

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
24 November 2005 (24.11.2005)

PCT

(10) International Publication Number
WO 2005/111918 A1

(51) International Patent Classification⁷: **G06K 9/62**

(74) Agents: **PASCHBURG, Donald, B.** et al.; Siemens Corporation- Intellectual Property Dept., 170 Wood Avenue South, Iselin, NJ 08830 (US).

(21) International Application Number:
PCT/US2005/015854

(22) International Filing Date: 5 May 2005 (05.05.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/569,552 10 May 2004 (10.05.2004) US
11/111,640 21 April 2005 (21.04.2005) US

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

(71) Applicants (for all designated States except US): **SIEMENS CORPORATE RESEARCH, INC.** [US/US]; 755 College Road East, Princeton, NJ 08540 (US). **SIEMENS AKTIENGESellschaft** [DE/DE]; Wiltelsbacherplatz 2, 80333 Munich (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BAHLMANN, Claus** [DE/US]; 255 Mt. Lucas Road, Princeton, NJ 08540 (US). **ZHU, Ying** [CN/US]; 88 Kingsland Circle, Monmouth Junction, NJ 08852 (US). **COMANICIU, Dorin** [RO/US]; 2 Stuart Lane West, Princeton Junction, NJ 08550 (US). **KÖHLER, Thorsten** [DE/DE]; Am Haslach 17, 93180 Deuerling (DE). **PELLKOFEr, Martin** [DE/DE]; Regensburger Str. 1, 93080 Pentling (DE).

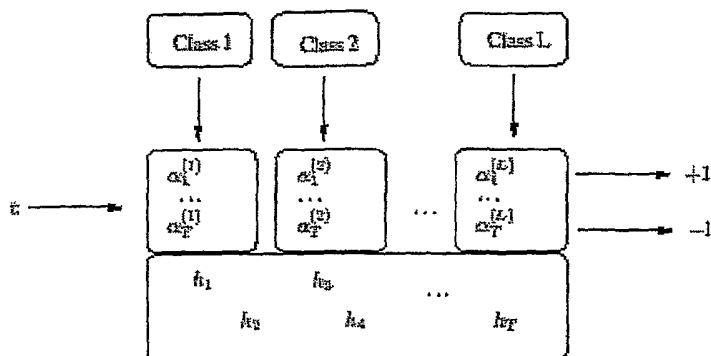
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR COMBINING BOOSTED CLASSIFIERS FOR EFFICIENT MULTI-CLASS OBJECT DETECTION



(57) Abstract: A method for training a system for detecting multi-class objects in an image or a video sequence is described. A common ensemble of weak classifiers for a set of object classes is identified. For each object class, a separate weighting scheme is adapted for the ensemble of weak classifiers. A method for detecting objects of multiple classes in an image or a video sequence is also disclosed. Each class is assigned a detector that is implemented by a weighted combination of weak classifiers such that all of the detectors are based on a common ensemble of weak classifiers. Then weights are individually set for each class.

METHOD FOR COMBINING BOOSTED CLASSIFIERS FOR EFFICIENT MULTI-CLASS OBJECT DETECTION

Cross Reference to Related Applications

[0001] This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/569,552, filed on May 10, 2004, which is incorporated by reference in its entirety.

Field of the Invention

[0002] The present invention is directed to a method for combining boosted classifiers for efficient multi-class object detection, and more particularly, to a system and method for training a detection system that identifies multi-class objects using boosted classifiers.

Background of the Invention

[0003] In recent years, the technique of boosting a set of simple or weak classifiers in order to obtain an overall strong classifier has evolved into a powerful solution, especially in the domain of image object detection. Image object detection is becoming increasingly more popular and can be used in a number of different detection scenarios. Examples of such applications include face detection, pedestrian detection, traffic sign detection, and vehicle detection.

