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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2012 (27.12.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/175321 A1

(51) International Patent Classification:
G06T 7/20 (2006.01)

(21) International Application Number:
PCT/EP2012/060512

(22) International Filing Date:
4 June 2012 (04.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11305768.1 20 June 2011 (20.06.2011) EP
11306127.9 12 September 2011 (12.09.2011) EP

(71) Applicant *(for all designated States except US)*: **ALCATEL LUCENT** [FR/FR]; 3, avenue Octave Gréard, F-75007 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TYTGAT, Donny** [BE/BE]; Karperstraat 100, B-9000 Gent (BE). **LIEVENS, Sammy** [BE/BE]; Leeuweriklaan 7, B-2930 Brasschaat (BE). **AERTS, Maarten** [BE/BE]; Leurshoek 8, B-9120 Beveren-Waas (BE).

(74) Agent: **ALU Antw Patent Attorneys**; Copernicuslaan 50, B-2018 Antwerp (BE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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, ML, MR, NE, SN, TD, TG).

Published:

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

(54) **Title**: METHOD AND ARRANGEMENT FOR 3-DIMENSIONAL IMAGE MODEL ADAPTATION

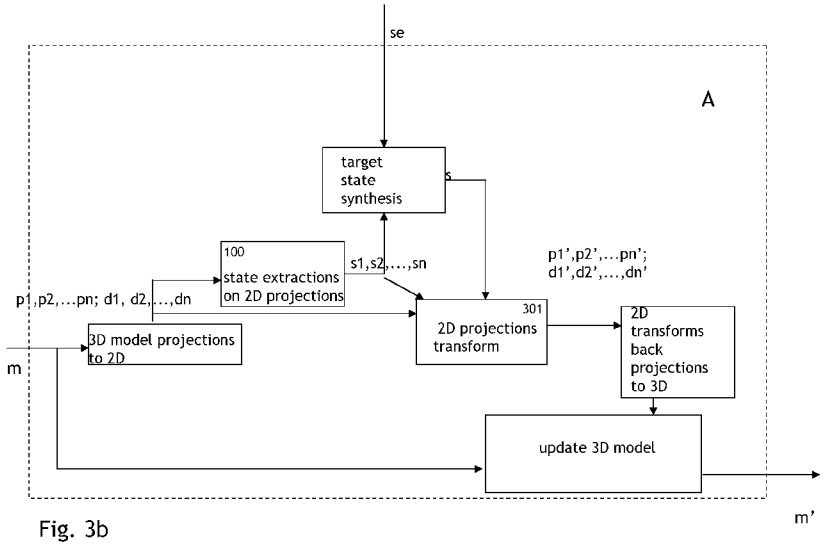


Fig. 3b

(57) **Abstract**: 1. Method for adapting a 3D model (m) of an object, said method comprising the steps of - performing at least one projection of said 3D model to obtain at least one 2D image model projection (p1) with associated depth information (d1), - performing at least one state extraction operation on said at least one 2D image model projection (p1), thereby obtaining at least one state (s1) - adapting said at least one 2D image model projection (p1) and said associated depth information (d1) in accordance with said at least one state (s1) and with a target state (s), thereby obtaining at least one adapted 2D image model (p1') and an associated adapted depth (d1') - back-projecting said at least one adapted 2D image model (p1') to 3D, based on said associated adapted depth (d1') for thereby obtaining an adapted 3D model (m').

WO 2012/175321 A1

METHOD AND ARRANGEMENT FOR 3-DIMENSIONAL IMAGE MODEL ADAPTATION

The present invention claims priority on the earlier filed European Patent Application 11305768 and relates to a method for adaptation of a three-dimensional, which, during the remainder of the text will be abbreviated by 3D, image model.

3D model adaptation is usually done in a manual way, which is generally not desirable. Another way to adapt a 3D model makes use of state adaptation, which concerns the adaptation of the 3D model in order to comply with a certain state . The state affects the 3D position of the shape and/or the appearance such as the texture of certain parts or features of the model. Again a major problem with present techniques for 3D model state adaptation is that the number of features to be adapted in 3D is usually very high, such that again manual intervention is often required due to insufficient computing resources . Moreover state-of-the-art techniques are limited to using rigged models, which presents a severe limitation for use in dynamic systems where models can be learned such that their shape can also vary during the learning process.

It is therefore an object of embodiments of the present invention to present a method and an arrangement for 3D image model adaptation, which can be used fully automatically and enables using dynamically adaptable models.

According to embodiments of the present invention this object is achieved by a method for adapting a 3D model of an object, said method comprising the steps of

- performing at least one projection of said 3D model to obtain at least one 2D image model projection (p1) with associated depth information (d1),

- performing at least one state extraction operation on said at least one 2D image model projection (p1), thereby obtaining at least one state (s1)

- adapting said at least one 2D image model projection (p1) and said associated depth information in accordance with said at least one state (s1) and with a target state (s), thereby obtaining at least one adapted 2D image model (p1') and an associated adapted depth (d1')

- back-projecting said at least one adapted 2D image model to 3D , based on said associated adapted depth (d1') for thereby obtaining an adapted 3D model (m').

By adapting the state of at least one 2D projection and its associated depth information of a 3D image model, less computing resources are used, therefore obviating the need for manual intervention in the process. The back-projection to 3D ensures that the 3D model itself is adapted as realistically as possible.

In an embodiment the adapted 3D model (m') is further determined based on the initial 3D model (m) information.

This enable a smooth morphing of the adapted model.

In another embodiment the target state (s) is determined by externally imposed restrictions.

This may e.g. comprise high level information with respect to the form of a nose, color of the eyes, etc.

In another embodiment the target state (s) is obtained from the state (se) of an external image input (IV).

This allows a 3D model to smoothly adapt to the changing features of e.g. an object on a live video, or to resemble this object as present on a still image, as the target state will be obtained by combining the state (se) of said external image input (IV) with said at least one state (s1) .

In a preferred variant said external image input (IV) comprises a 2D image input and one of the at least one 2D projections of said 3D model is performed in accordance with a virtual camera deduced from said external image input (IV) .

This is useful for obtaining an optimum relationship between the external image input and the 3D model.

In yet another variant the external image input may comprise a 2D+disparity input, with which is meant that both 2D as well as disparity information is externally provided e.g. by a stereoscopic camera. Depth information can then be directly derived from this disparity information by means of the formula depth x disparity = constant.

This allows to directly use the depth data from this input, for updating the associated depth.

The present invention relates as well to embodiments of an arrangement for performing this method, for image or video processing devices incorporating such an arrangement and to a computer program product comprising software adapted to perform the aforementioned or claimed method steps, when executed on a data-processing apparatus.

It is to be noticed that the term 'coupled', used in the claims, should not be interpreted as being limitative to direct connections only. Thus, the scope of the expression 'a device A coupled to a device B' should not be limited to devices or systems wherein an output of device A is directly connected to an input of device B. It means that there exists a path between an output of A and an input of B which may be a path including other devices or means.

It is to be noticed that the term 'comprising', used in the claims, should not be interpreted as being limitative to the means listed thereafter. Thus, the scope of the expression 'a device comprising means A and B' should not be

limited to devices consisting only of components A and B. It means that with respect to the present invention, the only relevant components of the device are A and B.

During the whole of the text two-dimensional will be abbreviated by 2D, while, as previously mentioned, three-dimensional will be abbreviated by 3D..

The above and other objects and features of the invention will become more apparent and the invention itself will be best understood by referring to the following description of an embodiment taken in conjunction with the accompanying drawings wherein:

Figs. 1a-b show a first variant of the method and apparatus ,

Figs.2a-b schematically show the geometrical model involved in embodiments of the invention

Figs. 3a-b show a second variant of the method,

Figs. 4a-b show a third , resp. fourth embodiment of the method,

Figs. 5a-c clarify the different steps as performed by the embodiment of Fig. 3a in case of an additional 2D video input ,

It should be appreciated by those skilled in the art that any block diagrams herein represent conceptual views of illustrative circuitry embodying the principles of the invention. Similarly, it will be appreciated that any flow charts, flow diagrams, state transition diagrams, pseudo code, and the like represent various processes which may be substantially represented in computer readable medium and so executed by a computer or processor, whether or not such computer or processor is explicitly shown.

Fig. 1a shows the steps as performed by a first variant of the method for adapting a 3D model, denoted m.

In a first step a projection of the 3D model is performed to 2D. The parameters for this projection, are the one used according to the well-known pinhole camera model as is for instance described in chapter 6 of the tutorial handbook "Multiple View Geometry in computer vision" by Richard Hartley and Andrew Zisserman, Cambridge University Press, second edition 2003, ISBN 0521 54051 8 .

This thus concerns the projection of points in a 3D space onto a plane, via a central "pinhole". In this model the plane corresponds to the projection plane of the camera, with the pinhole corresponding to the diafragma opening of the camera, often also denoted as the camera center. The result of the projection step is denoted p1, d1, with p1 indicating the 2D projection itself, which can be represented by a 2D matrix of pixel values containing color information, and with d1 indicating the projection depth map, which may also be represented by a 2D matrix of the associated depth values. These associated depth values are calculated from the original depth values and the camera position according to well known equations which will also be given in a later paragraph.

Alternatively the projection and the depth map can be represented within one large 2D matrix, wherein, for each projected pixel, both color information and associated depth information, is present in the corresponding matrix row and column.

The projection itself is schematically illustrated in Fig. 2a, showing a point A with 3 space coordinates x_A, y_A and z_A, with respect to an origin O defining these coordinates via three axes x,y,z defining a reference coordinate system. A pinhole camera is denoted by its camera center position C with

coordinates x_C, y_C and z_c with respect to this same reference origin and reference coordinate system. A projection of point A is made on a projection screen associated to this camera, and being denoted S. the projection to this screen via pinhole C of point A, is denoted p(A) with associated coordinates (x_{pA},y_{pA}) . These coordinates are however defined with respect to two-dimensional axes x_p and y_p as defined within this projection plane S.

