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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
09 January 2020 (09.01.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/007881 A1

(51) International Patent Classification:
G06F 3/01 (2006.01)

(21) International Application Number:
PCT/EP2019/067773

(22) International Filing Date:
02 July 2019 (02.07.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
LU100869 04 July 2018 (04.07.2018) LU

(71) Applicant: **IEE INTERNATIONAL ELECTRONICS & ENGINEERING S.A.** [LU/LU]; Zone industrielle, 12, rue Pierre Richardot, 6468 Echternach (LU).

(72) Inventors: **BEISE, Hans-Peter**; Im Kemmer 2, 66706 Perl (DE). **SCHRÖDER, Udo**; In der Rotheck 35, 54343 Föhren (DE). **DIAS DA CRUZ, Steve**; 21, Rue Widderbierg, 6675 Mertert (LU).

(74) Agent: **BEISSEL, Jean** et al.; Office Freylinger S.A., 234, Route d'Arlon, BP 48, 8001 Strassen (LU).

(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,

(54) Title: METHOD FOR TRAINING A GESTURE RECOGNITION SYSTEM

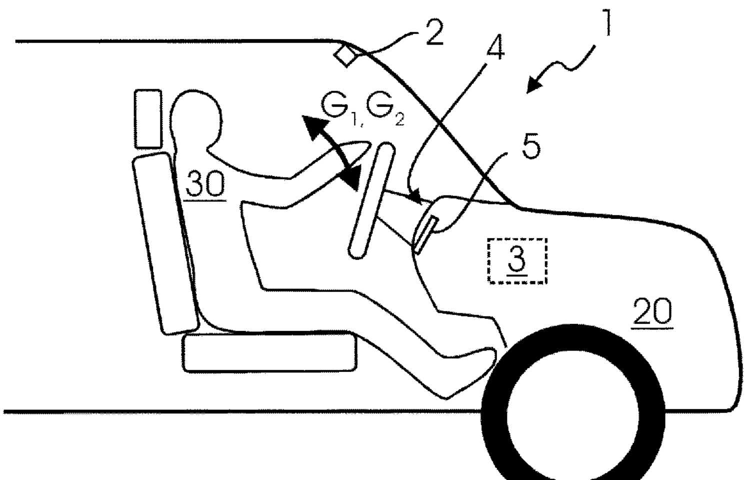

Fig. 1)

(57) Abstract: The invention relates to a method for training a gesture recognition system (3, 13). In order to provide means for adapting a gesture recognition system to a user while being online, the invention provides that the method comprises: - performing at least one sampling cycle, including: - receiving, from a sensor system (2, 12), sensor data (x) representing a first gesture (G_1) by a user (30) and at least temporarily storing the sensor data (x), - based on a mapping relation, interpreting the first gesture (G_1) as corresponding to a first action (α_x) to be performed, which first action (α_x) is selected from a predefined action set (S), - performing the first action (α_x), and - if, within a predefined reaction time (T) after beginning to perform the first action (α_x), a correcting user input is received indicating that the first gesture (G_1) corresponds to a different,

WO 2020/007881 A1

[Continued on next page]

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, 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))*

second action (α_y) from the action set (S), storing the second action (α_y) together with the sensor data (x) as a gesture sample, wherein after receiving the sensor data (x), a score (c) indicating a degree of correspondence between the first gesture (G_1) and the second action (α_y) is calculated, and the gesture sample is only stored if the score (c) is above a predefined threshold (c_{min}) and - after storing at least one gesture sample, adapting the mapping relation based on at least one stored gesture sample.

Method for Training a Gesture Recognition System

Technical field

[0001] The present invention generally relates a method for training a gesture recognition system and to an online trainable system.

Background of the Invention

[0002] Today, gesture recognition is used in a growing number of applications as a means of command input. The user performs a gesture, which is recognised and interpreted as a command. Gesture recognition systems utilize various types of sensors, like visual, radar, capacitive, etc., to detect the gesture. Among others, such systems are presently used to steer the infotainment system in a car interior or to activate an opening/closing mechanism for a vehicle door or a trunk.

[0003] However, such systems are usually trained offline, i.e. a supervised learning is performed by means of training data that have been generated/measured by the producer. Hence, they are not calibrated to the actual customer. As gestures can be very individual and depend on the individual body dimensions, this approach (which may be referred to as offline training) may significantly limit the performance of such a system in practice. In principle, it is conceivable to apply a training procedure for a specific user, where the user is asked to perform specific gestures and the recognition system learns how these gestures are being performed. However, such a training procedure can be tedious for the user and means an interruption to the normal service time.

Object of the invention

[0004] It is thus an object of the present invention to provide means for adapting a gesture recognition system to a user while being online.

[0005] This problem is solved by method for training a gesture recognition system according to claim 1 and by an online trainable system according to claim 12.

General Description of the Invention

[0006] The invention provides a method for training a gesture recognition system. The gesture recognition system is of course adapted to recognise a gesture of a user and to associate this gesture with an action to be performed. The respective gesture could e.g. correspond to a motion of a finger, a hand, an arm, a foot, a leg or even a head of the user. Optionally, combined motions of several body parts could be included in one gesture. In general, the gesture recognition system may be used in various applications. In particular, it can be associated with a car and the action to be performed may be an action of an active system of the car, e.g. a car infotainment system, a door opener or a trunk opener. However, the gesture recognition system could be associated with other devices, e.g. with a multimedia device like a tablet or a smartphone. At least some aspects of the gesture recognition system are normally software-implemented.

[0007] The method includes performing at least one sampling cycle. As will become apparent the following, the term "sampling cycle" refers to a sample that can be obtained during each sampling cycle. Performing the sampling cycle includes receiving, from a sensor system, sensor data representing a first gesture by a user and at least temporarily storing the sensor data. The sensor system could comprise any kind of sensor that can be used to detect a body motion, i.e. a gesture of the user. For example, it may comprise an optical sensor like a camera, a radar sensor, an ultrasonic sensor and/or a capacitive sensor. In general, the invention is in no way limited to a specific type of sensor system. The sensor data represent a first gesture of the user. In this context, the first gesture can be any kind of body motion, whether intentionally or unintentionally performed by the user. This gesture is detected by the sensor system, thereby generating sensor data. Primarily, the sensor data may even be an analogue signal, which normally is converted to a digital format. The sensor data are then at least temporarily stored. Of course, this includes the possibility that the format of the sensor data is converted, compressed or processed in any other way before they are stored. As will become apparent in the following, the sensor data are normally only stored temporarily, not permanently. They can be stored in a memory of the gesture recognition system or in some external memory.

