Katedra aplikovanej informatiky
Fakulta Matematiky, Fyziky a Informatiky
Univerzita Komenského v Bratislave
9.2.1 Informatika

Využitie sémantickej informácie pri určovaní významných oblastí v obraze

Rigorózna práca

Mgr. Júlia Kučerová

Bratislava 2012

Čestne prehlasujem, že som predloženú prácu vypracovala samostatne s použitím citovaných zdrojov.

. . . . . .

Poďakovanie

Chcem poďakovať RNDr. Elene Šikudovej PhD. za neoceniteľnú pomoc a hodnotné pripomienky pri písaní tejto práce a prof. Ing. Jaroslavovi Polecovi, PhD. za rady pri testovaní výsledkov. Ďakujem svojim rodičom, bratovi a kamarátom psychyckú podporu.

Chcem poďakovať Nadácií Slovenského plynárenského priemyslu Hlavička č. 61/2012 za štipendium, ktoré mi bolo udelené v školskom roku 2011/2012 a významne podporilo výskum súvisiaci s predloženou rigoróznou prácou.

Abstrakt

Predložená práca sa zameriava na problematiku detekcie významných oblastí v obraze s využitím sémantickej informácie, konkrétne detekcie pokožky. Tento prístup umožňuje rozpoznanie relevantných častí obrazu pri záznamoch obsahujúcich znakovú reč, pretože popri nájdení významných oblastí umožňuje i detekciu rúk a tváre. Takto detegované významné časti obrazu sú v práci využité pri návrhu nových metrík na ohodnocovanie kvality videa. Tieto metriky sú porovnané s bežne využívanými metrikami, pričom na základe testovania dosahujú veľmi dobré výsledky.

Kľúčové slová: vizuálna pozornosť, významné oblasti, detekcia pokožky, ohodnocovanie kvality videa, znaková reč

Abstract

This thesis focuses on the topic of salient regions detection utilizing an additional semantic information, namely the skin detection. This approach allows the recognition of the relevant parts of the sign language video recordings. Salient regions along with hands and face are detected. The obtained output is used to design several new video quality metrics. In comparison with existing widely used metrics our metrics achieve superior results.

Keywords: visual attention, salient regions, skin detection, video quality metrics, sign language

Obsah

Zoznam obrázkov

Zoznam tabuliek

Úvod

Problematika vizuálneho vnímania a vizuálnej pozornosti je dlhdobo predmetom rozsiahleho výskumu. Keďže ľudské vnímanie je obmedzené a ľudia nedokážu vnímať všetky vstupné informácie naraz, je z hľadiska aplikácií užitočné zistiť, na ktoré oblasti v scéne sa upriamuje ich pozornosť a kde v scéne sa tieto oblasti nachádzajú. Takéto oblasti sú označované ako významné.

Ľudské vizuálne vnímanie závisí od množstva faktorov – vizuálnych vlastností pozorovanej scény, medzi ktoré je zaraďovaná napr. farba, textúra, intenzita, orientácia. Z toho dôvodu je potrebné zohľadňovať tieto faktory pri skúmaní scény a určovaní významných oblasti v nej. V súčasnosti existuje značné množstvo štúdií tejto problematiky, ktoré sú založené na rôznych predpokladoch, ich kombináciách, vylepšeniach.

Na určenie významných oblastí v obraze bolo navrhnutých množstvo modelov, ktorých úspešnosť detekcie je rôzna, avšak všetky modely zaoberajúce sa danou problematikou sa snažia čo najlepšie aproximovať ľudské vnímanie. Každý človek vníma pozorovanú scénu sčasti subjektívne, čo ovplyvňuje aj percepciu týchto významných oblastí, a preto v súčasnosti nie je možná ich úplná detekcia.

Významné oblasti nesú pre pozorovateľa potenciálne dôležitú informáciu a poznatky o ich výskyte v obraze majú veľmi široké využitie v praxi, napr. pri kompresii a kódovaní obrazu, v rozpoznávaní, detekcii, kategorizácii objektov, pri segmentácii obrazu, odstraňovaní artefaktov, renderovaní, spracovaní multimédií, orezávaní obrazu, atď.

V minulosti sa väčšina modelov na detekciu významných oblastí zameriavala na získavanie a kombináciu nízkoúrovňových príznakov, ako sú farba, intenzita, orientácia. V súčasnosti sa však tieto modely upravujú a navrhujú sa nové, v ktorých hrá dôležitú úlohu sémantická informácia, ktorej využitie má veľký potenciál.

Predložená práca nachádza praktické uplatnenie v komunikácii sluchovo hendikepovaných ľudí prostredníctvom videa. Pretože primárnym dorozumievacím prostriedkom týchto osôb je znaková reč, je nesmierne dôležité, aby video so znakovou rečou bolo dostatočne zrozumiteľné. V práci je vybraný a modifikovaný základný model na detekciu významných oblastí v obraze. Ako sémantická informácia bola zvolená detekcia pokožky, keďže pri komunikácii znakovou rečou sú najdôležitejšie časti pozorovaného obrazu ruky a tvár.

Kapitola 1

Prehľad problematiky

Do ľudského oka vstupuje každú sekundu veľké množstvo dát ($10^8 - 10^9$ bitov). Spracovanie takéhoto množstva dát v reálnom čase je veľmi náročná úloha a preto je potrebné vedieť čo najefektívnejšie redukovať ich množstvo.

V posledných desaťročiach sa vedci snažili odpovedať na množstvo otázok súvisiacich s vnímaním a vizuálnou pozornosťou. Psychológovia študovali správanie spojené s vizuálnou pozornosťou, ako napríklad nevšímanie si zmeny, nepozornosť, žmurkanie. Neurofyziológovia ukázali, ako sa neuróny prispôsobujú kvôli lepšej odozve na objekty záujmu. Ďalší výskumníci sa pokúsili zostrojiť model neurónovej siete na simuláciu a vysvetlenie správania pri pozorovaní scény. Vedci zaoberajúci sa robotikou a počítačovým videním sa pokúsili o zachytenie komplexnosti celého problému a o vytvorenie systému, ktorý by pracoval v reálnom čase. Všetky tieto prístupy sa pokúsili pomôcť ľuďom pochopiť ľudské vizuálne vnímanie a tak umožniť modelovanie systémov, ktoré by toto vnímanie dokázali dostatočne presne napodobniť.

Modelovaniu systémov na simuláciu vizuálnej pozornosti sa vedci aktívne venujú už vyše 25 rokov. Bolo vytvorené veľké množstvo modelov, ktoré sa úspešne používajú v počítačovom videní, robotike, kognitívnych systémoch. Tieto modely sa zameriavajú na zachytenie významných oblastí v obraze a ich ďalšie použitie. Aj napriek desaťročiam výskumu v problematike vizuálnej pozornosti a vizuálneho vnímania je však ešte stále veľké množstvo otvorených problémov, ktoré sa vedci snažia pochopiť a riešiť. Jedným z nich je tvorba modelov na detekciu významných oblastí pomocou sémantiky. Riešenie tejto

komplexnej a zložitej úlohy si však vyžaduje ešte mnoho práce a ponúka veľa príležitostí na ďalší výskum [2], [11].

1.1 Vnímanie

Vizuálne vnímanie človeka je veľmi komplexný proces v ktorom ľudia pri rozoznávaní pozorovanej scény ešte stále dosahujú oveľa lepšie výsledky ako počítače. Proces vnímania je závislý na vstupe, ale aj na vedomostiach, ktoré má pozorovateľ pri sledovaní danej scény k dizpozícii.

Pri skúmaní vnímania je veľmi dôležité zodpovedať otázku: Ako ľudia vnímajú objekty?
Ľudia pri vnímaní scény využívajú vnemovú inteligenciu – vedomosť, ktorú získavajú skúsenosťami pri pozorovaní. Pri procese vnímania je vnemová inteligencia veľmi dôležitá a veľkou mierou prispieva ku konečnému vizuálnemu vnemu.

Vnímanie objektov je ústredná funkcia vizuálneho systému pri každodenných skúsenostiach človeka. V prípade úlohy, kde má pozorovateľ popísať, čo vníma práve teraz závisí od viacerých faktorov. Patria medzi ne napríklad prostredie, kde sa nachádza, psychologické rozpoloženie, atď. , avšak veľká časť toho, čo popíše bude súvisieť s tým, čo vidí [11].

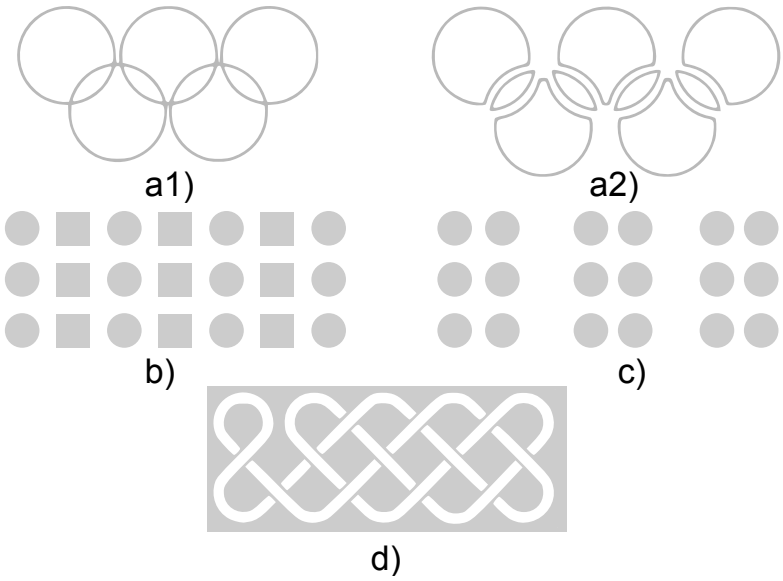
1.1.1 Systém vnímania

Ľudské vizuálne vnímanie je obmedzené, a preto pri sledovaní scény sú najprv vnímané jej časti, ktoré sú až následne spájané do celku.

Vďaka systému vnímania sú ľudia schopní zaraďovať prvky oblasti do objektov, a tým vytvoriť ucelenú scénu. Tento fenomén ako prvý skúmal psychológ Gestalt, ktorý základy Gestaltovej psychológie vytvoril v roku 1920.

Gestaltove zákony

Základom Gestaltovej psychológie sú Gestaltove zákony, ktoré sú uvádzané v zjednodušenej forme.



Obr. 1.1: Ukážka niektorých Gestaltových zákonov: a1) zákon jednoduchosti - ako vnímame objekty , a2) zákon jednoduchosti - ako nevnímame objekty b) zákon podobnosti c) zákon blízkosti d) zákon spojitosti

Zákon jednoduchosti pomáha vidieť každú časť vnemu čo najjednoduchšie, a preto napríklad pri pozorovaní objektu na obrázku 1.1 a1) človek nevidí 9 objektov (Obr. 1.1 a2)), ale 5 kruhov.

Na základe **podobnosti** sú veci zoskupené do skupín napríklad podľa tvaru, veľkosti, odtieňa, orientácie (Obr. 1.1 b)).

Veci, ktoré sú bližšie k sebe, vytvárajú dojem prepojenosti, preto sú podľa **zákona blízkosti** zaraďované do rovnakej skupiny (Obr. 1.1 c)).

Pri riešení hlavolamu (napríklad bludiska) sa ľudia riadia **zákonom dobrej spojitosti**, kedy vyberajú tú najľahšiu a najplynulejšiu cestu (Obr. 1.1 d)).

V prípade, že sa veci pohybujú rovnakým smerom, taktiež sú zaraďované do rovnakej skupiny; tento Gestaltov zákon sa nazýva **zákon spoločného určenia**.

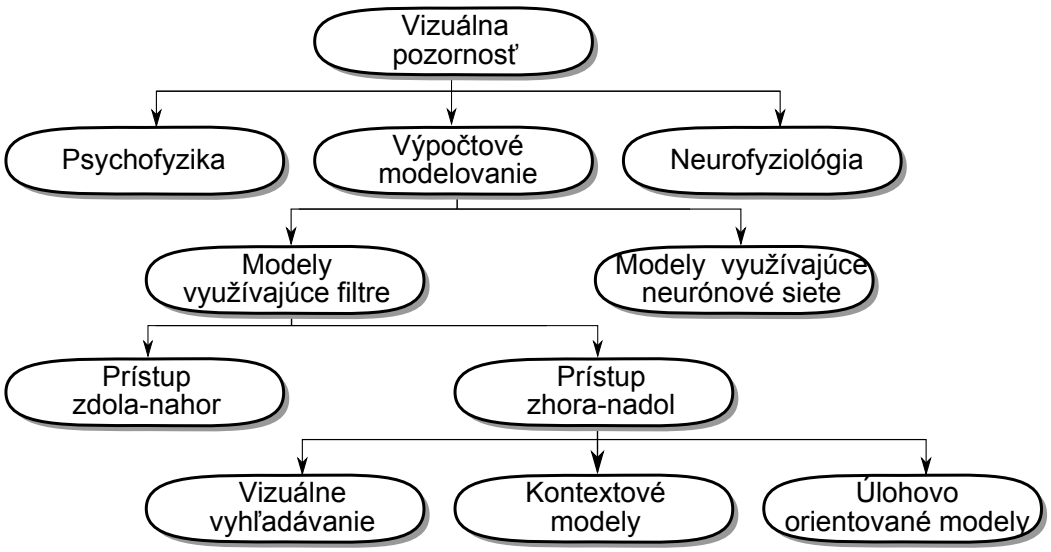
Ak sa veci zdajú byť podobné, podľa **zákona familiarity** pravdepodobne patria do jednej skupiny (napr. tváre v prírode tvorené z kameňov, listov).

Gestaltove zákony nedávajú vždy presné a dobré výsledky (napríklad, keď človek vidí tieň a podľa obrysov vníma zviera, hoci v skutočnosti to je iba konár) a preto by bolo

oveľa lepšie nazvať ich Gestaltove heuristiky, keďže heuristika udáva najlepšie riešenie na daný problém [11]. Pojem Gestaltove zákony je však zaužívaný, preto je používaný aj v tejto práci.

1.2 Vizuálna pozornosť

Vizuálna pozornosť veľmi vplýva na vnímanie, na pamäť (ľudia sú schopní si niečo zapamätať, len ak na to sústredia svoju pozornosť), na jazyk (čítanie zahŕňa pozornosť na text slovo po slove) aj na riešenie problémov (úspech v riešení problémov závisí od toho, ktorý problém zaujme pozornosť človeka) [11]. Na obrázku 1.2 je zobrazený graf, ktorý znázorňuje rozdelenie štúdií vizuálnej pozornosti. Z digramu je zrejmé, že vizuálna pozornosť je oblasť, ktorá je veľmi zaujímavá a zaoberajú sa ňou vedci z rôznych oblastí výskumu.



Obr. 1.2: Graf rozdelenia štúdií vizuálnej pozornosti [2]

Pri vnímaní scény sa rozlišujú dva hlavné prístupy: zdola-nahor (bottom-up) a zhora-nadol (top-down). Pomocou týchto dvoch prístupov je vizuálna pozornosť rozdelená na *endogénnu*, ktorá využíva prístup zhora-nadol a *exogénnu* založenú na prístupe zdola-nahor.

Exogénna pozornosť je zameraná na vonkajšie podnety, upozorňuje na výrazné časti automaticky a má rýchly časový priebeh. Existuje široká škála exogénnych vizuálnych atribútov na zachytenie pozornosti, ako napríklad priestorové vnemy, náhle vizuálne zmeny jasu atď.

V endogénnej časti je pozornosť určovaná subjektom. Táto pozornosť je známa ako cieľovo riadená pozornosť. Endogénna pozornosť je riadená pozorovateľom, závislá na ňom a je pomalá z dlhodobého časového hľadiska.

Existuje viacero modelov vizuálneho vnímania, ktoré využívajú tieto modely oddelene. Avšak vo všeobecnosti takmer každý model vizuálneho vyhľadávania je založený na vzájomnej interakciii medzi týmito prístupmi. [45] Oba hlavné prístupy sú navzájom nezávislé.

Prístup zdola-nahor

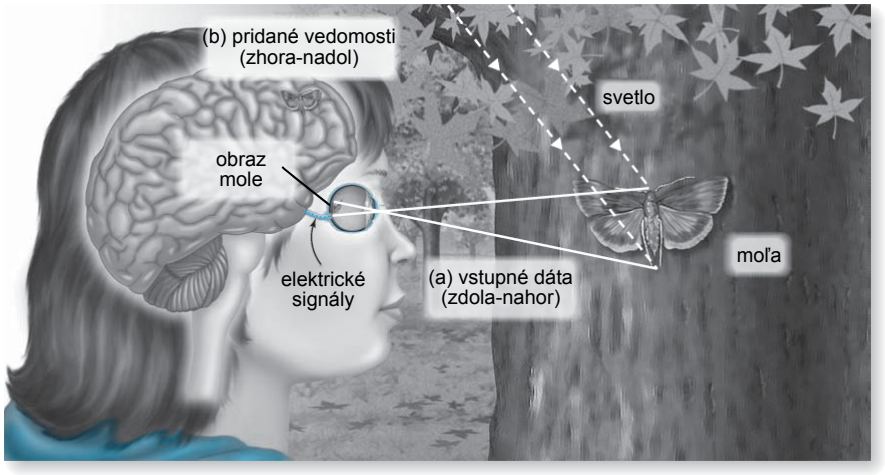
Postupnosť udalostí, ktoré začali stimuláciou receptorov je zaraďovaná medzi exogénne. Sú riadené stimulmi, kde sa automaticky, bez vonkajších podnetov, pozornosť zameriava na výraznejšie časti obrazu. Túto pozornosť môže ovplyvniť veľké množstvo podnetov, ako napríklad zmeny vo farbe, zmena polohy, intenzity. Všetky tieto zmeny ľudské oko zachytáva a podvedome upriamuje pozornosť práve na ne. Množstvo psychologických výskumov ukázalo, že vizuálny systém je vysoko citlivý na vlastnosti v obraze, ako sú napríklad hrany, náhle zmeny vo farbe, intenzite, náhle pohyby.

