

(12) CERERE DE BREVET DE INVENȚIE

(21) Nr. cerere: a 2021 00560

(22) Data de depozit: 17/09/2021

(41) Data publicării cererii:  
30/05/2022 BOPI nr. 5/2022

(71) Solicitant:  
• UNIVERSITATEA TEHNICĂ DIN  
CLUJ-NAPOCA, STR.MEMORANDUMULUI  
NR.28, CLUJ-NAPOCA, CJ, RO

(72) Inventatori:  
• POP MARIAN LEONTIN, NR.209A,  
SAT VIILE, SEINI, MM, RO;  
• TAMAS LEVENTE, STR.BISERICII  
ORTODOXE, NR.18, CLUJ-NAPOCA, CJ,  
RO

(54) CORECȚIA SUPRAFEȚELOR PLANE DIN IMAGINILE  
PROVENITE DE LA CAMERE CU INFORMAȚIE  
DE DISTANȚĂ FOLOSIND REȚELE NEURONALE  
CONVOLUȚIONALE

(57) Rezumat:

Invenția se referă la o metodă pentru corecția suprafețelor plane din imagini provenite de la camere dotate cu senzori de tip ToF (Time-of-Flight), folosind rețele neuronale convoluționale. Metoda conform invenției se implementează pe un sistem de calcul (1) prevăzut cu un program (2) de corecție a suprafețelor plane achiziționate de la o cameră (3) și constă într-o primă etapă în care este preluată imaginea și este creat setul de date necesare antrenării unei rețele neuronale, prin: - achiziția unei serii de imagini de adâncime ale unor spații interioare care conțin una sau mai multe suprafețe plane, cum ar fi: podea, tavan, pereți. etc. - transformarea imaginilor într-un nor de puncte, folosind parametrii intrinseci ai camerei (3), - identificarea planurilor majore din norul de puncte și proiectarea fiecăruia plan distorsionat pe un plan ideal, ale cărui coordonate au fost obținute prin folosirea iterativă a unor funcții din biblioteca PCL, rezultând planuri rectificatice cu ajutorul cărora norul de puncte este translatat înapoi într-o imagine 2D, obținând astfel imagini de referință pentru antrenarea modelului neuronal, această primă etapă fiind urmată de o a doua etapă în care se realizează modelul neuronal folosind modelul FPN (Feature Pyramid Network), având funcția de calculare a erorii adaptată pentru a putea servi la rectificarea suprafețelor plane, și de o a treia etapă în care modelul

neural antrenat este folosit de programul (2) de corecție care preia imaginile de la cameră (3), le introduce ca intrări ale modelului care, la rândul lui, le procesează și livrează la ieșire o imagine cu planurile rectificatice.

Revendicări: 4  
Figuri: 4

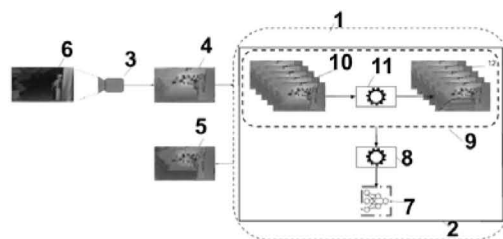


Fig. 1



OFICIUL DE STAT PENTRU INVENȚII ȘI MĂRCI
Cerere de brevet de invenție
Nr. a 2021 00560
Data depozit ....17-09-2021..

## Corecția suprafețelor plane din imaginile provenite de la camere cu informație de distanță folosind rețele neuronale convoluționale

Inventatori: Marian-Leontin Pop, Levente Tamas  
Universitate Tehnică din Cluj-Napoca, România

Invenția se referă la o metodă de eliminare, automată, a efectului de interferență multiplă, din suprafețele plane, care poate să apară în imaginile provenite de la camerele cu informație de distanță, camere cu senzori de tipul "Time-of-Flight" (ToF). Metoda cuprinde o componentă care folosește rețele neuronale convoluționale pentru eliminarea artefactelor survenite în urma efectului de interferență multiplă, cât și o componentă pentru crearea setului de date necesar antrenării rețelei neuronale.