[0008] In another step of the sampling cycle, the first gesture is interpreted, based on a mapping relation, as corresponding to a first action to be performed, which first action is selected from a predefined action set. The mapping relation in general comprises one or a plurality of rules or criteria according to which a gesture can be associated with an action to be performed. One might also say that these are criteria according to which the gesture can be mapped onto an action. For example, one gesture could be a kicking motion that the user performs next to the trunk of a car, and the associated action could be opening the trunk. Another example is a swiping motion that the user performs with a finger on a steering wheel, where the associated action is turning up the radio volume. The action set is a set of actions, each of which is associated with a certain gesture via the mapping relation, and vice versa. The mapping relation can be rather simple or complex. It may be applied by a gesture recognition module that can be regarded as the central part of the gesture recognition system. Of course, the gesture recognition module is at least partially software-implemented.

[0009] After the first action has been identified using the mapping relation, the first action is performed, e.g. by the above-mentioned active system like an infotainment system, a trunk opener or the like.

[0010] If, within a predefined reaction time after beginning to perform the first action, a correcting user input is received indicating that the first gesture corresponds to a different, second action from the action set, the second action together with the sensor data are stored as a gesture sample. According to the method, a reaction time is predefined, which is a time interval after beginning to perform the first action. As the first action is performed, the user can realise whether this was the action he intended to initiate with the first gesture or not. It is possible that the gesture recognition system misinterpreted the first gesture and erroneously associated it with the first action. The reaction time is an estimate of how long it may take the user at maximum to realise if the first action is the "correct" action or not and to react with a correcting user input. For example, the reaction time could be within one and 10 seconds or between 2 and 5 seconds. It is possible that the reaction time depends on the first action, e.g. because the user may need different amounts of time to realise the first action depending on its nature. By the user input, the user tries to correct the misinterpretation by the gesture recognition system and to initiate a second action that he originally intended to initiate with the first gesture. Therefore, the sensor data and second action are stored together as a gesture sample. The sensor data had already been stored before, wherefore normally only the second action has to be stored in a way that allows association with the sensor data. Although the second action is stored "together" with the sensor data, this normally does not mean that they are physically stored in the same region of a memory device, but only that they are stored so that they are both parts of the gesture sample.

[0011] In some cases, the gesture recognition system may fail to properly recognise a gesture because the user made a mistake. For example, the user could have performed a first gesture that does not even resemble the gesture associated with the second action, e.g. because the user was distracted or is simply unfamiliar with the gesture recognition system. In these cases, the system should not include the first gesture in the adapting process explained below. Therefore, it is useful if only those gesture samples are included where the first gesture at least resembles an "ideal" gesture (most clearly) associated with the second action. According to the invention, after receiving the sensor data, a score indicating a degree of correspondence between the first gesture and the second action is calculated, and the gesture sample is only stored if the score is above a predefined threshold. In other words, some criterion or criteria are applied to determine how well the first gesture indeed corresponded to the second action and a score is calculated that represents a probability. Of course, this score is higher if the first gesture has more resemblance to the above-mentioned "ideal" gesture. Preferably, the score indicates a degree of similarity between the first gesture and an ideal gesture associated with the second action. The ideal gesture can be - explicitly or implicitly - defined by the mapping relation. It is understood that the above-mentioned score could be assigned to a plurality of actions in the action set, in particular to each action in the action set, and when associating the first action with the first gesture, the first action is the one with the highest score. In other words, interpreting the first gesture as corresponding to the first action could be based on the first action having the highest score among the plurality of actions. The score is calculated after receiving the sensor data, which of course does not mean that it has to be calculated immediately afterwards. For example, it could be calculated only after the correcting user input has been received. As an alternative to determining the similarity of the first gesture and the ideal gesture, a “distance” or “similarity” could be defined (or calculated in advance) for each pair of actions. This distance/similarity would be based on a comparison of the respective ideal gestures associated with the actions. Then, the score could be calculated based on this distance/similarity.

[0012] As mentioned above, the method includes performing at least one sampling cycle. In another step of the method, after storing at least one gesture sample, the mapping relation is adapted based on at least one stored gesture sample. Storing at least one gesture sample requires completion of at least one sampling cycle. The at least one gesture sample contains information on at least one (first) gesture that is to be associated with a (second) action to be performed, but was misinterpreted by the gesture recognition system. Therefore, the at least one gesture sample can be used to adapt the mapping relation, i.e. to perform a learning process. There are numerous ways how this adaptation can be performed, at least some of which can be classified as "machine learning". Either way, by adapting the mapping relation based on the gesture sample(s), and adaptation to the individual gesture performance of a specific user can be achieved. After the adaptation has been completed, the respective first gesture, if repeated in the future, will be recognised correctly – i.e. associated with the correct action – with a higher probability. Adapting the mapping relation may at least implicitly correspond to adapting the ideal gestures associated with the individual actions. It should be noted that the inventive method can be performed as the recognition system is online and does not require any interruption by a dedicated training procedure.

[0013] Preferably, the score is calculated based on at least one criterion of the mapping relation. As mentioned above, the mapping relation is normally based on at least one criterion or several criteria. It will be understood that interpreting the first gesture as corresponding to the first action may be based on fulfilment of only some of the criteria or on partial fulfilment of at least one criterion. Therefore, this at least one criterion can also be applied for calculating a score that indicates a degree of correspondence. One could also say that the score is calculated at least partially based on the same criteria as the mapping relation. In fact, as already hinted above, the mapping relation could be based on calculating the score for a plurality of actions and selecting the action with the highest score as the one that the first gesture corresponds to.

[0014] It is preferred that if, within the reaction time, a correcting user input is received corresponding to the second action, the second action is stored together with the sensor data as the gesture sample. In other words, within the reaction time, a user input is received that indicates that the user intended to initiate a different action, namely the second action. One example of this could be that the user of a car infotainment system performs a first gesture which is interpreted as corresponding to starting a phone call. Within a reaction time, a correcting user input is received that corresponds to muting the radio, thereby indicating that his original first gesture was not intended to initiate the first action (starting the phone call). Rather, it can be assumed that the first gesture was also intended to mute the radio. This assumption can optionally be verified, as will be described below. Based on this assumption, the first gesture is associated with the second action (muting the radio) and both are stored as a gesture sample.