Pocesy zdola-nahor sú nezávislé na úlohe, ktorú pozorovateľ rieši. Snažia sa predvídať, ktoré časti pozorovanej scény môžu pritiahnuť viac pozornosti, a podľa toho vytvárajú mapy výrazných oblastí v obraze. Tieto procesy môžu byť využívané napríklad pri automatickej detekcii výrazných oblastí v scénach, v strojovom videní, pri inteligentnej kompresii obrazu, atď.

Pri analyzovaní scény pomocou prístupov zdola-nahor sú medzi výrazné objekty zaradené napríklad červené jablko na zelenom strome, horiaca sviečka v tmavej miestnosti, alebo pery a oči na tvári považované za jej najvýraznejšie časti.

Pri pozorovaní scény, v ktorej sa nachádza veľké množstvo rovnakých objektov, ktoré by samostatne boli určené ako výrazné, sa tieto objekty stávajú nevýraznými. Keď pozo-

rovateľ porovnáva dve fotografie jabloní s červenými jablkami, kde na jednej je iba jedno a na druhej je veľké množstvo jabĺk, v prvom prípade si jablko všimne, ale v druhom prípade sú jablká odfiltrované a pozorovateľ si všíma iné výrazné oblasti a objekty. [7]



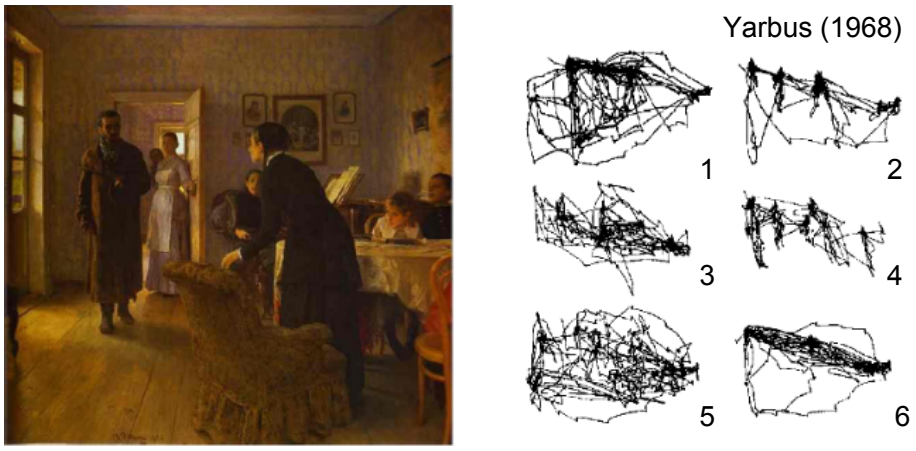
Obr. 1.3: Pozorovanie mole je ovplyvnené kombináciou (a)vstupných dát a (b) predošlých vedomostí [11]

Prístup zhora-nadol

Vizuálne vnímanie je oveľa komplikovanejšie, než len vnímanie energie na receptoroch, čo je prezentované pri procesoch založených na prístupe zdola-nahor. Preto je pri procesoch založených na prístupe zhora-nadol pozorovanie scény čiastočne ovládané pozorovateľom. Toto ovládanie sa deje pridaním dodatočnej informácie, ako napríklad tým, že človek je pri pozorovaní nových scén ovplyvnený vecami, ktoré už videl v minulosti. Toto pozorovanie zahŕňa skúsenosti pozorovateľa, avšak môže byť veľmi pomalé.

Ako príklad prístupov zhora-nadol je možné použiť pozorovanie kôry stromu. Pri procesoch zdola-nahor sa určí farba kôry, rozlišujú sa hrany a textúra, ale až pri pridaní ďalšej informácie je možné určiť, že na kôre stromu sa nachádza moľa. (Obr 1.3).

Pri využívaní prístupov zhora-nadol môže byť ako prídavná informácia použitá úloha, ktorú ma pozorovateľ vyriešiť. Avšak, ak má pozorovateľ napríklad nájsť určitý objekt v scéne, môže prehliadnuť iné výrazné oblasti v nej. Preto môže nastať situácia, kde niektoré



Obr. 1.4: Inštrukcie dané pozorovateľovi obrazu: 1) voľne pozorovať obraz 2) odhadnúť vek osôb 3) zistiť, čo robili osoby pred príchodom návštevníka 4) zapamätať si oblečenie osôb 5) zapamätať si pozíciu osôb a objektov v miestnosti 6) odhadnúť, ako dlho bol návštevník neprítomný [6]

objekty výrazné pri detekcii pomocou prístupov zdola-nahor nemusia byť považované za významné pri prístupe zhora-nadol [7].

Skúmaním vizuálnej pozornosti a procesov zhora-nadol sa zaoberal ruský psychológ Yarbus. Jeho výskum spočíval v zaznamenávaní fixácií a skokov oka pozorovateľov, ktorí mali úlohu skúmať danú scénu a odpovedať na položené otázky. Tieto otázky výrazne ovplyvňovali pohyby očí po danej scéne. Na Obr. 1.4 je znázornený pozorovaný obraz "An Unexpected Visitor" od Repina a pohyby očí, ktoré sa výrazne menili v závislosti od daných otázok.

1.3 Sémantická informácia

Význam vnímaných objektov a informácií sa nachádza v ľudskej mysli, nie v samotnom slove. Keď je povedané, že nejaké slovo má určitý význam, v podstate to znamená, že to slovo v človeku daný význam evokuje. Pokiaľ sa tento význam nedostane do ľudskej mysle, slovo je len súbor znakov. Rovnako je vnímaný aj obraz, vizuálnu informáciu. Pokým vizuálny vnem nie je spracovaný v ľudskej mysli, nedokáže človek plnohodnotne vnímať pozorovanú scénu.

Podľa výkladového slovníka [43] je sémantika veda o význame jednotlivých slov, znakov, symbolov a ich vzťahu ku skutočnosti, ktorú označujú. Presnejšie povedané, je to štúdia významu, ktorá sa zameriava na vzťahy medzi slovami a tým, ako sú využívané a čo znamenajú. Napríklad, keď človek povie, že ide zjesť malinu, každý si predstaví červenú, šťavnatú malinu. Avšak, vyjadrenie, že nejaká úloha bola malina, značí, že bola jednoduchá [28].

Vnímanie bez pridanej informácie je možné prirovnať k procesom zdola-nahor, kde pozorovateľ nevie komplexne určiť, čo sa v obraze nachádza. Jediné, čo vníma sú výrazné čiary, farby, textúry. Avšak pri pridaní ďalšej informácie, ako napríklad spomienky z minulosti vie tieto objekty pospájať tak, aby v ňom vyvolávali konkrétny vnem. Preto pri pozorovaní scény a taktiež pri zisťovaní významných oblastí v scénach je táto prídavná (sémantická) informácia veľmi dôležitá.

Kapitola 2

Prehľad podobných prác

Vizuálne vnímanie a mechanizmay spojené s videním sú predmetom dlhodobého a rozsiahleho výskumu. Počas skúmania tohto fenoménu sa kladú dve základné otázky:
Čo v obraze púta našu vizuálnu pozornosť?
Kde v obraze sa tieto objekty a oblasti nachádzajú?

Začiatky pozorovania vizuálneho vnímania sa zalkadali na jednoduchých pozorovaniach, neraz aj sebapozorovaniach. Táto situácia bola dôsledkom obmedzení v oblasti techniky, postupom času sa však možnosti skúmania vizuálnej pozornosti zlepšili a vo výskume sa začali používať poznatky z iných oblastí napr. psychológie, kognitívnej vedy.

Hermann Von Helmholtz [13] už začiatkom 20. storočia považoval vizuálnu pozornosť za základný mechanizmus vizuálneho vnímania. Pozoroval prirodzenú tendenciu zameriavať pozornosť na nové oblasti pozorovaného obrazu. Sústredil sa najmä na pohyby očí a jeho hlavnou výskumnou otázkou bolo, kam sa vizuálna pozornosť zameriava.

Narozdiel od Von Helmholtza, William James [17] predpokladal, že pozornosť je viac vnútorne zameraný mechanizmus, ktorý zahŕňa predstavivosť, očakávania, alebo myslenie vo všeobecnosti. James definoval pozornosť hlavne z hľadiska významu a očakávaní spojených so zameraním pozornosti. Jeho výskum bol zameraný na otázku "čo"v obraze priťahuje našu pozornosť. Uprednostňoval aktívne a spontánne aspekty pozornosti, avšak tiež rozoznával ich pasívne, podvedomé, nespontánne charakteristiky.

Tieto dva pohľady na danú problematiku sa nevylučujú, naopak, spoločne komplexne popisujú koncept vizuálnej pozornosti.

V polovici dvadsiateho storočia skúmal Donald Broadbent [3] fakt, že pozornosť sa v určitom zmysle dá definovať ako selektívny filter, zodpovedajúci za reguláciu zmyslovej informácie. Zaoberal sa akustickými pokusmi, ktoré smerovali k dokázaniu selektívneho charakteru sluchovej pozornosti, čo korešpondovalo s otázkou "kde" a Von Helmoltzovým návrhom.

Myšlienku selektívneho filtra vylúčili J. Anthony Deutsch a Diana Deutsch [8], ktorí uprednostňovali myšlienku, že všetky zmyslové vnemy sú perceptuálne analyzované na najvyššej úrovni. Navrhli centrálnu štruktúru s vopred nastavenými váhami dôležitosti, ktoré určujú výber. Váhy dôležitosti zodpovedajú Jamesovým predpokladom a otázke "čo".

V šesťdesiatych rokoch dvadsiateho storočia Anne Treisman spojila obe tieto myšlienky do jednej, kde prepojila modely Broadbenta a Deutscha určením dvoch komponentov pozornosti.

Veľmi významným medzníkom v skúmaní zameriavania vizuálnej pozornosti bol rok 1967, kedy ruský psychológ Yarbus zaznamenal fixácie a skoky oka pozorovateľov pri pozorovaní daných scén. Rôzne úlohy pri pozorovaní scény mali za následok podstatne odlišné skoky očí, kde sa pri každej úlohe tieto pohyby očí zameriavali na časti obrazu, ktoré pomohli zodpovedať kladenú otázku.

Posner [32], Noton a Stark [29] vylepšili teóriu vizuálnej pozornosti podobným spôsobom, ako navrhli Von Helmholtz a James. Anne Treisman následne spojila tieto pojmy s teóriou integrácie príznakov vizuálnej pozornosti [40] [41]. Podľa jej predpokladov poskytuje pozornosť tzv. "lepidlo", ktoré spája oddelené príznaky v určitej časti scény tak, aby ich spojenie bolo vnímané ako jeden objekt. Kosslyn [19] popísal pozornosť ako selektívny aspekt spracovania vnímania a navrhol "okno" zodpovedné za vyberanie vzoriek vo vizuálnom zásobníku [9].

V súčasnosti je väčšina modelov na detekciu významných oblastí inšpirovaných biologickými procesmi, sú založené na prístupe zdola-nahor (napr. [16], [14], [5], [31]). Detegujú nízkoúrovňové príznaky, ako sú farba, intenzita, orientácia, textúra, pohyb. Detegujú významné oblasti na rôznych úrovniach, používajú stratégiu víťaza (angl. winner take all, ozn. WTA), nemožnosť návratu (angl. inhibition of return, ozn. IOR).

Pre niektoré aplikácie sú používané dáta získané pomocou sledovania pohybu očí (eyetracking). Pozorovateľ sleduje scénu a systém zaznamenáva pohyby jeho oka, čím sa zisťuje, kam sa zameriava jeho pozornosť. Takýto spôsob zisťovania významných oblastí je však finančne a časovo náročný, a preto nie je vždy možné ho používať. Z týchto dôvodov je veľmi potrebné detegovať významné oblasti aj iným spôsobom. Ako alternatíva sú preto navrhované a používané systémy na detekciu pravdepodobnosti výskytu významných oblastí v obraze.

Pri mnohých aplikáciách z oblasti grafiky, dizajnu, HCI je veľmi dôležité porozumenie tomu, kam človek zameriava svoju vizuálnu pozornosť. Tieto informácie sa využívajú napríklad pri automatickom orezávaní obrazu, vytváraní náhľadov, pri vyhľadávaní konkrétneho obrazu v databáze. Môžu byť použité taktiež pri kompresii videa, nefotorealistickom renderingu, adaptívnom zobrazovaní na malých displejoch. Vo všeobecnosti je vzhľadom na ovedené aplikácie potrebné čo najlepšie, najpresnejšie a najrýchlejšie detegovať oblasti záujmu.

V nasledujúcej časti sú popísané vybrané modely na detekciu významných oblastí.
V prvej časti sú porovnané modely, ktoré sú založené na prístupe zdola-nahor. Ďalšiu časť tejto kapitoly tvorí popis dvoch modelov na detekciu významných oblastí. Prvý z popísaných modelov je Ittiho prístup, ktorý je založený na správaní a neurónovej architektúre vizuálneho systému primátov a radí sa medzi základné modely na detekciu významných oblastí v scéne. Ďalší model využíva dáta získané zo sledovania očí a taktiež porovnáva svoje výsledky s inými prístupmi.

2.1 Porovnanie najznámejších modelov na detekciu významných oblastí

Hlavnou úlohou modelov na detekciu je určovanie oblastí, ktoré priťahujú pohľad pozorovateľa. Modely využívajúce procesy zdola-nahor, sú ešte stále iba základným popisom ľudského vizuálneho systému. Najväčším obmedzením takýchto modelov je neprítomnosť ďalšej informácie,ktorá by v kombinácii s nízkoúrovňovými príznakmi pomohla určiť vý-

znamné oblasti v obraze [27].

V roku 1998 bol vytvorený prvý model založený na simulácii vnímania ľudským okom [16] využívajúci procesy zdola-nahor. Odvtedy vzrastal záujem o tému detekcie významných oblastí.

Modely je možné rozdeliť do troch skupín:

- Hierarchické modely
- Štatistické modely
- Bayesovské modely

Hierarchické modely

Tieto modely majú navzájom podobnú architektúru. Sú charakteristické postupným rozkladom, ktorý zahŕňa Gaussovskú, alebo Fourierovú vlnovú dekompozíciu. Väčšinou je na určovanie významnosti aplikovaný rozdiel gaussianov na vypočítaných podpásmach. Následne sú použité rôzne techniky na súhrn tejto informácie cez všetky úrovne, pomocou čoho sa vytvorí jedinečná mapa významných oblastí. Vybrané modely sú porovnané v Tabuľke 2.1.

Štatistické modely

Štatistické modely sú založené na pravdepodobnostnom prístupe, závislom na pozorovanom obraze. Významnosť oblastí je následne určená cez meranie odchýlky medzi príznakmi na súčasnej pozícii a príznakmi v okolí danej pozície. Na veľkosti okolia nezáleží, môže byť malé, ale taktiež môže byť porovnateľné s veľkosťou skúmaného obrazu.(Tabuľka 2.2)

Bayesovské modely

Bayesovská štruktúra ponúka viacero výhod, keďže povoľuje kombináciu prístupov zdola-nahor a prídavnej informácie. Príkladom takejto informácie je štatistika vizuálnych príznakov v scéne. Táto kombinácia je pravdepodobne jeden z dôležitých faktorov, ktorý má vplyv naše vnímanie. Prídavná informácia prichádzajúca z procesu nášho učenia sa počas vnímania môže pomôcť vizuálnemu systému pri porozumení pozorovaného prostredia.

