



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

OPTICKÉ METODY ROZEZNÁNÍ GEST

OPTICAL METHODS OF GESTURE RECOGNITION

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

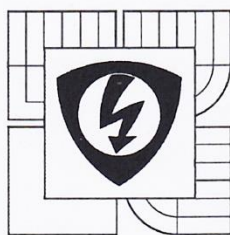
Bc. Jan Netopil

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Vratislav Čmiel

BRNO 2016



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav biomedicínského inženýrství

Diplomová práce

magisterský navazující studijní obor
Biomedicínské inženýrství a bioinformatika

Student: Bc. Jan Netopil

Ročník: 2

ID: 147525

Akademický rok: 2015/16

NÁZEV TÉMATU:

Optické metody rozeznání gest

POKYNY PRO VYPRACOVÁNÍ:

1) Prostudujte metody hodnocení gest pomocí obrazové analýzy. Seznamte se s různými aplikacemi využití optických metod k rozeznání gest. Zaměřte se zejména na IR oblast. Sestavte literární rešerši. 2) Navrhňte metodu detekce a hodnocení gest různé úrovně ve VIS a v IR oblasti s využitím optické kamery, IR obrazového senzoru a výpočetní jednotky. 3) Navrhňte testovací skupinu gest určených pro klasifikaci. Navrhňte vhodný algoritmus pro detekci a rozeznání gest ve VIS a IR oblasti. 4) Realizujte navrženou metodu a metodu ověřte na testovacích gestech. 5) Porovnejte rozdíly mezi hodnocením gest ve VIS oblasti a IR oblasti, zejména z hlediska HW a SW náročnosti a účinnosti metod. 6) Dosažené výsledky zhodnoťte a proveďte diskusi.

DOPORUČENÁ LITERATURA:

[1] WACHS J. P., M. KOLSCH, H. STERN a Y. EDAN. Vision-based hand-gesture applications. Communications of the ACM, vol. 54, no. 2, pp.60 -71, 2011.

[2] FREEMAN W. T. a C. D. WEISSMAN. Hand gesture machine control system. Computer Integrated Manufacturing Systems, vol. 10, no. 2, str. 175 -175, 1997.

Termín zadání: 8. 2. 2016

Termín odevzdání: 20. 5. 2016

Vedoucí práce: Ing. Vratislav Čmiel

Konzultanti diplomové práce:



prof. Ing. Ivo Provazník, Ph.D.

předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

ABSTRAKT

Tato práce se zabývá optickými systémy a metodami zpracování získaných dat pro rozeznání gest ruky. V rámci práce je popsáno rozdělení gest, možnosti využití rozeznání gest v různých oblastech a dostupné kontaktní a bezkontaktní technologie. Následně jsou představeny metody pro detekci ruky, získání příznaků a klasifikaci gest. Navržený systém pro rozeznání gest ruky se skládá z kamery FLIR A655sc a modulu FLIR Lepton, a webové kamery Logitech S7500, dále metody pro zpracování nasnímaných dat a databáze gest, sloužící pro závěrečnou klasifikaci. Pomocí vytvořené aplikace v MATLABU je pro každé zařízení realizováno hodnocení rozpoznání gest a to s ohledem na rychlost a úspěšnost rozpoznání v různých prostředích.

KLÍČOVÁ SLOVA

Rozeznání gest ruky, segmentace ruky, extrakce příznaků, FLIR kamery, termovizní systémy

ABSTRACT

This thesis deals with optical devices and methods image processing for recognizing hand gestures. The types of gestures, possible applications, contact based devices and vision based devices are described in thesis. Next, a review of hand detection, features extraction and gesture classification is provided. Proposed gesture recognition system consists of infrared camera FLIR A655sc, infrared FLIR Lepton module, webcam Logitech S7500, method for hand gesture analysis and a database of gestures for classification. For each of the devices, gesture recognition is evaluated in terms of speed and accuracy in different environments. The proposed method was implemented in MATLAB.

KEYWORDS

Hand gesture recognition, hand segmentation, feature extraction, FLIR cameras, thermal imaging systems

NETOPIL, J. *Optické metody rozeznání gest*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Ústav biomedicínského inženýrství, 2016. 69 s., Diplomová práce. Vedoucí práce: Ing. Vratislav Čmiel.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma Optické metody rozeznání gest jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Mé vřelé poděkování patří vedoucímu diplomové práce Ing. Vratislavu Čmielovi za odborné vedení, cenné rady a vstřícnost při konzultacích. Děkuji také Ing. Vratislavu Harabišovi, Ph.D. za pomoc s přípravou modulu FLIR Lepton a Ing. Marině Ronzhině za věcné připomínky k neuronovým sítím.