În metodele cunoscute, se încearcă eliminarea efectului de interferență multiplă din imaginile cu informație de adâncime provenite de la camere ToF cu modulație cu undă continuă, în care se compensează aceste erori folosind ecuațiile specifice pentru propagarea semnalului precum în [1-3, 7]. Alte invenții, precum [1, 4, 5, 8] folosesc seturi de date sintetice pentru antrenarea rețelelor neuronale, însă acestea nu produc rezultate foarte bune pentru datele reale, iar metodele care folosesc date reale [5, 9] folosesc și senzori de înaltă precizie, mult mai costisitori, pentru crearea setului de date. Invenția curentă diferă de cele menționate prin faptul că adresează problema eliminării interferențelor multiple din imaginile de adâncime provenite de la camerele ToF emițătoare de impulsuri folosind în același timp un set de date reale pentru antrenarea rețelei neuronale.

Problema tehnică pe care o rezolvă invenția este, în primul rând, realizarea setului de date reale folosit pentru antrenarea rețelei neuronale, folosind metode convenționale și fără a utiliza alți senzori costisitori, iar apoi antrenarea unei rețele neuronale convoluționale pe setul de date creat.

Metoda de realizare a setului de date implică trei pași importanți: translatarea imaginii de adâncime într-un set discret de puncte în spațiul tridimensional, procesarea punctelor din spațiul 3D pentru înlăturarea distorsiunilor prezente în planurile din imagine, iar apoi translatarea setului de puncte înapoi în spațiul bidimensional al imaginii de adâncime. Mai precis, folosind parametrii intrinseci ai camerei, translatăm punctele din imaginea 2D primită de la cameră, într-un set de puncte aflat într-un spațiu 3D. Acest lucru ne ajută, fiindcă artefactele care ne interesează sunt mult mai pronunțate în spațiul tridimensional. După translatare, folosind câteva din funcționalitățile prezente

în biblioteca de procesare a seturilor de puncte aflate în spațiul 3D denumită PCL, astfel încât să proiectăm toate punctele care aparțin unei suprafețe planare pe un singur plan, uniform, fără distorsiuni. Noul set de puncte, astfel obținut, este reprojectat înapoi în spațiul bidimensional al imaginii de adâncime. Imaginile de referință astfel obținute sunt servite rețelei neuronale convoluționale, împreună cu imaginile originale, iar astfel rețeaua este capabilă să distingă între cele două imagini, folosind o funcție de calculare a erorii, special construită pentru acest tip de date, iar apoi să învețe ce anume trebuie să îmbunătățească în imaginea de adâncime, fără a mai face anumite translații în spațiul tridimensional.

Ideea inovatoare iese în evidență prin faptul că este ușor reproductibilă pentru orice tip de cameră și nu necesită echipamente costisitoare pentru realizarea setului de imagini de referință necesare antrenării rețelei neuronale.

Implementarea metodei reiese mai clar din figurile prezentate mai jos, după cum urmează:

- Figura 1 - structura sistemului;
- Figura 2 - pașii necesari creării setului de date;
- Figura 3 - pașii necesari antrenării rețelei neuronale;
- Figura 4 - relația dintre punctele estimate de rețea și punctele de referință, în timpul procesului de antrenare.

Prezenta invenție se poate aplica la un sistem de forma celui prezentat în Figura 1. Sistemul este alcătuit dintr-un sistem ce calculează (1), având implementat un program (2), de corecție a suprafețelor plane achiziționate de la camera (3), astfel încât imaginea de adâncime (4) este transformată în imaginea de adâncime rectificată (5).

Camera (3) poate fi o cameră de tip Time-of-Flight emițătoare de impulsuri sau cu modulație cu undă continuă, și are rolul de a prelua imaginea unui spațiu interior (6) care conține suprafețe plane.

Programul (2) de corecție a suprafețelor plane constă dintr-o rețea neuronală (7), care a fost în prealabil antrenată de un program (8) pe un set de date (9).