Hierarchické modely	**Rozmer vizuálneho skúmania**	**Operácie**	**Pridaná informácia informácia**
Itti [16]	Intenzita, dva chromatické kanály, orientácia, flicker	Dvojčlenné Gaussovské a Gaborove pyramídy, filtre pre stred a okolie, normalizácia (peak–to–peak) delenie	žiadna
Le Meur et al. [23], [24]	jas, dva farebné kanály, pohyb	Orientovaná podpásmová dekompozícia vo Fourierovej oblasti, funkcie citlivé na kontrast, maskovanie filtre pre stred a okolie, normalizácia (long–term), delenie	žiadna
Bur et al. [5]	intenzita, dva farebné kanály, orientácia, pohyb	Dvojčlenné Gaussovské a Gaborove pyramídy, filtre pre stred a okolie, normalizácia (long–term), delenie	žiadna

Tabuľka 2.1: Hierarchické modely

Štatistické modely	**Rozmer vizuálneho skúmania**	**Operácie**	**Pridaná informácia**
Oliva et al. [31]	R, G, B	Významnosť oblastí je nepriamo úmerná pravdepodobnosti jej výskytu v obraze. Pravdepodobnosť rozdelenia je založená na štatistike pre pozorovaný obraz.	žiadna
Bruce et al. [4]	R, G, B	Významnosť oblastí je založená na výpočte Shanonovej informácii. Spojenie pravdepodobnosti pre príznaky odvodené z daného okolia.	žiadna
Gao et al. [10]	Intenzita, dva farebné kanály, orientácia a pohyb	Dvojčlenné Gaussovské a Gaborove pyramídy, filtre pre stred a okolie. Významnosť je určená pomocou Kullback-Leiblerovej divergencie, medzi lokálnou pozíciou a jej okolím.	žiadna

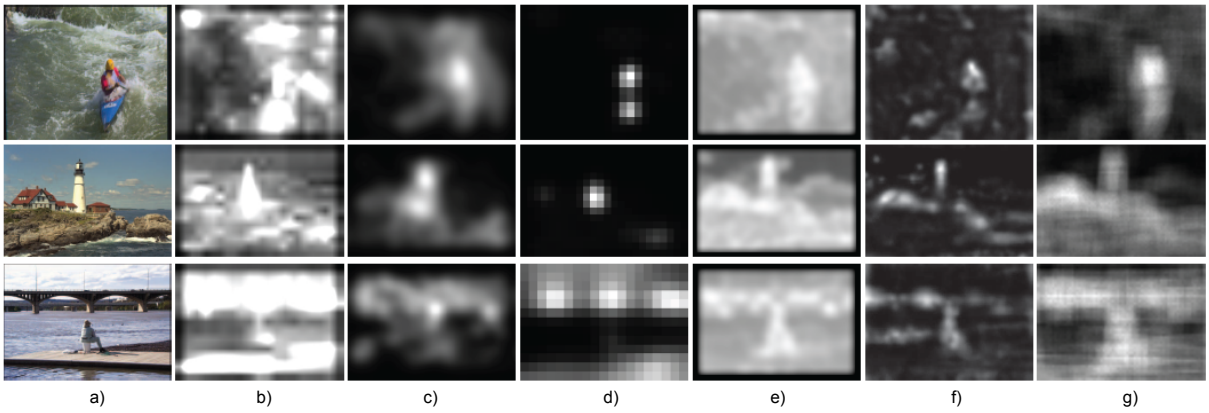
Tabuľka 2.2: Štatistické modely

Bayesovské modely	**Rozmer vizuálneho skúmania**	**Operácie**	**Pridaná informácia**
Zhang et al. [46]	jas, dva farebné kanály,	Významnosť je založená na výpočte Shanonovej informácie	odhad distribúcie pravdepodobnosti

Tabuľka 2.3: Bayesovské modely

Modely založené na Bayesovskom prístupe dávajú veľmi sľubné výsledky. Pri tvorbe takýchto modelov je dôležitá informácia prichádzajúca z nízkoúrovňových príznakov.

Všetky výpočtové modely popísané vyššie zachytávajú len veľmi základný popis ľudského vizuálneho vnímania. Na obrázku 2.1 sú znázornené výsledky detekcie významných oblastí porovnávanými metódami.



Obr. 2.1: Mapy významných oblastí detegované pomocou rôznych modelov: (a) pôvodný obrázok, a mapy významných oblastí získané prístupmi: (b) Itti [16], (c) Le Meur [23], [24] , (d) Bur [5], (e) Bruce [4], (f) Gao [10], (g)Zhang [46].

Pri navrhovaní nových modelov môže byť detekcia významných oblastí vylepšená z biologického hľadiska. Taktiež je pri navrhovaní nových a presnejších modelov veľmi dôležitý kognitívny proces a využitie sémantickej informácie.

Prvá otázka, ktorá sa naskytne pri uvažovaní nad využitím kognitívneho procesu je existencia miesta v ľudskom mozgu, kde sa lokalizuje mapa významných oblastí. Avšak

podľa doterajších výskumov také miesto neexistuje. Toto vysvetľuje fakt, že významnosť oblastí v pozorovanom obraze neurčujú iba nízkoúrovňové príznaky, ale taktiež naše vedomosti, spomienky, očakávania.

Vizuálne a kognitívne procesy sú veľmi úzko prepojené a v súčastnosti nie je možné určiť, ktoré z nich v danom čase pozorovania určujú rozloženie vizuálnej pozornosti. Pri určovaní presnosti modelov založených na procesoch zdola-nahor je preto využívané sledovanie očí. Avšak aj pri sledovaní očí nemôžme zabúdať na procesy, ktoré ovplyvňujú vnímanie zo sémantického hľadiska. Napríklad pri voľnom pozorovaní scény sú výrazné iné oblasti, ako pri pozorovaní scény s určitou úlohou. Výraznosť týchto oblastí je ovplyvnená aj správaním, myšlienkami a spomienkami pozorovateľa. Veľmi dobrý prístup je charakterizácia významnosti oblastí za použitia dát získaných sledovaním očí. Tento bol využitý v práci [18] a dáva veľmi dobré výsledky.

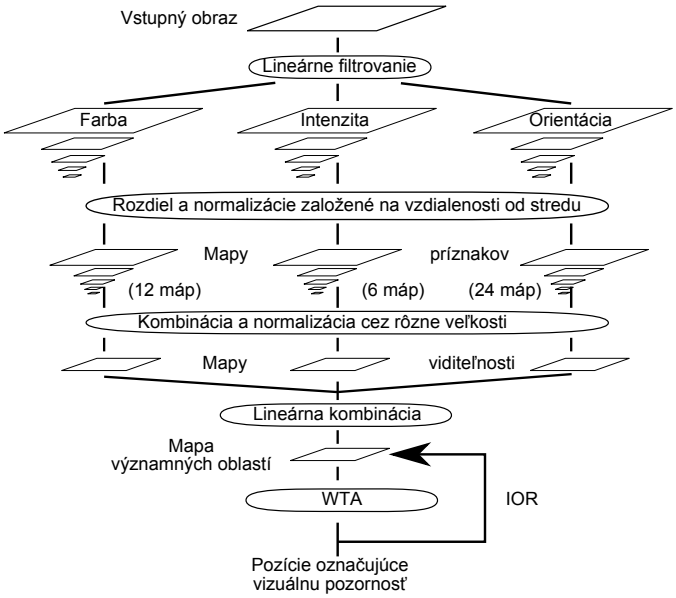
Prvý výpočtový model na detekciu významných oblastí využíval nízkoúrovňové príznaky. Niektoré ďalšie modely už využívajú kombináciu nízkoúrovňových a vysokoúrovňových príznakov. Tieto kombinácie dokážu pomôcť pri riešení veľmi špecifických úloh, ako napríklad detekcia chodcov. V súčastnosti sa výskum nachádza iba na začiatku tvorby výpočtových modelov, ktoré by popri nízkoúrovňových príznakoch využívali aj pridanú informáciu, čím by sa zlepšila a spresnila detekcia významných oblastí obrazu.

2.2 Model vizuálnej pozornosti založený na významných oblastiach pre rýchlu analýzu scény

Tento model [16] je založený na správaní a štruktúre neurónov vizuálneho systému primátov. Kombinuje viacúrovňové príznaky do jednej mapy výrazných oblastí. Používa dynamicku neurónovú sieť, ktorá vyberá oblasti záujmu v poradí od najvýraznejšej po menej výrazné. Tento systém rieši zložitý problém porozumenia scény výpočtovo efektívnym spôsobom.

Na vstupe sú farebné obrázky v rozlíšení 640×480 pixlov. Z nich je pomocou Gaussových pyramíd vytvorených 9 obrázkov určených na ďalšie spracovanie. Tieto sú zmenšenou kópiou pôvodného obrázka v pomere 1:1 (veľkosť 0) až 1:256 (veľkosť 8).

Každý príznak je vypočítaný pomocou operácií zameraných na stred (center-surround). Keďže vizuálne neuróny sú najcitlivejšie v malej oblasti (stred), zatiaľ čo na podnety zo širšej oblasti sú neuróny menej citlivé. Zameranie na stred je v tomto modeli implementované pomocou rozdielu medzi jemnými a hrubými mierkami. V strede je pixel s veľkosťou $c \in \{2, 3, 4\}$, v širšom okolí je pixel reprezentovaný veľkosťami $s = c + \delta$, kde $\delta \in \{3, 4\}$. Rozdiel medzi dvoma mapami "$\ominus$" je získaný pomocou interpolácie na jemnejšiu veľkosť a následným odčítaním. Na obrázku Obr 2.2, je znázornená schéma popisovaného modelu.



Obr. 2.2: Základná architektúra modelu [16]

2.2.1 Určovanie príznakov

Zo vstupného obrazu vo formáte RGB (r–červená, g–zelená, b–modrá) je získaná mapa pre intenzitu $I = (r + g + b)/3$, ktorá slúži ako vstup na vytvorenie Gausovskej pyramídy $I(\delta)$, kde $\delta \in \{0, ..., 8\}$. Kanály r,g,b sú normalizované.

Následne sú vytvorené farebné kanály: $R = r - (g + b)/2$ pre červenú, $G = g - (r + b)/2$ pre zelenú, $B = b - (r + g)/2$ pre modrú a $Y = (r + g)/2 - |r - g|/2 - b$ pre žltú (záporné hodnoty sú nastavené na 0). Z týchto kanálov sú vytvorené štyri Gausovské pyramídy $R(\delta)$, $G(\delta)$, $B(\delta)$ a $Y(\delta)$.

Neuróny cicavcov sú citlivé na náhlu zmenu intenzity, ako napríklad tmavý stred obklopený svetlým okolím, alebo svetlý stred obklopený tmavým okolím. Oba spomenuté typy citlivosti na zmenu intenzity model prepája, čím vznikne 6 máp $\mathcal{I}(c,s)$, kde $c \in \{2,3,4\}$ a $s = c + \delta$, kde $\delta \in \{3,4\}$:

$$\mathcal{I}(c,s) = |I(c) \ominus I(s)|. \tag{2.1}$$

Druhou množinou sú mapy pre farebné kanály, ktoré sú v mozgovej kôre reprezentované pomocou systému "dvojíc oponentov". V strede oblasti vnímania sú neuróny stimulované jednou farbou (napr. červená) a potláčané druhou farbou (napr. zelená), avšak v širšej oblasti je to opačne. Takto navzájom pôsobiace dvojice farieb sú červená/zelená, zelená-/červená, modrá/žltá, žltá/modrá. Viedlo to k vytvoreniu $\mathcal{RG}(c,s)$ mapy pre červenú a zelenú farbu a $\mathcal{BY}(c,s)$ mapy pre modrú a žltú farbu.

$$\mathcal{RG}(c,s) = |(R(c) - G(c)) \ominus (G(s) - R(s))|, \tag{2.2}$$

$$\mathcal{BY}(c,s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))|. \tag{2.3}$$

Posledná množina máp pre orientáciu je získaná z I pomocou orientovaných Gaborových pyramíd $\mathcal{O}(\delta,\theta)$, kde $\delta \in \{0,...,8\}$ reprezentuje veľkosť a $\theta \in \{0°, 45°, 90°, 135°\}$ orientáciu. Mapy pre orientáciu predstavujú lokálne zmeny medzi stredom a okolím:

$$\mathcal{O}(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)|. \tag{2.4}$$

Pomocou týchto vzťahov vznikne 42 máp príznakov: šesť pre intenzitu, dvanásť pre farbu a dvadsaťštyri pre orientáciu.

2.2.2 Mapa významných oblastí

Pri kombinácii všetkých máp nastáva situácia, ktorá spôsobuje, že oblasti, ktoré sú veľmi výrazné iba v niektorých z kombinovaných máp môžu zaniknúť. Z toho dôvodu je potrebný normalizačný operátor $\mathcal{N}(.)$, ktorého aplikácia pozostáva z troch krokov:

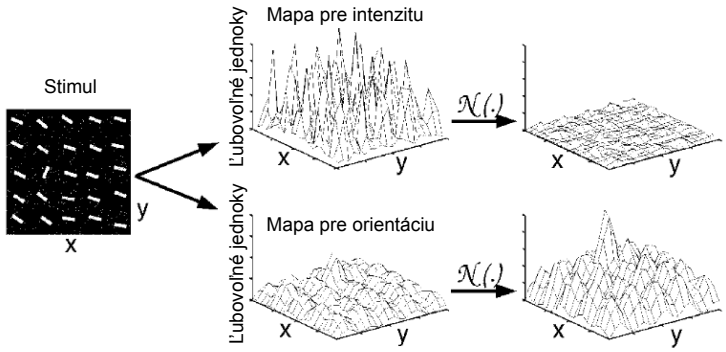
1. normalizovanie hodnôt mapy na pevný rozsah v rozmedzí $\{0,...,M\}$ s cieľom odstrániť rozdiely medzi mapami

2. nájdenie pozície lokálneho maxima M a vypočítanie priemernej hodnoty $\overline{m}$ všetkých týchto maxím

3. globálne vynásobenie mapy s $(M - \overline{m})^2$

V prípade, že je v mape veľký rozdiel, najvýraznejšia oblasť vystúpi do popredia. V prípade, že je rozdiel malý, mapa neobsahuje nič výrazné a je potlačená.

Mapy sú následne kombinované do troch "máp viditeľnosti" pre intenzitu $\overline{\mathcal{I}}$, farbu $\overline{\mathcal{C}}$ a orientáciu $\overline{\mathcal{O}}$ s veľkosťou $\delta = 4$. Tieto mapy sú získané pomocou sčítania "$\oplus$", ktoré sa skladá z redukcie každej mapy na veľkosť 4 a následnom sčítaní bod po bode.



Obr. 2.3: Normalizačný operátor $\mathcal{N}(.)$ [16]

Mapy pre intenzitu $\overline{\mathcal{I}}$ a farbu $\overline{\mathcal{C}}$ sú získané nasledovne:

$$\overline{\mathcal{I}} = \oplus_{c=2}^{4} \oplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{I}(c,s)) \tag{2.5}$$

$$\overline{\mathcal{C}} = \oplus_{c=2}^{4} \oplus_{s=c+3}^{c+4} [\mathcal{N}(\mathcal{RG}(c,s)) + \mathcal{N}(\mathcal{BY}(c,s))] \tag{2.6}$$

Pre orientáciu sú najprv vytvorené štyri mapy kombináciou šiestich máp s danou orientáciou θ, ktoré sú následne sčítané:

$$\overline{\mathcal{O}} = \sum_{\theta \in \{0°,45°,90°,135°\}} \mathcal{N}\left(\oplus_{c=2}^{4} \oplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{O}(c,s,\theta)) \right) \tag{2.7}$$

Tieto tri mapy sú normalizované a sčítané do výslednej mapy významných oblastí:

$$S = \frac{1}{3}\Big(\mathcal{N}(\overline{\mathcal{I}}) + \mathcal{N}(\overline{\mathcal{C}}) + \mathcal{N}(\overline{\mathcal{O}})\Big). \tag{2.8}$$

Maximum výslednej mapy v danom čase ukazuje najvýznamnejšie miesto, na ktoré by mala byť zameraná pozornosť.

V neurónovej implementácii bol vytvorený model mapy výrazných oblastí (angl. saliency map, ozn. SM) ako 2D vrstvy presakujúcich neurónov s vlastnosťou "integrate-and-fire" a s veľkosťou štyri. Tieto neuróny majú kapacitu, ktorá zahŕňa určitú hodnotu a prah. Keď je daný prah dosiahnutý, vytvorí sa "hrot" a kapacitný náboj je nastavený na nulu. Potom je SM daná 2D "winner-take-all" (angl. winner take all, ozn. WTA) neurónovej sieti, v ktorej nájdená výrazná oblasť ostáva a ostatné sú potlačené.

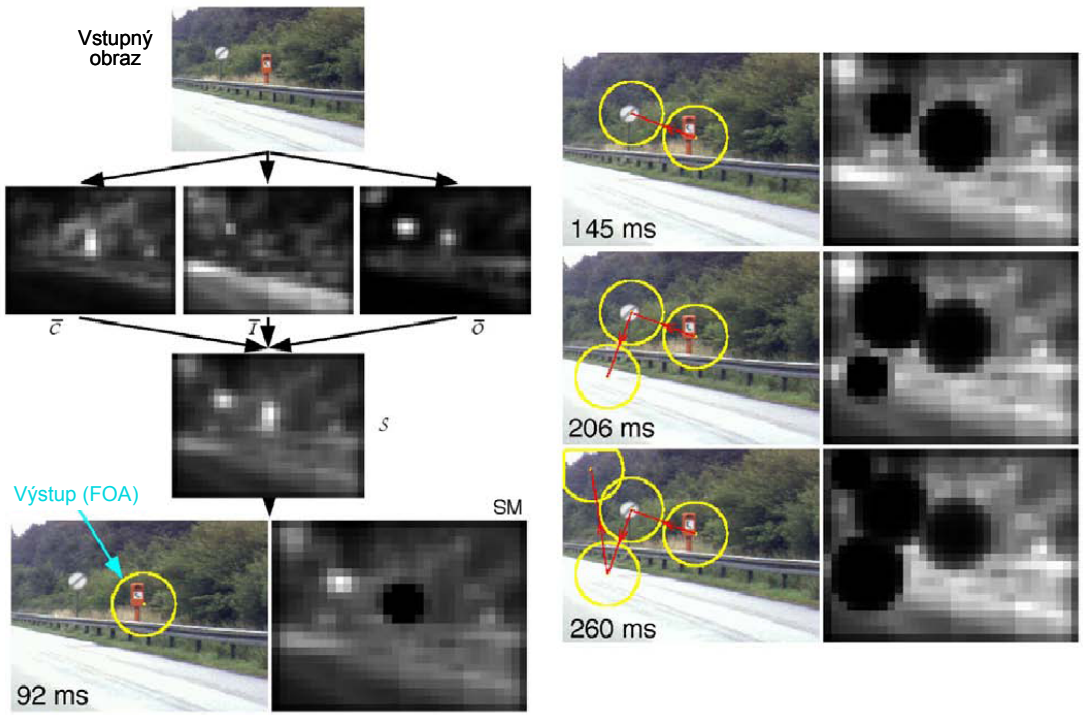
Neuróny dostávajú vstupy z mapy S a sú všetky nezávislé. Potenciál neurónov v SM patriacich najvýraznejšej oblasti vzrastá najrýchlejšie. Každý SM neurón excituje jeho príslušný WTA neurón. Všetky WTA neuróny sa tiež vyvíjajú nezávisle na sebe pokiaľ jeden ("víťaz") nedosiahne prahovú hodnotu. Toto vyvolá tri súbežné mechanizmy:

1. oblasť pozornosti (angl. focus of attention, ozn. FOA) je posunutá na miesto víťazného neurónu

2. globálne sa spustí WTA a inicializuje na predvolenú hodnotu všetky neuróny

3. v SM sa dočasne aktivuje lokálne potlačenie v oblasti s veľkosťou a umiestnením novej FOA. Takto je možná dynamická zmena FOA prostredníctvom povolenia nasledujúcej výraznej oblasti stať sa "víťazom". Taktiež sa nevraciame na už spracované FOA.

