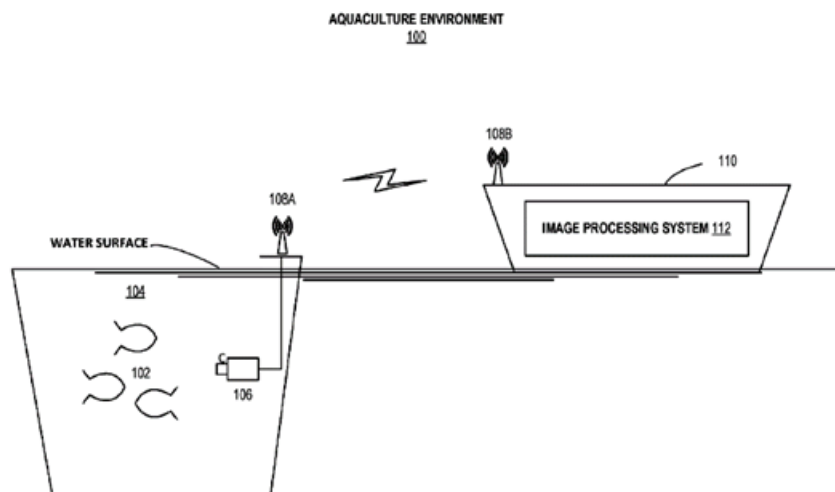
Example of a system for analysing underwater scenes.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Identifying an image sensor based on its output data	G06V 20/90
Recognition of biometric, human-related or animal-related patterns in images or video	G06V 40/00
Underwater vessels, e.g. submarines	B63G 8/00
Systems using the reflection or reradiation of acoustic waves, e.g. sonar systems	G01S 15/00
Image analysis in general	G06T 7/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

AUV	autonomous underwater vehicles
ROV	remotely operated vehicles
UUV	unmanned underwater vehicles
marine snow	presence of organic material falling from upper layers of the water column

G06V 20/10

Terrestrial scenes (scenes under surveillance with static cameras G06V 20/52; scenes perceived from the exterior of a vehicle G06V 20/56; scenes perceived from the interior of a vehicle G06V 20/59)

Definition statement

This place covers:

Arrangements and methods specifically adapted to recognise terrestrial scenes:

- Recognising urban or other man-made structures;
- Recognising network patterns such as roads or rivers;
- Recognising vegetation, agricultural fields, etc.;
- Deriving scene properties, e.g. the amount of clutter in terms of population with image objects, the type of background, the existence of various types of objects, detection of the skyline, clouds, weather conditions, etc.;
- Obtaining semantic attributes or information from the scene, such as types of objects and their inter-relations, quantifying the geometric placement of the objects;
- Recognising terrestrial objects in the context of simultaneous localisation and mapping [SLAM].

Illustrative example of subject matter classified in this place:
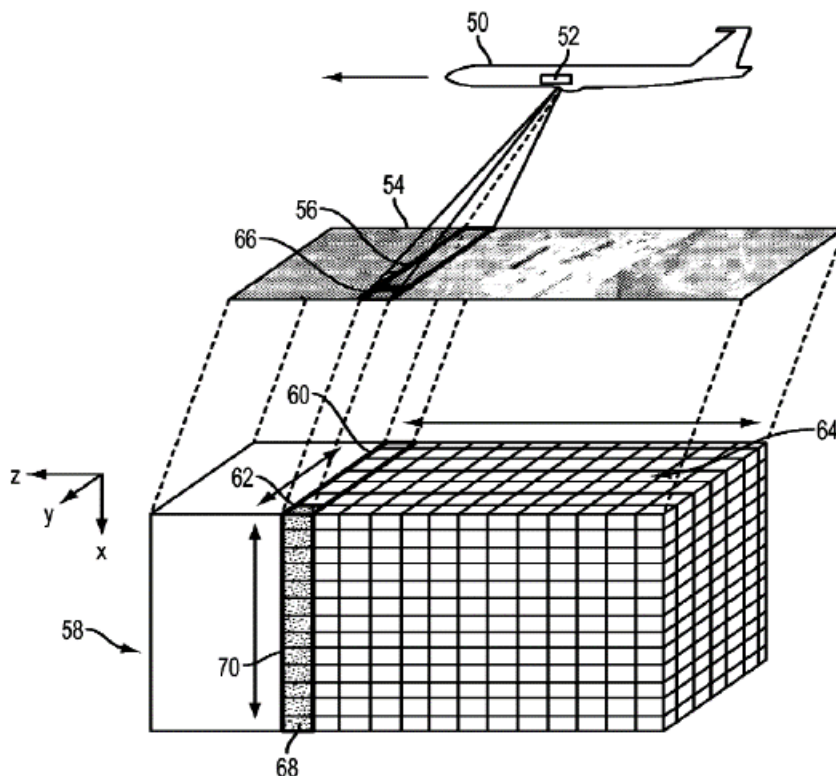
Perspective view of the region imaged by consecutive acquisitions of a hyperspectral sensor.

References

Limiting references

This place does not cover:

Surveillance or monitoring of activities, e.g. for recognising suspicious objects	G06V 20/52
Recognition or understanding of scenes outside a vehicle by using sensors mounted on the vehicle	G06V 20/56
Recognition or understanding of scenes inside of a vehicle	G06V 20/59

Informative references

Attention is drawn to the following places, which may be of interest for search:

Printing processes to produce particular kinds of printed work, e.g. patterns; Maps; Sea or meteorological charts	B41M 3/02
Navigation	G01C 21/00
Systems using the reflection or reradiation of radio waves, e.g. radar systems;	G01S 13/00
Systems using the reflection or reradiation of electromagnetic waves other than radio waves, e.g. lidar systems	G01S 17/00
Meteorology	G01W 1/00
Information retrieval of image data	G06F 16/50
Information retrieval of video data	G06F 16/70
Segmentation for general image processing	G06T 7/10
Motion image analysis	G06T 7/20

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

aerial imagery	images taken from an aircraft or other flying object (e.g. aircrafts, helicopters, UAVs, balloons, etc.)
band	response sensed by the optical sensor to a certain range of wavelength
endmember	material that has a spectrally unique signature in the wavelength bands used to collect the image
GIS	geographic information system
Hughes Phenomenon/Curse of dimensionality	when the dimensionality of the data increases, the volume of the data-space increases. Thus, if the dimensionality of a fixed amount of data is increased, the data becomes sparse in the increased data-space. This causes the classifier's performance to deteriorate. Increasing the amount of data or decreasing the dimensionality of the data will improve the performance of the classifier.
hyperspectral image	multi-band image where the z dimension corresponds to consecutive spectral wavelengths ranges
multispectral image	multi-band image where the z dimension corresponds to spectral wavelengths ranges (not necessarily consecutive)

remote sensing	process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation from a satellite or aircraft
SAR	Synthetic aperture radar
spectral image cubespectral image cubes	data having 3 dimensions, 2 spatial (x, y) and a third spectral dimension
UAV	unmanned aerial vehicles

G06V 20/13

Satellite images

Definition statement

This place covers:

Recognising patterns corresponding to different image structures (e.g., objects) in remotely sensed satellite images or video, e.g. optical data (images or video), GPS, radar, LIDAR measurement data or in combination.

Object detection, deriving hyperspectral signatures from objects within satellite images. Categorisation of man-made objects / image targets within satellite images.

Vegetation detection or monitoring canopy growth within satellite images.

Cloud detection and cloud mask segmentation within satellite images.

3D measurement of man-made objects, such as building roofs, within satellite images.

Change detection, e.g. assessing influence of natural disasters, presence of new objects (anomalies) against a known background within satellite images.

Weather condition monitoring by image or video analysis of satellite images.

Illustrative example of subject matter classified in this place:
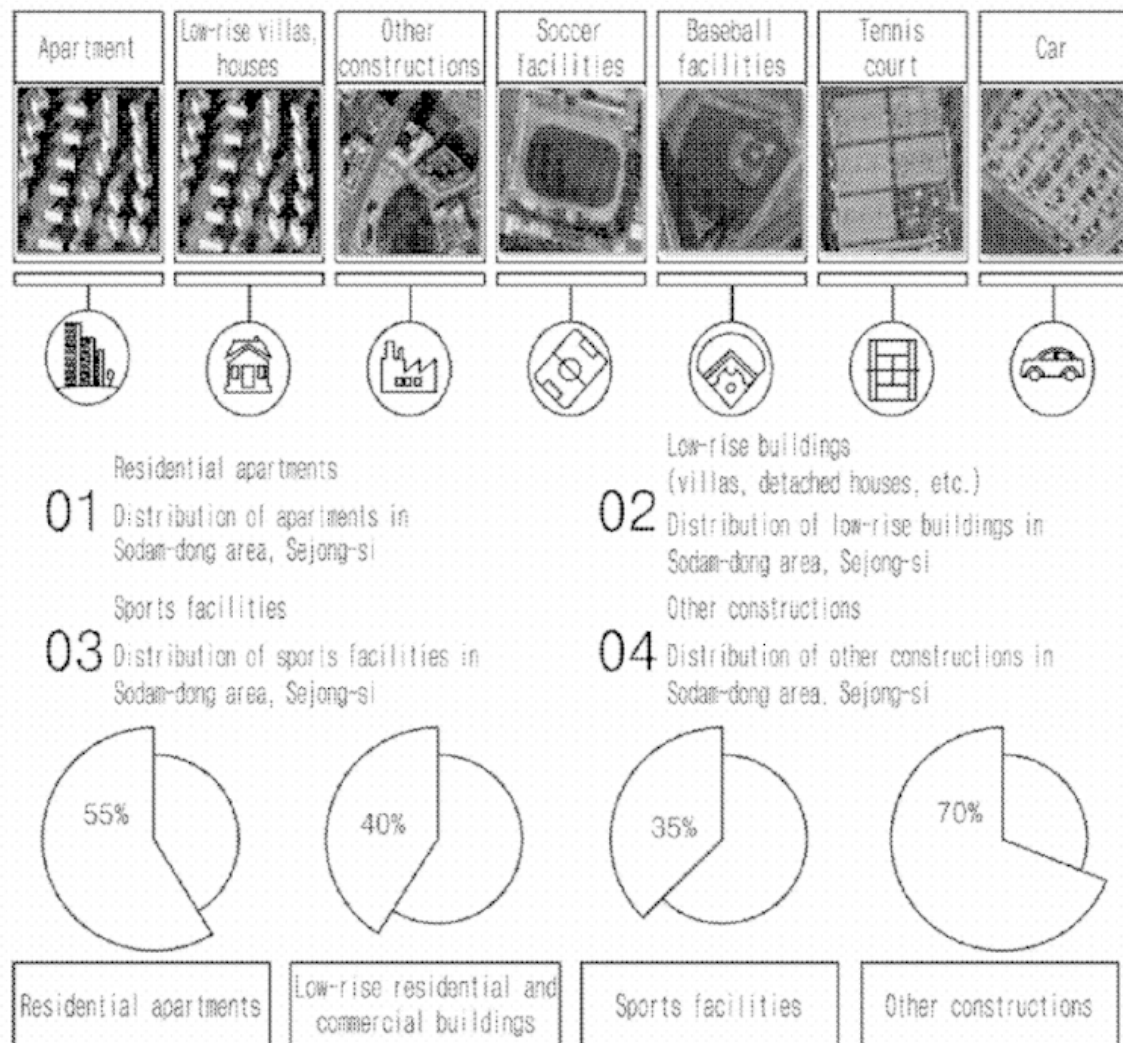
Automatic classification of objects in satellite images.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition and image understanding in terrestrial scenes with images taken by planes or drones	G06V 20/17
Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Recognition of technical drawings and geographical maps	G06V 30/422
Navigation	G01C 21/00
Systems using the reflection or reradiation of radio waves, e.g. radar systems	G01S 13/00
Systems using the reflection or reradiation of electromagnetic waves other than radio waves, e.g. lidar systems	G01S 17/00
Satellite radio beacon positioning systems, e.g. GPS	G01S 19/00
Information retrieval of image data	G06F 16/50

Information retrieval of video data	G06F 16/70
Segmentation for general image processing	G06T 7/10
Motion image analysis	G06T 7/20

Special rules of classification

This group covers techniques specifically adapted for remotely sensed satellite images or video. Recognising patterns in aerial images or video acquired from aircrafts, helicopters, unmanned aerial vehicles (UAVs), balloons, etc., is classified in group G06V 20/17. The difference between these two groups consists of how the images are acquired. Image or video techniques classified in group G06V 20/13 lack perspective (depth) information, while in the group G06V 20/17, images acquired from aircrafts, helicopters, UAVs, etc. contain the perspective (depth) information.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

hyperspectral images	images in which one continuous spectrum is measured for each pixel. Generally, the spectral resolution is given in nanometres or wave numbers.
remote sensing	process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation from a satellite or aircraft

G06V 20/17

taken from planes or by drones

Definition statement

This place covers:

Recognising patterns corresponding to different image structures (e.g. objects) in aerial images or video acquired from aircraft, helicopters, unmanned aerial vehicles (UAVs) or drones, balloons, etc.

Categorisation of man-made objects/image targets in aerial images or video.

Vegetation detection or monitoring canopy growth in aerial images or video.

3D measurement of man-made objects such as building roofs wherein the scene is taken from planes or by drones.

Inspection of buildings or other man-made objects, e.g. damage classification, wherein the scene is taken from planes or by drones.

Recognising flying entities, such as insects or birds, from images or video captured by drones.

Recognising or monitoring the activity of military targets in aerial images or video acquired from aircrafts, helicopters, UAVs or drones, balloons.

Illustrative example of subject matter classified in this place:

1.
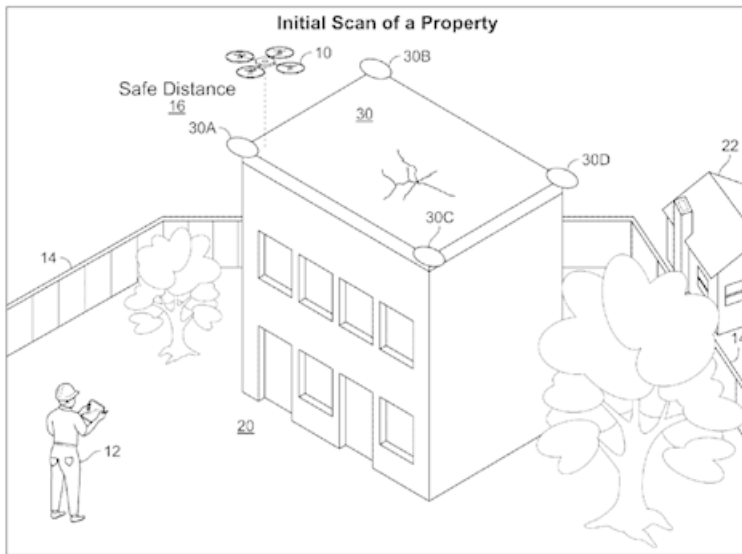
Recognising and assessing the damage to a building using a drone.

2.
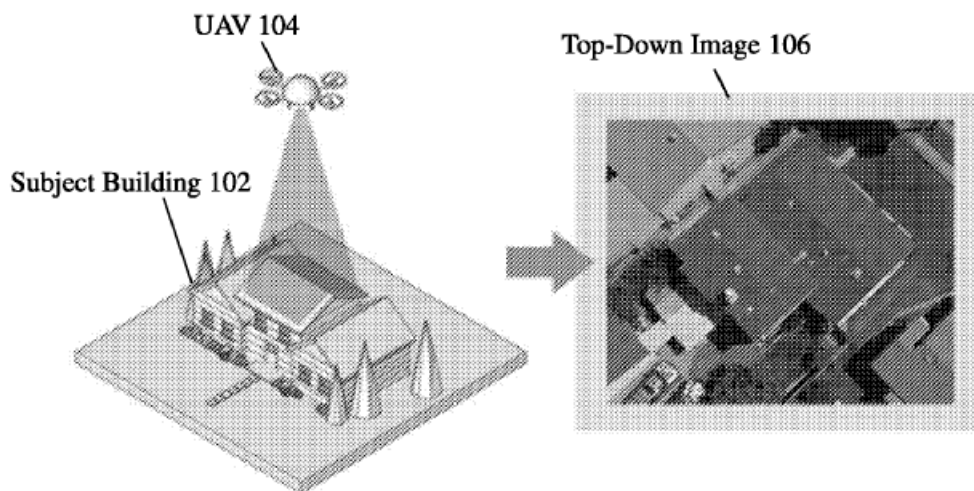
Determining the surface of roofs using UAVs.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Pattern recognition and image understanding in terrestrial scenes with images taken from satellites	G06V 20/13
Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Recognition of technical drawings and geographical maps	G06V 30/422

Photogrammetry or videogrammetry, e.g. stereogrammetry; Photographic surveying	G01C 11/00
Radar or analogous systems, specially adapted for specific applications for mapping or imaging	G01S 13/89
Lidar systems, specially adapted for mapping or imaging	G01S 17/89
Information retrieval of image data	G06F 16/50
Information retrieval of video data	G06F 16/70
Segmentation for general image processing	G06T 7/10
Analysis of motion in images	G06T 7/20

Special rules of classification

This group covers techniques specifically adapted for aerial images or video acquired from aircraft, helicopters, unmanned aerial vehicles (UAVs), balloons, etc.

Recognising patterns in remotely sensed satellite images or video are covered by group G06V 20/13. The difference between these two groups lies in the manner of acquisition of the images. Image or video techniques covered by group G06V 20/17 relate to perspective (depth) information, while techniques covered by group G06V 20/13, relate to images acquired from satellites, which lack perspective (depth) information.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

aerial imagery	images taken from an aircraft or other flying object (e.g. aircrafts, helicopters, UAVs, balloons, etc.)
remote sensing	process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation from a satellite or aircraft
UAV	unmanned aerial vehicle

G06V 20/20

in augmented reality scenes

Definition statement

This place covers:

Object recognition operating in an augmented reality environment and adapted to provide additional information about a scene to a user. The underlying processing may involve one or more of the following steps:

1. acquiring an image of a real scene by an image capture device;

2. detecting and recognising objects in the depicted scene;

3. acquiring additional information which is related to these objects (e.g. from a database);

4. presenting this information on the original image in an overlaid / superimposed manner.

The object detection and recognition processes need to be fast due to real-time constrains. For this reason, additional information provided by other sensors (e.g. accelerometers, gyroscopes, GPS, solid

state compasses or RFID) can be used to define or limit the analysis based on the information they provide.

Examples of adaptations include:
- the way in which objects of interest are detected and recognised in the image: feature-based detection, geometrical proximity to the object of interest or optical character recognition [OCR] of text in a scene, etc.;
- the way in which additional object related information is obtained, e.g. from a database stored locally in the device, or by internet search, etc.;
- the purpose for which the application is designed: e.g. for visually impaired people, for driver assistance systems, for chirurgical interventions, for presentation of chemical structures, as interactive guide for attractions and museums, or for use on construction sites, etc.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

Object recognition for augmenting scene information and adapted for head mounted displays and portable devices needs to be fast due to real-time constrains. For this reason, additional information provided by other sensors (e.g. accelerometers, gyroscopes, GPS, solid state compasses, RFID) may be used to define or limit the space subject to analysis.

Different approaches to camera pose estimation and registration may be essential to successful object recognition.

Illustrative example of subject matter classified in this place:
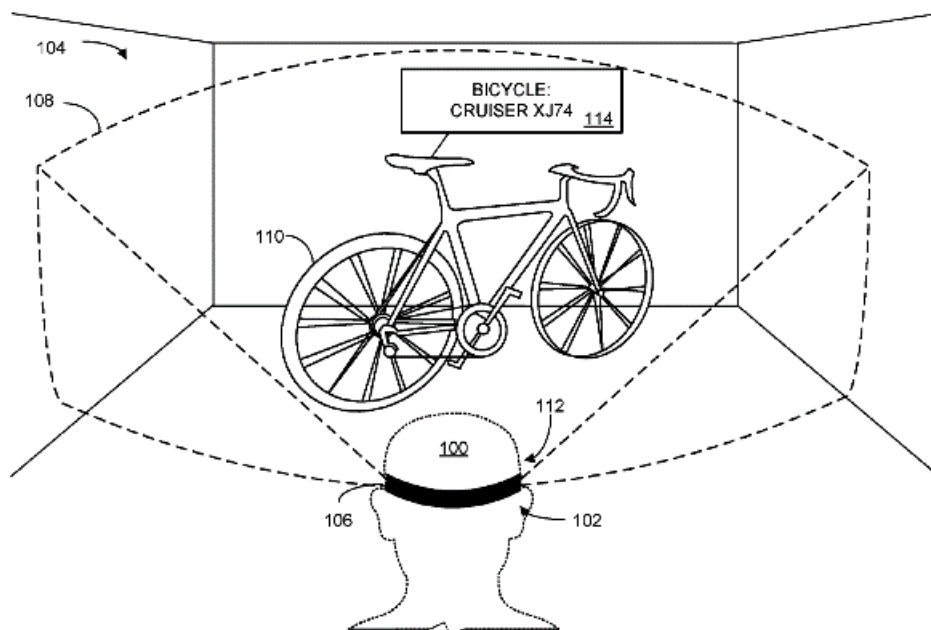
Recognition of a real-world object by a head-mounted computing device.

Relationships with other classification places

Determining position or orientation of objects or cameras is covered by group G06T 7/70. Input arrangements or combined input and output arrangements for interaction between user and computer is covered by group G06F 3/01.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Descriptors for shape, contour or point-related descriptions of extracted image or video features, e.g. scale invariant feature transform [SIFT] or bags of words [BoW]; Salient region features	G06V 10/46
Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Labelling scene content	G06V 20/70
Character recognition within an image or video; Document-oriented image-based pattern recognition	G06V 30/00
Recognition of biometric, human-related or animal-related patterns in image or video data	G06V 40/00
Input arrangements or combined input and output arrangements for interaction between user and computer	G06F 3/01
Digital output to display device	G06F 3/14
Analysis of motion in images	G06T 7/20
Manipulating 3D models or images for computer graphics	G06T 19/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

AR	augmented reality
AR overlay	images, videos, 3D or other information types superimposed over a target object
Field of View [FoV]	area that can be observed through a capture device lens. Depending on the lens focus, the field of view can be adapted and can vary in size.
OCR	optical character recognition
VR	virtual reality

G06V 20/30

in albums, collections or shared content, e.g. social network photos or video

Definition statement

This place covers:

Generating groups or clusters from images or video based on their similarity, based on events, backgrounds, identified individuals, etc.

Comparing and forming connections between image collections using matching, classification and clustering.

Detecting or recognising events in an image collection and ordering these events in an event timeline, based on image content.

Construction of a social network by analysis of an image collection.

Illustrative example of subject matter classified in this place:
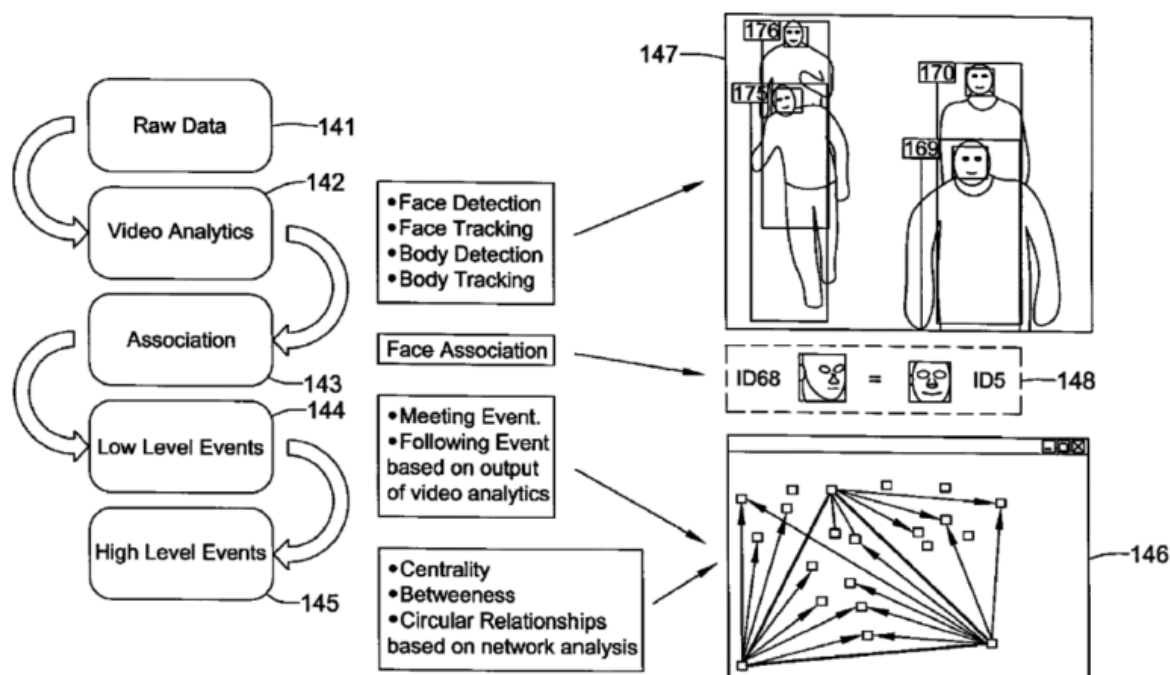
Steps for constructing a dynamic social network from raw video data of observations of people.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Global feature extraction by analysis of the whole pattern, e.g. global shape, global boundary descriptors or involving frequency domain transformations or autocorrelation	G06V 10/42
Local feature extraction by analysis of parts of the pattern, e.g. by detecting edges, contours, loops, corners, strokes or intersections; Connectivity analysis, e.g. of connected components, edge linking or neighbouring slice analysis	G06V 10/44
Recognition using clustering in general	G06V 10/762
Recognition of patterns in video content, e.g. in a film or a TV broadcasting	G06V 20/40
Recognition of scenes under surveillance or monitoring activities, e.g. recognising suspicious objects	G06V 20/52
Labelling scene content, e.g. semantic segmentation	G06V 20/70
Recognition of human bodies within image or video data	G06V 40/10
Recognition of human faces, e.g. facial parts, sketches or expressions within image or video data	G06V 40/16
Information retrieval of image data	G06F 16/50
Information retrieval of video data	G06F 16/70

Special rules of classification

Detection of events in video surveillance, in particular suspicious activities or objects, is classified in group G06V 20/52. Labelling of scene content, e.g. by semantic segmentation, is classified in group G06V 20/70.

G06V 20/40

in video content (extracting overlay text G06V 20/62; video retrieval G06F 16/70; processing of video elementary streams in video servers H04N 21/234; processing of video elementary streams in video clients H04N 21/44)

Definition statement

This place covers:

Video summarisation/abstraction, e.g. key-frame extraction, extracting of video features or fingerprints, extraction of representative shots, detecting important frames by analysing the reactions of the viewers or by monitoring parts of the video, such as the TV logo.

High-level semantic clustering, classification and understanding of video scenes, e.g. detection, labelling or Markovian modelling. Examples of video content subject to such analysis are sport broadcast events or TV news.

Low-level semantic clustering or determination of sections in videos such as scenes and shots; classification of shots, e.g. as close-up shot, medium shot or long shot.

Extraction of features, e.g. histogram similarity measures, manifolds, by use of video fingerprints, etc. Examples of low-level features are colour or texture-based features, local interest points (key-points), filter responses, edge features, local descriptors (SIFT, SURF, etc.) or combinations of them (see also group G06V 10/40). Examples of high-level features are features related to camera motion (tracking visual features), the presence of skin, the number of faces present, the size of faces or other human features visible, text or other objects that are identifiable in each frame.

Matching video sequences, e.g. by frame or temporal analysis.

Segmenting video sequences, e.g. parsing or cutting the sequence.

Video categorisation, e.g. classify video content into sport/music/news or recognise commercials in media content for substitution.

Sport games analysis, e.g. tactic analysis in sport videos for assistance of coaches and players; final pitching shot indexing for baseball game; indexing the important parts, such as shots, score points, etc; video monitoring the score table.

Generation of compact representations of the video sequence as a result of pattern recognition or image understanding, e.g. creating thumbnails or representative icons.

Detection and recognition of harmful/sexual/violent content.

Discovery of relationships between objects or persons in videos.

Detecting a key/anchor person from a video; characterising the main characters.

Association of a video with semantic information (e.g. keywords) to describe the content (using e.g. Markov random fields).

Generation of semantic labels using a graph which describes the video content, where the nodes are objects or activities and edges are the relationships between them.

