

Ing. Marek Vančo

Autoreferát dizertačnej práce

ROZPOZNÁVANIE GEST PRE OVLÁDANIE MULTIMEDIÁLNYCH ZARIADENÍ

na získanie akademickej hodnosti doktor (philosophiae doctor, PhD.)

v doktorandskom študijnom programe: **Telekomunikácie**

v študijnom odbore 5.2.15 Telekomunikácie

Miesto a dátum: Bratislava, 20.2.2017

**SLOVENSKÁ TECHNICKÁ UNIVERZITA
V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY**

Ing. Marek Vančo

Autoreferát dizertačnej práce

ROZPOZNÁVANIE GEST PRE OVLÁDANIE MULTIMEDIÁLNYCH ZARIADENÍ

na získanie akademickej hodnosti doktor (philosophiae doctor, PhD.)

v doktorandskom študijnom programe:

Telekomunikácie

Miesto a dátum: Bratislava, 20.2.2017

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia

Na Ústave multimediálnych informačných a komunikačných technológií,
Fakulta elektrotechniky a informatiky STU v Bratislave

Predkladateľ: Marek Vančo, Ilkovičova 3, 812 19 Bratislava

Školiteľ: prof. Ing. Gregor Rozinaj, PhD.

Oponenti: prof. Ing. Jozef Juhár, CSc. , Technická univerzita v Košiciach
doc. Ing. Jarina Roman, PhD., Žilinská univerzita

Autoreferát bol rozoslaný:

Obhajoba dizertačnej práce sa koná: o h.

Na

ANOTÁCIA

Slovenská technická univerzita v Bratislave

Fakulta elektrotechniky a informatiky

Ústav multimediálnych, informačných a komunikačných technológií

Študijný odbor: Telekomunikácie

Autor: Ing. Marek Vančo

Dizertačná práca: Rozpoznávanie gest

Školiteľ: prof. Ing. Gregor Rozinaj, PhD.

február 2017

Práca sa zaoberá problematikou návrhu algoritmov pre rozpoznávanie statických a dynamických gest ruky.

Hlavným cieľom je navrhnuť rozpoznávanie gest ruky pomocou snímania infračerveným senzorom. Samotný návrh spočíva vo vytvorení algoritmov rozpoznávania gest ruky tak, aby mohli byť aplikované v rámci multimodálneho rozhrania a multimediálnej TV. Na základe analýzy bola vytvorená metodika pre rozpoznávanie gest ruky a navrhnuté algoritmy, ktoré umožnili rozpoznávanie gest pre multimediálnu TV, ako je pohyb v menu, prepínanie kanálov, či zamykanie / odmykanie personalizovaného obsahu. Pri návrhu bola zohľadnená jednoduchosť implementácie a nízke hardvérové požiadavky. Pre rozpoznávanie statických gest ruky navrhnutý a overený postup detekcie spojených prstov. Taktiež bol navrhnutý systém intuitívnych gest. Navrhnuté algoritmy boli otestované a vyhodnotená ich úspešnosť. Z výsledkov vyhodnotenia bola vytvorená metodika pre používanie navrhnutých algoritmov v spojení s multimediálnou TV a pre jednotlivé navrhnuté princípy bolo definované ich použitie s ohľadom na funkčnosť.

ANNOTATION

Slovak University of Technology in Bratislava

Faculty of Electrical Engineering and Information Technology

Institute of Multimedia Information and Communication Technologies

Degree Course: Telecommunications

Author: Ing. Marek Vančo

Dissertation thesis: Gesture recognition for multimedia device control

Supervisor: prof. Ing. Gregor Rozinaj, PhD.

2017, February

The work deals with design of static and dynamic hand gesture recognition.

The main objective is to design gesture recognition application using infrared sensing. The design consists of creating of hand gesture recognition algorithms that can be applied within multimodal interface and multimedia TV. Based on analysis is needed to create methodics for hand gesture recognition and design algorithms that allow gesture recognition within multimedia TV such as menu navigation, channel control, lock / unlock personalized content. Within design it is necessary to consider simplicity of implementation and low hardware requirements. Within static gesture recognition to design and confirm detection of joint fingers. System of intuitive gestures has to be designed as well. Designed algorithms have to be confirmed and their success evaluated. Make a methodics for practical using of designed algorithms within multimedia TV based on results and define usage of designed principles based on utility.

Obsah

1. ÚVOD	4
2. MOTIVÁCIA A CIELE PRÁCE	5
3. VŠEOBECNÝ PREHĽAD GEST	6
3.1 METODIKA ROZPOZNÁVANIA GEST	6
3.2 METÓDY DETEKcie GEST	6
3.3 METÓDY SLEDOVANIA RÚK	7
3.4 METÓDY ROZPOZNÁVANIA GEST	7
4. VLASTNÁ PRÁCA	8
4.1 ROZPOZNÁVANIE PRSTOV POMOCOU KRUŽNÍC	8
4.1.1 TESTOVANIE ALGORITMU	9
4.2 ROZPOZNÁVANIE PRSTOV POMOCOU KONVEXNÝCH PORÚCH	10
4.3 K-CURVATURE ALGORITMUS	11
4.4 ROZPOZNÁVANIE PRSTOV NA ZÁKLADE ANALÝZY KRIVKY RUKY	12
4.4.1 APROXIMÁCIA STREDU DLANE	12
4.4.2 ODSTRÁNENIE PREDLAKTIA	13
4.4.3 REPREZENTÁCIA TVARU RUKY POMOCOU KRIVKY	14
4.5 TESTOVANIE VYBRANÝCH METÓD NAVZÁJOM	16
4.6 ALGORITMUS CIRCLE DYNAMIC GESTURE RECOGNITION	17
4.7 INKREMENTAČNÝ ROZPOZNÁVACÍ ALGORITMUS	18
4.8 VÝSLEDKY A TESTOVANIE	21
4.9 HBB-NEXT (NEXT-GENERATION HYBRID BROADCAST BROADBAND)	21
4.9.1 ANALÝZA POUŽÍVANIA GEST PRE MULTIMEDIÁLNU TV	22
5. VÝSLEDKY DIZERTAČNEJ PRÁCE	26
6. PREZENTÁCIA DOSIAHNUTÝCH VÝSLEDKOV	26
7. ZÁVER	27
8. PUBLIKÁCIE	28
9. ZOZNAM POUŽITEJ LITERATÚRY	31

1. ÚVOD

Pokroky ľudstva rastú čoraz rýchlejšie, no zároveň s pokrokmi rastú aj naše požiadavky v oblasti technológií, či už v medicíne, priemysle, vzdelávaní alebo domácnosti. Častokrát nestačíme spoznať novú technológiu a už na trh prichádza jej nasledovník. Toto je dôsledkom čoraz väčšej túžby ľudstva po nových objavoch a prekonávaní samých seba, či zvýšenia pohodlia a oddychu. Technológie a ich dopad sa dostal takmer všade, kde si vieme predstaviť. Od zobudenia budíkom až po večerné zaspávanie. Technológie používané ako pracovný nástroj sa transformovaním dostávajú v rovnakej alebo podobnej forme do domácností a z našich domácností sa stávajú malé technologické parky. Či už inteligentné kuchynské spotrebiče, ktoré vedia samé dopieť a pripraviť jedlo, alebo chladničky ktoré poznajú svoj obsah. Pomocou mobilného telefónu vieme v domácnosti ovládať osvetlenie, kúrenie, či analyzovať spotrebu energie celého domu. Samozrejmosťou sú aj inteligentné okná a dvere, či žalúzie, ktoré sa vedia prispôbovať podmienkam vonku či vnútri.

Čo sa stalo skutočne nečakaným trendom 21. storočia, je bezpochyby ovládanie gestami. Gestá patria už od počiatku ľudstva k jedným z najzákladnejších a najpoužívanejších dorozumievacím prostriedkom. V komunikácii dodávajú nášmu vyjadrovaniu kolorit, ktorý niekedy iba ťažko dokážeme opísať slovami. Samotné gesto neslúži iba ako sekundárny aspekt našej komunikácie. Môže sa stať majoritnou zložkou, kedy skutočne nesie najväčšiu časť informácie, ktorú chceme prijímateľovi odovzdať. Z tohto dôvodu ich funkciu ľudstvo začalo využívať na ovládanie stojov a zariadení. Počas niekoľkých rokov sa gestá stali novou, ale v krátkom čase prirodzenou komunikáciou človeka a digitálneho zariadenia. Gestami ovládame zariadenia, komunikujeme, hráme hry a tvoríme. Gestá začali mať svoj význam v rôznych oblastiach nielen v priemysle, ale aj v našich domácnostiach. Ovládanie robotov či časti výroby dotykovými displejmi je bežným výdobytkom dnešnej doby. I domácnosti začínajú dýchať takýmito vymoženosťami. Technológie rastú čoraz rýchlejšie a rýchlejšie a nám nezostáva nič iné iba sa s touto skutočnosťou zmieriť. Ako sa zmení naše zmýšľanie nástupom stále nových a nových zariadení a aplikácií, to asi povedať nevieme.

2. MOTIVÁCIA A CIELE PRÁCE

Rozpoznávanie gest predstavuje pomerne širokú oblasť nielen výskumu ale aj samotnej oblasti použitia. Ako bolo spomínané vyššie, v mojej dizertačnej práci sa venujem vývoju algoritmov pre rozpoznávanie gest v prostredí multimedialnej televízie. Motiváciou pre túto oblasť výskumu je potreba takéhoto systému, ktorý bude integrovateľný do platformy multimodálneho rozhrania v rámci projektu HBB-Next a zároveň bude otvoreným systémom vhodným na úpravy a modifikácie. Nakoľko komerčné systémy sú uzamknuté systémy a nemožno ich ľubovoľne modifikovať, modul gest s otvoreným rozhraním má veľké opodstatnenie. Taktiež väčšina komerčných systémov je založených na snímaní pomocou RGB kamery, čo v prostredí obývačky sa stáva často nepoužiteľné. Ďalšou motiváciou pre vytvorenie takéhoto systému bol fakt, že nový systém bude nezávislý na výrobcovi.

