

Guitar HandTracking

*Recunoaștere note muzicale

Butnaru Raimond-Eduard

Universitatea Tehnică "Gheorghe Asachi" din Iași

Iași, România

raimond-eduard.butnaru@student.tuiasi.ro

Guriuc Vlad Ionuț

Universitatea Tehnică "Gheorghe Asachi" din Iași

Iași, România

vlad-ionut.guriuc@sudent.tuiasi.ro

Abstract—Chitara este unul dintre cele mai populare și cele mai folosite instrumente. Popularitatea chitării o face unul din cele mai evidente alegeri pentru toți cei care își doresc să învețe să cânte la un instrument. Având în vedere popularitatea instrumentului, evident, putem discuta și despre gama largă de materiale online ce ajută la învățare și la perfecționare. Însă un mare dezavantaj al materialelor aflate în mediul online este lipsa flexibilității. Pe lângă aceasta, devine tot mai greu procesul de învățare online fără ajutorul unor profesioniști, de care nu toți dispunem. În acest document propunem o metodă de recunoaștere a notelor, respectiv a acordurilor cântate la chitară acustică/clasică. Metoda prin care ne-am propus să realizăm proiectul, exclude utilizarea procedurilor de machine learning pentru a crește portabilitatea produsului final.

I. INTRODUCERE

Procesul de învățare a unui nou instrument, cum este și chitara, poate fi interesant și plăcut pentru pasionații de muzică, din nefericire acest proces devine greoi și neplăcut în lipsa ajutorului unui chitarist profesionist. Chitaristul este cel responsabil pentru felul în care începătorii învață acordurile, chitaristul, ar trebui să corecteze greșelile pe care le face începătorul și să îi analizeze performanțele. Însă lecțiile de chitară pot fi destul de costisitoare și pot consuma destul de mult timp. Alternativa, ar fi cursuri de chitară online sau aplicații mobile care facilitează procesul de învățare, dar până și acestea au dezavantajele lor. Sistemul de cursuri online pentru chitară nu este metoda ideală de a învăța deoarece nimeni nu poate corecta greșelile celor ce participă în acest curs, ceea ce face ca feedback-ul interactiv să nu existe deloc. Posibilitatea de a învăța prin aplicații mobile vine și cu un set de dezavantaje, unul dintre cele mai mari dezavantaje sunt lipsa instrumentului propriu-zis, unele aplicații forțează utilizatorul spre a folosi instrumentele virtuale care sunt doar simulări ale variantelor reale, altele se folosesc de microfonul telefonului pentru a testa acuratețea, dar în funcție de alte zgomote din jur acuratețea aplicației poate scădea drastic, iar restul aplicațiilor permit doar un număr de lecții pe zi, devenind o povară și făcând în moduri tot mai insistente reclame la un abonament care ar scăpa de reclame și de limitele de lecții pe zi. Scopul acestei lucrări este a facilita un nou proces de învățare mult mai rapid și mai eficient care aduce și feedback interactiv.

Prin proiectul nostru, ne propunem o alternativă de facilitare a procesului de învățare eficientă, folosind camera stereo ZED Mini, aplicația va capta cadre ale feed-ului live și va recunoaște

acorduri și eventualele imperfecțiuni în execuția acestora. O posibilă extensie a acestei aplicații, ar fi încărcarea de videoclipuri cu recunoaștere în timp real, ajungând la generare automată de partituri.

A. Motivație

Există grupuri foarte mari de oameni care își doresc să învețe să cânte la chitară. Tot acest grup de oameni încep să își piardă motivația când procesul de învățare devine unul foarte greoi și neplăcut. Cel mai greu este pentru cei care nu merg la vreun curs de chitară sau au vreun ghid profesionist, pentru această categorie de persoane se pot crea întreprinderi greșite care rămân în timp neobservate.