V Brně dne

.....

(podpis autora)

OBSAH

Seznam obrázků	viii
Seznam tabulek	xi
Úvod	1
1 Gesta a jejich využití	2
1.1 Definice a členění gest	2
1.2 Aplikace využívající rozpoznání gest	4
2 Technologie pro rozeznání gest	9
2.1 Kontaktní zařízení pro rozpoznání gest	10
2.2 Optická zařízení pro rozpoznání gest	11
3 Metody používané pro rozpoznání gest ruky	15
3.1 Extrakční metody a předzpracování obrazu	15
3.2 Extrakce příznaků	20
3.3 Klasifikace gest	22
4 Hardware pro rozpoznání gest	28
4.1 Hardwarové možnosti pro optické rozpoznání gest	28
5 Navržený Algoritmus	33
5.1 Kalibrace	34
5.2 Detekce ruky	36
5.3 Úpravy po segmentaci	37
5.4 Úpravy pro Hammingovu síť	38
5.5 Hammingova síť	44
5.6 Sestavení databáze gest	45

6	Hodnocení	46
6.1	Ověření algoritmu na testovaných gestech	46
6.2	Srovnání algoritmu s ostatními přístupy a hodnocení zvoleného přístupu	49
6.3	Volba hardwaru, gest a parametry snímků pro hodnocení	51
6.4	Rozsah vzdáleností ruky od objektivu u jednotlivých zařízení	52
6.5	Rychlost a úspěšnost rozpoznání gest v ideálním prostředí při různých nastaveních	54
6.6	Problematická prostředí pro rozpoznání gest pro infračervenou oblast	58
6.7	Problematická prostředí pro rozpoznání gest pro viditelnou oblast	62
6.8	Srovnání výsledků hodnocení v infračervené a viditelné oblasti	65
7	Závěr	67
	Literatura	69
	Seznam symbolů, veličin a zkratk	74
	Seznam příloh	75

SEZNAM OBRÁZKŮ

Obr. 1.1:	Kategorie pro rozdělení jednotlivých gest [24]	4
Obr. 1.2:	Operátér používá systém Gestix pro procházení MRI obrázků [53]	5
Obr. 2.1:	Systém sloužící pro rozpoznání gest (převzato a upraveno z [4])	9
Obr. 2.2:	TOF systém měří čas návratu pozorováním fázového posunu signálu vyslaného a signálu přijatého (převzato a upraveno z [40])	13
Obr. 2.3:	Metoda strukturovaného světla – Strukturované světlo je promítáno na objekt, na kterém se objeví při pohledu ze strany kontury. (převzato a upraveno z [40]).....	14
Obr. 3.1:	Vyhledávání vzorů: a) zvolený vzor; b) správné a falešné detekce prstů; c) zvolené okno prohledávání [28].....	20
Obr. 3.2:	Použití kontury a konvexní obálky pro hledání příznaku d (převzato a upraveno z [36]).....	21
Obr. 3.3:	Získání příznaků z kontury pro hledání prstů [48].....	21
Obr. 3.4:	Rostoucí SGONG síť: a) počáteční stav, b) rostoucí síť, c) konečný výstup sítě neuronů [47]	22
Obr. 3.5:	Schéma neuronu [31]	24
Obr. 3.6:	Perceptronová vícevrstvá síť [31].....	25
Obr. 3.7:	Topologie Hammingovy sítě; $x_1 \dots x_n$ – binární vstupní vektor $\{-1,1\}$, w_m – váha pro spojení mezi neuronem m a vstupem n (váhy odpovídají zapamatovaným binárním vzorům $\{-1,1\}$), $y_1 \dots y_m$ – výstupy jednotlivých neuronů, z – výsledný výstup ze sítě (maximum z jednotlivých výstupů neuronů, minimum Hammingovy vzdálenosti) [43].....	26
Obr. 3.8:	Pětistavový levopravý HMM model, s – jednotlivé stavy, a – přechody mezi stavy [38]	27
Obr. 4.1:	Pořízené snímky gest infračervenou kamerou FLIR i7	29
Obr. 4.2:	Pořízené snímky gest infračervenou kamerou FLIR A655sc	30
Obr. 4.3:	Pořízené snímky gest modulem FLIR Lepton	31
Obr. 4.4:	Pořízené snímky gest kamerou Logitech S7500	32