Na základe analýzy potrieb a analyzovania existujúcich problémov rozpoznávania gest boli vytýčený hlavný cieľ navrhnúť systém rozpoznávania gest pre multimedialnú TV, ktorý sa rozčlenil do nasledujúcich podcieľov dizertačnej práce:

- Navrhnúť algoritmus pre rozpoznávanie statických gest
- Navrhnúť algoritmus pre detekciu spojených prstov pri statických gestách
- Navrhnúť algoritmus pre rozpoznávanie intuitívnych gest jednou rukou pre ovládanie multimedialnej TV s dôrazom na jednoduchosť algoritmu
- Modifikovať navrhnutý algoritmus intuitívnych gest na detekciu gest oboch rúk
- Navrhnúť metodiku pre používanie gest multimedialnej TV

3. VŠEOBECNÝ PREHĽAD GEST

Taxonómia gest je oblasť, ktorej sa výskumníci venujú už niekoľko desaťročí. Za tento čas vzniklo množstvo rozdelení, ktoré gestá analyzujú z najrôznejších pohľadov. V technickej literatúre možno gestá rozlišovať na základe ich dynamických charakteristík na statické a dynamické gestá. Vývojom tejto oblasti sa dospelo i ďalšiemu typu gest a tým sú kontinuálne gestá. Pri statických gestách ako už z vyplýva názvu, sa pohyb neprejavuje. Ich informačná hodnota je obsiahnutá v póze gestikulujúcej končatiny [1], ktorá je zachovaná v definovanom čase. Pri spracovaní gesta tak stačí analyzovať jedinú snímku, z ktorej je možné detegovať gesto. Dynamické gestá majú na rozdiel od predošlej kategórie časový priebeh, počas ktorého sa tvar gesta, pozícia a póza gestikulujúcej končatiny mení. Predstavujú teda kombináciu viacerých statických gest, ktoré sa menia v čase aj polohe. Kontinuálne gestá predstavujú ďalšie rozšírenie dynamických gest. Obe predošlé kategórie vyžadujú rozpoznanie začiatku a konca gesta, zatiaľ čo kontinuálne gestá sú analyzované neustále. Zmeny sledovanej končatiny (alebo častejšie celého tela) sa prenášajú zvyčajne do virtuálneho prostredia, kde slúžia na neustálu interakciu s aplikáciou.

Pre prácu so samotnými gestami je potrebné nejakým spôsobom reprezentovať získavané obrazové dáta, z ktorých treba detegovať jednotlivé časti tela prípadne ruky. Pre tento účel existuje niekoľko reprezentácií, ktoré sú členené do dvoch základných skupín: metódy založené na 3D modeloch a metódy založené na vzhľade [2].

3.1 METODIKA ROZPOZNÁVANIA GEST

Rozpoznávanie gest nie je zložené z jedinej fázy rozpoznávania, ale samotnému rozpoznávaniu predchádzajú ešte ďalšie fázy. Na to, aby sme mohli začať so samotným rozpoznaním gesta je potrebné extrahovať objekt (ruku) z obrazu, určiť jeho parametre a následne vykonať rozpoznávanie. Tri základné kategorizácie fáz rozpoznávania sú: detekcia, sledovanie pohybu a rozpoznávanie.

3.2 METÓDY DETEKIE GEST

Základným krokom systémov rozpoznávania gest je detekcia a segmentácia príslušnej obrazovej oblasti. Jej účel je vymedziť región, na ktorom sa nachádza ruka a tak zredukovať

veľkosť vyšetrovanej oblasti na minimum. Tým dosiahneme rýchlejšie spracovanie a nižšiu výpočtovú náročnosť. Následne z tohto obrazu je treba odstrániť pozadie, aby detegovaná oblasť obsahovala iba vyšetrovaný objekt, ktorý sa následne posunie do ďalších fáz (sledovanie, rozpoznávanie). Existuje niekoľko spoľahlivých metód na detekciu a ich rôzne kombinácie, z ktorých si predstavíme tie najvýznamnejšie.

3.3 METÓDY SLEDOVANIA RÚK

Sledovanie rúk je všeobecne zložitou problematikou, nakoľko medzi niekoľkými snímkami môžu vznikáť obrovské zmeny vzhľadu sledovanej ruky. Sledovanie je proces definovaný ako medzislímkové vyhodnocovanie pohybu ruky na základe zmeny jej vlastností v rámci vyšetrovaného segmentu. Robustnosť sledovania rúk je založená na dvoch hlavných spôsoboch. Prvým je medzislímkové spájanie rúk prípadne prstov do trajektórií zodpovedajúcich medzislímkovému pohybu. Tieto trajektórie poskytujú základné informácie súvisiace s gestom a neskôr môžu byť použité v surovej forme (kreslenie trajektórie pomocou ruky), alebo po analýze na rozpoznanie typu gesta ruky. Druhým spôsobom je metóda založená na modeloch, kde sa sledovanie vykonáva na základe odhadov jednotlivých parametrov a vlastností ruky, ktoré sa nedajú priamo zistiť v danom časovom momente.

3.4 METÓDY ROZPOZNÁVANIA GEST

Cieľom rozpoznávania gest ruky je interpretácia sémantiky ruky, ktorá reprezentuje jej umiestenie, pózu alebo gesto. Rozpoznávanie založené na počítačovom videní, môže byť rozdelené na statické a dynamické gestá. Na detegovanie statického gesta alebo pózy je možné použiť klasifikátor alebo šablóny. Statické gestá sú ďalej klasifikované do lineárnych alebo nelineárnych učení. Ďalšou cestou ako kategorizovať učiace algoritmy je na základe predpokladaných výstupov. Charakterizujeme ich ako učenie pod dohľadom (označované vzorky), učenie bez dohľadu (zhluky vzoriek bez označenia), polo-dohľadové učenie (hybridný prístup označovaných a neoznačovaných vzoriek dát) a zosilňovacie učenie (učenie pravidiel na základe pozorovania, prípadne učenie s učiteľom na základe predchádzajúcich skúseností). Keďže dynamické gestá majú dočasný charakter a vyžadujú techniky, ktoré vedia pracovať s touto veličinou ako napríklad Skryté Markove Modely (HMM), kde dočasná veličina je modelovaná pomocou reprezentáciu gesta ruky (ako napríklad: model založený na pohybe).

4. VLASTNÁ PRÁCA

Pri obrazovej detekcii a rozpoznávaní či už gest, tváre alebo postavy pomocou farebného (RGB) obrazu je dôležitým faktorom kontrast hrán, ktoré určujú črty objektu a o ktoré sa zväčša pri detekcii alebo rozpoznávaní opierame. Nízka či vysoká svetelnosť dokáže často extrémne zhoršiť kvalitu rozpoznávania a detekcie a to nehovoríme o neustálej zmene svetelnosti prostredia, ktorým môže byť napríklad pozeranie televízora pri vypnutom svetle v miestnosti. Uvedené dôvody vieme buď úplne alebo čiastočne riešiť pomocou kamier pracujúcich v infračervenej oblasti a nie v oblasti viditeľného svetla. Infračervená oblasť za štandardných podmienok v interiéri, kedy osvetlenie dosiahneme pomocou slnka (nie priamymi lúčmi) alebo umelým osvetlením, minimálne ovplyvní alebo vôbec neovplyvní infračervenú oblasť, v ktorej kamera pracuje.

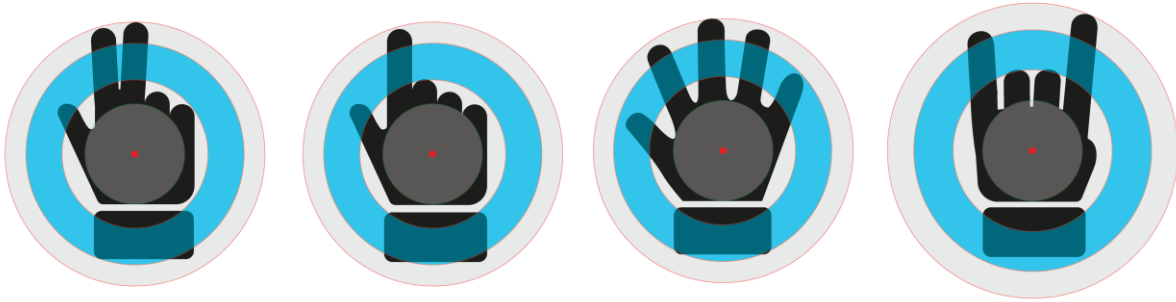
Pri rozpoznávaní prstov ruky nevieme nikdy zaručiť pozíciu snímanej ruky, čiže nevieme zabezpečiť ideálnu alebo aspoň optimálnu pozíciu ruky tak, aby bola jednoznačne separovateľná od prostredia. Pri snímaní ruky pomocou RGB kamery by separácia bola závislá nie len od osvetlenia ale aj od farby oblečenia osoby. Pokiaľ by optimálne podmienky neboli splnené, nastal by problém s určovaním obrysov ruky. Keďže pracujeme v infračervenej oblasti, aby sme v takejto situácii predišli zlievaniu objektov, využijeme poznatok tretieho rozmeru a to hĺbky. Použitím hĺbky vieme jednoducho separovať objekty na náklade ich vzdialeností od senzora. Jednoduchou filtráciou hĺbok dostávame z pôvodného obrázku nový filtrovaný obraz.

4.1 ROZPOZNÁVANIE PRSTOV POMOCOU KRUŽNÍC

Algoritmus pracuje s dvoma regiónmi rúk – pravej a ľavej. V rámci týchto regiónov sa pomocou opísanej kružnice ruky aproximuje stred dlane. Následne sa aproximovaný stred použije ako stred kružnice, ktorá pretína prsty. Polomer takejto kružnice musí byť menší ako polomer opísanej kružnice ruky a zároveň väčší ako polomer vpísanej kružnice dlane.

$$r_{dlane} < r_{detekcia} < r_{ruky} \text{ (platí: } r_{dlane} \ll r_{ruky} \text{)}$$

Následne sa z kružnice radiálne odčítavajú hodnoty, aby sa určil počet spojitých oblastí. Ten potom predstavuje odhadované určenie počtu prstov. Región hodnôt na intervale $\langle 0 - 2\pi \rangle$ sa filtruje pomocou morfologickej operácie „uzatvorenie“ [5], čím sa odstránia nedostatky rozlíšenia.



