Jedan pristup u zvučnoj interpretaciji slike

Nebojša Kolarić, dr Dragana Šumarac Pavlović

***Apstrakt*—U različitim domenima ljudskih aktivnosti postoji potreba da se izvrši zvučna interpretacija statičkih ili pokretnih slika. U savremenoj umetničkoj praksi se sreće često potreba za uspostavljanjem korelacije između slike i zvuka. Jedna od primena je i razvijanje grafičkog interfejsa za kreiranje zvuka. Sa druge strane promene u slici u sistemima za monitoring mogu se pratiti i na osnovu promena u zvuku. U radu je realizovan sistem za interpretaciju slike zasnovan na uspostavljenim korelacijama između slike kao dvodimnezionalnog objketa i zvuka prikazanog u vremenko-frekvencijskoj ravni. Osnovna ideja je bila da se realizuje zvučni signal uobličene melodijske i harmonijske strukture zasnovane na pojedinačnim tonskim elementima tako da se može uspostaviti korelacija između sadržaja slike i odgovarajućih prepoznatljivih elemenata u zvuku koji ih prati. U radu je realizovan sistem koji uz pomoć pokretne kamere formira niz slika koje se zvučno interpretiraju u realnom vremenu.**

***Ključne reči*—zvučna interpretacija slike, segmentacija slike, serijska komunikacija, Arduino ….**

I. Uvod

Zvučni signali predstavljaju vremenski promenljive jednodimenzionalne signale. Informacioni aspekt zvučnih signala sadržan je u različitim elementima kojima možemo opisati signal u vremenskom domenu, a koji se pre svega odnose na sporo promenljivu anvelopu signala kojom su određene mnoge značajne osobine signala. Jedna od važnih odlika zvučnih signala u načelu je njihova velika dinamika. Sa druge strane zvučni signali, bilo da se radi o govrnim signalima, muzičkim ili različitim oblicima ambijentalnog zvuka su vremenski i spektralno promenljivi i samo na kraćim segmentima možemo govoriti o njihovoj kvazi stacionarnosti.

Različite tehnike za analizu i sintezu zvučnih signala zasnovane su na kratkovremenoj Furijeovoj transformaciji. Kratkovremena Furijeova transformacija vrši preslikavanje jednog jednodimenzionalnog vremenskog niza u dvodimenzionalni prostor vremensko-frekvencijskih amplitudskih i faznih promena signala. Dvodimanzionalni prostor vremensko-frekvencijskih promena može se interpretirati kao slika, odnosno svaka slika može se pod određenim pretpostavkama interpretirati kao jedna predstava vremensko frekvencijskih promena zvučnog signala.

Polazeći od ovih pretpostavki moguće je na mnogo načina interpretirati sliku kao jedan oblik kratkovremene Furijeove transformacije nekog vremenskog signala. Svaka vrednost na slici s(x,y) može se pridružiti nekom vremenskom trenutku (x komponenta) i spektralom binu (y kooridinata).

Jedna od direktnih interpretacija slike može se svesti na posmatranje slike kao amplitudskog spektra dobijenog kratkovremenom Furijeovom transformacijom. Broj piksela slike po jednoj koordinati određuje vremensko trajanje zvučne interpretacije, dok broj piksela po drugoj kooordinati određuje frekvencijski sadržaj zvučnog signala. Moguće su različite korelacije koje se uspostavljaju između prostora zvučnih informacija i slike.

U radu je predložena interpretacija slike prema kojoj x koordinata određuje vremenske odrednice trajanja zvučnog segmenta, a y koordinata se interpretira preko niza tonova temperovane muzičke skale.

Osnovna ideja predložene interpretacije zasniva se na polaznim pretpostavkama da se može uspostaviti intuitivna korelacija između vizuelnog i zvučnog sadržaja, kao i da se obrada slike na predloženom hardveru može izvršavati u realnom vremenu. Generisanje muzičkog signala na bazi muzičkih tonova temperovane skale omogućava formiranje ritmičkih, melodijskih i harmonijskih struktura. Dodatna motivacije za realizaciju ovog sistema je proistekla iz nepostojanja nijednog ovakvog sistema. Nisu pronađeni nikakvi radovi niti projekti da je neko realizovao zvučnu interpretaciju slike tako da ona zvuči smisleno i još da radi u realnom vremenu.