[0004] Boosting techniques are particularly effective for detecting a single object class. However, when extending this approach from the detection of one single object class to the detection of multiple object classes, its complexity scales linearly with the number of classes. Many detection applications require multiple object class detection in order to be effective. An example of such an application is vehicle detection where separate object classes may be defined for vehicles, trucks, pedestrians and traffic signs. Another example of a detection application that requires multiple object classes is people detection. Particularly, if the people are in motion, it is more effective to define people sub-classes based on the difference poses or actions of the people. For example, such sub-classes could include sitting, standing and walking.

[0005] For the task of object detection in images, a known approach uses a learning/detection framework that is based on boosting. Boosting selects and combines a set

$$\mathbf{H} = \{h_1, \dots, h_T\} \quad (1)$$

of simple or weak classifiers $h_t : X \mapsto \{+1, -1\}$, each of it taken from a large set of classifier candidates to form a final or strong classifier. For the problem of object image detection, X is the set of all image patches, the class +1 corresponds to an object and the class -1 to a non-object. Given an additional set of weighting factors

$$\alpha = \{\alpha_1, \dots, \alpha_T\}, \quad (2)$$

the object detection is solved by evaluating the strong classifier h on candidate image patches $x \in X$. The decision $h(x)$ is computed from the weighted sum of the weak classifier decisions, that is,

$$f(x) = \sum_{t=1}^T \alpha_t h_t(x) - \theta \quad (3)$$

and

$$h(x) = \text{sign}(f(x)), \quad (4)$$

where θ is a threshold allowing the user to balance false alarm and miss detection rate. An optimal selection of the weak classifiers h_t and a proper weighting α_t is obtained from an AdaBoost training algorithm.

[0006] This technique has been applied very successfully to the detection of single class objects, for example, faces, vehicles, pedestrians, etc. However, situations exist where the detection of objects of a multitude $\{1, \dots, L\}$ of classes within one scene is desirable, for example, a combined detection of cars, trucks, pedestrians and traffic signs in traffic scenes.

[0007] Based on the above-described algorithm, a naïve solution would use AdaBoost to train one individual ensemble of weak classifiers $H^{(l)}$ and weights $\alpha^{(l)}$ for each class l , that is,

$$\{H^{(1)}, \dots, H^{(L)}\} = \{\{h_1^{(1)}, \dots, h_{T_1}^{(1)}\}, \dots, \{h_1^{(L)}, \dots, h_{T_L}^{(L)}\}\} \quad (5)$$

and

$$\{\alpha^{(1)}, \dots, \alpha^{(L)}\} = \{\{\alpha_1^{(1)}, \dots, \alpha_{T_1}^{(1)}\}, \dots, \{\alpha_1^{(L)}, \dots, \alpha_{T_L}^{(L)}\}\} \quad (6)$$

The memory and computational complexity for the detection task of this approach scales linearly with L and in many cases circumvents a real-time detection system. Such a computationally extensive approach is not feasible in time, cost or efficiency to be considered as a reasonable solution. There is a need for a detection system that is capable of real-time multi-class detection and which can perform in an efficient manner.

Summary of the Invention

[0008] In the present invention, a more efficient multi-class detection method and system is devised by a joint design of key components including feature and classifier design of individual detectors. In accordance with an embodiment of the present invention, a method for training a system for detecting multi-class objects in an image or a video sequence is described. A common ensemble of weak classifiers for a set of object classes is identified. For each object class, a separate weighting scheme is adapted for the ensemble of weak classifiers.

[0009] The present invention is also directed to a method for detecting objects of multiple classes in an image or a video sequence. Each class is assigned a detector that is implemented by a weighted combination of weak classifiers such that all of the detectors are based on a common ensemble of weak classifiers. Then weights are individually set for each class.

Brief Description of the Drawings

[0010] Preferred embodiments of the present invention will be described below in more detail, wherein like reference numerals indicate like elements, with reference to the accompanying drawings:

[0011] FIG. 1 is a system block diagram of a system for detecting and tracking multi-class objects in accordance with the present invention; and

[0012] FIG. 2 illustrates a multi-class object detection framework in accordance with the present invention.

