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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 May 2019 (31.05.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/101720 A1

(51) International Patent Classification:
G06K 9/00 (2006.01)

(21) International Application Number:
PCT/EP2018/081874

(22) International Filing Date:
20 November 2018 (20.11.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10 2017 127 592.4 22 November 2017 (22.11.2017) DE

(71) Applicant: CONNAUGHT ELECTRONICS LTD. [IE/IE]; Dunmore Road, Tuam, County Galway (IE).

(72) Inventors: FERNANDEZ, Joseph; c/o Connaught Electronics Ltd., Dunmore Road, Tuam, County Galway (IE). ENRIGHT, Catherine; c/o Connaught Electronics Ltd., Dunmore Road, Tuam, County Galway (IE).

(74) Agent: JAUREGUI URBAHN, Kristian; Laiernstr. 12, 74321 Bietigheim-Bissingen (DE).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: METHODS FOR SCENE CLASSIFICATION OF AN IMAGE IN A DRIVING SUPPORT SYSTEM

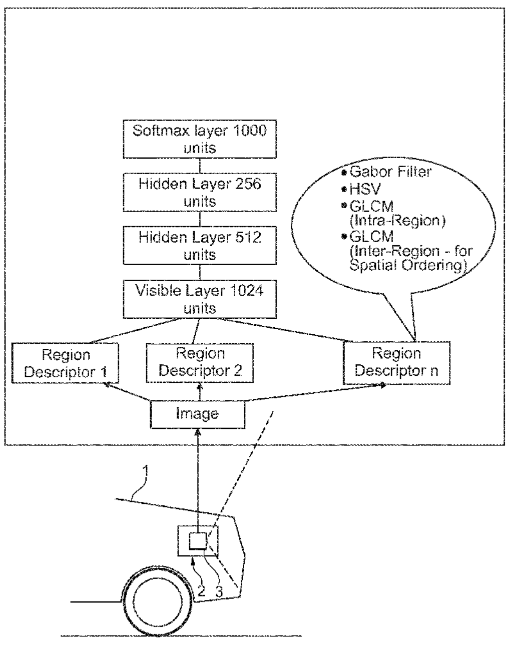

Fig. 2

(57) Abstract: The invention relates to a method for scene classification of an image in image processing in a driving support system (2) of an automotive vehicle (1) comprising the steps of spatially ordering regions of the image by clustering the image pixels into regions with high inter-class variance and low intra-class variance, modelling the underlying joint probability distribution of each scene category using a Deep Boltzmann Machine (DBM) as a generative model comprising a stack of Restricted Boltzmann Machines (RBMs) and a classifier such as a Radial Basis Filter Support Vector Machine (RBF-SVM), a second CNN or simply a softmax layer as the final layer, and classifying the scene based on the learning result of the Deep Boltzmann Machine (DBM).The invention further relates to a method for scene classification of an image in image processing in a driving support system (2) of an automotive vehicle (1) comprising the steps of providing a convolutional neuronal network (CNN) comprising a plurality of layers to learn what features of the image are most useful for the classification of scenes, wherein multiple resolutions of features are generated for capturing detail of features at higher resolution and "big picture" at lower resolution, modelling the joint probability distribution of each scene category using a Deep Boltzmann Machine (DBM) comprising a stack of Restricted Boltzmann Machines (RBMs) and a classifier such as a Radial Basis Filter Support Vector Machine (RBF-SVM), a second CNN or simply a softmax layer as the final layer, wherein the output of each layer of the convolutional neuronal network (CNN) is fed into the visible layer of the Deep Boltzmann Machine (DBM) as a separate input, followed by the stack of Restricted Boltzmann Machines (RBM) learning the underlying joint probability distribution of each scene category, and classifying the scene based on the learning result of the Deep Boltzmann Machine (DBM).

WO 2019/101720 A1

[Continued on next page]

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

Methods for scene classification of an image in a driving support system

The invention relates to a method for scene classification of an image in image processing in a driving support system of an automotive vehicle.

The invention further relates to a method for scene classification of an image in image processing in a driving support system of an automotive vehicle.

Driving support systems such as driver assistance systems are systems developed to automate, adapt and enhance vehicle systems for safety and better driving. Safety features are designed to avoid collisions and accidents by offering technologies that alert the driver to potential problems, or to avoid collisions by implementing safeguards and taking over control of the vehicle. In autonomous vehicles, the driving support systems provide input to perform a control of the vehicle. Adaptive features may automate lighting, provide adaptive cruise control, automate braking, incorporate traffic warnings, connect to smartphones, alert e.g. the driver in respect to other cars or different types of dangers, keep the vehicle in the correct lane, or show what is located in blind spots. Driving support systems including the aforementioned driver assistance systems, frequently rely on inputs from multiple data sources such as automotive imaging, image processing, radar sensors, LiDAR, ultrasonic sensors and other sources. Neural networks have recently been implicated in processing such inputs of data within driver assistance systems, or in general in driving support systems.

In recent times, there has been a surge of research on Deep Boltzmann Machines (DBMs) and convolution neural networks (CNNs). Their design has been aided by increase in computational power in computer architectures and the availability of large annotated datasets.

A Deep Boltzmann Machine (DBM) is a stochastic Hopfield net with hidden layers. A Hopfield net is an energy-based model. Whereas the Hopfield net is used as a content-addressable memory system, the Boltzmann Machine learns to represent its inputs. It is a generative model, which means that it learns the joint probability distribution of all its inputs. Once the Boltzmann Machine has learned its input (i.e. when it has reached thermal equilibrium), the configuration of weights at the (multiple) hidden layers constitute a representation of the inputs presented at the visible layer. RBMs are Restricted Boltzmann

Machines, wherein the restriction is that the neurons form a bipartite graph with no intra-layer connections. This restriction allows the use of the highly efficient Contrastive Divergence algorithm for training. A Deep Boltzmann Machine (DBM) is a stack of RBMs. A DBN (Deep Belief Net) also contains RBMs, but they have RBMs only in the top two layers and the layers below are sigmoid belief nets which are directed graphical models. In contrast, the DBM is an undirected graphical model through and through.

