

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DETEKCIA VÝZNAMNÝCH OBLASTÍ
VO VIDEU

Diplomová práca

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DETEKCIA VÝZNAMNÝCH OBLASTÍ
VO VIDEU

Diplomová práca

Študijný program: Aplikovaná informatika
Študijný odbor: 9.2.9. aplikovaná informatika
Školiace pracovisko: Katedra Aplikovanej Informatiky
Školiteľ: RNDr. Elena Šikudová, PhD.



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Martin Kuchyňár
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.2.9. aplikovaná informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický


Názov: Detekcia významných oblastí vo videu
Spatio-temporal salient object detection

Cieľ: Metódy na detekciu významných oblastí vo videu:
1. Naštudovanie
2. Návrh zlepšenia
3. Implementácia
4. Porovnanie výsledkov

Vedúci: RNDr. Elena Šikudová, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. PhDr. Ján Rybár, PhD.
Dátum zadania: 20.10.2014

Dátum schválenia: 24.10.2014

prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu


.....
študent


.....
vedúci práce

Čestné vyhlásenie

Čestne prehlasujem, že som túto diplomovú prácu vypracovala samostatne s použitím uvedených zdrojov.

V Bratislave

.....

Pod'akovanie

Ďakujem svojmu vedúcemu práce za...

Abstrakt

Tu je text slovenskej verzie abstraktu

Kľúčové slová: *slovo1, slovo2, slovo3, slovo4*

Abstract

Text anglickej verzie abstraktu

Keywords: *word1, word2, word3, word4*

Obsah

1	Úvod	18
2	Prehľad literatúry	20
2.1	Úvod do problematiky	20
2.2	Metody pre statické obrázky	20
2.2.1	Baseline Center	20
2.2.2	Hrany	20
2.2.3	Ittiho model	21
2.2.4	Spektrálne rezidua	21
2.2.5	Sun Model	21
2.2.6	Rare Model	21
2.3	Metody pre videá	22
2.3.1	Zohľadnenie audio informácie	22
2.3.2	Detekcia pohybu	23
2.4	Metódy Využívajúce neurónové siete	23
2.5	Metriky úspešnosti	23
2.6	Referčné datasety	23
2.7	Porovnanie štandardných Metód	23
3	Charakteristika - obrázky, zoznamy...	24
3.1	Podkapitola o zdrojoch	24
3.1.1	Este mensia podsekcia - zoznamy	24
3.1.2	Este jedna - obrázky vedľa seba	25
3.2	Ina podkapitola	25
3.2.1	A v nej subsekcia	25
4	Návrh - ako vložiť ukážku kódu	26
5	Záver	28

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DETEKCIA VÝZNAMNÝCH OBLASTÍ
VO VIDEU

Diplomová práca

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DETEKCIA VÝZNAMNÝCH OBLASTÍ
VO VIDEU

Diplomová práca

Študijný program: Aplikovaná informatika
Študijný odbor: 9.2.9. aplikovaná informatika
Školiace pracovisko: Katedra Aplikovanej Informatiky
Školiteľ: RNDr. Elena Šikudová, PhD.



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Martin Kuchyňár
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.2.9. aplikovaná informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický


Názov: Detekcia významných oblastí vo videu
Spatio-temporal salient object detection

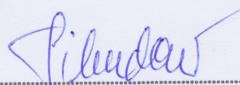
Cieľ: Metódy na detekciu významných oblastí vo videu:
1. Naštudovanie
2. Návrh zlepšenia
3. Implementácia
4. Porovnanie výsledkov

Vedúci: RNDr. Elena Šikudová, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. PhDr. Ján Rybár, PhD.
Dátum zadania: 20.10.2014

Dátum schválenia: 24.10.2014

prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu


.....
študent


.....
vedúci práce

Čestné vyhlásenie

Čestne prehlasujem, že som túto diplomovú prácu vypracovala samostatne s použitím uvedených zdrojov.

V Bratislave

.....

Pod'akovanie

Ďakujem svojmu vedúcemu práce za...

Abstrakt

Tu je text slovenskej verzie abstraktu

Kľúčové slová: *slovo1, slovo2, slovo3, slovo4*

Abstract

Text anglickej verzie abstraktu

Keywords: *word1, word2, word3, word4*

Obsah

1. Úvod

Ľudské oko je schopné spracovať 10^8 až 10^9 bitov obrazových dát za sekundu. Ľudský mozog nie je schopný spracovať také množstvo dát naraz, preto sa získané informácie filtrujú pomocou ľudského vizuálneho systému[9]. Ľudský vizuálny systém je pravdepodobne najzložitejším mechanizmom akým človek disponuje. Je natoľko kľúčovým pre fungovanie spoločnosti či jedinca, že psychológovia sa zaoberajú jeho výskumom. Už viacero dekád študujú vlastnosti tohoto mechanizmu z pohľadu psychológie, fyziológie alebo neurobiológie.