Detailed Description

[0001] The present invention is directed to a system and method for combining boosted classifiers for efficient multi-class object detection. The present invention is also directed to a method for training classifiers for multi-class object detection. The present invention can be used in a multitude of applications that require the detection of different objects. Such a system may, for example, be used for surveillance applications, such as for detecting and tracking a person or facial features.

[0002] The present invention could also be used to detect and track objects on an assembly line. Other applications could be created for detecting and tracking human organs for medical applications. It is to be understood by those skilled in the art that the present invention may be used in other environments as well. Another environment in which the present invention would be useful is in the detection and tracking of vehicles. In addition to detecting different types of vehicles, there is also sometimes a need for detecting pedestrians, traffic signs and other vehicle-environment related objects. For purposes of describing the present invention, it will be described in the on-road obstacle detection environment.

[0003] FIG. 1 illustrates a block diagram of a system for implementing the present invention. One or more cameras 102 are used to capture images of a road and its surroundings. As would be expected with a typical road image, the image includes background images, such as buildings, trees, and houses, and vehicles driving on the road. The images are communicated to a processor 104 which computes confidence scores using the component classifiers in a database 106. The images are analyzed in accordance with the present invention to identify different classes of objects. Examples of such objects include cars, trucks, pedestrians and traffic signs. Once an object is identified at a sufficient confidence level, the object is identified via an output device 108. The output device 108 provides an output signal which communicates to the user the presence of one or more objects. The output signal may be an audible signal or other type of warning signal. The output device 308 may also include a display for viewing the detected objects.

The display provides a view of the images taken by the cameras 302 which may then be enhanced to indicate objects that have been detected and which are being tracked.

[0013] In developing a technique for efficiently detecting multi-class objects, the present invention is directed to a number of training techniques which are used to identify and select a number of weak classifiers which are then stored in database 106 and used to detect the various objects. A premise of the present invention is an underlying principle that all L detectors share a common set of features or weak classifiers. In order to retain a high detection accuracy, each individual strong classifier is adapted to its individual set of weights $\alpha^{(1)}$ posterior to the weak classifier selection. By using individual weighting there is only a minimal increase in the computation complexity compared to single class detection. The majority of the complexity derived in the computations of Equations (3) and (4) of the prior art arise from the evaluation of the weak classifiers and not the weighting.

[0014] FIG. 2 illustrates a multi-class object detection framework in accordance with the present invention. Training of the system refers to the selection of a common set of weak classifiers H and the adaptation of individual weights $\alpha^{(1)}$. In accordance with the present invention, the following technique can be used to train the weak classifiers:

1. Given L object classes, select a common ensemble $H = H^{(1)} = \dots = H^{(L)} = \{h_1, \dots, h_T\}$ of weak classifiers.
2. **for** $l = 1, \dots, L$
3. Adapt $\alpha^{(1)}$ on the training set of class l using AdaBoost, where the candidate weak classifiers are given by H .

[0015] Given step 1 above, the remaining open issue is the selection of the common weak classifier ensemble H . In accordance with the present invention, three different techniques are described for selecting H . The first technique assumes that one distinguished class l_0 exists. The distinguished class can be, for instance, a class with a high occurrence of objects in the scene. The training set of class l_0 is used to select a common ensemble

$$H = H^{(l_0)} = \{h_1^{(l_0)}, \dots, h_{T_{l_0}}^{(l_0)}\} \quad (7)$$

of weak classifiers. Once the ensemble is selected, as indicated above, the individual class weights $\{\alpha^{(1)}, \dots, \alpha^{(L)}\}$ are re-adapted using AdaBoost, where the candidate weak classifiers are taken solely from H . An example of an application where this technique might be used is a manufacturing application. Since the predominant object being detected might be an item on an assembly line the weak classifiers associated with that object would be used as the training set.

[0016] A second technique for creating a training set is used when the assignment of one distinguished class is not reasonable. In such an instance, a more symmetric procedure would be appropriate. In this approach, a common classifier H is trained based on the union of all training samples. Next, each individual classifier l is fine-tuned by adjusting $\alpha^{(l)}$ with respect to H and the positive training samples solely from class l . In other words, in this approach H is optimized in order to discriminate all positive objects from the negatives. The individual detection problem can be specialized by adapting the respective impact of the individual weak classifiers.