[0015] Normally, the action set includes a neutral action, corresponding to inactivity. If a gesture is associated with the neutral action, this gesture has no effect whatsoever. Such gestures can be, for example, any random motion performed by the user without any intention to initiate an action. In particular, any gesture that cannot be recognised or associated with a "real" action, is interpreted as corresponding to the neutral action. However, it is also possible that the user performs a first gesture intended to initiate an action like e.g. opening the trunk, but the gesture recognition system interprets the first gesture as corresponding to the neutral action, for example because the motion performed by the user was too small to be recognised properly. Then, the user could try to correct this misinterpretation by a correcting user input, e.g. by manually pressing a trunk opener button. If the first action is the neutral action, “performing the neutral action” begins after interpreting the first gesture has been finished, i.e. the reaction time starts after the first gesture has been interpreted as corresponding to the neutral action. Especially in this case, the reaction time could be longer than for other actions, since it might take the user longer to realise that the neutral action is performed (i.e. that nothing happens).

[0016] However, it is also conceivable that the user performs some random motion as the first gesture. In this case, the user's intention is not to initiate any "real" action, but the first gesture still could be interpreted as corresponding to such an action, e.g. opening the trunk, turning on the radio or the like. In such a case, the normal reaction of the user could be to simply undo the first action, e.g. close the trunk, turn the radio off again or the like. This behaviour by the user can be interpreted in that the first gesture should have been associated with the neutral action. According to one embodiment the method includes that if, within the reaction time, a correcting user input is received corresponding to undoing the first action, the neutral action is stored together with the sensor data as the gesture sample. The idea is that the user would normally not input a first command by the first gesture and then shortly afterwards (i.e. within the reaction time) input a second command by the correcting user input to undo the first command if inputting the first command was intentional. Rather, the most likely interpretation is that the user did not want to input any command at all, i.e. the correct interpretation by the system would have been the neutral action. Accordingly, the first gesture is associated with the neutral action by storing these two as a gesture sample.

[0017] In order to further improve the reliability and accuracy of the learning process, it is preferred that before storing the gesture sample, the user is asked to confirm that the first gesture did not correspond to the first action. This could be done by outputting an audio or text message to the user, corresponding to e.g.: "Did you not want to…?" or "Was your gesture not intended to…?". The response of the user could then be input by voice or manually, e.g. by the user pressing a button or touching a sensor field. In principle, it is also conceivable that the response of the user is also input through the sensor system as another gesture. However, this may not be preferred because recognising the user's response should preferably be failsafe and independent of gesture recognition based on the sensor system. It is understood that in this embodiment, the gesture sample is only stored if the user confirms that the first gesture did not correspond to the first action.

[0018] If the user did not want any "real" action to be performed, i.e. if he wanted the neutral action to be performed, it can be sufficient to merely ask the user to confirm that the first gesture did not correspond to the first action. However, in other cases it may be preferable that before storing the gesture sample, the user is asked to confirm that the first gesture corresponds to the second action. This could be done by outputting an audio or text message to the user, corresponding to e.g.: "Did you want to…?" or "Was your gesture intended to…?". The response of the user can be input by any of the means mentioned above. It is understood that in this embodiment, the gesture sample is only stored if the user confirms that the first gesture corresponds to the second action.

[0019] The correcting user input may be received in various different ways. According to one embodiment, the correcting user input is sensor data representing a second gesture received from the sensor system. In other words, sensor data are received that are interpreted to correspond to the first gesture and within the (relatively short) reaction time, sensor data are received that are interpreted to correspond to the second (normally different) gesture. It should be noted that both gestures could be similar and the corresponding actions could be identical, though. For example, if the first gesture corresponds to a toggling action like e.g. "turn radio on/off", the user allegedly turning the radio on and turning it off immediately afterwards could indicate that the user did not want to turn the radio in the first place. Normally, however, the second gesture is different from the first gesture and can either correspond to a second action undoing the first action or it may correspond to a second action that is not related in any way to the first action. Receiving the correcting user input as a second gesture can be advantageous insofar as there is no need for an alternative input means or the user employing such input means for the correcting input.

[0020] Alternatively or additionally, it is possible that the correcting user input is received from an input device independent of the sensor system. This input device can be for example a sensor field, a button, a microphone or the like. Using such independent systems has the advantage that they are unaffected by difficulties of the gesture recognition system to appropriately recognise gestures.

[0021] After adapting the mapping relation, at least one stored gesture sample may be deleted. In particular, all stored gesture samples may be deleted. This is an approach where after the adaptation process, the stored gesture samples are regarded as already used and therefore obsolete. However, in some cases it may also be useful to keep at least some of the stored gesture samples for the next adaptation. Optionally, they could be included in the next adaptation process by applying a weight factor, with newer gesture samples having a greater weight than older gesture samples.

[0022] In principle, the number of sampling cycles that are performed before the mapping relation is adapted can be varied. It is understood that for an adaptation of the mapping relation, at least one stored gesture sample is necessary. Apart from this, the adaptation could be performed after a certain time interval, e.g. every day or after a certain number of service hours. According to one embodiment, the mapping relation is adapted when a predefined number of gesture samples have been stored. This predefined number could have various values, e.g. somewhere between 10 and 100. It could be chosen so that on the one hand, there are enough gesture samples to provide a reliable basis for the adaptation, while on the other hand, the adaptation is not postponed too long if the gesture recognition system e.g. fails to correctly recognise certain gestures too often.

[0023] The invention further provides an online trainable system comprising a gesture recognition system, a sensor system and an active system, which online trainable system is configured to perform at least one sampling cycle. Each sampling cycle includes receiving sensor data representing a first gesture by a user from the sensor system and at least temporarily storing the sensor data; based on a mapping relation, interpreting the first gesture as corresponding to a first action to be performed by the active system, which first action is selected from a predefined action set; the active system performing the first action; and if, within a predefined reaction time after beginning to perform the first action, a correcting user input is received indicating that the first gesture corresponds to a different, second action from the action set, storing the second action together with the sensor data as a gesture sample, wherein after receiving the sensor data, a score indicating a degree of correspondence between the first gesture and the second action is calculated, and the gesture sample is only stored if the score is above a predefined threshold. The system is further configured to, after storing at least one gesture sample, adapt the mapping relation based on at least one stored gesture sample.