Obr. 5.1:	Algoritmus pro rozpoznání gest.....	33
Obr. 5.2:	Aplikační rozhraní pro kalibraci	35
Obr. 5.3:	Snímek získaný webovou kamerou Logitech a zobrazení jeho jednotlivých složek v různých barevných prostorech.....	36
Obr. 5.4:	Snímky s postupnými úpravami po segmentaci (získáno webovou kamerou Logitech S7500).....	38
Obr. 5.5:	Obrázek ruky po počáteční rotaci, ořezu a otočení (bílá barva). Šedá barva zobrazuje rotovanou původní ruku. Snímek před zpracováním byl nasnímán kamerou FLIR A655sc v rozlišení 640 x 480 pixelů.....	40
Obr. 5.6:	Ukončení eroze – bílá barva označuje výslednou erozi; vlevo je objekt rozdělen na dva vlivem zápěstí, vpravo je eroze ukončena po dosažení dostatečně malého objektu. Původní snímek byl nasnímán kamerou FLIR A655sc v rozlišení 640 x 480 pixelů.....	41
Obr. 5.7 :	Získané hodnoty pro několik variant předloktí – vypočtená výsledná hodnota (rozdíl vypočítaný algoritmem) a zjištěná požadovaná velikost ořezu (manuálně) a proložená funkce	42
Obr. 5.8 :	Získané hodnoty pro několik variant předloktí – Vypočtené hodnoty poměru (algoritmem), zjištěné velikosti ořezu (manuálně) a proložená lineární funkce	43
Obr. 5.9:	Navržená databáze gest s pojmenováním gest dle pořadí sloužící pro klasifikaci (získané snímky jsou z kamery FLIR A655sc)	45
Obr. 5.10:	Vliv rotace gesta na výsledné zpracování obrázku pro čtvrté gesto. Lze si povšimnout i mírně jiného úhlu palce mezi zbytkem ruky, což má v součtu nepříznivý dopad na výsledné rozpoznání.....	48
Obr. 5.11:	Znázornění provedených gest s různým natočením, vzdáleností od kamery a s rozdílnou délkou předloktí zasahující do scény. Všechna zobrazená gesta byla rozpoznána a správně zařazena	49
Obr. 6.1 :	Gesta vybraná pro hodnocení algoritmu – gesto jedna, dva, sedm a deset..	51
Obr. 6.2:	Zvolené prostředí pro jednotlivá zařízení, kde probíhalo rozpoznání gest (snímky nahoře). Spodní snímky zobrazují maximální vzdálenost prováděných gest pro jednotlivá zařízení.	52
Obr. 6.3:	Problematické prostředí pro IR oblast obsahující v pozadí napájecí zdroje. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po	

- segmentaci prahováním a na třetím je vyobrazen snímek po úpravách..... 58
- Obr. 6.4: Problematické prostředí pro IR oblast obsahující v pozadí 3D tiskárnu. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po segmentaci prahováním a třetím je vyobrazen snímek po úpravách 59
- Obr. 6.5: Problematické prostředí pro IR oblast obsahující v pozadí rozsvícenou lampu. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po segmentaci prahováním a třetím je vyobrazen snímek po úpravách 60
- Obr. 6.6: Problematické prostředí pro IR oblast obsahující v pozadí aktivní reproduktor a tablet. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po segmentaci prahováním a třetím je vyobrazen snímek po úpravách 61
- Obr. 6.7: Nedostatečně prokrvená ruka a první tři varianty gesta..... 62
- Obr. 6.8: Problematické prostředí pro VIS oblast obsahuje v pozadí červenou sedačku, list rostliny a obrázek. První snímek zobrazuje zvolené prostředí, snímek napravo pak segmentaci v normalizovaném prostředí RGB a složek R,G,B a poslední snímek pak výsledek po úpravách..... 62
- Obr. 6.9: Druhé problematické prostředí: První snímek v horním řádku zobrazuje zvolené prostředí, snímky napravo pak segmentaci v normalizovaném RGB všech tří složek a následné úpravy. První snímek ve spodním řádku zobrazuje segmentaci v YC_bC_r prostoru a to dle složek C_bC_r , další pak snímek po úpravách a poslední snímek pak segmentaci získanou kombinací složek R a C_r (zde jsou nastavené prahy odlišné jako v přechozí variantě barevných prostorů zvlášť) 63
- Obr. 6.10: Třetí problematické prostředí: První snímek zobrazuje zvolené prostředí, snímky napravo pak segmentaci v normalizovaném RGB všech tří složek a následné úpravy. První snímek ve spodním řádku zobrazuje segmentaci v YC_bC_r prostoru a to složek C_bC_r , další pak snímek po úpravách a poslední snímek pak segmentaci získanou kombinací složek R a C_r (zde jsou nastavené prahy odlišné jako v přechozí variantě barevných prostorů zvlášť) 64