Obr. 1 Princíp detekcie prstov

Výsledkom je množina spojitých oblastí, ktoré zodpovedajú všetkým výčnelkom ruky – vrátane zápästia. Na základe šírky spojitých oblastí a za pomoci vektora medzi stredom dlane a predlaktia je možné odfiltrovať spojitý región predstavujúci predlaktie. Teda vieme povedať, že počet prstov je o jedno menší ako počet detegovaných regiónov. Ruka môže mať ľubovoľné natočenie v troch osiach. My budeme uvažovať o natáčaní ruky tak, aby vždy čelila dlaň ku kamere. I tak dostávame množstvo pootočení ruky, kedy prsty môžu ukazovať horizontálne alebo až vertikálne. Toto je akceptované natáčanie ruky v plnom rozsahu 2π .

4.1.1 TESTOVANIE ALGORITMU

Pri optimálnej pozícii ruky (otvorená dlaň čelí senzoru) algoritmus pracuje pomerne presne 84% (množina 120 gest) a jeho implementácia je pomerne nenáročná. Obrázok nižšie zobrazuje ideálne pozície otvorenej a zatvorenej dlane. Celá výpočtová náročnosť sa zjednodušila rýchlym výberom regiónu rúk a pomerne jednoduchou detekciou spojitých oblastí na kružnici. Ak však začneme robiť rozpoznania počtu prstov mimo optimálnej pozície s náklonmi ruky v ďalších dvoch osiach, rozpoznávací schopnosť sa radikálne zhorší. Miesta na kružnici, ktoré pretínajú prsty sú vyznačené červenou farbou. Je vidno, že pri miernom naklonení ruky počty prstov nie sú správne rozpoznané. Počet prstov, ktorý je rozpoznávaný nie je veľmi presný a nie je možné určiť, kedy je počet prstov rozpoznávaný správne.

4.2 ROZPOZNÁVANIE PRSTOV POMOCOU KONVEXNÝCH PORÚCH

Konvexný obal objektu je pojem, ktorý bol definovaný už v osemdesiatych rokoch minulého storočia [6]. Konvexné poruchy [7] môžeme voľne definovať ako plochy, ktoré vzniknú rozdielom plochy konvexného obalu a vnútornej plochy kontúry. Jednoduchým príkladom konvexného obalu môže byť gumička natiahnutá medzi prstami. Miesta konvexného obalu, ktoré sú delené nepravidelnosťou objektu sú nazývané konvexné poruchy. Práve teória konvexných porúch bude použitá na stavbu algoritmu rozpoznávania prstov ruky.

Predtým, než bude možné hľadať konvexné poruchy, je potrebné vytvoriť z objektov na obraze kontúry. Následne si popíšeme postupnosť celého algoritmu v krokoch, kedy na konci dostaneme počet ukázaných prstov.

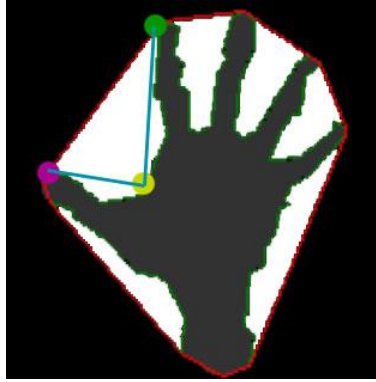
A. Nájdenie stredu ruky

Na určenie stredu otvorenej ruky použijeme kontúru ruky, ktorá bola zistená v predošlých krokoch. Toto určenie sa urobilo pomocou obrazových momentov. Obrazový moment je vážený priemer intenzít pixelov obrazu, ktorý sa aplikuje na šedotónový obraz s intenzitami pixelov $I(x, y)$, kde momenty $M_{i,j}$ sú počítané ako

$$M_{i,j} = \sum_x \sum_y x^i y^j I(x, y)$$

B. Výpočet konvexných porúch

Určovanie konvexných porúch je možné robiť v smere hodinových ručičiek alebo v protismere hodinových ručičiek pričom výsledok bude v oboch prípadoch zhodný. Výsledkom aplikovania tejto metódy na konvexný obal (Obr. 2– červená línia) budú trojice bodov pre každú konvexnú poruchu konvexného obalu a to začiatkový bod (zelená farba), hĺbkový bod (žltá farba) a koncový bod (fialová farba).



Obr. 2 Vyšetovanie konvexných porúch ruky

C. Nájdenie maximálnej vzdialenosti bodu prstu

Pre nájdenie najvzdialenejšieho bodu od stredu ruky bol použitý algoritmus prehľadávania medzi všetkými bodmi, kde sa prehľadávalo a spomedzi všetkých bodov sa hľadal taký bod, ktorého vzdialenosť bola maximálna.

D. Nájdenie minimálnej vzdialenosti bodu prstu

Podobne ako v predchádzajúcom kroku sa vykoná hľadanie ale tento krát minimálnej vzdialenosti medzi bodmi.

E. Hľadanie výšky trojuholníka

Po určení maximálnych a minimálnych hodnôt ruky musíme určiť strednú prahovú hodnotu, ktorá bude medznou hranicou pre odfiltrovanie všetkých parazitných porúch, nakoľko konvexných porúch na ruke môže byť určených viac ako vyžadujeme.

F. Filter parazitných porúch

Prvý filter sa zaoberá filtráciou všetkých porúch ktoré nemajú minimálnu stanovenú výšku trojuholníka (určuje sa na základe veľkosti výrezu okna a priemernej veľkosti ruky). Všetky takéto konvexné poruchy a ich body budú odfiltrované.

4.3 K-CURVATURE ALGORITMUS

Princíp algoritmu K-Curvature aplikovaného na detekciu prstov ruky je založený na hľadaní požadovaného uhla medzi dvomi vektormi na kontúre ruky. Vo všeobecnosti sa hľadá ostrý uhol. Parameter uhla medzi dvomi vektormi na kontúre ruky pre detekciu prstov na zariadení Kinect v.1 bol zvolený $\alpha = 50^\circ$.

Vzdialenosť od stredového bodu $d(P(i)) = n$. Hodnota k sa počítaná dynamicky v závislosti na veľkosti vzdialenosti ruky od snímača, nakoľko sa veľkosť ruky so vzdialenosťou od zariadenia mení, musí sa dynamicky prepočítavať aj hodnota n .

$$|P(i), S| - |P(i - n), S| > 0$$

a zároveň platí

$$|P(i), S| - |P(i + n), S| > 0$$

Algoritmus je možné optimalizovať tak, že sa výpočet nevykonáva na každý bod kontúry ruky, ale vykonáva sa s istým krokom (k). Tak dosiahneme nie m výpočtov na jednej kontúre ruky ale m/k výpočtov. Algoritmus vráti niekoľko bodov ruky, ktoré treba filtrovať. Na filtráciu použijeme pravidlo o vzdialenosti od stredu ruky, po ktorej zostáva jediný bod P_F pre každý prst.

$$P_F = \max\{d(P_1) \dots d(P_j)\}$$

Pre pomerne nízku náročnosť a zjednodušenie algoritmu je možná implementácia za predpokladu nízkych hardvérových požiadaviek, čiže nižšieho výpočtového výkonu i nižšieho rozlíšenia hĺbkovej kamery (640 x 320 pixelov).

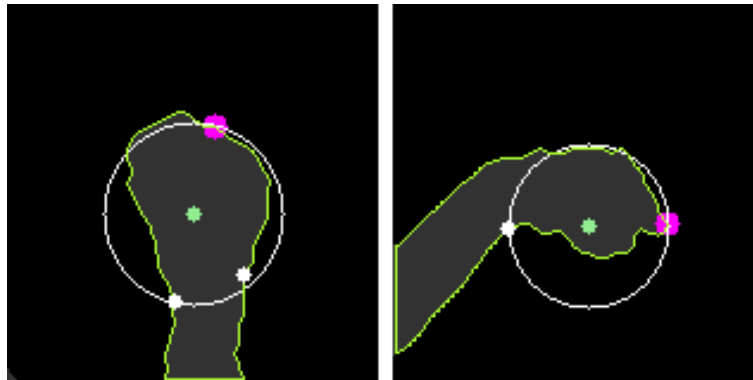
4.4 ROZPOZNÁVANIE PRSTOV NA ZÁKLADE ANALÝZY KRIVKY RUKY

Nasledujúci algoritmus pracuje na odlišnom princípe ako predchádzajúce algoritmy a aj z toho dôvodu sa vykonávajú niektoré kroky predspracovania, ktoré v predchádzajúcich metódach neboli uvedené. Keďže obraz získaný z hĺbkovej kamery je nepresný a skreslený, je vhodné o jeho čiastočnú rekonštrukciu. Na vyrezaný binárny obraz teda použijeme metódu mediánového filtra [8]. Parameter filtra sa mení dynamicky na základe vzdialenosti ruky od hĺbkového senzora a to z dôvodu, že pri väčších vzdialenostiach by pri konštantnej hodnote mohlo prichádzať k strate informácií.