Convolutional neuronal networks (CNNs) are highly successful at classification and categorization tasks, but much of the research is on standard photometric RGB images and is not focused on embedded automotive devices. Automotive hardware devices need to have low power consumption requirements and thus low computational power.

In machine learning, a convolutional neural network is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. Convolutional networks were inspired by biological processes, in which the connectivity pattern between neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage. CNNs have applications in image and video recognition, recommender systems and natural language processing.

For the methods described herein, scene classification may be performed e.g. based on distinguishing between one or all of the following three categories:

a. Types of scenes
    i. Country-side
    ii. City
    iii. Parking lot in the open
    iv. Parking lot in the basement of a mall
b. Weather conditions
    i. Snow
    ii. Sunny

c. Density of scenes

i. Sparse

ii. Dense / Busy Scene.

The above classification can be used by a layer that runs across all algorithms in a computer vision product. The classification can thus be used:

a) to determine the activation logic of an algorithm variant. For instance3 Dimensional Object Detection (3DOD) can have an algorithm for sparsely populated scenes, which runs most times and consumes fewer resources (CPU, Memory), and can further have an intensive variant for densely populated scenes. So, if the “master / supervisor algorithm” knows that the scene is densely or sparsely populated, it can activate the corresponding variant of the 3DOD algorithm.

b) In addition, snow and rain are well-known difficult conditions for Computer Vision algorithms. It is made all the more difficult for the algorithm, because it has to have the same configuration and learning parameters for sunny weather as for snowy conditions. However, if the supervisor algorithm knows that the scene is snowy, raining or sunny, it can activate different variants of 3DOD, Pedestrian Detection (PD), Parking Slot Marker Detection (PSMD) and so on, while each of those variants only learns to handle one weather condition.

c) Similarly, for a Parking Slot Marker Detection algorithm, parking slots in the open are very different from parking slots in an underground basement with artificial lights all around. With a supervisor algorithm and a Scene Classification algorithm to guide it, the PSMD algorithm then only needs to learn one scenario and learn it well.

In this respect, US 2007/0282506 A1 discloses a method for image processing for vehicular applications that determines edges of objects in images and feeds these data to a neural network that can provide a classification, identification and/or location of an object. The method comprises the steps of obtaining information about objects in an environment in or around a vehicle comprising training pattern recognition algorithm, such as a neural network, to provide information about objects in the environment upon receiving as input information about edges of unknown objects, installing the pattern recognition algorithm in a processor on the vehicle, operatively obtaining images of the environment deriving data about edges of objects in the obtained images and providing the data to the pattern recognition algorithm in the processor to receive as output, information about the object, such as a classification, identification and/or location of an object.

US 2008/0144944 A1 discloses a method for obtaining information about an occupying item in a vehicle, such as a human being, comprising obtaining images of an area above a seat in the vehicle in which the occupying item is situated and classifying the occupying item by inputting signals derived from the images into a trained neural network form which is trained to output an indication of the class of occupying item from one of a predetermined number of possible classes. The images may be pre-processed to remove background portions of the images and then converted into signals for input into the neural network form.

Driving support systems like driver assistance systems are one of the fastest-growing segments in automotive electronics and there is a need for improved methods and systems for image processing in driving support assistance systems.

It is an objective of the present invention to provide methods for classifying scenes in driving support systems more accurately than current methods and classifying scenes in a better way in order to eliminate hand-crafted features being used as inputs.

This objective is addressed by the subject-matter of the independent claims. Preferred embodiments are described in the sub claims.

The invention provides a method for scene classification of an image in image processing in a driving support system of an automotive vehicle comprising the steps of:

- spatially ordering regions of the image by clustering the image pixels into regions with high inter-class variance and low intra-class variance,
- modelling the underlying joint probability distribution of each scene category using a Deep Boltzmann Machine (DBM) as a generative model comprising a stack of Restricted Boltzmann Machines (RBM) and a classifier such as a Radial Basis Filter Support Vector Machine (RBF-SVM), a second CNN or simply a softmax layer as the final layer, and
- classifying the scene based on the learning result of the Deep Boltzmann Machine (DBM).

Thus, it is an essential idea of this embodiment of the invention to combine the following three main steps in a unique manner: spatially ordering regions of the image, modelling the underlying joint probability distribution with a generative model, i.e. the Deep Boltzmann Machine (DBM), and then classifying scenes based on that generative model. One advantage of the invention is no loss of region order (i.e. discarding improbable explanations for a scene), which contains a lot of information useful for classifying scenes, e.g. “sky above ground", "road below sky" and "tree top above road". The human brain uses such information to understand a scene and to discard the improbable explanations for what is being seen. In addition, the invention uses an unsupervised, generative model, i.e. the Deep Boltzmann Machine (DBM), which provides the advantage of reducing the amount of labeled data necessary. Labeled data are only used to fine-tune the DBM. Thus, very little annotated data is required, which will significantly reduce cost and reduce effort of annotation. A further advantage of the invention making use of a DBM is that the method comprises more applicability to a broader range of tasks, which means that one does not have to go through the expensive step of annotating images for applying this to another task, e.g. segmentation. In addition, there are a lot more transformations (e.g. lighting, perspective, and occlusion) of scenes available than there are annotated data at hand. Given that fact, a generative model, i.e. the DBM, is more likely to give a better classification. Furthermore, using an unsupervised, generative method provides the further advantage of providing a much better and more expansive representation, as specifically with respect to scene classification, there are numerous combinations that can make up the same scene, e.g. i) a scene can be a motorway at night, which can be composed of a lot of combinations of regions and of features, ii) there is also a lot of overlap among various kinds of scenes and these overlaps are not as well captured through a human-made annotation as they are through learning the underlying probability distributions of various scenes, iii) the annotated data at hand for scene classification is not likely to have enough representation.