Illustrative examples of subject matter classified in this place:

1.
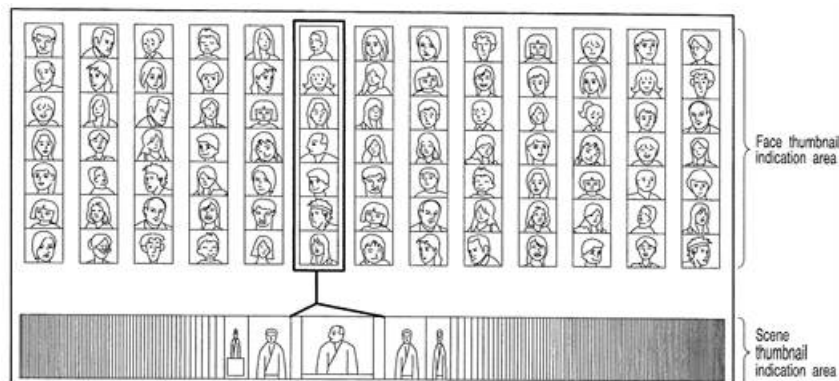
Clustering of the representative frames containing a given face and creation of face thumbnails of a video sequence containing faces.

2.
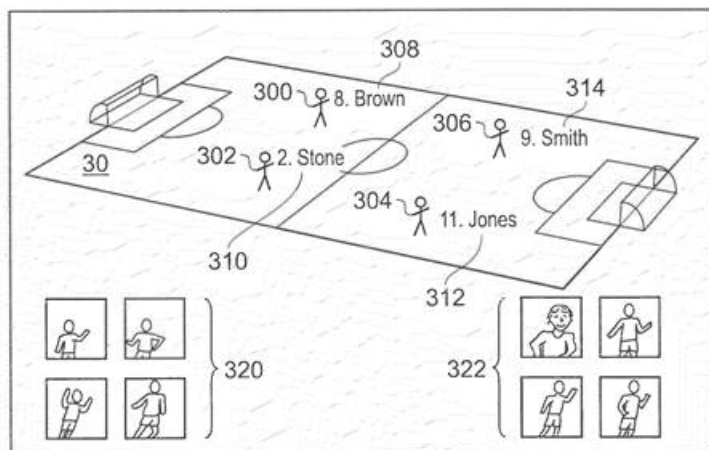
Recognising football players in a football match and displaying the representative shots in which a certain player was active.

References

Limiting references

This place does not cover:

Extracting overlay text	G06V 20/62
Information retrieval of video data	G06F 16/70
Processing of video elementary streams in video servers	H04N 21/234
Processing of video elementary streams in video client devices	H04N 21/44

Informative references

Attention is drawn to the following places, which may be of interest for search:

Arrangements for image or video understanding in general	G06V 10/00

Global feature extraction by analysis of the whole pattern, e.g. global shape, global boundary descriptors or involving frequency domain transformations or autocorrelation	G06V 10/42
Local feature extraction by analysis of parts of the pattern, e.g. by detecting edges, contours, loops, corners, strokes or intersections; Connectivity analysis, e.g. of connected components, edge linking or neighbouring slice analysis	G06V 10/44
Pattern recognition or machine learning in images or video using clustering	G06V 10/762
Recognition of scenes under surveillance or monitoring activities, e.g. recognising suspicious objects	G06V 20/52
Labelling scene content, e.g. deriving syntactic or semantic representations	G06V 20/70
Recognition of human or animal bodies	G06V 40/10
Recognition of human faces, e.g. facial parts, sketches or expressions	G06V 40/16
Recognition of movements or behaviour, e.g. gesture recognition	G06V 40/20
Analysis of motion in images	G06T 7/20
Image analysis using motion-based segmentation	G06T 7/215
Editing; Indexing; Addressing; Timing or synchronising; Monitoring; Measuring tape travel	G11B 27/00
Television picture signal circuitry for video frequency region	H04N 5/14

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

video fingerprinting	class of dimension reduction techniques for identifying, extracting and summarising characteristic components of a video enabling that video to be uniquely identified
video summarisation	generation of a short summary of the content of a longer video by selecting and presenting the most informative or interesting video frames

G06V 20/50

Context or environment of the image

Definition statement

This place covers:

Recognising and understanding scenes according to the context or the environment of the scene, e.g. the type of scene or the situation in which it is acquired.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition based on the type of objects	G06V 20/60

Labelling scene content, e.g. deriving syntactic or semantic representations	G06V 20/70
Recognition of biometric, human-related or animal-related patterns in image or video data	G06V 40/00
Recognising movements or behaviour, e.g. gesture recognition	G06V 40/20
Radar or analogous systems for traffic control	G01S 13/91
Radar or analogous systems for anti-collision purposes of land vehicles	G01S 13/931
Lidar systems for anti-collision purposes of land vehicles	G01S 17/931
Analysis of motion in images	G06T 7/20
Image analysis for determining position or orientation of objects or cameras	G06T 7/70
Traffic control systems for road vehicles	G08G 1/00

Special rules of classification

Recognising different types of objects is classified in group G06V 20/60. Classification in groups G06V 20/60 and G06V 20/50 or subgroups is applied when a certain type of object is recognised in a specific scene context. For example, recognition of license plates, covered by group G06V 20/62, is classified also in group G06V 20/52 when the recognition is performed in the context of a scene under surveillance, such as a parking lot.

G06V 20/52

Surveillance or monitoring of activities, e.g. for recognising suspicious objects (recognising microscopic objects G06V 20/69)

Definition statement

This place covers:

Detection and recognition of objects or events in scenes under surveillance, for example by:
- detecting activity in a restricted area/zone, e.g. detecting intrusion, unsafe situations around working equipment or machines, monitoring of trespassing of a specific area;
- excluding certain spatial/temporal fragments, for example, for privacy protection (e.g. input of a PIN for a cash dispenser surveyed by a camera);
- detecting hazards: fire, explosions, smoke, contamination, fluid spills, etc.; contamination from pollutants, e.g. petroleum; dangers for occupational injuries; detecting flashes originating from machine guns;
- applying pattern recognition or image understanding techniques for counting various image objects (objects of interest, people, etc.), monitoring queues;
- detecting and recognising hidden objects, ammunition, explosives, e.g. as in airport luggage scanner;
- recognising movements/trajectories, determining paths of the objects in the surveyed scene, e.g. detect the flows of persons in public places;
- identification of certain image objects/persons based on prior information; selection of the relevant surveyed scenes;
- identification and re-identification of image objects/persons, i.e. identification of the same person at different times or in different places along image sequences;
- occupancy or presence detection, e.g. monitoring the filling state of the shelves in a supermarket, keeping track of empty places in a parking lot, elevator occupancy monitoring, seat occupancy in public spaces, e.g. cinemas, concerts, etc.;

- detecting presence for intelligent building control, e.g. for switching off light, for controlling the air conditioning systems, etc.;
- detecting anomalous activities or suspicious behaviour, such as vandalism, robbery, loitering, etc.;
- detecting and recognising suspicious objects or objects left behind;
- monitoring people habits, i.e. in a wearable computing setting (eating patterns, sleeping patterns, washing habits, etc.);
- monitoring queues, predicting queue waiting time, etc.;
- recognising static or dynamic crowd, e.g. crowd congestion.

The subgroup G06V 20/54 concerns recognising and understanding of traffic scenes, by detection, identification, classification and recognition of traffic patterns, e.g. cars on the roads, traffic junctions, traffic jams, estimating the travel time.

Illustrative examples of subject matter classified in this place:
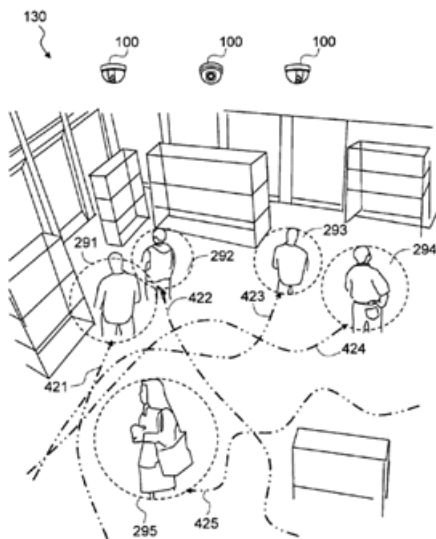
1.

Monitoring activities for a scene under surveillance.

2.
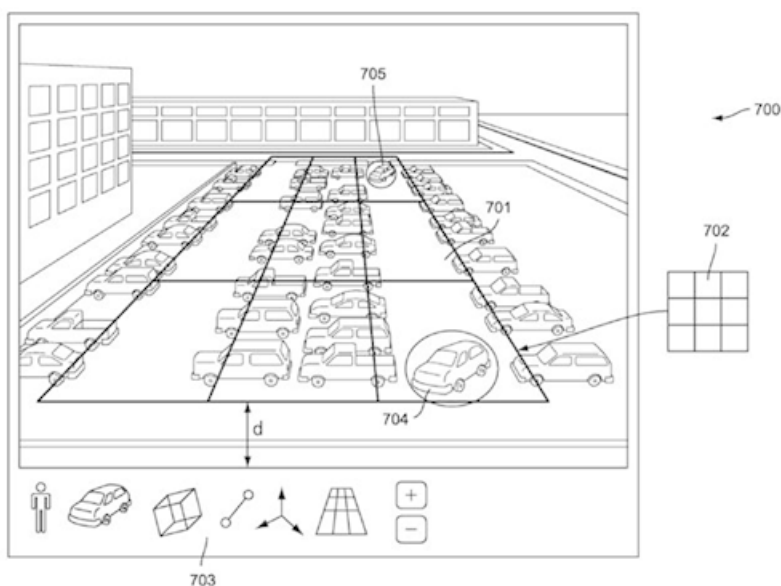
Monitoring the occupancy of a parking lot for a scene under surveillance.

References

Limiting references

This place does not cover:

Recognising microscopic objects	G06V 20/69

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition based on the type of objects	G06V 20/60
Labelling scene content, e.g. deriving syntactic or semantic representations	G06V 20/70
Recognition of biometric, human-related or animal-related patterns in image or video data	G06V 40/00
Recognising movements or behaviour, e.g. gesture recognition	G06V 40/20
Analysis of motion in images	G06T 7/20
Burglar, theft or intruder alarm	G08B 13/00
Details of television systems	H04N 5/00
Closed-circuit television systems, i.e. systems in which the signal is not broadcast	H04N 7/18

G06V 20/56

exterior to a vehicle by using sensors mounted on the vehicle

Definition statement

This place covers:

Detection, identification and recognition of road lanes, lane markings and borders, free road ahead.

Lane and road marking categorisation, e.g. solid lines, dashed lines, markings at pedestrian crossings, direction indicating arrows, etc.

Estimation of road geometry characteristics, such as curvature, slope or elevation, using e.g. disparity maps of road surfaces or the relative motion of surrounding objects using a clothoidal lane model.

Detection of physical entities located at the side of the road, such as structural barriers (e.g. wall, guardrail), delineators and markers.

Recognition of the drivers' driving pattern in relation to the road lanes perceived from the vehicle.

Recognising the trajectory of a car relative to the road.

Detecting the drivable area ahead of the host vehicle, or of the clear path.

Detection or recognition of road surface characteristics, e.g. cracks, holes.

Detection, classification and recognition of road signs, indicators, etc.

Detection or recognition of potential obstacles, e.g. vehicles ahead, pedestrians.

Recognising surrounding objects by the analysis of their relative position or velocity, possibly with the aid of additional sensors.

Recognition of surrounding objects with the aid of a map of the environment. Categorising vehicles, e.g. car, lorry, bicycle.

Detection or recognition of available parking places; parking assistance by recognising surrounding objects and producing an image of the environment during the parking process with an overview of the host vehicle surroundings, such as a bird-eye view.

Detection of foreign matter on the windshield, e.g. water, dirt, snow.

Adapting the recognition according to the weather conditions, e.g. rain, fog, snow.

Recognition of illumination non-uniformities, e.g. discriminating between objects and shadows.

Recognition of scene objects using special illumination, e.g. infrared light for night vision.

Recognition and compensation for the effects of non-uniformities in illumination, e.g. shadows.

Recognition of light-casting objects, such as traffic lights, lights of the cars ahead, etc.

Illustrative examples of subject matter classified in this place:

1.
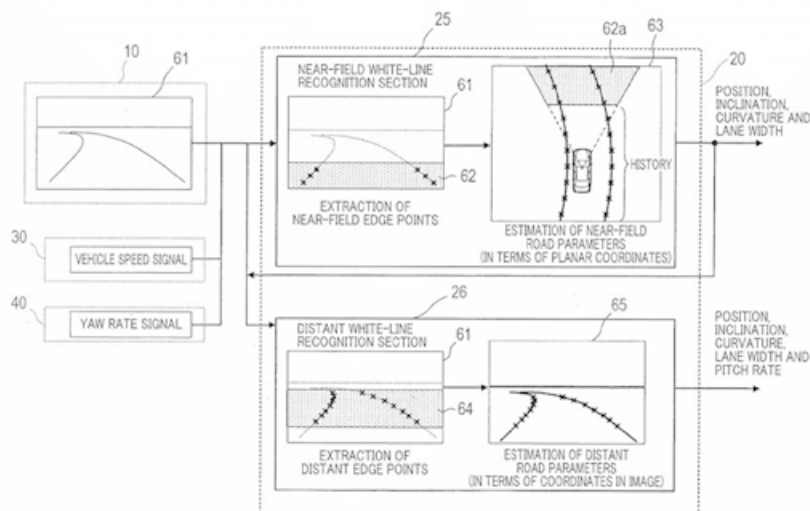
Recognition of lane markers for autonomous driving.

2.
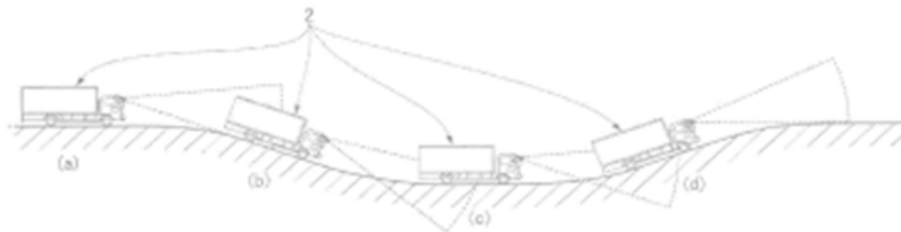
3.
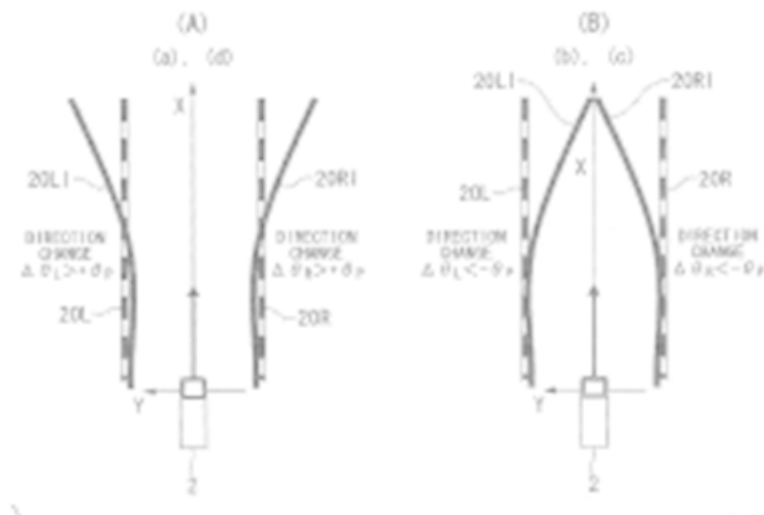
Recognition of the road geometry (e.g. its slope) by image analysis.

Relationships with other classification places

Recognising and understanding of scenes for autonomous driving makes extensive use of a mix of sensors or modalities which are classified in different places in IPC (see the informative references indicated below).

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Drive control systems specially adapted for autonomous road vehicles	B60W 60/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition or understanding of scenes inside of a vehicle	G06V 20/59

Recognition based on the type of objects	G06V 20/60
Character recognition	G06V 30/00
Recognition of human or animal bodies, e.g. pedestrians	G06V 40/10
Navigation	G01C 21/00
Radar or analogous systems for anti-collision purposes of land vehicles	G01S 13/931
Lidar systems for anti-collision purposes of land vehicles	G01S 17/931
Control of position, course or altitude of land, water, air or space vehicles, e.g. automatic pilot	G05D 1/00
Analysis of motion in images	G06T 7/20
Image analysis for determining position or orientation of objects or cameras	G06T 7/70
Analysis of captured images to determine intrinsic or extrinsic camera parameters, i.e. camera calibration	G06T 7/80
Traffic control systems for road vehicles	G08G 1/00

Special rules of classification

The functions of acquisition, pre-processing, feature extraction, pattern recognition and machine learning classified in group G06V 10/00 are also classified in group G06V 20/56 according to the function. For instance, special illumination (e.g. infrared) used for night vision, is classified in group G06V 10/143 and in group G06V 20/56. Other examples are techniques for determination of a region of interest (ROI) defining the obstacles ahead, classified in group G06V 10/25 and in group G06V 20/56.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

ADAS	"advanced driver-assistance systems": technologies that assist drivers in driving and parking functions
AV	"autonomous vehicle": vehicle that is capable of driving itself
ECU	"electronic control unit": an embedded unit in the vehicle that controls one or more electrical systems, such as the engine control unit or the human-machine interface

G06V 20/59

inside of a vehicle, e.g. relating to seat occupancy, driver state or inner lighting conditions

Definition statement

This place covers:

Recognising seat occupancy, e.g. forward or rearward facing child seat.

Recognising driver or occupant position, e.g. for automatic seat adjustment, adjustment of the driving wheel or mirrors.

Recognising the drivers' state, behaviour, emotions, e.g. attention, drowsiness, hands on the wheel, drivers' gaze ("eyes-off-road"), potential alcohol consumption, etc.

Recognising the state of vehicle controls, e.g. dashboard indicators such as speedometers, fuel meters, etc.

The recognition may be performed on images taken from an on-board camera located within the vehicle or from images taken from cameras located outside of the vehicle.

Illustrative examples of subject matter classified in this place:

1.
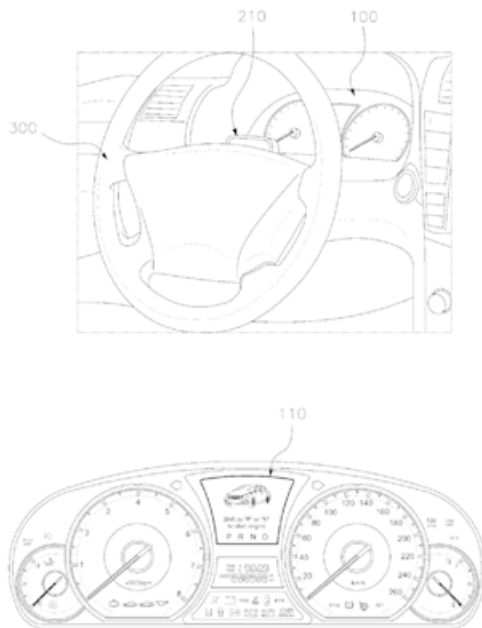
Recognising indicators on the dashboard.

2.
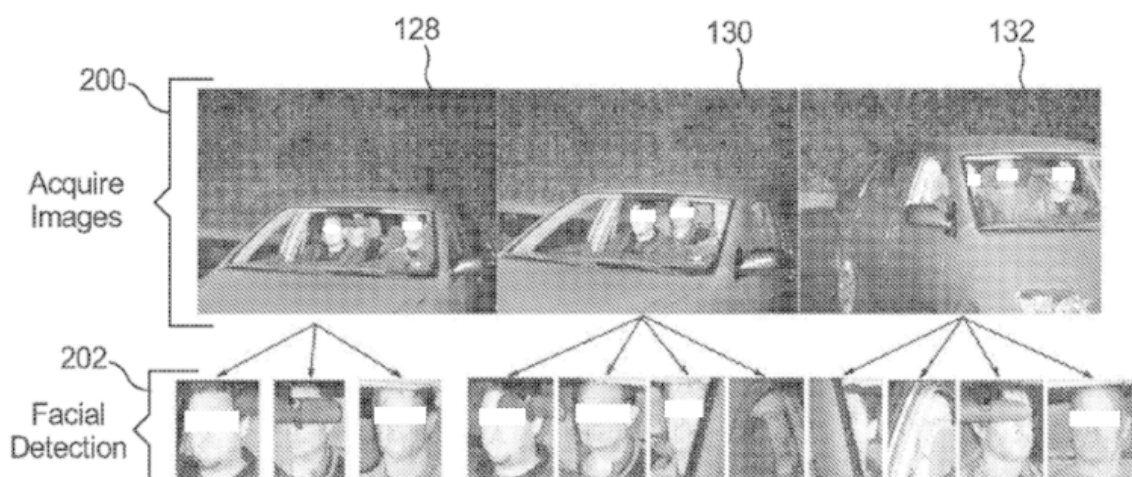
Detecting faces within a vehicle, when the camera is located outside of the vehicle.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition of human or animal bodies, e.g. pedestrians	G06V 40/10
Recognition of eye characteristics within images or video, e.g. of the iris	G06V 40/18
Recognition of movement or behaviour	G06V 40/20
Measuring devices for psychotechnics for vehicle drivers	A61B 5/18
Safety devices for propulsion unit control, specially adapted for, or arranged in, vehicles, responsive to condition of driver	B60K 28/02
Estimation or calculation of driving parameters for road vehicle drive control systems not related to the control of a particular sub-unit, related to drivers or passengers	B60W 40/08
Analysis of motion in images	G06T 7/20
Alarms for indicating a condition of sleep, e.g. anti-dozing alarms	G08B 21/06

G06V 20/60

Type of objects

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Context or environment of the image	G06V 20/50
Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition	G06V 30/00
Recognition of biometric, human-related or animal-related patterns in image or video data	G06V 40/00
Measuring arrangements characterised by the use of optical means	G01B 11/00
Three dimensional [3D] modelling for computer graphics	G06T 17/00
Manipulating 3D models or images for computer graphics	G06T 19/00

G06V 20/62

Text, e.g. of license plates, overlay texts or captions on TV images

Definition statement

This place covers:

Detection and recognition of text or logo regions in scene imagery, e.g. detection and recognition of street names, business logos or names, license plate numbers, or numbers on the clothing of players in a sporting activity.

Localising and recognising text regions on postal items, parcels or containers.

Detection and recognition of overlay text in broadcast video, including embedded captions in TV videos or images.

Illustrative examples of subject matter classified in this place:

1.
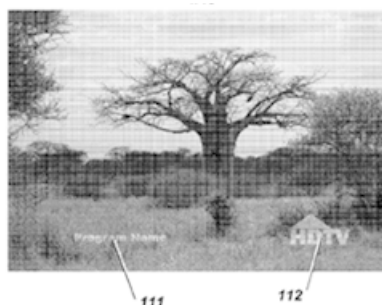
Image wherein the overlay text (detection and recognition of the TV station) is recognised.

2.
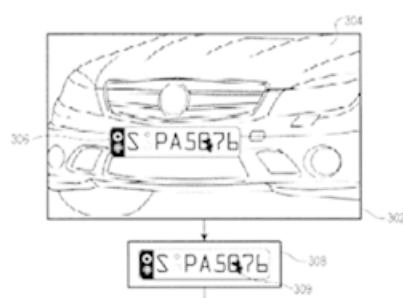
Image wherein the overlay text (the logo License plate) is recognised.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding involving the determination of region of interest [ROI] or a volume of interest [VOI]	G06V 10/25
Character recognition; Recognising digital ink; Document-orientated image-based pattern recognition	G06V 30/00
Image analysis in general	G06T 7/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

overlay text	text elements superimposed over a video stream
regions of interestROI	samples within images or video identified for a particular purpose

Synonyms and Keywords

In patent documents, the following abbreviations are often used:

AOI	area of interest

ROI	region of interest
VOI	volume of interest

G06V 20/64

Three-dimensional objects

Definition statement

This place covers:

Recognition of objects based on their three-dimensional geometric structure ("3D shape"), potentially also exploiting other visual cues such as surface texture, grey-level image values or colours.

Note:

The analysed data is three-dimensional in nature, or the reference/template is three-dimensional. The three-dimensional representation can be very varied: depth/range images, also called 2.5D-images (potentially including texture information), point cloud representations, meshes/tessellations/ wire frames or finite element representations, voxel representations, representations as manifolds (continuous, smooth or Riemannian manifolds; using local charts; as null sets of a certain set of functions, etc.). The majority of the techniques involved recognise the 3D-surface, or part of the 3D-surface ("front side relative to a camera") of the three-dimensional object rather than the interior or its volume.