4.4.1 APROXIMÁCIA STREDU DLANE

Pre presnejšie určenie stredu využijeme znalosti o konvexných poruchách prezentované vyššie, ktoré sa nachádzajú na kontúre ruky v blízkosti dlane a pomocou nich definujeme hranice pre minimálnu opísanú kružnicu dlane. Pri takejto detekcii pomocou opísanej kružnice

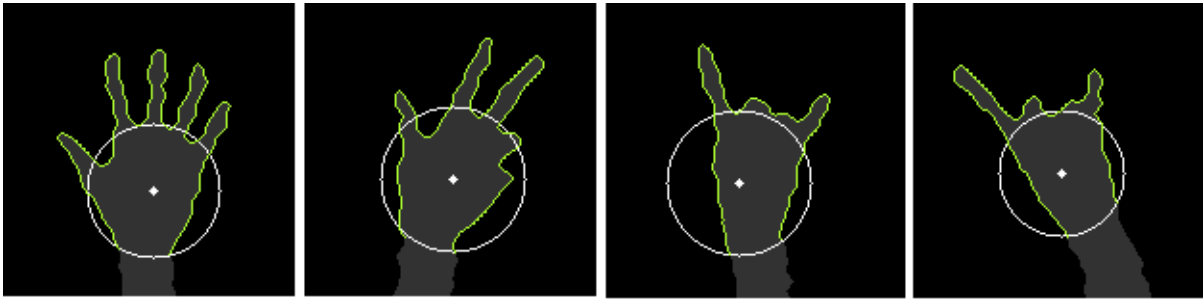
nastávajú prípady, kedy nemožno správne detegovať stred dlane. Tieto situácie je možné vyriešiť pridaním ďalšieho bodu na kontúru ruky tak, aby to bol bod konvexného obalu, ktorého vzdialenosť je čo najpodobnejšia vzdialenosti detegovaných porúch od stredu kružnice. Takouto korekciou dosiahneme správne určovanie stredu dlane aj pri neštandardných situáciách a náklonoch ruky aj v prípade, že analýzou konvexných porúch sme dokázali detegovať iba jeden bod konvexnej poruchy na kontúre ruky.



Obr. 3 Riešenie špeciálnych prípadov

4.4.2 ODSTRÁNENIE PREDLAKTIA

Pre odstránenie predlaktia je potrebná správna detekcia niektorého kĺbu ruky zápästia alebo lakťa z toho dôvodu aby sme vedeli určiť orientáciu ruky a dlane. Metóda, ktorá bola popísaná v práci [9] rieši hľadanie zápästia tak, že definuje os ruky. Postupným krokováním po osi vodorovnej s osou ruky merali šírku kontúry ruky a kolmo na os ruky. Zápästie definovali tak, že je to miesto, kde šírka kontúry začína prudko narastať. Tento prístup mal niekoľko medzier a aj pri rôznych vlastných modifikáciách sa nepodarilo dosiahnuť touto metódou uspokojujúce výsledky. Vlastným riešením je využitie minimálnej opísanej kružnice. Úlohou je teda nájsť dva body kontúry také, ktoré sú priesečníkom s kružnicou. Samotná kontúra ruky je krivka, ktorá negarantuje práve dva priesečníky s kružnicou, nakoľko sa môže stať že priesečník nedosiahne žiadny alebo dokonca má niekoľko priesečníkov. Je potrebné ešte definovať otočenie ruky a to za pomoci pozície lakťa. Riešenie sa ukazuje v podobe nájdenia bodu konvexného obalu, takého, ktorý má najmenšiu vzdialenosť od bodu lakťa. Od tohto bodu postupne prechádzame jednotlivé body kontúry dovtedy, pokiaľ nenájdeme bod, ktorý spadá do vnútra minimálnej opísanej kružnice dlane. Tento bod môžeme považovať za priesečník.



Obr. 4 Odstránenie predlaktia

4.4.3 REPREZENTÁCIA TVARU RUKY POMOCOU KRIVKY

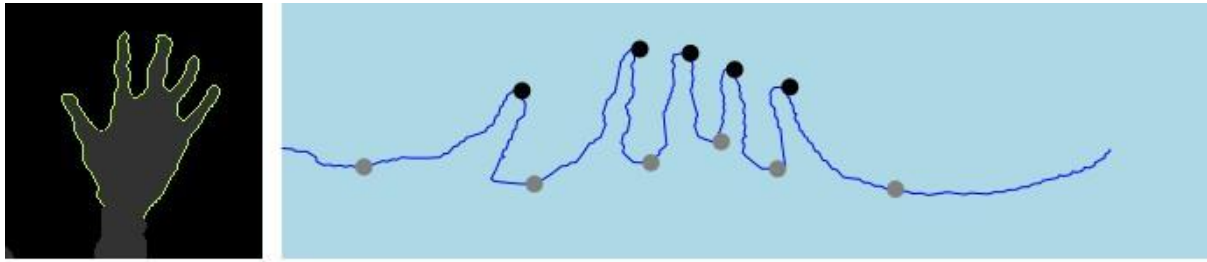
Pri reprezentácii ruky pomocou krivky budeme používať funkciu ako závislosť vzdialenosti bodu kontúry ruky od stredu dlane a uhla opísanej kružnice. Výsledkom je krivka, ktorá reprezentuje kontúru ruky v závislosti na vzdialenosti od stredu opísanej kružnice. Analýzou získanej krivky chceme dosiahnuť detekciu prstov, ktorú je možné jednoducho pracovať. Na to aby sme dokázali správne určiť jednotlivé prsty, ktoré predstavujú náhle zmeny krivky, je využitie znalostí o lokálnych extrémoch. Analýza lokálnych extrémov je však možná u funkcií, ktoré majú matematický predpis, čo v našom prípade neplatí. Z tohto dôvodu je potrebné definovať princíp určovania lokálnych extrémov na danej krivke. Jej definícia je uvedená nižšie.

$$x(n - k) < x(n) > x(n + k) \rightarrow k \in \{1..m\}, m \in Z^+$$

Kde $x(n)$ je vzdialenosť vyšetrovaného bodu krivky od stredu opísanej kružnice, $x(n - k)$ je m predchádzajúcich bodov a $x(n + k)$ je m nasledujúcich bodov, m vyjadruje počet porovnávaných bodov pred a za vyšetrovaným bodom krivky. Pokiaľ vzťah platí, bod prehlasujeme za lokálne maximum. Pri analýze lokálnych miním sa vyššie uvedený vzťah upraví nasledovne:

$$x(n - k) > x(n) < x(n + k) \rightarrow k \in \{1..m\}, m \in Z^+$$

Takýmto postupom je však možné na krivke nájsť niekoľko desiatok lokálnych extrémov. Z tohto dôvodu je potrebné pomocou vhodnej filtrácie odstrániť nadbytočnosť a ponechať extrémny, ktoré definujú požadované hľadané body. Filtrácia je založená na princípe rozdielu funkčných bodov jednotlivých za sebou idúcich lokálnych extrémov. Zvolením vhodného parametra dostávame len lokálne extrémny definujúce vrcholy prstov a údolia medzi nimi.



Obr. 5 Filtrácia lokálnych extrémov (sivá - minimá, čierna - maximá)

Po analýze krivky ruky a vyšetrení potrebných extrémov, pristúpime k analýze prstov ruky. Po získaní lokálnych extrémov krivky nasleduje extrakcia prstov ruky. Jednou z možností je jednoduché pospájanie prislúchajúcich maxím a miním, čím vytvoríme trojuholníky predstavujúce prsty. Ďalším prístupom je detekcia prstov na základe minimálnej relatívnej vzdialenosti a prahovania [10]. Takýto postup nám zabezpečí spoľahlivejšie výsledky a vyššiu robustnosť. Aplikácia takéhoto prístupu, však nesie so sebou nedostatok. Pri použití globálneho parametra pre všetky prsty vznikalo stále množstvo prípadov, kedy dopad globálneho parametra výrazne ovplyvňovalo výsledok. Dôsledkom je občasná detekcia prsta v časti zatvorenej dlane. Je potrebné určiť množinu parametrov, tvorenú z piatich parametrov, kde každý parameter zastupuje jediný prst ruky. Počíta sa ako $1/65$ z relatívnej vzdialenosti špičky každého prsta. Takto dostávame robustnú sústavu parametrov adaptovanú pre každý prst.

V istých prípadoch sa stáva, že do detegovanej množiny sa berie aj proporcia ruky, ktorá nie je prstom alebo je zohnutým prstom. Z tohto dôvodu vykonáme filtráciu, ktorá zabezpečí odstránenie takýchto neželaných proporcií ruky.

$$\forall E_{max} \begin{cases} E_{max}(d) \geq 1/3 * d_{max} \rightarrow E_{max}(d) \\ E_{max}(d) < 1/3 * d_{max} \rightarrow 0 \end{cases}$$

Preto túto procedúru vykonávame iba pre tie lokálne maximá, ktorých relatívna vzdialenosť od stredu ruky je väčšia než $1/3$.

Po analýze krivky ruky môžeme pristúpiť k determinácii váh a intervalov jednotlivých prstov ruky, čím určíme jednotlivé vyšetované segmenty. Interval segmentu je šírka výčnelku krivky, ktorá spĺňa podmienku lokálneho maxima a jej interval je daný ako $\langle X_1, X_2 \rangle$. Váhou segmentu nazývame plochu výčnelku ohraničenú prahovacou priamkou zospodu. Následne po definovaní váh segmentu V_S sa vykoná samotná detekcia prstov. Priemerná váha segmentu V_P a koeficienty pre korektnú detekciu počtu prstov N sú definované nasledovne:

$$N = \begin{cases} 1, & V_S \leq 1,5 * V_P \\ 2, & 1,5 * V_P < V_S \leq 2,4 * V_P \\ 3, & 2,4 * V_P < V_S \end{cases}$$

Na zabránenie kolísania počtu prstov v krátkom časovom intervale je použitý zásobník s modusovým filtrom závislý na posledných m meraniach.

4.5 TESTOVANIE VYBRANÝCH METÓD NAVZÁJOM

Pre vybrané algoritmy statických gest popísané vyššie bolo vykonané krížové testovanie a porovnávanie robustnosti v jednotlivých spoločných vlastnostiach. Gestá boli vykonávané vo vzdialenosti 110 cm od snímacieho zariadenia a celkovo bolo každou osobou vykonaných 100 gest a celková množina vykonaných gest bola o veľkosti 400.