[0017] An example where this technique might be used is in application where the classes share some commonality. For example, facial detection uses classes that are made up of different facial poses. However, many of the classifiers for the poses are similar (e.g., eyes, nose, and mouth). Likewise, vehicle detection uses classes that are made up of different vehicles. In each case, normally certain features such as the edges and corners of the vehicle are looked at.

[0018] A third technique obtains H by collecting the most appropriate weak classifiers for each object class l . In this respect, individual sets of weak classifiers $\{H^{(1)}, \dots, H^{(L)}\}$ are selected in a first training step. Then, the final set H is obtained from a combination of $\{H^{(1)}, \dots, H^{(L)}\}$. One approach is to use the union

$$H = \bigcup_l H^{(l)}, \quad (8)$$

other options including using the intersection of $\{H^{(l)}\}$, or using a subset of $\{H^{(l)}\}$ decided by a more complex voting scheme.

[0019] This technique is best used for applications that require a distribution of data resources. For example, in a more complex vehicle detection scheme, in addition to vehicles, the application may also detect, traffic signs and pedestrians. Additional objects

that may be detected include road lanes, guard rails and background objects like trees, houses and other buildings. In such an instance, each of the classes uses weak classifiers that are distinct from the other classes. However, there still may be some common weak classifiers among all of the classes.

[0020] The framework shown in FIG. 2 can be extended to be applicable for multi-class classification. In this scenario, the classifier assigns the most probable class label \hat{l} to x . A natural procedure for this task would be to choose the class with maximum activation, that is,

$$\hat{l} = \arg \max_l \{f^{(l)}(x)\} \quad (9)$$

with $f^{(l)}$ the activation function of class l .

[0021] The present invention uses a set of boosted weak classifiers that is common within all individual class detectors and thus has to be computed only once. A possible major loss in detection accuracy by this restriction is prevented by a posterior re-training of the weak classifier weights, individually to each object class to reduce its misclassification error. The training complexity of the present invention does not increase significantly compared to the prior art approaches since re-adjustment of the individual class weights with AdaBoost is based only on \mathbb{H} and thus on a much small set of weak classifier candidates.

[0022] Having described embodiments for a method for using boosted classifiers to detect multi-class objects, it is noted that modifications and variations can be made by persons skilled in the art in light of the above teachings. It is therefore to be understood that changes may be made in the particular embodiments of the invention disclosed which are within the scope and spirit of the invention as defined by the appended claims. Having thus described the invention with the details and particularity required by the patent laws, what is claimed and desired protected by Letters Patent is set forth in the appended claims.

We claim:

1. A method for training a system for detecting multi-class objects in an image or a video sequence comprising the steps of:
identifying a common ensemble of weak classifiers for a set of object classes; and
for each object class, adapting a separate weighting scheme for the ensemble of weak classifiers.
2. The method of claim 1 wherein one of the object classes is a distinguished class and the ensemble of weak classifiers are selected from the distinguished class.
3. The method of claim 2 wherein the distinguished class has a high occurrence of at least one object in the image or the video sequence.
4. The method of claim 1 wherein the ensemble of weak classifiers is learned from a data set comprising all of the object classes.
5. The method of claim 4 wherein the ensemble of weak classifiers is optimized to discriminate positive objects from negative objects.
6. The method of claim 1 wherein the ensemble of weak classifiers is a combination of weak classifier ensembles, each of which is individually learned from a single object class.
7. The method of claim 6 wherein the combination is a mathematical union of the ensembles of weak classifiers.
8. The method of claim 6 wherein the combination is a mathematical intersection of the ensembles of weak classifiers.
9. The method of claim 6 wherein the combination of weak classifier ensembles is a subset determined by voting.