Illustrative examples of subject matter classified in this place:

1A.
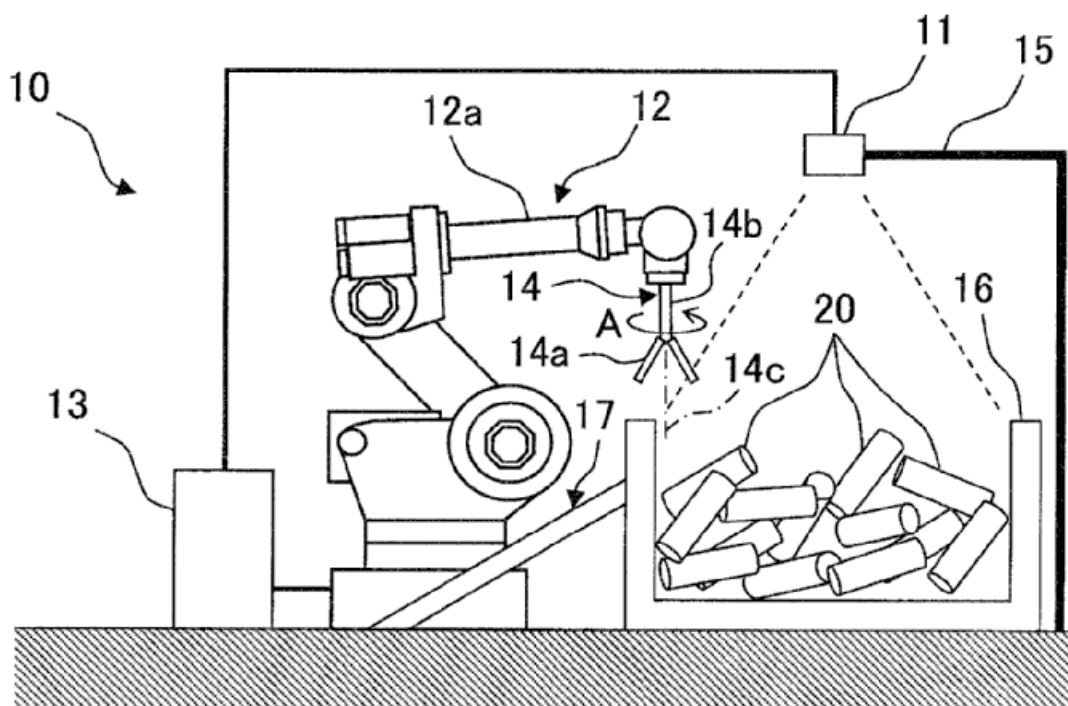
1B.
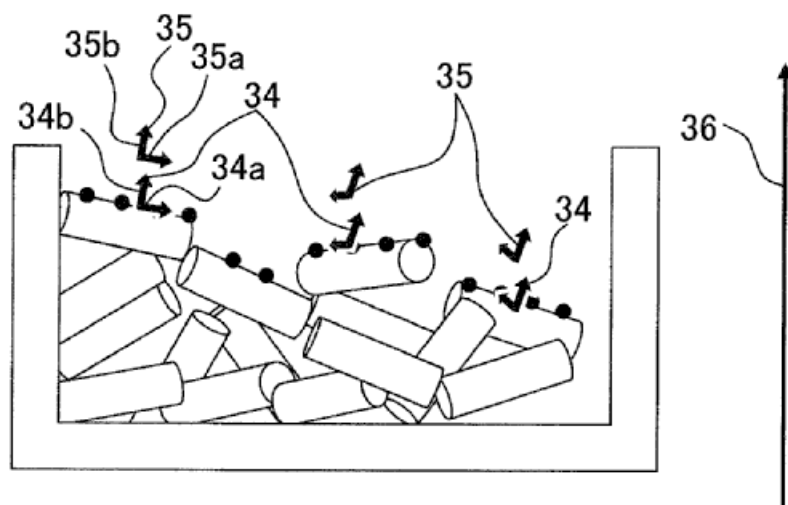
1C.
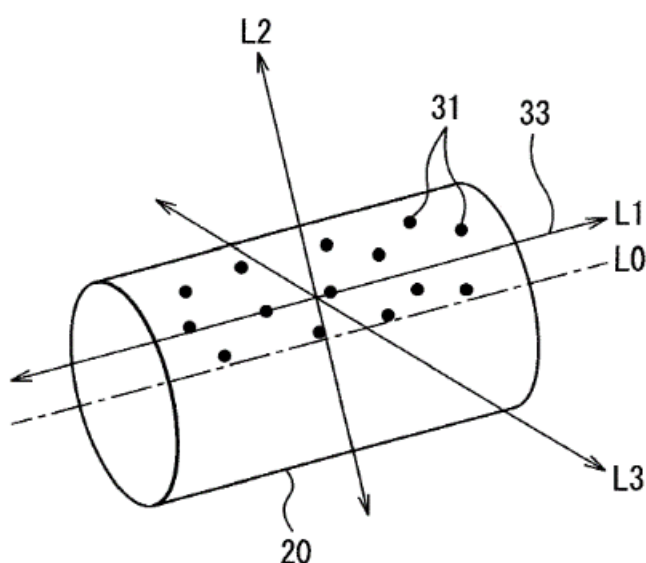
3D object recognition for guiding a robot gripper.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Pattern recognition and image understanding in terrestrial scenes with images taken by planes or drones	G06V 20/17
Surveillance or monitoring of activities, e.g. for recognising suspicious objects	G06V 20/52
Recognition of traffic patterns, e.g. cars on the road, trains or boats	G06V 20/54

Recognition of scenes exterior to a vehicle by using sensors mounted on the vehicle	G06V 20/56
Recognition or understanding of scenes inside of a vehicle	G06V 20/59
Recognition of trinkets, e.g. jewellery items, buttons, gun bullets, medication pills	G06V 20/66
Recognition of food in scenes	G06V 20/68
Recognition of microscopic objects in scenes	G06V 20/69
Recognition of biometric, human-related or animal-related patterns	G06V 40/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Extraction of image or video features using descriptors for shape, contour or point-related descriptors	G06V 10/46
Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition	G06V 30/00
Measuring arrangements characterised by the use of optical means	G01B 11/00
Measuring contours or curvatures by projecting a pattern, e.g. moiré fringes on the object	G01B 11/25
Systems using the reflection or reradiation of electromagnetic waves other than radio waves, e.g. lidar systems	G01S 17/00
Image analysis in general	G06T 7/00
Two dimensional [2D] image generation	G06T 11/00
Three dimensional [3D] image rendering	G06T 15/00
Three dimensional [3D] modelling, e.g. data description of 3D objects	G06T 17/00
Manipulating 3D models or images for computer graphics	G06T 19/00

Special rules of classification

Sometimes special illumination (e.g. that produced by grating patterns) is cast into the scene to gather local 3D shape information. In such cases, classification in groups G06V 10/145 and G06V 20/64 is applied.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

2.5D image	image that simulates the appearance of being three-dimensional when in fact it is 2D
manifold	topological space with the property that each point has a neighbourhood that is homeomorphic to an open subset of n-dimensional Euclidean space
mesh	collection of vertices, edges and faces that defines the shape of a polyhedral object. Also known as a polygon mesh.
tessellation	dividing of data sets of polygons, i.e. vertex sets, presenting objects in a scene into suitable structures for rendering. In real-time rendering, the data is tessellated into triangles, also known as polygon triangulation

topology	properties of a geometric object that are preserved under continuous deformations

G06V 20/66

Trinkets, e.g. shirt buttons or jewellery items (recognising microscopic objects G06V 20/69)

Definition statement

This place covers:

Detection, recognition (e.g. clustering, classification) of personal accessories or small objects of personal use such as:

- Shirt buttons;
- Stamps;
- Gun bullets;
- Jewellery items;
- Coins;
- Drugs, pills, ampoules.

Recognition of keys for door locks.

Recognition of such objects for counting and tracking of these objects.

Illustrative example of subject matter classified in this place:
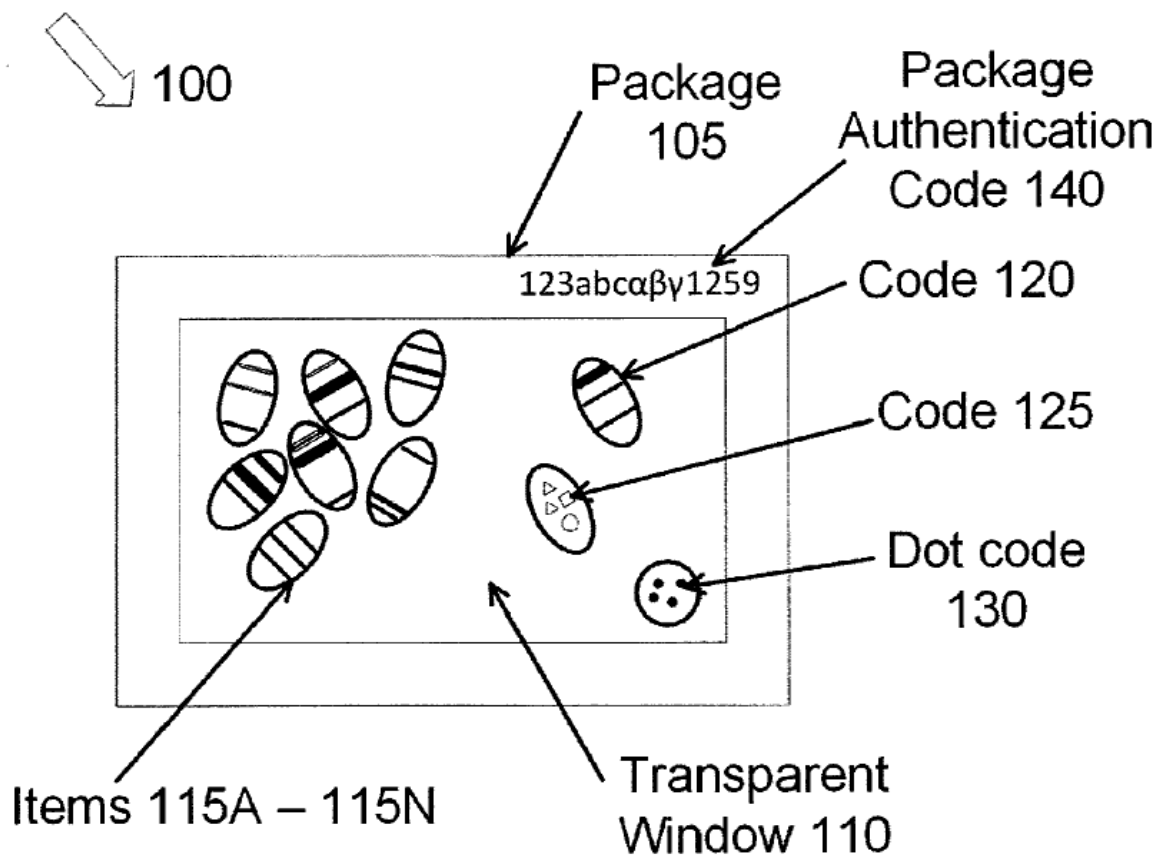
References

Limiting references

This place does not cover:

Recognising microscopic objects in scenes	G06V 20/69

Informative references

Attention is drawn to the following places, which may be of interest for search:

Local feature extraction by analysis of parts of the pattern, e.g. by detecting edges, contours, loops, corners, strokes or intersections; Connectivity analysis, e.g. of connected components, edge linking or neighbouring slice analysis	G06V 10/44
Recognising image objects characterised by unique random patterns	G06V 20/80
Methods or arrangements for sensing record carriers	G06K 7/00
Image analysis in general	G06T 7/00
Testing specially adapted to determine the identity or genuineness of valuable papers or for segregating those which are unacceptable, e.g. banknotes that are alien to a currency	G07D 7/00

G06V 20/68

Food, e.g. fruit or vegetables

Definition statement

This place covers:

Detection, recognition or classification of food items, e.g. on shelves in a supermarket, at the cashier, inside the cart, inside a fridge, inside an oven, etc.

Example applications include determining the freshness of the food, determining the portion size, computing the calories intake based on the recognised ingredients.

Illustrative example of subject matter classified in this place:
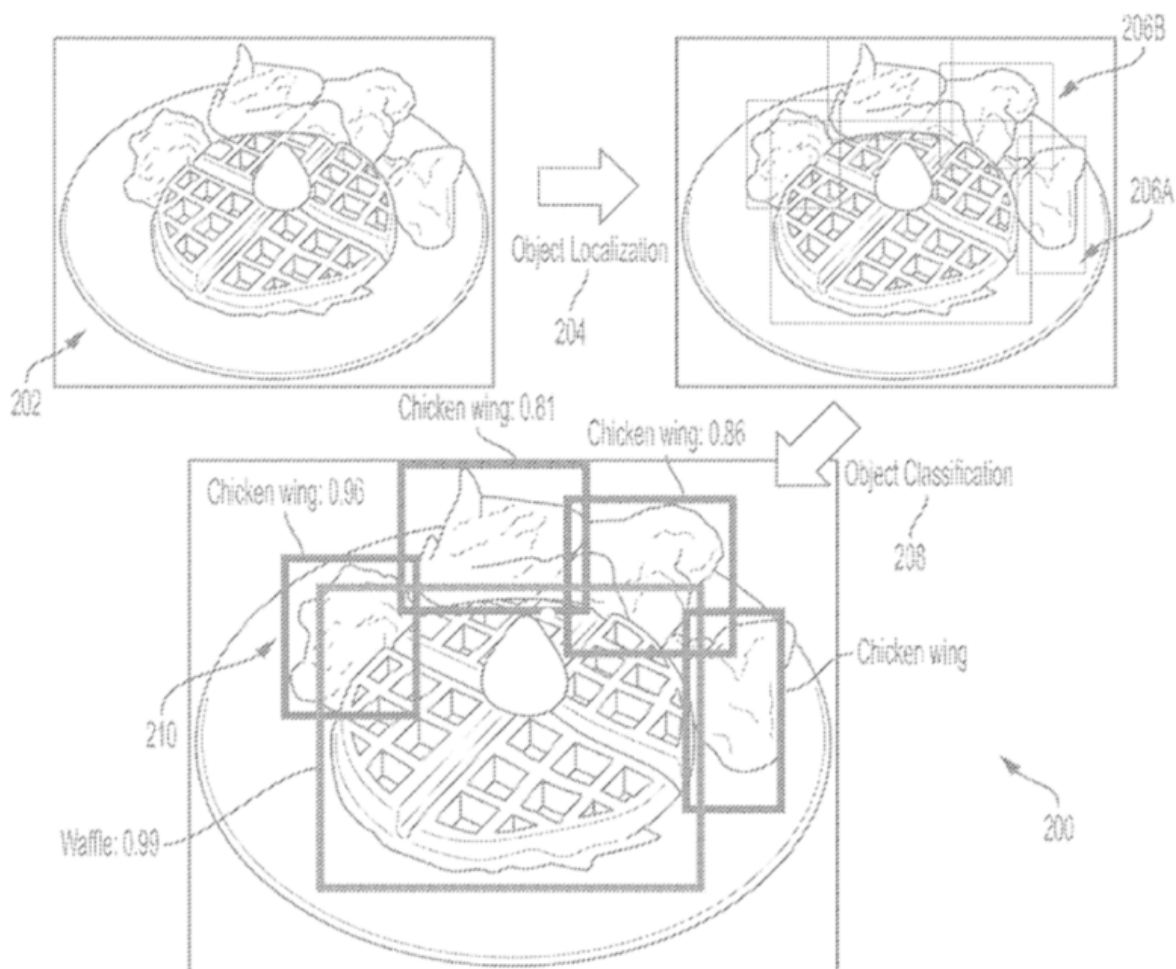
Recognition of the food on the plate using a two-stage process, object localisation followed by object classification.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Object recognition in augmented reality scenes	G06V 20/20
Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Recognising image objects characterised by unique random patterns	G06V 20/80
Foods, foodstuffs	A23L
Hand carts having more than one axis carrying transport wheels; Steering devices therefor; Equipment therefor	B62B 3/00
Payment architectures, schemes or protocols	G06Q 20/00
Commerce, e.g. shopping or e-commerce	G06Q 30/00
Cash registers	G07G 1/00

Special rules of classification

Food recognition is usually performed in an interactive fashion by displaying it on the screen of a mobile phone or in an augmented-reality set-up. In such case, classification in groups G06V 20/20 (augmented reality scenes) and G06V 20/68 is applied.

G06V 20/69

Microscopic objects, e.g. biological cells or cellular parts

Definition statement

This place covers:

Detection, recognition, clustering, or classification of:

- biological cells and cellular parts, e.g., cytoplasm, nucleus, cell membrane, chromosomes, cilia, flagella, etc. of all kinds of cells: prokaryotes, eukaryotes, bacteria, etc.;
- other microscopic biological material such as pollen grains;
- images of virus strains;
- crystals.

Recognition of such objects for counting and tracking.

Detection and classification of certain events (e.g. cellular division, development of an anomaly, detection of replication etc.).

Illustrative examples of subject matter classified in this place:

1A.
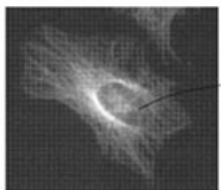
1B
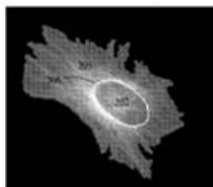
1C.

1D.
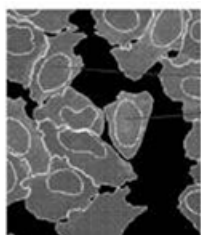
Detection and recognition of cells in microscopic images.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Measuring or testing for enzymology or microbiology with condition measuring or sensing means, e.g. colony counters	C12M 1/34
Measuring or testing processes involving enzymes, nucleic acids or microorganisms; Compositions thereof; Processes of preparing such compositions	C12Q 1/00
Spectrometry; Spectrophotometry; Monochromators; Measuring colours	G01J 3/00
Investigating characteristics or properties of individual particles using electro-optical means	G01N 15/14
Investigating or analysing materials by the use of optical means by use of fluorescence or phosphorescence	G01N 21/64
Investigating or analysing materials by specific methods not covered by groups G01N 1/00-G01N 31/00; Analysis of biological material, e.g. blood, urine	G01N 33/48
Microscopes	G02B 21/00
Image analysis in general	G06T 7/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

hyperspectral image	multi-band image where the z dimension corresponds to consecutive spectral wavelengths ranges
multispectral image	multi-band image where the z dimension corresponds to spectral wavelengths ranges (not necessarily consecutive)

G06V 20/70

Labelling scene content, e.g. deriving syntactic or semantic representations

Definition statement

This place covers:

Automatic annotation or labelling of scenes.

Semantic segmentation of scenes, e.g. by means of labelling each pixel of an image with a corresponding class of what is being represented. This process can be seen as image classification at pixel level.

Syntactic segmentation of scenes, e.g. by means of using the structural representation of an image. Examples of structural representations include grammars or graphs. This process can be used instead of statistical pattern recognition when there is a clear structure in the pattern.

Illustrative examples of subject matter classified in this place:

1.
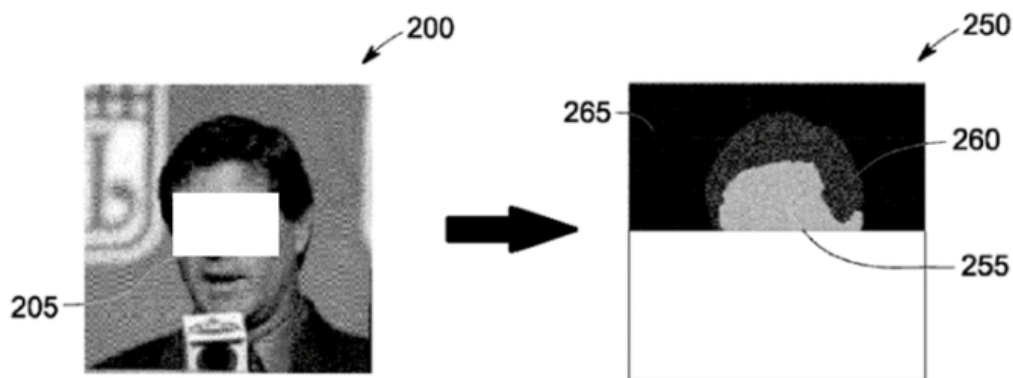
Semantic segmentation of hair wherein a tiered structure constraint has been used for determining the labels of the pixels.

2.
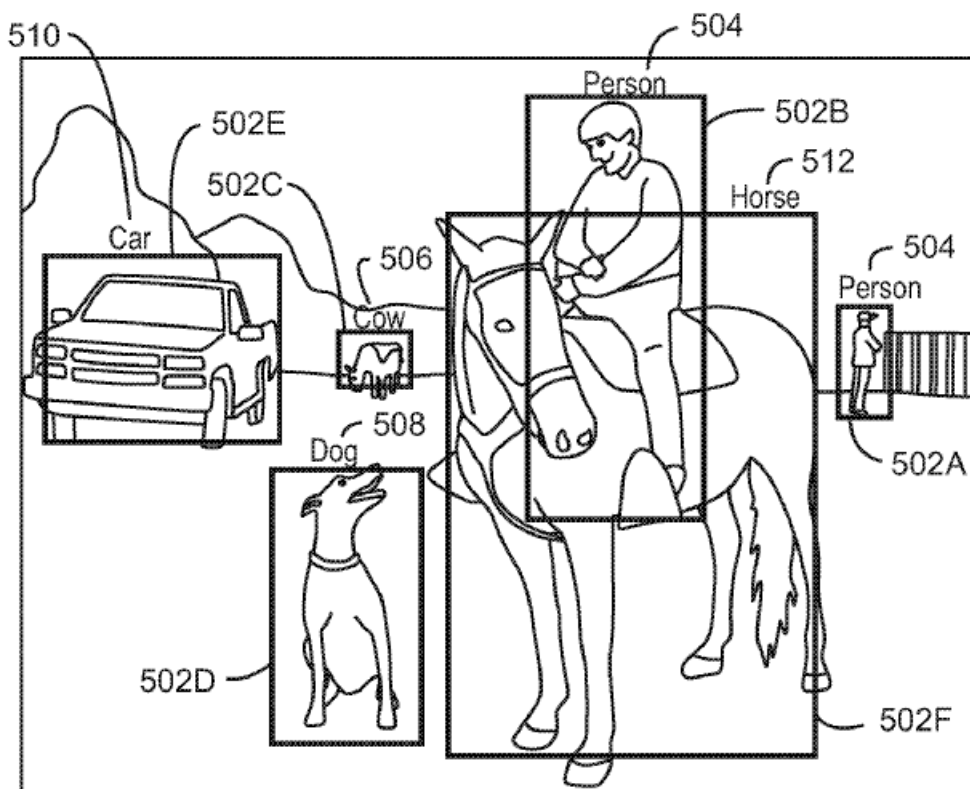
Labelling of image objects according to known object classes.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding involving the determination of region of interest [ROI] or a volume of interest [VOI]	G06V 10/25
Segmentation of patterns in the image field; Cutting or merging image elements to establish the pattern region, e.g. region growing, watershed or clustering-based techniques; Detection of occlusion	G06V 10/26
Recognition using syntactic or structural representations of the image or video pattern, e.g. symbolic string recognition; Graph matching	G06V 10/86
Techniques for post-processing in character recognition using context analysis, e.g. lexical, syntactic or semantic context	G06V 30/262
Information retrieval; Database structures therefor; File system structures therefor	G06F 16/00
Image analysis by segmentation or edge detection in general	G06T 7/10

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

dense prediction	labelling each pixel of an image or video with a corresponding class of what is being represented.
semantic image segmentation	labelling regions (e.g. set of pixels) of an image with a corresponding object class of what is being represented.
syntactic pattern recognition	a form of pattern recognition, in which each object can be represented by a variable-cardinality set of symbolic, nominal features. This allows for representing pattern structures, taking into account more complex interrelationships between attributes than is possible in the case of flat, numerical feature vectors of fixed dimensionality, that are used in statistical classification.

G06V 20/80

Recognising image objects characterised by unique random patterns

Definition statement

This place covers:

Authentication of objects or products by physically unclonable function [PUF].

Identification of counterfeit goods by PUF.

Identification by micro-random structures naturally occurring on the surface of an object.

Identification by applying specially designed micro-structures to the surface of an object, e.g. quantum dots or nano-barcodes or ink containing magnetic particles.

Encoding the extracted PUF and digitally storing the code for retrieval or printing the code on surface of the object for authentication.

Recognition of PUF which change their appearance depending on the incident angle of the illumination.

Recognition of PUF by dedicated or general purpose devices; mostly microscopes; they can be fixed, for example in an industrial context, or mobile (e.g. microscope attached to mobile phone or mobile phone with very large zoom), for example for identifying counterfeit goods by user.

Examples of objects and products that may be authenticated by this technique include:
- pharmaceutical and cosmetics products;
- individual pills or packaged substances;
- electronics;
- luxury goods, e.g. watches;
- text documents and certificates;
- weapons;
- agricultural products, e.g. fruits;
- recipients for bio-medical probes.

Illustrative example of subject matter classified in this place:
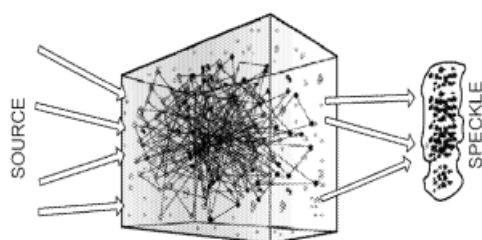
Analysis of the random patterns in a material by casting light and encoding the resulting speckle pattern using PUF.

Relationships with other classification places

While group G06V 20/80 aims at recognising objects depicted in the image from their random patterns, group G06V 20/90 assumes that the image is analysed without necessarily identifying the image objects. The purpose of the analysis of group G06V 20/90 is to assess, based on image imperfections generated by the sensor, whether the image has been captured by the same sensor or not.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Investigating or analysing materials by the use of optical means	G01N 21/00
Commerce, e.g. shopping or e-commerce	G06Q 30/00
General purpose image data processing	G06T 1/00
Testing specially adapted to determine the identity or genuineness of valuable papers or for segregating those which are unacceptable, e.g. banknotes that are alien to a currency	G07D 7/00

Special rules of classification

Highlighting the non-uniformities in the objects subject to analysis usually involves casting special light (e.g. having a certain spectral content which matches those of the non-uniformities) or using special sensors, which is classified in group G06V 10/10 and its subgroups. In such case, classification in groups G06V 10/10 and G06V 20/80 is applied.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

PUFdigital fingerprintsecurity markphysical dispersion patternphysical scatter pattern	physical unclonable function
physically unclonable features	unique features on the surface of objects, products or documents which uniquely identify the object, in a manner similar to how fingerprints uniquely identify a person; these unique features may be naturally occurring or purposefully added random microstructures on the physical object surface.

G06V 20/90

Identifying an image sensor based on its output data

Definition statement

This place covers:

Identifying an image sensor based on characteristic sensor noise patterns, sensor imperfections, artifacts, or optical defects. Defective pixels which, individually, are normally not perceptible to the human eye may be detected, and the repeatability of their occurrence at the same spatial position may be used for sensor/camera identification.

Notes – technical background

These notes provide more information about the technical subject matter that is classified in this place:

The process of digital camera identification may involve three steps:

1. Photo response non-uniformity (PRNU) noise extraction. The PRNU-pattern from the image under investigation is extracted using a denoising filter;

2. Extraction of sensor pattern noise (SPN), also known as the camera fingerprint, is obtained by taking a series of flat-field images with the camera under investigation.

From each image the PRNU-pattern is extracted and then these patterns are combined to estimate the SPN;

3. Comparison. The SPN-pattern of the camera and the PRNU-pattern of the image are compared by calculating for example a correlation metric.

Illustrative example of subject matter classified in this place:
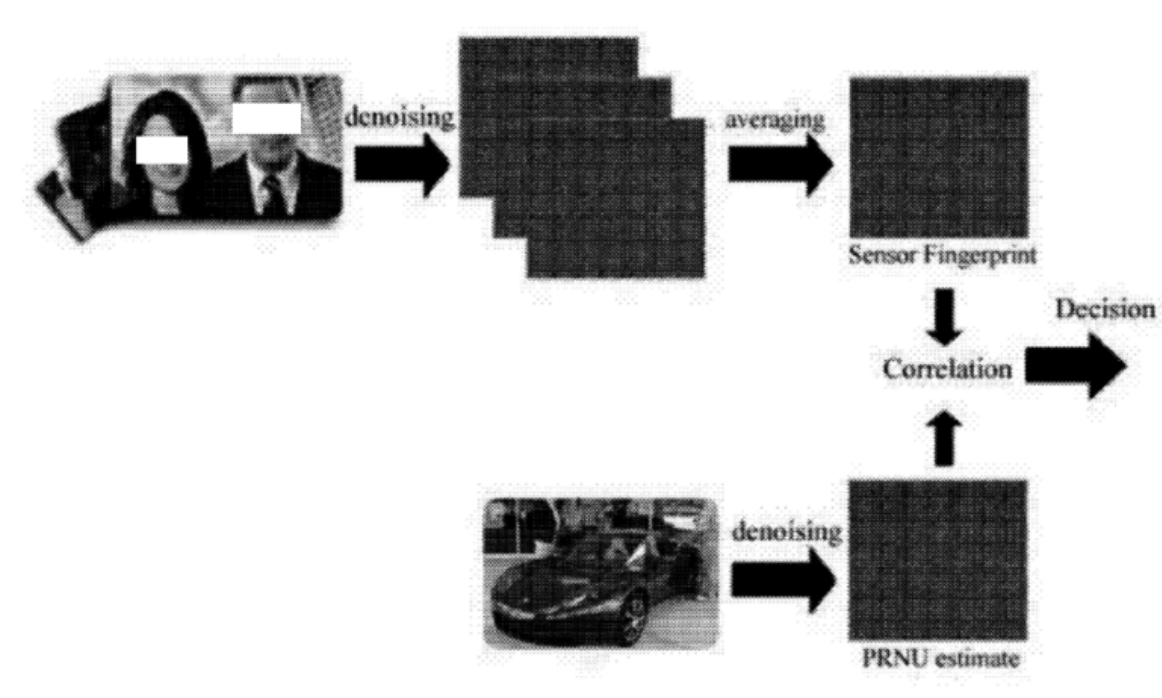
Determining whether the two images are taken by the same camera, by implementing the basic three-step process described above.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image enhancement or restoration	G06T 5/00
Speaker identification or verification	G10L 17/00
Details of television systems	H04N 5/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

fixed pattern noise [FPN]	additive noise caused by dark currents when the sensor array is not exposed to light
hardwaremetry	process of searching for characteristic features for identifying an image sensor
photo response non-uniformity [PRNU]	major source of noise caused when pixels have different light sensitivities caused by the inhomogeneity of silicon wafers
SNR	signal-to-noise ratio

G06V 30/00

Character recognition; Recognising digital ink; Document-oriented image-based pattern recognition (scanning, transmission or reproduction of documents or the like H04N 1/00)

Definition statement

This place covers:

Acquisition, preprocessing, segmentation, feature extraction and recognition of characters that are represented as an image:

- optical character recognition [OCR] if the text to be recognised consists of machine printed characters;
- offline handwriting symbol and character recognition for different alphabets (e.g. Latin, Kanji, Hiragana, Katakana, etc.).

Preprocessing, segmentation, feature extraction and recognition of digital ink (i.e. online handwritten character recognition), where the characters are represented as temporal sequences of handwritten position coordinates, in the form of order-dependent strokes.

The analysis may rely on order-independent strokes where point coordinates are represented without temporal information (i.e. offline handwritten character recognition).

The above representations include representations in three dimensions, e.g. as written by performing gestures in the air.

Document analysis, recognition and understanding, where the document is represented as an image. Possible application scenarios are business forms, standard forms, graphical technical drawings, geographical maps, parcels, letters, credit cards, cheques, etc.