Tab. 1 Porovnanie algoritmov statických gest

	Konvexné poruchy	K-Curvature	Analýza krivky
Akceptovaná vzdialenosť od senzora	119 cm	160 cm	175 cm
Zložitosť algoritmu (počet špecifických krokov)	7	6	7
Detekcia spojených prstov	Nie	Nie	Áno
Stupeň voľnosti	40%	70%	95%
Úspešnosť	80%	92%	90%

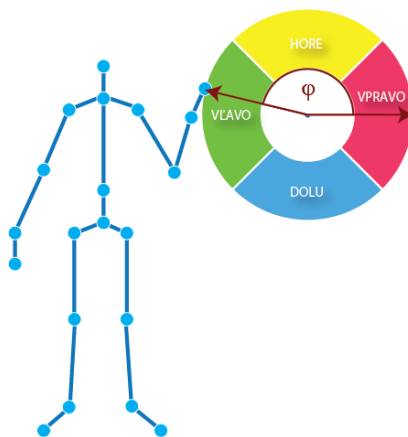
V ďalšom kroku boli vykonávané testy na robustnosť rotácie ruky v troch osiach a ich vplyv na správne rozpoznanie ukazovaného gesta. Podmienky testovania boli rovnaké ako v predchádzajúcom prípade.

Tab. 2 Porovnanie algoritmov statických gest - voľnosti rotácie

Hranice natočenia ruky ^a						
Konvexné poruchy	X os		Y os		Z os	
		35°	75°	25°	30°	180°
K-curvature	X os		Y os		Z os	
		35°	75°	25°	40°	175°
Analýza krivky	X os		Y os		Z os	
		50°	85°	25°	40°	150°

4.6 ALGORITMUS CIRCLE DYNAMIC GESTURE RECOGNITION

Navrhnutá metóda je založená na princípe kružníc, preto bola nazvaná ako metóda CDGR (Circle Dynamic Gesture Recognition). Jej princípom je vyšetovanie na jednotkovej kružnici. Algoritmus deteguje uhol φ od začiatku gesta až po koniec gesta. Začiatok gesta je definovaný ako pretnutie vnútornej kružnice, koniec gesta je definovaný ako pretnutie vonkajšej kružnice. Pretnutím kružnice sa myslí dosiahnutie absolútnej vzdialenosti od spoločného stredu. V okamihu, kedy sa bod ruky dostane za vnútornú kružnicu, meria sa čas kedy dosiahne hranicu vonkajšej kružnice. Diferencia týchto dvoch časov nám poskytuje informáciu o rýchlosti gesta. Samozrejme okrem času sa zmeria aj uhol, pod ktorým bod ruky prešiel vonkajšiu kružnicu. Takýmto spôsobom vieme určiť smer gesta aj jeho rýchlosť. Algoritmus má definovanú aj minimálnu rýchlosť, ktorou treba prekonať vnútornú kružnicu, aby sa zamedzilo falošným gestám. Následne je možné kružnicu rozdeliť do 4 kvadrantov, ktoré korešpondujú jednotlivým smerom (hore, dolu, vpravo, vľavo). Algoritmus bol ďalej upravený tak, aby vedel pracovať s oboma rukami súčasne. Touto úpravou sme získali možnosť rozpoznávať väčšie množstvo gest a zároveň sa vytvorila množina komfortných gest ako napríklad (priblíženie / oddialenie, rýchly posun vpravo / vľavo).



Obr. 6 CDGR (Circle Dynamic Gesture Recognition)

Testovanie algoritmu bolo vykonané 10 osobami, kedy každá osoba vykonala gesto 10-krát. Zariadenie ktoré bolo použité na snímanie osôb bolo Kinect v.1. Každý testovaný subjekt stál vo vzdialenosti 1,8 metra od snímacieho zariadenia a vykonával sled definovaných gest. Takto bola získaná vzorka 400 gest. Po vyhodnotení výsledkov bola vypočítaná úspešnosť metódy na 94%.

4.7 INKREMENTAČNÝ ROZPOZNÁVACÍ ALGORITMUS

Inkrementačný rozpoznávací algoritmus (IRA) [11] bol primárne navrhnutý pre dotykové obrazovky. Tento algoritmus som sa rozhodol použiť pre výskum. Idea je modifikovať inkrementačný rozpoznávací algoritmus a pretvoriť princípy dotykového ovládania pre bezdotykové zariadenia. Princípom je vytvorenie virtuálneho povrchu pred používateľom, kde je možné pomocou ruky kresliť krivky. Ako bolo uvažované vyššie, senzor Kinect pracuje so vzdialenosťami a práve táto vlastnosť bola využitá na vytvorenie imaginárneho povrchu. Ak užívateľ vloží ruku do imaginárneho priestoru, ktorý je definovaný špecifickou relatívnou vzdialenosťou, je možné interagovať s virtuálnym povrchom.

Tento algoritmus bol modifikovaný tak, aby spĺňal požiadavky rozpoznávania bezdotykových gest. V tomto prístupe je vzor definovaný ako množina segmentov popísaných vzorovým gestom. Každý segment popisuje progresívne narastajúce časti vzorového gesta tak, že prvý segment je podmnožinou druhého segmentu, ktorý je podmnožinou tretieho segmentu atď. a posledný segment reprezentuje celý vzor gesta. Každý segment je reprezentovaný ako séria časovo-zoradených bodov. S každým novým bodom pozorovaného gesta systém počíta Bayesovskú aposteriórnu pravdepodobnosť, že gesto sa zhoduje so vzorovým gestom pre každý zo vzorov v databáze a je daný vzťahom:

$$P(\omega_j | \mathbf{I}_i) = \frac{P(\omega_j)P(\mathbf{I}_i | \omega_j)}{\sum_k P(\omega_k)P(\mathbf{I}_i | \omega_k)},$$

kde $P(\omega_j)$ je apriórna pravdepodobnosť, $P(\mathbf{I}_i | \omega_j)$ je pravdepodobnosť a menovateľ je marginálnym výrazom. Pravdepodobnosť merania je daná ako pravdepodobnosť, že časť sledovaného gesta sa zhoduje so vzorom:

$$P_l(\mathbf{I}_i | \omega_j) = P_l(\mathbf{I}_i | \omega_j)E(\mathbf{I}_i | \omega_j),$$

kde $P_l(\mathbf{I}_i | \omega_j)$ je pravdepodobnosť sledovaného gesta a príslušná časť vzoru gesta. Je daná ako maximum funkcie vzdialenosti D v závislosti na Euklidovej vzdialenosti [12] medzi normalizovanými bodmi sledovaného gesta a vzorovým segmentom a medzi uhol otočenia a dvomi bodovými sekvenciami. Funkcia vzdialenosti je daná ako:

$$D(\mathbf{I}, \mathbf{S}) = \exp\left(-\left[\lambda \left(\frac{x_e^2}{\sigma_e^2}\right) + (1 - \lambda) \left(\frac{x_t^2}{\sigma_t^2}\right)\right]\right)$$

Funkcia vzdialenosti závisí od Euklidovej vzdialenosti x_e medzi prislúchajúcimi bodmi zaznamenanej trajektórie \mathbf{I} a známym vzorom \mathbf{S} , a stredným uhol otočenia x_t medzi príslušnými čiarovými segmentami \mathbf{I} a \mathbf{S} sekvenciami.

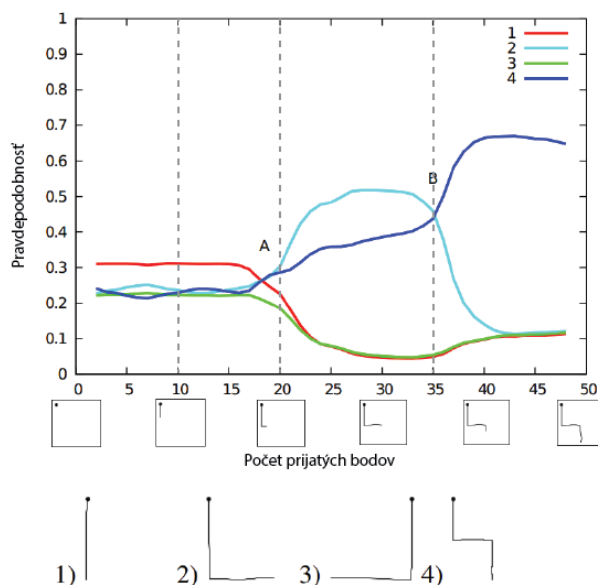
Hľadanie podobností [13] je vykonávané pomocou porovnávania dvoch hlavných vlastností: Euklidovej vzdialenosti a uhla natočenia. Prvá podobnosť je daná ako stredná Euklidova vzdialenosť medzi všetkými zodpovedajúcimi bodmi, pričom obe sekvencie bodov musia byť normalizované:

$$x_e = \frac{1}{n} \sum_{i=1}^n \|a_i - b_i\|$$

Ďalšou pravdepodobnosťou, ktorá sa meria je stredný uhol otočenia medzi dvomi bodovými sekvenciami použitím fixnej referenčnej osi:

$$x_t = \frac{1}{n-1} \sum_{i=2}^n dt(a_i, a_{i-1}, b_i, b_{i-1})$$

Aposteriórne pravdepodobnosti sú potom filtrované oknom, závislým na posledných piatich predikciách, aby sa stabilizovali.



Obr. 7Princíp sekvenčného rozpoznávania gesta

Pre účel bezdotykového rozpoznávania bol algoritmus upravený tak, aby bolo možné rozpoznávať gestá vo vzduchu, vykonané na imaginárnom plátne. V prvom rade je dôležité detegovať začiatok a koniec gesta, ktorá má byť rozpoznaná. Pre tento účel bola použitá vyššie spomenutá metóda na detekciu prstov. Užívateľ otvorením dlane dá najavo, že začína vykonávať gesto. Po vykonaní trajektórie dľaň zatvorí a zaznamenaná trajektória sa vyhodnotí. Pred samotným rozpoznávaním je potrebné vykonať úpravu dát tak, aby bolo možné na nich vykonať rozpoznávanie. V prvom kroku treba dáta filtrovať, nakoľko výsledná trajektória obsahuje množstvo nepotrebných bodov a duplicity. Následne sa po filtrácii body normalizujú na základe vzdialenosti. Potom sa vytvorí matica 1000 x 1000 bodov do ktorej sa body trajektórie zaznamenajú. Následne sa táto matica posiela na spracovanie a porovnávanie s existujúcimi vzormi v databáze. Ako bolo spomenuté vyššie, výsledkom nie je jedno konkrétne gesto, ale množina gest, pričom pre každé gesto v databáze je počítaná samostatná pravdepodobnosť.