Da bi se omogućilo kontinualno generisanje muzičkog signala bilo je neophodno sukcesivno snimanje slika promenljivog sadržaja. U tom cilju predložena je hardverska realizacija sistema za generisanje slika zasnovana je na pokretnoj kameri koja se nalazi na motorizovanom obrtnom postolju čime se omogućava kontinuirano prikupljanje slika iz velikog prostornog ugla različitog sadržaja. Slike sa kamere se u realnom vremenu obrađuju i sukcesivno koriste za generisanje kontinualnog zvučnog signala.

U radu je dat opis hardverske realizacije sistema za prikupljanje, obradu i diskretizaciju slika kao i algoritam za interpretaciju podataka iz obrađene slike u zvučni signal.

II. Hardverska realizacija

Hardverska realizacija celokupnog sistema za generisanje slika prikazana je na slici 1. Sistem treba da obezbedi kontinuirano prikupljanje slika promenljivog sadržaja u diskretnim vremesnkim trenucima. Sistem čine: arduino kontroler (1), sistema za napajanje servo motora za pokretanje kamere (2) i kamera postavljena na obrtno postolje sa servo motorom (3). Arduino platforma, kamera i napajanje servo motora priključeni su na računar.

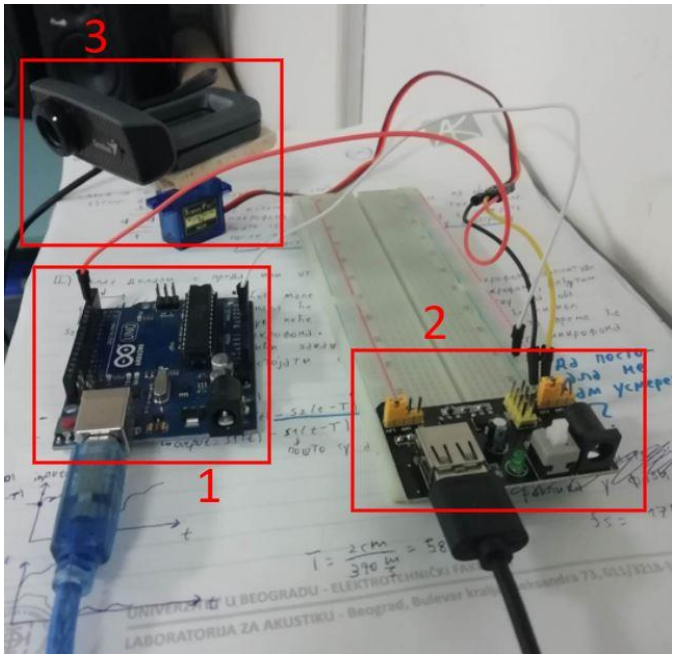
Kamera je postavljena na obrtno postolje kojim upravlja servo motor koji se napaja iz posebnog izvora. Arduino

Nebojša Kolarić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11120 Beograd, Srbija (e-mail: nebojsa.kolaric99@gmail.com)

Dragana Šumarac Pavlović – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11120 Beograd, Srbija (e-mail: dsumarac@etf.rs)

 ISBN 978-86-7466-969-3

platforma je upotrebljena za davanje instrukcija servo motoru za obranje kamere, međutim Arudino kontroler ne zna kada treba kamera da se zarotira jer se cela obrada vrši na računaru. Ove informacije računar šalje Arduino kontroleru preko USB porta tako što se šalje „$“ simbol, Arduino sve vreme isčitava poruke koje dobija sa USB porta i u trenutku kada pronađe „$“ simbol rotira kameru za fiksan ugao (koji je ranije definisan). Na računaru se osim obrade i slanja informacija Arduino kontroleru vrši i generisanje zvučnih signala koji se kasnije šalju na zvučnik. Obrtno postolje je upotrebljeno kako bi se obezbedio drugačiji sadržaj slike i određena dinamika u generisanom zvučnom signalu. U predloženoj proceduri jedna slika se konvertuje u zvučni zapis dužine 1min. To je vreme u kome sistem može da obradi sliku i pripremi ulazne podatke za generisanje zvučnog signala, odnosno to je perioda sa kojom se generišu nove slike.