Setul de date (10) a fost construit prin achiziționarea unui set de imagini de adâncime (10), de la camera (3), și procesarea acestor imagini de către programul (11), pentru obținerea imaginilor de adâncime de referință (12).

Metoda conform invenției presupune realizarea următoarelor etape:

1. Preluarea imaginii cu ajutorul camerei și crearea setului de date necesare antrenării rețelei neuronale;
2. Realizarea modelului neuronal;
3. Exploatarea modelului neuronal pentru corecția erorilor;

În prima etapă, pentru crearea setului de date necesar antrenării rețelelor neuronale, s-au achiziționat, folosind camera ToF, o serie de imagini de adâncime ale unor spații interioare care conțin una sau mai multe suprafețe planare, cum ar fi podea, tavan, pereți, etc. Imaginile astfel achiziționate le-am transformat, folosind parametrii intrinseci ai camerei, într-un nor de puncte, în spațiul tridimensional. În spațiul 3D, am folosit funcții din biblioteca PCL pentru identificarea planurilor majore din norul de puncte și proiectarea fiecărui plan distorsionat pe un plan ideal, ale cărui coordonate au fost obținute prin folosirea iterativă a unor funcții din biblioteca PCL. După rectificarea planurilor majore în spațiul 3D, norul de puncte este translatat înapoi într-o imagine 2D, obținând astfel imaginile de referință pentru antrenarea modelului.

În cea de a doua etapă, am utilizat un model deja existent, Feature Pyramid Network (FPN), care oferă rezultate bună în procesele de segmentare a imaginilor RGB, în sarcinile necesării identificării și clasificării obiectelor prezente în imagini, la nivel de pixel. Din acest motiv am ales acest model, pe care l-am modificat astfel încât să proceseze imagini de adâncime, modificându-i astfel și funcția de calculare a erorii pentru a putea servi obiectivului nostru de rectificare a suprafețelor plane.

În ultima etapă, modelul neuronal antrenat este folosit de către un program care preia imaginile de la camera ToF și le servește pe intrarea modelului care mai apoi le procesează și oferă la ieșire o imaginile cu planurile rectificate din imagine.

În Figura 2 se prezintă organigrama pașilor necesari generării setului de date necesar antrenării și exploatării modelului neuronal, în care se pune în evidență modul de procesare a punctelor în spațiul tridimensional folosind biblioteca PCL. Ideea principală din această etapă constă în folosirea algoritmului RANSAC, al bibliotecii, de două ori: prima dată pentru a extrage coordonatele planului care aproximează cel mai bine planul din imagine, iar a doua oară pentru a extrage cât mai multe puncte care aparțin planului major din imagine. Setul de puncte rezultat în urma celei de-a doua iterații a algoritmului este proiectat pe planul ce are coordonatele determinate în prima iterație, rezultând astfel un plan ideal care nu conține artefacte.

Blocurile de calcul necesare generării setului de date din Figura 1 sunt:

Blocul (13), preluarea datelor de adâncime de la camera (2). În acest bloc, fluxul de imagini achiziționat de la cameră prin intermediul ROS (Robot Operating System) este împărțit în cadrele componente ale acestuia, salvat sub formă de imagini 9.

Blocul (14), are rolul de translatare a punctelor din spațiul 2D în spațiul 3D, folosind parametrii intrinseci ai camerei. Translatarea se face cunoscând matricea intrinsecă a camerei, care conține distanța focală a camerei, precum și centrul sensorului camerei (2), folosind o funcție din biblioteca PCL.

Blocul (15), are rolul de a prelua norul de puncte obținut în blocul anterior, pentru a identifica, folosind algoritmul RANSAC, implementat în biblioteca PCL, toate planurile majore existente și de a procesa fiecare plan în parte, astfel încât să identifice un set de puncte care aparțin unui plan ideal ce aproximează cel mai bine planul curent, care conține artefacte. Selectarea punctelor care aparțin unui plan ideal se realizează prin setarea unor constrângeri mai mari în algoritmul de identificare al planurilor, prin setarea corespunzătoare a parametrilor acestuia.