In order not to overload Fig. 2a it is assumed here that the camera is not rotated with respect to the three reference axes x,y,z. . However well –known formula's exist also for this more general case, and these are used in embodiments according to this invention for the calculation of the projections and associated depth maps. These rotations of the camera are denoted θ_x , θ_y , θ_z for respectively denoting the rotation of the camera center around the x,y,z axis, as schematically denoted in Fig. 2b, where only these rotations are shown, for cases where the original O is coinciding with the camera center C.

In the most general case C may thus both be translated and rotated with respect to the reference origin O and the reference axes x,y,z.

In embodiments according to the present invention the projection of a 3D model then will consist of the color or texture information of the projected 3D points of this model, as long as these are falling within the contours of the screen area S, and as long as they are not occluded by another projection of another 3D point of this model. Occlusion indeed occurs almost inherently with all 2D projections of a 3D object, and relates to the fact that more than one 3D point of this model will be projected to the same 2D point on the projection

The depth map associated to this projection will then consist, for each of the projected pixels p(A), of their respective relative depth value, with respect to the position of the camera. This is denoted as

$$d_z = \cos\theta_x \cdot (\cos\theta_y \cdot (a_z - c_z) + \sin\theta_y \cdot (\sin\theta_z \cdot (a_y - c_y) + \cos\theta_z \cdot (a_x - c_x))) - \sin\theta_x \cdot (\cos\theta_z \cdot (a_y - c_y) - \sin\theta_z \cdot (a_x - c_x)) \quad (1)$$

with θ_x , θ_y , θ_z indicating the respective rotations of the camera around the reference axes as indicated on Fig. 2b,

with a_x, a_y and a_z representing the coordinates of a point a in a reference coordinate system,

c_x, c_y and c_z representing the coordinates of the camera center c in this reference coordinate system, and

with d_z representing the associated depth of a point a with respect to the camera center c

In case where there is no rotation of the camera with respect to the reference coordinate system x,y,x in the reference origin O, these rotation angles are zero, such that the equation (1) will be reduced to

$$d_z = a_z - c_z \quad (2)$$

which, using the notations as in Fig. 2a , corresponds to

$$d(A) = z_A - z_c \quad (3)$$

as is also indicated in Fig. 2a

In general the projection is selected such that the features of the 3D model which are to be adapted in 3D will be part of the projection, at a sufficiently high resolution or such that they optimally fill the projection image.

This may be done heuristically, by trying a set of predetermined projection positions, and selecting the one giving the best results.

In another embodiment this can be further determined via an intermediate step wherein the 3D surface of the model will be approximated by means of 3D triangles. In general only the parts of the model related to the features to be adapted will then be approximated by such 3D triangles. For each of these triangles the normal related to the perpendicular direction is determined. For an ideal projection the direction of this normal should be 180 degrees with respect to the direction of the camera to this triangle. For each camera position, the summation, over all triangles, of the cosine of this angle between the normal on the respective triangle and the direction of the camera to the center of the triangle, should then be minimal. By calculating this summation over a number of possible camera positions, and selecting the position yielding the minimum value for this summation, an optimum direction can be calculated. Alternatively the minimization problem itself can be solved, such as to determine the optimum camera direction.

Of course a lot of other techniques can be used, as is well known by a person skilled in the art.

In a next step the state is extracted from this projection. With state a configuration of object features is meant, which features are themselves represented by a set of values. These values may thus describe the possibly variable properties or features of the object. This set of values can be arranged into a vector, but other representations for such a state are of course also possible. State extraction thus means that state parameters, for representing the state of an object of an image , in this case being a projection of a 3D model, are determined. This can be done via some calculations based on the 3D model information, as will be shown in the examples described in a further paragraph,

or by using more general methods e.g. first involving a step of recognition/ detection of the object under consideration, possibly but not necessarily by performing segmentation operations, followed by a further in depth analysis of the thus recognized/detected object.

However in most embodiments according to the invention the 3D model itself is already known such that the state extraction can be seriously reduced to calculations based on the state of the 3D model. In case this 3D state relates to coordinates of certain features, which can be facial features in the case of a 3D model of a human head the 2D projections of these 3D points may immediately lead to the state parameters of the 2D images.

In case the state of the 3D model is not yet known, the earlier described recognition step may be followed by a further analysis e.g. involving usage of the Active Appearance Model, abbreviated by AAM. This allows, e.g. in case of a human head as object model to be updated , the determination of the shape and appearance of facial features on the 2D projected image via a fit with a 2D AAM internal shaping model. It may start with comparing the 2D projection with the starting value of a 2D AAM model, which AAM model itself is then further gradually altered to find the best fit. Once a good match is found, the parameters such as face_expression_1_x, face_expression_1_y, etc. thus determined based on this AAM adapted model are output.

In fig. 1a the state of the projection image is denoted s1, and this is used during a target state synthesis step. The target state s is obtained from the state s1 of the 2D projection, and from external state information. This external state information, denoted se, may have been determined beforehand, either offline e.g. from a still image input, or based on other descriptive information, e.g. high level semantic information with respect to e.g. the shape of a nose or

colors of the eyes, facial expressions etc. In this case this external state information may also be stored beforehand within a memory.

Alternatively this external state information se can be determined "on the fly" e.g. based on changing external video image input data, which can thus rapidly change over time. In such situations the external state se will generally be determined on successive frames of a video sequence.

The external state information is used together with the state s1 of the 2D projection for obtaining the target state.

Methods for determining the target state, denoted by s in Fig. 1a, out of the input state s1 and se, may comprise performing a weighted combination of the value of s1 and se, with the weights reflecting the confidence of the states, which confidence levels themselves were determined during the state extraction itself. For the aforementioned example of the AAM method for determining the s1 parameters, parameters identifying the matching result can then e.g. be selected as such confidence measures.

Another method for determining the target state may simply consist of selecting e.g. se, which option can be preferred in case a check of the result of the interpolation or weighted combination as explained in the previous example, of the different states, indicates that such interpolated result is lying outside predetermined limits.

Specific implementations for the determination of the state and target states will be further described during the description of the embodiments of Figs. 4a-b.

Upon determination of the target state, denoted s in Fig. 1a, the 2D projection p1 as well as the associated depth map d1 will be transformed in accordance with the target state s. In an example a method making use of triangles for representing e.g. facial features may be used . By means of

interpolating distances as defined by these triangles, and attributing features to the pixels as these new positions, which were previously attributed to the pixels at their previous position, an image transform may result. Such method is very useful in case a lot of such triangles are used.

In a similar method the updated 2D coordinates of the pixels of the projection images, associated to the features, will be calculated in accordance with the new state. The color and texture information of pixels lying in between triangles defined on the original 2D projection, will be attributed to the pixels lying in between the triangles at these new positions in the updated images. If thus two points on the 2D projection have internal coordinates (100,100) and (200,200), and these will be transformed to coordinates (50,50) and (100,100) on the transformed projections, the color of original pixel at coordinate (150,150) will be attributed to the pixel in the transformed image at coordinate (75,75) .

Another more detailed implementation will be further described when describing Figs. 4a-b.

The adapted 2D projection is denoted p1'.

In parallel also the associated depth values of the associated depth map are adapted in accordance to the target state. In some embodiments the target state determination directly involves the calculations of adapted depth values, for some of the pixels of the projection. Adaptation of the other depth values in accordance with the target state may then also take place via an interpolation between the already calculated adapted depth, as was explained in the previous paragraph with respect to the adaptation of the color values for the adapted projected pixels.

The adapted depth map is denoted d1' .

Based on the transformed depth map and transformed 2D projection, which generally includes the adapted 2D image model, a re-projection or back-projection to 3D can be performed, using the reverse transformations as these used during the 3D to 2D projections themselves, but now using the adapted associated depth values for each 2D pixel of the adapted projection image.

The result of this back-projection is denoted p3d_1.

In some cases the back-projected points in 3D are sufficient for forming an updated 3D model.

In other embodiments the back-projection to 3D is merged with the original 3D model m, to obtain the updated or adapted 3D model m'.

Fig. 1b shows an arrangement A for performing an embodiment of the method.

Fig. 3a shows a variant embodiment wherein more than 1 projection is performed from the initial 3D model m. The projections themselves may be selected in accordance with the form and shape of the model, and the amount of occlusions which occur by selection of a first projection, or using one of the methods as previously described for the determination of the projection parameters itself. A possible implementation can thus be based on approximations of the 3D surface which is to be modeled by means of a set of triangles in 3D. For each of these triangles the perpendicular direction is calculated. This may be represented by a 3D "normal" vector pointing outside the 3D model body. By calculating the difference between this 3D vector and the camera projection direction, a simple way for determination of occlusion is obtained, as for non-occluded surfaces, the projection direction should be opposite to the normal vector. As such the camera projection can be tuned, and it may thus also turn out that, for obtaining a sufficiently good projection, thus with sufficient resolution, of all features to be modeled, that several projections

may be needed. Alternatively, a default number of 3 predetermined projections may also be used, alleviating a trial and error calculation of the most optimum camera position.

These different projections are denoted p1,p2 to pn, with associated depth maps d1,d2 to dn. Each of these projections is thus associated with a virtual camera with a certain position , rotation,and associated screen width and length, as denoted in Figs. 2a-b.

Each of these different projections p1 to pn will also undergo state extraction or operations, leading to respective determined states s1,s2 to sn. In some embodiments the states of these respective projections can be calculated, as earlier described, especially in these situations where the features to be adapted directly relate to the coordinates or pixel positions of the features under consideration.

These respective determined states s1 to sn are used as respective input, possibly but not necessarily together with external state input se, for determination of a target state s. This determination of the target state may comprise performing a weighted combination of the various input states, with the weights reflecting the confidence of the states, which confidence levels themselves were determined during the state extraction itself. For the earlier example of the AAM method for determining the s1 parameters, parameters identifying the matching result can then e.g. be selected as such confidence measures.