Sl. 1. Izgled hardvera za upravljanje kamerom

III. OPIS ALGORITMA ZA SINTEZU ZVUKA IZ SLIKE

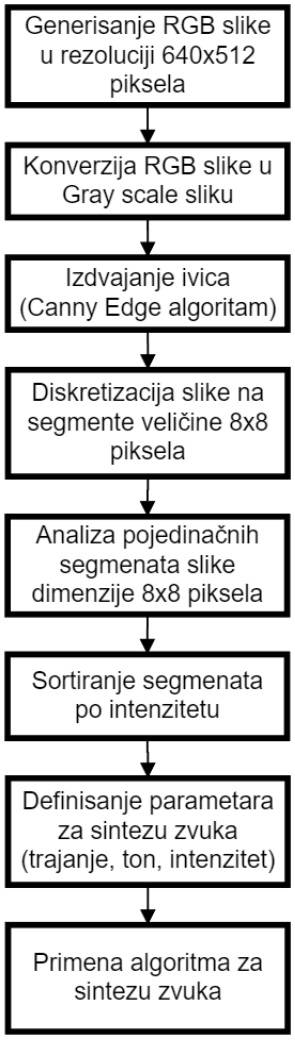
Sistem za sintezu zvuka na bazi slike koju generiše kamera zasniva se na više etapa obrade. Na slici 2 je prikazana blok šema obrade slike i princip generisanja zvuka. Algoritam za obradu slike i sintezu zvuka obuhvata sledeće etape:

1. Generisanje RGB slike u rezoluciji 640x512 piksela
2. Konverzija RGB u Gray scale sliku
3. Izdvajanje ivica (slika 3)
4. Diskertizacija slike na segmente veličine 8x8 piksela (slika 4)
5. Analiza pojedinačnih segmenata slike dimenzija 8 x 8 piksela
6. Sortiranje segmenata po intenzitetu
7. Definisanje parametara za sintezu zvuka: trajanje, frekvencija (ton) i intenzitet
8. Primenu algoritma za sintezu zvuka

Kompletan algoritam za obradu slike i sintezu zvuka realizovan je programskom okruženju python.

Kamerom se u diskretnim vremenkim trenucima generiše RGB slika rezolucije 640 x 512 piksela. Da bi se smanjila količina informacija sadržanih u slici i izdvojile informacije potrebne za sintezu zvuka, slika se najpre konvertuje u Gray-scale format.

Na sliku u Gray scale formatu se primenjuje Canny Edge algoritam [1] za detekciju ivica. Na ovaj način se smanjuje količina informacija u slici na osnovu kojih je moguće formirati zvučni signal na bazi muzičkih tonova. Rezultat Canny Edge algoritma je crno bela slika, gde beli pikseli označavaju ivice (slika 3). Jos jedan od razloga za upotrebu ovog algoritma je taj što je krajnji cilj bio upotrebiti neku transformaciju koja će redukovati broj informacija, ali očuvati dovoljno informacija da posmatrač može sa sigurnošću da kaže da je to i dalje ista slika.

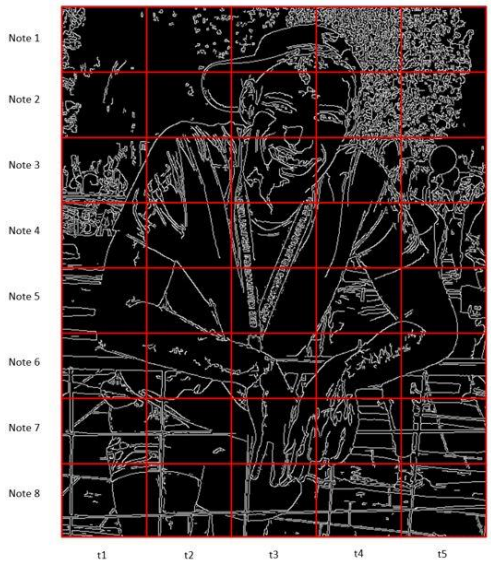


Sl. 2. Algoritam za zvučnu interpretaciju slike

 ISBN 978-86-7466-969-3

Sl. 3. Originalna slika (levo) i slika nakon Canny Edge algoritma (desno)

Sledeća faza je diskretizacija slike na manje segmente. Diskretizacija po x koordinati slike premapirana je u vremenski sled zvučnih objekata koji će se pojaviti u zvučnom signalu, a diskretizaijom po y koordinati određujemo frekvencijski sadržaj sintetizovanog zvuka. Originalna slika je diskretizovana na polja dimenzija 8 x 8 piksela. Veličina polja definisana je mogućnošću obrade u realnom vremenu sa raspoloživim hardverom. Na slici 4 je prikazan uprošćen primer diskretizacije slike na 8 mogućih tonova (nota) i 5 vremenskih intervala. Crvenim linijama označene su granice segmenata.