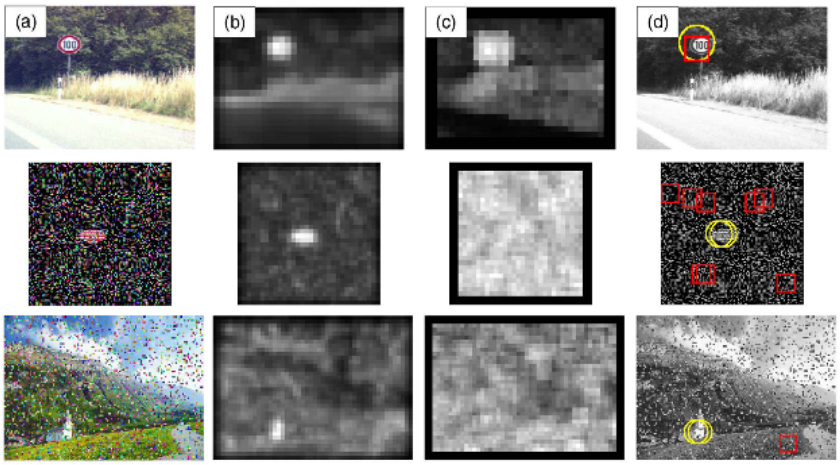
Ľudská visuálna psychika funguje na princípe skúmanie scény so zákazom vracať sa na už navštívené oblasti (angl. inhibition-of-return, ozn. IOR). Obr 2.4 je znázorňuje príklad operácií modelu v pozorovanej scéne. Najprv je spustená súbežná extrakcia príznakov, výstupom sú tri mapy kontrastov pre farbu $\overline{\mathcal{C}}$, intenzitu $\overline{\mathcal{I}}$ a orientáciu $\overline{\mathcal{O}}$. Tieto mapy sú kombinované, čím vznikne mapa výrazných oblastí (SM). Najvýraznejšou časťou obrazu je oranžová oblasť s telefónom, ktorá je veľmi výrazná v mape $\overline{\mathcal{C}}$. Preto sa stáva prvou oblasťou pozornosti (výpočet trval 92 ms simulovaného času). Pomocou kroku "inhibition-of-return" sa toto miesto už znova nespracováva, čím môžu byť vybrané ďalšie významné oblasti.



Obr. 2.4: Schéma výberu významných oblastí [16]

Autori [16] porovnávali výsledky vytvoreného modelu s SFC (angl. Spatial Frequency Content Model, ozn. SFC). Na získanie SFC bolo vytvorené jednoduché meranie. V danej polohe v obraze bol extrahovaný blok s rozmermi 16×16 pixlov z máp $I(2)$, $R(2)$, $G(2)$, $B(2)$, a $Y(2)$ a na tieto bloky boli aplikované 2D FFT (angl. Fast Fourier Transform) a prahovanie. SFC je priemer čísel nezanedbateľných koeficientov v piatich zodpovedajúcich blokoch. Veľkosť a počet blokov sú vybrané tak, aby bolo SFC meranie približné navrhnutému modelu. Použitím SFC s veľkosťou štyri je vytvorená mapa, ktorá je porovnávaná s navrhnutým modelom (Obr. 2.5).

Tento model je výpočtovo nenáročný, má jednoduchú architektúru. Dokáže dobre detegovať oblasti vizuálnej pozornosti v scéne. Rýchlo určuje napríklad dopravné značky rôznych tvarov, farieb, a textúru aj napriek tomu, že nebol navrhnutý na tento účel. Z výpočtového hľadiska je hlavnou výhodou tohto modelu parelelná implementácia, nielen vo výpočtovo náročnej fáze extrakcie príznakov, ale aj v systéme zaostrovania na výrazné oblasti. Využitie neurónových sietí sa ukázalo ako účinné pri reprodukcii niektorých výkonov vizuálneho systému primátov. Jeho účinnosť však závisí na implementovaných príznakoch.



Obr. 2.5: Časť (a) označuje vstupné obrázky, (b) predstavuje korešpondujúce mapy významných oblastí, (c) SFC mapy. V (d) sú znázornené oblasti, kde bol vstup mapy výrazných oblastí väčší ako 98 percent maxima (žlté kružnice) a bloky obrazu pre ktoré bolo SFC vyššie ako 98 percent maxima (červené čtvorce). Mapy výrazných oblastí sú odolné na šum, zatiaľ čo SFC nie [16].

2.3 Model založený na spracovaní dát získaných sledovaním očí

Veľké množstvo prístupov na detekciu významných oblastí je založených na prístupe zdola-nahor. Neberú do úvahy sémantiku obrazu a často nekorešpondujú s reálnymi pohybmi očí. Autori prístupu [18] sa preto rozhodli zamerať práve na sémantiku. Zozbierali unikátne dáta zo sledovania očí a použili ich na trénovanie a testovanie modelu založeného na nízko, stredné a vysokoúrovňových príznakoch.

Súčasné modely na detekciu významných oblastí nedokážu presne určiť fixácie oka. Na obrázku 2.6 je znázornená mapa významných oblastí získaná pomocou Ittiho prístupu [16]. Na obrázku vpravo sú zobrazené fixácie oka získané prostredníctvom systému na sledovanie očí. Ittiho prístup jasne deteguje oblasti s vysokou úrovňou jasu, ostré hrany, avšak len veľmi malé množstvo z týchto oblastí je pre človeka zaujímavých. Preto je popisovaný prístup vhodný na detegovanie významných oblastí v kontexte, kde je vytvorený model s veľkým množstvom príznakov [18].

Autori zostavili vlastnú databázu, ktorá obsahuje obrázky s popismi a analýzou, vytvorili riadený model, ktorý kombinuje príznaky získané prístupmi zdola-nahor a sémantické príznaky, čím vytvára presnejšie mapy významných oblastí.

Obr. 2.6: Porovanie detegovaných významných oblastí [16] (v strede) s fixáciami oka získanými sledovaním očí (červené bodky).

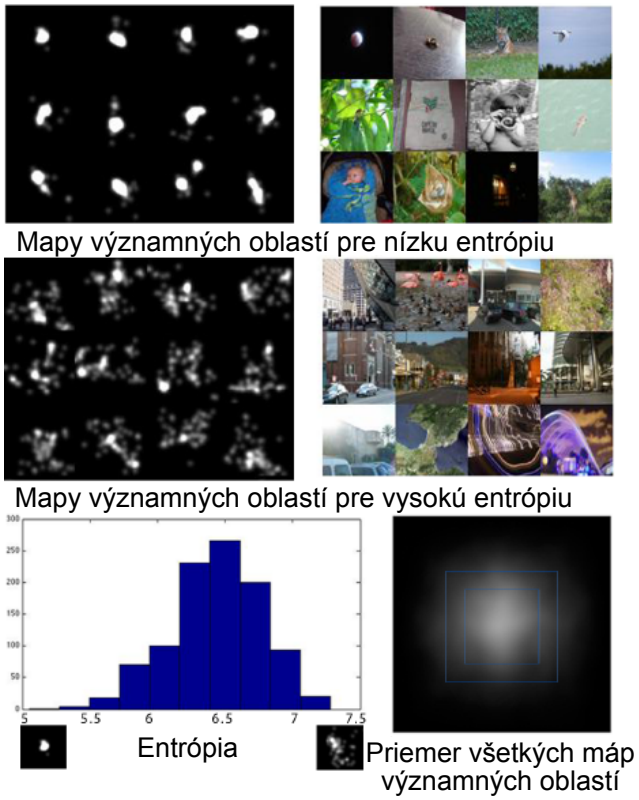
2.3.1 Databáza a jej spracovanie

Použitá databáza je jedinečná svojím obsahom, keďže zahŕňa obrázky a k nim príslušné dáta. Dáta ku každému obrázku boli získané sledovaním pohybu očí 15 respondentmi. Databáza pozostáva z 1003 obrázkov, z toho 779 prírodných obrázkov a 229 portrétov. Rozmery obrázkov sa pohybujú v rozmedzí od 405 do 1024 pixlov pre najdlhšiu stranu, kde väčšina má veľkosť 768 pixlov. Vek pozorovateľov sa pohyboval v rozmedzí od 18 do 35 rokov. Každému pozorovateľovi boli obrázky zobrazené 3 sekundy a oddelené jednosekundovou šedou obrazovkou. Kvôli presnosti bol systém kalibrovaný každých 50 obrázkov.

Spracovanie databázy

Prvým krokom v práci bola analýza dát z databázy. Pomocou nej sa zistilo, že pri niektorých obrázkoch pozorovatelia upriamujú svoju pozornosť na rovnaké oblasti, avšak pri iných je ich pozornosť rozptýlená po celom obraze. Analýzou konzistencie fixácií v obraze

pomocou merania entrópie priemerných máp sa zistilo, že tieto dáta ukazujú silné odchýlky pre fixácie oči v blízkosti stredu obrazu. Obrázok 2.7 znázorňuje histogram entrópií pre obrázky z databázy. Taktiež sú na ňom zobrazené vzorky z 12 máp významných oblastí pre nízku a vysokú entrópiu a k nim zodpovedajúce obrázky. Podľa merania entrópie zistili, že obrázky s vysokou konzistenciou/nízkou entrópiou obsahujú väčšinou objekty v strede. Obrázky s nízkou konzistenciou/vysokou entrópiou často obsahujú niekoľko textúr.



Obr. 2.7: Analýza fixácií pozícií

Na obrázku 2.7 je zobrazená priemerná mapa významných oblastí pre všetky skúmané obrazy. Je možné vidieť, že 40% fixácií sa nachádza v 11% stredu obrazu a až 70% fixácií sa nachádza v 25% obrazu, ktoré tvoria jeho stred. Toto je dôsledkom faktu, že pri pozorovaní bol obraz umiestnený priamo pred pozorovateľom a tiež tým, že na fotografiách sú väčšinou objekty záujmu umiestnené v strede, alebo v jeho tesnom okolí [46].

Na ohodnotenie toho, ako dokážu modely predpovedať skutočné fixácie očí, sa autori rozhodli použiť ROC (angl. Receiver Operating Characteristic) krivky, významné oblasti

sa spraovávajú vždy pre jedného pozorovateľa. Pomocou prahovania sa určí, ktoré časti obrazu obsahujú významné časti a ktoré nepútajú pozornosť. Dáta od ostatných pozorovateľov sú následne použité ako ground truth. Pomocou postupnej zmeny prahu je vykresľovaná ROC krivka.

Veľké množstvo fixácií je zameraných na ľudí (vrátane reprezentácií ľudí ako sú sochy a obrazy) aj v prípade, že ich tváre nie sú viditeľné (tváre tvoria 10% z celkových fixácií). Fixácie na text tvoria 11% z celkového počtu fixácií - znaky sú väčšinou navrhnuté tak, aby boli výrazné, najmä v prípade dopravných značiek, názvov obchodov, reklám, atď. Ďalšie fixácie sú zamerané na zvieratá, autá, či časti ľudského tela (oči a ruky).

Rozborom obrázkov s tvárami sa tiež zistilo, že pozorovatelia zameriavajú svoju pozornosť na tváre v prípade že zaberajú v obraze primeranú plochu. V prípade detailu tváre sa zameriavajú na jej časti, ako sú oči, nos, alebo pery. Z toho vyplýva, že pri výbere objektov, na ktoré pozoroveteľ zameriava svoju pozornosť, zohráva dôležitú úlohu ich veľkosť (Obr. 2.8).

Obr. 2.8: Na obrázku vľavo sú tváre dostatočne malé na to, aby sa pozorovateľ zameral na ne ako na celok. Na obrázku vpravo je však tvár dostatočne veľká, takže pozorovetelia sa zameriavajú na jej časti, ako sú napríklad oči, ústa a nos.

2.3.2 Príznaky

Väčšina navrhnutých modelov využíva na určenie významných oblastí kombináciu filtrov založených na fyziologickej štruktúre ľudského vizuálneho systému. Na rozdiel od takejto kombinácie filtrov, autori v tejto práci využívajú proces učenia na trénovanie klasifikátora

priamo z dát získaných sledovaním fixácií očí.

Na vytvorenie modelu boli použité tri typy príznakov – nízkoúrovňové, stredne– a vysokoúrovňové. Pre každý skúmaný obrázok sú predspracované príznaky pre každý bod obrazu a tieto informácie sú použité na učenie modelu.

Nízkoúrovňové príznaky

Nízkoúrovňové príznaky sú založené na fyziológii ľudského vizuálneho systému. Ako príznaky boli okrem iných použité jednoduché modely významných oblastí (napríklad Torabla [30] a Rosenholtz [34]). Ďalšie použité nízkoúrovňové príznaky sú intenzita, orientácia, farebný kontrast. K príznakom boli zahrnuté aj hodnoty červeného, zeleného a modrého farebného kanála a pravdepodobnosť každej farby vypočítaná z 3D farebného histogramu obrazu, ktorý bol filtrovaný s mediánovým filtrom na 6 rôznych úrovniach.

Stredoúrovňové príznaky

Veľké množstvo objektov sa nachádza na zemi, nie vo vzduchu, a preto ľudia pri pozorovaní scény upriamujú svoju pozornosť na horizont. Model využíva detektor horiznotu, tento príznak zaraďujeme medzi stredoúrovňové.

Vysokovňové príznaky

V prípade, že sa v obraze nachádzajú ľudia, pozorovatelia na nich upriamujú svoju pozornosť. Model používa detekciu tváre pomocou Viola-Jones detektora a taktiež Felzenszwalbovu detekciu osôb.

Uprednostňovanie stredu

Pri fotografovaní ľudia prirodzene umiestňujú objekty záujmu do blízkosti stredu fotografií. Ako ďalší príznak bola preto použitá vzdialenosť každého obrazového bodu od stredu obrazu.

2.3.3 Učenie

Databáza bola rozdelená na dve časti, kde na trénovanie bolo použitých 903 a na testovanie zvyšných 100 obrázkov.

Ku každému obrazu v databáze prislúchali dáta získané sledovaním očí pozorovateľov, ktoré označovali významné oblasti obrazu. Pomocou týchto dát autori rozdelili obraz na oblasti s rôznou úrovňou významnosti. Pre každý obraz z databázy bolo vybraných 10 pozitívne ohodnotených vzoriek z prvých 20% oblastí obrazu ohodnotených ako významné a 10 vzoriek ohodnotených ako negatívne oblasti. Ako negatívne boli označené oblasti, na ktoré pozorovateľ neupriamuje svoju pozornosť.

Takýmto spôsobom bolo vytvorených 18060 vzoriek na testovanie a 2000 vzoriek na trénovanie.

Na trénovanie na 9030 pozitívnych a na rovnakom počte negatívnych vzoriek bol použitý systém SVM (Support vector machine).

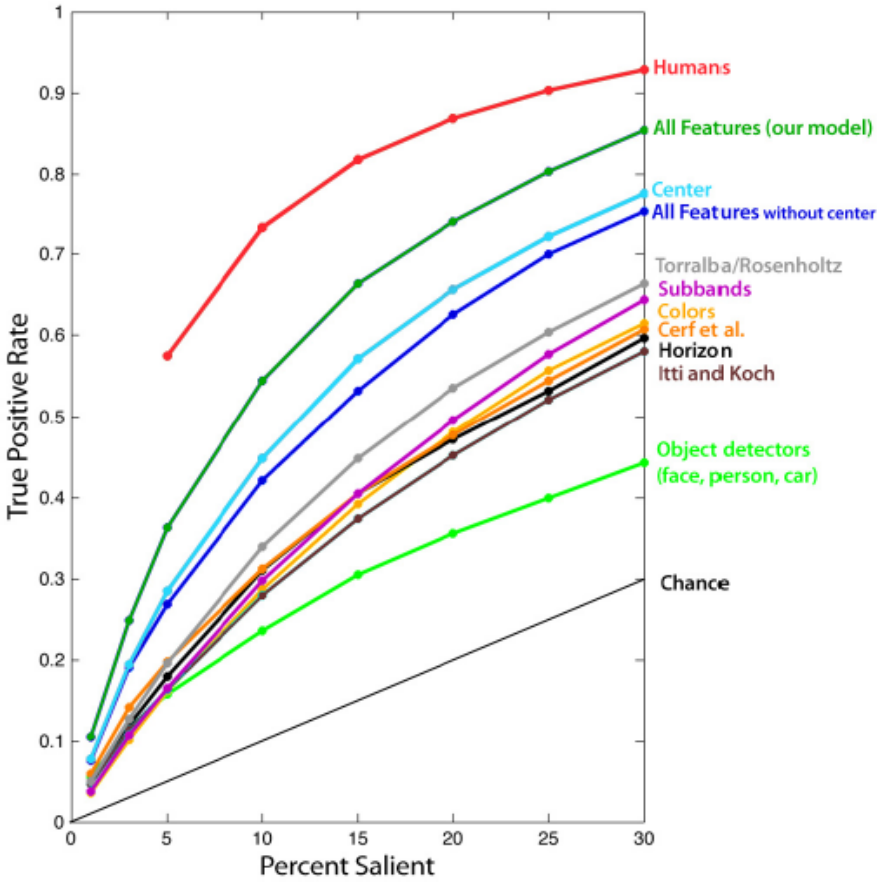
2.3.4 Výsledky a pozorovania

Presnosť a výkonnosť modelu autori vyhodnocovali dvoma spôsobmi. Najprv merali výkon modelu pre detekciu na celých obrázkoch pomocou ROC kriviek. Neskôr testovali detekcie porovnávaných modelov na špecifických vzorkách, ktoré boli získané zo stredovej oblasti obrazu, mimo nej a na tvárach.