Tvorba databázy založená na vytvorení „odtlačku“ trajektórie. Trajektória je definovaná ako sled bodov vo vymedzenom priestore. Tie sa následne normalizujú a prevádzajú do priestoru 1000 x 1000 bodov a uložia sa do databázy.

4.8 VÝSLEDKY A TESTOVANIE

Na snímanie trajektórie bol použitý senzor Kinect v.1. Trajektória bola následne prevedená do sekvencie bodov v matici 1000 x 1000 a uložená medzi vzory. Predvolená množina gest boli veľké písmená anglickej abecedy (26 znakov). Každé gesto bolo vykonané 4 osobami 5-krát zo vzdialenosti 1,8 metra od snímacieho zariadenia. Každé gesto bolo vykonávané zľava doprava. Celkovo bolo získaných 520 záznamov do testovanej množiny. Treba však pripomenúť, že nie všetky gestá boli vykonané s vysokou precíznosťou. Napriek tomu, že gestá neboli vykonávané precízne, metóda dosiahla úspešnosť na úrovni 91%.

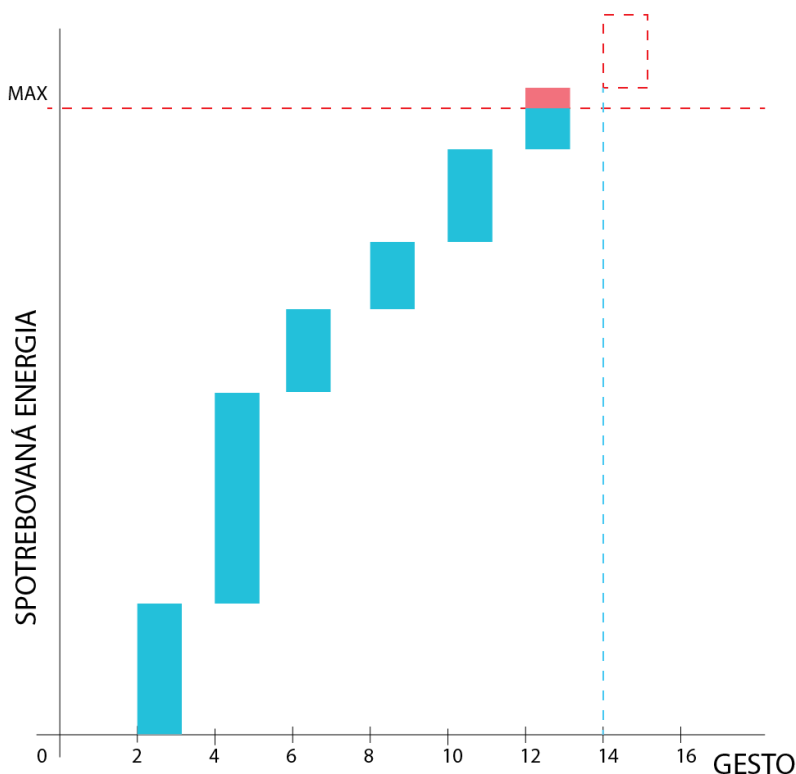
4.9 HBB-NEXT (NEXT-GENERATION HYBRID BROADCAST BROADBAND)

Pojem HBB-NEXT predstavuje štandard pre televíziu, kde cieľom je používateľovi poskytnúť nielen nadštandardný digitálny obsah, ale aj nadštandardnú komunikáciu a ovládanie systému v prostredí obývačky. V rámci projektu HBB-NEXT bolo vytvorené zjednotené rozhranie schopné okrem tradičných foriem ovládania naraz pracovať s vizuálnymi aj nevizuálnymi vstupmi, ako sú rozpoznanie tváre, rozpoznávanie gest, hlasových povelov a rozpoznanie rečníka. Všetky tieto modality spoločne tvoria systém, ktorý používateľsky príjemnou formou rozpozná používateľov a umožní im prirodzené a intuitívne ovládanie. Jeho názov je multimodálne rozhranie tzv. MMI. V platforme MMI boli v rámci mojej dizertačnej práce zaintegrované niektoré z vyššie navrhnutých algoritmov rozpoznávania gest za účelom návrhu vhodného ovládania pre multimedialnú TV. Pre účel jednoduchého prepínania kanálov bola implementovaná metóda ťahových gest, vďaka ktorej je možné pomocou ťahového gest vpravo / vľavo prechádzať TV kanály, gesta hore / dolu vyberať TV kanály alebo sa vrátiť do menu, prípadne gestom oboch rúk od seba / k sebe maximalizovať alebo minimalizovať obraz. Ťahovými gestami bola nahradená funkcia ovládača na prepínanie kanálov. Takisto je možné použiť na jednoduché ovládanie aj statické gestá ruky, kedy počtom prstov jednej ruky vieme pracovať s hlasitosťou, či rýchle prepnúť na obľúbený kanál. Ukážku grafického rozhrania je možné vidieť na obrázku nižšie. Zároveň bol integrovaný ďalší z navrhnutých algoritmov, ktorý zabezpečoval odmykanie obsahu bez potreby zadávania PIN kódu. Princíp spočíva v tom, že algoritmus dynamických gest bol implementovaný tak, aby vytvoril funkcionality prístupu k obsahu pomocou vzoru. Používateľ po vyzvaní systémom zadal vzor na odomknutie požadovaného obsahu. Výhodou odmykania pomocou vzoru je, že systém je citlivý na rukopis,

ale zároveň toleranciu na rukopis je možné jednoducho upravovať. Následne po integrácii boli vykonané testy, ktorých účelom bolo vytvorenie správnej metodiky pre ovládanie multimediálnej TV pomocou gest.

4.9.1 ANALÝZA POUŽÍVANIA GEST PRE MULTIMEDIÁLNU TV

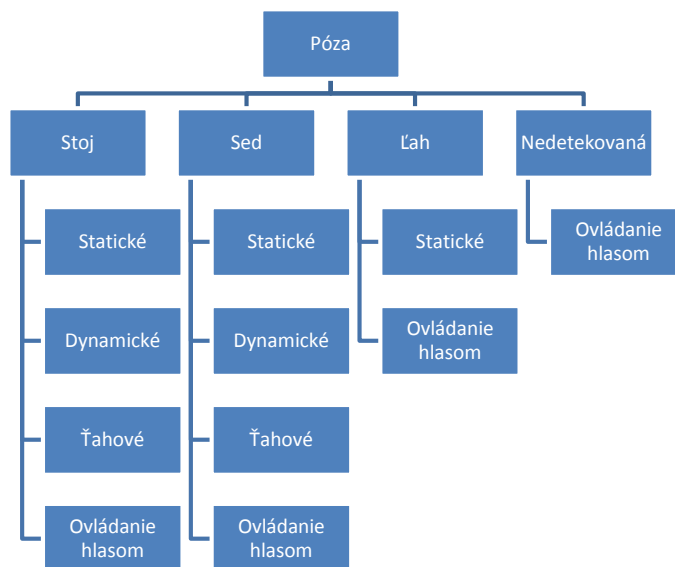
Gestá a multimediálna TV musia tvoriť symbiózu, ktorá musí evokovať u používateľa skutočne príjemné a pohodlné ovládanie. Nahradenie diaľkového ovládača preto nie je možné vždy a všade. Treba hľadať spôsoby a situácie, kedy je možné túto zámenu urobiť a nájsť pre ňu ten najlepší scenár. Počas môjho výskumu som zistil, že gesto je pohodlné pre používateľa vtedy, ak na jeho vykonanie nepotrebuje veľa energie. Čím menej energie stojí telo vykonať gesto, tým je vyššia pravdepodobnosť, že vykonám ďalšie gesto, inými slovami povedané, telo má obmedzené množstvo energie, ktoré môže a chce minúť a jeho spotreba a vôľa je závislá na množstve energie vydaného pre každé gesto. Táto hypotéza bola potvrdená pri testovaní. Dĺžka ovládania systému pomocou gest bola závislá na zložitosti gesta.



Obr. 8 Vplyv energie na vykonávanie gest

Obrázok (Obr. 8) zobrazuje situáciu, kedy dosiahnutím maximálnej hranice spotreby energie sa gestá stavajú nepohodlnými. Táto hranica je veľmi individuálna u každej osoby a taktiež je závislá od psychologického efektu (celkového dojmu), ktorý môže pozitívne i negatívne posúvať krivku maxima. To znamená, pokiaľ je pre konkrétnu osobu ovládanie atraktívne, vie si hranicu maxima nevedomky posunúť vyššie, v opačnom prípade krivka maxima spadne nižšie. Obrázok vyššie ukazuje, že osoba po dosiahnutí hranice 14 gest, nie je ochotná vykonávať ďalšie gestá.