SEZNAM TABULEK

Tab. 4.1:	Parametry infračervené kamery FLIR i7 [50].....	28
Tab. 4.2:	Parametry infračervené kamery FLIR A655sc [51].....	29
Tab. 4.3:	Parametry modulu FLIR Lepton	30
Tab. 4.4:	Parametry kamery Logitech S7500.....	31
Tab. 5.1:	Doporučené hodnoty pro práh a dynamickou kriteriální váhu	35
Tab. 5.2:	Parametry notebooku	46
Tab. 5.3:	Úspěšnosti rozpoznávaných a správně zařazených gest (vyjádřené v procentech) pro jednotlivé varianty nastavení	48
Tab. 6.1:	Rozsah vzdáleností použitelných pro rozpoznání gest u jednotlivých zařízení	53
Tab. 6.2:	Změřené časy algoritmu (v sekundách) a jeho částí pro různá zařízení a zvolená čtyři gesta při první variantě nastavení.....	55
Tab. 6.3:	Změřené časy algoritmu (v sekundách) a jeho částí pro FLIR A655c a Logitech S7500 a zvolená čtyři gesta při druhé variantě nastavení.....	56
Tab. 6.4:	Změřené časy algoritmu (v sekundách) a jeho částí pro různá zařízení a zvolená čtyři gesta při třetí variantě nastavení.....	57
Tab. 6.5:	Naměřené časy (v sekundách) a úspěšnost algoritmu.....	59
Tab. 6.6:	Naměřené časy (v sekundách) a úspěšnost algoritmu.....	60
Tab. 6.7:	Naměřené časy (v sekundách) a úspěšnost algoritmu.....	60
Tab. 6.8:	Naměřené časy (v sekundách) a úspěšnost algoritmu.....	61
Tab. 6.9:	Naměřené časy (v sekundách) a úspěšnost algoritmu.....	63

ÚVOD

Tato diplomová práce se věnuje optickým metodám rozeznání gest. Cílem práce je seznámit se s metodami hodnocení gest pomocí obrazové analýzy a s různými aplikacemi využití těchto metod. Dále je cílem navrhnout a realizovat metodu detekce a hodnocení gest ve viditelné a infračervené oblasti spektra s využitím optické kamery, IR obrazového senzoru a výpočetní jednotky a vytvořit vhodnou databázi gest, která bude sloužit pro klasifikaci analyzovaných gest. Na závěr je nutné provést hodnocení metody ve viditelné a infračervené z hlediska hardwarové a softwarové náročnosti a účinnosti metod a výsledky zhodnotit.

Nejprve je kvůli komplexnosti celé problematiky představena kapitola věnující se samotné definici gest a jejich rozdělení. Pro správné navržení metody je nutné rozlišovat mezi gesty statickými a dynamickými a také dalšími možnými způsoby provádění gest. V následující kapitole věnující se aplikacím využívajících rozpoznání gest jsou podrobněji popsány jednotlivé oblasti možnosti využití jako jsou medicínské a asistenční technologie, počítačové hry a ovládání robotických zařízení. Kapitola Technologie pro rozpoznání gest představuje celý systém sloužící pro zpracování gest, nastiňuje problematiku kontaktních zařízení pro rozpoznání gest a popisuje především optická zařízení v této oblasti se zaměřením na vhodné parametry zařízení. Další kapitola se věnuje metodám zpracování nasnímaných gest popsaných v dostupné literatuře. Jsou zde probrány algoritmy pro detekci ruky, získání příznaků ruky a klasifikaci pro statická gesta a dynamická gesta. Kapitola Hardware pro rozpoznání gest pak popisuje dostupná zařízení a jejich parametry a věnuje se vhodnosti použití zařízení v této oblasti.