References

Limiting references

This place does not cover:

Scanning, transmission or reproduction of documents or the like	H04N 1/00

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Sorting of mail or documents using means for detecting the destination	B07C 3/10
Input arrangements for interaction between user and computer	G06F 3/01
Testing specially adapted to determine the identity or genuineness of valuable papers or for segregating those which are unacceptable	G07D 7/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Optical elements, systems or apparatus	G02B
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements	G06F 3/00
Handling natural language data	G06F 40/00
Image contour coding, e.g. using detection of edges	G06T 9/20

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

OCR	optical character recognition, recognising a machine printed symbol or character based on an image
offline handwriting recognition	recognising a handwritten symbol or character based on an image, i.e. without any temporal information. The difference with respect to OCR is that the symbols or characters are handwritten.
online handwriting recognition	recognising a handwritten symbol or character based on time-series of handwritten coordinates, i.e. with temporal information
strokes	basic components of characters that are either separated spatially and/or temporally, e.g. contiguous segments left by a writing instrument during handwriting

Synonyms and Keywords

In patent documents, the following words/expressions are often used with the meaning indicated:

stroke order independent	analysis where the temporal order of the strokes is not relevant (i.e. offline handwriting recognition)
stroke order dependent	analysis where the temporal order of the strokes is relevant (i.e. online handwriting recognition)

G06V 30/10

Character recognition

Definition statement

This place covers:

Recognition of characters, wherein the characters have been generated by machine or by handwriting.

Recognition of characters, wherein the characters have been generated by machine or by handwriting.

Image acquisition specially adapted for OCR and/or recognition of handwritten text using handheld instruments (e.g. with touch screens).

Instruments generating sequences of position coordinates corresponding to handwriting.

OCR of symbols and characters for any language.

Stroke segmentation and recognition of whole cursive handwritten words, i.e. whose letters are not separated but are linked together, whether from offline (image representation) or from online (digital ink, e.g. pen input) acquisition:

Preprocessing, feature extraction, matching, recognition and classification of all kinds of handwritten characters, symbols, drawings, except signatures, on the basis of trajectories as a function of time of a stylus, finger, etc. Trajectories can be acquired by a touch pad/screen or by a stylus like device (in collaboration with a passive or active surface).

Segmentation of strokes, characters or words.

Text and character recognition using temporal information, e.g. free-form handwriting.

Recognition of drawings using temporal information, e.g. sketches, flow charts, graphical or mathematical symbols or formulae, chemical structure formulae, editorial notes, proof marks.

Illustrative examples of subject matter classified in this place:

1.
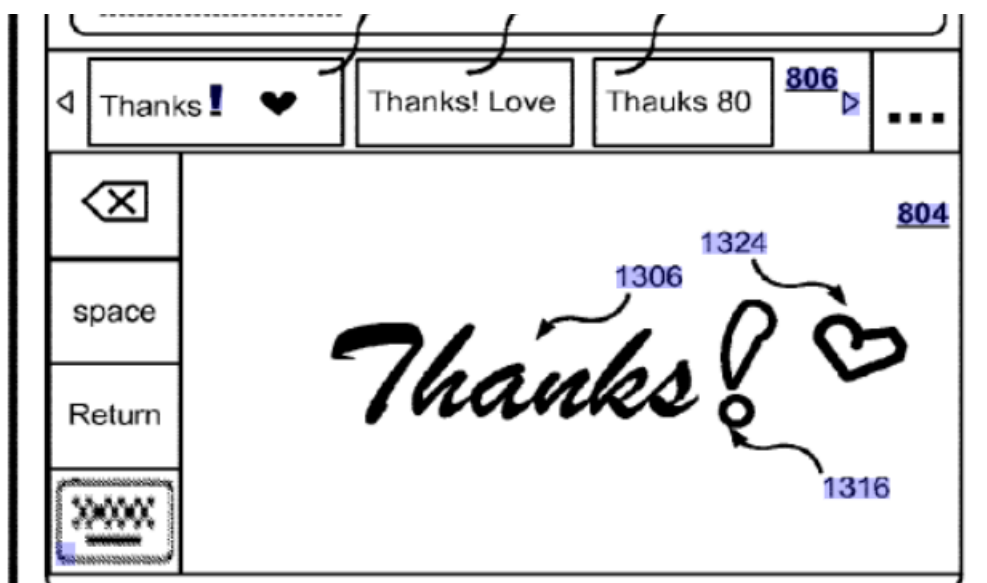
Recognition of handwritten English text input via a touch screen.

2.
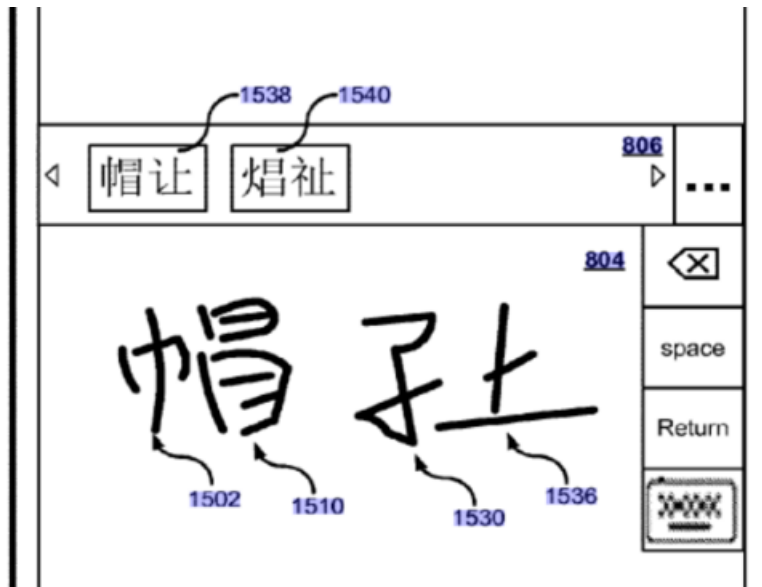
Recognition of handwritten Chinese text input via a touch screen.

3.
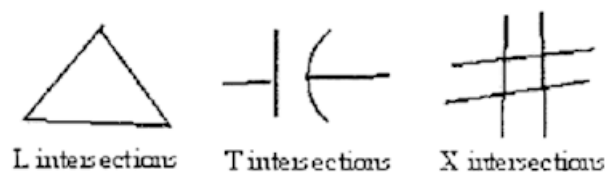
Recognition of different symbols.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing in arrangements for image or video recognition or understanding	G06V 10/20
Noise filtering in arrangements for image or video recognition or understanding	G06V 10/30
Image acquisition of characters, digital ink or documents using hand-held devices for recognition purposes	G06V 30/142
Context analysis as post-processing after provisional recognition	G06V 30/262
Writer recognition; Reading or verifying signatures	G06V 40/30
Arrangements for converting the position or the displacement of a member into a coded form, for input arrangements or input/output arrangements for user-computer interaction	G06F 3/03
Inputting data by handwriting, e.g. gestures or text, to a computer via a graphical user interface [GUI] using a touch-screen or digitiser	G06F 3/0488

User authentication by graphic or iconic representation, in security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/36

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

digital inkelectronic inke-ink	technology that digitally represents handwriting, e.g. using a finger or a stylus, in its natural form using temporal information. Digital ink may also be referred to as electronic ink or e-ink.

G06V 30/12

Detection or correction of errors, e.g. by rescanning the pattern

Definition statement

This place covers:

Processes and devices for detecting or correcting character recognition errors after image input. The errors can be detected or corrected automatically (e.g. by a computer program), with the help of an operator, or as a semi-automatic process (e.g. by displaying an image on a graphical user interface and requesting human intervention).

Detecting errors in particular comprises evaluating the quality of given image or video data with regard to the suitability for being subjected to an automated character recognition process. Typical quality criteria include image sharpness/blurriness, resolution, contrast or brightness.

Monitoring print quality by performing character recognition on the prints, wherein quality relates to a property of the characters or of the digital ink insofar as its effect on the accuracy or performance of the character recognition process. Quality within the meaning of this group is a property of the characters or of the digital ink insofar as it has an effect on the accuracy or performance of the pattern recognition process.

Illustrative examples of subject matter classified in this place:

1.
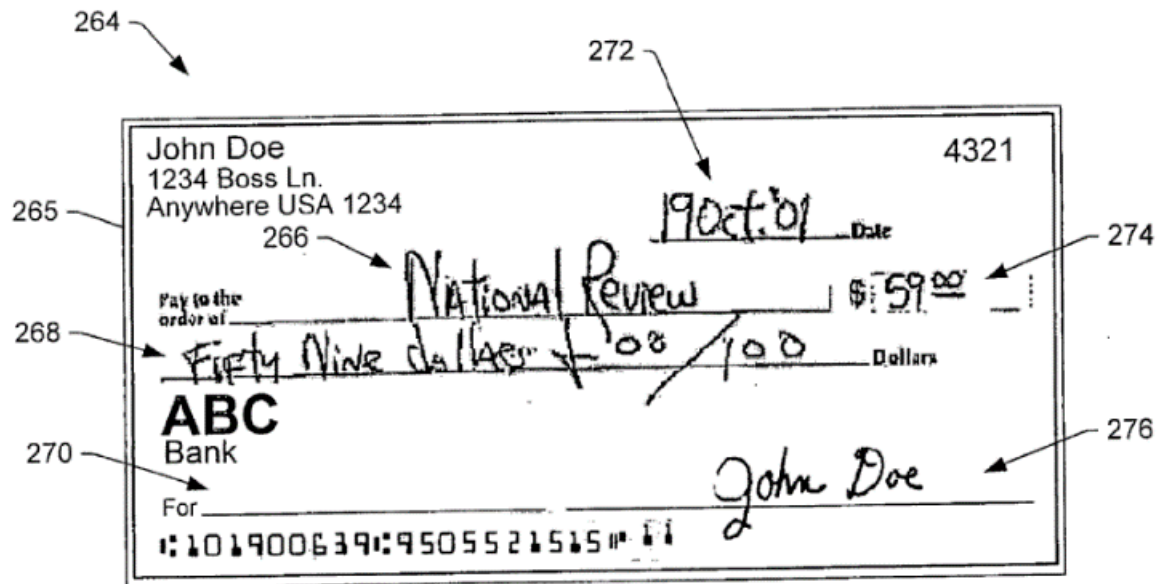
Evaluation of the quality of recognition of the different fields of a bank cheque.

2.
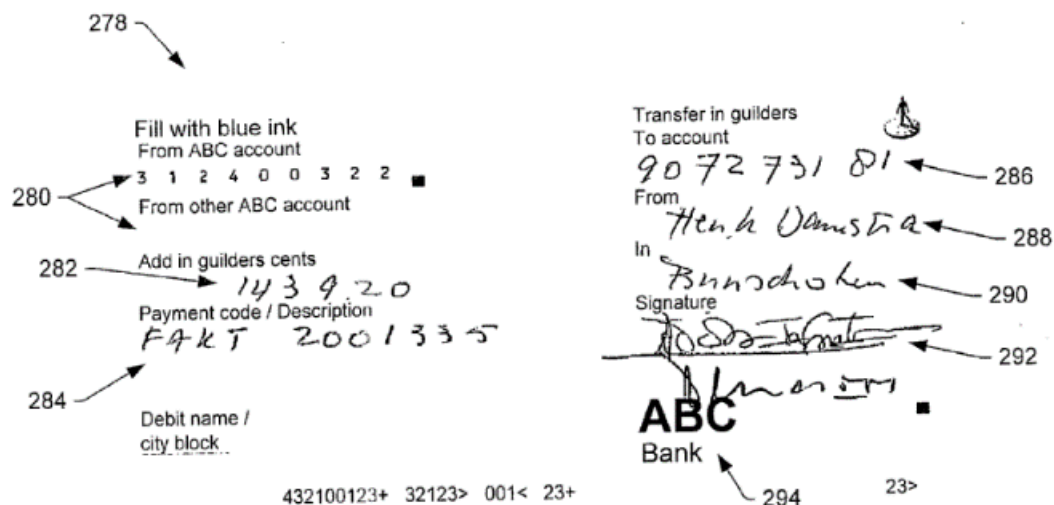
Evaluation of the quality of recognition of the different fields of a bank cheque.

Relationships with other classification places

Methods or arrangements for detection or correction of errors covered by this group involve such correction after the acquisition step with the aim of having a reliable input before the subsequent recognition. In contrast, group G06V 30/26 covers the methods and arrangements used after recognition with the aim of correcting the final output, by using additional information such as context.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Scenes; Scene-specific elements	G06V 20/00
Recognition of biometric, human-related or animal-related patterns in image or video data	G06V 40/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Aligning, centring, orientation detection or correction of the image	G06V 10/24
Validation or performance evaluation for pattern recognition or understanding in images or video, in general	G06V 10/776
Detection or correction of errors in image or video recognition or understanding, in general	G06V 10/98
Techniques for post-processing	G06V 30/26
Investigating the presence of flaws, defects or contamination in materials by the use of optical means	G01N 21/88
Investigating the presence of flaws in materials by the use of thermal means	G01N 25/72
Image analysis	G06T 7/00
Arrangements for detecting or preventing errors in the digital information received via transmission	H04L 1/00
Scanning, transmission or reproduction of documents or the like	H04N 1/00
Details of television systems	H04N 5/00

Special rules of classification

Group G06V 30/12 may be regarded as relevant to subject matter also classified in other subgroups of group G06V 30/00 and so the principles of multiple classification apply.

G06V 30/14

Image acquisition

Definition statement

This place covers:

Image acquisition specially adapted for character recognition.

Image acquisition for character recognition using handheld instruments (e.g. with touch screens).

Image acquisition for character recognition using handheld instruments generating sequences of position coordinates corresponding to handwriting.

Image acquisition for character recognition using a slot moved over the image, discrete sensing elements at predetermined points or automatic curve-following means.

Image acquisition for character recognition using alignment or centring of the image pick-up or image-field, e.g. skew correction.

Image acquisition for character recognition using segmentation of character regions.

Illustrative examples of subject matter classified in this place:

1.
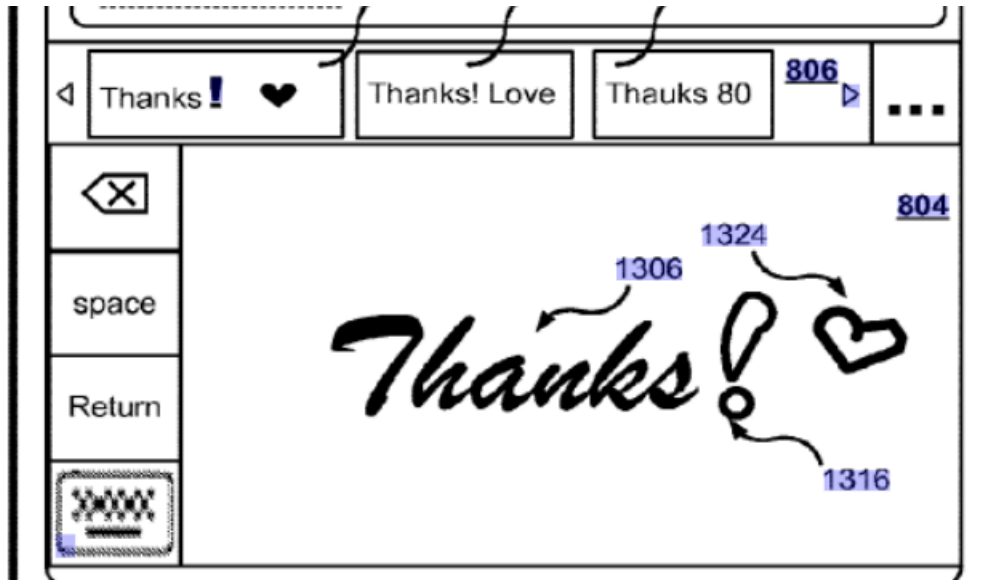
Acquisition of handwritten text input via a touch screen written with a finger.

2.

Acquisition of handwritten text input via a touch screen written with a digital pen/stylus.

3.
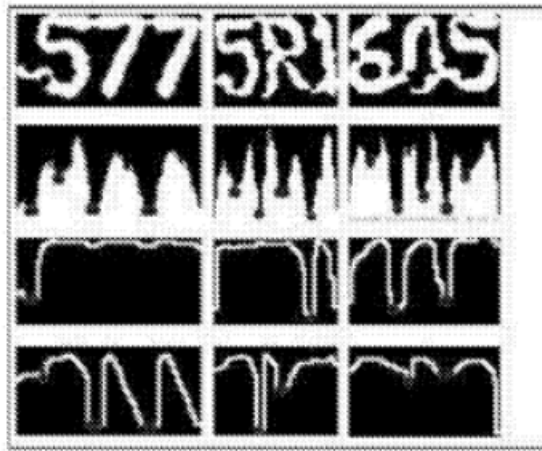
Segmentation of characters based on projection profiles.

4.

Inclination detection and correction before recognition, original slanted text (left), text after correction (right).

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Arrangements for converting the position or the displacement of a member into a coded form, for input arrangements or input/output arrangements for user-computer interaction	G06F 3/03
Inputting data by handwriting, e.g. gestures or text, to a computer via a graphical user interface [GUI] using a touch-screen or digitiser	G06F 3/0488
User authentication by graphic or iconic representation, in security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/36

G06V 30/16

Image preprocessing

Definition statement

This place covers:

Image preprocessing specially adapted for character recognition, in particular:
- Quantising the image signal;
- Noise filtering;
- Normalisation of pattern dimensions;
- Smoothing or thinning of the pattern; skeletonisation.

Illustrative examples of subject matter classified in this place:

1A-1D.
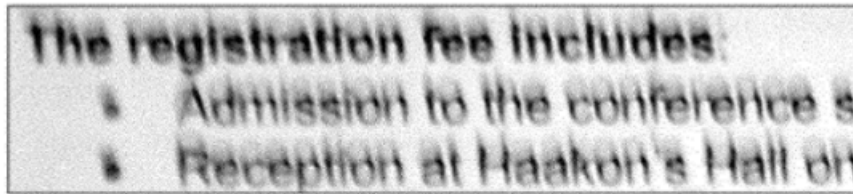
FIG. 1A
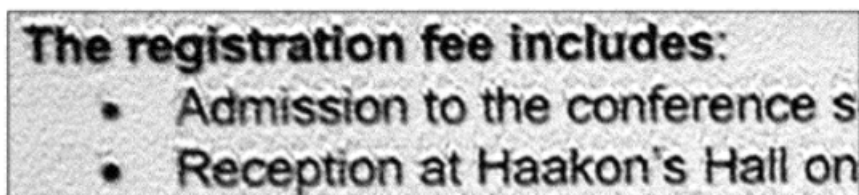
FIG. 1B
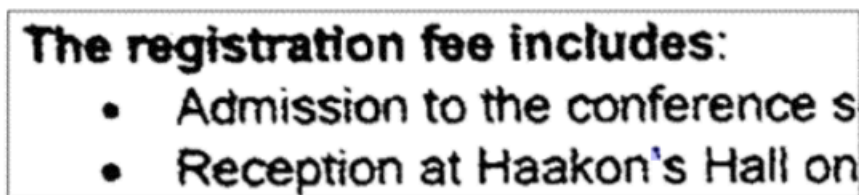
FIG. 1C
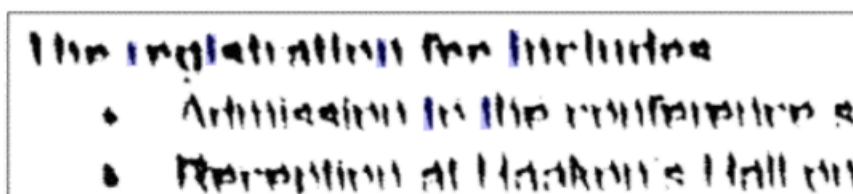
FIG. 1D

FIG. 1A shows a fragment of an example image with blur in its original state.

FIG. 1B shows the fragment shown in FIG. 1A in a restored state after applying a method for restoring blurred images.

FIG. 1C shows the fragment shown in FIG. 1B after being binarised.

FIG. 1D shows the fragment shown in FIG. 1A after being binarised without applying a method as described herein.

2.
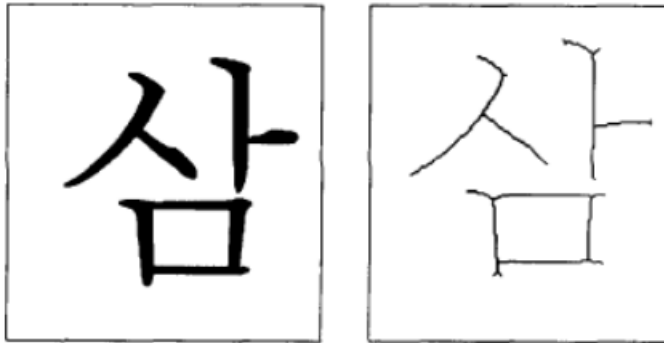
Left: original character; right: skeletonised version

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing in image or video recognition or understanding	G06V 10/20
Noise filtering in image or video recognition or understanding	G06V 10/30
Image enhancement or restoration, in general	G06T 5/00

G06V 30/18

Extraction of features or characteristics of the image

Definition statement

This place covers:

Extraction of features or characteristics of the image specifically adapted for character recognition:
- by coding the contour of the character pattern, e.g. contour-related features;
- by analysing segments intersecting the character pattern, e.g. the segments being obtained with lines, circles drawn on the pattern;
- by deriving mathematical or geometrical properties from the whole character pattern or image, e.g. centre of mass, moments of inertia, etc.

Illustrative examples of subject matter classified in this place:

1.

(Stroke)   (Polygonal Approximation)
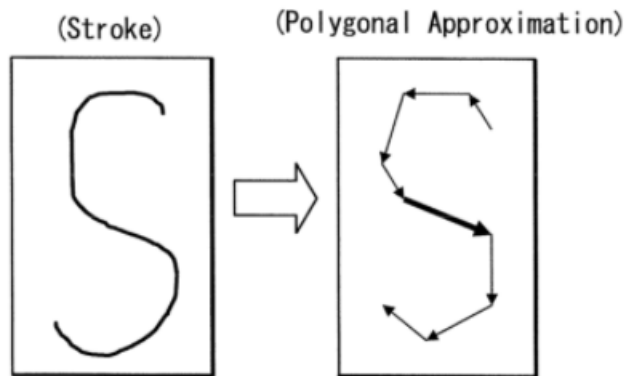
An explanatory drawing showing an example of each stroke and polygonal approximation.

2.

Cumulative Angle Characteristic Function (1)

Length for Predetermined Cumulative Angle/Winding Angle

< Linear Interpolation Function by Exterior Angle/ Δ Sequence and Length Sequence >

0.448

0   0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8   0.9   1

Obtain Length For Predetermined Cumulative Angle/Winding Angle

Length of 90° From Start Portion (a)
(90° Length of Start Portion):0.448

Length of 90° From End Portion (b)
(90° Length of End Portion):0.448

Min [(a),(b)]Max[(a),(b)]
(90 Degree Length Ratio):1.0

(From Start Portion: Forward Direction)

Length Sequence	0.16	0.15	0.08	0.11	0.08	0.11	0.14	0.17
Cumulative Length Sequence	0.16	0.31	0.39	0.50	0.58	0.69	0.83	1.00
Exterior Angle/ Δ Sequence	14.73	26.94	29.20	36.03	32.15	29.71	15.51	
Cumulative Angle/ Winding Angle	14.73	41.67	70.87	106.90	139.05	168.76	184.27	

(From End Portion: Backward Direction)

Length Sequence	0.17	0.14	0.11	0.08	0.11	0.08	0.15	0.16
Cumulative Length Sequence	0.17	0.31	0.42	0.50	0.61	0.69	0.84	1.00
Exterior Angle/ Δ Sequence	15.51	29.71	32.15	36.03	29.20	26.94	14.73	
Cumulative Angle/ Winding Angle	15.51	45.22	77.37	113.40	142.60	169.54	184.27	

Example of the features that are extracted from a "U" handwritten sign based on the cumulative angle feature function.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image analysis in general	G06T 7/00

G06V 30/19

Recognition using electronic means

Definition statement

This place covers:

Recognition of symbols and characters using electronic means specifically adapted to character recognition:

- using simultaneous comparisons or correlations of the image signals with a plurality of references, including references that are adjustable by an adaptive method, e.g. learning;
- using sequential comparisons or correlations of the image signals with a plurality of references, wherein at any stage the selection of a reference depends on the result of the preceding comparison.

Illustrative examples of subject matter classified in this place:

1.

Handwritten pattern recognition based on comparison with respect to reference stroke data.

2.
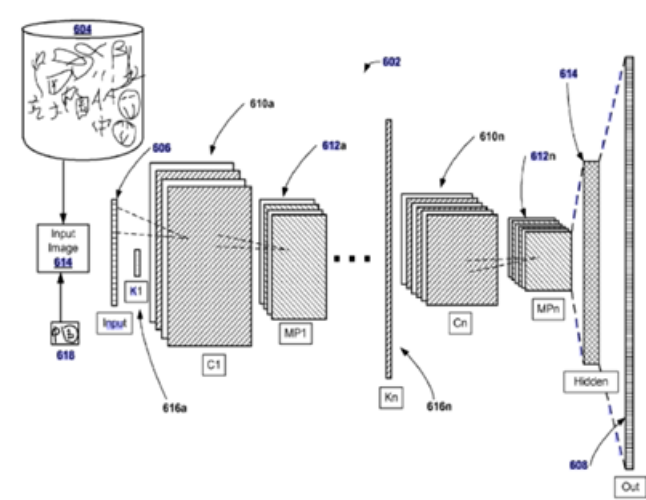
A convolutional neural network for the recognition of handwritten symbols and characters.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Text, e.g. of license plates, overlay texts or captions on TV images	G06V 20/62
Arrangements for character recognition using optical reference masks, e.g. holographic masks	G06V 30/199

G06V 30/22

characterised by the type of writing

Definition statement

This place covers:

Character recognition characterised by the type of writing, including:
- recognition of characters separated by spaces, i.e. non-connected characters;
- recognition of printed characters having additional code marks or containing code marks, e.g. the character being composed of individual strokes of different shape, each representing a different code value or having associated magnetic codes;
- recognition of whole cursive handwritten words, i.e. whose letters are not separated but are linked together, whether from offline (scanning) or from online (digital ink, e.g. pen input) acquisition;
- recognition of three-dimensional handwriting, e.g. writing in the air.

Illustrative examples of subject matter classified in this place:

1.
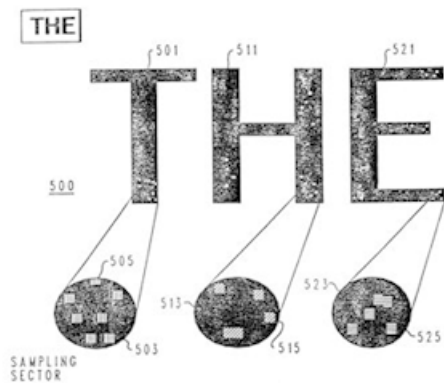
Colour codes embedded in characters to assist their recognition.

2.
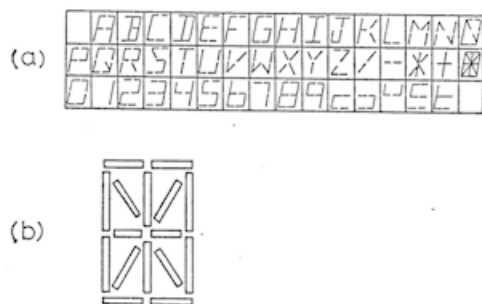
Letters and numbers composed of a combination of sixteen segments each.

3.

Characters composed of vertical bars, the shape of the bars assisting the optical character recognition.

4A.
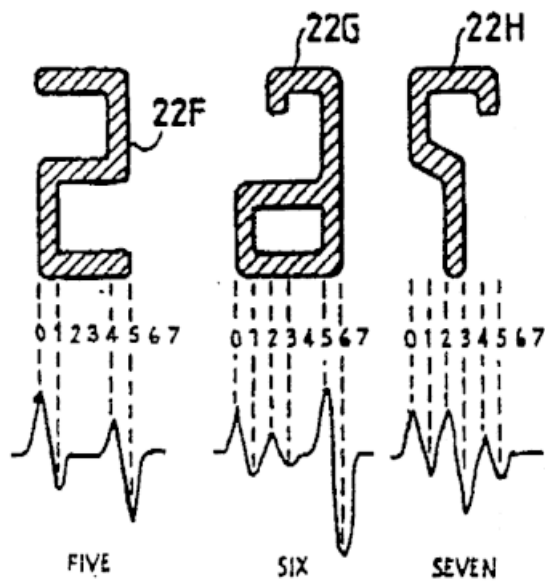
4B.
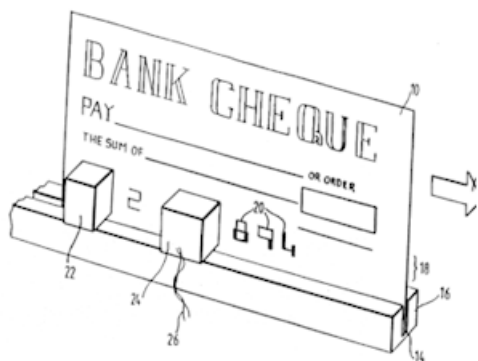
The characters (22F, 22G and 22H) have a different stroke width/length, which results in a characteristic waveform when scanned from right to left by a magnetic reader (22), an application is frequently used for bank cheques.