Blocul (16), este aproape identic cu blocul (15), doar că de această dată se folosesc parametrii cu constrângeri mai slabe, care îi permit algoritmului să selecteze întregul plan cu toate punctele sale, chiar dacă acesta conține și artefacte.

Blocul (17), preia toate punctele selectate în blocul (16) și le proiectează pe planul ideal, ce are coordonatele extrase în blocul (15).

Blocul (18), realizează înlocuirea punctelor din blocul (16), adică punctele planului ce conține artefacte, cu punctele din blocul (17), adică planul rectificat, lăsând restul punctelor care nu aparțin planului neatins.

Blocul (19), realizează proiectarea norului de puncte înapoi în imaginea de adâncime, folosind tot parametrii intrinseci ai camerei (3), dar folosind o funcție inversă din librăria PCL.

În Figura 3 sunt prezentați pașii necesari pentru antrenarea rețelei neuronale convoluționale, în care am ales un model de tipul FPN [6]. După ce am încărcat în model perechile de imagini de adâncime folosim o funcție personalizată de calculare a erorii dintre imaginea prezisă de către rețea și imaginea de referință, în spațiul 3D. Fiindcă artefactele pe care dorim să le înlăturăm se observă într-un mod mai pronunțat în spațiul tridimensional am ales să folosim o rețea convoluțională care este capabilă să obțină rezultate bune în procesarea de imagini, dar pentru care calculăm o eroare în spațiul 3D.

Blocurile de calcul necesare antrenarea rețelei neuronale din Figura 3 sunt:

Blocul (20), are rolul de încărcare în programul (9) a imaginilor de adâncime (4) de la camera (3), împreună cu imaginile de adâncime de referință (10) pentru a putea fi servite pe intrarea rețelei neuronale (7).

Blocul (21), are rolul de a normaliza imaginile în intervalul  $[0, 1]$ , pentru a optimiza procesul de antrenare al rețelei (7).

Blocul (22), inițializează modelul (8) prin încărcarea acestuia în memoria sistemului de calcul (4).

Blocul (23), preia imaginea de pe ieșirea modelului neuronal ca răspuns la imaginile de intrare și o trimite în funcția de calculare a erorii, din blocul (24).

Blocul (24), are rolul de a prelua imaginea de ieșire de la blocul (23), împreună cu imaginea de referință încărcată în blocul (20) și translatarea lor în spațiul tridimensional. Odată realizată translatarea, se calculează eroarea dintre cele două seturi de puncte, pentru a putea determina cât de departe sau cât de aproape sunt punctele din imaginea de referință față de cele din imaginea estimată de rețea.

Blocul (25), are rolul de a actualiza parametrii interni ai modelului neuronal (8), în funcție de eroarea estimată în blocul (24), după care procesul se reia de la blocul (23) până când eroarea calculată este suficient de mică.

În Figura 4 se prezintă relația dintre punctele din planul estimat de către rețea și punctele din planul de referință. Practic, în calcularea erorii se calculează într-o primă fază o radicalul unei erori medii pătratice logaritmice realizate între fiecare punct din imaginea estimată de rețea și imaginea de referință, denumită  $RMSE\_log$ , iar mai apoi se calculează și câte o eroare medie pătratică a punctelor, în funcție de coordonata X, coordonat Y și coordonata Z, denumite  $lossX$ ,  $lossY$ , respectiv  $lossZ$ . Aceste patru ecuații sunt mai apoi combinate pentru a obține o ecuație generală de calculare a erorii.