Another method for determining the target state may simply consist of selecting one of the input states, or of the external state which option can be preferred in case a check of the result of the interpolation or weighted combination as explained in the previous example, of the different states, indicates that such interpolated result is lying outside predetermined limits.

The target state s forms the basis of which the n respective projections and their respective associated depth maps are updated. The updated projections are denoted p1', p2' to pn', and the updated depth maps are denoted d1', d2' to dn' .

Each of these updated projections p1', p2' to pn' is then back-projected to 3D based on the updated depth map values associated to each 2D pixel in the projections. These back-projections are merged together with the original model to create an updated or adapted model.

Fig. 3b shows an embodiment of an arrangement for performing this variant method.

Fig. 4a describes an embodiment for adapting a 3D model of a head of a person. In this embodiment the state of this model relates to the expressions of the face, but in other embodiments the state may as well relate to colours of the hair, eyes, skin etc. The goal in this particular embodiment is to animate the 3D model using facial features provided by an input 2D video.

This input video is denoted IV on Fig. 3a. For each frame of the video the scale and orientation of the object are estimated with respect to these of the 3D model. This is preferred for determining a first projection , being related to the virtual camera viewpoint of the 3D model to a 2D plane, which projection should resemble as much as possible the 2D projection used in the camera capturing the 2D video. This particular choice of first projection is not needed as such, but may be beneficial for an easy update. For this particular projection, the projection of the 3D model to a 2D plane should thus use a virtual camera with associated projection parameters which resemble as close as possible those of the camera used for having taken the 2D images of the input video.

This calculation of these projection parameters is done in accordance with known techniques such as will be described here below :

Input to the process of determining the parameters for this virtual camera is a 3D database model of a human face and a live 2D video feed. As the 3D positions of the facial features of the 3D database model, the 2D positions of the facial features in the live video feed and the projection matrix of both the webcam and the virtual camera are known, these data should be sufficient to calculate the 3D position of the facial features of the face in the live video feed. If the 3D positions of the facial features in the live video feed are thus known, together with the 3D location of the corresponding facial features of the database model, the 3D transformation (translation and rotation) between the corresponding 3D positions can be calculated. Alternatively the 3D transformation (translation and rotation) needed on a virtual camera in order to capture the same 2D viewport of the 3D database model, as is seen in the live video feed can thus also be calculated. The minimal amount of feature points needed, for this calculation of transformation to be applied on the virtual camera, is 3. Because the human face isn't a rigid object due to the changing and different emotions, taking more facial features would require solving minimization problems. Therefore 3 stable points, e.g. the left edge of the left eye, the right edge of the right eye and the top of the mouth, are used. The 3D position of these 3 facial features in the database model, together with the 2D position of the corresponding facial features in the live video feed and the webcam projection matrix are next inputted to the well known Grunert's algorithm. This algorithm will provide the calculated 3D positions of these corresponding 3 facial features. This can in turn be used to move the virtual camera around the 3D database model in order to capture the same 2D view of the database model as is provided by the face in live video feed.

In some embodiments, as the one shown in Fig. 4a, it may be preferred to use yet another projection of the 3D model. This may be desirable in

case the first projection using camera parameters resulting in an optimum projection resemblance with the image of the video feed, still does not result in sufficient pixel data, e.g. when at the projection image a portion of the face is occluded by the nose .

This is illustrated on Fig. 5a, depicting in the left rectangle the video captured by a "real" camera of the "real" person, while the left part of the right rectangle shows the projection of the 3D model with the first virtual camera, denoted virtual camera 1. As can be observed, the projection of the 3D model by this virtual camera matches the projection conditions used by the "live" 2D camera. Yet still some pixels of the left part of the face are occluded by the nose . Therefore another projection, by another virtual camera is performed, this camera being denoted "virtual camera 2". Its parameters are determined based on the occluded pixels of the other camera position. This can e.g. be determined based on the intrinsic parameters such as focal point and the extrinsic parameters of the virtual cameras and on the knowledge of the 3D model. This information will enable to determine whether or not two voxels or 3D points of the features to be modeled of 3D model will be projected to the same pixel in a 2D projection. If this is the case, it is clear that occlusion will occur. Based on this information, another virtual camera position can then be calculated, allowing different projections at least for this voxel. By performing this check on all projected pixels, the presence of occlusion can be determined, and another virtual camera position and rotation can be determined based on this .

In another embodiment a number of predetermined virtual cameras can be used, or a selection out of these, for getting projections of the features of interest. Alternatively also a standard configuration of virtual cameras for providing respectively a front view, and two side views at 90 degrees may be

used, and dependent on which features are to be modeled, all projections, or a subset of them can be used.

In case only two projections are used, the result of this second projection is shown in the right part of the right rectangle of fig. 5a. Together with the projections p1 and p2, also associated depth maps are created, denoted d1 and d2. These indicate, for each 2D projected pixel, the relative depth, including rotational information by virtue of equation (1) with respect to the respective camera positions, as observed from the point of view of the respective virtual camera 1 or 2. The depth map for each of the two projections is denoted in the bottom figures of the right rectangle.

In a next step the state is to be extracted on both projections p1 and p2 as well as on the successive frames of the input video. As in this embodiment the state relates to the facial expressions, these are thus to be characterized. Features relating to these facial expressions are extracted both on the successive frames on the input video as on the 2D projections using state of the art techniques such as the aforementioned AAM technique. It is also possible to calculate the states of the projections as earlier explained based on the 3D state of the model and on the corresponding voxel projections. This is shown in Fig. 5b, in the left rectangle indicating positions of different pixels of edges of the mouth and eyes on the live 2D frame. These positions of these same features are thus also determined on the projections. In the right part of Fig. 5b, this is only shown for projection p1, but it is clear that this also takes place on projection p2, which is not shown on this figure in order not to overload the drawing. In this particular embodiment the respective states correspond to the positions of the pixels associated to these features as present on p1, p2 and on an input frame. These states are respectively denoted s1, s2 and se. As only p1 is shown on Fig. 5b, also only s1 is shown. These 3 states are used for determining the target

state, which in this embodiment corresponds to the state se. While in this embodiment the respective states s1 and s2 are thus not used for the determination of the target state, these respective states s1 and s2 are nevertheless used during the transformation of the projections in accordance with the target state. This target state is thus also used for adapting the 2D projections p1 and p2 . For the virtual camera, corresponding to the "real" video camera, this adaptation can be easily done by replacing the pixel locations of the selected features by the corresponding pixel locations of these features as present in the video frame. By virtue of the selection of virtual camera 1 as mapping to the real camera, this can be done very easily . For adapting the 2D projection p2 obtained by the other virtual camera 2, a possible method involves calculating the locations of the adapted features of p2 first determined in 3D. This can be done based on the adapted projection p1' and adapted depth map d1'. This allows to determine, for these features which were visible on p1', to calculate their position in 3D. By using the projection parameters for the second projection, their corresponding positions on p2' can be identified . For the occluded features from p1 and p1', interpolation techniques may be used for calculating the adapted projections and adapted depth map.

Once the new locations of the key features for p1 and p2 are known, morphing techniques such as weigthed interpolation can be used for determining the color and depth of the pixels that were no key features.

The adaptations of the projection p1 is shown in the bottom figures of the right rectangle on Fig. 5b. It is clear that this projection is now adapted to the "laughing" face expression, as present on the input video frame of the left hand rectangle. This will also occur on projection p2 (not shown on Fig. 5b)

Both adapted projections p1' and p2' are then re-projected to 3D using the adapted depth maps and merged, to replace or update the old data.

The data for d1' may be calculated based on the approximation that the adapted depth equals the initial depth, thus that the initial depth d(A)for pixel A, related to a feature under consideration and with projection coordinate x_{pA},y_{pA} will now be attributed to the pixel with coordinate x_{pA}',y_{pA}' for x_{pA}' and y_{pA}' being the adapted coordinates of the feature under consideration.

In this respect it is to mention that all back-projections of the adapted 2D images should be consistent in the 3D domain. This basically means that when back-projecting a transformed feature that is visible in more than one 2D projected image, that this feature should be back-projected to the same 3D location from all projections . So if the corner of the mouth is transformed, and this corner of the mouth is present in several of these projections, all back-projected coordinates should be the same.

Say x_3d is a certain feature on the 3D object that is considered (e.g. the tip of the nose). x_3d is a vector with information (x, y, z, color). x_2dz is a certain feature in the 2D+Z domain, it is a vector containing information (x_2d, y_2d, depth, color).

The projection of 3D to 2D+Z according to a certain virtual camera c1 is modelled with the function p:

$$p(c1, x_3d) = x_2dz_c1$$

Lets now consider the state adapted 3D model. The expected 3D feature after state adaptation is called x'_3d. The 3D state transfer function is m_3d:

$$x'_3d = m_3d(x_3d)$$

this means that

$$x'_2dz_c1 = p(c1,x'_3d) = p(c1,m_3d(x_3d))$$

As the adaptation with respect to the state is performed on the projections, thus in the 2D+Z domain, the m_3d function is not available. This can be approximated by using a m_2dz function:

$$x''_2dz_c1 = m_2dz(c1, x_2dz_c1)$$

which can only be 3D state consistent if

$$x'_2dz_c1 = x''_2dz_c1$$

this means that the functions p(c1,m_3d) and m_2dz(c1) are effectively the same within the considered domains.

If this is the case, there is no issue and the aforementioned method can be used without any problems. If not, an additional step has to be implemented.

To take this into account a careful selection of the projection parameters could solve this issue from the beginning.