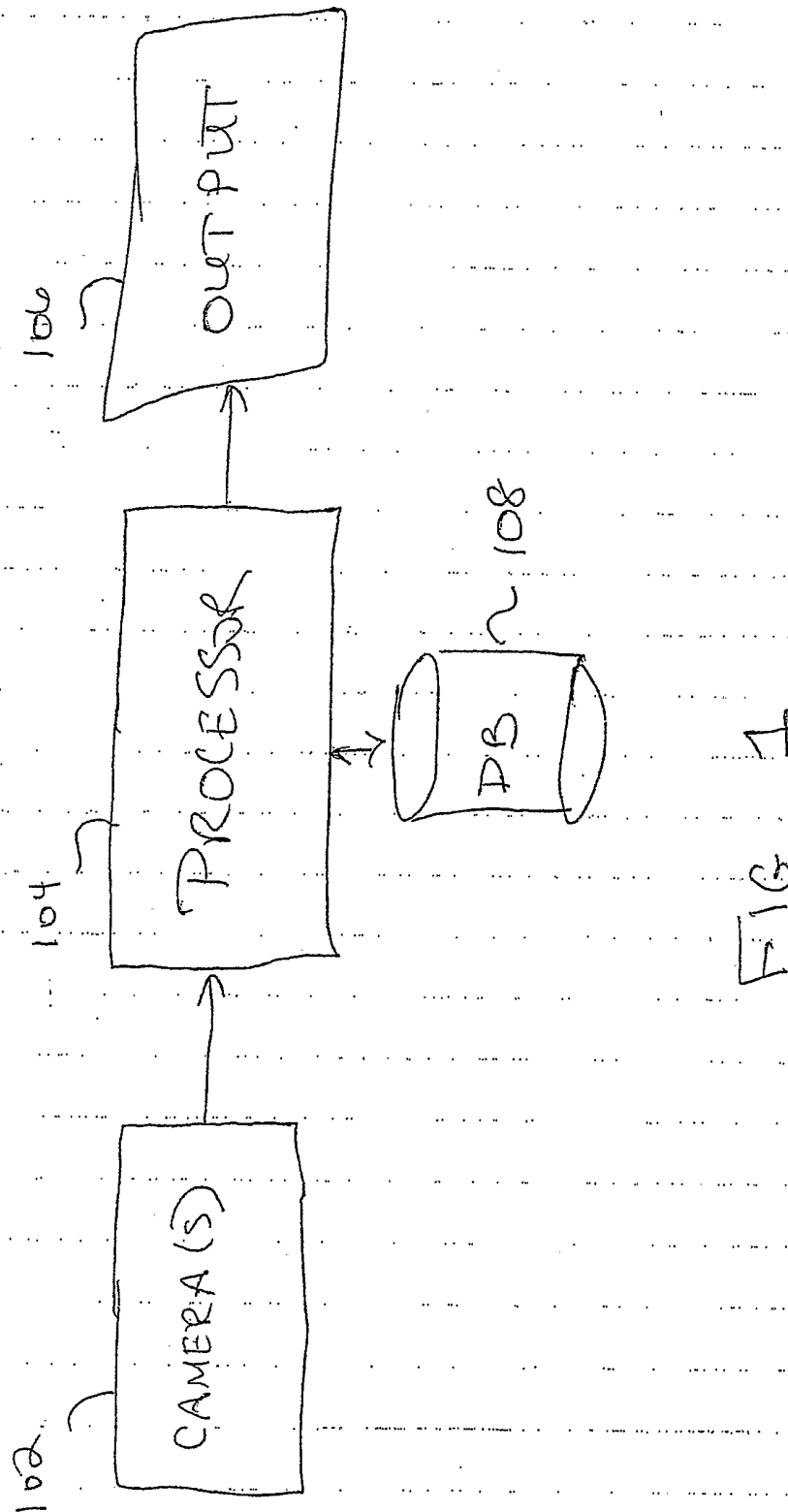
10. A method for detecting objects of multiple classes in an image or a video sequence, the method comprising the steps of:

assigned each class a detector that is implemented by a weighted combination of weak classifiers wherein all of the detectors are based on a common ensemble of weak classifiers, and

individually setting weights for each class.