Praktická část obsahuje návrh a realizaci celého výpočetního algoritmu v programovém prostředí MATLAB. Algoritmus využívá segmentaci pomocí prahování, kde jsou prahy získány vhodnou kalibrací. Součástí algoritmu jsou i veškeré úpravy prováděné před a po segmentaci, dále pak zpracování snímku pro Hammingovu síť, která jako poslední krok provádí na základě sestavené databáze rozpoznání gest. V poslední kapitole je provedeno hodnocení jednotlivých zařízení s ohledem na rozsah vzdáleností ruky od kamery, rychlost zpracování a rozpoznání gesta a účinnosti rozpoznání v různých prostředích a podmínkách. Závěr obsahuje komentář jednotlivých výsledků a návrh na volbu vhodného zařízení.

1 GESTA A JEJICH VYUŽITÍ

Rozeznání lidských gest v oblasti počítačového zpracování probíhá za pomoci analýzy gest a automatického rozpoznání při použití několika technik a hardwaru (sensorů – kamer, speciálních rukavic atd.). Před samotným automatickým rozpoznáním gest je nutné si definovat samotný pojem gesto, do jakých kategorií je možné gesta začlenit, z jakého důvodu a za jakým účelem jsou gesta používána a jaký druh informace je možné ze samotných gest získat.

1.1 Definice a členění gest

Gesta mohou být různými autory definovány odlišně. Gesto lze vyjádřit jako sérii pozic částí těla v určitém časovém intervalu, nebo jako pohyby částí těla, tváře a ruky jako vyjádření významu nebo záměru komunikujícího [8][27]. Z lingvistického hlediska lze popsat gesta jako pohyby rukou a paží [37]. Gesta budou dále chápána jako pohyb těla či uspořádání částí těla mluvčího, za účelem vyjádření nějaké informace.

Gesta jsou hlavními prostředky neverbální komunikace, doplňují verbální komunikaci, nebo se používají samostatně. Při komunikaci využívá většina jedinců 35 % verbální složky a 65 % neverbální. V neverbální komunikaci bylo vyčísleno na 700 tisíc signálů, které zahrnují kolem padesáti tisíc mimických gest a pět tisíc gest ruky. Na složitosti rozpoznání gest se navíc podílí skutečnost, že význam gest silně závisí na kulturních odlišnostech. Například kývání hlavou téměř v celé Evropě ve smyslu *ano* znamená v Bulharsku *ne*. Gesta je možné členit dle toho, jestli jsou vědomá (úmyslná), nebo nevědomá (reflexy), nebo zda se jedná o dynamická či statická gesta. [24]

Za statické gesto lze pokládat specifickou polohu ruky (také části těla), její orientaci či ohnutí pozorované v určitém čase. Tato gesta mohou být analyzována za použití jednoho nebo série obrázků v konkrétním čase. Jako příklad může být úsměv na tváři, nebo gesto ruky zobrazující znaky *OK* nebo *stop*. [32]

Dynamické gesto lze chápat jako sekvence pozic spojených pohybem v krátkém časovém úseku (například ve video signálu). Dynamickými gesty mohou být například pohyby hlavy vyjadřující *ano* nebo *ne*, pohyby ruky znamenající *na shledanou* nebo *pojd' sem*. [32]