5.
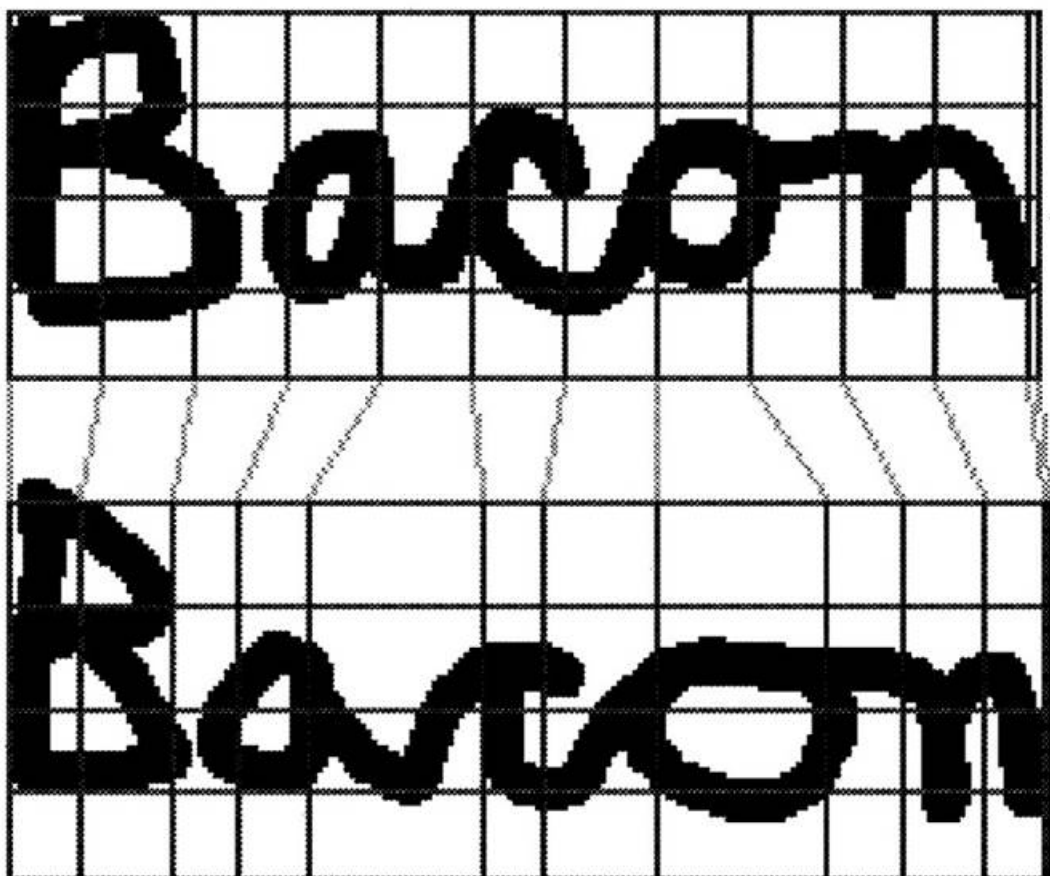
Recognition of cursive words by fitting the characters on a deformable grid.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image acquisition for character recognition using a slot moved over the image, discrete sensing elements at predetermined points or automatic curve following means	G06V 30/144
Methods and arrangements for sensing record carriers	G06K 7/00
Testing specially adapted to determine the identity or genuineness of valuable papers, e.g. banknotes	G07D 7/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

CMC-7	special font used for printing characters for magnetic and optical character recognition systems
magnetic ink	ink containing particles of magnetic material used for printing characters to facilitate magnetic character recognition
MICR	magnetic ink character recognition

G06V 30/24

characterised by the processing or recognition method (segmentation of character regions G06V 30/148)

Definition statement

This place covers:

Character recognition characterised by the processing or recognition method, including:
- Division of character sequences into groups prior to recognition; selection of dictionaries;
- Using graphical properties, e.g. alphabet type, font or type of print when performing recognition;
- Alphabet recognition;
- Font recognition;
- Discrimination between machine-print, hand-print or cursive writing;
- Analysis of linguistic properties, e.g. English or German.

Illustrative examples of subject matter classified in this place:

1A.
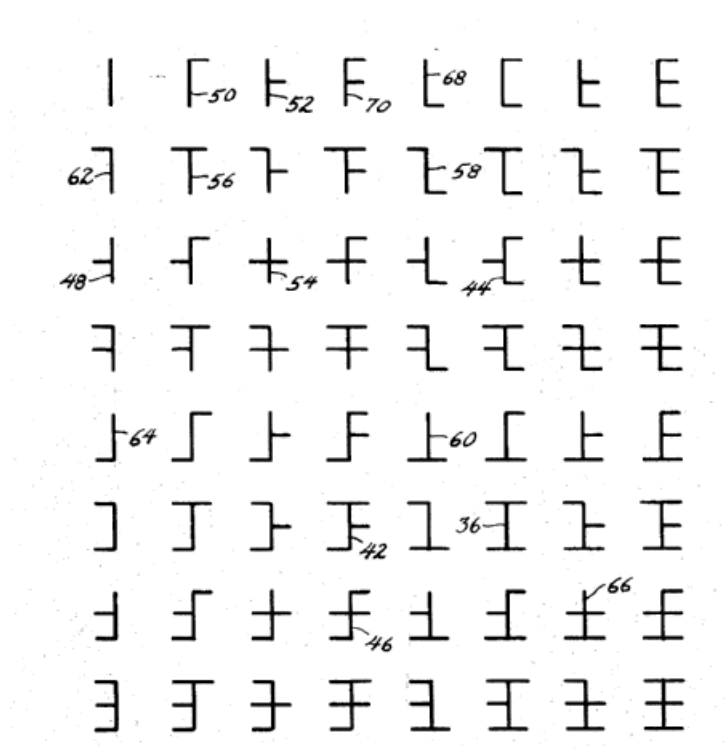
1B.
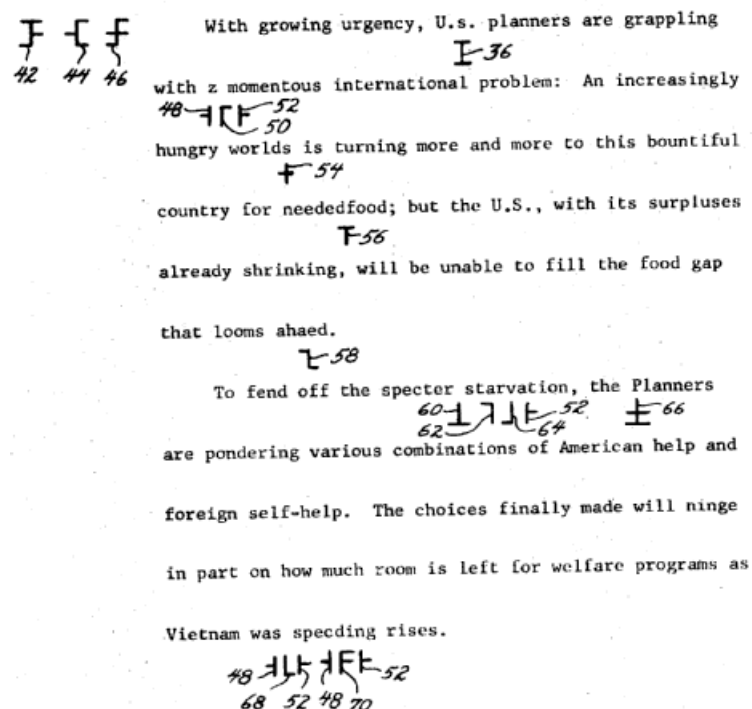
Example of correction symbols forming an alphabet (fig. 1A), each symbol having a predefined meaning which allows the text to be automatically processed (fig. 1B).

References

Limiting references

This place does not cover:

Segmentation of character regions, for character recognition	G06V 30/148

Informative references

Attention is drawn to the following places, which may be of interest for search:

Methods and arrangements for sensing record carriers	G06K 7/00

G06V 30/26

Techniques for post-processing, e.g. correcting the recognition result

Definition statement

This place covers:

Techniques for post-processing, e.g. by correcting the recognition result. This usually involves the context analysis of a certain character or word by taking into account neighbouring characters or words (bi-grams, tri-grams, etc.). The specific context can be analysed:
- lexically, e.g. with a help of a dictionary to correct for mis-recognised characters in a word;
- syntactically, e.g. by considering the syntax rules of a phrase containing the words recognised;
- semantically, e.g. by analysing the intrinsic meaning of the word when considered in the recognised context.

Illustrative example of subject matter classified in this place:
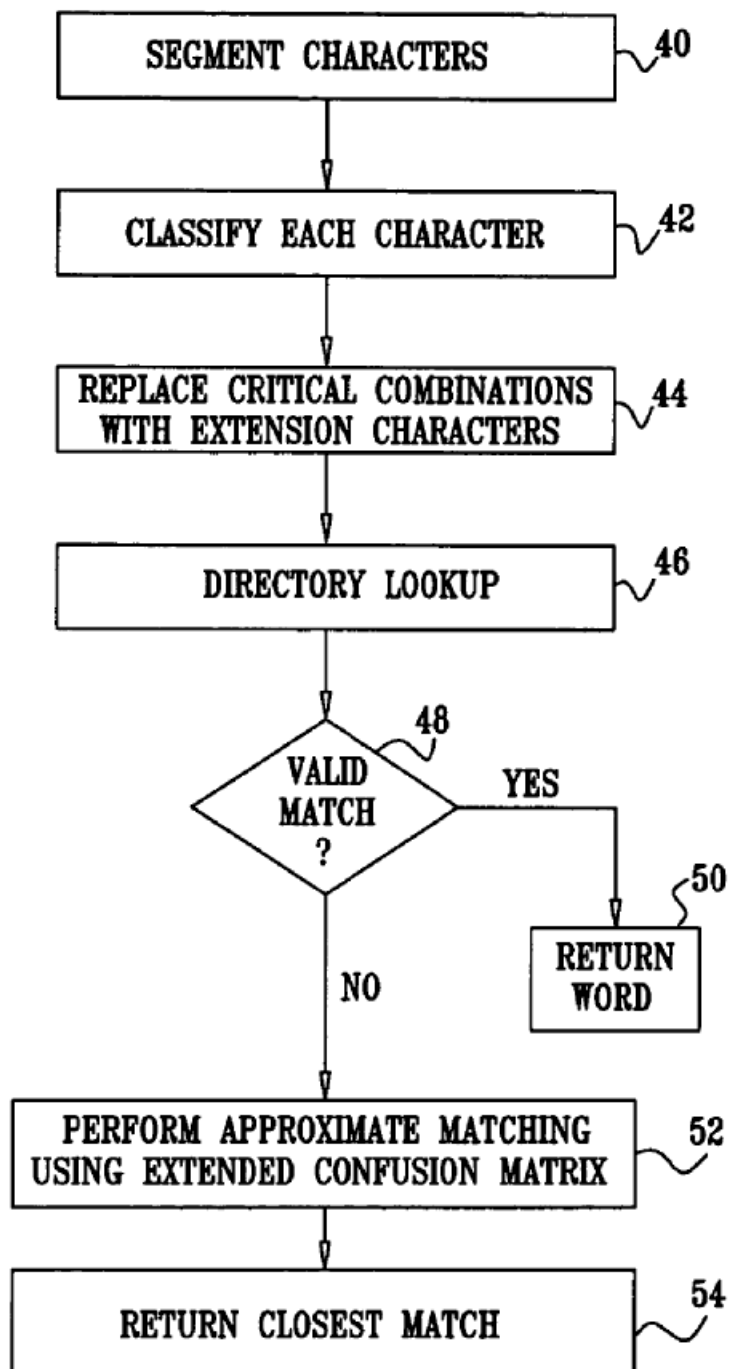
Example of OCR correction using a directory lookup; in case of non-valid matches, an approximate matching is output as closest match using a confusion matrix.

Relationships with other classification places

Methods or arrangements for detection or correction of errors provided by group G06V 30/12 involves such correction after the acquisition step with the general aim to have a reliable input before the subsequent recognition. In contrast, the present group covers the methods and arrangements used after recognition that aim at correcting the final output by using additional information, such as context.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image or video pattern matching, using syntactic or structural representations	G06V 10/86
Segmentation of character regions, for character recognition	G06V 30/148
Information retrieval of unstructured textual data	G06F 16/30
Information retrieval of still image data	G06F 16/50
Handling natural language data	G06F 40/00
Image analysis using region-based segmentation	G06T 7/11

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

n-gram	a contiguous sequence of n items from a given sample of text. The items can be syllables, letters, words or base pairs according to the application. Typical examples are bi-grams and tri-grams.
trie	also called "digital tree" or "prefix tree", is a type of tree data structure used for locating specific keys (items) from within a set of characters or words. In order to access a key (to recover its value, change it, or remove it), the trie is traversed (usually in a depth-first fashion), following the links between nodes.

G06V 30/28

specially adapted to the type of the alphabet, e.g. Latin alphabet

Definition statement

This place covers:

OCR techniques specifically adapted to the type of the alphabet, e.g. Latin or Asian alphabets; alphabet recognition.

Illustrative examples of subject matter classified in this place:

1A.
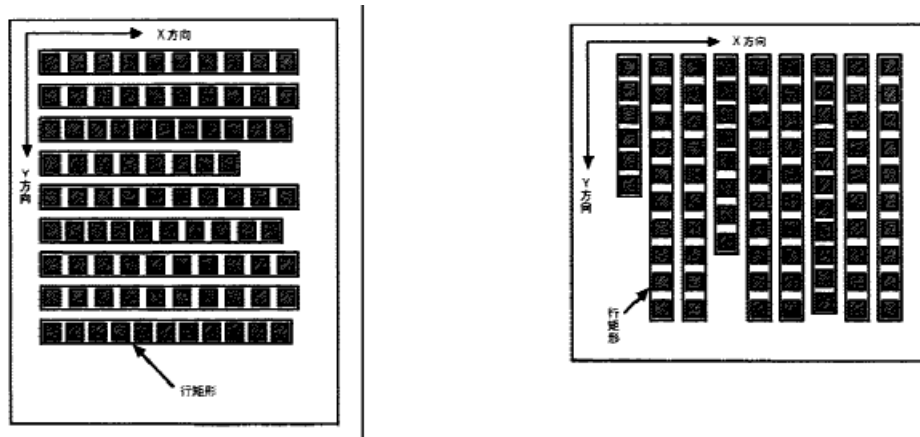
1B.
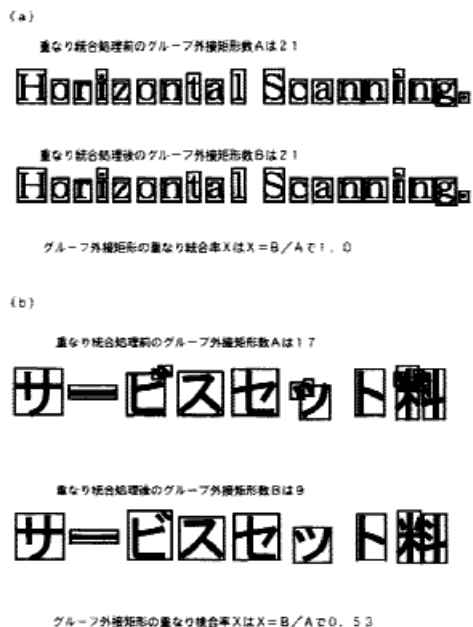
The scanning direction for OCR (horizonal or vertical) is adapted according to the detected alphabet.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Segmentation of character regions, for character recognition	G06V 30/148
Handling natural language data	G06F 40/00
Scanning, transmission or reproduction of documents or the like	H04N 1/00

G06V 30/30

based on the type of data

Definition statement

This place covers:

Recognition according to the type of data (specific images). Examples are:
- images containing characters for discriminating human versus automated computer access ("Completely Automated Public Turing test to tell Computers and Humans Apart" - CAPTCHA);
- musical notations.

Illustrative examples of subject matter classified in this place:

1.
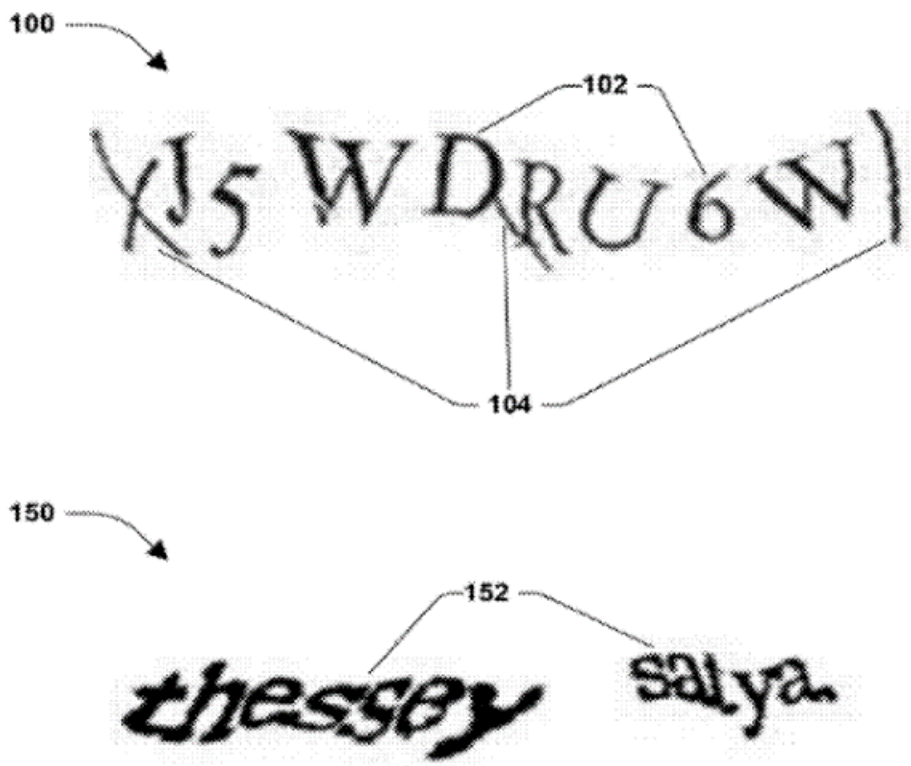
Examples of CAPTCHA images.

2A.
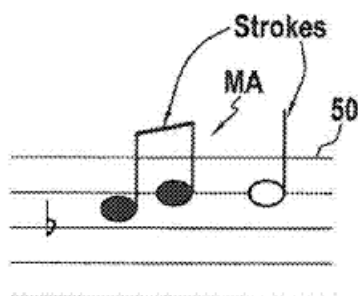
2B.
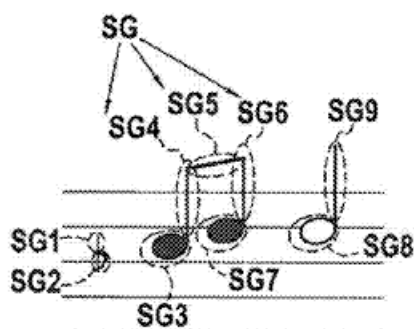
2C.
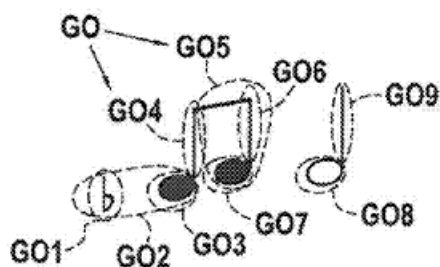
Recognition of music notations by stroke extraction and segmentation of each note.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

User authentication by graphic or iconic representation, in security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/36
Teaching music	G09B 15/00
Means for the representation of music	G10G 1/00

G06V 30/32

Digital ink

Definition statement

This place covers:

Preprocessing, feature extraction, matching, recognition and classification of all kinds of handwritten characters, symbols, drawings, except signatures, on the basis of trajectories as a function of time of a stylus, finger etc. Trajectories can be acquired by a touch pad/screen or by a stylus-like device (in collaboration with a passive or active surface).

Segmentation of strokes, characters or words.

Text and character recognition using temporal information, e.g. free-form handwriting, Asian scripts.

Recognition of drawings using temporal information, e.g. sketches, flow charts, graphical or mathematical symbols or formulae, chemical structure formulae, editorial notes, proof marks.

Illustrative examples of subject matter classified in this place:

1A.
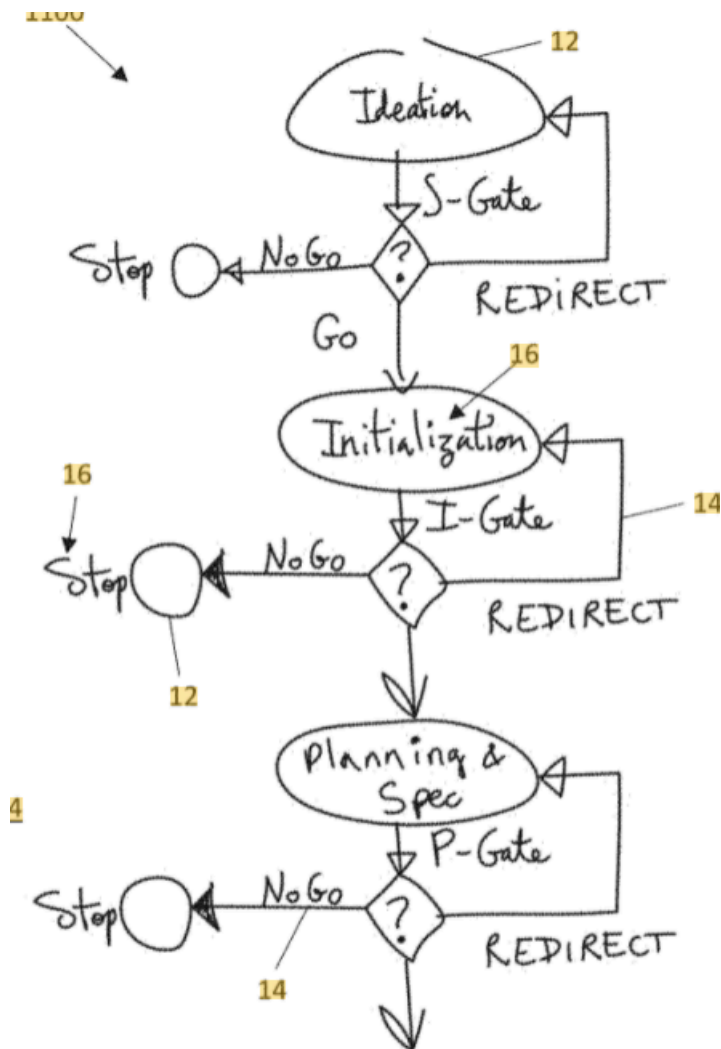
1B.
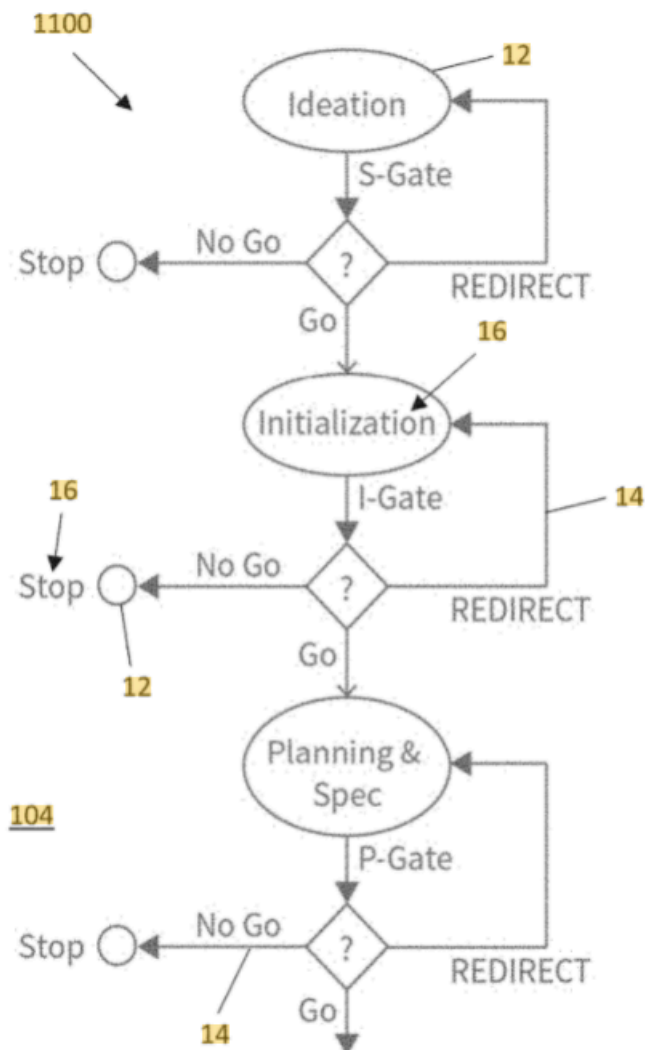
Example of recognising a handwritten flow-chart.

2.
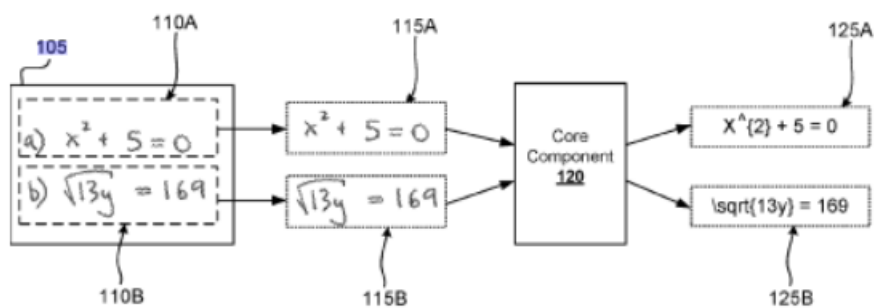
Recognising handwritten mathematical symbols.

3.
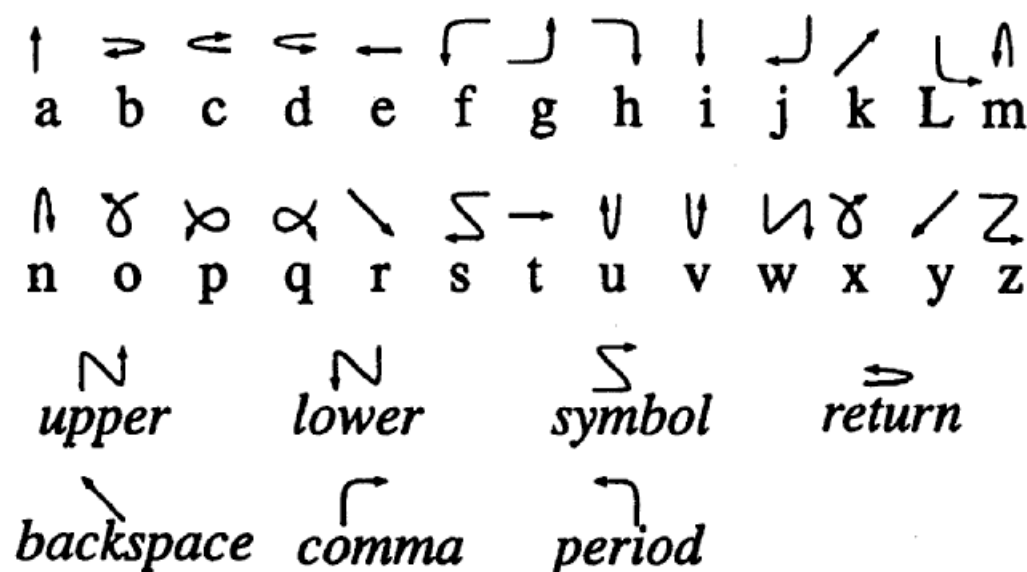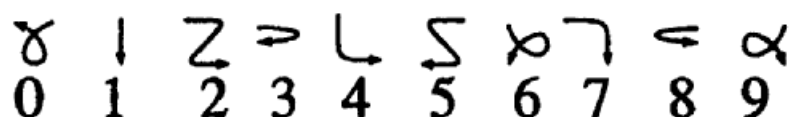
Uni-strokes for computerised interpretation of handwriting.

