

RAPORTARE ȘTIINȚIFIC ȘI TEHNIC 2021

Contract nr. 23PTE/2020

Etapă nr. 2/2021

Cod proiect: PN-III-P2-2.1-PTE-2019-0367;

Titlu proiect: Servicii de transfer de cunoștințe pentru sisteme de percepție 3D bazate pe AI

Beneficiar: Analog Devices GMBH, România

Partener: Universitatea Tehnică din Cluj-Napoca

Rezumatul etapei 1 (Valorificare algoritmi de recunoaștere de obiecte în aplicații client): Procesarea datelor 3D prin extinderea capabilităților ale camerei 3D de la ADI prin intermediul soluțiilor bazate pe IA. Rezultatele preconizate: primele publicații științifice, studii de caz, brevete.

Avizat:

Director de proiect:
Dr. Andrei Cozma

Activitate 2.1 Dezvoltare modul de transfer learning

Această activitate avea drept scop îmbunătățirea calității normalelor de suprafață generate pentru nori de puncte 3D obținute de la camera ToF (vezi Fig. 1). Acesta metodă se bazează pe abilitatea de a captura caracteristici multi-scalare de la camerele ToF folosind Feature Pyramid Networks (FPN) [1] pentru estimarea normalelor. Din cauza tipului specific al camerelor ToF, măsurătorile spațiului 3D sunt codificate sub forma imaginilor de adâncime, care se pot proiecta în 3D ca un nor de puncte organizat. Chiar dacă între straturile piramidei și bazele de suport a normalului, din norul de puncte, nu este o legătură strictă definită, de fapt modelul imită estimarea normalelor folosind metode multiscalare în spațiul 3D.

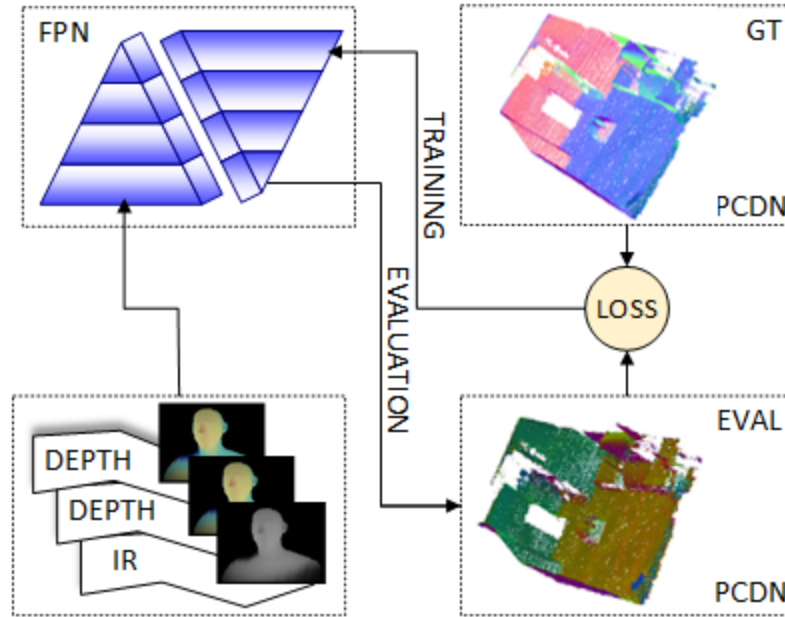


Fig. 1 Calcul optim de normale bazat pe transfer learning

Codificarea informațiilor referitoare la normale înseamnă că în spațiul 3D un vector are trei coordonate: x , y și z . Vectorii sunt considerați ca fiind vectori normalizați, adică valorile coordonatelor sunt în intervalul $[-1,1] \in \mathbb{R}$, iar lungimea vectorului este de o unitate. După stabilirea limitelor în spațiul 3D, putem să stabilim limitele în spațiul RGB. La o imagine color, de obicei sunt 3 canale (R-roșu (red), G-verde (green), B-albastru (blue)), fiecare canal fiind pe 8 biți, cu valori posibile în intervalul $[0,255] \in \mathbb{N}$. Folosindu-ne de informațiile menționate, se poate face o mapare între coordonatele x,y,z și canalele R,G,B respectiv, cu o pierdere minimală, care este sub un grad.