However, is this is not taken care of, such an inconsistency might be the case. One of the issues is that when using multiple 2D+Z sources to re-build the 3D model, that the back-projections of these sources need to "agree" on the state transfer function. When the functions are 3D state consistent, this is no

problem (as all 2dz functions actually implement specific 2dz version of the 3d state transfer function). When they are not 3d state consistent, we need to force their consistency , either via the "correct" 3d state transfer function, or an approximation thereof. This can be done for instance by choosing one reference 2DZ state transfer function, and projecting all other state transfer functions onto this reference:

x'_2dz_c1ref = m_2dz (c1ref, x_2dz_c1ref)

Now we consider m_2dz(c1ref) to be our reference 2dz state transfer function. We can build the other functions by moving via the 3D domain:

x'_3d = p_inv(c1ref, x'_2dz_c1ref) = p_inv(c1ref, m_2dz (c1ref, x_2dz_c1ref)

m_2dz (c2, x'_2dz_c2) = p(c2, x'_3d) = p(c2, p_inv(c1ref, m_2dz (c1ref, x_2dz_c1ref)))

Note that not all features from the object in 3D will have valid values after moving through p(c, x_3d). For instance points that are not within the virtual camera view , or points that are occluded by other features in the object. In order to have a consistent transfer function for such points, other reference cameras will be needed .

A second embodiment is a variant to the first that also involves the state adaptation of a 3D model of the face of a person; but as opposed to the previous embodiment it used a 2D+Z camera instead of a 2D camera e.g. using

a stereo camera or a time-of-flight camera such as the Microsoft Kinect. In this case we can use the facial feature points in 3D coordinates instead of 2D as external. We again take as many 2D+Z projections of the offline model as needed to cover all points that are modified by the live data and infer the state onto these projections. One can for example merge the data by using the morphing technique of the previous embodiment on the 'offline' 2D+Z data, but now also use the modified Z data for the feature points

In these embodiments we were able to reduce the problem of 3D state adaptation. Where we started from transferring state from one or multiple 2D images to a full 3D model, it is now reduced to transferring state from 2D to 2D+Z making these operations tractable for real-time applications.

While the principles of the invention have been described above in connection with specific apparatus, it is to be clearly understood that this description is made only by way of example and not as a limitation on the scope of the invention, as defined in the appended claims. In the claims hereof any element expressed as a means for performing a specified function is intended to encompass any way of performing that function. This may include, for example, a combination of electrical or mechanical elements which performs that function or software in any form, including, therefore, firmware, microcode or the like, combined with appropriate circuitry for executing that software to perform the function, as well as mechanical elements coupled to software controlled circuitry, if any. The invention as defined by such claims resides in the fact that the functionalities provided by the various recited means are combined and brought together in the manner which the claims call for, and unless otherwise specifically so defined, any physical structure is of little or no importance to the novelty of the claimed invention. Applicant thus regards any means which can provide those functionalities as equivalent as those shown herein.

CLAIMS

1. Method for adapting a 3D model (m) of an object, said method comprising the steps of

- performing at least one projection of said 3D model to obtain at least one 2D image model projection (p1) with associated depth information (d1),

- performing at least one state extraction operation on said at least one 2D image model projection (p1), thereby obtaining at least one state (s1)

- adapting said at least one 2D image model projection (p1) and said associated depth information (d1) in accordance with said at least one state (s1) and with a target state (s), thereby obtaining at least one adapted 2D image model (p1') and an associated adapted depth (d1')

- back-projecting said at least one adapted 2D image model (p1') to 3D , based on said associated adapted depth (d1') for thereby obtaining an adapted 3D model (m').

2. Method according to claim 1 wherein the adapted 3D model (m') is further determined based on the initial 3D model (m) information.

3. Method according to claim 1 or 2 wherein the target state (s) is obtained from externally imposed semantic information.

4 . Method according to claim 1 or 2 wherein the target state (s) is obtained from the state (PS) of an external image input (IV) .

5. Method according to claim 4 wherein said target state is obtained by combining the state (PS) of said external image input (IV) with said at least one state (s1) .

6. Method according to claim 4 wherein one of the at least one 2D projections of said 3D model are performed in accordance with a virtual camera deduced from said external image input (IV).

7. Method according to any of the previous claims 4 to 6 wherein the transform is performed on key features extracted from external said live video and the projected 2D images, and wherein new positions of the key features for said projections are determined based on the locations of the key features of the live video.

8. Arrangement (A1) adapted to perform the method according to any of the previous claims 1 to 7.

9. Image processing apparatus comprising an arrangement according to claim 8 .

10. A computer program product comprising software adapted to perform the method steps in accordance to any of the claims 1 to 6, when executed on a data-processing apparatus.

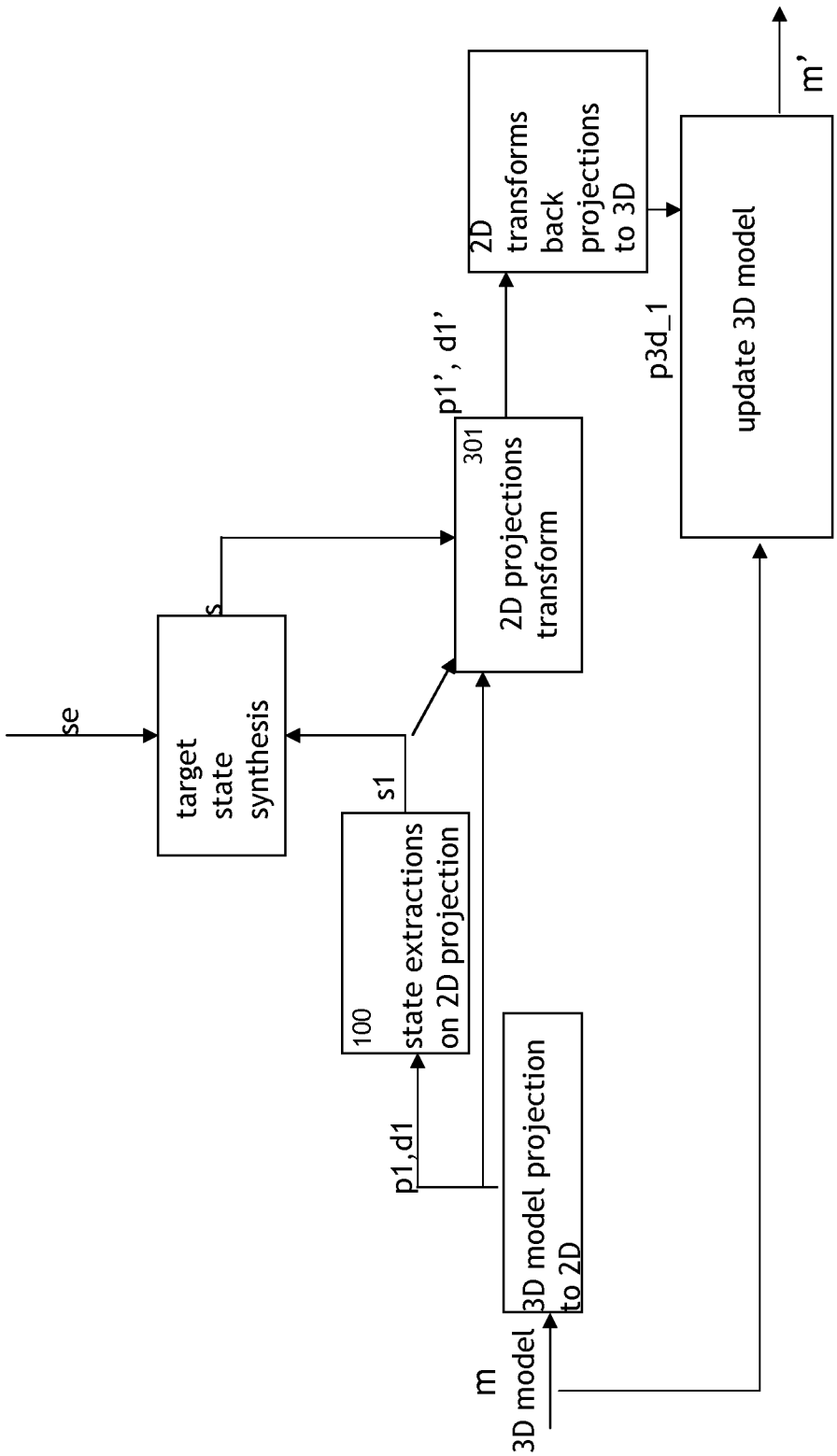

m
3D model
3D model projection to 2D
p1,d1
100
state extractions on 2D projection
s1
target state synthesis
se
s
2D projections transform
301
p1', d1'
2D transforms back projections to 3D
p3d_1
update 3D model
m'