Relationships with other classification places

The recognition of signatures is considered as a biometric trait and it is covered by group G06V 40/30. If functional details concerning the temporal analysis of the digital ink used for signature recognition are present, double classification with the present group is recommended. The present group assumes that the digital ink is inherently provided with temporal information which is relevant during processing. If the temporal information is not relevant, the character recognition groups provided under group G06V 30/10 apply.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Arrangements for converting the position or the displacement of a member into a coded form, for input arrangements or input/output arrangements for user-computer interaction	G06F 3/03
Inputting data by handwriting, e.g. gestures or text, to a computer via a graphical user interface [GUI] using a touch-screen or digitiser	G06F 3/0488
Digital computing or data processing equipment or methods, specially adapted for specific functions	G06F 17/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

digital ink, electronic ink, e-ink	technology that digitally represents handwriting, e.g. using a finger or a stylus, in its natural form using temporal information. Digital ink may also be referred to as electronic ink or e-ink.

G06V 30/40

Document-oriented image-based pattern recognition

Definition statement

This place covers:

Document analysis, understanding and recognition of document images, involving the analysis of the document content, such as analysis of the geometrical or logical structure. Different types of documents can be involved, such as technical drawings, geographical maps, postal images, e.g. labels on parcels, addresses on postal envelopes.

Illustrative examples of subject matter classified in this place:

1A.
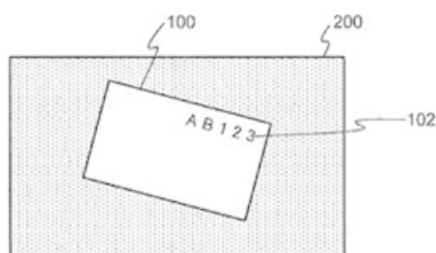
1B.
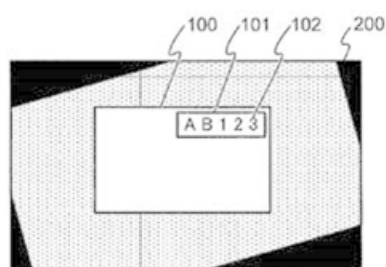
Identification of the text region of a document after its skew correction.

2.
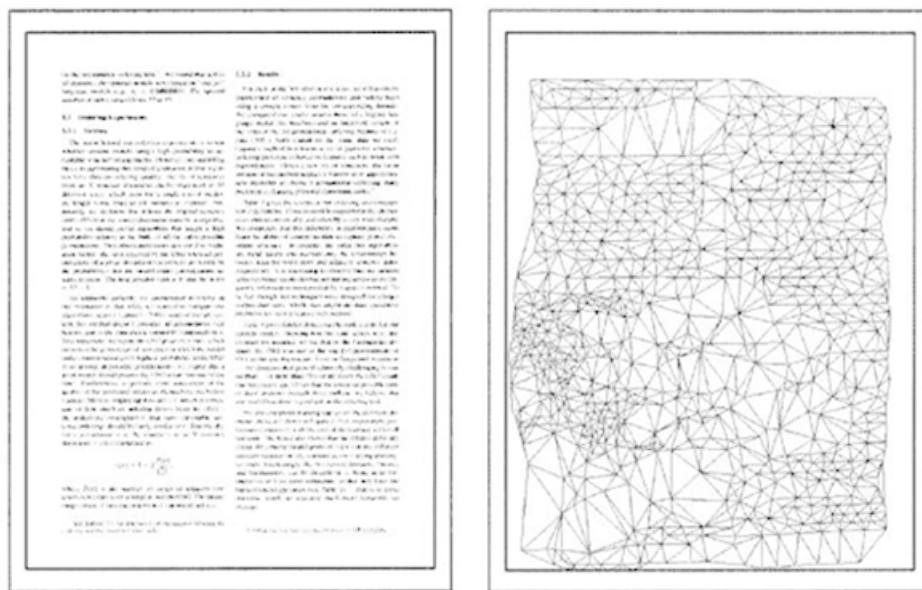
Extraction of image key points, considering them as nodes and constructing a graph representation by connecting with edges the neighbouring nodes; the graph-based representation is later used for document matching.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Sorting of mail or documents using means for detecting the destination	B07C 3/10
Input arrangements for interaction between user and computer	G06F 3/01
Testing to determine the identity or genuineness of valuable papers, e.g. banknotes	G07D 7/00

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition of printed characters based on code marks	G06V 30/224
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements	G06F 3/00
Information retrieval of unstructured textual data	G06F 16/30
Information retrieval of still image data	G06F 16/50
Handling natural language data	G06F 40/00

G06V 30/41

Analysis of document content (recognition of printed characters based on code marks [G06V 30/224](#))

Definition statement

This place covers:

Document analysis, recognition and understanding by processing:

- Structured documents such as business forms, bank checks whose layout is provided with printed lines or input boxes, bounding boxes, checkboxes, straight lines or tables;
- Classification of document image content by identification of text regions, photographs, tables;
- Extracting and analysing the geometrical structure, e.g. the layout tree representation in which different entities such as paragraphs, images, etc. are represented as nodes of a tree or a graph;
- Extracting and analysing the logical structure, e.g. identification of the chapter headings, sections, columns, titles, paragraphs, captions, page numbers, or identification of the constituting elements such as authors, keywords, postal codes, money amounts;
- Document matching, e.g. by establishing the degree of (dis)similarity between two document images, one reference/template document image and one input document image.

Illustrative examples of subject matter classified in this place:

1.

Identification of the text and image regions; the image tiles are marked with an "I" and text tiles are marked with a "T".

2.
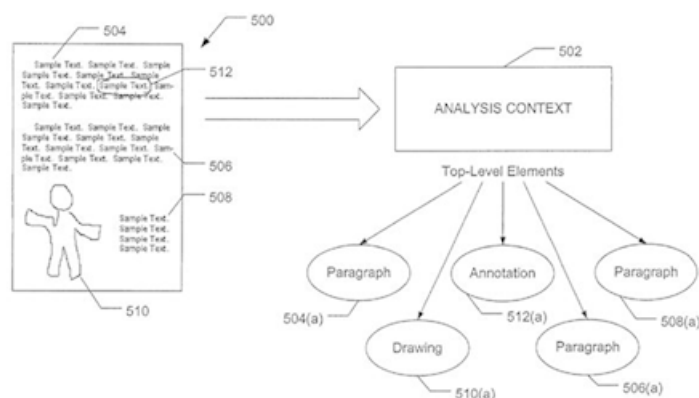
Extraction of document structure by analysis of its content, resulting in the identification of elements such as paragraphs, drawings, handwritten annotations.

References

Limiting references

This place does not cover:

Recognition of printed characters based on code marks	G06V 30/224

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding, by selection of a specific region containing or referencing a pattern; Image preprocessing for image or video recognition by locating or processing of specific regions to guide the detection or recognition, e.g. highlights, fiducial marks or predetermined fields	G06V 10/22
Local feature extraction for image or video recognition or understanding by analysis of parts of the pattern, e.g. by detecting edges, contours, loops, corners, strokes or intersections; Extraction of image or video features for image or video recognition or understanding using connectivity analysis, e.g. of connected components, edge linking or neighbouring slice analysis	G06V 10/44
Analysis of text in scene images, e.g. of license plates, overlay texts or captions on TV images	G06V 20/62
Aligning or centring of the image pick-up or the image field, for character recognition	G06V 30/146
Segmentation of character regions, for character recognition	G06V 30/148
Information retrieval of unstructured textual data	G06F 16/30
Information retrieval of still image data	G06F 16/50
Handling natural language data	G06F 40/00
Image analysis using region-based segmentation	G06T 7/11
Details of scanning heads for optical reproduction of scanned documents or the like	H04N 1/036

G06V 30/42

based on the type of document

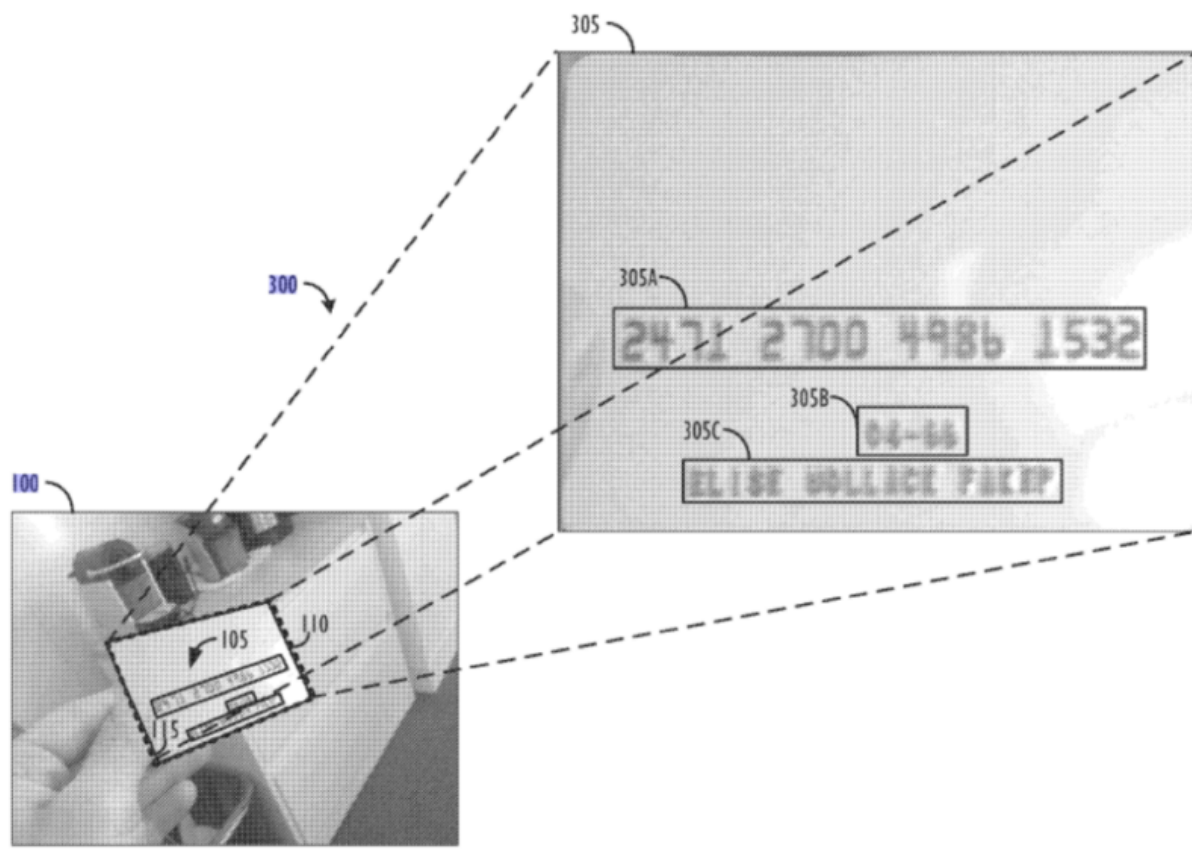## Definition statement

This place covers:

Document analysis, understanding and recognition, based on the type of document.

Examples include:
- technical drawings and geographical maps;
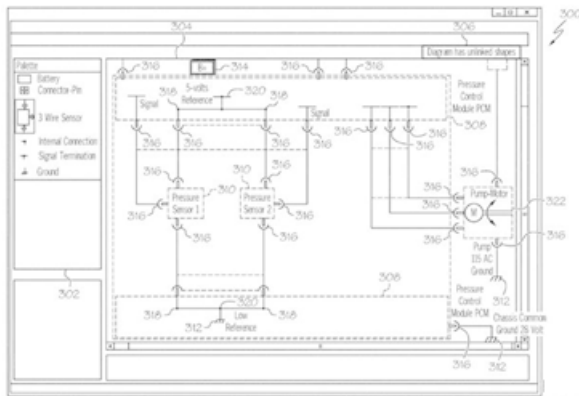- postal images, e.g. labels or addresses on parcels or postal envelopes.

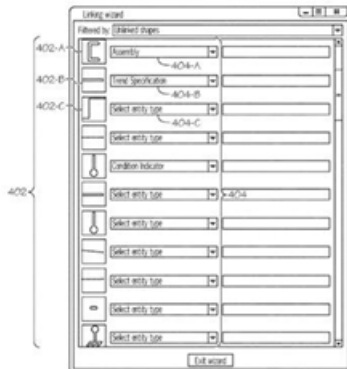Illustrative examples of subject matter classified in this place:

1.

Credit card detection (bottom left), perspective mapping (top right), extraction of the relevant fields and recognition of the information.

2A.

2B.

Acquisition and recognition of elements in a schematic drawing (fig. 2A) and mapping the recognition results into a database (fig. 2B).

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding involving the determination of region of interest [ROI] or a volume of interest [VOI]	G06V 10/25
Aligning or centring of the image pick-up or image-field, for character recognition	G06V 30/146
Map- or contour-matching specially adapted for navigation in a road network using correlation of data from several navigational instruments	G01C 21/30
Information retrieval of unstructured textual data	G06F 16/30
Information retrieval of still image data	G06F 16/50
Handling natural language data	G06F 40/00
Image analysis using region-based segmentation	G06T 7/11

G06V 40/00

Recognition of biometric, human-related or animal-related patterns in image or video data

Definition statement

This place covers:

Detection, feature extraction, classification, identification, authentication of human-related or animal-related patterns in images or video. It includes monitoring behaviour, habits or activities, such as eating and sleeping patterns, sport activities, gait recognition, hand gestures (both static and dynamic), including those performed on a touch screen or freely in the air.

Biometric identification and authentication using body parts, e.g. fingerprints, palmprints, footprints, using faces or eye characteristics such as vessel patterns of the eye sclera or eye fundus, or iris patterns. Other examples of biometric traits include measurements obtained from the hand geometry or the limbs, or personal signatures.

Writer recognition, i.e. establishing the identity of the person who wrote a piece of text.

References

Application-oriented references

Examples of places where the subject matter of this place is covered when specially adapted, used for a particular purpose, or incorporated in a larger system:

Means to switch the anti-theft system of vehicles on or off, using biometry	B60R 25/25
User authentication using biometric data for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/32
Identity check of a pass holder for individual registration on entry or exit using biometric data	G07C 9/25
Identity check for individual registration on entry or exit without a pass using biometric data	G07C 9/37
Acquiring the identification of end-users of distributed media content using their biometric characteristics	H04N 21/4415

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding	G06V 10/20
Extraction of image or video features, e.g. computing feature vectors, image or video descriptors	G06V 10/40
Recognition of moving objects or obstacles, e.g. vehicles or pedestrians; Recognition of traffic objects, e.g. traffic signs, traffic lights or roads	G06V 20/58
Recognition or understanding of scenes inside of a vehicle	G06V 20/59
Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Recognition of digital ink within image or video data	G06V 30/32
Arrangements or fittings on vehicles for protecting or preventing injuries to occupants or pedestrians in case of accidents or other traffic risks, including means for detecting the presence or position of passengers, passenger seats or child seats	B60R 21/015

Digitisers as the input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
Analysis of motion in images	G06T 7/20
Image analysis for determining position or orientation of objects	G06T 7/70
Checking-devices for individual entry or exit registers	G07C 9/00
Burglar, theft or intruder alarms	G08B 13/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

gesture	posture or hand movement denoting a certain meaning, e.g. deaf sign language.

G06V 40/10

Human or animal bodies, e.g. vehicle occupants or pedestrians; Body parts, e.g. hands

Definition statement

This place covers:

Detection, feature extraction, classification, identification and recognition of:

- human bodies;
- human body parts e.g. arms, hands, legs;
- vehicle occupants or pedestrians as perceived by a camera inside or outside a vehicle;
- animal bodies;
- biometric identification based on hand measurements, e.g. distances between joints, length of the fingers, etc.;
- static gestures, e.g. pose recognition.

Illustrative examples of subject matter classified in this place:
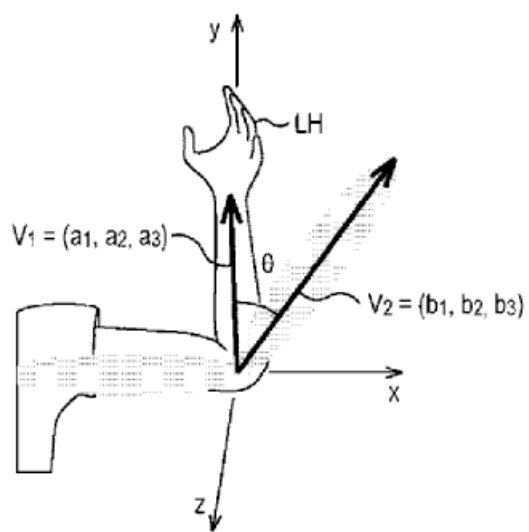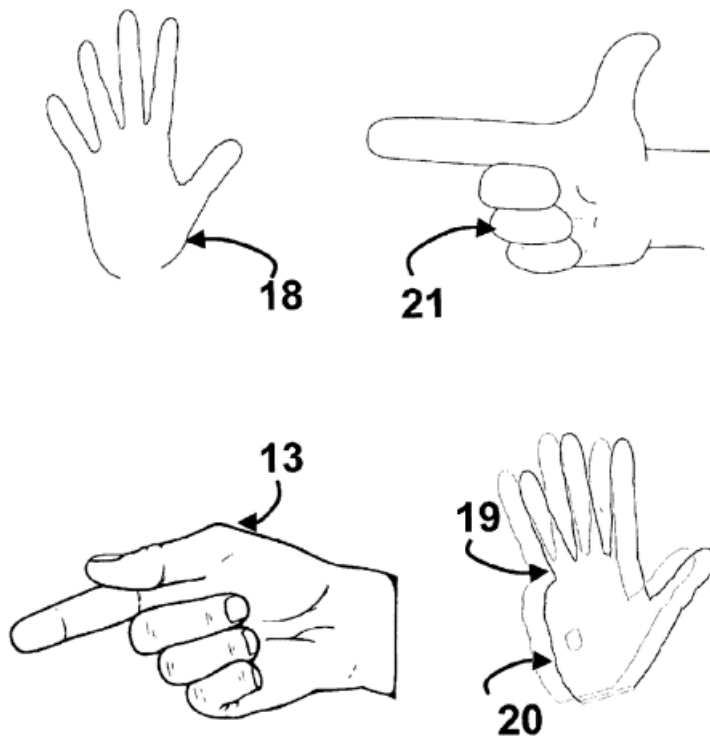
1A.
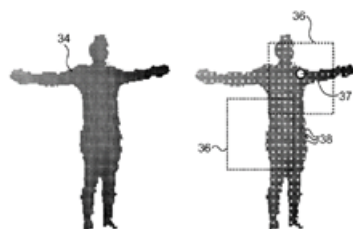
1B.

Successive stages in the process of pose estimation.

2.

Pose recognition of the arm by quantifying its direction as a vector.

3.

Different hand configurations used for a secret sign.

Relationships with other classification places

Recognition of body movements, e.g. gesture recognition in a temporal image sequence, or monitoring sport training in video is classified in group G06V 40/20.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Extraction of image or video features by performing operations within image blocks or by using histograms, e.g. histogram oriented gradients [HoG]	G06V 10/50
Recognition of moving objects or obstacles in scenes, e.g. vehicles or pedestrians; Recognition of traffic objects, e.g. traffic signs, traffic lights or roads	G06V 20/58
Recognition or understanding of scenes inside of a vehicle	G06V 20/59
Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Arrangements or fittings on vehicles for protecting or preventing injuries to occupants or pedestrians in case of accidents or other traffic risks, including means for detecting the presence or position of passengers, passenger seats or child seats	B60R 21/015
Image analysis for determining position or orientation of objects	G06T 7/70
Burglar, theft or intruder alarms	G08B 13/00

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

| dynamic gesture	movement of the hand encoding a certain meaning, e.g. deaf sign language
static gesture	posture of a hand denoting a certain meaning

G06V 40/12

Fingerprints or palmprints

Definition statement

This place covers:

Acquisition, pre-processing, feature extraction and

- matching for biometric identification purposes; or
- classification into types; or
- detecting the live character of the finger, i.e. distinguishing from a fake or cadaver finger by using either specialised acquisition arrangements or by image processing.
- Pre-processing and feature extraction for the purpose of fingerprint recognition, e.g.:
- denoising/filtering, enhancement, normalisation;
- minutiae extraction;
- ridge properties extraction, such as ridge spatial frequency and ridge orientation.

Illustrative examples of subject matter classified in this place:

1.
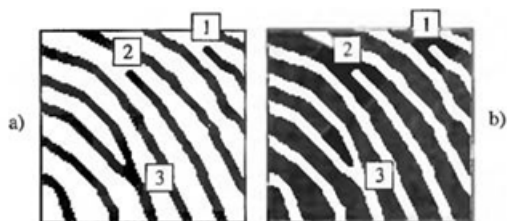
Fingerprint representations by ridges (thin and thick lines) and minutiae (ridge endings (1) and (2) and bifurcations (3)).

2.
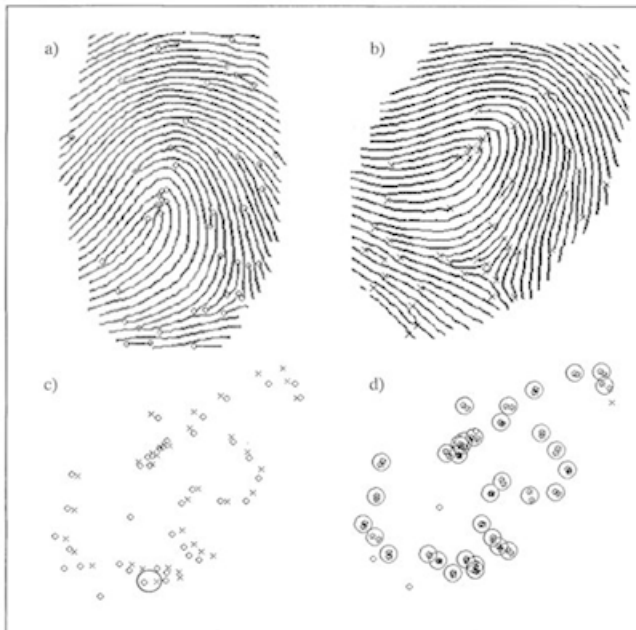
The sets of minutiae extracted from two fingerprint images are matched to establish the person's identity; the small circles denote matched pairs.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding	G06V 10/20
Noise filtering for image or video recognition or understanding	G06V 10/30
Writer recognition; Reading and verifying signatures	G06V 40/30
Spoof detection in image or video recognition	G06V 40/40
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70
Identification of persons	A61B 5/00
Fittings or systems for preventing or indicating unauthorised use or theft of vehicles	B60R 25/00
Digitisers as the input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
Security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/00
Checking-devices for individual entry or exit registers	G07C 9/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00
Means for preventing unauthorised calls from a telephone set	H04M 1/667

Special rules of classification

Detection of the static pose of the hand or biometrics obtained from hand geometrical arrangement of the fingers, e.g. distance between the finger joints is classified in group G06V 40/10.

Techniques involving multiple biometrics are classified in group G06V 40/70.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

fingerprints or palmprints	2D or 3D images of the (sub-)surface (sub-)epidermal structures of fingers or palm

G06V 40/13

Sensors therefor

Definition statement

This place covers:

Fingerprint or palmprint sensors of all kinds:
- optical sensing, e.g. through reflection in optical elements such as prisms;
- non-contact direct (distance) sensing;
- capacitive/RF (active impedance) sensing;
- ultrasonic sensing;
- thermal sensing;
- pressure sensing;
- piezoelectric sensing;
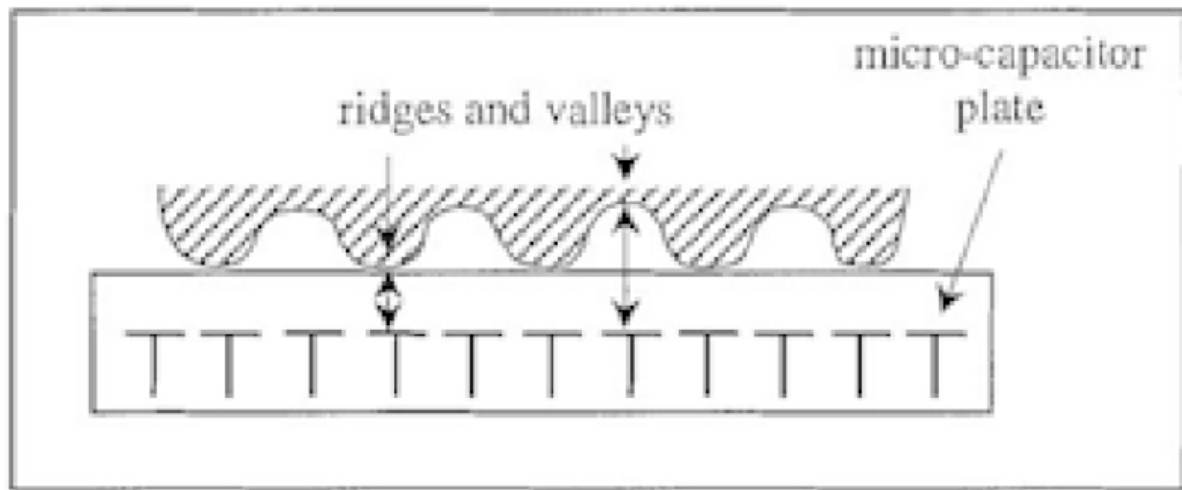- sweep sensing etc.

Protecting the fingerprint sensors against wear and tear.

Illustrative examples of subject matter classified in this place:
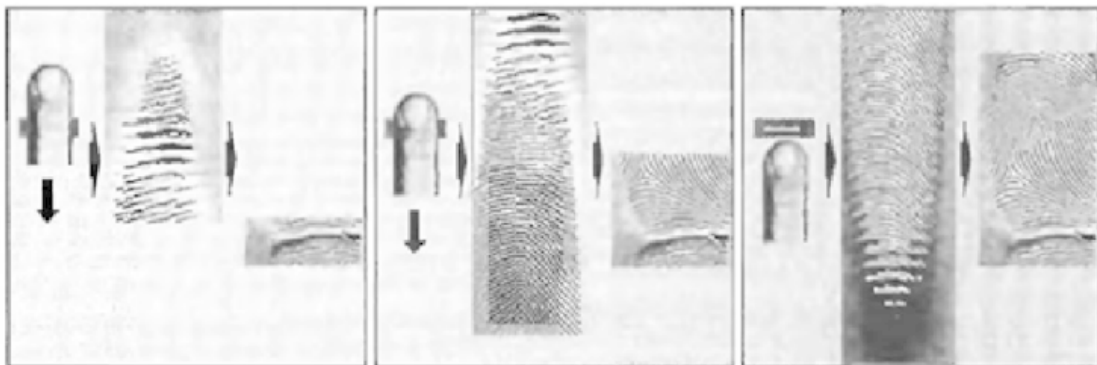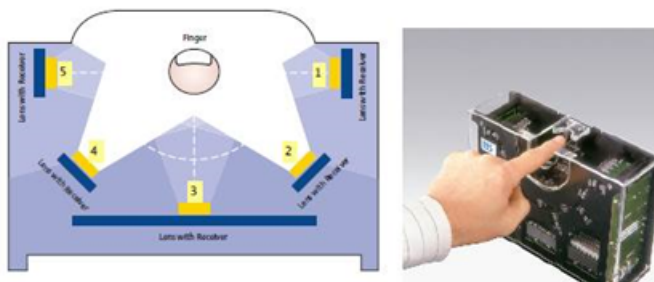
1A.
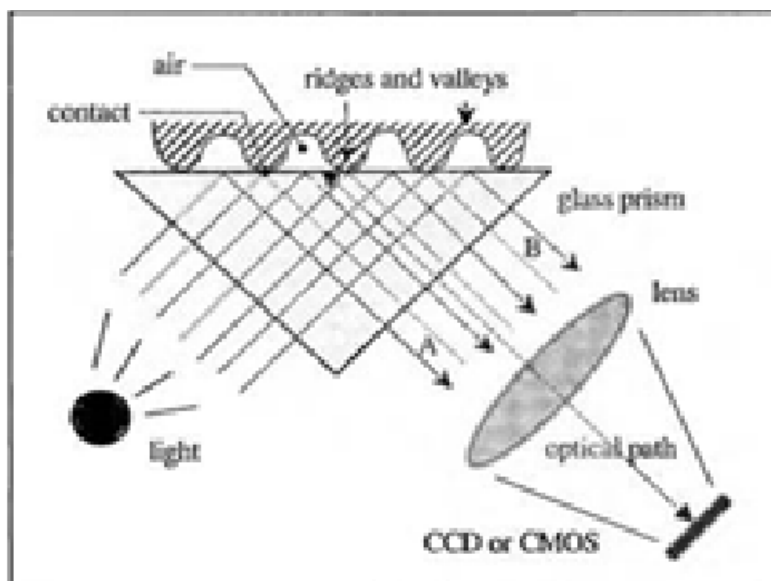
1B.