[0024] The online trainable system, in this context, is simply any kind of system that comprises the gesture recognition system, the sensor system and the active system and is adapted as described above. In particular, this may be a system in a car, a multimedia device like a tablet or a smartphone. All other terms have been described above with respect to the inventive method and therefore will not be explained again. It is understood that the first gesture is interpreted by the gesture recognition system.

[0025] Preferred embodiments of the inventive system correspond to those of the inventive method.

Brief Description of the Drawings

[0026] Further details and advantages of the present invention will be apparent from the following detailed description of not limiting embodiments with reference to the attached drawing, wherein:

Fig.1 is a schematic view of a first embodiment of an online trainable system according to the invention;

Fig.2 is a flowchart illustrating a first embodiment of an inventive method for training a gesture recognition system;

Fig.3 is a flowchart illustrating a second embodiment of an inventive method for training a gesture recognition system; and

Fig.4 is a schematic view of a second embodiment of an online trainable system according to the invention.

Description of Preferred Embodiments

[0027] Fig.1 shows a part of a car 20 with a first embodiment of an online trainable system 1 according to the invention. Among various other systems, the car 20 comprises an infotainment system 4 that can be operated basically in two different ways. On the one hand, a user 30 can operate the infotainment system 4 manually via a touchscreen 5. It is understood that pushbuttons or other sensor fields could be used instead of the touchscreen 5. On the other hand, the user 30 can operate the infotainment system 4 by gestures, which are recorded by a camera 2 that is installed in front of the user 30. The camera 2 is in communication with a gesture recognition system 3, which in turn is connected to the infotainment system 4.

[0028] The camera 2 is a sensor system that provides sensor data x corresponding to a gesture G_1, G_2 by the user 30 to the gesture recognition system 3. Depending on the gesture G_1, G_2 recognised by the recognition system 3, it can output a command to the infotainment system 4 to perform a certain action a_o-a_m. The recognition system 3, which of course is at least partially software-implemented, comprises a gesture recognition module $M_{A,\theta}$, which is an implementation of a machine learning based method for gesture recognition (e.g. an artificial neural network) with trainable parameters θ. When receiving sensor data x as input, the gesture recognition module $M_{A,\theta}$ outputs an action or a subset of possible actions a_o-a_m from an action set S=$\{a_0, a_1, \ldots, a_m\}$. The latter contains all possible actions a_o-a_m that could be intended by gestures G_1, G_2 and that are supported by the gesture recognition system 3. Here and in the following, a_0 denotes the neutral action, corresponding to no command being sent to the infotainment system 4. This can mean e.g. that the gesture recognition system 3 has detected a gesture but is not able to decode it. In general, one can say $M_{A,\theta}$ (x)∈S. In other words, $M_{A,\theta}$ represents a mapping relation from the space of gestures G_1, G_2 to the action set $\{a_0, a_1, \ldots, a_m\}$.

[0029] No two drivers 30 will perform a certain gesture G_1, G_2 in exactly the same manner. Therefore, there is some room for misinterpretation of a certain gesture G_1, G_2, so that the gesture recognition system 3 may be unable to identify any specific gesture G_1, G_2 (resulting in output of the neutral action a_0) or may misinterpret the gesture G_1, G_2 so that the infotainment system 4 performs an action a_o-a_m that is different from the one intended by the user 30. In order to better adapt to the gestures G_1, G_2 performed by different drivers 30, the online trainable system 1 is adapted to perform the method illustrated by the flowchart in fig. 2.

[0030] During operation, the system 1 performs a plurality of sampling cycles. At the beginning of the sampling cycle, sensor data x are received, which represent a first gesture G_1 by the user 30. These sensor data x are temporarily stored. In a next step, the gesture recognition system 3 applies the mapping relation represented by the gesture recognition module $M_{A,\theta}$, thereby interpreting the first gesture G_1 as corresponding to a first action a_x=$M_{A,\theta}$ (x). After that, the first action a_x is performed by the infotainment system 4. Now, within a reaction time T, a correcting user input may be received that corresponds to a different, second action a_y to be performed by the infotainment system 4. The correcting user input is hereinafter denoted as $act(T)$. It may either be a second gesture G_2 that is recorded by the camera 2 or it may be an input through the touchscreen 5. If no correcting user input related to at least one of $\{a_1, \ldots, a_m\}$ is detected within the reaction time T, then $act(T)=a_O$. If a correcting user input corresponding to one of $\{a_1, \ldots, a_m\}$ is detected within the reaction time T, this could indicate that the first gesture G_1 was misinterpreted by the gesture recognition system 3, namely in that the user 30 in fact intended to initiate the second action a_y with the first gesture G_1. Normally, it should be checked whether the correcting user input corresponds to a second action a_y that is actually different from the first action a_x.

[0031] In order to better assess this possibility, a score c is calculated that indicates a degree of correspondence between the sensor data x and the second action a_y. Herein, $c = E(act(T), x)$, where E is an estimator function. In particular, the estimator function E and the score c are based at least partially on the same criteria as the mapping relation represented by $M_{A,\theta}$. If c is greater than a predefined threshold c_{min}, it can be assumed that the first gesture G_1 corresponds to the second action a_y. Optionally, the user 30 can be asked to confirm this correspondence. This can be done by a voice or text output to the user 30 corresponding to e.g.: "Did you want to…?" or "Was your gesture intended to…?". The answer of the user 30 could be input e.g. via the touchscreen 5 or via a microphone (not shown in the figure). If the user 30 confirms, the second action is stored together with the sensor data x as a gesture sample $(x, act(T_x))$ to a database (X_{online}, Y_{online}). One could also say that the sensor data x are labelled by the second action a_y

[0032] There are a variety of possibilities how the estimator function could be calculated. One example would be

$$E(act(T), x) := d_O - dist(M_{A,\theta}(x), act(T))$$

where $dist()$ is a suitable distance function and d_O is a constant (which could be zero). The function could be designed in an optimal way, which means that gestures which are very similar could be rated differently from gestures which are not similar at all. Due to the fact that the number of possible actions $\{a_0, \ldots, a_n\}$ of the system is limited, the distance function could be realized via a matrix, e.g. $dist(a_i,a_j)=a_i^T D\, a_j$, where $D \in \mathbb{R}^{n+1\times n+1}$. In this context, $dist()$ is assumed to yield a higher value for dissimilar gestures, so that $E(act(T), x)$ as defined above has its maximum value for identical gestures. It should be understood that while $dist()$ is formally a function of two actions, its function value is based on the distance (or similarity, respectively) of the "ideal" gestures associated with these actions.