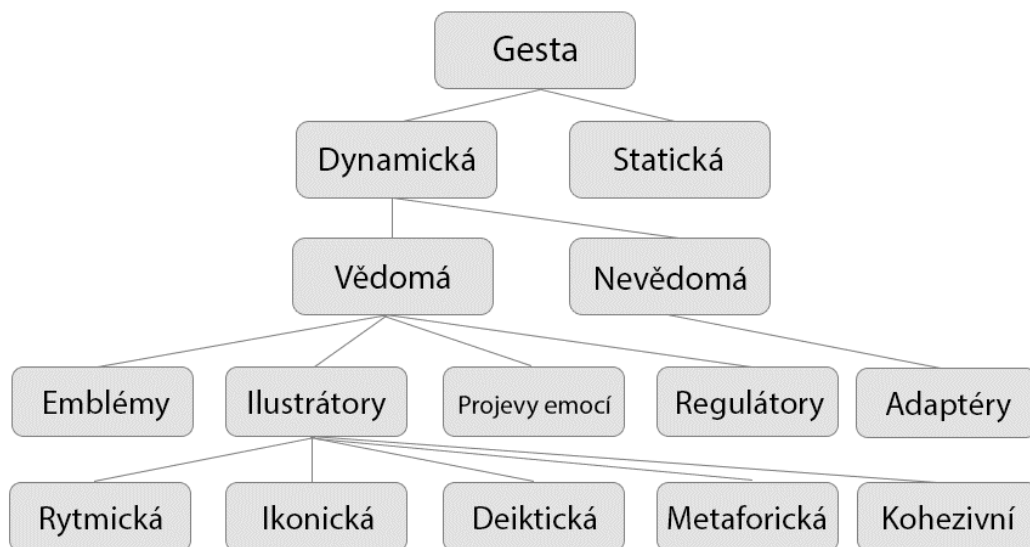
Rozdělení jednotlivých gest lze pozorovat na obrázku (Obr. 1.1). Nejedná se o univerzální dělení, v literatuře lze dohledat více variant členění, ale pro zařazení téměř veškerých gest je toto schéma dostačující. Gesta jsou zde členěna na dynamická a statická a dále na vědomá a nevědomá. [24]

Vědomá se dělí na emblémy, ilustrátory, projevy emocí („affect display“) a regulátory, nevědomá pak na adaptéry. [30]

Emblémy, nebo též sémantická gesta, jsou pohyby lidského těla, které mají svůj přesný verbální ekvivalent (rozsah jednoho či dvou slov). Gesty přenášený význam je pro většinu příslušníků dané kultury dobře známý, ale může mít specifickou platnost v kontextu rozdílných kultur, sociální či geografické oblasti. Používají se především v případech, kdy je znesnadněna verbální komunikace. Všechny emblémy nevyžadují signalizaci za použití prstů, rukou nebo paží, je možné komunikovat pouze pohybem nosu, ramen, či potřesením hlavy.[30]

Ilustrátory jsou ilustrativní gesta spojená s řečí, používaná pro zdůraznění slov a udávající rytmus mluvenému slovu (rytmický typ), ilustrující pomocí kreslení ve vzduchu různé předměty a činnosti (ikonický typ) nebo abstraktní pojmy (metaforický typ), spojující různá témata a prohlášení (kohezivní typ) a vyjadřující někoho nebo něco pomocí posunků a ukazování (deiktický typ). [30][10]

K dalším kategoriím patří projevy emocí (affect display), které jsou nejčastěji vyjadřovány pomocí lidské mimiky, ale může se zde podílet i celá řada jiných gest. Regulátory jsou regulující gesta související s verbálním projevem se vztahem k průběhu vlastní personální interakce. Jedná se tedy o signály změn, které mají často důležitou roli pro započetí nebo ukončení komunikace. Nejčastěji se zde uplatňují pohyby hlavy, pohledy vedené určitým směrem apod. Poslední kategorií jsou nevědomá gesta – adaptéry, což jsou pohyby pro zvládnutí našich pocitů a k řízení reakcí. Objevují se zpravidla ve stresových situacích. [30]



Obr. 1.1: Kategorie pro rozdělení jednotlivých gest [24]

1.2 Aplikace využívající rozpoznání gest

Aplikací, které využívají rozpoznání gest, lze nalézt v odborné literatuře mnoho. Gesta mohou u počítačů nahradit standardní vstupní zařízení jako je myš, klávesnici nebo případně joystick a pomocí gest ruky lze pak ovládat například kurzor myši, manipulovat s grafickými objekty, upravovat dokumenty nebo prohlížet webové stránky. Uplatnění lze nalézt u aplikací pro virtuální realitu. Velký potenciál má využití gest v medicínských systémech a asistenčních technologiích, jako je například ovládání invalidního vozíku či rozpoznání znakové řeči u hluchoněmých. Perspektivní je také využití gest v počítačových hrách, při ovládání robotických zařízení a systémů krizového řízení.