Vyfiltrované oblasti obrazu už je možné spracovať v výrazne rýchlejšie ako nefiltrované, ideálne v reálnom čase. Takéto oblasti sa nazývajú významné alebo charakteristické (v literatúre salient - prebrané v angličtine). Významné oblasti sú vyberané pomocou mnoha faktorov. Najznámejšími sú prechody vo farbe, intenzite alebo orientácií. Ľudský vizuálny systém taktiež využíva skúsenosti pri pozorovaní. Oblasť takto vyfiltrovaná nesú pre pozorovateľa viac potencionálnych informácií ako ostatné oblasti obrazu a preto sa stávajú salietnými.

V systémoch počítačového videnia sa snažíme využívať primárne tieto oblasti pre pridelenie väčšej časti zdrojov. Z tohoto dôvodu je zistenie zaujímavých oblastí častým prvým krokom mnohých algoritmov v oblasti počítačového videnia.

Algoritmy na detekciu významných oblastí sa delia do 3 skupín podľa princípu akým spracovávajú dáta[5]

1. Zdola-nahor: Prístup je cielený na nezávislosť od používateľa. Zameriava sa fyziologicky významné oblasti vizuálneho systému ako výrazné zmeny v tvare, jase alebo farbe.
2. Zhora-nadol: Prístup je založený na čiastočnom riadení zo strany používateľa (konanie je podmienené úlohou). Riadenie je prínosom pretože obsahuje aj informáciu používateľa a jeho prechádzajúcich vedomostí či skúsenosti, ktoré ovplyvňujú vnímanie.
3. Algoritmy využívajúce neurónové siete.

Cieľom práce je štúdium a výskum nových metód na detekciu významných oblastí vo videu. Následne porovnanie nových metód s existujúcimi v rôznych štandardných oblastiach ako aj

v rýchlosti výpočtu.

V prvej časti sa nachádza prehľad metód na detekciu významných oblastí vo videu, alebo metód na detekciu v statických obrazoch ktoré majú potenciál pre použitie aj vo videu. Ďalej detailné vysvetlenie fungovania metód ktoré budú použité v implementácii zlepšenia.

V druhej časti je popísaný postup a princíp zlepšenia. Následne porovnanie s metódami uvedenými v prvej časti.

V závere....

2. Prehľad literatúry

2.1 Úvod do problematiky

Saliency a teda detekcia významných oblastí je využívaná rôznych oblastiach. Počínajúc automatizáciou je ťažiskom pri segmentácii obrazu alebo detekcií špecifických objektov. Od saliency modelov sú taktiež závislé aj programy ovládajúce zabezpečovacie zariadenia. Až po reklamu kde je vizuálna pozornosť kľúčovým parametrom čo môže rozhodnúť o úspechu produktu, veď aký význam by mala reklama kde si nevšimnete prezentovaný produkt, alebo si všimnete iba jeho "menej" dokonalé časti.

2.2 Metody pre statické obrázky

Algoritmy pre statické obrázky tvoria základ všetkých saliency modelov a tvoria najstaršiu oblasť výskumu. V tejto časti uvediem prehľad algoritmov pre výpočet saliency modelov od najjednoduchších cez najznámejšie až po nejefektívnejšie. Na záver uvediem porovnanie všetkých metód pomocou všeobecne uznávaných metrík a dát získaných z zariadení merajúcich pohyb očí používateľa (eyetrackera).

2.2.1 Baseline Center

Baseline center je triviálny model ktorý sa vypočítava pomocou Gaussovej krivky vzhľadom na pomer strán čím predpokladá salientné oblasti presne v strede obrazu. Nezachytáva však žiadne sémantické aspekty videa ako ani podvedome informácie vnímanania obrazu iba rozlíšenie dané optikou skenujúcou scénu. TODO jeden obrazok-povodny/mapa/eyetracker vizualizacia

2.2.2 Hrany

Skupina algoritmov využívajúca význačné prechody v obraze inak nazývané hrany. TODO

2.2.3 Ittiho model

Najznámejším modelom pre výpočet významných oblastí pre statické farebné obrazy je ittiho model navrhnutý v roku 1998. Model zakladá na rozložení obrazu na 3 základné charakteristiky obrazu a to farbu, intenzitu, orientáciu.