Preferably, the spatially ordering regions of the image comprises using one or more region descriptors to capture a semantically uncorrelated low-level representation of each region based on the features i) Gabor filter, ii) Hue, Saturation and Value (HSV) color space features, and iii) Co-occurrence features that are Haralick features derived from the Gray Level Co-occurrence Matrix (GLCM).

The captured low-level representation of each region is semantically uncorrelated in that the above used features i), ii) and iii) are mutually uncorrelated in a semantic sense. The i) Gabor filter is a linear filter used for texture analysis, which analyses whether there is any specific frequency content in the image in specific directions in a localized region around the point or region of analysis. Frequency and orientation representations of Gabor filters are proven to be a feature used by the human vision. The ii) HSV color space is a color space that defines the localization of a color by the features Hue, Saturation and Value.

The iii) Co-occurrence features can be used for intra-region co-occurrence statistics (mean and range), wherein said features are selected from the group consisting of Angular Second Moment, Contrast, Sum Average, Sum Variance and Difference Variance.

The driving support systems include driver assistance systems are systems, which are already known and used in state of the Art vehicles. The developed driving support systems are provided to automate, adapt and enhance vehicle systems for safety and better driving. Safety features are designed to avoid collisions and accidents by offering technologies that alert the driver to potential problems, or to avoid collisions by implementing safeguards and taking over control of the vehicle. In autonomous vehicles, the driving support systems provide input to perform a control of the vehicle. Adaptive features may automate lighting, provide adaptive cruise control, automate braking, incorporate traffic warnings, connect to smartphones, alert e.g. the driver in respect to other cars or different types of dangers, keep the vehicle in the correct lane, or show what is located in blind spots. Driving support systems including the aforementioned driver assistance systems, frequently rely on inputs from multiple data sources such as automotive imaging, image processing, radar sensors, LiDAR, ultrasonic sensors and other sources.

Further, according to a preferred embodiment of the invention, spatially ordering regions of the image further comprises adding spatial relationships among neighboring regions to create a Spatially Ordered Region Descriptor (SORD), wherein further Haralick features are used that are selected from the group consisting of the mean and range of iv) Correlation, v) Entropy, vi) Sum Variance, vii) Difference Variance and viii) Information Measures of Correlation.

Preferably, the Spatially Ordered Region Descriptor (SORD) comprises the features i) Gabor filter, ii) Hue, Saturation and Value (HSV) color space features, iii) Co-occurrence features, iv) Correlation, v) Entropy, vi) Sum Variance, vii) Difference Variance and viii) Information Measures of Correlation. That means that the aforementioned eight features are added to the Region Descriptor to convert it into the SORD.

The Spatially Ordered Region Descriptor (SORD) can be the input to the neurons of the visible layer of the Deep Boltzmann Machine (DBM) when modelling the underlying joint probability distribution of each scene category followed by processing through a stack of Restricted Boltzmann Machines (RBMs) that learn the underlying joint probability distributions of each scene category.

Preferably, classifying the scene is done by a softmax layer on top of the stack of Restricted Boltzmann Machines (RBMs). That means that the DBM preferably is topped off with a softmax layer in order to perform the classification of the scene. Once the stack of RBMs has learned the underlying joint probability distributions of each scene category, the softmax layer added on top does the actual classification. In place of the softmax layer, a RBF-SVM or a CNN can also be used for the classification. This final classification layer requires a far smaller training set because its weights are initialized by the output of the DBM and hence require only fine-tuning for the specific scene categories in the intended application.

The invention further provides a method for scene classification of an image in image processing in a driving support system of an automotive vehicle comprising the steps of:

- providing a convolutional neuronal network (CNN) comprising a plurality of layers to learn what features of the image are most useful for the classification of scenes, wherein multiple resolutions of features are generated for capturing detail of features at higher resolution and “big picture” at lower resolution,
- modelling the joint probability distribution of each scene category using a Deep Boltzmann Machine (DBM) comprising a stack of Restricted Boltzmann Machines (RBM) and a classifier such as a Radial Basis Filter Support Vector Machine (RBF-SVM), a second CNN or simply a softmax layer as the final layer, wherein the output of each layer of the convolutional neuronal network (CNN) is fed into the visible layer of the Deep Boltzmann Machine (DBM) as a separate input, followed by the stack of Restricted Boltzmann Machines (RBM) learning the underlying joint probability distribution of each scene category, and
- classifying the scene based on the learning result of the Deep Boltzmann Machine (DBM).

The combination of a CNN with a DBM in a method for scene classification of an image in image processing in a driving support system of an automotive vehicle is unique and provides several advantages. It is a particular advantage of this embodiment of the invention that features used as input to the DBM are not hand-crafted but discovered by the CNN as the most useful features, which significantly reduces cost and efforts. In addition, this embodiment of the invention uses a single CNN at multiple resolutions. The features at higher resolution capture detail, whereas the features learned at lower resolution capture the “big picture”, i.e. the region-level and scene-level information in the image. The single CNN provides these features at multiple resolutions or scales. Since the lower resolution images capture the “big picture”, a CNN trained at that resolution and performing inference at that low resolution will provide the region-level and scene-level features to the DBM, which then performs the scene classification in its softmax layer that is added on top.

In addition, the DBM has advantages over a DBN (Deep Belief Network) in that the DBN is a DAG (Directed Acyclic Graphical model), whereas the DBM is an undirected graphical model. Unlike DBNs, the approximate inference procedure in DBMs, in addition to an initial bottom-up pass, can incorporate top-down feedback, allowing DBMs to better propagate uncertainty about and hence deal more robustly with ambiguous inputs. Also, through greedy layer-wise pre-training, this embodiment of the invention can achieve fast approximate inference in DBMs. That is, given a data vector on the visible units, each layer of hidden units can be activated in a single bottom-up pass by doubling the bottom-up input to compensate for the lack of top-down feedback (except for the very top layer, which does not have a top-down input). This fast approximate inference is used to initialize the mean-field method, which converges much faster than with random initialization.