Fig. 1a

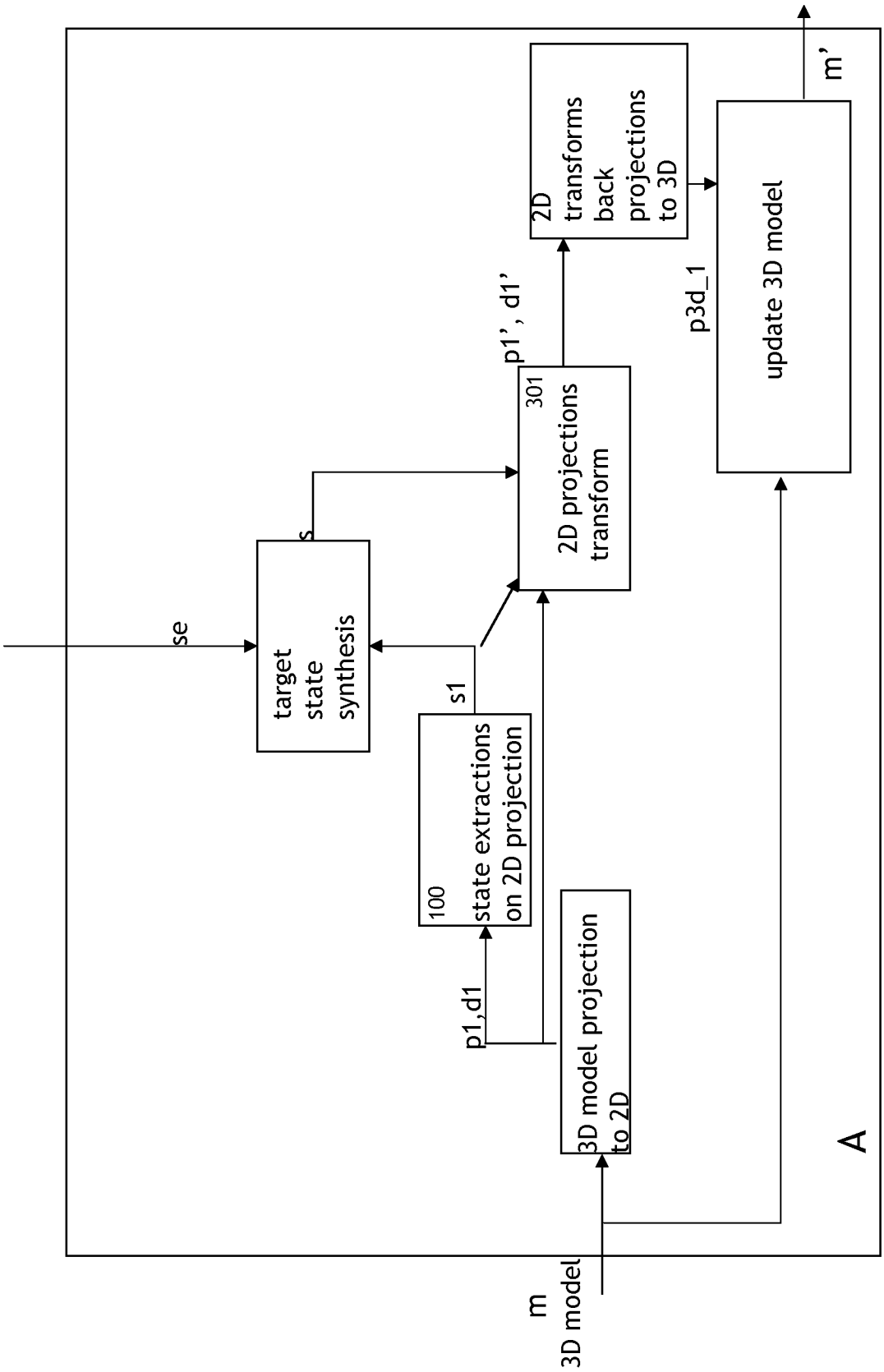

m
3D model
A
3D model projection to 2D
p1,d1
100
state extractions on 2D projection
s1
target state synthesis
se
s
301
2D projections transform
p1', d1'
2D transforms back projections to 3D
p3d_1
update 3D model
m'

Fig. 1b

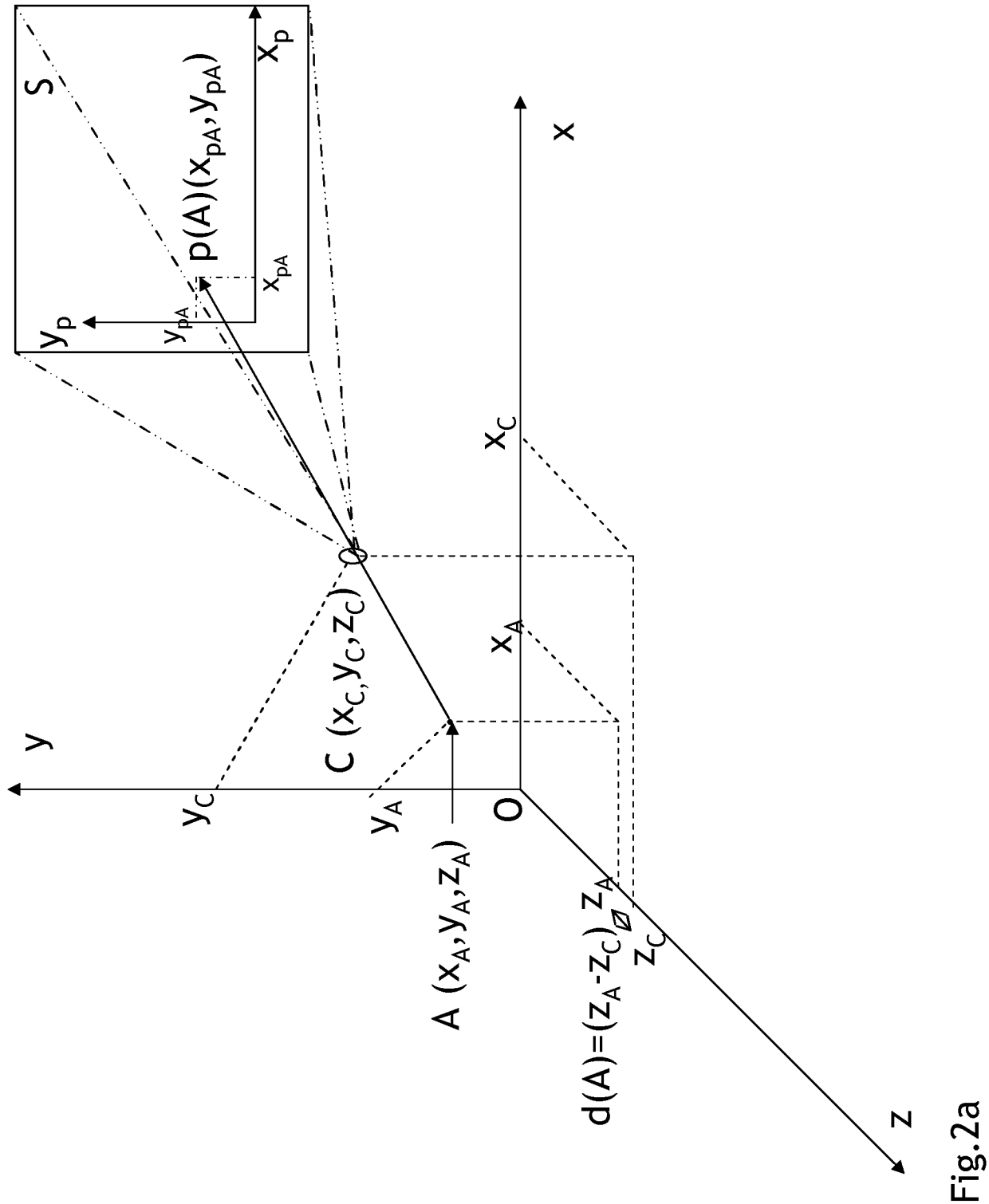

S
x_p
$p(A)(x_{pA},y_{pA})$
x_{pA}
y_p
y_{pA}
x
x_C
x_A
$C\ (x_C,y_C,z_C)$
y
y_C
y_A
O
$A\ (x_A,y_A,z_A)$
$d(A)=(z_A-z_C)$
z_A
z_C
z

