

Drop-D

Guitar note recognition

Paniş Alexandru Onofrei Grigore

Abstract—În acest document noi propunem o metodă de determinare a notelor cântate la chitară (de diferite tipuri, ex. : număr variabil de string-uri, fret-uri) din semnal video. Ca scopuri ne-am pus realizarea proiectului neapelând la camere stereo ,sisteme cu două camere, senzori de poziție sau marker-e colorate astfel modelul poate fi folosit mai de către o audiorie mai largă. La fel ca constrângere am renunțat la metode bazate pe machine learning. Ca rezultat aplicația poate fi folosită având doar un laptop cu o camera destul de modestă.

1 ÎNTRUDUCERE

Memorarea poziționării notelor pe neck-ul chitarei este prima problemă ce apare în calea începătorilor chitariști. Modelul prezentat în curentul document propune o soluționare a acesteea, folosind o cantitate modestă de resurse: un calculator cu camera (în teorie s-ar putea folosi și pe telefoane mobile care au destulă putere de procesare pentru folosire comodă).

2 LUCRĂRI ASEMĂNĂTOARE

Unul dintre primele modele pe această temă este propus de Chutisant Kerdvibulvech și Hideo Saito [1], însă acesta include un marker (ARTag) pe vârful gâtului chitarei pentru a determina poziția ei, marker-e colorate pentru fiecare deget la mâna de selecție, machine learning pentru a identifica degetele be baza acelor marker-e și , cel mai important, un sistem cu două camere pentru vedere spațială. Evident o asemenea abordare nu se încadrează în constrângerile menționate.

O lucrare foarte asemănătoare [2], realizată de Joseph Scarr și Richard Green, la fel exclude echipamente sau sisteme complexe, însă unele etape ale prelucrării sunt descrise foarte vag sau duc lipsă de detalii. Diferențele dintre metoda prezentată și cea descrisă în [2] vor fi discutate în descrierea metodei la etapele respective.

Merită menționată și lucrarea [-1], fiind o abordare foarte asemănătoare, doar că nefinisată și descrisă foarte succint.

3 DESCRIEREA METODEI

Metoda propusă folosește un algoritm de prelucrare a imaginilor live captate de camera calculatorului, și include trei etape generale:

- eliminarea background-ului.
- Segmentarea neck-ului.
- recunoașterea notelor.

3.1 Eliminarea background-ului

Aceasta este o problemă deseori întâlnită sistemele de prelucrare a imaginilor, de regulă pentru a evidenția obiecte mobile într-un set de imagini, spre deosebire de modelul curent

unde este folosit pentru a exclude informații redundante care pot altera rezultatul de la etapa următoare, mai multe detalii în paragraful 3.2.

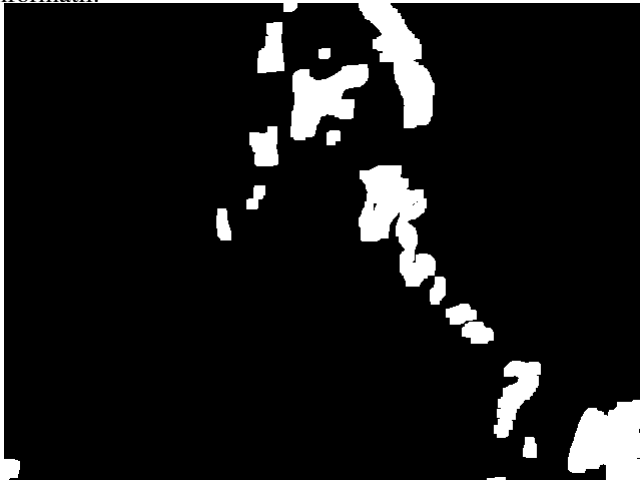
Metoda aleasă trebuie să îndeplinească următoarele cerințe: eliminarea background-ului fără alterarea informației necesare, adaptare la variații de luminozitate, cu o complexitate redusă pentru a putea fi folosit în real time.

Ne-am propus să testăm 3 algoritmi: Frame Difference, Running average Gaussian și Mixture of Gaussians(OpenCV).

Frame difference este foarte eficient, dar schimbul ulterior al background-ului duce la erori, acesta necesită actualizarea periodică a background-ului în regim manual când acesta suportă variații.



Running average Gaussian [3] - eficient, suportând schimb de background, însă după un anumit timp pixelii care nu au suportat o variație mare devin background, deci pierdere de informații.



MoG - folosește un set de imagini pentru construirea unui model de extragere. Oferă precizie (nu foarte necesară cazului

nostru) si ofera posibilitatea de actualizare a modelului in real time, inasa cu o performanta foarte scazuta.



Ne propunem sa incercam o combinatie de Running average Gaussian combinat cu Frame Difference [4] sau combinat cu Edge-Based [5].

3.2 Segmentarea neck-ului

După eliminarea informațiilor redundante este necesară segmentarea gâtului chitarei. Fiind cunoscut faptul că în imagine, după etapa anterioară, avem doar chitaristul în sine și instrumentul său, concludem că gâtul chitarei va reprezenta cel mai mare cluster de linii drepte din imagine, deci putem aplica o combinație dintre algoritmul Canny (OpenCV) de detecție a muchiilor și transformata Hough probabilistică (OpenCV). Pentru detecția muchiilor se calculează un threshold adaptiv pentru fiecare frame pe baza metodei Otsu(OpenCV), în urma unor experimente am dedus valoarea optimă pentru threshold-ul low Canny - 1/5 din cel calculat. La fel printr-un set de experimente am dedus și valori optime pentru aplicarea transformatei Hough:



Apoi folosind valoarea medie a orientării liniilor detectate, se face o rotire a imaginii, neck-ul fiind orizontal, aceasta va fi util pentru următoarea etapă.

3.3 Recunoașterea notelor(urmează a fi implementată și pot apărea modificări)

Folosind datele liniilor calculate cu ajutorul transformatei Hough se segmentează din imagine doar regiunea neck-ului. Apoi folosind datele introduse de utilizator la început: numărul string-urilor și fret-urilor, se poate aproxima foarte bine poziția lor. Apoi se determină zonele din imagine unde se află degetele, ca rezultat avem notele în imaginea curentă.

REFERENCES

- [1] Chutisant Kerdvibulvech and Hideo Saito: Guitarist Fingertip Tracking by Integrating a Bayesian Classifier into Particle Filters
- [2] Joseph Scarr and Richard Green: Retrieval of Guitarist Fingering Information using Computer Vision
- [3] Cerman Martin : SUBTRACTION USING RUNNING GAUSSIAN AVERAGE : A COLOR CHANNEL COMPARISON
- [4] Zhen Tang, Zhenjiang Miao and Yanli Wan : Background Subtraction Using Running Gaussian Average and Frame Difference
- [5] Sumer Jabri, Zoran Duric, Harry Wechsler and Azriel Rosenfeld : Detection and Location of People in Video Images Using Adaptive Fusion of Color and Edge Information