In this embodiment of the invention scene classification is realized by using a CNN to learn what features are most useful for the classification of scenes of an image, wherein a CNN of multiple resolutions is used for capturing detail and “big picture” as separate inputs followed by modeling the joint probability distribution of each scene category using a DBM, which is then followed by classifying the scenes based on the learning result of the DBM that has been generated based on the inputs of the separate layers of the CNN.

Thus, this embodiment of the invention can be described as a Hybrid CNN-DBM model, in which a DBM learns the inter-relationships among regions and multi-resolution features of the same region with an unsupervised, generative method based on the separate inputs of the various layers of the CNN. That means that the DBM learns an internal representation of the scene category using features at multiple resolutions extracted from the image that are fed by each layer of the CNN to the visible layer of the DBM as separate inputs. In other words, each layer’s output of the CNN is fed separately to the visible layer of the DBM, which is then followed by the stack of Restricted Boltzmann Machines (RBMs) learning the underlying joint probability distribution of each scene category followed by classifying the scene based on the learning result of the DBM. This architecture of this Hybrid CNN-DBM model of the invention allows the DBM to not only learn the inter-relationships among regions in a scene, but also to learn the inter-relationships among multi-resolution features of the same regions. This is a major advantage of this embodiment of the invention over purely discriminative networks.

Thus, in this Hybrid CNN-DBM model of the invention, the primary modeling of scene categories is done by an unsupervised, generative method, i.e. the DBM, thus

a. naturally providing a representation that incorporates inter-feature relationship. The DBM learns the joint probability distribution of all its inputs. The DBM inputs are composed of the discriminative features learned at various abstraction levels. So, the DBM not only learns the inter-relationships among regions in a scene, but also the inter-relationships among the multi-resolution features of the same region;

b. providing a much better, more expansive representation as specifically with respect to scene classification, there are numerous combinations that can make up the same scene;

i. for example, the scene can be a motorway at night, which can be composed of a lot of combinations of regions and of features;

ii. there is also a lot of overlap among various kinds of scenes and these overlaps are not as well captured through a human annotation as they are through learning the underlying probability distribution of various scenes.

Preferably, the convolutional neuronal network (CNN) is pre-trained using supervised training and labeled data, wherein classifying the scene is done by a temporary softmax layer as the final layer of the Deep Boltzmann Machine (DBM), wherein the temporary softmax layer is stripped off once the convolutional neuronal network (CNN) has learned the features, followed by feeding the output of each layer of the convolutional neuronal network (CNN) into the visible layer of the Deep Boltzmann Machine (DBM) as a separate input.

Further, according to a preferred embodiment of the invention, the Deep Boltzmann Machine (DBM) is pre-trained using greedy layer-wise pre-training to learn the internal representation of the combination of multiple features in a scene and multi-resolution features of the same region, followed by adding on and pre-training the softmax layer using labelled data.

Preferably, classifying the scene is done by a softmax layer on top of the stack of Restricted Boltzmann Machines (RBMs). In other words, the DBM preferably is topped off with a softmax layer, where classifying the scene based on the learning result of the DBM is performed, once the stack of RBMs has learned the underlying joint probability distributions of each scene category. In place of the softmax layer, a RBF-SVM or a CNN can also be used for the classification. This final classification layer requires a far smaller training set because its weights are initialized by the output of the DBM and hence require only fine-tuning for the specific scene categories in the intended application.

The invention also provides the use of the methods described herein in a driving support system of an automotive vehicle. Specifically, the invention provides the use of the methods for scene classification of an image in image processing in a driving support system of an automotive vehicle described above.

The invention further provides a driving support system for an automotive vehicle comprising a camera for providing images for classification, wherein the driving support system is configured for performing the methods described herein.

The invention further provides a non-transitory computer-readable medium, comprising instructions stored thereon, that when executed on a processor, induce a driving support system to perform the methods described herein.

The invention also provides an automotive vehicle comprising:

a data processing apparatus,

a non-transitory computer-readable medium comprising instructions stored thereon, that when executed on a processor, induce a driving support system to perform the methods described herein, and

a driving support system for an automotive vehicle comprising a camera for providing images for classification, wherein the driving support system is configured for performing the methods described herein.

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments and examples described hereinafter. Individual features disclosed in the embodiments constitute alone or in combination an aspect of the present invention. Features of the different embodiments can be carried over from one embodiment to another embodiment. Embodiments of the present disclosure are further described in the following examples, which are offered by way of illustration, and are not intended to limit the invention in any manner.

In the drawings:

Fig. 1 schematically depicts an automotive vehicle with a driving support system and a camera according to a first, preferred embodiment of the invention,

Fig. 2 schematically depicts the classification of an image in an image processing in a driving support system of the automotive vehicle according to the first embodiment of the invention, and

Fig. 3 schematically depicts a second embodiment of the classification of an image in image processing in a driving support system of an automotive vehicle that is based on a Hybrid-CNN-DBM model.

Example 1

Fig. 1 schematically depicts an automotive vehicle 1 with a driving support system 2 and a camera 3 according to a first, preferred embodiment of the invention. The camera 3 provides images of a scene, e.g. a motorway at night, for classification by the driving support system 2, which is configured for performing the methods for scene classification of an image in image processing, as described herein.

Fig. 2 schematically depicts the classification of an image in an image processing in a driving support system 2 of an automotive vehicle 1 according to a preferred embodiment of the invention. The image of a scene, e.g. a motorway at night, which is taken by the camera 3 is spatially ordered into regions by clustering the image pixels into regions with high inter-class variance and low intra-class variance.