Diagrama 2 prezintă detaliile referitoare la conexiuni între straturile FPN-ului folosit pentru estimarea normalelor: jos în sus (portocaliu), sus în jos (galben), conexiunile laterale (albastru), doi operatori convoluționali consecutivi (verde), predicția finală (rosu)

Folosind tehnica propusă în această etapă de dezvoltare în proiect, s-au obținut estimări de suprafețe pentru date 3D folosind echipamente înglobate cu GPU dedicat, cu rate de calcul peste 10Hz. Această tehnică dezvoltată a fost publicată în (1) și (2).

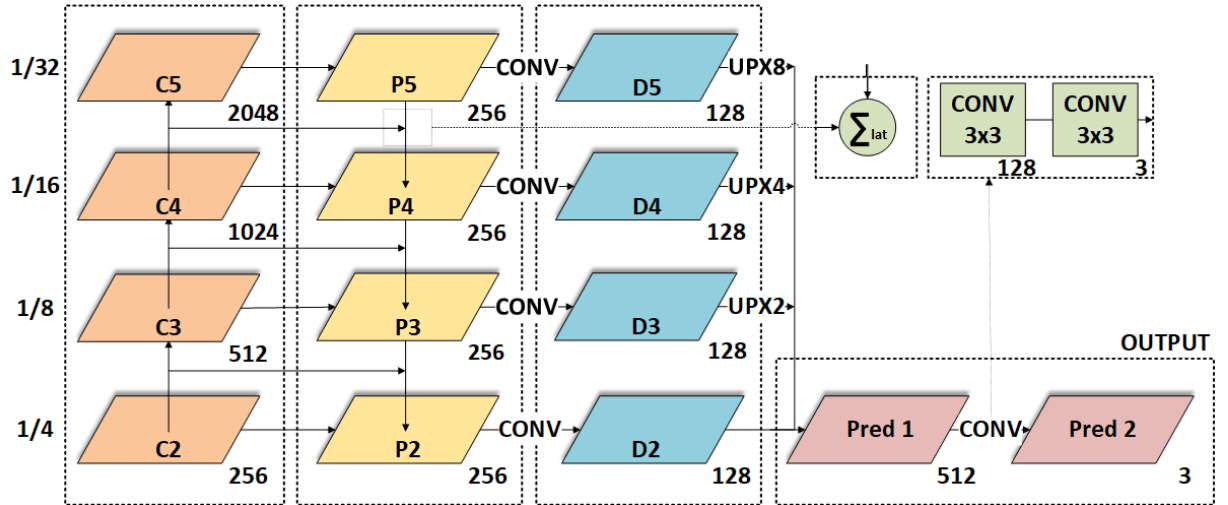


Fig.2.

Fig. 2 prezinta diagrama de conexiuni intre straturi a arhitecturii FPN folosite pentru estimarea normalelor.

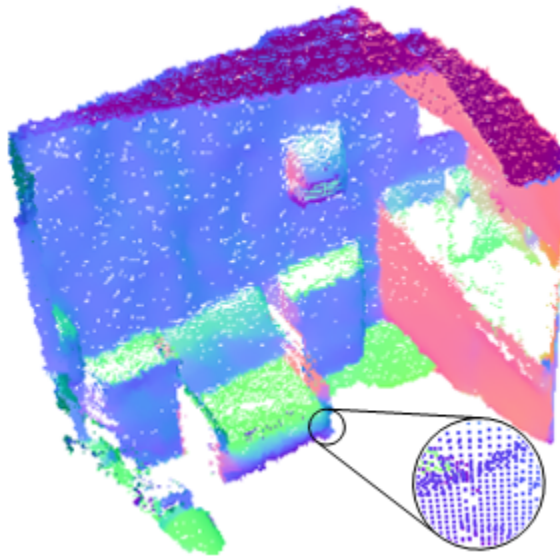


Fig. 3 Rezultatele obținute pentru estimarea normalelor

În Figura 3 se poate observa rezultatul codificat în spațiul RGB obținut în urma folosirii metodei noastre.

Activitate 2.2 Validare sistem de recunoaștere de obiecte cu modul de transfer learning

Recunoașterea obiectelor este o problemă relevantă atât în cercetare cât și în domeniul industrial. În acest domeniu o ramură specială reprezintă recunoașterea obiectelor din nori de puncte 3D. Acesta din urmă are avantajul de a avea informație topologică și metrică pentru obiectele în fața camerei.

Focusul pentru această activitate din cadrul proiectului era pe recunoașterea formelor 3D pentru a clasifica obiectele în diferite clase geometrice (e.g. sferă, plan, cilindru, etc).