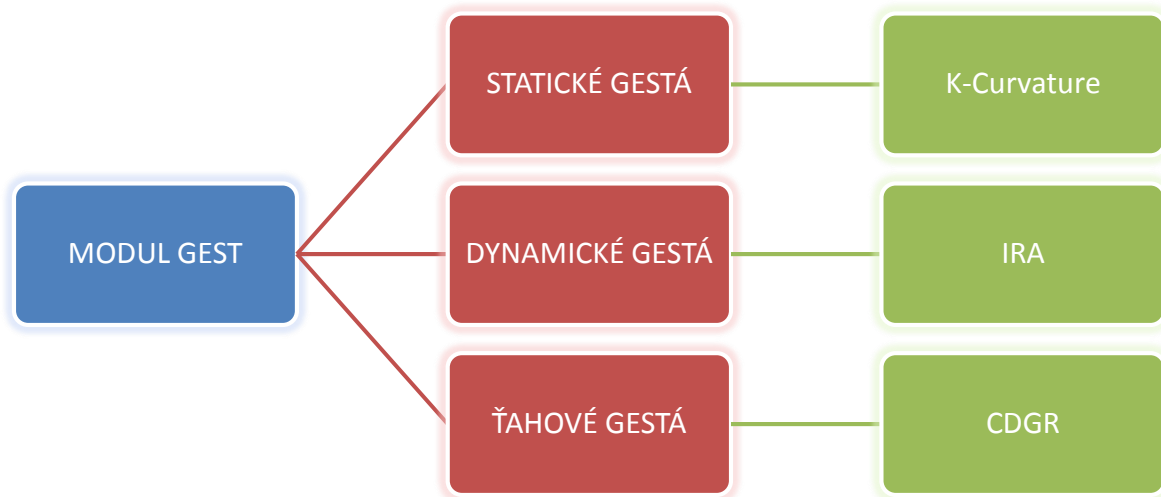
Ďalej je dôležité rozlišovať schopnosť používania gest podľa vzdialenosti od TV a aktuálnej pózy používateľa. To znamená, nemôžeme vyžadovať od používateľa vykonávanie precízneho gesta odmykania obsahu z veľkej vzdialenosti, pokiaľ nám to kvalita hardvéru nedovoľuje. Zároveň si treba uvedomiť, že aj póza používateľa pôsobí na komfort zadávaného gesta. Určite najpohodlnejšie zadávanie gest je v stoji. Vtedy nie sme limitovaný žiadnym predmetom či operadlom pohovky. Diagram nižšie (Obr. 9 Závislosť pózy a gesta) ukazuje, aké možnosti ponúkajú jednotlivé pózy pri vykonávaní gesta.



Obr. 9 Závislosť pózy a gesta

Modul rozpoznávania gest je logický celok, ktorý je rozdelený v skutočnosti na tri funkčné celky: modul rozpoznávania statických gest, modul rozpoznávania dynamických gest

a modul ťahových gest. Pre každý funkčný blok bol implementovaný jeden algoritmus, ako to ukazuje obrázok nižšie.



Obr. 10 Použité metódy pre kategórie gest

Účel modulu statických gest v systéme je mnohoraký. Pod pojmom statické gesto sa rozumie gesto, ktoré sa nemení v čase a jeho význam je jednoznačný. V tomto koncepte sa statické gestá vyšetrojú pomocou prstov a dlane. Počet ukázaných prstov pravej a ľavej ruky reprezentuje statické gesto, pričom ľubovoľnou kombináciou počtu prstov pravej a ľavej ruky vieme dostať väčšie množstvo možností pre statické gesto. Ak sa statické gesto skombinuje napríklad s rotáciou dlane, je možné vytvoriť prirodzené ovládanie hlasitosti zvuku, či intenzity osvetlenia natáčaním dlane. Prípadným ukázaním počtu prstov je možnosť sa pohotovo prepnúť na požadovanú aplikáciu. Ak sa pozrieme na tabuľku porovnania algoritmov pre statické gestá (Tab. 1), pravdepodobne by sme sa rozhodli pre algoritmus analýzy krivky, nakoľko je veľmi robustný a dokonca sa vie vysporiadať so spojenými prstami. Avšak treba pripomenúť, že implementácia pre multimedialnú TV úzko súvisí s hardvérovými požiadavkami. Z toho dôvodu bola do systému MMI zaintegrovaná metóda K-Curvature.

Modul dynamických gest slúži na rozpoznávanie gest, kde dôležitým faktorom je trajektória vykonaného gesta. Samotné rozpoznávanie trajektórie je spustené na začiatku statickým gestom (otvorená dlaň s piatimi prstami), ktoré je detegované statickým

algoritmom, a ukončené uzavretím dlane. Táto trajektória sa následne porovnáva s databázou všetkých vzorov a vyberá sa vzor s najvyššou pravdepodobnosťou. Tento modul obsahuje aj tréningovú časť, kde je možné nahráť ľubovoľné gesto do databázy a priradiť mu úlohu. Dynamické gestá sa používajú v rôznych situáciách ako napríklad na uzamykanie súkromného obsahu, kde používateľ je vyzvaný zadať dynamické gesto a ak sa jeho vykonané dynamické gesto podobá gestu zadanému na otvorenie prístupu, používateľ získa prístup.

Ťahové gestá a ich implementácia korení z poznatkov rôznych výskumov a z riešení komerčných spoločností. Ťahové gestá majú vysokú mieru pohodlia, vysokú presnosť a prirodzenosť. Z toho dôvodu je ich využitie presne zacielené na ovládacie prvky zodpovedajúce daným pohybom. Pri ovládaní multimédií sa používajú na prepínanie kanálov (ťah vľavo či ťah vpravo), maximalizovanie či minimalizovanie prezeraného obsahu na obrazovke (spojením či roztiahnutím rúk). Samozrejme je možné funkcionality mapovať oddelene pre pravú ruku (ťah doprava, ťah doľava, ťah hore, ťah dolu) a zvlášť pre ľavú ruku, kedy získame niekoľko možných kombinácií pre používateľa.

S gestami sú spojené a náhodné detekcie, ktoré predstavujú nežiadany alebo neúmyselný povel, nakoľko systém musí vedieť počiatok a koniec gesta. Pri statických gestách nie je problém sa vysporiadať s takouto situáciou, keďže statické gesto vyžaduje konkrétnu pózu. Väčší problém nastáva u dynamických gest, kedy ľubovoľný pohyb ruky môže znamenať gesto. Tento problém bol vyriešený prístupom použitia kombinácie statických a dynamických gest. Používateľ začiatok gesta naznačí ukázaním konkrétneho počtu prstov a následným vykonaním pohybu. Po zatvorení dlane sa sledovanie trajektórie ukončí a začne sa analýza.

Ťahové gestá majú obdobný problém ako gestá dynamické. Tu bol problém náhodnou detekciou vyriešený pomocou detekčnej zóny. Jej veľkosť a vzdialenosť od tela sa počíta relatívne voči postave. Vytvorením detekčnej zóny získavame trojrozmerný priestor, v ktorom ruka môže gesto vykonávať. Ďalej sa dajú neúmyselné detekcie potlačiť zadaním minimálnej relatívnej rýchlosti gesta, ktorá definuje minimálnu kinetiku. Tu treba brať do úvahy minimálnu a maximálnu kinetiku.

Neposledným, ale veľmi dôležitým faktorom je zohľadňovanie svetelných podmienok. Pri používaní gest s multimediálnou TV sa tomuto problému nevyhneme. Keďže televízna

obrazovka svojim svetelným vyžarovaním radikálne mení svetelné podmienky a to najmä pri nízkom okolitom svetle, nie je možné používať senzory založené na princípe RGB kamery. V takomto prípade by žiaden algoritmus nebol úspešný. Riešením je použitie senzorov pracujúcich v oblasti infračerveného svetla ako je to uvádzané v práci.

5. VÝSLEDKY DIZERTAČNEJ PRÁCE

1. Boli navrhnuté tri algoritmy na rozpoznávanie statických gest a bolo vytvorené ich vzájomné porovnanie.
2. Algoritmus statických gest bol obohatený o detekciu spojených prstov.
3. Bol navrhnutý algoritmus pre rozpoznávanie intuitívnych gest pomocou jednej ruky tak aby mohol byť využitý pri navigácii multimediálnej TV.
4. Algoritmus intuitívnych gest bol modifikovaný na prácu gest oboch rúk súčasne.
5. Pre navrhnuté algoritmy bola na základe implementácie do systému multimodálneho rozhrania vytvorená metodika používania.

6. PREZENTÁCIA DOSIAHNUTÝCH VÝSLEDKOV

Začlenenie vyvinutých algoritmov rozpoznávania gest do projektu HBB-Next (7. rámcový projekt) znamenalo, pripraviť komplexné a funkčné aplikačné riešenie, prezentovateľné na medzinárodných fórach. Navrhnuté algoritmy okrem prezentácií na medzinárodných konferenciách boli demonštrované na nasledujúcich podujatiach:

1. **IBC Amsterdam (International Broadcasting Convention)** – celosvetová výstava televíznych technológií, kde v rámci demonštrácie výsledkov projektu HBB-Next bolo prezentované aj multimodálne rozhranie ovládané gestami.
2. **Finálny Workshop HBB-Next konzorcia (Mníchov)** – demo ukážka multimodálneho rozhrania ovládaného gestami pre všetkých členov medzinárodného konzorcia HBB-Next.
3. **Medzinárodné konferencie IWSSIP (Dubrovnik) a ELMAR (Zadar)** – ukážky funkcionalít multimodálneho rozhrania a ovládania pomocou gest v rámci medzinárodnej konferencie.
4. Integrovanie algoritmov do laboratória **Smart Room na STU**.
5. Prezentácia funkcionalít pre **vedeckú reláciu RTVS**.

7. ZÁVER

V mojej dizertačnej práci som sa venoval analýze a návrhu bezdotykových gest pre multimediálnu TV. V práci je popísaný všeobecný prehľad v oblasti gest, problematika detekcie gest, sledovania rúk ale aj rozpoznávacích prístupov. Ďalej je v práci predstavená a popísaná problematika snímacích zariadení, ktoré môžu byť použité pre rozpoznávacie algoritmy.