Spatially ordering regions of the image preferably is done by using region descriptors to capture a semantically uncorrelated low-level representation of each region based on the features i) Gabor filter, ii) Hue, Saturation and Value (HSV) color space features and iii) Co-occurrence features that are Haralick features derived from the Gray Level Co-occurrence Matrix (GLCM, intra-region and inter-region). The captured low-level representation of each region is semantically uncorrelated in that the above used features i), ii) and iii) are mutually uncorrelated in a semantic sense. The i) Gabor filter is a linear filter used for texture analysis, which analyses whether there is any specific frequency content in the image in specific directions in a localized region around the point or region of analysis. Frequency and orientation representations of Gabor filters are proven to be a feature used by the human vision. The ii) HSV color space is a color space that defines the localization of a color by the features Hue, Saturation and Value.

The iii) Co-occurrence features are used for intra-region co-occurrence statistics (mean and range), wherein said features are selected from the group consisting of Angular Second Moment, Contrast, Sum Average, Sum Variance and Difference Variance. Spatially ordering the regions of the image is further done by adding spatial relationships among neighboring regions to create the Spatially Ordered Region Descriptor (SORD), wherein further Haralick features are used that are selected from the group consisting of the mean and range of iv) Correlation, v) Entropy, vi) Sum Variance, vii) Difference Variance and viii) Information Measures of Correlation.

The SORD that is created in this way comprises the features i) Gabor filter, ii) Hue, Saturation and Value (HSV) color space features, iii) Co-occurrence features, iv) Correlation, v) Entropy, vi) Sum Variance, vii) Difference Variance and viii) Information Measures of Correlation. That means that the aforementioned eight features are added to the Region Descriptor to convert it into the SORD.

The SORD is then fed into the neurons of the visible layer (having e.g. 1024 units) of the DBM followed by processing through the hidden layers (a Hidden Layer 1 having e.g. 512 units and a Hidden Layer 2 having e.g. 256 units) of the DBM for modelling the underlying joint probability distribution of each scene category. The hidden layers of the DBM are a stack of RBMs.

The final step of classifying the scene is then performed by the final softmax layer having e.g. 1000 units on top of the hidden layers (stack of RBMs) of the DBM. That means that the DBM is topped off with a softmax layer in order to perform the classification of the scene. Once the stack of RBMs has learned the underlying joint probability distributions of each scene category, the softmax layer added on top does the actual classification.

Example 2

Fig. 3 schematically depicts a second embodiment of the classification of an image in image processing in a driving support system 2 of an automotive vehicle 1 that is based on a Hybrid-CNN-DBM model. The image of a scene, e.g. a motorway at night, that is taken by the camera 3 is fed into a CNN comprising a plurality of layers (Layer 1, Layer 2, …, Layer n) to learn what features of the image are most useful for the classification of scenes.

Multiple resolutions of features are generated for capturing detail of features at higher resolution as well as “big picture” at lower resolution. The output of each individual layer of the CNN is then fed into the visible layer of the DBM as a separate input for modelling the joint probability distribution of each scene category. In other words, this embodiment of the invention uses a single CNN at multiple resolutions. The features at higher resolution capture detail, whereas the features learned at lower resolution capture the “big picture”, i.e.

the region-level and scene-level information in the image. The single CNN provides these features at multiple resolutions or scales. Since the lower resolution images capture the “big picture”, the CNN trained at that resolution and performing inference at that low resolution provides the region-level and scene-level features for the DBM.

This is followed by processing through the hidden layers (Hidden Layer 1, Hidden Layer 2) of the DBM. The hidden layers of the DBM are a stack of RBMs, which learn the underlying joint probability distribution of each scene category.

The CNN is pre-trained using supervised training and labeled data, wherein classifying the scene is done by a temporary softmax layer as the final layer of the DBM, wherein the temporary softmax layer is stripped off once the CNN has learned the features, followed by feeding the output of each layer of the CNN into the visible layer of the DBM as a separate input.

The DBM is pre-trained using greedy layer-wise pre-training to learn the internal representation of the combination of multiple features in a scene and multi-resolution features of the same region, followed by adding on and pre-training the softmax layer using labelled data.

The final step of classifying the scene is then performed by the final softmax layer on top of the hidden layers (stack of RBMs) of the DBM. In other words, the DBM preferably is topped off with a softmax layer, where classifying the scene based on the learning result of the DBM is performed, once the stack of RBMs has learned the underlying joint probability distributions of each scene category.

Reference signs list

1	automotive vehicle
2	driving support system
3	camera

Claims

1. A method for scene classification of an image in image processing in a driving support system (2) of an automotive vehicle (1) comprising the steps of:
   - spatially ordering regions of the image by clustering the image pixels into regions with high inter-class variance and low intra-class variance,
   - modelling the underlying joint probability distribution of each scene category using a Deep Boltzmann Machine (DBM) as a generative model comprising a stack of Restricted Boltzmann Machines (RBMs) and a classifier such as a Radial Basis Filter Support Vector Machine (RBF-SVM), a second CNN or a softmax layer as the final layer, and
   - classifying the scene based on the learning result of the Deep Boltzmann Machine (DBM).

2. The method according to claim 1, wherein spatially ordering regions of the image comprises using one or more region descriptors to capture a semantically uncorrelated low-level representation of each region based on the features i) Gabor filter, ii) Hue, Saturation and Value (HSV) color space features, and iii) Co-occurrence features that are Haralick features derived from the Gray Level Co-occurrence Matrix (GLCM).

3. The method according to claim 2, wherein the iii) Co-occurrence features are used for intra-region co-occurrence statistics (mean and range), wherein said features are selected from the group consisting of Angular Second Moment, Contrast, Sum Average, Sum Variance and Difference Variance.