Optical fingerprint sensing and capacitive sensing.

2.

Sweep-type sensing

3.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Sensors for the recognition of vascular patterns	G06V 40/145
Sensors for the recognition of eye characteristics	G06V 40/19
Identification of persons	A61B 5/00
Fittings or systems for preventing or indicating unauthorised use or theft of vehicles	B60R 25/00
Sonar systems specially adapted for mapping or imaging	G01S 15/89
Digitisers as the input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
User authentication using biometric data for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/32
Checking-devices for individual entry or exit registers	G07C 9/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00
Means for preventing unauthorised calls from a telephone set	H04M 1/667
Scanning, transmission or reproduction of documents or the like	H04N 1/00
Television systems	H04N 7/00

Special rules of classification

Techniques which combine fingerprint sensors and vein (vascular) sensors are classified in groups G06V 40/13 and G06V 40/145.

Acquisition of fingerprint images generally requires specialised hardware which is essentially different from normal cameras. For this reason, fingerprint sensors are not classified in the generic group G06V 10/10.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

(RF) active sensing	active measure of the impedance formed between the finger and an electrode plate in the sensor, typically using RF band waves
capacitive sensing	static measure of the capacitance formed between the skin and an electrode plate in the sensor
FTIR sensing	frustrated total internal reflection sensing – the finger is imaged at the Brewster angle (air/glass); light rays are reflected only from the valley zones of the fingerprint, the ridges (partly) absorb the light
sweep sensor	sensor acquiring partial fingerprint images and stitching them together to form a full fingerprint image

G06V 40/14
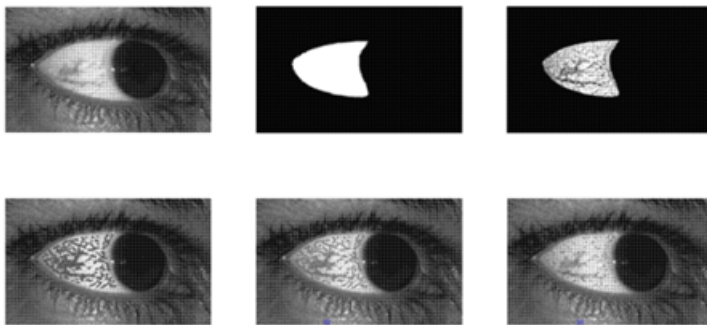
Vascular patterns

Definition statement

This place covers:

Vascular pattern acquisition, pre-processing, feature extraction and matching for biometric identification or classification purposes. The steps of pre-processing and feature extraction for the vascular pattern recognition may include:

- de-noising / filtering, enhancement or normalisation of vein / vessel images;
- detection, segmentation or thinning in vein / vessel images;
- pattern or signature matching in vein / vessel images.

Illustrative examples of subject matter classified in this place:
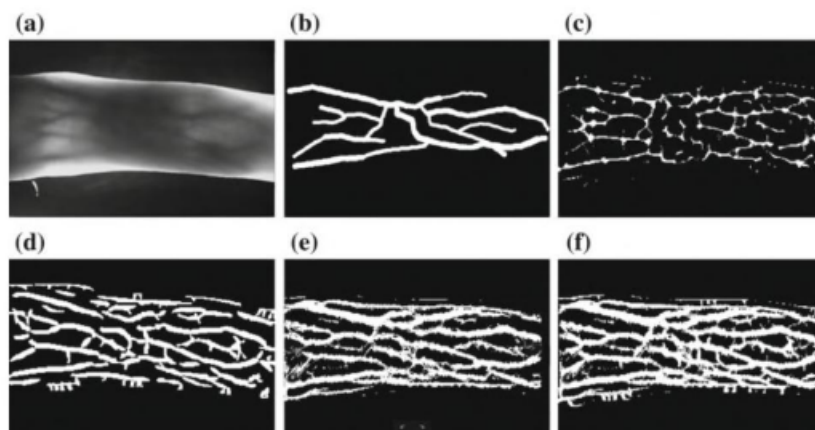
1.

Vascular patterns of the eye used for biometric identification.

2.
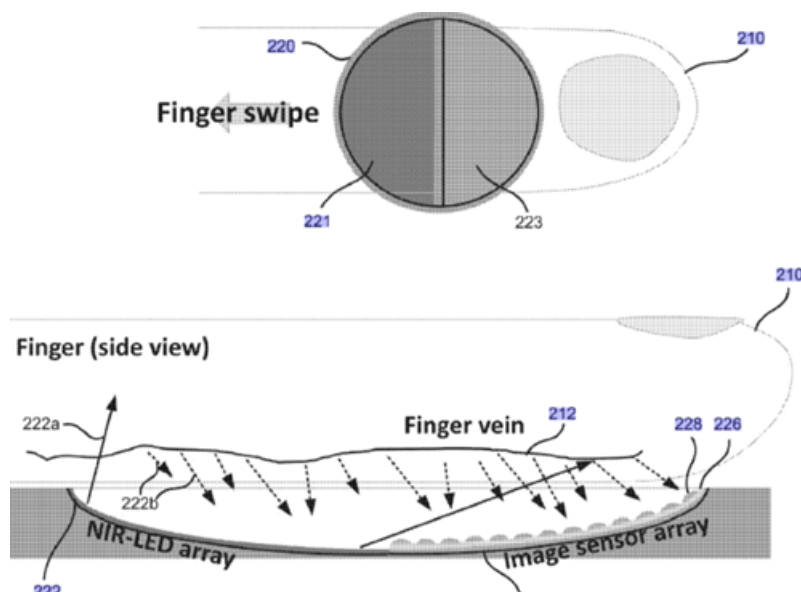
Vascular patterns of the finger used for biometric identification.

3.

Vascular patterns of the hand used for biometric identification.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image preprocessing for image or video recognition or understanding	G06V 10/20
Recognition of fingerprints or palmprints within images or video	G06V 40/12
Recognition of faces within images or video	G06V 40/16
Recognition of eye characteristics within images or video, e.g. of the iris	G06V 40/18
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70
Identification of persons	A61B 5/00
Fittings or systems for preventing or indicating unauthorised use or theft of vehicles	B60R 25/00
Digitisers as input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
User authentication using biometric data for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/32
Checking-devices for individual entry or exit registers	G07C 9/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00
Means for preventing unauthorised calls from a telephone set	H04M 1/667

Special rules of classification

Techniques involving multiple biometrics are classified in group G06V 40/70.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

vascular patterns	2D or 3D images of the (sub-)surface of fingers, palm or sclera showing the vessels/veins

G06V 40/145

Sensors therefor

Definition statement

This place covers:

Vascular imagers such as a finger vein scanner or palm vein scanner which use near infrared lights combined with a special camera to capture vein patterns.

Finger, palm or eye vessels sensors of all kinds.

Near infrared cameras used for making the vascular pattern visible.

Illustrative example of subject matter classified in this place:

(a) Overhead illumination   (b) Side illumination
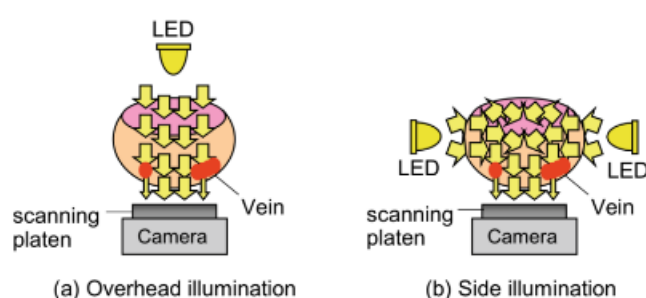
Relationships with other classification places

Techniques involving acquisition of finger movements on a digitiser are classified in group G06F 3/041.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition of fingerprints or palmprints within images or video	G06V 40/12
Recognition of faces within images or video	G06V 40/16
Recognition of eye characteristics within images or video, e.g. of the iris	G06V 40/18
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70
Identification of persons	A61B 5/00
Fittings or systems for preventing or indicating unauthorised use or theft of vehicles	B60R 25/00
Digitisers as input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
User authentication using biometric data for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/32
Checking-devices for individual entry or exit registers	G07C 9/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00

Means for preventing unauthorised calls from a telephone set	H04M 1/667
Scanners in general	H04N 1/00
Cameras in general	H04N 7/00

Special rules of classification

Techniques which combine fingerprint sensors and vein (vascular) sensors are classified in groups G06V 40/13 and G06V 40/145.

Acquisition of vascular patterns generally requires specialised hardware which is essentially different from normal cameras. For this reason, vascular sensors are not classified in the generic group G06V 10/10.

G06V 40/16

Human faces, e.g. facial parts, sketches or expressions

Definition statement

This place covers:

Detection, localisation, representation and recognition of the face or of facial parts.

Detection of multiple faces in an image or video, e.g. for video-conferencing.

Feature extraction based on the facial image taken as a whole, e.g. holistic features such as colour of the face region, eigenfaces, Fisherfaces, etc., or based on facial parts, e.g. local features such facial components (eyes, nose, etc.), their geometric configuration.

Face occlusion detection.

Race, gender and age detection based on facial features (e.g. skin wrinkles).

Recognition of facial expressions, e.g. static or dynamic expressions.

Spoof-by-picture using an image of the face.

Detection of faces using different types of acquisition modalities, e.g. infrared (thermal) images, or their combination.

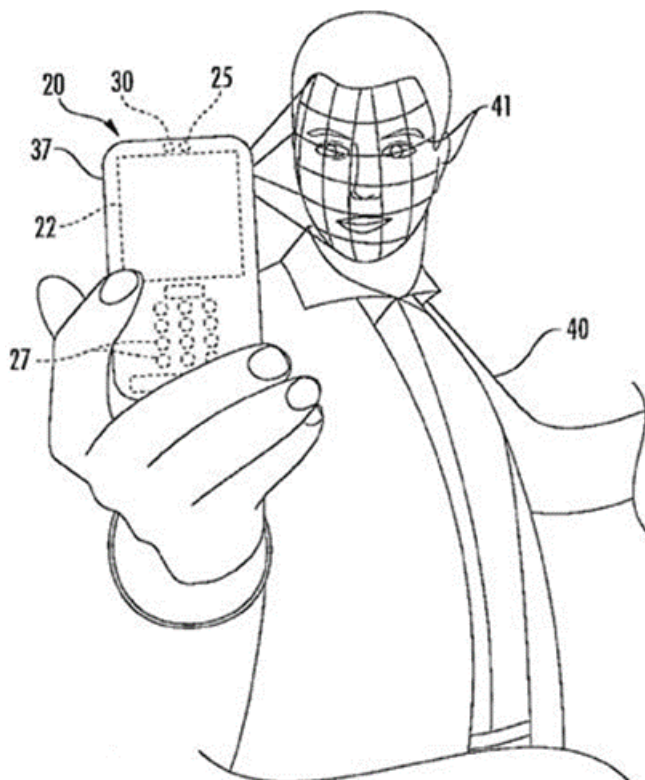Facial skin detection based on skin properties, e.g. skin colour.

Illustrative examples of subject matter classified in this place:

1.
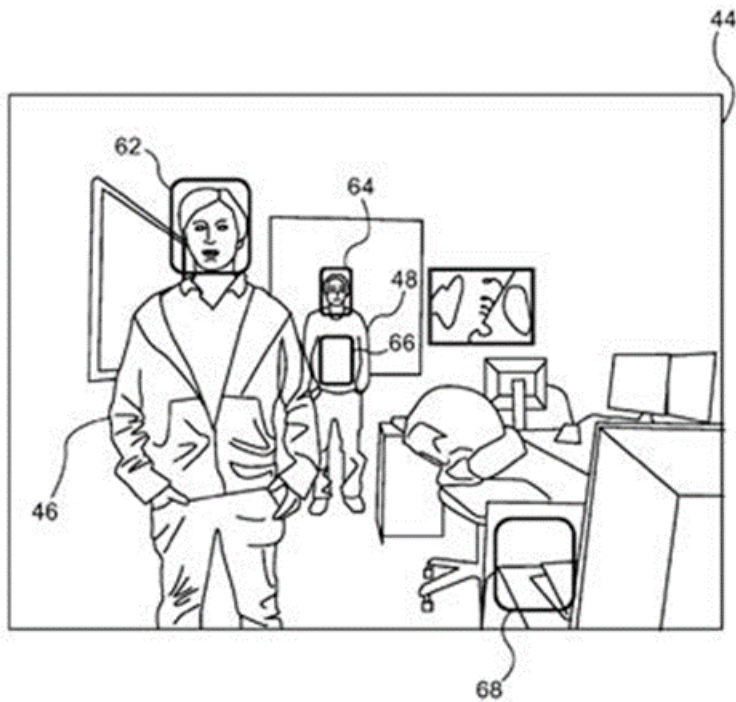
Faces detected in an image.

2.

Acquisition of a face in 3D by means of a smartphone.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognising three-dimensional [3D] objects in scenes	G06V 20/64
Recognition of human or animal bodies in images or video, e.g. vehicle occupants or pedestrians; Body parts, e.g. hands	G06V 40/10
Recognition of fingerprints or palmprints within images or video	G06V 40/12
Recognition of eye characteristics within images or video, e.g. of the iris	G06V 40/18
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70

Special rules of classification

Recognition using iris patterns of the eye are classified in group G06V 40/18. If the technical aspects of a document cover aspects relevant both for face recognition and iris recognition, both aspects are classified in groups G06V 40/16 and G06V 40/18.

Techniques for spoof detection of faces, e.g. spoof-by-picture, are classified in groups G06V 40/16 and G06V 40/40.

Techniques for face recognition using 3D models are also classified in group G06V 20/64.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

eigenface	face representation using a principal component analysis in a high-dimensional space created from images of faces. The eigenvectors of the representation are derived from the covariance matrix of the probability distribution computed in this high-dimensional vector space.
fisherface	linear discriminant analysis [LDA] applied in a multi-dimensional representation space created from a set of face images, resulting in a set of basis vectors defining that space
frontal face recognition	face images are generally obtained by placing a camera in front of the subject who is asked to look at the camera while the picture is taken
illumination-invariant recognition	recognition insensitive to changes in lighting conditions
multiview face recognition	employs a gallery of images of every face at various poses to cover multiple views for each face
pose-invariant recognition	recognition insensitive to changes in pose

G06V 40/18

Eye characteristics, e.g. of the iris
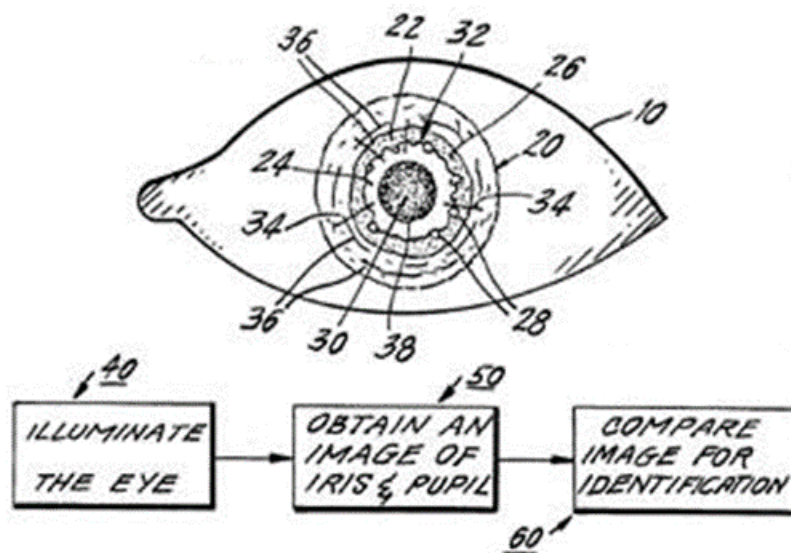
Definition statement

This place covers:

Acquisition, pre-processing, feature extraction, clustering, classification of eye regions or eye components (e.g. iris, pupil, eyelids, eyelashes, sclera) for:

- Biometric identification and authentication by eye characteristics, e.g. iris recognition;
- Recognition of eye movements (e.g. fixation, saccade, smooth pursuit) and detection of eye blink;
- Eye tracking, gaze estimation and correction, by acquiring the image of the eye or in combination with the analysis of the scene (e.g. using saliency models) for biometric purposes. The techniques involved may use specialised hardware, such as head-mounted systems, infrared or visible light, or may use computer vision methods, such as modelling eye and scene geometry, appearance-based methods, etc.;
- Red eye detection due to image acquisition using a camera flash;
- Monitoring attention-based eye movements, e.g. for measuring the time spent when looking at products for advertisements purposes;
- Detecting and monitoring the eye open and eye closed states, e.g. for monitoring driver fatigue.

Illustrative examples of subject matter classified in this place:

1. Iris recognition

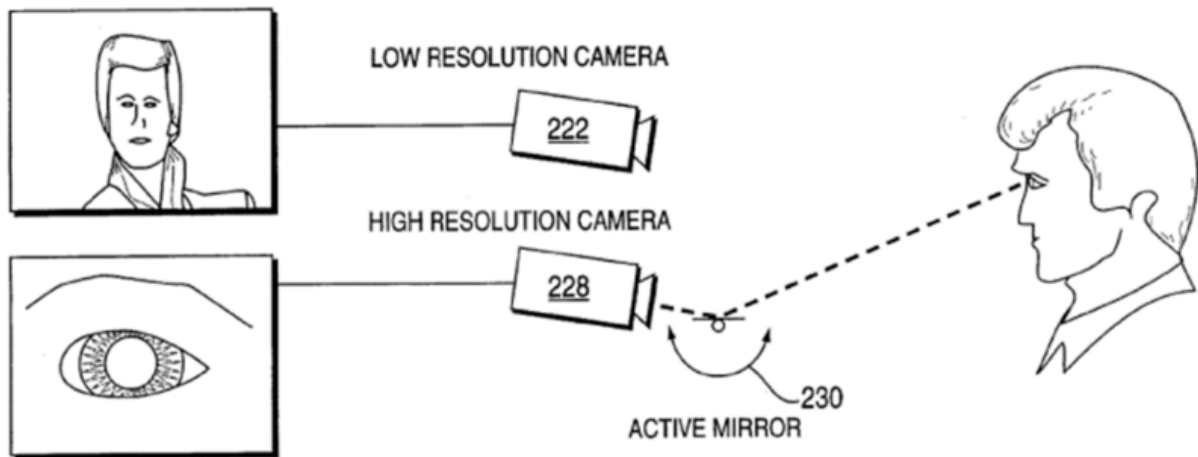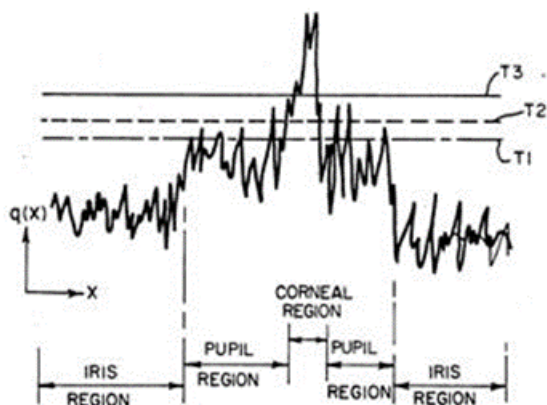Patterns of the eye are extracted and used for personal identification.

2.

The IrisCode (a binary sequence which characterises the texture of the iris) may be used for personal identification.
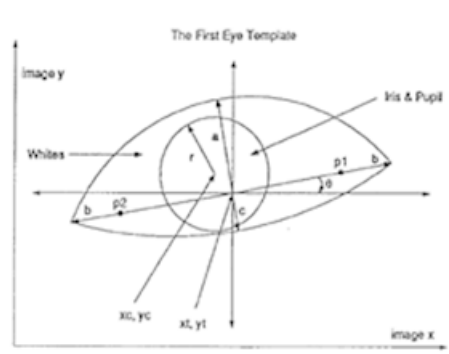
3.

Acquisition of the eye using a dual system based on a low-resolution acquisition to detect the face and high-resolution camera to detect the iris.
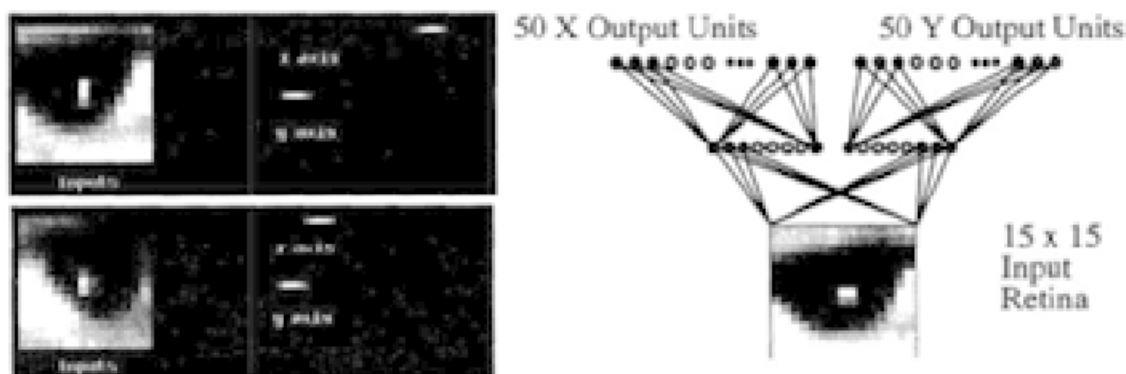
4.

Detection of the iris region using the variations of the image grey levels along a crossing line.

5.

Geometrical representation of different eye components.

6.

Eye detection using a neural network.

References

Informative references

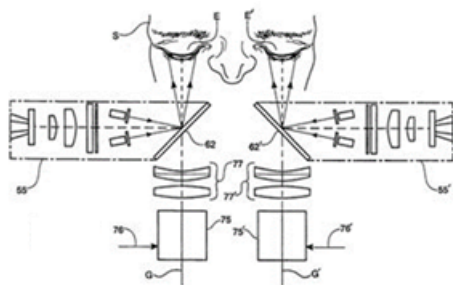*Attention is drawn to the following places, which may be of interest for search:*

Computing image salient features for recognition purposes	G06V 10/46
Recognition or understanding of scenes inside of a vehicle	G06V 20/59
Recognition of human or animal bodies within images or video, e.g. vehicle occupants or pedestrians; Body parts, e.g. hands	G06V 40/10
Recognition of fingerprints or palmprints in images or video	G06V 40/12
Recognition of human faces in images or video, e.g. facial parts, sketches or expressions	G06V 40/16
Apparatus for testing the eyes; Instruments for examining the eyes	A61B 3/00
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements	G06F 3/00
Image analysis for determining position or orientation of objects or cameras	G06T 7/70
Scanning, transmission or reproduction of documents or the like; Colour correction or control; Red eye correction	H04N 1/62

Special rules of classification

Recognition using iris patterns of the eye are classified in group G06V 40/18. If the technical aspects of a document cover aspects relevant both for face recognition and iris recognition, both aspects are classified in groups G06V 40/16 and G06V 40/18.

Techniques for spoof detection of faces, e.g. spoof-by-picture, are classified in groups G06V 40/16 and G06V 40/40.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

LoG	line of gaze (optical axis)
LoS	line of sight (visual axis)
PCR	pupil corneal reflection
PoR	point of regard
WFOV (camera), WFOV	wide field of view (camera provided with a relatively large view to roughly detect the position of the eye)
NFOV (camera), NFOV	narrow field of view (camera provided with a narrow field of view which acquires a more precise eye image)
Purkinje images	reflections of objects present in the environment which can be seen on the structure of the eye, e.g. sclera
saliency map	map displaying areas of higher visual importance, e.g. luminance contrast, semantic contrast, etc.

G06V 40/19

Sensors therefor

Definition statement

This place covers:

Special sensors or acquisition arrangements adapted to acquire the image of an eye or its anatomical components (iris, eye fundus, etc.) for biometric purposes.

Illustrative examples of subject matter classified in this place:

1.

Optical system for eye and iris acquisition.

2.

Optical system for eye and iris acquisition.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition of fingerprints or palmprints within image or video data	G06V 40/12
Recognition of faces within image or video data	G06V 40/16
Recognition of eye characteristics within image or video data, e.g. of the iris	G06V 40/18
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70
Apparatus for testing the eyes; Instruments for examining the eyes	A61B 3/00
Identification of persons	A61B 5/00
Fittings or systems for preventing or indicating unauthorised use or theft of vehicles	B60R 25/00
Digitisers as the input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
Security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/00
Checking-devices for individual entry or exit registers	G07C 9/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00
Means for preventing unauthorised calls from a telephone set	H04M 1/667
Scanners in general	H04N 1/00
Cameras in general	H04N 7/00

Special rules of classification

Acquisition of eye patterns generally requires specialised hardware which is essentially different from normal cameras. For this reason, eye sensors are not classified in the generic group G06V 10/10.

G06V 40/20

Movements or behaviour, e.g. gesture recognition (recognition of facial expressions G06V 40/16)

Definition statement

This place covers:

Detection, tracking, recognition of:
- Gestures, e.g. whole body, upper body, hand, arm, head, free movements for sport activities, hand movements for interface control;
- Hand or arm movements, e.g. for deaf sign language recognition;
- Gait recognition, e.g. walking, running;
- Lip movement, e.g. for lip-reading.

Recognising human behaviour, e.g. daily activities; monitoring eating patterns or calories intake.

Recognition of movements during sport activities.

Recognising touch or drawing movements on a surface or in a three-dimensional space, e.g. patterns on a touch screen, smart tables, smart whiteboards, etc.