Fig.2a

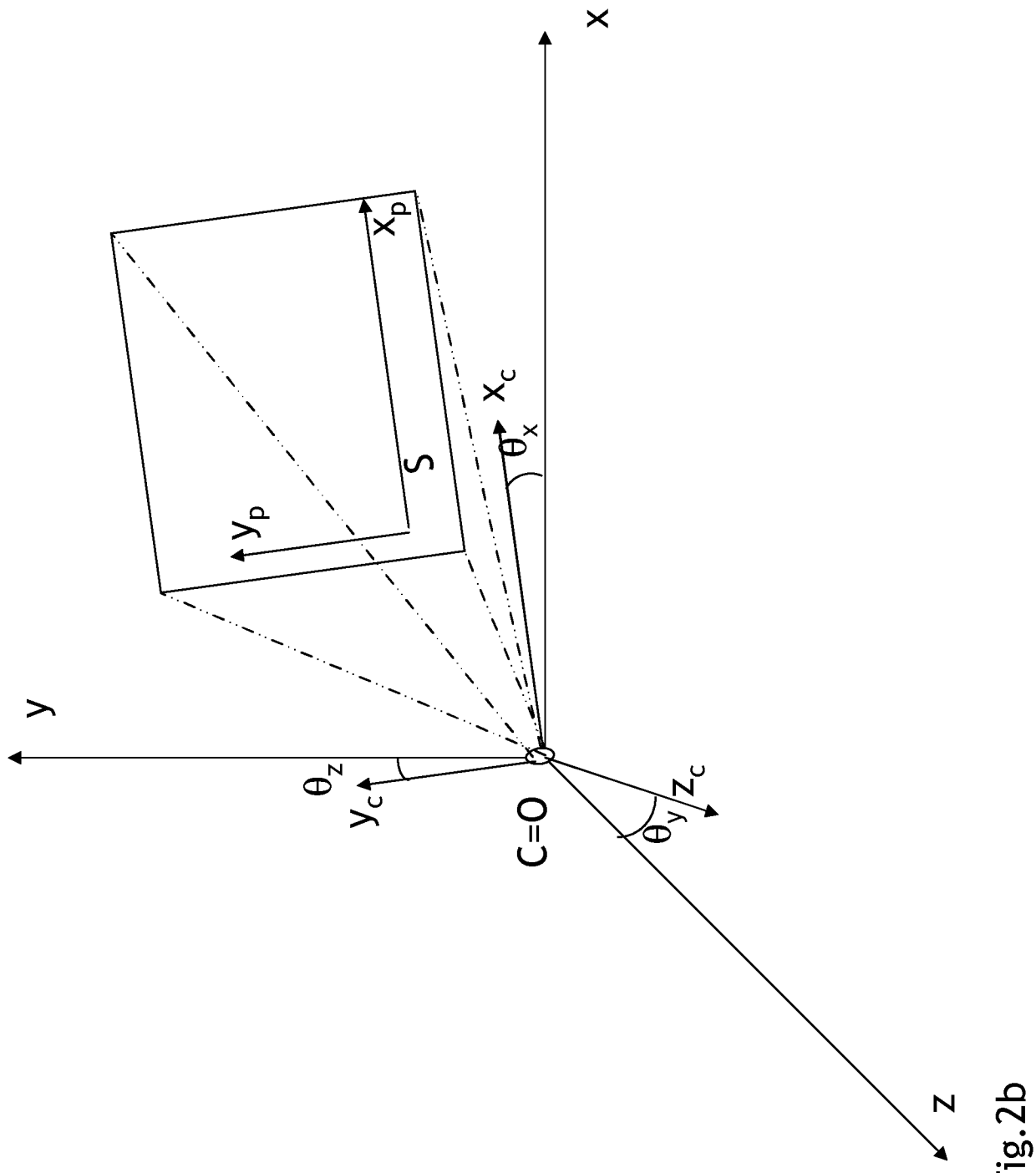


Fig.2b

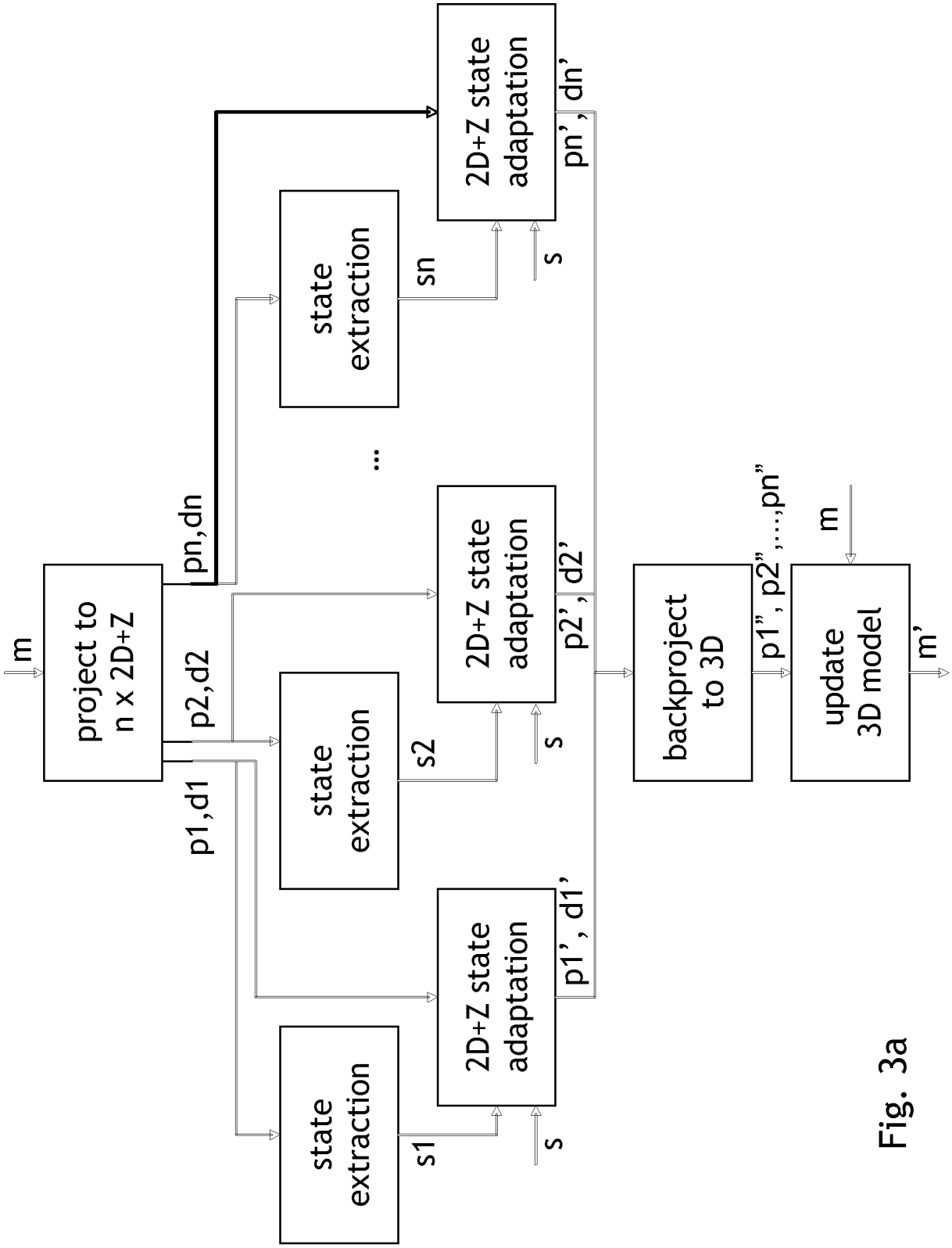

m
project to n x 2D+Z
p1,d1
p2,d2
pn,dn
state extraction
state extraction
state extraction
s1
s2
sn
2D+Z state adaptation
2D+Z state adaptation
2D+Z state adaptation
s
s
s
p1', d1'
p2', d2'
pn', dn'
backproject to 3D
p1", p2",...,pn"
update 3D model
m
m'

Fig. 3a

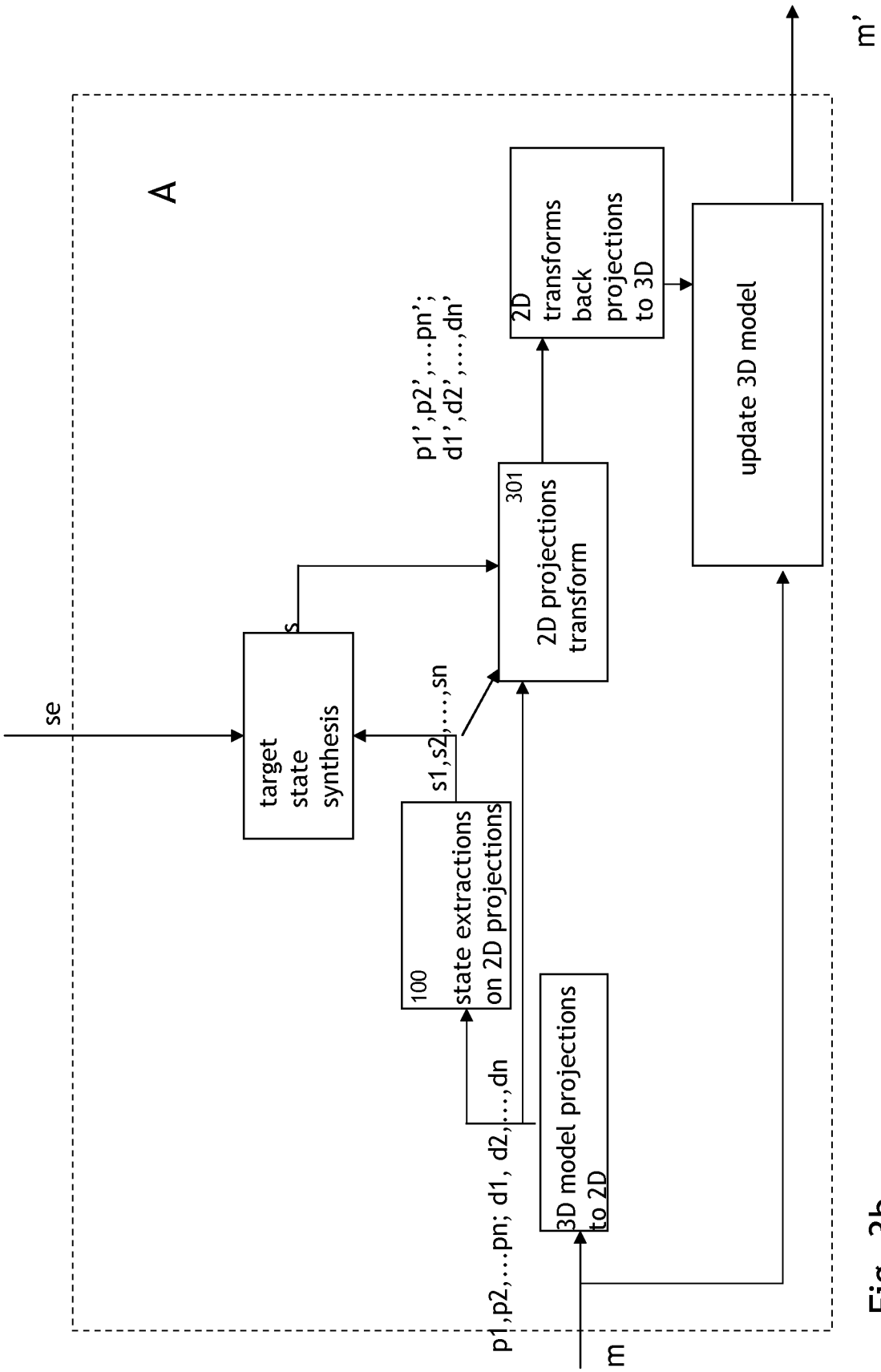

A
se
target state synthesis
s
s1,s2,...,sn
100
state extractions on 2D projections
p1,p2,...pn; d1, d2,...,dn
3D model projections to 2D
m
301
2D projections transform
p1',p2',...pn';
d1',d2',...,dn'
2D transforms back projections to 3D
update 3D model
m'