Na obrázku 2.9 sú znázornené ROC krivky popisujúce výsledky detekcií rôznych modelov získané priemerovaním na všetkých obrázkoch z databázy. Na horizontálnej osi sú prahy znázorňujúce, koľko percent z detegovaných oblastí bolo považovaných za významné.

Pomocou analýzy týchto ROC kriviek boli vyslovené nasledujúce zistenia týkajúce sa celých obrázkov:

- Navrhnutý model kombinujúci všetky príznaky dosiahol lepšie výsledky ako model používajúci iba jednu množinu príznakov. Tiež dosiahol lepšie výsledky ako napríklad modely Torralba [30], Rozenholtz [34], Itti a Koch [16].
- Navrhnutý model dosiahol 88% z výsledkov získaných pomocou sledovania očí.



Obr. 2.9: Graf znázorňujúci ROC krivky pre navrhnutý model (Tmavozelená krivka) a viacero iných modelov. Pre porovnanie znázornená krvka zobrazujúca hodnoty namerané z reálnych dát (červenou farbou).

- Model, ktorý obsahoval všetky príznaky okrem vzdialenosti od stredu (2.9: tmavomodrá krivka) dosahoval podobné výsledky ako model berúci do úvahy vzdialenosť od stredu.

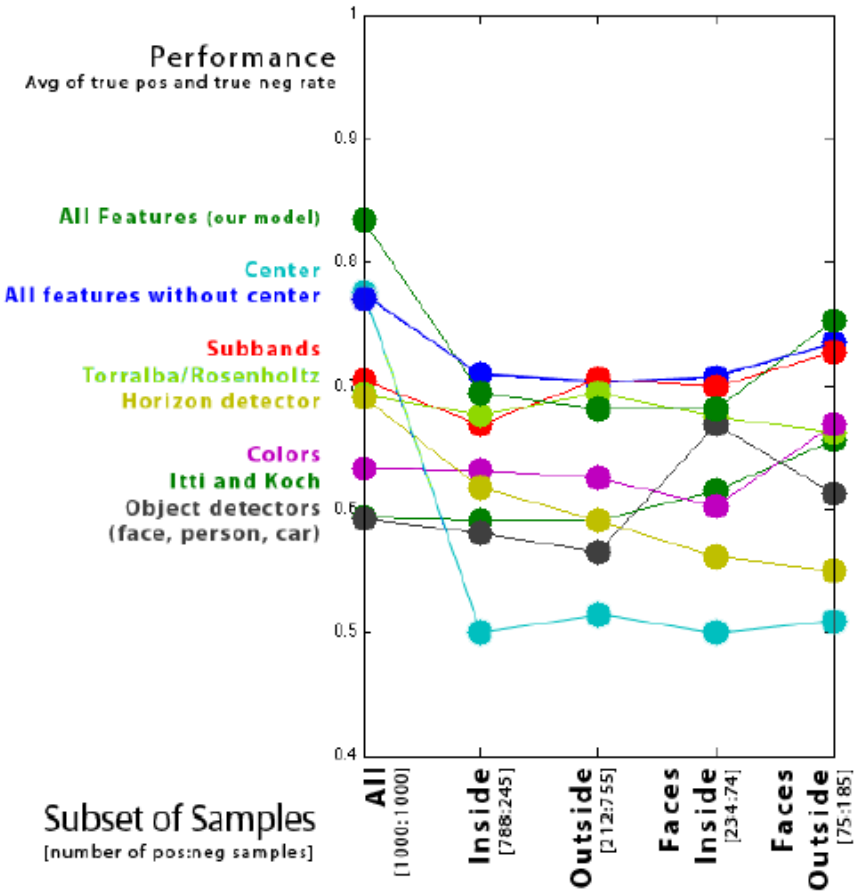
- Model obsahujúci všetky príznaky okrem vzdialenosti od stredu dosahoval lepšie výsledky ako model používajúci len jednu podmnožinu príznakov.

- Detektory objektov (2.9: svetlozelená krivka) dávajú veľmi dobré výsledky v prípade, že sa dané objekty v obraze nachádzajú. Vo všeobecnosti sú však dosahujú veľmi slabé výsledky, a preto by malo byť využívané iba v kombinácii s inými príznakmi.

- Výsledky všetkých modelov boli výrazne lepšie ako šanca určiť, že každý z príznakov má samostatne schopnosť predpovedať významné oblasti (2.9: čierna krivka).

Kvôli porozumeniu vplyvu odklonu od stredu boli obrázky rozdelené do oblastí stredu (kruh) a okolia. Stredné oblasti boli definované modelom založeným iba na príznaku, ktorý udával vzdialenosť od stredu obrazu. V tomto modeli každá snímka, ktorá bola od stredu vzdialenejšia ako 0.42 bodov (kde vzdialenosť od stredu k okrajom je 1) bola označená ako negatívna a akákoľvek bližšia snímka bola označená ako pozitívna. Pomocou tohto prahu boli vzorky rozdelené do dvoch skupín.

Na obrázku 2.10 je graf, ktorý zobrazuje výsledky modelov pre rozličné vzorky. Výsledky sú tu definované ako priemer true positive a true negative ohodnotení.



Obr. 2.10: Priemerné hodnoty nameraných true positive a true negative pre SVM, ktorý bol trénovaný s na rôznych množinách vzoriek.

Pri testovaní vzoriek sa zistilo:

- Aj napriek tomu, že model založený na detekcii stredu dával veľmi dobré výsledky pri všetkých vzorkách, jeho výkon pri ostatných vzorkách bol porovnateľný so šancou určiť to, že každý z príznakov má samostatne schopnosť predpovedať významné oblasti.

- Kým pri všetkých vzorkách model zameraný na detekciu stredu a model so všetkými príznakmi okrem detekcie stredu dávali rovnaké výsledky, pri ďalších množinách vzoriek druhý model dával oveľa lepšie výsledky.

- Model trénovaný na príznakoch z detekcie objektov pre tváre, ľudí a autá dával najlepšie výsledky pri podmnožinách s tvárami.

- SVM používajúce príznak uprednostňovania stredu a tie, ktoré používajú všetky príznaky, dávali veľmi dobré výsledky na množine všetkých vzoriek, ale v prípade ďalších vzoriek dosahoval druhý detektor lepšie výsledky.

Analyzovaním nameraných dát zo zostavenej databázy sa zistilo, že pri pozorovaní scény ľudia zameriavajú svoju pozornosť na text a iných ľudí, hlavne na ich tváre. Ak sa v scéne nenachádzajú ľudia, pozornosť sa upriamuje na ostatné živé bytosti, napr. zvieratá. Pri absencii takýchto špecifických objektov a textu sa ľudia zameriavajú na stred obrazu, alebo oblasti, kde sú výrazné nízkoúrovňové príznaky. Popisovaný model založený na rôznoúrovňových príznakoch je založený na týchto pozorovaniach a teda sa približuje reálnemu vnímaniu scény [18].

Kapitola 3

Špecifikácia diela

Táto kvalifikačná práca je zameraná na použitie sémantickej informácie pri hľadaní významných oblastí v obraze. V tejto časti je popísaná motivácia pre jej vytvorenie, základný model, ako aj navrhnuté riešenie.

3.1 Motivácia a cieľ práce

Motiváciou pre túto rigoróznu prácu bol fakt, že získané informácie o vizuálnej pozornosti a o tom, kam ľudia zameriavajú svoju pozornosť, je možné využiť pri veľkom množstve aplikácií. Pri kompresii obrazu je možné zamerať sa na významné oblasti, ktoré sú skomprimované bezstratovo a ostatné stratovo, čím sa nestráca dôležitá informácia. Taktiež môžme informácie o významných oblastiach využiť pri retargetingu, kódovaní obrazu, odstraňovaní nežiadúcich častí obrazu [38].

Cieľom tejto práce je využitie vedomostí o detegovaní významných oblastí v scéne a pridanie sémantickej informácie, ktorá by pomohla pri bližšom určovaní relevantných významných oblastí. Práca je plynulým pokračovaním diplomovej práce [20], kde sa autorka zamerala na detekciu významných oblastí s pridaním informácie o detegovaných tvárach v obraze, kde podľa zistení v práci [22] sú tváre v obraze veľmi významné.

Uvedené východiská tvoria veľmi dobrý základ predloženej práce, ktorá sa primárne zameriava na spracovanie obrazu zachytávajúceho osobu komunikujúcu posunkovou rečou. V ďalších kapitolách je používný model implementovaný v prostredí MATLAB R2007b.

Tento je následne optimalizovaný a je k nemu pridaná sémantická informácia relevantná pre daný výskum. V kapitole 5 sú výsledky porovnávané a je vybraný najlepší prístup pre ďalší výskum.

Na implementáciu metód a modelov bolo zvolené použitie softvéru MATLAB spoločne s jeho rozšírením*Image Processing Toolbox*. Softvér MATLAB bol zvolený predovšetkým preto, že umožňuje jednoducho a prehľadne formulovať riešenia technických výpočtových problémov, špeciálne tých, ktoré zahŕňajú maticovú implementáciu a taktiež kvôli tomu, že veľké množstvo nových prác v oblasti spracovania obrazu obsahuje implementáciu v MATLABe.

3.2 Základný model

V tejto časti je popísaný model navrhnutý v [14], ktorý je použitý ako východiskový model pre daný výskum.

Pôvodní autori použili pre detekciu výrazných oblastí vo vstupnom obraze ako prídavný príznak pozornosti textúru. Taktiež navrhli stratégiu kombinácie príznakov, ktorá v kontrastnej mape potláča oblasti nesprávne označené ako významné. Táto stratégia používa na potlačenie sporných oblastí lokálnu kontextovú informáciu a zvýraznenie pravých výrazných oblastí.

3.2.1 Získanie textúrneho príznaku

Textúra je veľmi vhodná na zachytenie vizuálnej pozornosti v obrazoch obsahujúcich malé objekty. Textúrny príznak pozornosti použitý v základnom modeli bol získaný nasledovne.

Vstupný obraz bol rozdelený na bloky, nazvané *textúrne bloky (texture patches)*, kde každý blok obsahoval $p \times q$ pixlov, $p, q \in N$. Pomocou Gaborovej vlnovej transformácie aplikovanej na rôznych veľkostiach boli pre každý textúrny blok získané stredné hodnoty μ_{sk} a štandardné odchýlky σ_{sk}, kde s označuje veľkosť a k označuje orientáciu. Pre S veľkostí a K orientácií tak vzniklo SK máp stredných hodnôt $MM_{s,k}$ a SK máp štandardných odchýlok $SDM_{s,k}, s = 1, ..., S$ a $k = 1, ..., K$. Následne boli vytvorené mapy Average Mean Difference (AMD) a Average Standard Deviation Difference (ASDD) využitím susedných

blokov, kde AMD pre blok na pozícii (i, j) je daná ako

$$AMD_{s,k}(i,j) = \frac{1}{N} \sum_{u,v} |MM_{s,k}(i+u, j+v) - MM_{s,k}(i,j)| \tag{3.1}$$

a ASDD pre rovnaký blok je daná ako

$$ASDD_{s,k}(i,j) = \frac{1}{N} \sum_{u,v} |SDM_{s,k}(i+u, j+v) - SDM_{s,k}(i,j)|, \tag{3.2}$$

kde N predstavuje počet susedov. Kontrast textúry (angl. texture contrast, ozn. TC) v bloku (i, j) v spracovanej mape veľkosti s a orientácii k bol dopočítaný pomocou vzťahu ako

$$TC_{s,k}(i,j) = AMD_{s,k}(i,j).ASDD_{s,k}(i,j). \tag{3.3}$$

Následne bol výsledný kontrast textúry v bloku (i, j) vypočítaný pomocou

$$TC(i,j) = \sum_{s} \sum_{k} TC_{s,k}(i,j). \tag{3.4}$$

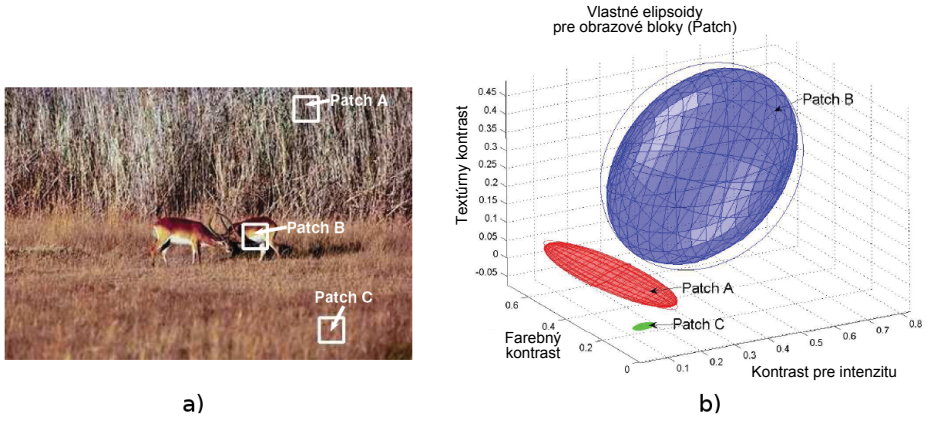
Takto vzpočítaný textúrny kontrast určuje oblasti záujmu aj v prípade, keď iné príznaky ako farba a intenzita zlyhajú.

3.2.2 Kombinácia príznakov

V tejto časti je popísaná stratégia kombinácie príznakov a tvorba faktora potlačenia (angl. suppression factor, ozn. SF) v lokálnom kontexte pre adaptívnu kombináciu násobných príznakov pozornosti ako intenzita, farba a textúra. Obraz bol rozdelený do blokov nazvaných *Attention Patches*. Každý blok obsahuje $p \times q$ pixlov. Zmena kontrastu príslušného príznaku v bloku centrovanom v (i, j) bola počítaná ako

$$FV(i,j) = \frac{1}{N} \sum_{u,v} |MF(i,j) - MF(i+u, j+v)|, \tag{3.5}$$

kde $MF(i, j)$ je stredná hodnota príznaku v bloku (i, j) a N je počet susediacich blokov. Hodnotz kontrastu na bloku (i, j) pre n príznakov sú normované na interval $[0, 1]$. Každý blok je teda reprezentovaný pomocou n rozmerného kontrastného vektora príznakov, ktorý je porovnávaný s ostatnými v jeho susedstve. Kontrast spracovávaného bloku je potlačený v prípade, že susedné bloky sú podobné. Táto podobnosť je určená pomocou variacií dát



Obr. 3.1: a) Originálny obrázok b) Vlastné elipsoidy pre každý blok [14]

pozdĺž vlastných vektorov $n \times n$ kovariančnej matice. Táto matica je získaná z kontrastných vektorov príznakov pre blok (i, j) a jeho susedov. Vlastné hodnoty $\bar{\lambda}$ tejto matice reprezentujú podobnosť resp. rozdielnosť pozdĺž príznakov pozornosti. Napríklad veľká (malá) vlastná hodnota určuje veľkú (malú) varianciu pozdĺž smeru zodpovedajúceho vlastného vektora, čo určuje vysokú (nízku) mieru spoľahlivosti.

Na obrázku Obr. 3.1 a) je znázornený vstupný obrázok s vyznačenými tromi blokmi. Blok A patrí trstine v pozadí, C je z oblasti obsahujúcej trávu a B je z výraznej oblasti obsahujúcej antilopy. Máme trojrozmerný vektor príznakov, ktorého zložky predstavujú farbu, intenzitu a textúru. V Obr. 3.1 b) sú znázornené vlastné elipsoidy prislúchajúce blokom v Obr. 3.1 a). Osi elipsoidu sú v smere hlavných vektorov a ich dĺžka je úmerná zodpovedajúcim vlastným číslam. Ako možno vidieť na obrázku, blok C ma malú varianciu vo všetkých smeroch, zatiaľ čo blok A má varianciu v jednom smere väčšiu ako ostatné. Blok B má veľkú varianciu vo všetkých troch smeroch, čo určuje veľkú schopnosť odlíšenia v závislosti na jeho susedstve a mal by patriť k výrazným oblastiam obrazu. Preto je potrebné, aby bloky A a C boli potlačené, no blok B zvýraznený.