Mai jos este prezentată formula folosită pentru calcularea erorii:

$$lossX = 1i|xPi-xPi*|n, in (1)$$

$$lossY = 1i|yPi-yPi*|n, in (2)$$

$$lossZ = 1i|zPi-zPi*|n, in (3)$$

$$RMSE\_log = 1i(\ln|zPi|-\ln|zPi*|)2n12, in (4)$$

Unde: -  $xPi, yPi, zPi$  sunt coordonatele spațiale ale punctului  $Pi$  din imaginea de referință;

-  $xPi*, yPi*, zPi*$  sunt coordonatele spațiale ale punctului  $Pi*$  din imaginea estimată de rețeaua neuronală;

-  $n$  este numărul total de puncte din imaginea de referință.

Folosind ecuațiile (1), (2), (3) și (4) se ajunge la forma finală a erorii:

$$\mathcal{E} = 10RMSE\_log^{10}(3^{-e_{lossX}-e_{lossY}-e_{lossZ}}) (5)$$

Ideea inovatoare iese în evidență prin faptul că metoda este capabilă să corecteze erorile sistematice de la camerele de tip ToF în timp real chiar și pe dispozitive înglobate cu procesor GPU folosind tehnici de inteligență artificială. Un video cu caracter demonstrativ și codul sursă va fi disponibilă pe site-ul autorilor după ce se va publica brevetul.

**Referințe**

- [1] Guo, Q., Frosio, I., Gallo, O., Zickler, T., & Kautz, J. (2018). Tackling 3d tof artifacts through learning and the flat dataset. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 368-383).
- [2] Jiménez, D., Pizarro, D., Mazo, M., & Palazuelos, S. (2014). Modeling and correction of multipath interference in time of flight cameras. *Image and Vision Computing*, 32(1), 1-13.
- [3] Patent US10215856B1
- [4] Marco, J., Hernandez, Q., Munoz, A., Dong, Y., Jarabo, A., Kim, M. H., ... & Gutierrez, D. (2017). DeepToF: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 36(6), 1-12.
- [5] He, Y., Liang, B., Zou, Y., He, J., & Yang, J. (2017). Depth errors analysis and correction for time-of-flight (ToF) cameras. *Sensors*, 17(1), 92.
- [6] Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S., 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).
- [7] Fuchs, S., 2010, August. Multipath interference compensation in time-of-flight camera images. In *2010 20th International Conference on Pattern Recognition* (pp. 3583-3586). IEEE.
- [8] Agresti, G. and Zanuttigh, P., 2018. Deep learning for multi-path error removal in ToF sensors. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops* (pp. 0-0).
- [9] Son, K., Liu, M.Y. and Taguchi, Y., 2016, May. Learning to remove multipath distortions in time-of-flight range images for a robotic arm setup. In *2016 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3390-3397). IEEE.
- [10] Patent US10234561B2
- [11] Patent US10311378B2

### REVEDICĂRI

1. Metodă pentru corecția suprafețelor din imaginile provenite de la camere cu informație de distanță folosind rețele neuronale convoluționale implementată pe un sistem de calcul (1), având implementat un program (2), de corecție a suprafețelor plane achiziționate de la o cameră (3), **caracterizată prin aceea că** pentru rectificarea imaginilor provenite de la camera (3) se parcurg următoarele etape:

a. Preluarea imaginii cu ajutorul camerei și crearea setului de date necesare antrenării rețelei neuronale, în următorii pași:

- Achiziția unei serii de imagini de adâncime ale unor spații interioare care conțin una sau mai multe suprafețe planare, cum ar fi podea, tavan, pereți, etc.
- Transformarea imaginilor într-un nor de puncte, folosind parametrii intrinseci ai camerei
- Identificarea planurilor majore din norul de puncte și proiectarea fiecărui plan distorsionat pe un plan ideal, ale cărui coordonate au fost obținute prin folosirea iterativă a unor funcții din biblioteca PCL rezultând planuri rectificate;
- Folosind planurile majore rectificare (corectate) norul de puncte este translatat înapoi într-o imagine 2D, obținând astfel imaginile de referință pentru antrenarea modelului.

b. Realizarea modelului neuronal folosind modelul Feature Pyramid Network (FPN), cu funcția de calculare a erorii adaptată pentru a putea servi la rectificarea suprafețelor plane

c. Exploatarea modelului neuronal în care modelul neuronal antrenat este folosit de către un program care preia imaginile de la camera ToF și le servește pe intrarea modelului care mai apoi le procesează și oferă la ieșire o imaginile cu planurile rectificate din imagine.