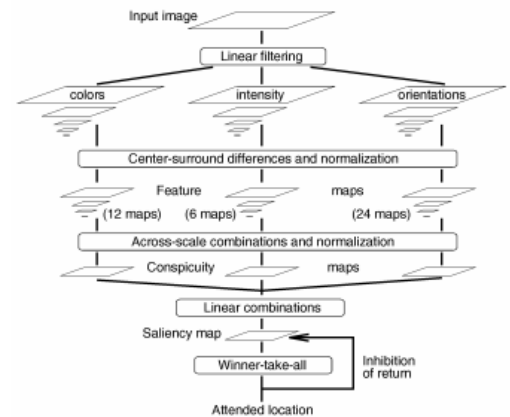
Chrakteristika farby obsahuje 12 máp (šedotónové obrazy), pričom model používa farebný model RGB.

Nazačiatku sa vypočíta intenzita podľa vzťahu $I = (R + G + B)/3$. Pomocou mapy I sa následne normalizujú všetky farebné kanáli modelu RGB.

Model extrahuje 4 farebné kanáli červený (r), zelený (g), modrý (b), žltý (y) a pomocou Gausových pyramíd vytvorí 3 rôzne mapy každej farebnej zložky separátne.

Červená zložka sa počíta difenčným spôsobom ako $R = r - (g + b)/2$, zelená ako $G = g - (r + b)/2$, modrá ako $B = b - (r + g)/2$ a žltá ako $Y = (r + g)/2 - |r - g|/2 - b$. Chrakteristika intenzity

obsahuje 6 máp. Získaná je pomocou orientovaných gáborových filtrov s orientáciou 0° , 45° , 90° , 135° . Dokopy 42 máp charakteristík je následne linárne skombinovaných do jednej saliency mapy[7].



Obr. 2.1: Itti model general workflow.

2.2.4 Spektrálne rezidua

Metóda využíva princíp, že potláča štatisticky často opakujúce sa časti obrazu a do popredia stavia časti obrazu ktoré sa štatisticky odlišujú od ostatných. Na detekciu používa rýchlu fourierovu transformáciu. Pomocou nej rozdelí obrázok na amplitúdovú časť a fázovú časť. Amplitúdová zložka sa následne vyhladí čím sa do popredia dostanú iba informácie ktoré sa vymykajú z priemeru. Odčítaním od pôvodnej amplitúdoj zložky dostaneme iba časti obrazu ktoré sú významné [6] TODO obrazok asi porovnanie s itty

2.2.5 Sun Model

2.2.6 Rare Model

...Context-Aware saliency, Weighted Maximum Phase Alignment Model, Torralba saliency, Murray model,

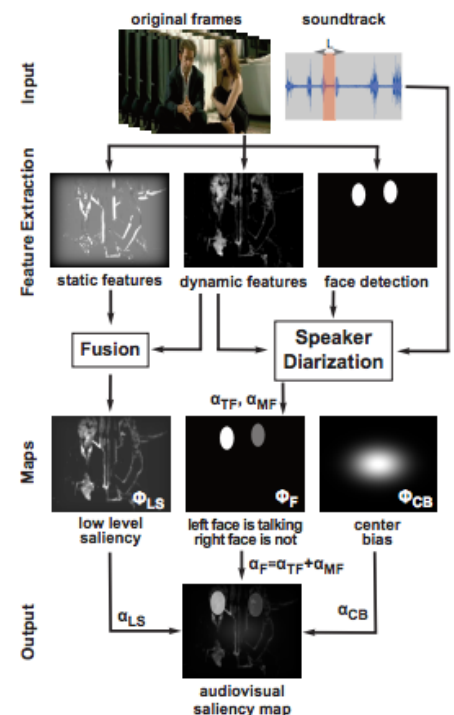
2.3 Metody pre videá

Video obsahuje rozsiahlejšie možnosti ako iba obrazová informácia, pribúdajú ďalšie rozmery ako je pohyb objektov na obraze alebo vplyv zvuku na ľudské vnímanie. Avšak oproti obrazu obsahuje je potrebné spracovávať vedšie množstvo dát. Navyše vo väčšine algoritmov využívajúcich saliency modely je potrebné aby model dával výsledky v reálnom čase. Používané hlavne v oblasti zabezpečovacej techniky.