Výsledkom mojej práce je návrh rozpoznávacích algoritmov pre statické gestá, dynamické gestá a ťahové gestá. Pri statických gestách boli navrhnuté tri algoritmy. Všetky tri algoritmy pre rozpoznávanie statických gest boli implementované v jazyku C# a následne testované. Pri testovaní všetkých troch algoritmov som sa zameral nielen na celkovú úspešnosť rozpoznávania, ale boli vyhodnotené faktory ako stupeň voľnosti pri používaní gesta, či aspekt detekcie spojených prstov. Pri riešení problému intuitívnych gest som využil vlastnosti ťahových gest, ktoré poskytujú dostatočnú voľnosť a jednoduchosť používania. Tu bol navrhnutý vlastný algoritmus s názvom CDGR (Circle Dynamic Gesture Recognition). Jeho úspešnosť dosiahla hranicu 94%, čo sa v konečnom nasadení v multimediálnej TV ukázalo ako veľmi uspokojivé. V oblasti dynamických gest bol navrhnutý algoritmus, ktorého princípy majú počiatok v dotykových displejoch. Tu sa podarilo využitím základných princípov rozpoznávania gest na dotykovom displeji, vytvoriť algoritmus s úspešnosťou 91%. Podarilo sa modifikovať prístup tak, aby dokázal rozpoznávať bezdotykové gestá snímané infračerveným senzorom. Pri dynamických gestách bol navrhnutý a implementovaný princíp definovania vlastného gesta. Ďalej sa v práci hovorí o problematike HBB-Next a začlenení mojich výsledkov do tohto európskeho projektu. V systéme MMI boli mnou navrhnuté algoritmy použité nielen na navigáciu multimediálnej TV, ale aj na prístup k uzamknutým obsahom používateľa, kedy gesto slúži ako zámok. Na základe implementácie a testovania nasadených algoritmov bola vytvorená metodika používania gest pre multimediálnu TV.

Rozpoznávanie gest prináša dnešnému svetu silný nástroj na komunikáciu medzi človekom a strojom. Čas ukáže, kam až siahla ľudská túžba po objavoch a čo všetko prinesú ďalšie pokroky v tejto oblasti.

8. PUBLIKÁCIE

- [1] **AFC** VANČO, Marek - MINÁRIK, Ivan - ROZINAJ, Gregor. Gesture Identification for System Navigation in 3D Scene. In *Proceedings ELMAR-2012 : 54th Symposium ELMAR-2012, 12-14 September 2012 Zadar, Croatia*. Zadar : Croatian Society Electronics in Marine, 2012, s.45-48. ISBN 978-953-7044-13-8.
- [2] **AFD** ŠTOFAŇÁK, Martin - ČAVOJSKÝ, Maroš - VANČO, Marek. Image Preprocessing for 3D Face Recognition using Kinect. In *Redžúr 2013 : proceedings; 7th International Workshop on Multimedia and Signal Processing; Smolenice, Slovakia, 1. Máj 2013*. 1. vyd. Bratislava : Nakladateľstvo STU, 2013, s.57-60. ISBN 978-80-227-3921-4.
- [3] **AFD** VANČO, Marek - ROZINAJ, Gregor. Výslovnosť čísloviek pri syntéze reči. In *ŠVOČ 2011 : študentská vedecká a odborná činnosť. Zborník vybraných prác. Bratislava, Slovak Republic, 4.5.2011*. Bratislava : STU v Bratislave FEI, 2011, s.646-651. ISBN 978-80-227-3508-7.
- [4] **AFD** VANČO, Marek - ROZINAJ, Gregor. Pronunciation of Numerals in Speech Synthesis. In *Redžúr 2011 : proceedings; 5th International Workshop on Multimedia and Signal Processing. May 12, 2011, Bratislava, Slovak Republic*. Bratislava : Nakladateľstvo STU, 2011, s.89-92. ISBN 978-80-227-3506-3.
- [5] **AFC** MOLČAN, Ľubomír - VANČO, Marek - TÓTH, Ján. Touchless Computer Control with Hands. In *Redžúr 2012 : proceedings; 6th International Workshop on Multimedia and Signal Processing. April 11, 2012, Vienna, Austria*. Bratislava : Nakladateľstvo STU, 2012, s.77-80. ISBN 978-80-227-3686-2.
- [6] **AFD** VANČO, Marek - MINÁRIK, Ivan - ROZINAJ, Gregor - PAVLOVIČOVÁ, Jarmila. Robustné rozpoznávanie gest pre bezdotykové multimediálne prostredie. In *EE časopis pre elektrotechniku, elektroenergetiku, informačné a komunikačné technológie : konferencia ELOSYS, Trenčín, 15.-18.10.2013*. Roč. 19, mimoriadne č (2013), s.263-266. ISSN 1335-2547.
- [7] **AFD** MOLČAN, Ľubomír - MITRO, Martin - VANČO, Marek. Gesture Recognition for HBB-Next. In *Redžúr 2013 : proceedings; 7th International Workshop on Multimedia and Signal Processing; Smolenice, Slovakia, 1. Máj 2013*. 1. vyd. Bratislava : Nakladateľstvo STU, 2013, s.49-52. ISBN 978-80-227-3921-4.

- [8] **AFC** VANČO, Marek - MINÁRIK, Ivan - ROZINAJ, Gregor. Evaluation of static hand gesture algorithms. In *IWSSIP 2014 : 21st International Conference on Systems, Signals & Image Processing. Dubrovnik, Croatia, May 12-15, 2014*. Zagreb : University of Zagreb, 2014, s. 83-86. ISBN 978-953-184-191-7.
- [9] **AFC** VANČO, Marek - MINÁRIK, Ivan - ROZINAJ, Gregor. Dynamic Gesture Recognition for Next Generation Home Multimedia. In *Proceedings ELMAR-2013 : 55th International Symposium. Zadar, Croatia, 25-27 September 2013*. Zadar : Croatian Society Electronics in Marine, 2013, s.219-222. ISBN 978-953-7044-14-5.
- [10] **AFC** VANČO, Marek - MINÁRIK, Ivan - RYBÁROVÁ, Renáta. Evolution of static gesture recognition. In *Redžúr 2014 [elektronický zdroj] : proceedings; 8th International Workshop on Multimedia and Signal Processing; 13 May 2014, Dubrovnik, Croatia*. Bratislava : Nakladateľstvo STU, 2014, CD-ROM, p. 41-44. ISBN 978-80-227-4162-0.
- [11] **AFC** ROZINAJ, Gregor - RYBÁROVÁ, Renáta - VANČO, Marek. Interactive speech synthesizer. In *Proceedings ELMAR-2014 : 56th International Symposium. Zadar, Croatia. 10-12 September 2014*. Zagreb : University of Zagreb, 2014, S. 51-54. ISBN 978-953-184-199-3.
- [12] **AFD** ŠUGÁR, Matúš - VANČO, Marek. Identification of users based on MAC addresses of their devices. In *Redžúr 2015 [elektronický zdroj] : 9th International workshop on multimedia and signal processing. Smolenice, Slovakia. April 22-23, 2015*. 1. vyd. Bratislava : Nakladateľstvo STU, 2015, CD ROM, S. 99-102. ISBN 978-80-227-4346-4.
- [13] **AFD** ŠUGÁR, Matúš - VANČO, Marek. Identifikácia používateľov na základe fyzických adries ich zariadení. In *ELOSYS. Elektrotechnika, informatika a telekomunikácie 2015 [elektronický zdroj] : Konferencia s medzinárodnou účasťou. Trenčín, Slovakia. 13. – 15. október 2015*. 1. vyd. Bratislava : Nakladateľstvo STU v Bratislave, 2015, CD-ROM, S. 178-181. ISBN 978-80-227-4437-9.

- [14] **AFD** LUŇÁČEK, Branislav - RYBÁROVÁ, Renáta - VANČO, Marek. Wireless configuration using smart phone application. In *Redžúr 2016 [elektronický zdroj] : 10th International workshop on multimedia and signal processing. Bratislava, Slovakia. May 24, 2016.* 1. ed. Bratislava : Nakladateľstvo STU, 2016, CD-ROM, S. 85-90. ISBN 978-80-227-4560-4.
- [15] **ADE** ROZINAJ, Gregor - VANČO, Marek – MINÁRIK, Ivan – DROZD, Ivan - RYBÁROVÁ, Renáta. Extending System Capabilities with Multimodal Control. In *Acta Polytechnica Hungarica, Vol. 13, No. 4, 2016.*, S. 185-204. ISSN 1785-8860.

9. ZOZNAM POUŽITEJ LITERATÚRY

- [1] S. Mitra and T. Acharya, "Gesture recognition: A survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 3, pp. 311–324, 2007.
- [2] A. Bourke, J. O'Brien, and G. Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm," *Gait & posture*, vol. 26, no. 2, pp. 194–199, 2007.
- [3] X. Zabulis, H. Baltzakis, and A. Argyros, "Vision-based hand gesture recognition for human-computer interaction," *The Universal Access Handbook. LEA*, pp. 34–1, 2009.
- [4] M. Côté, P. Payeur, and G. Comeau, "Comparative study of adaptive segmentation techniques for gesture analysis in unconstrained environments," in *IEEE Int. Workshop on Imaging Systems and Techniques*, 2006, pp. 28–33.
- [5] I. GLADIŠOVÁ and J. MIHALÍK, "Aplikácia vybraných morfológických transformácií na spracovanie obrazu."
- [6] K. Homma and E. Takenaka, "An image processing method for feature extraction of space-occupying lesions.," *Journal of nuclear medicine: official publication, Society of Nuclear Medicine*, vol. 26, no. 12, pp. 1472–1477, 1985.
- [7] T. Koma, "Ovládanie počítača pohybom - Bakalárska práca." 2013.
- [8] D. Brownrigg, "The weighted median filter," *Communications of the ACM*, vol. 27, no. 8, pp. 807–818, 1984.
- [9] M. Bhuyan, D. R. Neog, and M. K. Kar, "Fingertip detection for hand pose recognition," *International Journal on Computer Science and Engineering*, vol. 4, no. 3, pp. 501–511, 2012.
- [10] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 1093–1096.
- [11] P. O. Kristensson and L. C. Denby, "Continuous recognition and visualization of pen strokes and touch-screen gestures," in *Proceedings of the Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling*, 2011, pp. 95–102.
- [12] P.-O. Kristensson and S. Zhai, "SHARK 2: a large vocabulary shorthand writing system for pen-based computers," in *Proceedings of the 17th annual ACM symposium on User interface software and technology*, 2004, pp. 43–52.
- [13] W. Niblack and J. Yin, "A pseudo-distance measure for 2D shapes based on turning angle," in *Image Processing, 1995. Proceedings., International Conference on*, 1995, vol. 3, pp. 352–355.