4. The method according to claims 2 or 3, wherein spatially ordering regions of the image further comprises adding spatial relationships among neighboring regions to create a Spatially Ordered Region Descriptor (SORD), wherein further Haralick features are used that are selected from the group consisting of the mean and range of iv) Correlation, v) Entropy, vi) Sum Variance, vii) Difference Variance and viii) Information Measures of Correlation.

5. The method according to claim 4, wherein the Spatially Ordered Region Descriptor (SORD) comprises the features i) Gabor filter, ii) Hue, Saturation and Value (HSV) color space features, iii) Co-occurrence features, iv) Correlation, v) Entropy, vi) Sum Variance, vii) Difference Variance and viii) Information Measures of Correlation.

6. The method according to claims 4 or 5, wherein the Spatially Ordered Region Descriptor (SORD) is the input to the neurons of the visible layer of the Deep Boltzmann Machine (DBM) when modelling the underlying joint probability distribution of each scene category followed by processing through a stack of Restricted Boltzmann Machines (RBMs) that learn the underlying joint probability distributions of each scene category.

7. The method according to any of the previous claims, wherein classifying the scene is done by a softmax layer on top of the stack of Restricted Boltzmann Machines (RBMs).

8. A method for scene classification of an image in image processing in a driving support system (2) of an automotive vehicle (1) comprising the steps of:
- providing a convolutional neuronal network (CNN) comprising a plurality of layers to learn what features of the image are most useful for the classification of scenes, wherein multiple resolutions of features are generated for capturing detail of features at higher resolution and “big picture” at lower resolution,
- modelling the joint probability distribution of each scene category using a Deep Boltzmann Machine (DBM) comprising a stack of Restricted Boltzmann Machines (RBMs) and a classifier such as a Radial Basis Filter Support Vector Machine (RBF-SVM), a second CNN or a softmax layer as the final layer, wherein the output of each layer of the convolutional neuronal network (CNN) is fed into the visible layer of the Deep Boltzmann Machine (DBM) as a separate input, followed by the stack of Restricted Boltzmann Machines (RBM) learning the underlying joint probability distribution of each scene category, and
- classifying the scene based on the learning result of the Deep Boltzmann Machine (DBM).

9. The method according to claim 8, wherein the convolutional neuronal network (CNN) is pre-trained using supervised training and labeled data, and wherein classifying the scene is done by a temporary softmax layer as the final layer of the Deep Boltzmann Machine (DBM), wherein the temporary softmax layer is stripped off once the

convolutional neuronal network (CNN) has learned the features, followed by feeding the output of each layer of the convolutional neuronal network (CNN) into the visible layer of the Deep Boltzmann Machine (DBM) as a separate input.

10. The method according to claim 8 or 9, wherein the Deep Boltzmann Machine (DBM) is pre-trained using greedy layer-wise pre-training to learn the internal representation of the combination of multiple features in a scene and multi-resolution features of the same region, followed by adding on and pre-training the softmax layer using labelled data.

11. The method according to any of claims 8 to 10, wherein classifying the scene is done by a softmax layer on top of the stack of Restricted Boltzmann Machines (RBMs).

12. Use of the method according to any of claims 1 to 7 or any of claims 8 to 11 in a driving support system (2) of an automotive vehicle (1).

13. A driving support system (2) for an automotive vehicle (1) comprising a camera (3) for providing images for classification, wherein the driving support system (2) is configured for performing the method according to any of claims 1 to 7 or any of claims 8 to 12.

14. A non-transitory computer-readable medium (4), comprising instructions stored thereon, that when executed on a processor, induce a driving support system (2) to perform the method of any of claims 1 to 7 or any of claims 8 to 12.

15. An automotive vehicle (1) comprising:
a data processing apparatus (5),
the non-transitory computer-readable medium (4) according to claim 14, and
the driving support system (2) according to claim 13.