B. Problema

Înainte ca cineva să înceapă vreun curs de specialitate pentru chitară, sigur a căutat cândva pe YouTube: *How to play the guitar* și au descoperit o sumedenie de clipuri care nu au nicio legătură între ele și au în comun același principiu: "Reproduce ceea ce vezi chiar dacă nu știi cât de corect este". De menționat este și posibilul zgomot des întâlnit în videoclipuri, care poate crea disconfort sau poate presa utilizatorul spre a repeta aceeași secțiune a clipului până când aude ceea ce crede el că ar trebui să audă, proces care consumă foarte mult timp de altfel. Este nevoie de un sistem portabil care face învățarea acestui instrument accesibilă pentru toate lumea, oricând și oriunde.

C. Cum se formează acordurile

Acordurile se formează doar pe brațul chitării, urmând teoriile muzicale, avem 12 acorduri de bază, majore și minore, cu forme unice ce pot fi executate pe intervalul de taste 1-5, ele sunt cunoscute și ca "Base chords". Mai departe pe gâtul chitării avem "Barre chords", aceste acorduri au forma mâinilor foarte similare între ele, însă diferă poziția acestora pe gâtul chitării. Spre exemplu, acordul Fa major (F), de regulă se afla între tastele 1 și 3, însă, executat pe intervalul de taste 3-6 acesta devine Sol major (G).

II. CONCEPTE ASEMĂNĂTOARE

În lucrarea lui Yingdong Ru intitulată *Computer Assisted Chord Detection Using Deep Learning and YOLOV4 Neural Network Model*[1] este prezentată detectarea acordurilor de

pe partituri folosind machine learning, predispunând aplicația la diferite seturi de date care aveau de recunoscut unde se află notele pe portativ. Un alt articol asemănător este lucrarea lui ALbert Mitjans Coma intitulată *Visual recognition of guitar chords using neural networks*[2]. În acest articol este pusă în vedere presupunerea mai multor acorduri în urma detectării mâinii, urmând mai apoi intersectarea cu rezultatul cel mai apropiat de adevăr, adică cel mai credibil. Lucrarea cea mai asemănătoare este *GuitarGuru: A Realtime Guitar Chords Detection System* [3]. În această lucrare, într-adevăr, se folosește machine learning, însă apare doar în procedura de comparare, nu și în cea de recunoaștere. Recunoașterea acordurilor este realizată pe baza calculului formelor mainilor și a punctelor de discontinuitate aflate pe fiecare tastă a fiecărei corzi. Din moment ce este foarte importantă mimica mâinilor pentru detectarea acordurilor, următoarea lucrare se referă la un sistem de recunoaștere a mimicii mâinilor care captura gesturile făcute de către mâinile utilizatorului pentru a controla un sistem computerizat[4].

III. METODA UTILIZATĂ

Metoda propusă spre implementare utilizează un algoritm de procesare a imaginilor captate dintr-un feed live, oferit de către camera stereo, iar acest algoritm se poate împărți în 3 etape principale:

- Ștergerea fundalului și a secțiunilor redundante
- Recunoașterea și secționarea corzilor și a tastelor
- Recunoașterea acordurilor/notelor

A. Ștergerea fundalului și a secțiunilor redundante

Pentru a evidenția chitara în interiorul imaginii, vom elimina fundalul și alte "informații" redundante din imaginea captată. Metoda finală va trebui să elimine aceste informații inutile fără modificarea secțiunii principale, va trebui să se adapteze la variații de lumină și să aibă un complexitate redusă pentru a nu apărea desincronizări în timpul rulării aplicației. Pentru scopul nostru, am testat algoritmii Running average Gaussian, Mixture of Gaussians (specific OpenCV) și Frame difference. Frame difference este eficient dar schimbul fundalului poate duce la erori majore, fiind nevoia de actualizare frecventă a fundalului în manieră manuală pentru a reduce erorile.

Mixture of Gaussian folosește un set de imagini pentru a construi un model de extragere, fiind foarte precis și cu posibilitatea de actualizare a modelului în timp de real, acesta în schimb scade destul de mult performanțele oferite de aplicație.