Astfel s-a creat o aplicație care returnează clasa geometrică pentru obiectele din jur folosind algoritmi standard de tip RANSAC. Scopul era de a detecta în mod corect aceste clase indiferent de înclinația camerei. Pentru acesta s-a creat o pistă specială înclinată pe care se deplasa un robot mobil echipat cu camera ToF. Astfel am reușit să testăm algoritmul dezvoltat în diferite condiții, după cum se poate observa în imagine. Pentru lanțul de procesare de recunoaștere de obiecte s-a folosit o variantă specifică de PVCNN de pe Figura 4.

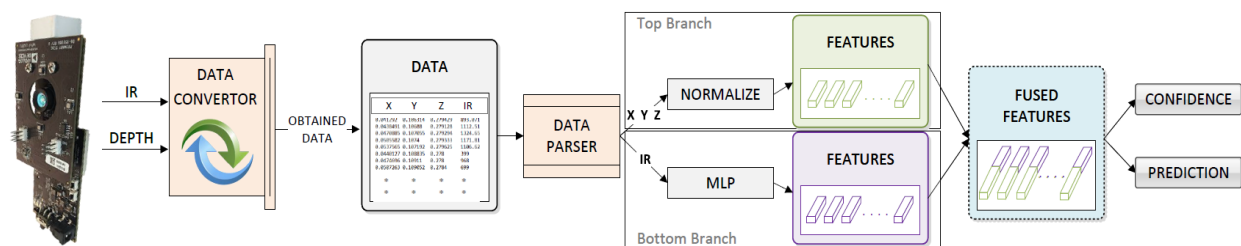


Fig. 4 Arhitectura generică de recunoaștere de obiecte bazată pe PVCNN

Rezultatele acestei etape au fost publicate în (3).

Activitate 2.3 Transfer de know-how pentru modulele dezvoltate

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 translatare 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:

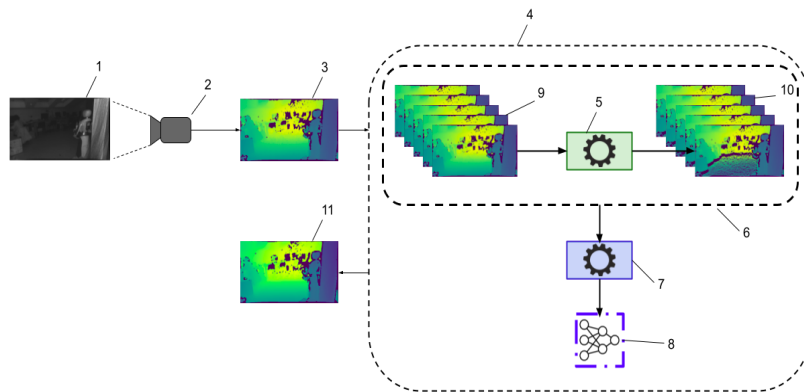


Fig. 5 - pașii necesari creării setului de date

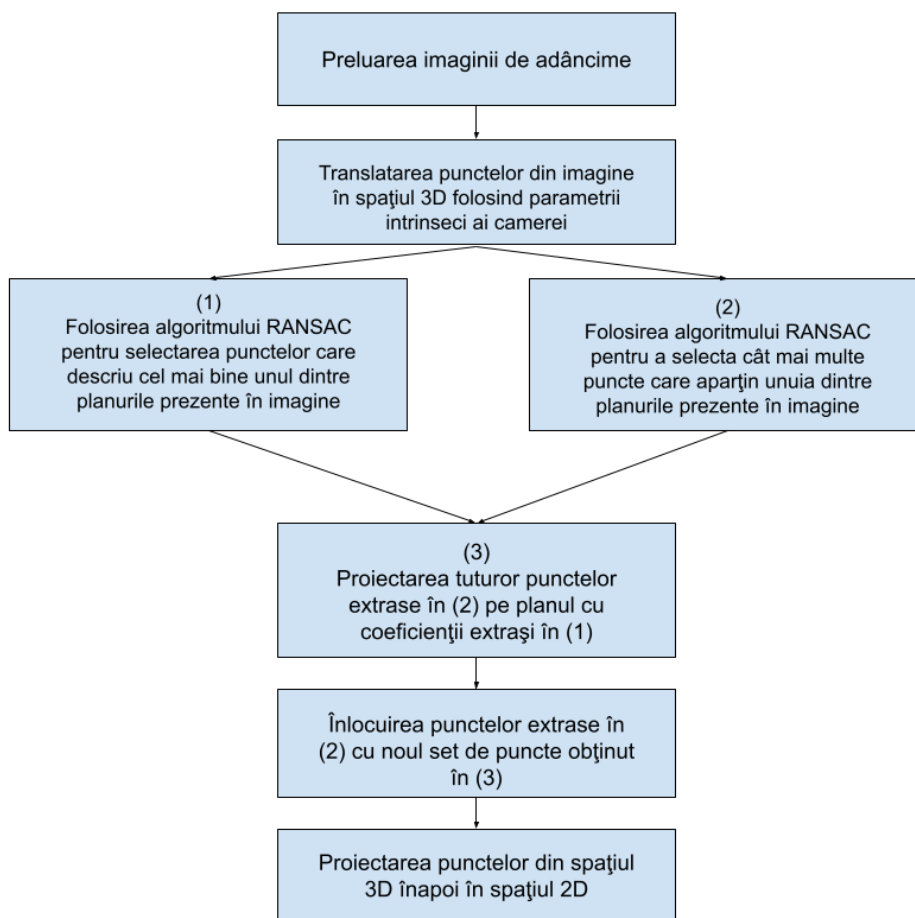


Fig. 6 - pașii necesari antrenării rețelei neuronale

În Figura 5 se prezintă o formă schematică a pașilor necesari generării setului de date, 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 iterație a algoritmului este proiectat pe planul ce are coordonatele determinate în prima iterație, rezultând astfel un plan ideal care nu conține artefacte.

În Figura 6 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.

Activitate 2.4 Validarea modulelor dezvoltate în studiu de caz (la client)

Pentru a valida rezultatele obținute s-au construit demo-uri specifice pentru roboți autonomi folosind platformele P3 disponibile la partenerul academic.

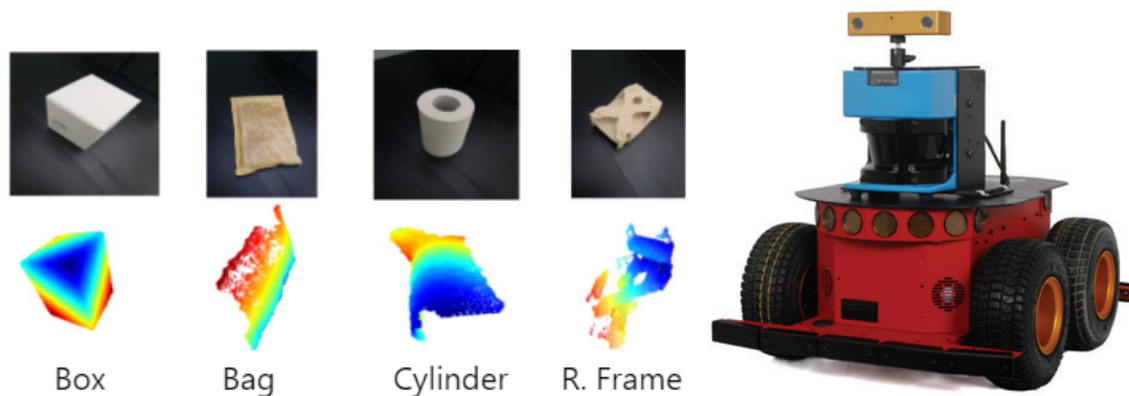


Fig. 7 Obiectele considerate pentru teste și platforma de AGV folosit pentru testare