Faktor potlačenia SF pre blok (i, j) je získaný yo vzťahu $\tau(i, j) = \prod_{u=1}^{p} \bar{\lambda}_u$, kde sú hodnoty $\bar{\lambda}$ zoradené vzostupne a parameter p riadi mieru potlačenia. Pre získanie mapy výynamných oblastí $S(i, j)$ pre blok (i, j) sú násobné príznaky pozornosti lineárne kombi-

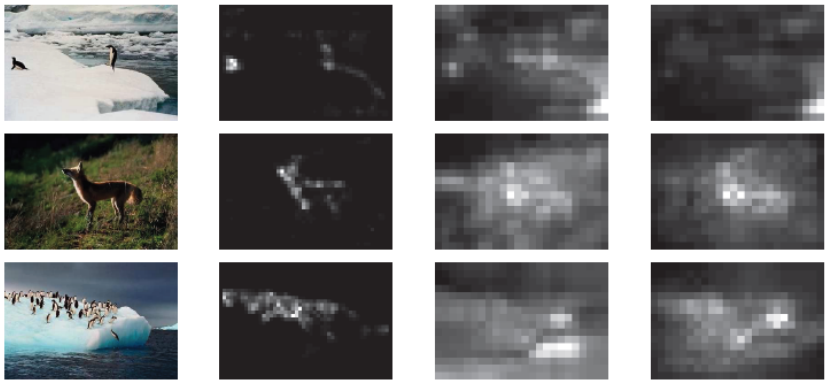
nované a výsledok je vynásobený SF

$$S(i,j) = \tau(i,j).\sum_{u=1}^{k} FV_u(i,j). \tag{3.6}$$

Výsledkom je mapa výrazných oblastí, ktorá obsahuje požadované oblasti záujmu (AR). Lineárna kombinácia kontrastných máp je implementovaná ako ich maticové sčítanie. SF pre každý blok je znázornený ako obraz s tmavými oblasťami reprezentujúcimi vysoký SF a svetlými oblasťami reprezentujúcimi nízky SF.

Príznaky ako farba aj textúra sú schopné zachytiť výrazné oblasti, avšak mapa zodpovedajúca farbe obsahuje viac sporných AR, ako mapa zodpovedaúca textúre. SF tieto sporné oblasti odstraňuje.

Na obrázku Obr. 3.2 sú výsledky tohto modelu znázornené a porovnané s ďalšími modelmi ([16], [15]) na detekciu výrazných oblastí v obraze [14].



Obr. 3.2: a) Originálny obrázok b) Mapa výrazných oblastí získaná pomocou popísaného modelu Hu [14] c) Itti [16] d) CSI [15]

3.3 Sémantická informácia

Algoritmy na detekciu významných oblastí v obraze môžu veľmi dobre detegovať časti obrazu s výraznou zmenou vo farbe, orientácii, textúre, avšak zlyhávajú pri detekcii sémanticky dôležitých častí obrazu.

V predchádzajúcej práci [20] sa autorka zamerala na využitie detekcie tvárí ako prídavného príznaku. Predkladaná práca sa zameriava na spracovanie obrazu zachytávajúceho posunkovú reč. Pri posunkovej reči sú najdôležitejšie časti obrazu ruky a tvár, čo predchádzajúci algoritmus nebral do úvahy. Preto je v tejto práci zamenená detekcia tvárí detekciou pokožky, čo by malo detekcii tvárí i rukám priradiť rovnakú dôležitosť.

Detekcia pokožky je dôležitá pri veľkom množstve aplikácií pracujúcich s obrazom. Používa sa napríklad pri detekcii tvárí v obraze, taktiež môže byť využívaná pri detekcii osôb. Problém pri detekcii pokožky v obraze je v ostatnom čase predmetom veľmi intenzívneho výskumu, postupne tiež našiel široké praktické uplatnenie. Existuje veľké množstvo prístupov na detekciu pokožky, v krátkosti predstavíme prahovanie, využitie histogramu, používanie Gaussianu.

Prahovanie je najjednoduchšia metóda na detekciu pokožky využívajúca ohraničenia v danom farebnom priestore. Pri prahovaní môže byť použitých niekoľko jednoduchých i zložitejších prahov, aplikovaných na rôzne zložky jednotlivých farebných priestorov. Detekciu pokožky pomocou prahovania možno uskutočniť v rôznych farebných modeloch, napr. RGB, HSV, YCbCr, YIQ, atď. Skúmané prahy sú používané napríklad pri odtieni, jase, v farebných súradniciach (jas a odtieň, alebo zeleno-červené a žlto-modré kanály), v normalizovaných červeno–zelených súradniciach, resp. vo všetkých troch.

Práca [36] využíva kombináciu RGB a HSV farebných modelov spolu so špecifickým prahom pre hodnoty r,g,H,S,V. Detekcia pokožky v tomto prípade slúžila ako predspracovanie pre odstraňovanie červených očí na fotografiách.

Farba pokožky môže byť jednoducho určená aj pomocou **histogramu** generovaného z obrazových bodov. Zvyčajne sa využíva 2D histogram. Tieto histogramy sú používané ako model na konvertovanie skúmaného obrazu na obraz s detegovanými oblasťami pokožky. Najpoužívanejšie farebné modely pri detekcii pokožky pomocou histogramov sú HSV, RGB, HSI, YUV. Niektoré prístupy používajú HSV model, kde je kontrolný histogram porovnávaný s histogramom pohybujúceho sa okna a pokiaľ je zhoda vyššia ako určitý prah, je táto časť obrazu detegovaná ako pokožka. Uvedené prístupy umožňujú detekciu tváre, rúk, ako aj sledovanie tváre.

Ďalšou metódou na detekciu pokožky je **unimodálny Gaussian**. Tento sa často využíva pri reprezentovaní farby pokožky buď pomocou jej strednej hodnoty, alebo použitím kovariančnej matice, resp. kombináciou oboch. Metóda súvisiaca s unimodálnym Gaussianom je tu *model eliptického ohraničenia*. Použitím nemenného prahu pravdepodobnosti sa získava eliptické okolie so stredom v strednej hodnote všetkých trénovaných farebných vektorov. Pomocou Gaussianu možno získať pravdepodobnosť výskytu pokožky pre každý pixel v obraze.

Kombinácia Gaussianov sa používa na lepšie modelovanie distribúcie farby pokožky, špeciálne v prípadoch osôb prislúchajúcich k rôznym etnickým skupinám. Detekcii pokožky pomocou kombinácie Gaussianov sa venuje veľké množstvo prác, kde využívajú rôzne farebné modely. Na lepšiu detekciu pokožky využívajú ďalšie techniky, napr. cross-validation, adaptívny učiaci sa algoritmus a ďalšie.

Okrem horeuvedených prístupov možno použiť na detekciu pokožky aj neurónové siete, samoorganizujúce sa mapy, SVM. Tieto techniky majú špecifické pravidlá na určovanie oblastí obsahujúcicj pokožku a neobsahujúcich. [35].

3.4 Východiská navrhovaného riešenia

Najdôležitejšou zložkou komunikácie sluchovo hendikepovaných ľudí je jej vizuálna časť. V prípade prenosu videa zaznamenávajúceho znakovú reč sú pre jeho zrozumiteľnosť najdôležitejšie oblasti, ktoré obsahujú ruky a tvár. Tieto možno v danom obraze naraz identifikovať metódami detekcie pokožky (pozri časť 3.3). Okrem rúk a tváre sa však v obraze nachádzajú aj iné významné oblasti, preto je vhodné kombinovať metódy ich detekcie s technikami rozpoznávania pokožky.

Navrhovaný model je modifikáciou základného modelu, ktorý na detekciu významných oblastí využíva lokálnu kontextovú informáciu a potláča sporné oblasti záujmu (pozri časť 3.2). Oproti pôvodnému je navyše rozšírený o sémantickú informáciu (pokožka), na detekciu ktorej využíva tri rôzne prístupy (pozri časť 4.3).

Kvalita navrhovaného riešenia sa meria pomocou špecálne upravených objektívnych

metrík ohodnocovania kvality videa, tieto modifikácie navrhla autorka. Nové metriky sú následne podrobené sérii testov a vyhodnotené vzhľadom na bežne používané metódy merania kvality videa (pozri časť 5.2).

Kapitola 4

Implementácia metódy

Základný model popísaný v predchádzajúcej kapitole bol naprogramovaný v prostredí MATLAB v rámci práce [20]. Tento bol modifikovaný, spresnený a zrýchlený približne trojnásobne. Namiesto detekcie tváre bola použitá detekcia pokožky, pričom boli na jej určenie využívané tri prístupy.

Zvolený model [14] je založený na tvorbe kontrastných máp pre farbu, intenzitu a textúru, poskytuje základ pre navrhnutý prístup, hoci je vhodný najmä pre detegovanie menších výrazných oblastí. V predchádzajucej práci [20] autorka použila bloky s veľkosťou 16×16 pixlov, v tejto boli použité bloky veľkosti 8×8 a 4×4 pixle. Použitie inej veľkosti blokov boli zvolené na základe veľkosti vstupu.

Na detekciu pokožky boli použité tri prístupy. Faktor potlačenia (SF) bol vytvorený pomocou kombinácie máp pre farbu, intenzitu, textúru a detekciu pokožky.

Vstupom pre testovanie navrhnutého modelu boli videosekvencie, zachytávajúce osobu komunikujúcu znakovou rečou. Každá videosekvencia obsahovala 298 snímok s veľkosťou 288×352 pixlov. Videosekvencie boli pomocou programu VirtualDub [25] rozdelené na jednotlivé snímky, ktorých rozmery boli upravené na rozmery 256×256 a následne boli spracovávané samostatne. Na každej snímke sa detegovali významné oblasti a výsledné snímky boli spojené do novej videosekvencie, ktoré slúžili ako vstup na porovnávanie.

Pri implementácii navrhnutého modelu boli použité niektoré prídavné kódy: 2D Gabor Filter [33], kód na spracovanie cirkulárnych dát [1], kovariančná matica [39].

4.1 Farba a intenzita

Prvým krokom pri spracovaní vstupnej snímky bola jej konverzia na obrázok v HSV farebnom modeli.

Farba a intenzita patria medzi základné zložky potrebné na vytvorenie mapy významných oblastí vstupného obrazu. Ich kontrastné mapy boli získané pomocou vzťahu 3.5. Pri tvorbe kontrastnej mapy pre farbu boli použité zložky HSV farebného modelu: odtieň H a saturácia S. Dáta pre odtieň sú cirkulárne, preto sa hodnoty pre štandardné odchýlky a stredné hodnoty počítali špecificky. Pre každú túto zložku bola vytvorená mapa stredných hodnôt. Následne boli tieto mapy spracované a sčítané, čím vznikla mapa príznakov pre farbu.

V pôvodnom modeli [14] bola na získanie intenzity použitá kombinácia $I = (R + G + B)/3$, kde R, G a B sú zložkami RGB farebného modelu, v predkladanej práci na bola určenie intenzity použitá Value zložka HSV modelu. Pre túto bola vytvorená mapa stredných hodnôt a následným aplikovaním vzťahu 3.5 vznikla výsledná mapa pre intenzitu.

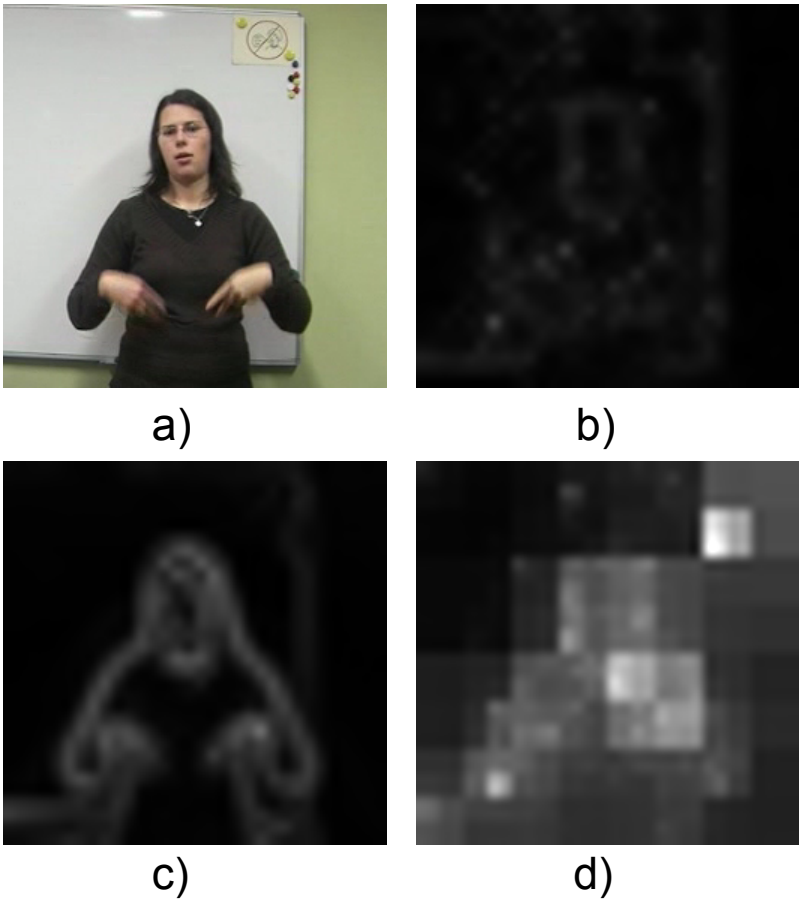
Pre extrakciu máp pre oba príznaky boli vytvorené samostatné funkcie. Na obrázku 4.1 b) je znázornená mapa pre farbu, na obrázku 4.1 c) je mapa pre intenzitu. Ako vstup bola použitá 188. snímka z videosekvencie.

4.2 Textúra

Postup získania mapy pre príznak textúry je podrobne popísaný v časti 3.2.1.

Pre každý kanál (H, S a V) boli získané a následne sčítané mapy textúrneho kontrastu TC. Funkcia na získanie TC pozostávala z vytvorenia 24 máp (6 orientácií x 4 veľkosti) a vypočítania stredných hodnôt a štandartných odchýlok pre každú z nich. Ďalším krokom bolo vytvorenie AMD a ASDD máp, kde pri ich tvorbe bolo použité osem-susedné okolie. Vynásobením príslušných máp pre AMD a ASSD vznikla množina 24 máp.

Pri tvorbe výslednej mapy (Obr. 4.1 d)) pre textúru boli najprv spočítané mapy s rovnakou orientáciou, čím vznikli 4 mapy rôznych veľkostí. Tieto mapy boli prepočítané na rovnakú veľkosť a sčítané, čím vznikla výsledná mapa pre H, S, alebo V.



Obr. 4.1: a) Vstupná snímka; Mapy príznakov pre b) farbu, c) intenzitu a d) textúru

4.3 Detekcia pokožky

V predkladanej práci je spracovávané video, ktoré zachytáva znakovú reč. Pri modeloch na detekciu významných častí obrazu je veľmi zložité používať sémantickú informáciu, keďže každý pozorovateľ je jedinečný a môžu ho zaujímať rôzne oblasti v obraze. Pri obraze zachytávajúcom znakovú reč je však veľmi pravdepodobné, že pozorovateľ bude zameriavať svoju pozornosť okrem oblastí s výraznou farebnou oblasťou a textúrou, aj na ruky a tvár osoby na videu. Boli zvolené tri metódy detekcie, pri ktorých sa porovnali výsledky a pomocou týchto zistení sa vybral najlepší model pre daný problém.

Prvé dve metódy sú založené na prahovaní, avšak na rôznej úrovni zložitosti. Tretia metóda využíva Gaussian.

4.3.1 Prahovanie

Prahovanie predstavuje najjednoduchší prístup. Bolo aplikované obraz reprezentovaný v HSV farebnom modeli, ktorý je kompatibilný s ľudským vnímaním farieb. Tento model je hexagonálny, kde odtieň H je reprezentovaný uhlom, saturácia farby S je definovaná hodnotou v rozmedzí od 0 po 1 a intenzita farby je definovaná hodnotou V. Pomocou kombinácie odtieňa H a saturácie S farby možno definovať farbu pokožky. Pri použitom prahovaní boli definované naslednove:

$$S_{min} = 0.23, S_{max} = 0.68;$$
$$H_{min} = 0°, H_{max} = 50°.$$

Uvedené hodnoty boli vybrané na základe práce [37]. Po aplikovaní prahov boli na výsledné oblasti aplikované morfologické operácie (dilatácia, erózia), čím vznikli hladšie a ucelenejšie oblasti. Na obrázku (Obr. 4.2 b)) sú znázornené výsledky pre detektor prahovania.

4.3.2 Kombinácia viacerých farebných modelov

V súčasnosti neexistuje jednotný prístup na určovanie optimálneho prahu pre RGB a HSV priestory tak, aby sa plne využili výhody oboch modelov a súčasne sa znížil i vplyv osvetlenia. Kombinácou oboch modelov sa získajú veľmi dobré výsledky pri detekcii pokožky v rôznych svetelných podmienkach s malým množstvom chybne rozpoznaných častí obrazu.

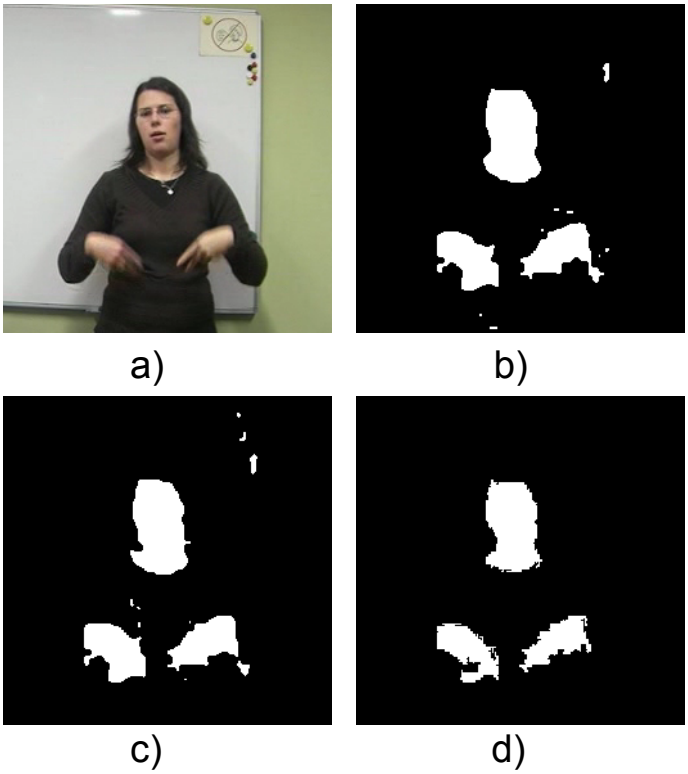
V tomto prístupe bola použitá kombinácia výstupov prahovania v oboch farebných priestoroch (RGB, HSV), pričom hodnoty r a g boli normalizované. Pri tomto prístupe boli hodnoty pre potrebné zložky definované naslednove:

$$r \in [35, 55], g \in [25, 28];$$
$$H \in [0, 50] \bigcup [340, 360];$$
$$S > 0,2; V > 0,35.$$

Táto metóda bola pôvodne použitá ako predspracovanie pre odstraňovanie červených očí [36]. V predkladanej práci je tento prístup použitý ako jedna z troch metód na detekciu pokožky (Obr. 4.2 c)).