Fig. 3b

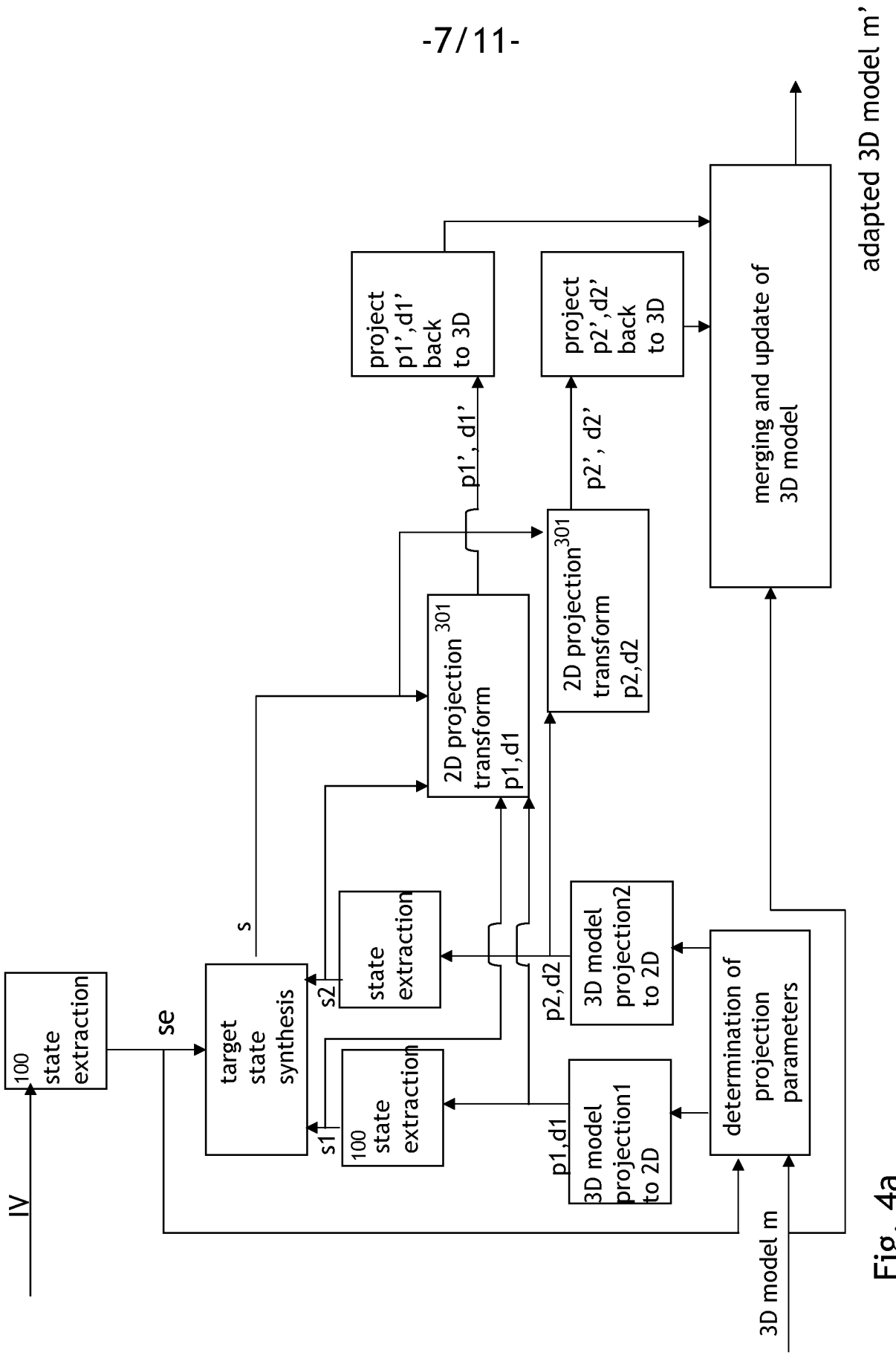

IV
100 state extraction
se
target state synthesis
s
s1
s2
100 state extraction
state extraction
3D model projection1 to 2D
3D model projection2 to 2D
p1,d1
p2,d2
2D projection transform p1,d1
2D projection transform p2,d2
301
301
p1', d1'
p2', d2'
project p1',d1' back to 3D
project p2',d2' back to 3D
merging and update of 3D model
adapted 3D model m'
determination of projection parameters
3D model m

Fig. 4a

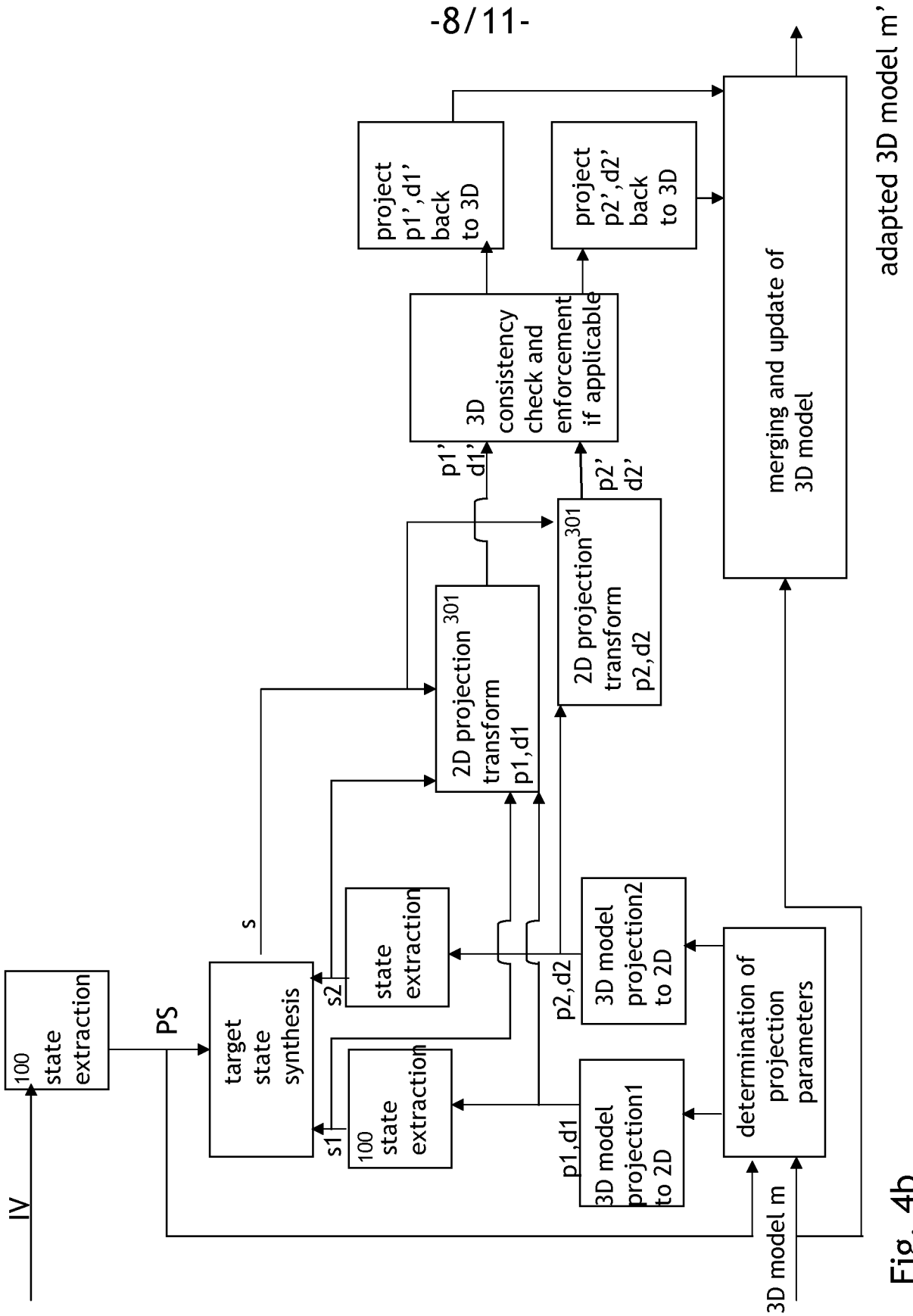

IV
100 state extraction
PS
target state synthesis
s
s2
state extraction
s1
100 state extraction
2D projection 301 transform p1,d1
p1,d1
3D model projection1 to 2D
p2,d2
3D model projection2 to 2D
2D projection 301 transform p2,d2
p1' d1'
p2' d2'
3D consistency check and enforcement if applicable
project p1',d1' back to 3D
project p2',d2' back to 3D
merging and update of 3D model
adapted 3D model m'
determination of projection parameters
3D model m