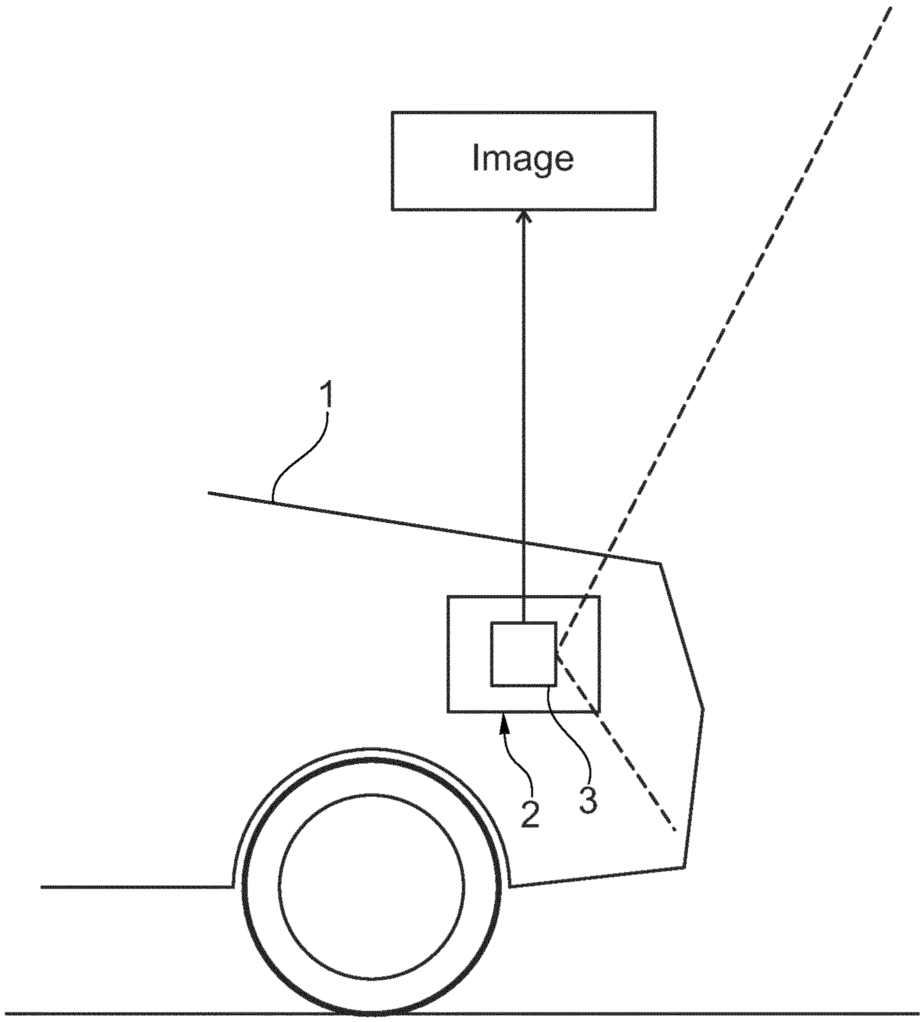

Fig. 1

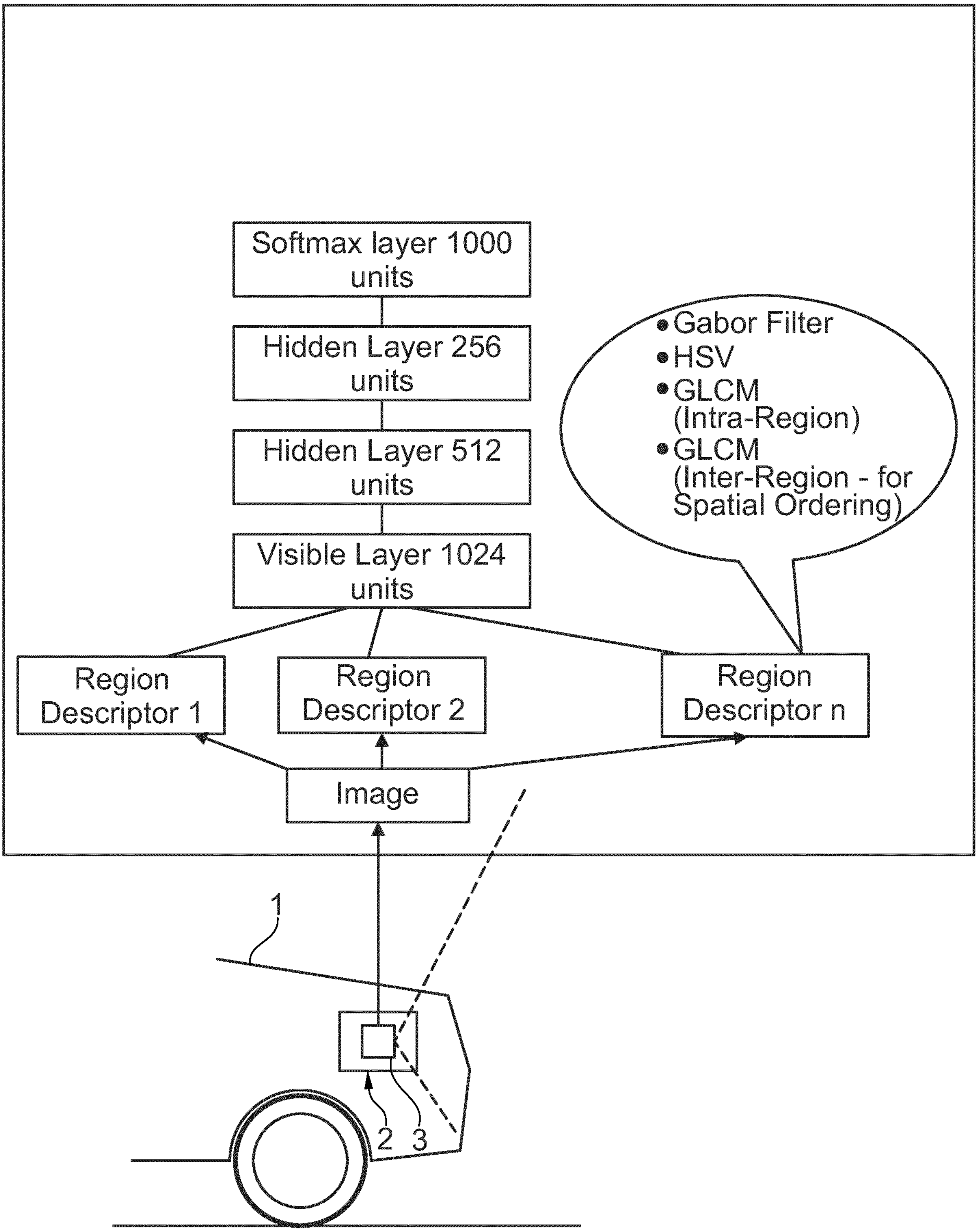
Softmax layer 1000 units
Hidden Layer 256 units
Hidden Layer 512 units
Visible Layer 1024 units
Gabor Filter
HSV
GLCM (Intra-Region)
GLCM (Inter-Region - for Spatial Ordering)
Region Descriptor 1
Region Descriptor 2
Region Descriptor n
Image
1
2
3