Sl. 4. Primer diskretizacije slike po vremenu i po frekvenicjama (tonovima)

Ulazna slika u ovaj sistem za sintezu zvuka, slika sa kamere ima rezoluciju 640x512 piksela. U predloženoj realizaciji sistema, slika je diskretizovana na 80 frekvencijskih opsega (tonova, nota) i 64 vremenska intervala. Svaki vremenski segment ima određeno trajanje koje će kasnije biti objašnjeno. Izbor 80 tonova se može translirati na frekvencijskog skali proizvoljno.

Vertikalna kolona diskretizovanih delova slike određuje ono što će činiti zvučni sadržaj u prvoj sekundi. U svakom od segmenata sumira se broj belih piksela na osnovu čega se vrši sortiranje i 4 segmenta sa najvećim brojem belih piksela usvajaju se kao ulazni parametri za sintezu zvuka. Svaki od segmentiranih delova slike kodiran ukupnim brojem belih piksela (intenzitetom) na osnovu koga se određuje intenzitet visina i trajanje tona koji se generišu.

Za sintezu pojedinačnih tonova (80), upotrebljena je scamp.py python biblioteka. Ova biblioteka ima mogućnost generisanja muzičkih nota na različitim instrumentima (gitara, klavir, violina, ...) kao i mogućnost reprodukcije više tonova odjednom.

Nakon završene obrade i donete odluke na koji način će biti sintetizovan zvuk u prvom vremenskom segmentu, zvuk se generiše i reprodukuje na zvučniku, a za to vreme se vrši obrada nad segmentima koji odgovaraju drugom vremenskom intervalu i tako sve dok se ne dođe do kraja slike.

Nakon toga se zapamte 4 segmenta (note koje će potencijalno biti odsvirane) sa najviše broja belih piksela. Razlog zašto je broj segmenata ograničen na četiri je to što bi mnogo segmenata prošlo, samim tim veliki broj nota bi bio odsviran odjednom i to ne bi zvučalo ni na šta. Naravno ovo važi za slike sa puno detalja, kao što je prikazano na slici 2, dok za slike sa dosta manje detalja ovo ne bi bilo neophodno. Naredni korak je sortiranje nota, od note sa najviše belih piksela do note sa najmanje belih piksela. Ove informacije o redosledu nota se čuvaju u jednom nizu, a informacije o brojevima belih piksela za svaku notu se čuvaju u drugom nizu. Nakon toga se primenjuje generator slučajnih brojeva koji generiše broj od 1 do 4. Ovaj broj označava koliko nota će biti odsvirano (od četiri koje su u opticaju). Na primer ako generator generiše broj 2 onda će prve dve note biti odsvirane. Uloga ovog generatora je da eliminiše monotonost melodije koja nastaje, jer da njega nema skoro uvek bi bile odsvirane četiri note odjednom. Na kraju dolazi deo za odlučivanje jačine i trajanja nota koje trebaju da budu odsvirane, a to se vrši prema sledećim formulama:

$$Amplitude = 0.3 + n * 0.025 \quad (1)$$

$$Time = 0.1 + n * 0.0125 \quad (2)$$

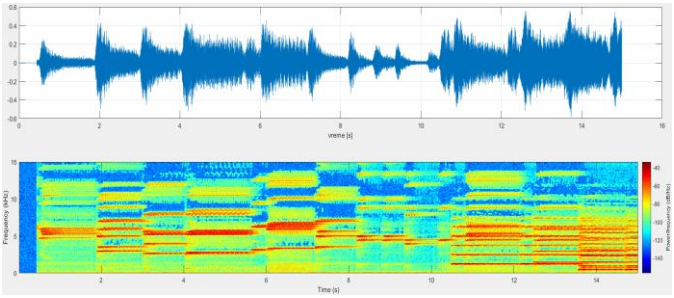
Amplitude označava jačinu note, *Time* označava vreme trajanja note i *n* označava zabeležen broj belih piksela za notu koja treba da bude odsvirana. Takođe, ovaj broj *n* se ograničava na vrednost 24 ako je zabeleženo preko 24 belih piksela u segmentu. Ovaj broj nije slučajno izabran već je eksperimentalno određen. Analizirano je pet različitih slika i beležene su vrednosti promenljive *n* za svaku notu koja bi bila odsvirana prema prethodno opisanom algoritmu. Kada se ti rezultati usrednje dobija se vrednost od 24. Pošto je maksimalna vrednost promenljive *n* 24, onda se može izračunati minimalna jačina note (0.325), maksimalna jačina note (0.9), minimalno trajanje note (0.1125 sekundi) i maksimalno trajanje note (0.4 sekundi). Formule prema kojima se računaju jačina i trajanje note su podešene tako da melodija zvuči što zanimljivije.