[0033] The gesture recognition module $M_{A,\theta}$, for instance in case of neural network, can provide a score indicating a similarity or probability of correspondence for each possible action $\{a_0 \ldots, a_n\}$. This score could be calculated in a similar way or even identical to the estimator function E. The one with the highest score is considered as the first action a_x. Thus, in case a second action a_y is manually performed by the user 30 within the reaction time T, the system 1 could evaluate the original score resulting from $M_{A,\theta}$ for this second action a_y and compare it to a threshold value.

[0034] As a variant, a mechanism could be implemented in $M_{A,\theta}$ according to which its parameters θ are slightly changed, either deterministically or randomly, but in a way that the classification performance stays in certain range that is considered acceptable. Then, each sensor data x would be run through several versions of $M_{A,\theta}$, say $M_{A,\theta,1}, \ldots, M_{A,\theta,l}$ (l being the total number of versions), each of which yields a score for all possible actions $\{a_0 \ldots, a_n\}$. Then, if the scores for the second action a_y (performed by the user 30) are in mean above a certain threshold, the system 1 could accept the gesture to correspond to that action. Additionally to the average score, the system 1 could also take into account the variance of the scores and only accept the gesture as corresponding to the second action a_y if the variance is below another predefined threshold.

[0035] The system 1 may also take into account the duration of the gesture, e.g. assigning a lower score to shorter gestures, since they are more likely the result of random motion. Additionally or alternatively, the system 1 may take into account the distance of the user 30 from the camera 2 (or another sensor used to detect the gestures G_1, G_2). If the distance is great, this could impair the detection accuracy, wherefore a lower score could be assigned.

[0036] If a gesture e.g. of a hand is to be detected, a “skeleton” hand model could be fit to the image of the hand in the camera images. Then, trace and velocity of particular parts of the skeleton could then be determined. By comparison of theses trace and velocity values to “ideal” values stored for each action $\{a_0 \ldots, a_n\}$, the score c can be determined and/or the first action a_x can be identified. The ideal values correspond to an ideal gesture associated with the respective action.

[0037] As an example, the user 30 performs a first gesture G_1 that is detected by the system 1 and interpreted as a command to start a phone call. If, within the reaction time T, the user 30 rejects to start a phone call but instead mutes the radio, the system 1 can take this correcting user input as an indicator that the preceding gesture G_1 could have been intended to mute the radio. Then, if $E(a = mute\ radio, x)$ exceeds the threshold c_{min}, the system 1 could ask the user 30 to confirm this interpretation and, if he does, store $(x, mute\ radio)$ in (X_{online}, Y_{online}). The function E could for instance be a machine learning module dedicated to estimate if in certain wide sense a gesture-action pair fits to corresponding pairs that have been used to train the system.

[0038] The online trainable system 1 performs a plurality of sampling cycles until a predefined number N of gesture samples have been stored. Then, the mapping relation of the gesture recognition module $M_{A,\theta}$ is adapted by applying an online (re-)training function $R(M_{A,\theta}, X, Y)$ of the gesture recognition module $M_{A,\theta}$ with current parameters θ. This function $R(M_{A,\theta}, X, Y)$ returns new parameters $\tilde{\theta}$ defining the trained model. After the adaptation has been completed, the previously recorded gesture samples can be deleted and the method can resume performing a plurality of sampling cycles.

[0039] The method can be described by the following pseudo code:

While lifetime of the system

If a signal x that indicates a gesture is detected

Perform $M_{A,\theta_0}(x)$

If $M_{A,\theta_0}(x) \neq act(T_x)$ **And** $act(T_x) \neq a_0$ **And** $E(act(T\_x), x) > c_{min}$

If driver confirms the correctness of the prediction $E(act(T_x), x)$

Write the pair $(x, act(T_x))$ to (X_{online}, Y_{online})

End if

End if

End if

If number of samples (X_{online}, Y_{online}) exceeds N

$\theta_0 = R(M_{A,\theta_0}, X_{online}, Y_{online})$

Clear (X_{online}, Y_{online})

End if

End while

[0040] As mentioned above, the confirmation by the user 30 is only optional.

[0041] An optional modification of the method is shown by the flowchart in fig. 3. This accounts for the possibility that the user 30 e.g. performs some random motion that leads to sensor data *x* and a first gesture G_1 which are interpreted to correspond to an action different from the neutral action a_O, while the user 30 did not intend to initiate any action at all. In such a case, a common scenario will be that the correcting user input corresponds to undoing the first action a_x. For example, the first action a_x could be "start a phone call", while the correcting user input corresponds to "end the phone call". If such a correcting user input is received within a relatively small reaction time T, this normally indicates that the user 30 did not want to initiate a phone call in the first place. This can optionally be confirmed by asking the user 30: "Did you not want to start a phone call?". Since the user 30 actually did not want to initiate any action with the first gesture G_1, the neutral action a_O is stored together with the sensor data x.

[0042] A second embodiment of an online trainable system 1 is shown in fig. 4. Here, the sensor system is a capacitive sensor 12 by which can be used to detect a gesture by a foot of the user 30. The capacitive sensor 12 generates sensor data X that can be received by a gesture recognition system 13. The gesture recognition system 13, in turn, can send a command to a trunk opener 14. Alternatively, the trunk opener 14 can be operated via a pushbutton 15. In this embodiment, the gesture recognition system 13 only has to distinguish between two actions, namely open/close and doing nothing, i.e. $\{a_0, a_1\} = \{do\ nothing, trigger\ trunk\ opening/closure\}$. Opening or closing can be triggered by a kick-like movement. If the gesture recognition system 13 receives sensor data x, but is not confident about whether this was an intended kick-like movement, then $M_{A,k}(x) = a_0$. If, however, the user 30 opens/closes the trunk within a reaction time T (via manual opening/closing with the pushbutton 15 or re-kicking), e.g. 5 seconds after detecting x and interpreting it as corresponding to a_0, $act(T_x) = a_1$. Then, if $E(a_1, x) > c_{min}$, the gesture sample(x, a_1) is stored for online training. Here $E(a_1, x)$ could be for instance derived from the confidence by which a_1 has been rejected by M_{A,θ_0}. That is if the machine learning algorithm underlying M_{A,θ_0} assigns a certain probability to a_1.