2.3.1 Zohľadnenie audio informácie

Saliency modely využívajú rôznorodé druhy príznakov a to od geneticky zakorenených ako sú prechody farieb, alebo intenzít, až po sémantické príznaky ako je detekcia tváre [8]. Majoritná väčšina saliency modelov využíva iba obrazovú zložku ale zvuková stopa býva ponechaná stranou ale úplne zanedbaná. Použitie zvuku je známim trikom filmovej scény už desiatročia, kde režiséri posilujú kontrolu nad diváckou pozornosťou pomocou práve pomocou zvukového doprovodu. Prvé štúdie sa zaoberali detekciou reči a tváre, kde je spojitost' jednoznačná [3]. Neskôršie štúdie dokazujú korelácie aj na všeobecnejšej úrovni a pokusy o extrakciu samotnej charakteristiky zo zvukovej stopy [4]. Tieto pokusy viedli aj zostaveniu modelov zohľadňujúcich zvukovú stopu ako samostatnú charakteristiku spolu s kombináciou s nízko-urovnňovým príznakmi obrazu [2].

Model extrahuje video na sekvenciu obrazov (framy) a audio stopu v tvare grafu vlnovej dĺžky. Potom extrahuje 3 typy rôznych charakteristík. Nízko-úrovňové príznaky založené na biologicky inšpirovaných saliency modeloch rozdelených na dynamickú časť a statickú časť. Statická časť sa zameriava na najjasnejšie a najkontrastnejšie časti obrazu. Dynamická časť sa zameriava na relatívny pohyb objektov vzhľadom na pozadie (eliminácia pohybu kamery). Tieto 2 časti sa nakoniec spoja. Ďalšou charakteristikou použitou v tomto modeli je detekcia tváre. Každý objekt klasifikovaný ako tvár je v saliency mape nahradený oválnym objektom, intenzita daných objektov je daná pomocou metódy Speaker Diarization, ktorá detekuje podľa zvukovej stopy objekt ktorý generuje zvuk. Metóda predpokladá striedavú konverzáciu n objektov oddelenú pauzou. Následne spojí vyššie spomínané charakteristiky do jeden výslednej mapy. Ako posledný krok preloží cez celú



Obr. 2.2: Audiovisual model workflow.[2]

mapu baseline center model popísaný v časti 2.2.1.

2.3.2 Detekcia pohybu

V tejto časti sa zmeriame na segmentáciu objektov ktoré sa na scéne pohybujú. Taktéto obrazy sú v ľudskom vizuálnom systéme vysoko hodnotené. Dôvody, prečo takto ľudský vizuálny systém pridáva prioritu práve takýmto oblastiam môžeme nájsť v antropológii (najšť zdroj!). Vysvetlenie je jednoduché a to snaha zabezpečiť bezpečné prostredie okolo seba a všetko pohybujúce sa narušuje pocit bezpečnosti. V nasledujúcom texte rozoberieme 2 najpožívanejšie algoritmy používané na detekciu oblastí pohybu v obraze. Prvým z nich bude LUCAS KANADE[anotacia!], a druhým Horn Schunck[anotacia!]. TODO:

2.4 Metódy Využívajúce neurónové siete

2.5 Metriky úspešnosti

Zobrat z MIT saliency

2.6 Refenčné datasety

RSD, SAVAM, AUDITORY DATASET

2.7 Porovnanie štandardných Metód

3. Charakteristika - obrázky, zoznamy...

Lorem ipsum dolor sit amet, consectetur adipiscing elit a ďalší text...

V nasledujúcich častiach...

3.1 Podkapitola o zdrojoch

Text s uvedeným zdrojom [**vis**]: Zdroje sa ako bibtex(ten sa da vacsinou rovno stiahnut) pridavaju do textoveho suboru *bibliography.bib* (ktory treba nasledne tiez prekompilovat) a potom sa mozu citovat. Zdroj moze byt napr typu article [1], online link a ine. Pri online linku [**anemone**] treba uviesť datum citovania, ten sa nastavuje v subore *fmph.bbx* a pri online zdrojoch sa potom uz automaticky vypise.

3.1.1 Este mensia podsekcia - zoznamy

Zoznam poloziek:

- **prve** – popis
- **dalsie** – Aliquam erat volutpat. Proin vel mi aliquet, dignissim metus quis, elementum lectus. Quisque ac est enim. Interdum et malesuada fames ac ante ipsum primis in faucibus.
- **este nieco dalsie** - dalsi popis

Číslovaný zoznam:

1. **Analýza dát**
2. **Niečo ďalšie**
3. **Ešte niečo**
4. **Aby toho bolo viac**