Na slici 5 je prikazan jedan frejm (slika) kamere nakon upotrebe Canny Edge algoritma, a na slici 6 je prikazan talasni oblik generisanog zvuka i njegov spektogram u prvih 15 sekundi. Pošto 15 sekundi odgovara četvrti trajanja cele obrade nad jednim frejmom (slikom), na slici 5 crveni pravougaonik predstavlja deo slike koji je zaslužan za

 ISBN 978-86-7466-969-3

generisanje talasnog oblika i spektograma prikazanog na slici 6. Mali podsetnik, najniža nota odgovara vrhu slike a najviša odgovara dnu slike. Sada kada analiziramo malo više od prve polovine crvenog pravougaonika, vidimo da se beli pikseli nalaze tek na dnu slike. Zbog toga očekujemo samo više frekvencije u generisanom zvuku, što se i može videti na slici 6 u spektogramu. Dok u poslednjem delu crvenog pravougaonika postoje beli pikseli na sredini slike i generisan zvuk će biti srednjih frekvencija, što se takođe može videti na spektogramu sa slike 6. Još jedna stvar koja može da se vide sa slike 6 je da trajanje note, određenih frekvencija u spektogramu, odgovaraju uočljivim oblicima anvelopa u vremenskom domenu.

Slika 5



Prvih 15 sekundi

IV. ZAKLJUČAK

U ovom radu objašnjen je implementirani algoritam za zvučnu interpretaciju slike, kao i njegova hardverska realizacija optimizovana za rad u realnom vremenu. Realizovani sistem ispunjavana tri najbitnija zahteva koji su autori postavili za cilj. Prvi je da radi u realnom vremenu. Drugi je da je generisani zvuk smislen i prijatan za slušanje, pošto su autori pronašli jedan rad gde se slika posmatra kao spektogram ali je generisani zvuk veoma loš i zvuči kao šum. Treći je da posmatrač može da pretpostavi šta će čuti na osnovu slike posle Canny Edge detektora ivica. Naravno postoje i određena ograničenja ovog sistema, ali i načini da se ona prevaziđu. Samim tim što sistem radi u realnom vremenu, postoji ograničenje u kompleksnosti primenjenog algoritma zato što je neophodno da on bude što jednostavniji i da se što pre izvrši da ne bi program unosio dodatno kašnjenje, zato i postoji ograničenje od maksimalno četiri note koje mogu da se odsviraju. Ovo bi moglo da se izbegne kada bi se koristio programski jezik C umesto Python-a jer je on optimizovan za rad u realnom vremenu. Druga mana je to što je sistem ograničen na tonove iz scamp.py python biblioteke, ali to se može rešiti snimanjen 64 različita tona koji bi zamenili tonove iz biblioteke.

Za kraj autori žele da napomenu da nisu pronašli nijedan rad ili sistem koji radi bilo šta slično, pogotovo ne u realnom vremenu i smatraju da je ovo začetak nove oblasti. Takođe smatraju da će ovaj rad mnogo značiti ljudima koji žele ovime da se bave i da ima puno mesta za nadogradnju i proširenje, na šta i ohrabruju sve zainteresovane.

LITERATURA

.

[1] Canny Edge detector Wikipedia

ABSTRACT

Abstract- In various domains of human activities, there is a need to perform sound interpretation of static or moving images. In contemporary artistic practice, there is often a need to establish a correlation between image and sound. One of the applications is the development of a graphic interface for creating sound. One the other hand, changes in the image in monitoring systems can be monitored based on changes in the sound. In the paper, a system for image interpretation was implemented based on established correlation between the image as a two-dimensional object and the sound displayed in the time-frequency plane. The basic idea was to realize a sound signal of a shaped melodic and harmonic structure based on individual tonal elements so that a correlation could be established between the content of the image and the corresponding recognizable elements in the accompanying sound. In the paper, a system was implemented that, with the help of a moving camera, forms a series of images that are interpreted sonically in the real time.

One approach in the sound interpretation of the image

Nebojša Kolarić, Dragana Šumarac Pavlović

 ISBN 978-86-7466-969-3