4.3.3 Detekcia pokožky na základe Gaussianu

Predchádzajúce dva spôsoby detekcie pokožky boli založené na prahovaní. Tento spôsob využíva Gaussian. Takto možno získať pravdepodobnosť výskytu pokožky pre každý pixel v obraze, využíva sa detekcia pokožky použitá v práci [35]. Výsledok detekcie je zobrazený na Obr. 4.2 d).



Obr. 4.2: a) Vstupná snímka; Detekcia pokožky použitím b) prahovania, c) kombinácie viacerých farebných modelov a d) Gaussianu

4.4 Kombinácia príznakov a tvorba faktora potlačenia

Na získanie výslednej mapy významných oblastí bola potrebná kombinácia všetkých použitých príznakov. Kombinácia konečného počtu príznakov je veľmi zložitý problém. Niektoré modely využívajú jednoduchú lineárnu kombináciu príznakov, iné navrhujú váhovanú kombináciu, postprocessing. Výsledkom týchto kombinácii je mapa významných oblastí, avšak

táto obsahuje aj nepravo významné oblasti, ktoré sú nežiadúce. V predkladanej práci bola preto využitá modifikovaná kombinácia príznakov navrhnutá v [14].

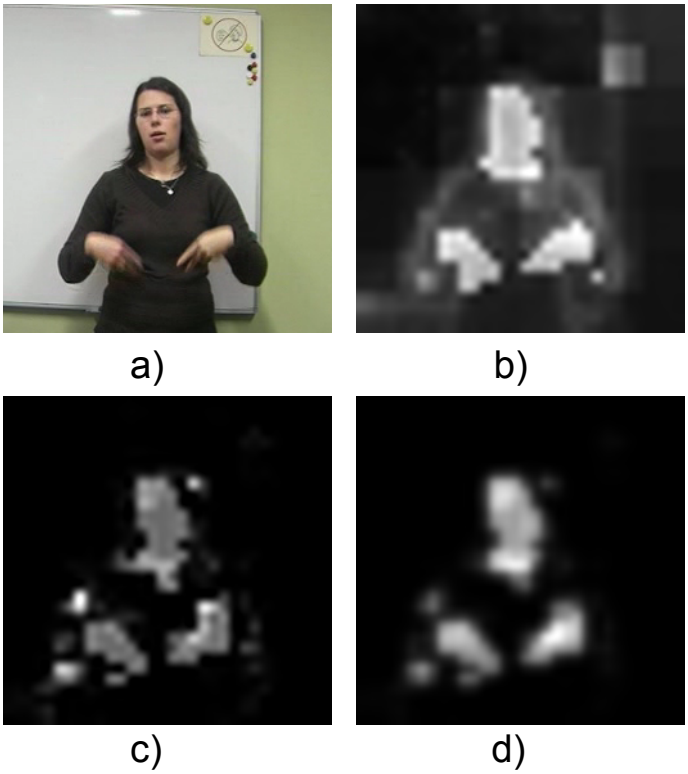
Vstupom pre túto kombináciu boli mapy kontrastov pre štyri príznaky: farba, intenzita, textúra a detekciu pokožky. Mapy kontrastov pre intenzitu, farbu a textúru získané na rovnakom princípe ako v [14] sú popísané v častiach 4.1 a 4.2. Mapa detekcie pokožky bola získaná vždy jedným z troch vyššie popísaných detektorov pokožky.

Prvým krokom kombinácie príznakov bolo spočítanie a normalizácia máp pre farbu, intenzitu, textúru a detekciu pokožky, čím vznikla mapa kombinácií (Obr. 4.3 b)). Táto mapa obsahuje výrazné oblasti, avšak niektoré z nich sú nesprávne označené ako výrazné. Na rozpoznanie pravých významných oblastí a potlačenie tých ktoré sú nesprávne označené ako výrazné bol navrhnutý faktor potlačenia (angl. suppression factor, ozn. SF).

Použitím faktora potlačenia bola vytvorená *mapa potlačenia*, ktorá sa skladá z máp pre farbu, intenzitu a textúru. Mapa potlačenia pozostáva z hodnôt z intervalu $[0, 1]$. Čím je hodnota vyššia, tým je faktor potlačenia nižší a naopak. To znamená, že svetlejšie oblasti sú významnejšie ako tmavé oblasti.

Ďalší krok bol kombinácia mapy potlačenia s mapou pre detekciu pokožky. Týmto sa zvýraznili významné oblasti v obraze s rovnakou prioritou, ako časti detegované ako pokožka. Pri kombinácii týchto máp sa použila podmienená kombinácia. Týmto prístupom sa docielilo to, že významné oblasti boli dostatočne výrazné a oblasti obsahujúce pokožku sa taktiež zvýraznili (Obr. 4.3 c)). Hodnota výsledného pixla mapy potlačenia bola určená ako minimum hodnôt súčtu sčítavaných pixlov a pôvodnej hodnoty pixla.

Posledným krokom pri tvorbe mapy významných oblastí bola kombinácia mapy potlačenia a mapy kombinácii, ktorá spočívala vo vzájomnom vynásobení týchto máp. Výsledná mapa významných oblastí (Obr. 4.3 d)) takto obsahuje významné oblasti i oblasti obsahujúce pokožku, pričom potláča nepravé oblasti záujmu.



Obr. 4.3: a) Vstupná snímka; b) Mapa kombinácií, c) Mapa pre faktor potlačenia a d) Výsledná mapa významných oblastí

Kapitola 5

Validácia a výsledky

Predkladaná práca slúži ako základ pre ďalší výskum, ktorý sa zameriava na detekciu sémanticky významných častí obrazu a ich použitie pri jeho kódovaní. Z tohto dôvodu sa autorka rozhodla pre špecifické testovanie výsledných modelov. Pomocou porovnávaných modelov boli upravené viaceré bežne používané metriky na ohodnocovanie kvality videa. Porovnaním týchto metrík sa získala informácia o relevantnosti daných modelov pre ďalšie využitie a následne bol identifikovaný model a taktiež metrika, ktoré dosahovali najlepšie výsledky.

Na testovanie a porovnanie použitých prístupov boli použité videosekvencie zachytávajúce osobu komunikujúcu posunkovou rečou. K dispozícii boli videosekvencie v pôvodnej a zníženej kvalite. Porovnávaním pôvodnej videosekvencie so snímkami so zníženou kvalitou boli získané potrebné dáta, ktoré slúžili ako základ pre porovnanie daných prístupov.

5.1 Úprava dát pre porovnávanie

5.1.1 Rôzne prístupy na detekciu významných oblastí

Prvým krokom pri testovaní výsledkov bolo rozdelenie videosekvencie pôvodnej kvality na jednotlivé snímky. Toto rozdelenie umožnil program VirtualDub [25].

Ďalším krokom bola detekcia významných oblastí v snímkach, pričom bolo použitých viacero prístupov. Prvý bol založený na Ittiho návrhu [12]. Ďalšie sa zakladali na modeli

[14] a jeho modifikáciách, ktoré sú popísané v časti 3.2. Modifikácie boli založené na použití rôznych detektorov pokožky a taktiež rôznych veľkostí blokov použitých pri detekcii významných oblastí.
Výsledné modifikácie sú:

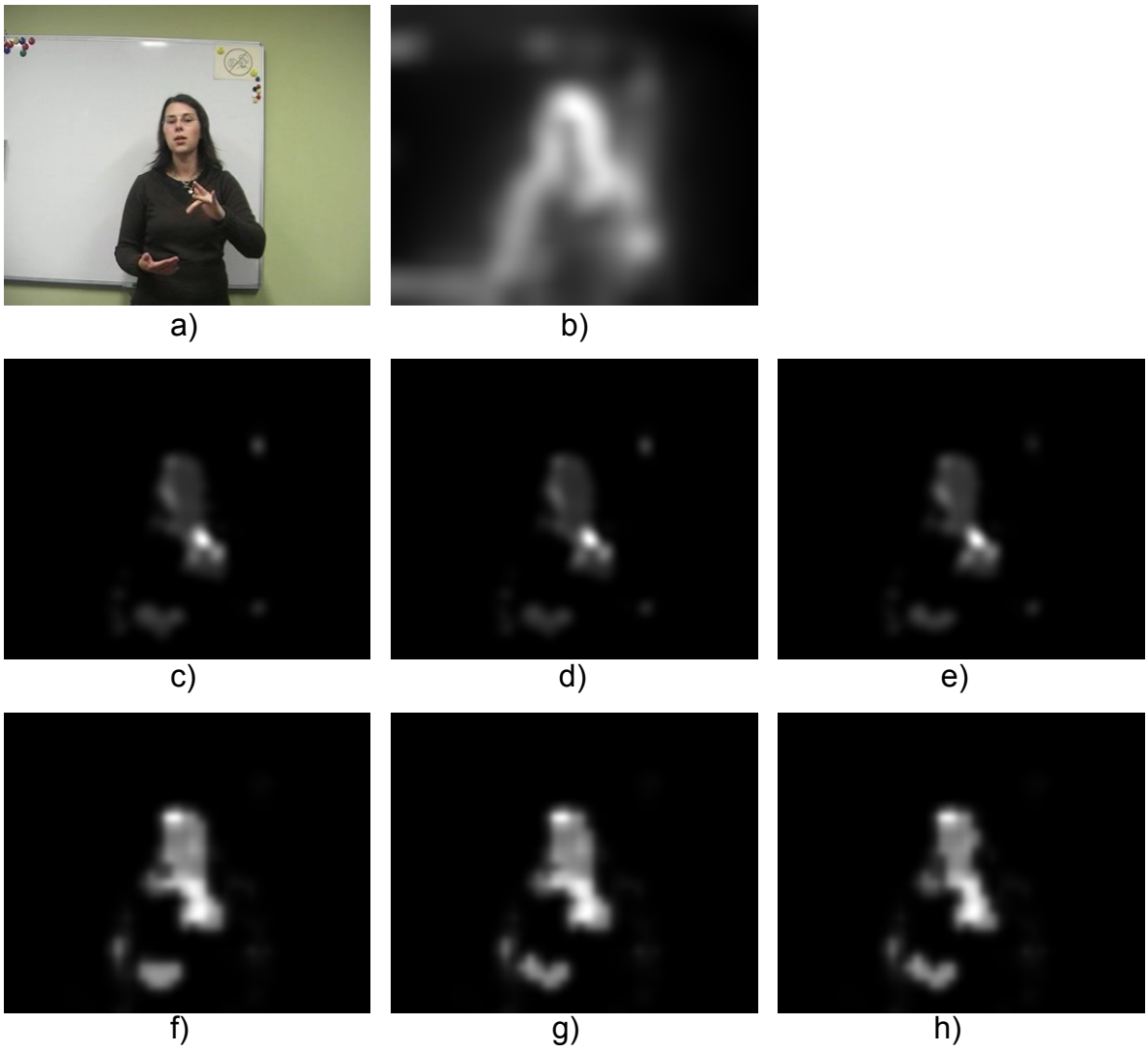
- $4P$: Model využívajúci bloky veľkosti 4x4 s použitím detekcie pokožky založenej na prahovaní.
- $4K$: Model využívajúci bloky veľkosti 4x4 s použitím detekcie založenej na prahovaní s využitím kombinácie RGB a HSV farebných modelov.
- $4G$: Model využívajúci bloky veľkosti 4x4 s detekciou využívajúcou Gaussian.
- $8P$: Model využívajúci bloky veľkosti 8x8 s použitím detekcie pokožky založenej na prahovaní.
- $8K$: Model využívajúci bloky veľkosti 8x8 s použitím detekcie založenej na prahovaní s použitím kombinácie RGB a HSV farebných modelov.
- $8G$: Model využívajúci bloky veľkosti 8x8 s detekciou využívajúcou Gaussian.

Pri porovnávaní boli využité pôvodné snímky a snímky modifikované každým z vyššie spomenutých prístupov.

Na obrázku 5.1 je zobrazená 210. snímka z videosekvencie a príslušné mapy významných oblastí získané všetkými porovnávanými prístupmi. Niektoré z máp sú na pohľad veľmi podobné, tento fakt však neovplyvňuje výsledky a aj na pohľad podobné mapy dávajú pri testovaní rôzne výsledky.

5.1.2 Kombinácie

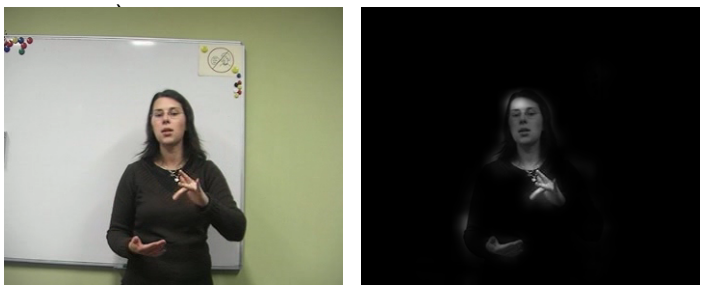
Kvôli testovaniu a návrhu nových metrík bolo nutné na pôvodné video aplikovať masky tvorené mapami významných oblastí. Mapa významných oblastí pozostáva z hodnôt intervalu $[0, 1]$, pričom vyššia hodnota označuje významnejšiu oblasť.



Obr. 5.1: a) Pôvodná snímka; Mapy významných oblastí pre b) Ittiho prístup [16], c) 4P d) 4K e) 4G f) 8P g) 8K h) 8G

Snímky z videa sa postupne kombinovali s mapami významných oblastí. Pôvodné snímky sa konvertovali do farebného modelu YUV a Y zložka sa násobila s mapou významných oblastí.

Na obrázku 5.2 je znázornená kombinácia 210. snímky z pôvodného videa s mapou významných oblastí získanou prístupom označeným $4P$.
Rovnakým spôsobom sa kombinovali snímky z príslušných videí rôznej kvality (Obr. 5.3). Rozdielne úrovne kvality videa vznikli modifikovaním pôvodného videa použitím kódera H.264 s rôznymi bodovými hodnotami (QP: 30, 35, 40, 45, 50, čo zodpovedá hodnotám od 87 do 18 kbps) [21].



Obr. 5.2: Snímka č. 210. z videosekvencie: a) Pôvodná snímka b) snímka kombinovaná s mapou významných oblastí získanou prístupom $4P$

5.2 Objektívne metriky na ohodnocovanie kvality videa

Na porovnanie modifikácií pôvodného modelu boli použité rôzne metriky na ohodnocovanie kvality videa. Nové metriky vznilkli kombináciou výsledkov z porovnávaných modifikácií s rôznymi metrikami.

Vo všeobecnosti je pri vyhodnocovaní novej metriky potrebné riadiť sa podmienkami ITU [44]. Nová metrika by sa mala čo najviac tvarovo zhodovať so subjektívnym hodnotením.

V predkladanej práci bolo namiesto subjektívneho hodnotenia použité hodnotenie zrozumiteľnosti respondentmi. Každá videosekvencia bola ukázaná skupine 14 sluchovo hen-

dikepovaných pozorovateľov. Každý pozorovateľ určil časti videa, ktorým nerozumel. Nakoniec jeden nezávislý pozorovateľ ohodnotil zrozumiteľnosť videa.

Obr. 5.3: Snímka č. 210 videosekvencie: a) Pôvodná kvalitaô snímky dekódované pomocou H.264 s rôznou kvalitou b) QP = 30 c) QP = 35 d) QP = 40 e) QP = 45 f) QP = 50

Zrozumiteľnosť jazyka Z môže byť definovaná percentuálne ako pomer správne prijatých prvkov, alebo častí reči a a ich celkového počtu b: $Z = \frac{a}{b}.100\%$.
Použitím uvedeného vzťahu sa zisťuje zrozumiteľnosť oblastí vo videu.

V tabuľke 5.1 sú zobrazené namerané hodnoty pre zrozumiteľnosť jednotlivých testovaných sekvencií, kde hodnota 5 označuje najhoršiu zrozumiteľnosť a 1 najlepšiu. Použité hodnotenia zrozumiteľnosti videa a testovacie videosekvencie boli získané v spolupráci s Pedagogickou Fakultou UK.

Pre každú hodnotu QP bola vypočítaná priemerná zrozumiteľnosť, ktorá slúžila ako referenčné kritérium pre objektívne metódy použitím Pearsonového korelačného koeficientu. Tento koeficient je najčastejšie využívaný a vyjadruje presnosť s akou porovnávaná metrika ohodnotila kvalitu vzhľadom na použité subjektívne hodnotenie. Absolútna hodnota

QP	Zrozumiteľnosť
30	1,17
35	1,17
40	1,67
45	1,83
50	3

Tabuľka 5.1: Zrozumiteľnosť videosekvencií pri rôznych kvalitách videa

korelačného koeficientu nadobúda hodnoty z intervalu $[0, 1]$, pričom hodnota 1 vyjadruje maximálnu zhodu medzi porovnávanou metrikou a hodnotením respondentmi.