Running average Gaussian [5] este eficient, suportă schimb de fundal, dar pierde informații în condițiile în care pixelii nu suferă variații observabile.

În această lucrare ne propunem să folosim o metodă combinată dintre acești algoritmi pentru a menține performanțele și a scădea din pierderile de informații.

B. Recunoașterea și secționarea corzilor și a tastelor

Informațiile redundante fiind eliminate, aplicația se poate concentra asupra obiectului în discuție, chitara, fiind necesară



Fig. 1. Imagine cu fundal eliminat



Fig. 2. Imagine cu filtru Canny aplicat

segmentarea tastelor de pe brațul chitării. Rezultatul anterior va fi trimis mai departe spre procesare, considerând doar chitaristul și chitara, tastele chitării pot fi filtrate evidențiind liniile paralele verticale aplicând filtru Canny, respectiv filtru Sobel vertical.

Pentru minimizarea erorilor ce pot apărea la filtrare, vom decupa brațul chitării din imaginea principală, concentrându-ne și mai mult pe chitară în sine. Cel mai mare obstacol pe care îl putem întâlni este iluminarea slabă, dacă lumina nu bate într-un mod specific pe chitară, există foarte mare șanse ca aplicația să nu vadă corzile sau să nu vadă tastierele.

În continuare, dorim să aplicăm o combinație între filtrul Canny și transformata Hough probabilistică. Pe baza liniilor orizontale create de transformata Hough, vom roti imaginea până când acestea ajung să fie perfect orizontale. Rotirea este realizată în urma aproximării unui centru pe brațul chitării, apoi sunt calculate punctele extreme care sunt rotite în jurul centrului până se ajunge la o poziție orizontală.

C. Recunoașterea acordurilor/notelor

Pentru pasul de recunoaștere ar trebui să ajungem cu 2 imagini grayscale, una dintre ele are evidențiate liniile verticale, adică tastierele, iar cealaltă are evidențiată liniile orizontale, adică corzile. Recunoașterea acordului va avea la bază o



Fig. 3. Imagine cu filtru Canny decupată

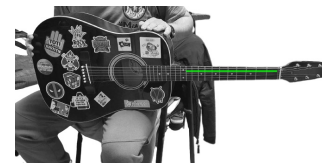


Fig. 4. Aplicarea transformatei Hough probabilistică

intersecție între întreruperi, spre exemplu, în imaginea cu linii verticale se va calcula numărul tastelor pe care se află mâna, iar în imaginea cu liniile orizontale se vor calcula corzile pe care se afla degetele. Comparând cu schemele acordurile scrise în prealabil, va apărea recunoașterea propriu-zisă.

IV. DESCRIEREA MODULELOR UTILIZATE

A. *imageModule*

Scopul funcțiilor prezente în acest modul este unul simplu, dar vital. Aceste funcții au rolul de a returna un mesaj de eroare în cazul în care imaginea sau calea sa nu pot fi citite, de a returna imaginea dorită sau de a aplica imaginii date, după caz, un filtru Canny, Sobel, Laplace sau transformata Hough. Practic în acest modul descrie clasa imaginilor ce urmează a fi procesate.

B. *Rotate and crop*

În acest modul regăsim o serie de funcții denumite sugestiv, ce ne ajută să rotim o imagine dată astfel încât să putem aduce gatul chitării într-o poziție standard pentru a facilita identificarea acordurilor prin aplicarea ulterioară a filtrelor CannyEdge și a transformatei Hough. Odată ce imaginea este rotită, gatul chitării va fi decupat din imagine pentru a optimiza procesul de identificare. Pentru a asigura stabilitatea algoritmului am introdus o funcție de resize ce va redimensiona imaginile pe o rezoluție de maxim 1920x1080.