2. Metodă conform revendicării 1, **caracterizată prin aceea că** în procesul de creare al setului de date necesar antrenării rețelei neuronale, se folosește doar informația provenită de la camera ToF, fără a se folosi un alt senzor adițional.

3. Metodă conform revendicării 2, **caracterizată prin aceea că** funcția de rectificare a suprafețelor plane folosește într-un mod iterativ funcții de bază ale bibliotecii PCL pentru procesare punctelor în spațiul 3D, pentru a putea obține suprafețe planare fără distorsiuni în mod automat.

4. Metodă conform revendicării 1, **caracterizată prin aceea că** antrenarea rețelei neuronale se face folosind imagini 2D, în timp ce eroarea, dintre imaginile cu planurile rectificate estimate de către rețea și imaginile de referință din setul de date, este calculată în spațiul 3D.



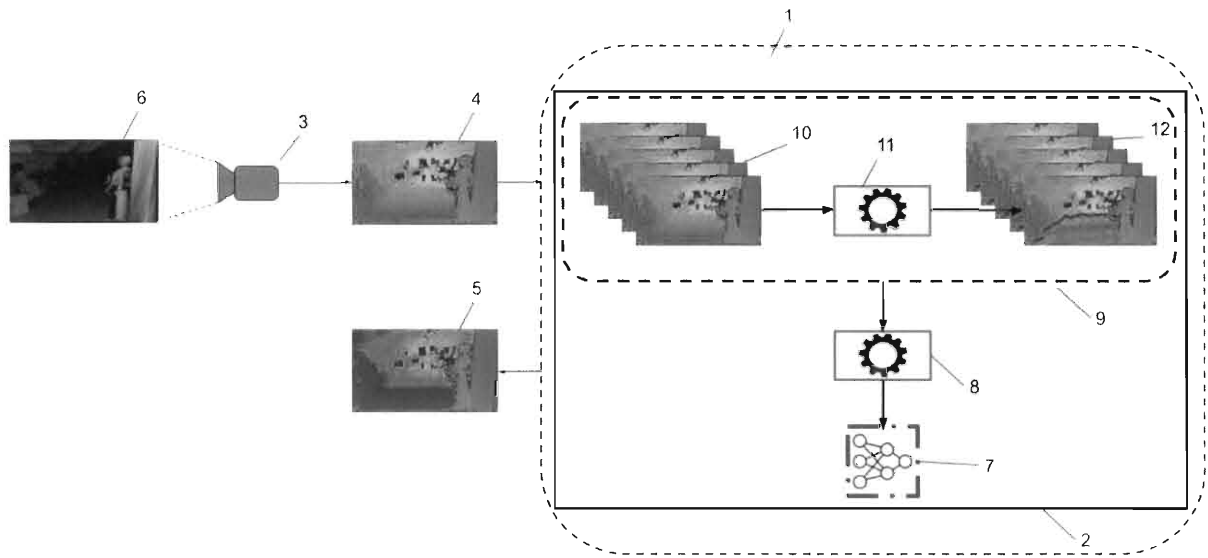


Fig. 1

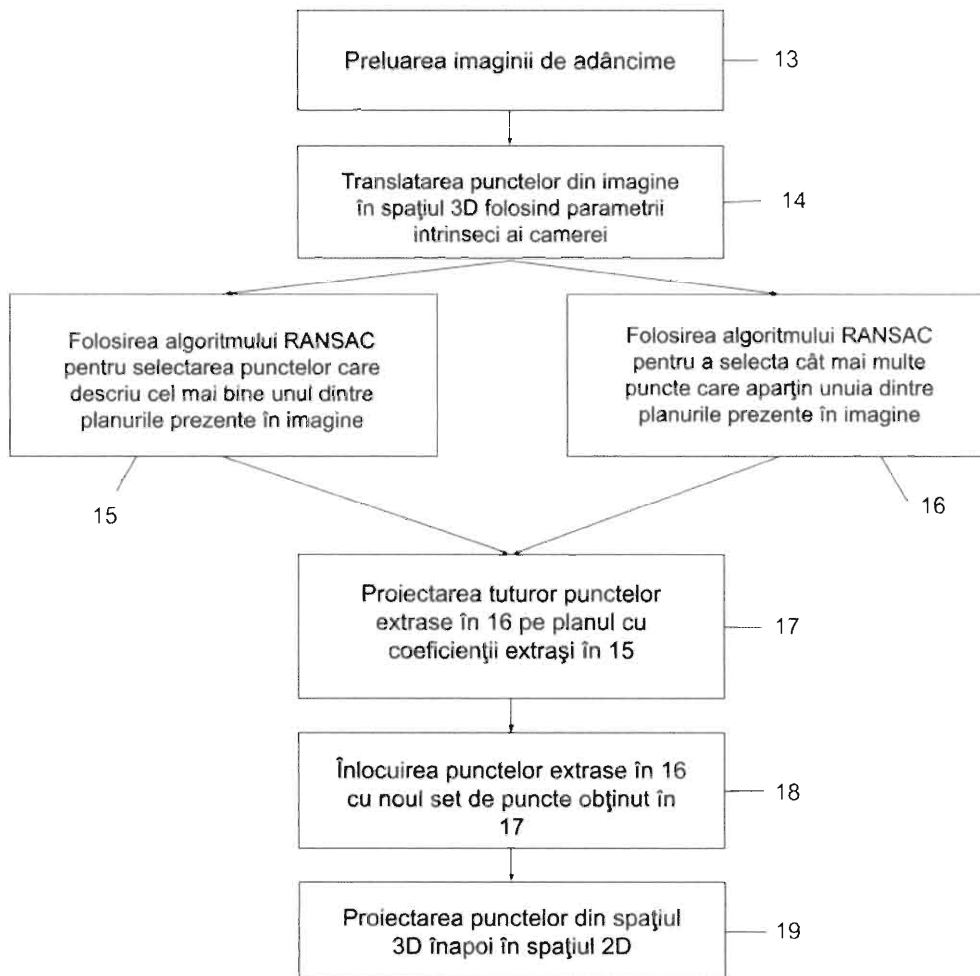


Fig. 2

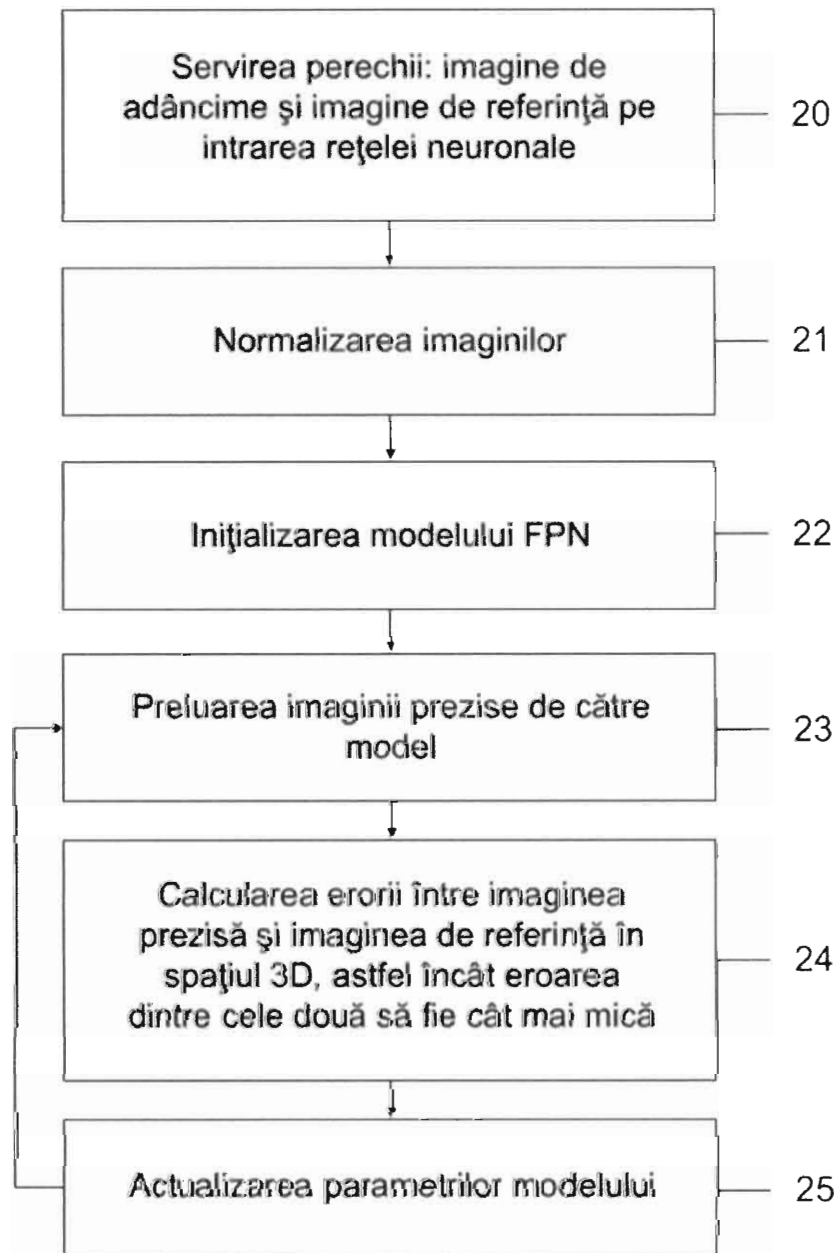


Fig. 3

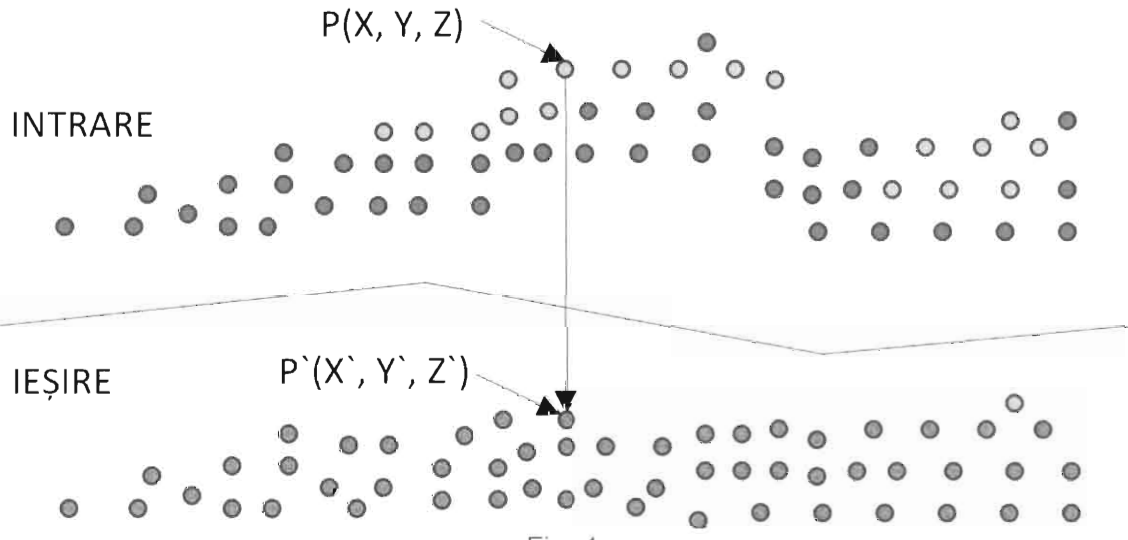


Fig. 4