List of Reference Symbols

1	online trainable system
2	camera
3, 13	gesture recognition system
4	infotainment system
5	touchscreen
6	photodiode
12	capacitive sensor
14	trunk opener
15	pushbutton
20	car
30	user
G_1, G_2	gesture

Claims

1. A method for training a gesture recognition system (3, 13), comprising
   - performing at least one sampling cycle, including:
     - receiving, from a sensor system (2, 12), sensor data (x) representing a first gesture (G1) by a user (30) and at least temporarily storing the sensor data (x),
     - based on a mapping relation, interpreting the first gesture (G1) as corresponding to a first action (a_x) to be performed, which first action (a_x) is selected from a predefined action set (S),
     - performing the first action (a_x), and
     - if, within a predefined reaction time (T) after beginning to perform the first action (ax), a correcting user input is received indicating that the first gesture (G1) corresponds to a different, second action (a_y) from the action set (S), storing the second action (a_y) together with the sensor data (x) as a gesture sample, wherein after receiving the sensor data (x), a score (c) indicating a degree of correspondence between the first gesture (G_1) and the second action (a_y) is calculated, and the gesture sample is only stored if the score (c) is above a predefined threshold (c_{min})

     and
   - after storing at least one gesture sample, adapting the mapping relation based on at least one stored gesture sample.

2. A method according to claim 1, characterised in that the score (c) is calculated based on at least one criterion of the mapping relation.

3. A method according to claim any of the preceding claims, characterised in that if, within the reaction time (T), a correcting user input is received corresponding to the second action (a_y), the second action (a_y) is stored together with the sensor data (x) as the gesture sample.

4. A method according to any of the preceding claims, characterised in that the action set (S) includes a neutral action (a_O), corresponding to inactivity.

5. A method according to claim 3, characterised in that if within the reaction time (T), a correcting user input is received corresponding to undoing the first action (a_x), the neutral action (a_o) is stored together with the sensor data (x) as the gesture sample.

6. A method according to any of the preceding claims, characterised in that before storing the gesture sample, the user (30) is asked to confirm that the first gesture (G_1) does not correspond to the first action (a_x).

7. A method according to any of the preceding claims, characterised in that before storing the gesture sample, the user (30) is asked to confirm that the first gesture (G_1) corresponds to the second action (a_y).

8. A method according to any of the preceding claims, characterised in that the correcting user input is sensor data representing a second gesture (G_2) received from the sensor system (2, 12).

9. A method according to any of claims 1 to 7, characterised in that the correcting user input is received from an input device (5, 15) independent of the sensor system (2, 12).

10. A method according to any of the preceding claims, characterised in that after adapting the mapping relation, at least one stored gesture sample is deleted.

11. A method according to any of the preceding claims, characterised in that the mapping relation is adapted when a predefined number (N) of gesture samples have been stored.

12. An online trainable system (1) comprising a gesture recognition system (3, 13), a sensor system (2, 12) and an active system (4, 14), which system (1) is configured to perform at least one sampling cycle, including:

- receiving sensor data (x) representing a first gesture (G_1) by a user (30) from the sensor system (2, 12) and at least temporarily storing the sensor data (x),
- based on a mapping relation, interpreting the first gesture (G_1) as corresponding to a first action (a_x) to be performed by the active system (4, 14), which first action (a_x) is selected from a predefined action set (S),
- the active system (4, 14) performing the first action (a_x), and

- if, within a predefined reaction time (T) after beginning to perform the first action (a_x), a correcting user input is received indicating that the first gesture (G_1) corresponds to a different, second action (a_y) from the action set (S), storing the second action (a_y) together with the sensor data (x) as a gesture sample, wherein after receiving the sensor data (x), a score (c) indicating a degree of correspondence between the first gesture (G_1) and the second action (a_y) is calculated, and the gesture sample is only stored if the score (c) is above a predefined threshold (c_{min})

and, after storing at least one gesture sample, adapt the mapping relation based on at least one stored gesture sample.

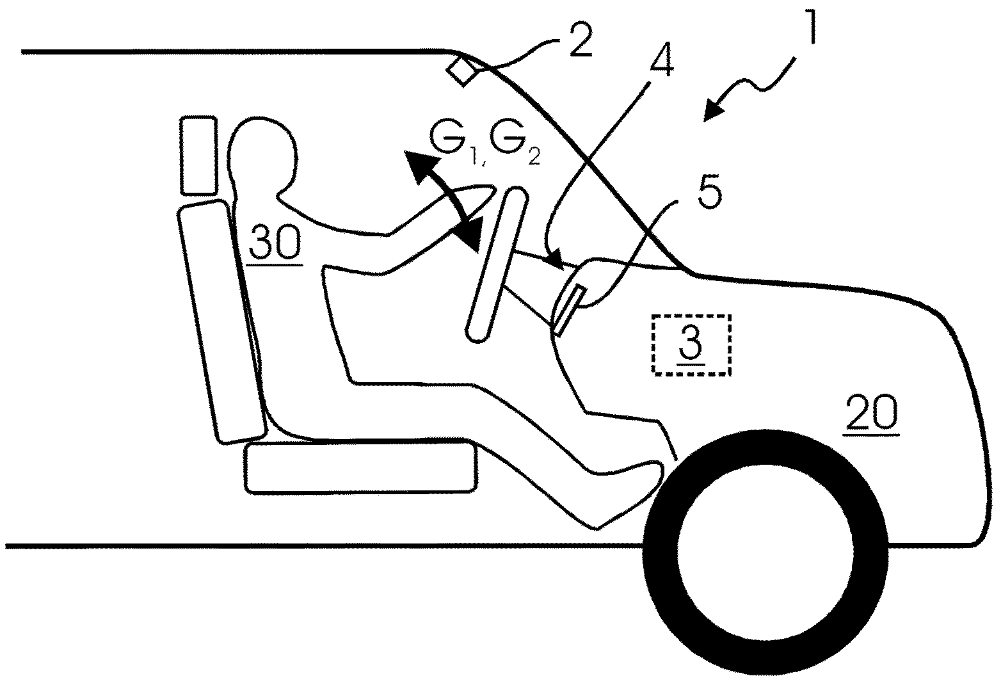

Fig. 1)