C. *Watershed segmentation*

În prima instanță, acest modul convertește imaginea dată ca parametru într-o imagine cu tonuri de gri ce va fi binarizată cu ajutorul binarizării Otsu/threshold. După ce erodăm (reducem numărul de obiecte din imagine) și dilatăm (extindem dimensiunea obiectelor rămase) imaginea, creăm o imagine de fundal. Combinăm imaginea de fundal cu cea dilatăta și vom obține marcatorii cu ajutorul cărora se va aplica filtrul Canny. După ce aplicăm algoritmul watershed pentru a segmenta imaginea, o binarizăm cu Otsu și apoi o inversăm. Aplicăm operația AND pe biți între imaginea originală, unde imaginea binară va fi mască și repetăm procesul dar de această dată mască va fi imaginea binară inversată. Bazându-ne pe aceste diferențe putem contura imaginea finală.

D. *Grid detection*

Acest modul primește un parametru reprezentat de o imagine cu gatul chitării. În funcția "string detection" pe această imagine se detectează corzile folosind transformata Hough. Se creează o imagine cu corzile pe baza detecției și se delimitează imaginea în porțiuni verticale pentru a calcula spațiile dintre corzi. Calculează diferențe de pixeli pentru a identifica pozițiile unde pot fi prezente corzi și trasează linii orizontale în dreptul corzilor. Funcția "fret detection" primește ca parametru aceeași imagine și e folosit pentru a detecta tastele prin decuparea imaginii în segmente orizontale și calcularea spațiilor dintre ele. Odată ce aceste operații sunt terminate, se vor afișa linii verticale deasupra tastelor detectate.

V. DIFICULTĂȚI ÎNTÂMPINATE

Primele dificultăți au început să apară după rezultatele intermediare. Pentru început, detecția de corzi are șanse foarte mari de eșec la ultima, respectiv cea mai subțire coardă, recunoscând astfel 5 din 6 corzi, o soluție ar putea fi presupunerea în mod implicit a 6 corzi și căutarea continuă a exact 6 corzi. Mai apoi avem detecția tastelor de pe brațul chitării care are secțiuni desincronizate, soluție ar fi segmentarea pe spații albe. Ultima dificultate, este detectarea vârfurilor degetelor.

VI. CONCLUZIE

Chitara este un instrument foarte popular și multă lume dorește să învețe să cânte la acesta. Dar evident, există dificultăți în procesul de învățare și există foarte puține platforme user-friendly care să faciliteze acest proces. De aceea am luat inițiativa de a dezvolta un sistem ce ajută chitariștii începători.

VII. OBIECTIVE PE VIITOR

Ne propunem ca în viitor să ne reîntoarcem la această aplicație și să îi sporim portabilitatea și accesibilitatea, dar și să îi mărim repertoriul de acorduri la acorduri și mai complicate cum sunt cele cu septimă sau cele suspendate.

VIII. RECUNOȘȚINȚĂ

Suntem recunoscători de această oportunitate și mulțumim Facultății de Automatică și Calculatoare a Universității Tehnice "Gheorghe Asachi" din Iași pentru suportul și pentru facilitățile oferite în cadrul dezvoltării proiectului.

REFERENCES

- [1] Computer Assisted Chord Detection Using Deep Learning and YOLOV4 Neural Network Model by Yingdong Ru, 2021.
- [2] Visual recognition of guitar chords using neural networks by Albert Mitjans Coma and Artur Carnicer.
- [3] GuitarGuru: A Realtime Guitar Chords Detection System by Varsha Nagpurkar, Neha Pattankar, Tripti Nayak Allan D'Souza and Nipun Henriques, 2023.
- [4] B., Abhishek, Krishi, Kanya M., Meghana, Daaniyaal, Mohammed S., Anupama. "Hand gesture recognition using machine learning algorithm" Computer Science and Information Technologies 2020.
- [5] Cerman Martin : SUBTRACTION USING RUNNING GAUSSIAN AVERAGE : A COLOR CHANNEL COMPARISON