Fig. 4b

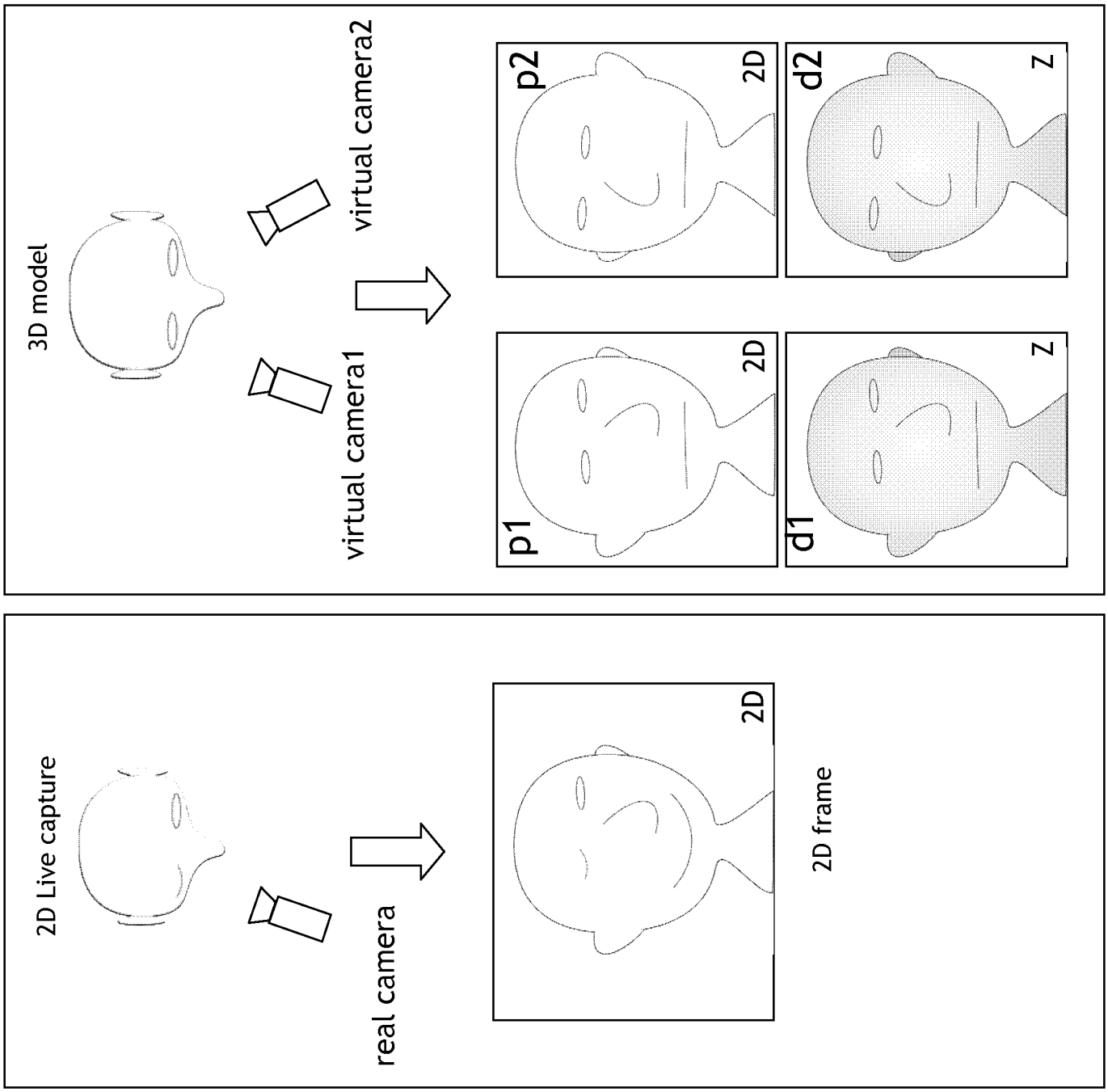

3D model
virtual camera2
virtual camera1
p2
2D
d2
z
p1
2D
d1
z
2D Live capture
real camera
2D
2D frame

Fig. 5a

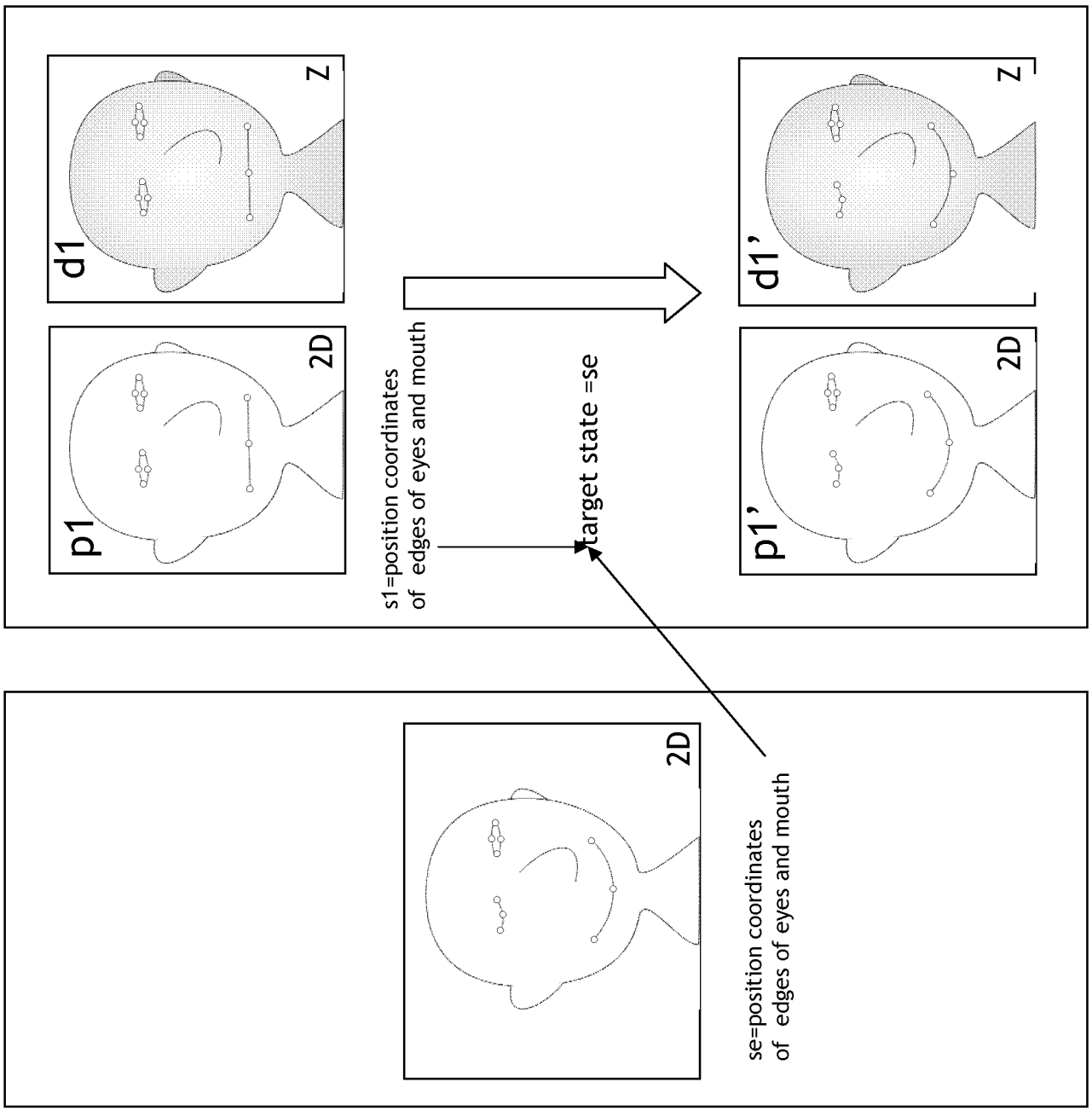

d1
z
p1
2D
s1=position coordinates
of edges of eyes and mouth
target state =se
d1'
z
p1'
2D
2D
se=position coordinates
of edges of eyes and mouth

Fig. 5b

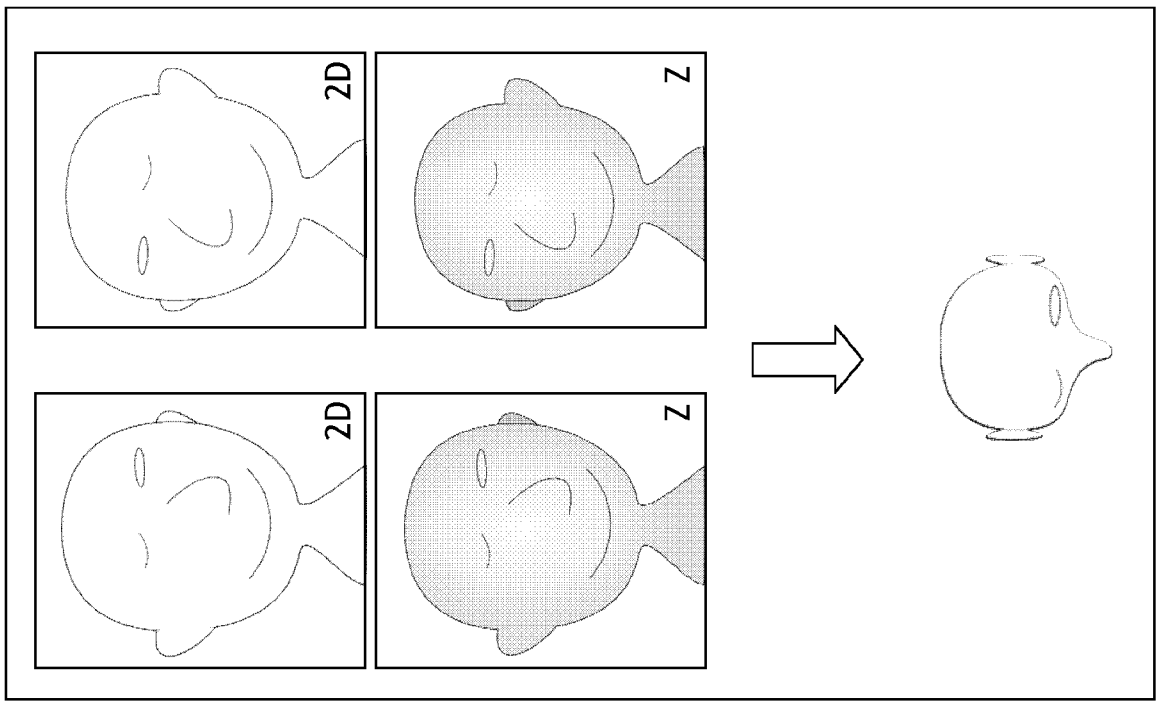


Fig. 5c

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/060512

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06T7/20
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)
G06T 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, 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	EP 1 669 933 A2 (SONY CORP [JP]) 14 June 2006 (2006-06-14) abstract figures 3,25-28,31-33,41-42,48 paragraphs [0273] - [0278], [0290] - [0294], [0300] - [0313], [0327] - [0340], [0405], [0408] -----	1-10
A	US 2010/215255 A1 (XIAO JING [US] ET AL) 26 August 2010 (2010-08-26) abstract figures 12,13,23,26 paragraphs [0059] - [0060], [0064] - [0067], [0077] - [0081], [0096] - [0101], [0189], [0222] -----	1-10

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

☒ See patent family annex.

* Special categories of cited documents :

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

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

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of 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 August 2012	13/08/2012

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	Scholz, Volker

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2012/060512

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1669933 A2	14-06-2006	CN 1787012 A	14-06-2006
		CN 101055646 A	17-10-2007
		CN 101055647 A	17-10-2007
		EP 1669933 A2	14-06-2006
		EP 1847967 A1	24-10-2007
		EP 1847968 A1	24-10-2007
		JP 4449723 B2	14-04-2010
		JP 2006163871 A	22-06-2006
		KR 20060064553 A	13-06-2006
		US 2006188144 A1	24-08-2006
US 2010215255 A1	26-08-2010	NONE	