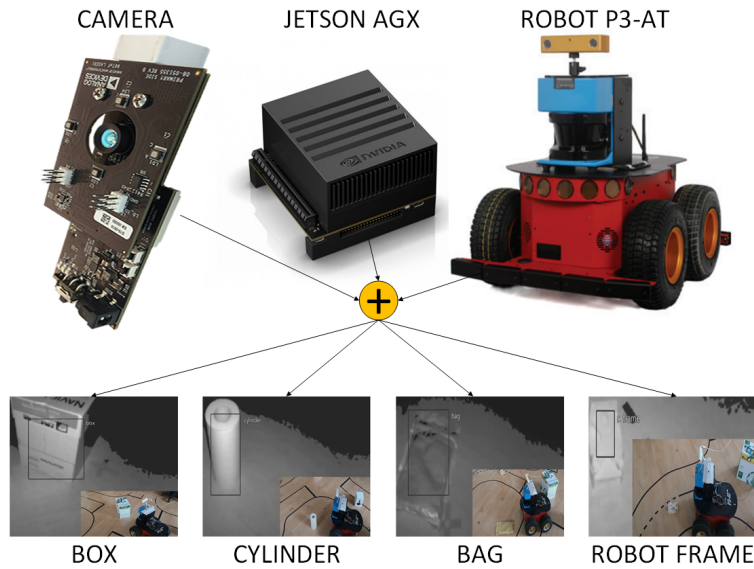


Fig. 8 Exemple de înregistrare din experimente cu echipamente înglobate

S-au analizat mai multe tipuri de camere ToF pentru algoritmi propuși, rezultatele acestor investigații fiind prezentate în Fig 9:

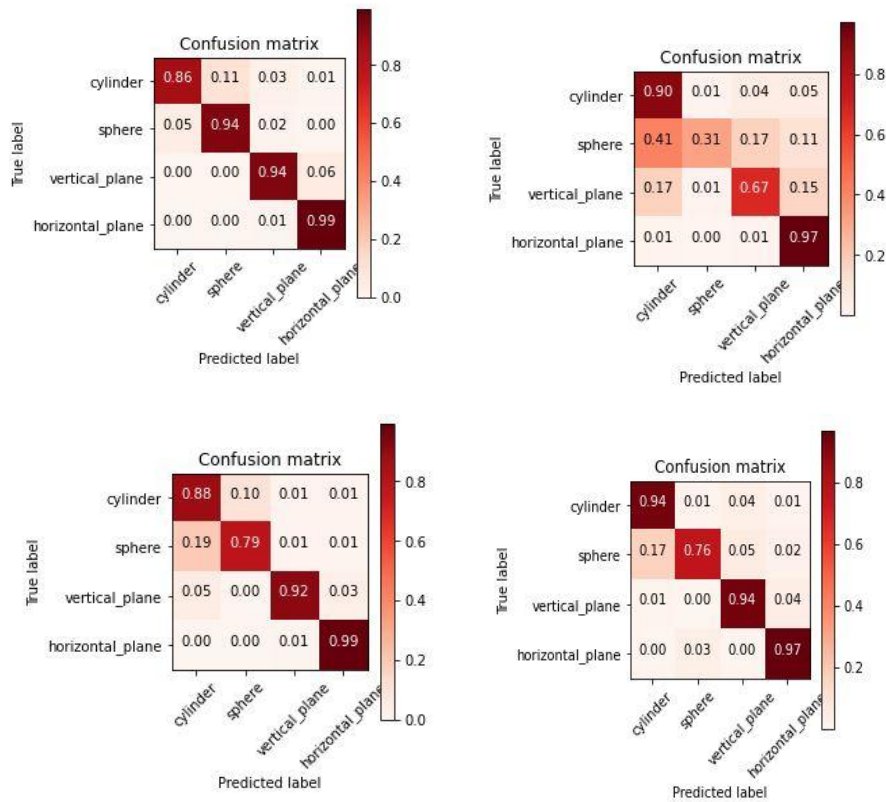


Fig. 9 Comparație pentru diferite camere 3D pentru algoritmi propuși

Concluzii

În concluzie, activitățile propuse în etapa 2 a proiectului au fost atinse cu succes (inclusiv pe partea de recunoaștere de obiecte, transfer de cunoștințe și diseminare), iar aceste rezultate vor servi ca și bază pentru activitățile propuse în etapa 3.

Bibliografie

1. Tamas, Levente; Cozma, Andrei; Embedded real-time people detection and tracking with time-of-flight camera Proc. of SPIE Vol 11736 117360B-1 2021
2. Pop, Marian-Leontin; Molnar, Szilard; Pop, Alexandru; Kelenyi, Benjamin; Tamas, Levente; Cozma, Andrei; CNN Based ToF Image Processing SciPy 2021
3. Molnár, Szilárd; Kelényi, Benjamin; Tamas, Levente; Feature Pyramid Network Based Efficient Normal Estimation and Filtering for Time-of-Flight Depth Cameras Sensors 21 18 6257 2021 Multidisciplinary Digital Publishing Institute
4. Molnar, Szilard; Kelenyi, Benjamin; Tamas, Levente; ToFNest: Efficient normal estimation for time-of-flight depth cameras Proceedings of the IEEE/CVF International Conference on Computer Vision 1791-1798 2021