3.1.2 Este jedna - obrázky vedľa seba



Obr. 3.1: Dva obrázky so spoločným popisom
obr



zi

Obr. 3.2: Popis obrázku s menším fontom popisu
by sa to bilo so zvýšením textom tak nech to odli

uz

vložený väčší vertikálny priestor medzi obrázkami



Obr. 3.3: Ďalší obrázok s veľkosťou fontu popisu ako ostatný text

3.2 Ina podkapitola

a ďalší múdry text

3.2.1 A v nej subsekcia

4. Návrh - ako vložiť ukážku kódu

farby a ďalšie kľúčové slova na zvýraznenie sa dajú určiť v definícii `lstset` ktorá je v `main.tex` nasleduje niečo ako JQuery....

```
(new MojObj());          //vytvorenie
MojObj = function() {
  var preml = [];
  var bool = 0;
  var construct = function() {
    var preml = $('#id').find('div');
    preml.each(function() {
      preml.push(new obj(this));
    });
    $(document).click(nejakaFunkcia);
  }
  var funkcia = function() {
    for (var i in preml) {
      if ( bool ) { //tak niečo sprav
      }
    }
    bool = !bool;
  }
  $(document).ready(construct);
};
```

Momentálne je `lstset` nastavený na jazyk PHP, tak automaticky zvýrazní syntax aj bez pridávania kľúčových slov

ukážka php kódu

```
<?php
function randomcode() {    //komentarik
  $var = "abcdefghijklmnopqrstuvwxyz0123456789";
  $rand = (double)microtime()*1000000;
  $i = 0;
  while ($i <= 7) {
    $num = rand() % 33;
    $tmp = substr($var, $num, 1);
    if (isset($code)) {
      $code = $code . $tmp;
    }
  }
}
```

```
    $i++;  
  }  
  return $code;  
}  
?>
```

5. Záver

Cieľom diplomovej práce bolo...

V práci som.....

Ďalší možný rozvoj....

Zoznam použitej literatúry

- [1] AIGNER, Wolfgang - MIKSCH, Silvia - MÜLLER, Wolfgang - SCHUMANN, Heidrun - TOMINSKI, Christian, “Visualizing time-oriented data-a systematic view”, *Comput. Graph.*, vol. 31, no. 3, s. 401–409, Jun. 2007, ISSN: 0097-8493.
- [2] COUTROT, A. - GUYADER, N. “An audiovisual attention model for natural conversation scenes”, in *Image Processing (ICIP), 2014 IEEE International Conference on* : Oct. 2014. S. 1100–1104. DOI: 10.1109/ICIP.2014.7025219.
- [3] COUTROT, Antoine - GUYADER, Nathalie, “How saliency, faces, and sound influence gaze in dynamic social scenesshort title??”, *Journal of Vision*, vol. 14, no. 8, p. 5, 2014. DOI: 10.1167/14.8.5. eprint: /data/Journals/JOV/933549/i1534-7362-14-8-5.pdf. Dostupné na internete: +%20http://dx.doi.org/10.1167/14.8.5.
- [4] COUTROT, Antoine - GUYADER, Nathalie - IONESCU, Gelu - CAPLIER, Alice, “Video viewing: do auditory salient events capture visual attention?”, *annals of telecommunications-Annales des télécommunications*, vol. 69, no. 1-2, s. 89–97, 2014.
- [5] DUNCAN, K. - SARKAR, S. “Saliency in images and video: a brief survey”, *Computer Vision, IET*, vol. 6, no. 6, s. 514–523, Nov. 2012, ISSN: 1751-9632. DOI: 10.1049/iet-cvi.2012.0032.
- [6] HOU, Xiaodi - ZHANG, Liqing, “Saliency detection: a spectral residual approach”, in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* : Jun. 2007. S. 1–8. DOI: 10.1109/CVPR.2007.383267.
- [7] ITTI, L. - KOCH, C. - NIEBUR, E. “A model of saliency-based visual attention for rapid scene analysis”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, s. 1254–1259, Nov. 1998, ISSN: 0162-8828. DOI: 10.1109/34.730558.
- [8] SHARMA, P. - CHEIKH, F.A. - HARDEBERG, J.Y. “Face saliency in various human visual saliency models”, in *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on* : Sep. 2009. S. 327–332. DOI: 10.1109/ISPA.2009.5297732.

- [9] ŠIKUDOVÁ, E. - ČERNEKOVÁ, Z. - BENEŠOVÁ, W. - HALADOVÁ, Z. - KUČEROVÁ, J. 2014. *Počítačové videnie. Detekcia a rozpoznávanie objektov*, first : Wikina, Livornská 445, 109 00 Praha 10, 2014. .

Prílohy

CD obsahujúce:

- Elektronickú verziu
- Zdrojáky
- atď