Illustrative examples of subject matter classified in this place:
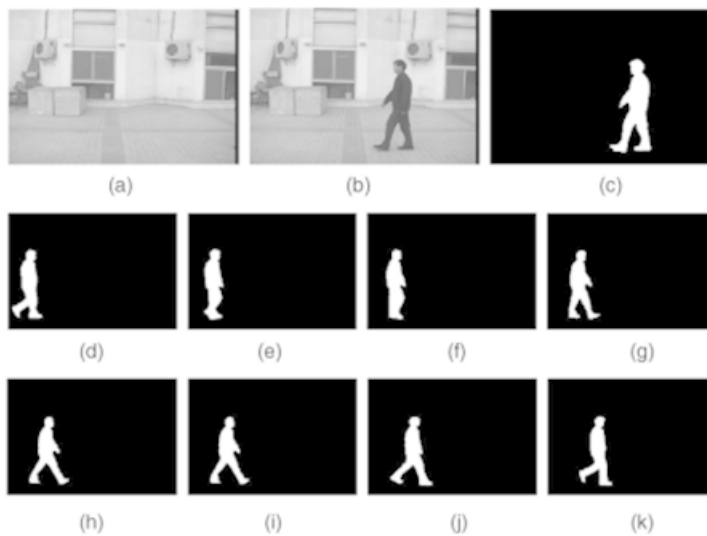
1.
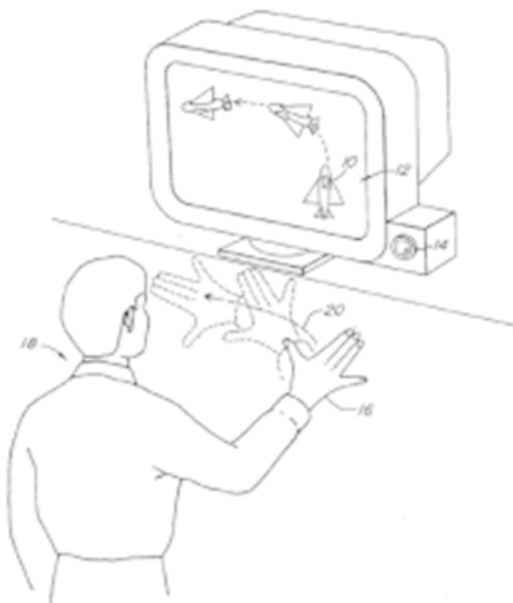
Recognising the movement of a hand for controlling an object on the screen of a computer.
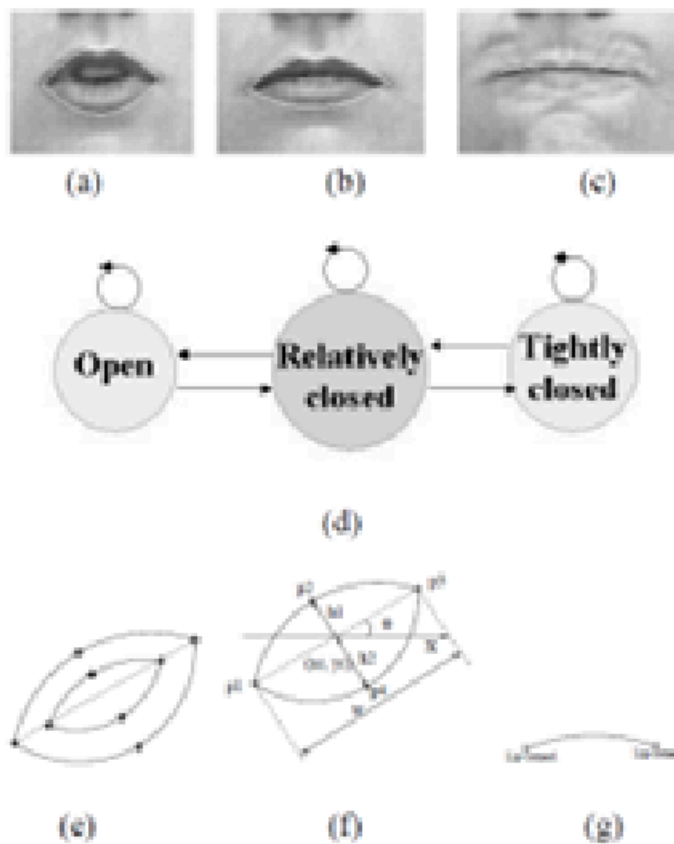
2.
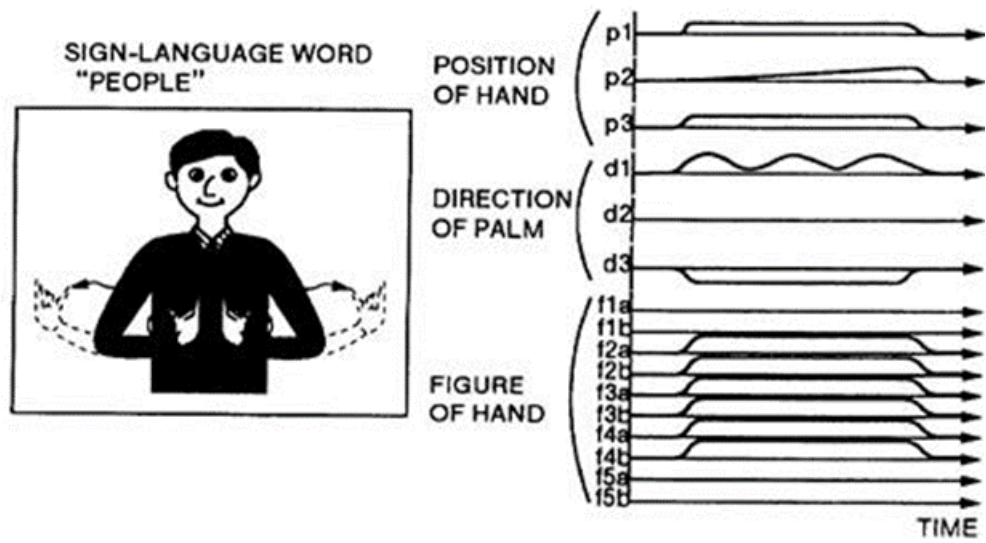
Recognising deaf-sign language by movement analysis.

3.

Recognising human activities, e.g. walking.

4.

Recognising lips states and their motion.

References

Limiting references

This place does not cover:

Facial expression recognition	G06V 40/16

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition of scenes; Scene-specific elements	G06V 20/00
Recognition of human or animal bodies within image or video data	G06V 40/10
Recognition of fingerprints or palmprints within image or video data	G06V 40/12
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Input arrangements or combined input and output arrangements for interaction between user and computer	G06F 3/01
Analysis of motion in images	G06T 7/20
Speech recognition using position of the lips, movement of the lips or face analysis	G10L 15/25

Special rules of classification

Recognising activities for scenes under surveillance (e.g. suspicious activities, occupancy, etc.) is classified in group G06V 20/52.

Static gesture recognition, e.g. recognition of deaf signs is classified in group G06V 40/10.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

dynamic gesture	movement of the hand encoding a certain meaning.
gait recognition	recognising a person's manner of walking.
static gesture	posture of a hand denoting a certain meaning.

G06V 40/30

Writer recognition; Reading and verifying signatures

Definition statement

This place covers:

Acquisition, pre-processing, feature extraction and classification of handwritten signatures and handwritten text input to identify the writer.

The processing may be based on a bit map image showing the signature (called static or offline signature recognition) or a signal representing the position, velocity, acceleration or pressure of the writing tip (called dynamic or online signature recognition).
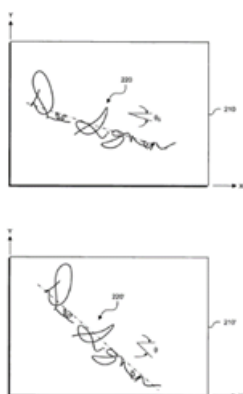
Illustrative examples of subject matter classified in this place:
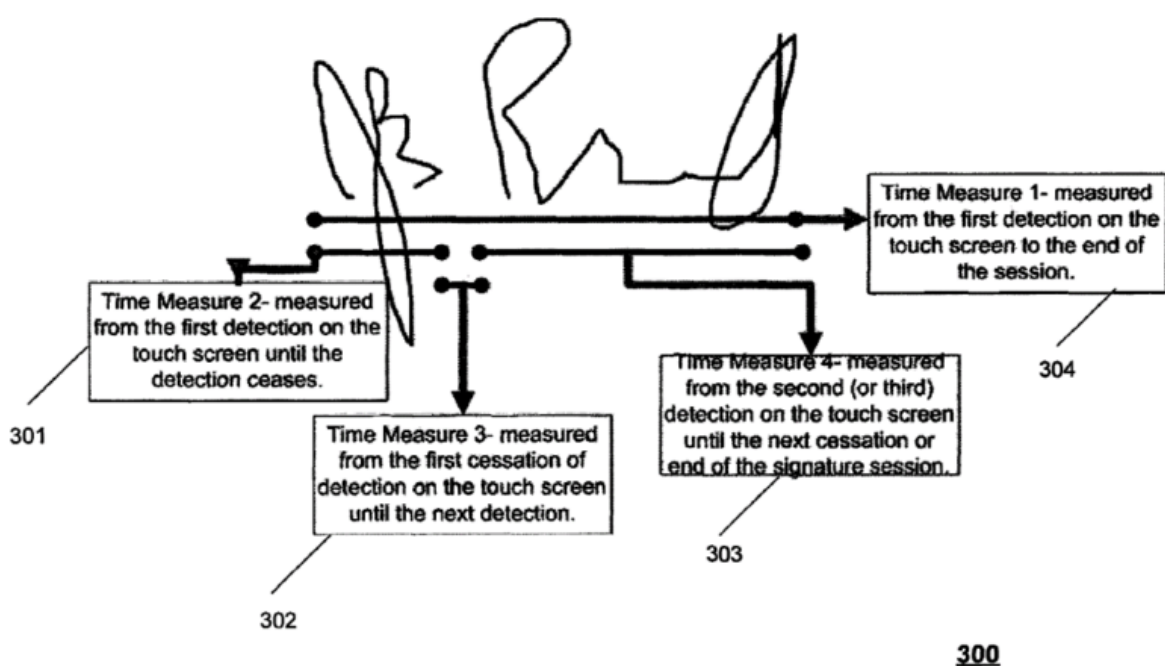
1.

Handwriting input using a grid defined on the screen of a mobile phone.

2.

Transforming a signature to a consistent angle of inclination for recognition purposes.

3.

Temporal analysis of a pen stroke for signature encoding.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Image acquisition for image or video recognition or understanding	G06V 10/10
Image preprocessing for image or video recognition or understanding	G06V 10/20
Arrangements for image or video recognition using probabilistic graphical models, e.g. Markov models or Bayesian networks	G06V 10/84
Image-based acquisition using hand-held instruments for character recognition; Constructional details of the instruments	G06V 30/142

Character recognition of cursive writing	G06V 30/226
Character recognition; Recognition of three-dimensional handwriting, e.g. writing in the air	G06V 30/228
Recognising digital ink, i.e. recognising handwritten individual characters or symbols represented by temporal sequences of position coordinates	G06V 30/32
Image-based pattern recognition of technical drawings or geographical maps	G06V 30/422
Input arrangements for converting the position or the displacement of a member into a coded form	G06F 3/03
Interaction techniques based on a graphical user interface; Using specific features provided by the device; Entering handwritten data, e.g. gestures, text	G06F 3/04883
Security arrangements for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/00
Security arrangements in wireless communication networks, e.g. access security or fraud detection; Authentication, e.g. verifying user identity or authorisation; Protecting privacy or anonymity	H04W 12/00

Special rules of classification

Details about the temporal aspects in the acquisition, preprocessing, feature extraction or recognition of the digital ink are classified in group G06V 30/32.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

off-line signature	analysis of a (static) image characterising the signature
on-line signature	analysis of a temporal sequence of position, velocity, acceleration or pressure values characterising the signature

G06V 40/40

Spoof detection, e.g. liveness detection

Definition statement

This place covers:

Spoof detection, i.e. detecting an attempt to fool a biometric system by presenting data which is not genuine. An example is the detection of inanimate replicas of living tissue, and the distinguishing of such replicas, e.g. a rubber model of a finger, from parts of living beings.

Spoof detection can be performed using acquisition arrangements in which the sensor is provided with specialised hardware to assess or highlight the genuineness of the acquired data (e.g. using special illumination in infrared) or by performing image processing operations (e.g. colour analysis to discriminate the genuine skin against a copy). Multiple biometric modalities can be involved:

- Signals such as blood pressure, pulse and perspiration at the fingertips, hippus movement of the pupil, brain waves [EEG] and electrical heart signals [ECG] in combination with other biometric images;
- Reflexive signals such as pupillary light reflex (pupil dilation), corneal reflex (blink reflex) and patellar reflex (knee-jerk);
- Voluntary signals given unconsciously or as a response to a "challenge" such as blinking, mouth movements and facial expressions.
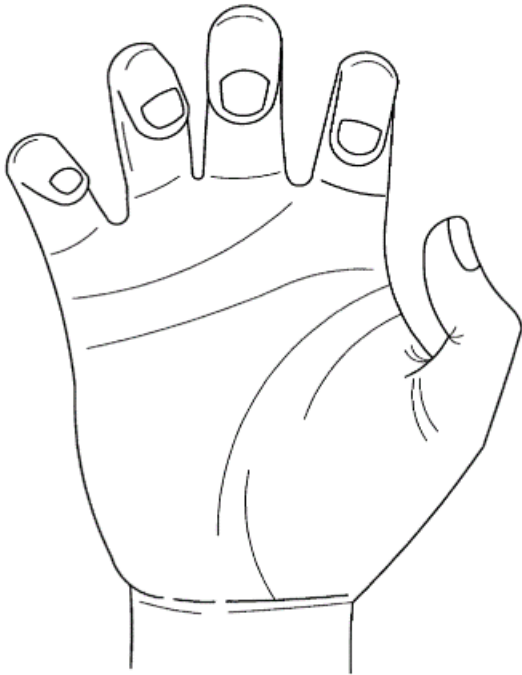
Other properties of a body can be assessed:
- Determination of the flatness of a face to detect use of a picture to challenge a biometric system ("spoof-by-picture");
- Light distribution in a real finger which differs from a fake finger.

Illustrative examples of subject matter classified in this place:
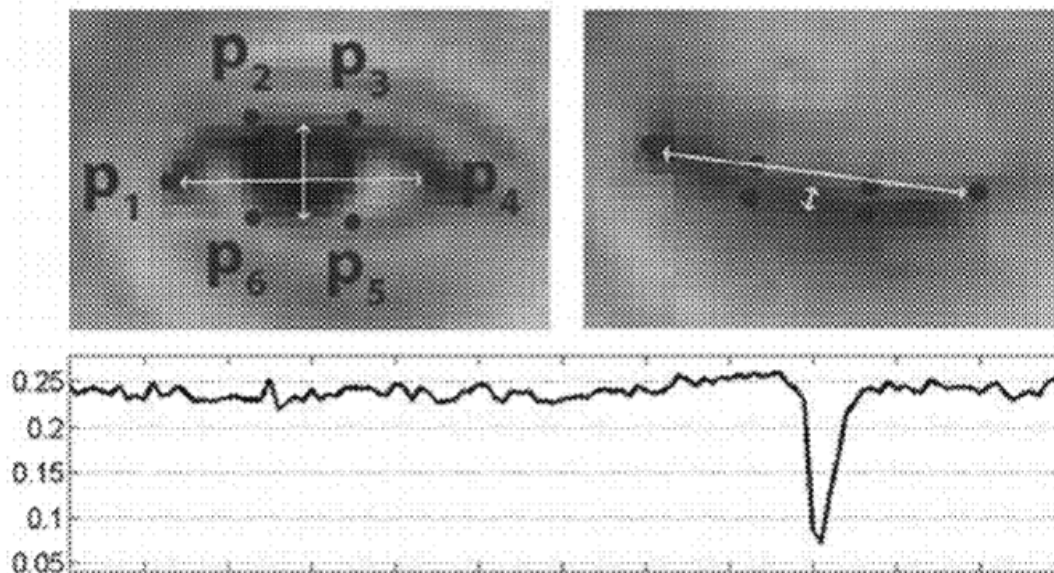
1A.

1B.

Recognition method using hand biometrics with anti-counterfeiting. The user is asked to perform randomly selected gestures with the hand, e.g. rotate the hand to the left, clench into a fist. The

gestures are recognised, allowing the method to determine that a real user is standing in from of the camera.

2.

When the eye opens, the eye aspect ratio is roughly constant, and only fluctuates around the range 0.25. Once the eye blinks and closes, because the vertical distance is almost zero, the eye aspect ratio is correspondingly reduced to zero. When the eye opens again, the eye aspect ratio rises to the range 0.25 again. These measurements may indicate if the person is a real, genuine person or a fake.

Relationships with other classification places

Authentication of users for computer access is classified in group G06F 21/32.

Authentication of financial documents is classified in group G07D 7/00.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Detection or correction or errors, e.g. by rescanning the pattern; Evaluation of the quantity of an acquired biometric pattern	G06V 10/98
Recognition of fingerprints or palmprints in images or video	G06V 40/12
Recognition of vascular patterns in images or video	G06V 40/14
Recognition of human faces in images or video, e.g. facial parts, sketches or expressions	G06V 40/16
Recognition of eye characteristics in images or video, e.g. of the iris	G06V 40/18
Recognition of movements or behaviour in images or video, e.g. gesture recognition	G06V 40/20
Recognition of signatures	G06V 40/30
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70
Fittings or systems for preventing or indicating unauthorised use or theft of vehicles	B60R 25/00

Digitisers as the input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
User authentication using biometric data for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/32
User authentication by graphic or iconic representation for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/36
Checking-devices for individual entry or exit registers	G07C 9/00
Testing specially adapted to determine the identity or genuineness of valuable papers	G07D 7/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00
Means for preventing unauthorised calls from a telephone set	H04M 1/667

Special rules of classification

This group is used alone when no technical contribution can be identified in the processing associated with biometric authentication. If, however, a technical contribution can be identified in biometric authentication, the respective groups are allocated in combination with this group. In other words, anti-spoofing is usually a part of an authentication process which acts as a verifier of liveliness, thus anti-spoofing inventions often rely on processing biometric data of a certain modality provided in the following groups:

- Fingerprints or palmprints – group G06V 40/12;
- Human faces, e.g. facial parts, sketches or expressions – group G06V 40/16;
- Eyes – group G06V 40/18;
- Movements or behaviour, e.g. gesture recognition – group G06V 40/20;
- Signatures – group G06V 40/30.

For example, in order to assure safe biometric authentication, the face matching process classified in group G06V 40/16 combined with a liveness detection, e.g. by determining if the user in front of the camera is moving his mouth when requested so that a spoof-by-picture attack can be prevented, is classified also in group G06V 40/40.

G06V 40/50

Maintenance of biometric data or enrolment thereof

## Definition statement

This place covers:

Maintenance of biometric data which includes, e.g., enrolment of a user using biometric information or updating the biometric information stored in a database for each user.

The enrolment process may include the decision which of the plurality of templates should be stored and used for future authentication of the user.

Typical examples include replacement of enrolment data with the more recent one using temporal criteria (e.g. compensating for the aging of the person) or using quality criteria (e.g. a reference with higher quality has become available during system use).
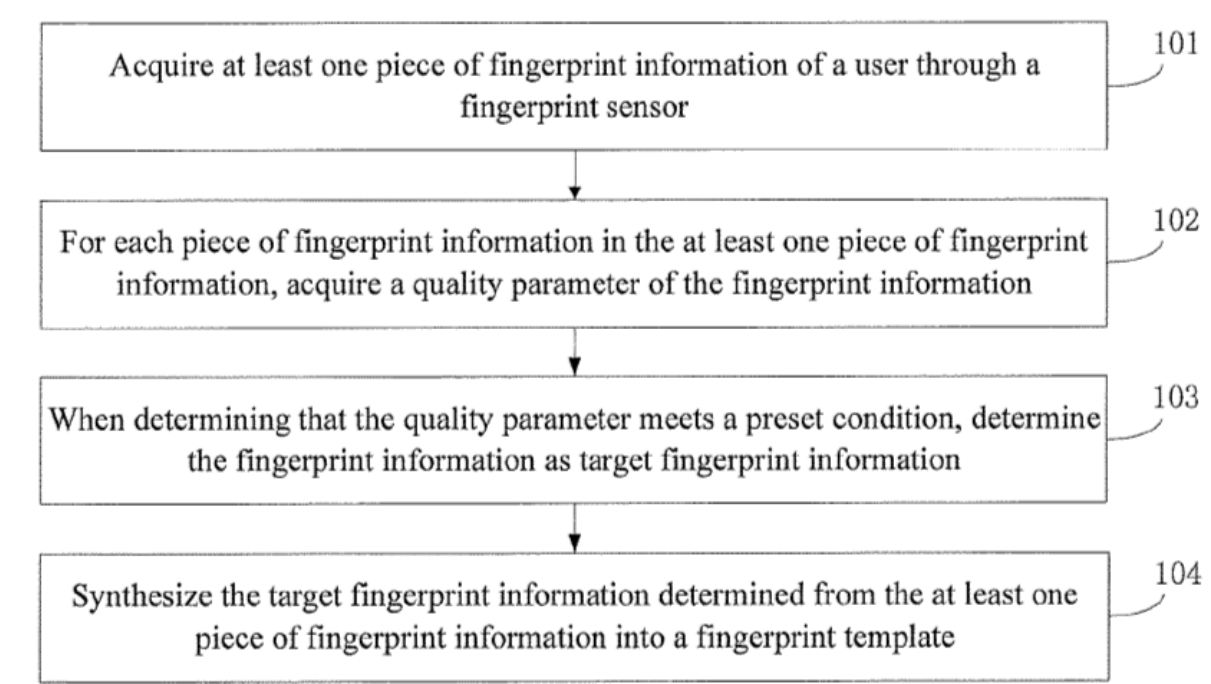
Illustrative examples of subject matter classified in this place:

1.
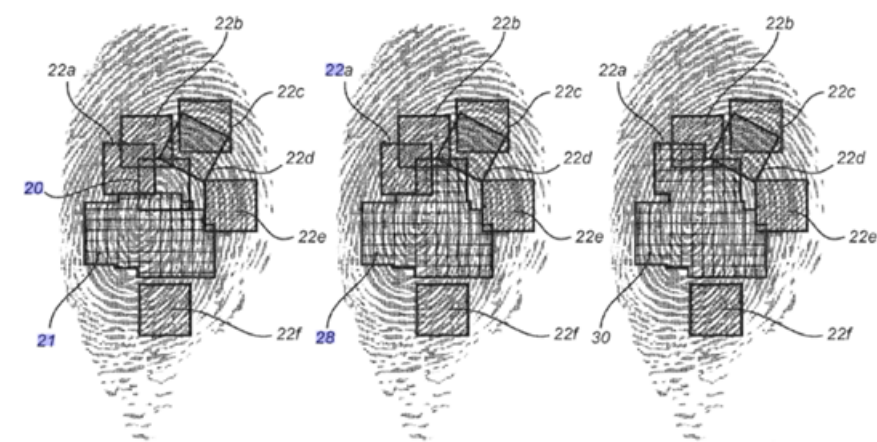
Fingerprint authentication with template updating.

2.

Fingerprint template update based on its quality.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Detection or correction or errors, e.g. by rescanning the pattern; Evaluation of the quality of an acquired biometric pattern	G06V 10/98
Recognition of fingerprints or palmprints in images or video	G06V 40/12
Recognition of vascular patterns in images or video	G06V 40/14

Recognition of human faces in images or video, e.g. facial parts, sketches or expressions	G06V 40/16
Recognition of eye characteristics within images or video, e.g. of the iris	G06V 40/18
Recognition of movements or behaviour in images or video, e.g. gesture recognition	G06V 40/20
Recognition of signatures	G06V 40/30
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70
Fittings or systems for preventing or indicating unauthorised use or theft of vehicles	B60R 25/00
Digitisers as the input arrangement for user-computer interaction, e.g. touch screens or touch pads	G06F 3/041
User authentication using biometric data for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/32
User authentication by graphic or iconic representation for protecting computers, components thereof, programs or data against unauthorised activity	G06F 21/36
Checking-devices for individual entry or exit registers	G07C 9/00
Testing specially adapted to determine the identity or genuineness of valuable papers	G07D 7/00
Arrangements for secret or secure communications; Network security protocols	H04L 9/00
Means for preventing unauthorised calls from a telephone set	H04M 1/667

Special rules of classification

This group is used alone when no technical contribution can be identified in the processing associated with authentication. If, however, a technical contribution can be identified in user authentication, the respective groups are allocated in combination with this group. In other words, the maintenance of the biometric information improves the authentication process by updating the templates stored or initially storing templates during enrolment which ensure a certain quality, thus such inventions often rely on the following groups:

- Fingerprints or palmprints – group G06V 40/12;
- Human faces, e.g. facial parts, sketches or expressions – group G06V 40/16;
- Eyes – group G06V 40/18;
- Movements or behaviour, e.g. gesture recognition – group G06V 40/20;
- Signatures – group G06V 40/30.

If the maintenance or enrolment involves quality-based criteria, classification in groups G06V 10/98 and G06V 40/50 is applied.

G06V 40/60

Static or dynamic means for assisting the user to position a body part for biometric acquisition

### Definition statement

This place covers:

Means for assisting the user to position body parts such as face, eye(s), hand(s) for the purpose of biometric identification using either static means, e.g. a finger guide for fingerprint acquisition, or dynamic means, e. g. a visual indication on an interactive screen.

Illustrative examples of subject matter classified in this place:
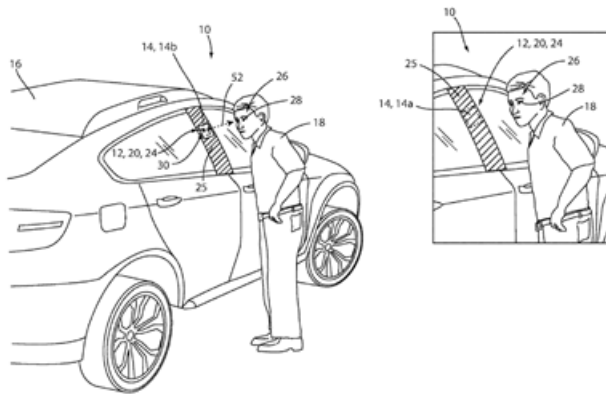
1A.
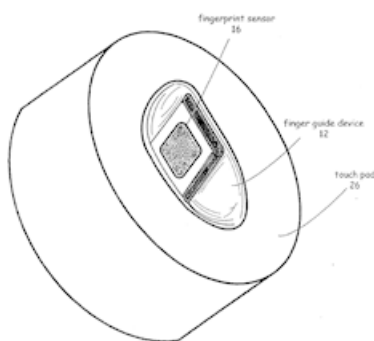
1B.
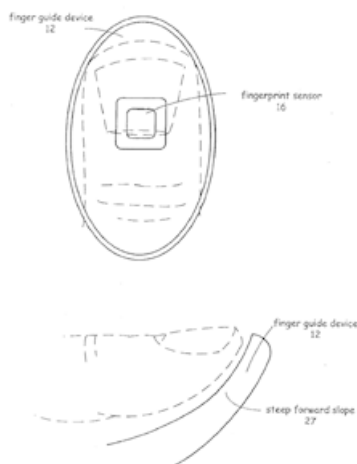
Special cradle used for finger positioning to allow reproducibility during subsequent acquisitions.

2.

Determination of the liveness of a person by giving him visual feedback on a screen provided on the side of the car and giving directions to move the face in a certain way.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition of human or animal bodies in images or video, e.g. vehicle occupants or pedestrians; Body parts, e.g. hands	G06V 40/10
Recognition of fingerprints or palmprints in images or video	G06V 40/12
Recognition of human faces in images or video, e.g. facial parts, sketches or expressions	G06V 40/16
Recognition of eye characteristics in images or video, e.g. of the iris	G06V 40/18
Multimodal biometrics, e.g. combining information from different biometric modalities	G06V 40/70
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Input arrangements or combined input and output arrangements for interaction between user and computer	G06F 3/01

Special rules of classification

When fingerprint acquisition is performed by requiring the user to place the finger in a recess specially provided to guide the acquisition, classification in groups G06V 40/60 and G06V 40/13 is applied. Similarly, when the face acquisition is guided by user feedback, classification in groups G06V 40/16 and G06V 40/60 is done.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

fingerprint guide	mechanical component specially provided to guide the placement of the finger during fingerprint acquisition
visual feedback	visual information indicating the position of a body part during image acquisition

G06V 40/70

Multimodal biometrics, e.g. combining information from different biometric modalities

Definition statement

This place covers:

Biometric identification and authentication using multiple modalities at the same time, e.g. fingerprint and face, iris and face, etc. The image-based biometric modalities can be combined with non-image based modalities, such as voice or physiological measurements (heart rate).
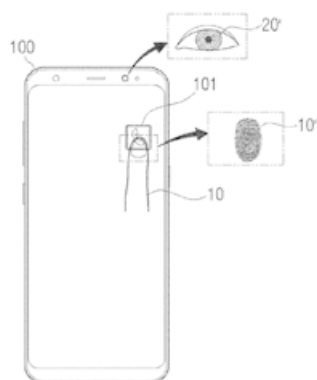
Illustrative examples of subject matter classified in this place:

1.

Multiple biometric modalities are encoded in a database and used for personal identification.

2.

Multiple biometric modalities analysis on a smartphone.

References

Informative references

Attention is drawn to the following places, which may be of interest for search:

Recognition of fingerprints or palmprints in images or video	G06V 40/12
Recognition of vascular patterns in images or video	G06V 40/14
Recognition of human faces in images or video	G06V 40/16

Recognition of eye characteristics in images or video, e.g. of the iris	G06V 40/18
Recognition of movement or behaviour in images or video	G06V 40/20
Writer recognition; Reading and verifying signatures	G06V 40/30
Identification of persons	A61B 5/00
Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Input arrangements or combined input and output arrangements for interaction between user and computer	G06F 3/01
Speech recognition	G10L 15/00
Speaker identification or verification	G10L 17/00

Special rules of classification

If the fusion between the different biometric modalities is performed, classification in groups G06V 10/80 and G06V 40/70 is applied.

Glossary of terms

In this place, the following terms or expressions are used with the meaning indicated:

multimodal biometrics	using multiple biometric traits for identification or authentication