1.2.1 Virtuální realita

Virtuální realita (virtuální prostředí) je technologie, která umožňuje uživateli interagovat se simulovaným prostředím. Tímto prostředím mohou být například v lékařství prostorové modely orgánů nebo celého těla, u počítačových her herní prostředí, v armádě například simulátor létání. Pomocí gest při využití stereoskopického zobrazení je možná realistická manipulace a uspořádání virtuálních objektů.

Hlavním cílem je zde podpora přirozené, efektivní, rychlé a uživatelem přizpůsobitelné interakce. Použití tradiční klávesnice a myši ve virtuální realitě není příliš vhodné. Ideální se jeví zařízení schopné detekovat polohu a orientaci těla (s tím spojená různá gesta), směr očí, řeč a zvuk a aspekty lidského chování. [3]

Interakce s virtuální realitou je v současné době provedena jednoduchým způsobem a to například pomocí 3D myši, space ballu a různých speciálních rukavic. Přírozenějším a efektivnějším mechanismem se zdá být právě použití gest, kde k řízení navigace s 3D objekty lze například použít pohyby ruky dopředu a dozadu, rotaci a její umístěním v prostoru. [3]

1.2.2 Medicínské systémy a asistenční technologie

Aby se zabránilo vážným chybám s možnými fatálními následky, je v medicínské oblasti nutno dbát především na bezpečnost použitých přístrojů. Pro ovládání počítače zdravotnickými pracovníky zůstávají standardními vstupními zařízeními myš a klávesnice. Nedostatkem použití těchto zařízení je možnost šíření infekce z jedné osoby na druhou, což může být na jednotkách intenzivní péče, nebo operačních sálech velice závažný problém. V takovýchto prostorech musí být displeje neprodyšně uzavřeny (z důvodu zabránění nahromadění nečistot), povrch musí být hladký a snadno čistitelný běžnými čisticími prostředky. Pro dodržení těchto přísných hygienických standardů lze s výhodou využít pro ovládání počítačů, zdravotních nástrojů a displejů právě gesta. Například pro snadné bezdotykové procházení obrázků z MRI (Magnetic resonance imaging) na operačním sále bylo vytvořeno zařízení sledující gesta (systém Gestix viz obrázek (Obr. 1.2)), díky němuž se efektivně řeší rychlost procházení snímků a zmíněné zamezení přenosu infekce.[54][53]



Obr. 1.2: Operatér používá systém Gestix pro procházení MRI obrázků [53]

U laparoskopických operací je možnost využít ovládání a kontrolu laparoskopu na

základě příslušného obličejového gesta a to bez další nutnosti zapojení rukou, nohou nebo hlasu. Pro léčbu pacientů s různými pohybovými potížemi, jako je například porucha koordinace ramen, lze provádět biofeedback, který jednotlivá gesta rozeznává. Taktéž pro rehabilitaci může sloužit speciální hmatová rukavice nasazená na ruce, která pomáhá rehabilitovat pacientům v chronické fázi po mrtvici.[54][39]

Rozpoznání gest lze uplatnit také u asistenčních technologií. V případě invalidních vozíků je klasické ovládání pro různé úkony realizováno tlačítky, joysticky a pákami. Při tomto způsobu ovládání musí být hendikepovaný dostatečně pohyblivý. Řešením pro méně pohyblivé může být řízení vozíku pomocí pohybu hlavy, mimickými gesty, otevíráním a uzavíráním úst, jednoduchými gesty ruky či ovládání pomocí hlasu. V případě použití gest u invalidních vozíků nastává problém vysokého stupně volnosti gest, různého osvětlení a měnícího se pozadí. Gesta lze uplatnit u hendikepovaných také pro ovládání domácích přístrojů, osvětlení, teploty místnosti a tísňové linky. [54]

Atraktivním odvětvím může být vývoj rozpoznání gest u znakového jazyka. Znakový jazyk je přirozený komunikační systém tvořený vizuálně – pohybovými prostředky jako jsou tvar, postavení a pohyb ruky, mimika a pozice hlavy, pozice horní části trupu. Znakový jazyk není univerzální a existuje mnoho různých znakových jazyků, které jsou spolu navzájem rozdílné (nesrozumitelné). Jelikož je to jeden z nejpřirozenějších prostředků pro výměnu informací u většiny sluchově postižených, je velkou motivací vyvíjet systémy, které mohou znakový jazyk efektivně zpracovat a podpořit komunikaci mezi sluchově postiženými a zdravou společností. Z důvodu obrovského souboru lidských gest je třeba rozlišovat mezi mnoha gesty, které mohou být velmi podobné (v americké znakové řeči se rozlišuje mezi gesty dle mírně jiného natočení ruky), což je pomocí současných řešení a zařízení prozatím nesnadno řešitelný úkol. [3][5]