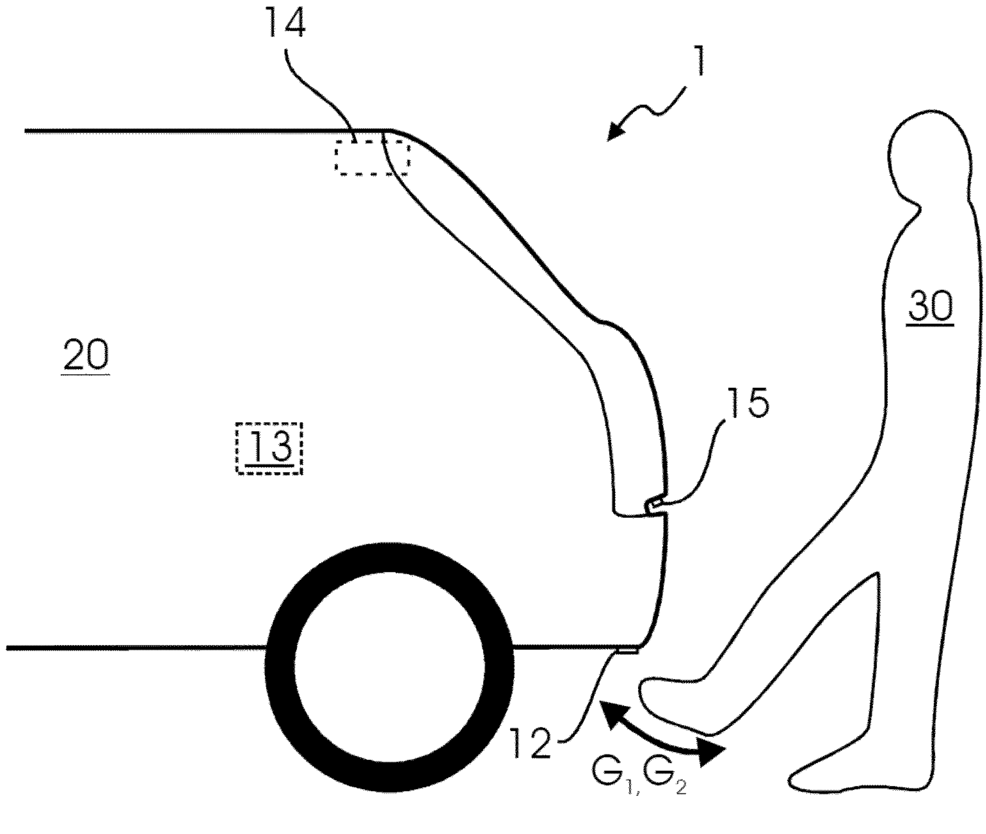

Fig. 4

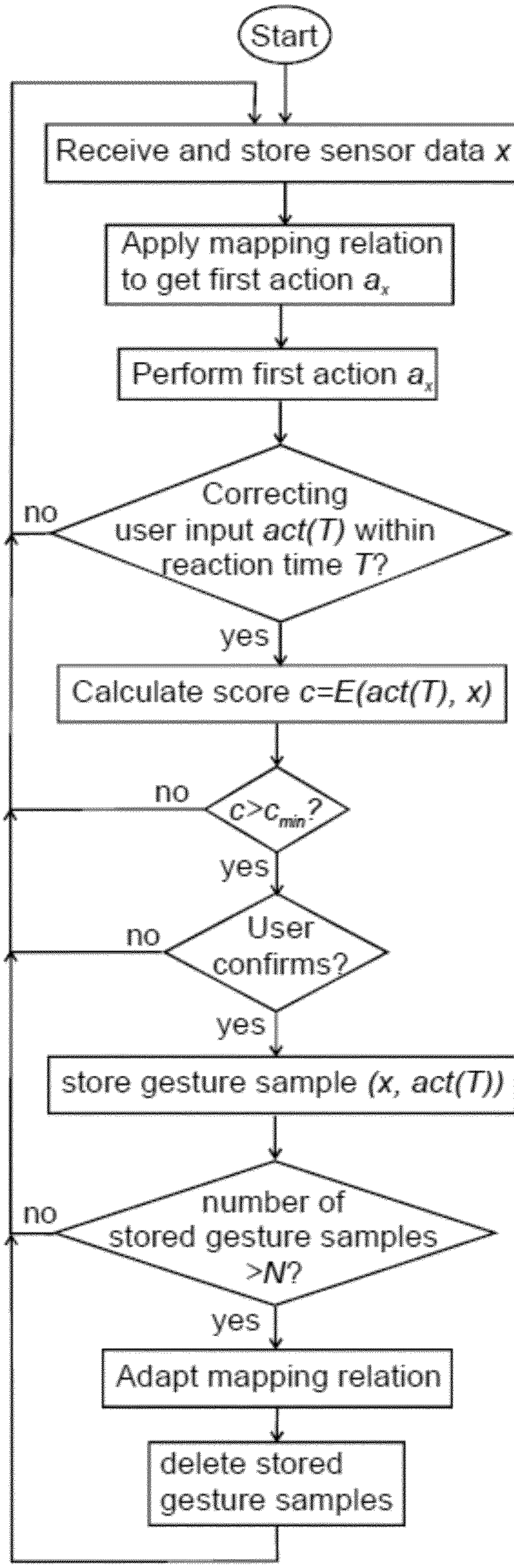
Start
Receive and store sensor data x
Apply mapping relation to get first action a_x
Perform first action a_x
Correcting user input act(T) within reaction time T?
no
yes
Calculate score c=E(act(T), x)
no
$c>c_{min}$?
yes
no
User confirms?
yes
store gesture sample (x, act(T))
no
number of stored gesture samples >N?
yes
Adapt mapping relation
delete stored gesture samples