5.2.1 Použité metriky

V predkladanej práci boli porovnané rôzne metriky na ohodnocovanie kvality videa. Množinu týchto metrík tvorí PSNR, SSIM, 3SSIM, VQM a MSE, ich voľba bola ovplyvnená rozšírenosťou používania.

PSNR (Peak signal-to-noise ratio) metrika patrí medzi objektívne metódy bez triedenia obrazových dát. PSNR meria, ako sa obraz so zníženou kvalitou podobá nepoškodenému originálu.

SSIM (Structural Similarity Index) metrika patrí medzi objektívne metódy s kategorizáciou dát. Vyhodnocuje vizuálny dopad posunu jasu v obraze, zmien štruktúry a zmien v kontraste. Je založená na predpoklade, že ľudský vizuálny systém je prispôsobený na výber štrukturálnych informácií zo scény. Toto meranie štrukturálnych informácií by malo zaručiť lepšiu koreláciu so subjektívnym vnímaním.

3SSIM (3 Structural Similarity Index) je ďalšou metrikou zo skupiny objektívnych metód s kategorizáciou pixlov. Zakladá sa na rozpoznávaní oblastí záujmu (Region of Interest ROI), čo napomáha pri odhadovaní kvality videa. Výpočet tejto metriky sa skladá zo štyroch krokov. Najprv sa vypočíta mapa metriky SSIM, následne je pôvodný obraz rozdelený do troch oblastí: hladké oblasti, textúry a hrany. Na body prislúchajúce jednotlivým oblastiam sú aplikované váhy. Nakoniec sa tieto váhy sčítajú a výsledky sa spriemerujú.

VQM (Video Quality Metric) patrí medzi objektívne metódy s jednoduchým triedením

obrazových dát. Využíva vlastnosti ľudského vnímania obrazu. Je založená na funkcii priestorovo–časovej citlivosti kontrastu, kde je možné reprezentovať informáciu s menšou presnosťou, keďže ľudské oko nie je citlivé na stratu informácie.

Poslednou porovnávanou metrikou je **MSE** (Mean squared error), ktorá je jednou z najrozšírenejších metrík pri spracovaní obrazu a videa. Vyjadruje kvadratický rozdiel hodnôt jasových úrovní pixlov medzi dvoma obrazmi alebo sekvenciami [26].

5.3 Porovnávanie vytvorených metrík

Na porovnanie spomenutých metrík bol použitý program MSU VQMT [42]. Vstupom pre porovnávanie boli dve videá, pričom prvé bolo nekomprimované video a druhé s rôznou úrovňou kompresie. Pre porovnávanie videí pôvodnými metrikami boli ako vstup použité pôvodné videá. Použitím kombinácie masiek vytvorených z máp významných oblastí a už existujúcich metrík boli vytvorené nové metriky na ohodnocovanie kvality videa. Kombinácia masiek a pôvodných videí je popísaná v časti 3.2.1.

Pre nové metriky boli zvolené označenia kombináciou označení pre použité prístupy s označeniami pre použité metriky, napr. $4P$-$PSNR$, $4K$-$PSNR$, atď. Metriky s použitím Ittiho prístupu sú označované ako I-$PSNR$, I-$SSIM$, I-VQM a I-MSE.

Výsledné hodnoty Pearsonovho korelačného koeficientu pre nové metriky sú zobrazené v tabuľke 5.2. Porovnávané metriky sú v tabuľke zoradené podľa úspešnosti od najlepšej. Pomocou analýzy výsledných hodnôt korelačného koeficientu pre porovnávané metriky na daných testovacích dátach je možné určiť prístup detekcie významných oblastí, ktorý dosahoval najlepšie výsledky.

Podľa výsledného poradia metrík sa javí ako najlepší prístup detekovania významných oblastí model s použitím veľkosti blokov 4x4. Metriky využívajúce prístupy s takouto veľkosťou blokov obsadili, až na dve, prvých 10 miest.

Na prvých troch miestach sa umiestnili prístupy využívajúce rovnakú základnú metriku a všetky tri detektory pokožky. Tento fakt je spôsobený tým, že detegované oblasti pokožky boli veľmi podobné, a teda neovplyvňovali až do takej veľkej miery výsledné mapy významných oblastí. Podľa tohto zistenia môžme vysloviť tvrdenie, že pre nasledujúcu

Objektívna metrika	**Pearsonov korelačný koeficient**	**Poradie**	**Objektívna metrika**	**Pearsonov korelačný koeficient**	**Poradie**
4K-MSE	0,98309	1	8K-SSIM	0,97641	13
4P-MSE	0,98277	2	8P-SSIM	0,97640	14
4G-MSE	0,98241	3	8G-SSIM	0,97630	15
8G-MSE	0,97867	4	4K-VQM	0,97558	16
8P-MSE	0,97858	5	8P-VQM	0,97448	17
4K-SSIM	0,97822	6	8G-VQM	0,97427	18
4P-SSIM	0,97818	7	4P-PSNR	0,91133	19
4G-SSIM	0,97792	8	4G-PSNR	0,91124	20
4P-VQM	0,97765	9	4K-PSNR	0,91039	21
4G-VQM	0,97744	10	8P-PSNR	0,90965	22
8K-VQM	0,97737	11	8K-PSNR	0,90912	23
8K-MSE	0,97732	12	8G-PSNR	0,90906	24

Tabuľka 5.2: Porovnanie nových metrík na určovanie kvality videa zoradených od najlepšej

prácu je najvhodnejšie vybrať z týchto detekcií výpočtovo najmenej náročnú.

Prvých 5 miest v tabuľke 5.2 obsadili nové metriky, využívajúce bloky veľkosti 4x4, ktorých základ bola metrika MSE.

Pre overenie relevantnosti výberu základného modelu [14] a jeho modifikácií boli porovnané výsledné metriky s metrikami založenými na Ittiho [16] prístupe. Tabuľka 5.3 zobrazuje päť najlepších metrík vybraných na základe porovnaní v tabuľke 5.2, najlepšie modifikácie všetkých metrík a metriky vytvorené s použitím Ittiho prístupu. Metriky sú zoradené podľa úspešnosti a až na $4P\text{-}PSNR$ dávajú metriky využívajúce navrhnutý prístup lepšie výsledky.

V tabuľke 5.4 sú porovnávané nové metriky s už existujúcimi metrikami, ktoré sú zoradené podľa úspešnosti. V tabuľke je zobrazených prvých 5 najúspešnejších metrík a

Objektívna metrika	**Pearsonov korelačný koeficient**	**Poradie**
4K-MSE	0,98309	1
4P-MSE	0,98277	2
4G-MSE	0,98241	3
8G-MSE	0,97867	4
8P-MSE	0,97858	5
4K-SSIM	0,97822	6
4P-VQM	0,97765	7
I-VQM	0,97737	8
I-SSIM	0,97304	9
I-MSE	0,96672	10
I-PSNR	0,92090	11
4P-PSNR	0,91133	12

Tabuľka 5.3: Porovnanie nových metrík na určovanie kvality videa s metrokami založenými na Ittiho prístupe

najúspešnejšie modifikácie všetkých porovnávaných metrík. Okrem metriky $4P - PSNR$ dosahujú všetky navrhnuté metriky lepšie výsledky ako už existujúce metriky.

Najlepšie výsledky dosiahla metrika $4K\text{-}MSE$, ktorá bola vytvorená modifikáciou metriky MSE. MSE, ktorá patrí medzi najpoužívanejšie, obsadila posledné miesto. Tento výsledok naznačuje, že navrhovaným prístupom sa podarilo pre dané dáta veľmi výrazne vylepšiť jednu z najpoužívanejších metrík.

Použitím detekcie významných oblastí bol obraz rozdelený na 256 úrovní, ktoré určujú významnosť každého bodu v obraze. Zvyšovanie počtu váh na pixel dáva lepšie výsledky ako použitie troch váh pri metrike 3SSIM, ktorá využíva podobné delenie.

V tejto kapitole boli porovnávané navrhnuté modifikácie základného modelu [14]. Ako

Objektívna metrika	**Pearsonov korelačný koeficient**	**Poradie**
4K-MSE	0,98309	1
4P-MSE	0,98277	2
4G-MSE	0,98241	3
8G-MSE	0,97867	4
8P-MSE	0,97858	5
4K-SSIM	0,97822	6
4P-VQM	0,97765	7
VQM	0,96960	8
3SSIM	0,96546	9
SSIM	0,95879	10
PSNR	0,91370	11
4P-PSNR	0,91133	12
MSE	0,77740	13

Tabuľka 5.4: Porovnanie nových metrík s už existujúcimi metrikami

vstup slúžilo video zachytávajúce osobu komunikujúcu posunkovou rečou. Pri takomto type videa je veľmi dôležité zachytenie oblastí, ktoré obsahujú ruky a tvár, predstavujúce základ pri komunikácii posunkovou rečou. Preto bol základný model modifikovaný pridaním príznaku detekcie pokožky, čo predstavuje využitie sémantickej informácie - vedomosti, že v danom obraze sú dôležité ruky a tvár. Takto modifikovaný model je schopný zachytávať významné oblasti v obraze rovnako, ako aj ruky a tvár.

Pridaním detekcie významných oblastí boli vytvorené nové metriky na ohodnocovanie kvality videa. Tie boli porovnávané s už existujúcimi metrikami. Metriky využívajúce navrhnuté modifikácie modelu na detekciu významných oblastí dosahujú pre dané testovacie dáta výrazne lepšie výsledky, ako súčasne využívané metriky.

Záver

Využitie sémantickej informácie pri určovaní významných oblastí v obraze je veľmi zložitý problém a v súčasnosti nie je možné určiť jednotnú sémantickú informáciu pre všetky možné typy videa a obrazu. Preto i predložená rigorózna práca využíva na detekciu významných oblastí v obraze konkrétnú sémantickú informáci, a to detekciu pokožky.

Predložená práca nachádza praktické uplatnenie v komunikácii sluchovo hendikepovaných ľudí prostredníctvom videa. Pretože primárnym dorozumievacím prostriedkom týchto osôb je znaková reč, je nesmierne dôležité, aby video so znakovou rečou bolo dostatočne zrozumiteľné. Najdôležitejšou časťou obrazu sú v tomto prípade ruky a tvár.

Na základe predpokladov pre zrozumiteľnosť videa sluchovo hendikepovaných bolo vytvorených niekoľko modifikácií pôvodného modelu na detekciu významných oblastí v obraze, čím vzniklo viacero prístupov. Tieto boli následne porovnané a vyhodnotené na základe úspešnosti. Pomocou daných modifikácií boli vytvorené a ohodnotené nové metriky na ohodnocovanie kvality videa, ktoré boli navzájom porovnávané.

Dosiahnuté výsledky predloženej práce naznačujú, že využitie detekcie významných oblastí s použitím sémantickej informácie pri ohodnocovaní kvality videa zachytávajúcom ľudí komunikujúcich posunkovou rečou môže byť v budúcnosti veľmi efektívne pre komunikáciu nielen sluchovo hendikepovaných, ale i celej populácie.

Literatúra

[1] P. Berens and M. Velasco. Circstat: A matlab toolbox for circular statistics. [Online] `http://www.jstatsoft.org/v31/i10`, oktober 2009.

[2] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(PrePrints), 2012.

[3] D. E. Broadbent. *Perception and Communication*. Pergamon Press, 1958.

[4] N. D. B. Bruce. *Saliency, attention and visual search: an information theoretic approach*. PhD thesis, Canada, 2008. AAINR45988.

[5] A. Bur and H. Hügli. Optimal cue combination for saliency computation: A comparison with human vision. In *Proceedings of the 2nd international work-conference on Nature Inspired Problem-Solving Methods in Knowledge Engineering: Interplay Between Natural and Artificial Computation, Part II*, IWINAC '07, pages 109–118, Berlin, Heidelberg, 2007. Springer-Verlag.

[6] K. Cater, A. Chalmers, and G. Ward. Detail to attention: Exploiting visual tasks for selective rendering. In *Eurographics Symposium on Rendering*, pages 270–280. Eurographics, 2003.

[7] A. Chalmers, K. Cater, and D. Maflioli. Visual attention models for producing high fidelity graphics efficiently. In *Proceedings of the 19th spring conference on Computer graphics*, SCCG '03, pages 39–45, New York, NY, USA, 2003. ACM.

[8] J. A. Deutsch and D. Deutsch. Attention: Some theoretical considerations. *Psychological Review*, 70:80–90, 1963.

[9] A. Duchowski. Eye-based interaction in graphical systems. In *SIGGRAPH*. 2000.

[10] D. Gao, V. Mahadevan, and N. Vasconcelos. On the plausibility of the discriminant centersurround hypothesis for visual saliency. *Journal of Vision*, pages 1–18, 2008.

[11] E. Bruce Goldstein. *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*. Wadsworth Publishing, 2007.

[12] J. Harel. A saliency implementation in matlab. [Online] `http://www.klab.caltech.edu/~harel/share/gbvs.php`, august 2010.

[13] H. Von Helmholtz. *Handbuch der Physiologischen Optik (Treatise on Physiological Optics), Translated from the Third German ed.* The Optical Society of America, Rochester, 1925.

[14] Y. Hu, D. Rajan, and L. Chia. Adaptive local context suppression of multiple cues for salient visual attention detection. In *IEEE International Conference on Multimedia and Expo*, pages 1–4, 2005.

[15] Y. Hu, X. Xie, W. Ma, L. Chia, and D. Rajan. Salient region detection using weighted feature maps based on the human visual attention model. In *In Proceedings of the Fifth IEEE Pacifc-Rim Conference on Multimedia*, November 2004.

[16] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[17] W. James. *The Principles of Psychology, vol. I of The Works of William James*. Harvard University Press, 1981.

[18] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *IEEE International Conference on Computer Vision (ICCV)*, 2009.

[19] S. M. Kosslyn. *Image and Brain*. The MIT Press, 1994.

[20] J. Kucerova. Visual attention models. Master's thesis, Comenius University, Bratislava, Slovakia, April 2011.

[21] J. Kucerova, J. Polec, and D. Tarcsiova. Video quality assessment using visual attention approach for sign language. volume 65, pages 194–199. World Academy of Science, Engineering and Technology, 2012.

[22] S. Langton, A. Law, M. Burton, and S. Schweinberger. Attention capture by faces. *Cognition*, 107:330–342, 2008.

[23] O. Le Meur, P. Le Callet, and D. Barba. Predicting visual fixations on video based on low-level visual features. *Vision Res*, 2007.

[24] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(5):802–817, May 2006.

[25] A. Lee. Virtualdub. [Online] `http://www.virtualdub.org/`, december 2011.

[26] M. Mardiak. *Kvalitativne charakteristiky videa.* PhD thesis, Slovak University of Technology, Bratislava, Slovakia, September 2012.

[27] O. Le Meur and P. Le Callet. What we see is most likely to be what matters: Visual attention and applications. *ICIP 2009*, pages 3085–3088, 2009.

[28] neznamy autor. Semantics. [Online] `http://www.alleydog.com/glossary/definition.php?term=Semantics`, august 2012.

[29] D. Noton and L. Stark. Eye movements and visual perception. *Scientific American*, pages 34–43, 1971.

[30] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 42:145–175, 2001.

[31] A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson. Top-down control of visual attention in object detection. In *Proc. of the IEEE Int'l Conference on Image Processing (ICIP '03)*, 2003.

[32] M. Posner, Ch. Snyder, and B. Davison. Attention and the detection of signals. *Experimental Psychology: General*, 109:160–174, 1980.

[33] A. Poursaberi. 2d gabor filter(ver1,2,3). [Online] http://www.mathworks.com/matlabcentral/fileexchange/5237-2d-gabor-filterver123, november 2010.

[34] R. Rosenholtz. A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, 39:3157–3163, 1999.

[35] E. Sikudova. Comparison of color spaces for face detection in digitized paintings. In *Spring Conference on Computer Graphics : SCCG 2007 : Conference Proceedings*, pages 135–140, 2007.

[36] B. Smolka, K. Hardeberg, J. Plataniotis, M. Szczepanski, and K. Wojciechowski. Towards automatic red eye effect removal. *Pattern Recognition Letters*, 24:1767–1785, 2003.

[37] K. Sobottka and I. Pitas. Face localization and facial feature extraction based on shape and color information. In *IEEE International Conference on Image Processing*, volume 3, pages 483–486, 2005.

[38] F. Stentiford. A visual attention estimator applied to image subject enhancement and colour and grey level compression. In *International conference on Pattern Recognition (ICPR(3))*, pages 638–641. IEEE, 2004.

[39] K. Teknomo. Covariance matrix. [Online] http://www.mathworks.com/matlabcentral/fileexchange/29256-covariance-matrix, oktober 2010.

[40] A. Treisman. Features and objects in visual processing. *Scientific American 255*, pages 114B–125, 1986.

[41] A. Treisman and G. Gelade. A feature integration theory of attention. *Cognitive Psychology 12*, pages 97–136, 1980.

[42] D. Vatolin, A. Moskvin, O. Petrov, S. Putilin, S. Grishin, and A. Marat. Msu video quality measurement tool. [Online] http://www.compression.ru/video/quality_measure/video_measurement_tool_en.html, august 2012.

[43] P. Vrabel. Čo je semantika. [Online] `http://www.hugomedia.sk/co-je-semantika`, aaugust 2012.

[44] A. Webster. *Objective perceptual assessment of video quality: Full reference television.* ITU, 2004.

[45] J. Wolfe. *Visual attention.* CA Academic Press, 2000.

[46] L. Zhang, M. Tong, T. Marks, H. Shan, and G. Cottrell. Sun: A bayesian framework for saliency using natural statistics. *J Vis*, 8(7):32.1–20, 2008.