11. The method of claim 10 wherein the scores of the detectors are utilized to classify the detected object into one of the object classes.

1/2



2/2

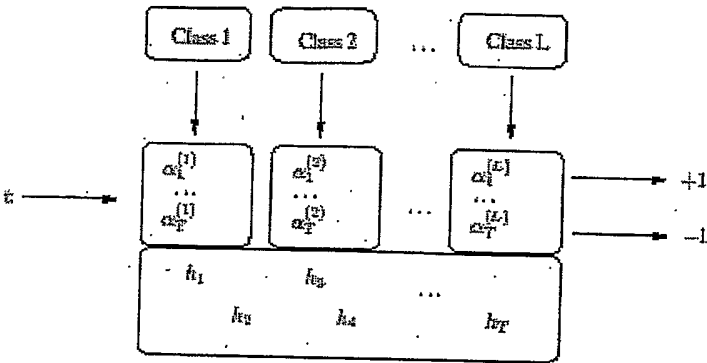


FIG. 2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/015854

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06K9/62

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)
IPC 7 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 practical, search terms used)

EPO-Internal, INSPEC, 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	<p>TORRALBA A ET AL.: "Sharing Visual Features for Multiclass and Multiview Object Detection"</p> <p>MASSACHUSETTS INSTITUTE OF TECHNOLOGY, COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY, AI MEMO, no. 2004-008, April 2004 (2004-04), XP002337421</p> <p>Cambridge, Massachusetts, USA</p> <p>the whole document</p> <p style="text-align: center;">----- -/--</p>	1-11

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in 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 document 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

22 July 2005

Date of mailing of the international search report

08/08/2005

Name and mailing address of the ISA
 European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

Müller, M

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/015854

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>XIAO R ET AL: "ROBUST MULTIPOSE FACE DETECTION IN IMAGES" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE INC. NEW YORK, US, vol. 14, no. 1, January 2004 (2004-01), pages 31-41, XP001186968 ISSN: 1051-8215 abstract; figures 1,6,11,12 page 32, left-hand column, paragraph 5 - right-hand column, paragraph 4 page 33, left-hand column, paragraph 7 - right-hand column, paragraph 1 page 35, right-hand column, paragraph 4 - page 36, left-hand column, paragraph 2</p>	1,5,10, 11
X	<p>VIOLA P ET AL: "Rapid object detection using a boosted cascade of simple features" PROCEEDINGS 2001 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. CVPR 2001. KAUAI, HAWAII, DEC. 8 - 14, 2001, PROCEEDINGS OF THE IEEE COMPUTER CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, LOS ALAMITOS, CA, IEEE COMP. SOC, US, vol. VOL. 1 OF 2, 8 December 2001 (2001-12-08), pages 511-518, XP010583787 ISBN: 0-7695-1272-0 abstract; figure 3 pages I-511, paragraph 4 pages I-512, left-hand column, paragraph 6 - right-hand column, paragraph 2</p>	1,5,10, 11
A	<p>SCHNEIDERMAN H ET AL: "A statistical method for 3D object detection applied to faces and cars" PROCEEDINGS IEEE COMPUTER SOCIETY CONFERENCE OM COMPUTER VISION AND PATTERN RECOGNITION, vol. 1, September 2000 (2000-09), pages 746-751, XP002313459 abstract</p> <p style="text-align: center;">----- -/-</p>	5

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/015854

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>TIEU K ET AL: "BOOSTING IMAGE RETRIEVAL" PROCEEDINGS 2000 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. CVPR 2000. HILTON HEAD ISLAND, SC, JUNE 13-15, 2000, PROCEEDINGS OF THE IEEE COMPUTER CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, LOS ALAMITOS, CA : IEEE COMP. SO, vol. VOL. 1 OF 2, 13 June 2000 (2000-06-13), pages 228-235, XP001035602 ISBN: 0-7803-6527-5 abstract page 229, left-hand column, paragraph 4 - page 230, right-hand column, paragraph 3 -----</p>	