Fig. 2)

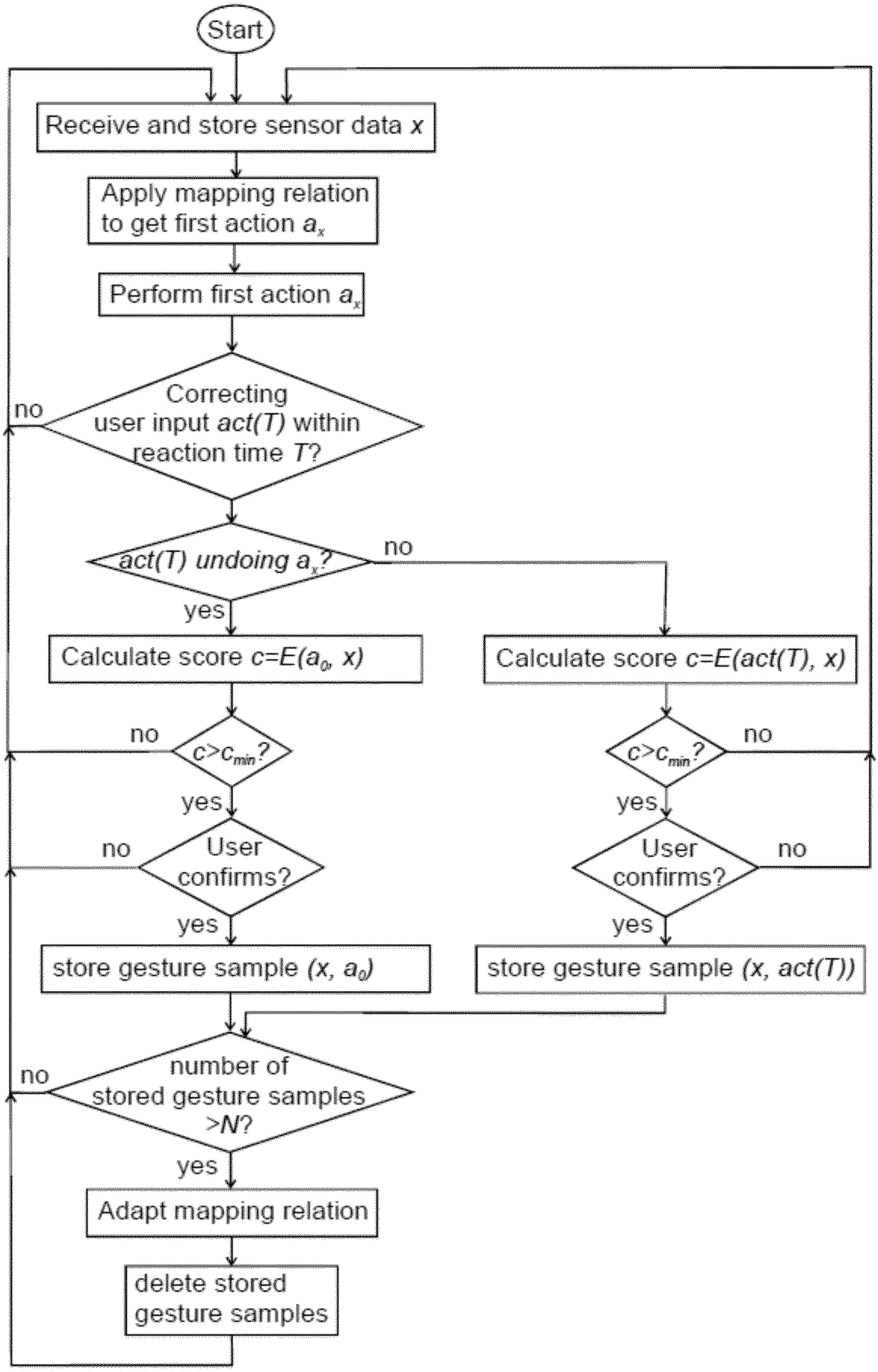
Start
Receive and store sensor data x
Apply mapping relation to get first action a_x
Perform first action a_x
Correcting user input act(T) within reaction time T?
no
act(T) undoing a_x?
no
yes
Calculate score $c=E(a_0, x)$
Calculate score $c=E(act(T), x)$
no
$c>c_{min}$?
yes
no
User confirms?
yes
store gesture sample (x, a_0)
$c>c_{min}$?
no
yes
User confirms?
no
yes
store gesture sample $(x, act(T))$
no
number of stored gesture samples $>N$?
yes
Adapt mapping relation
delete stored gesture samples

Fig. 3)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/067773

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F3/01
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)
G06F G06N B60W G06K B60K

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	WO 2015/153835 A1 (HONDA MOTOR CO LTD [JP]; EL DOKOR TAREK A [US]; KING JOSHUA [US]; CLUS) 8 October 2015 (2015-10-08) figures 1-19 paragraph [0033] paragraph [0035] - paragraph [0038] paragraph [0061] - paragraph [0078] paragraph [0081] paragraph [00149] - paragraph [00151] -----	1-12
X	EP 2 652 579 A2 (MICROSOFT CORP [US]) 23 October 2013 (2013-10-23) figures 1-21 paragraph [0033] - paragraph [0037] paragraph [0041] - paragraph [0042] paragraph [0049] paragraph [0081] - paragraph [0102] ----- -/--	1-12

[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
3 September 2019	13/09/2019

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	Roth, Lucia

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014/354527 A1 (CHEN ZHE [CA] ET AL) 4 December 2014 (2014-12-04) figures 1-7B paragraph [0013] - paragraph [0020] paragraph [0043] - paragraph [0078] -----	1-12

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2019/067773

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015153835 A1	08-10-2015	EP 3126934 A1	08-02-2017
		WO 2015153835 A1	08-10-2015
EP 2652579 A2	23-10-2013	CN 103262005 A	21-08-2013
		EP 2652579 A2	23-10-2013
		KR 20130129220 A	27-11-2013
		US 2012154293 A1	21-06-2012
		WO 2012083215 A2	21-06-2012
		WO 2012083223 A2	21-06-2012
US 2014354527 A1	04-12-2014	US 2014354527 A1	04-12-2014
		US 2019018500 A1	17-01-2019