Fig. 2

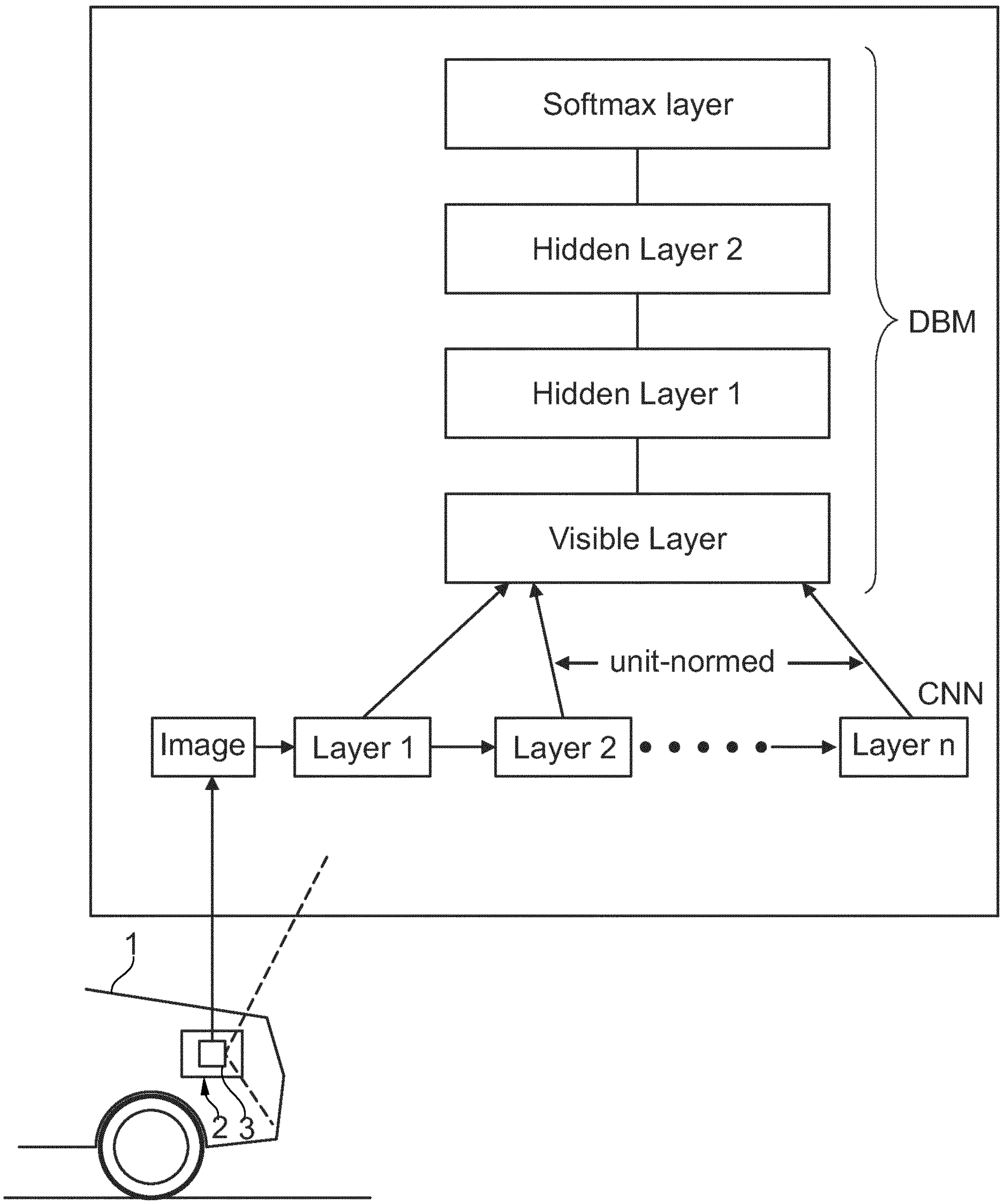
Softmax layer
Hidden Layer 2
Hidden Layer 1
Visible Layer
DBM
unit-normed
CNN
Image
Layer 1
Layer 2
Layer n
1
2
3

Fig. 3

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2018/081874

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06K9/00
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06K

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2017/206440 A1 (SCHRIER MADELINE JANE [US] ET AL) 20 July 2017 (2017-07-20) paragraphs [0012], [0023], [0026], [0040], [0046] -----	1-15

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 February 2019	01/03/2019

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Philips, Petra

4

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2018/081874

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	Ruslan Salakhutdinov ET AL: "Deep Boltzmann Machines", Proceedings of AISTATS 2009, 18 April 2009 (2009-04-18), pages 448-455, XP055556013, United States Retrieved from the Internet: URL:http://proceedings.mlr.press/v5/salakhutdinov09a/salakhutdinov09a.pdf [retrieved on 2019-02-13] abstract figures 1,2 Section "1 Introduction" Section "4.2 NORB" -----	
T	Anonymous: "Otsu's method", Wikipedia , 23 October 2017 (2017-10-23), XP002788905, Retrieved from the Internet: URL:https://en.wikipedia.org/w/index.php?title=Otsu%27s_method&oldid=806653495 [retrieved on 2019-02-13] page 1, paragraph 1 -----	
T	"Chapter 8.3.3 Statistical approaches" In: Nixon, Mark and Aguado, Alberto: "Feature Extraction and Image Processing", 1 January 2002 (2002-01-01), Elsevier, Newnes, XP002788906, ISBN: 0750650788 pages 297-299, item 8.3.3item ; the whole document first paragraph of item 8.3.3 -----	
T	"Output layer for classification" In: Patterson, Josh and Gibson, Adam: "Deep Learning: A Practitioner's Approach", 28 July 2017 (2017-07-28), O'Reilly Media Inc., XP002788908, ISBN: 1491914238 page 180, the whole document ----- -/--	

4

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2018/081874

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GAO JINGYU ET AL: "Natural scene recognition based on Convolutional Neural Networks and Deep Boltzmannn Machines", 2015 IEEE INTERNATIONAL CONFERENCE ON MECHATRONICS AND AUTOMATION (ICMA), IEEE, 2 August 2015 (2015-08-02), pages 2369-2374, XP033216758, DOI: 10.1109/ICMA.2015.7237857 ISBN: 978-1-4799-7097-1 [retrieved on 2015-09-02] abstract -----	1-15

4

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2018/081874

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2017206440 A1	20-07-2017	CN 106980813 A	25-07-2017
		DE 102017100198 A1	20-07-2017
		GB 2548456 A	20-09-2017
		US 2017206440 A1	20-07-2017