1.2.3 Počítačové hry

Použití gest v počítačových hrách je díky současným technologiím a ekonomické návratnosti velmi perspektivní a stává se inovativním rozhraním pro zábavnou formu přirozeného ovládání počítače. Pro hráče je takové rozhraní atraktivní, jelikož se díky němu dokáží lépe ponořit do náročné hry. Jednou z hlavních podmínek funkčnosti herního systému, který je založený na počítačovém vidění a rozeznání gest, je rychlá reakce a robustnost. Vývojáři se tedy zaměřují na sledování gesta a jeho pozice s vysokou snímkovací frekvencí (zpravidla větší než 10 snímků za sekundu). Dalším požadavkem je rozlišení užitečných gest od neúmyslných pohybů. Možným řešením může být vybrání konkrétního gesta nebo slova, které označí začátek a konec rozpoznávání hledaných gest.

Například v bojové hře *Mind Warping*, kde hrají hráči mezi sebou pomocí gest ruky, je problém ohraničením gest řešen vyslovením určitého slova. Nintendo Wii používá jako řešení stisknutí tlačítka na WiiMote ovladači, což je podobné jako u systému „*push to talk*“. Je to ale i drobná nevýhoda, jelikož hráči tento ovladač musí držet stále v ruce. Požadavkem na ovládání her je také intuitivnost. Hry jsou navrženy tak, aby gesta napodobovaly skutečné lidské pohyby jako je u golfu, tenisu či bowlingu (Nintendo Wii, EyeToy od Sony, Kinect sensor od Microsoftu). Potřeba je také přizpůsobit rozhraní uživateli a dávat mu zpětnou vazbu. Před samotnou hrou se uživatelé musejí často naučit gesta ze slovníku, která jsou implementovaná do hry. Trénink obvykle vyžaduje správné provedení gest včetně rychlosti, trajektorie, polohy prstů a držení těla. Zatímco začátečníci potřebují na naučení gest určitý čas, zkušený uživatel ovládá hry přinejmenším stejně rychle, jako by to bylo pomocí klávesnice, myši nebo jiných vstupních zařízení. [54][46][6]

Inteligentní uživatelská herní rozhraní, využívající technologie rozpoznání gest rukou a těla, musí tedy řešit specifické problémy jako je spolehlivost, rychlost, intuitivnost ovládání, omezení falešných detekcí, stanovení počátku a konce gesta a také implementaci přístroje s rozhraním za co možná nejnižší cenu. [54]

1.2.4 Ovládání robotických zařízení

Rozpoznání gest ruky je jednou z metod používaných pro ovládání pohyblivých i nepohyblivých robotů s vhodně umístěnými kamerami. Důležitá gesta mohou být pro zlepšení robustnosti kombinována s hlasovými příkazy. Pohyby ruky lze robot například navigovat v prostoru (ukazující gesto symbolizuje, kam má robot jít) nebo využít pro operační manipulace (uchopení a uvolnění). Ovládání tímto přístupem je vhodné například pro lidi s tělesným postižením, kde jsou jiné přístupy interakce omezené, nebo nemožné. Rozhraní přináší také schopnost ovládání robotů pro začátečníky, pro které je zpočátku těžké řídit roboty pomocí sofistikovaného ovládání. Jako u jiných aplikací je nutné řešit řadu problémů, jako jsou požadavky na rychlost (real-time zpracování), intuitivnost, přesnost (problémem je spleť pozadí, dynamické světelné podmínky, deformovatelný tvar lidské ruky), interakci v prostoru a přizpůsobivost pro uživatele. Významným faktorem pro rozvoj systémů a praktické využití v různých aplikačních doménách je pokles ceny hardwaru a nákladů na zpracování. [54][55]

V literatuře je v této oblasti popsána řada přístupů. Například pro ovládání robotických manipulátorů použil Kawarazaki systém se stereo vizí, který rozpoznává čtyři statická gesta. Při ukázání ukazováčkem na objekt robot objekt detekuje a následně s ním dle pokynů manipuluje. Gesta rukou lze kombinovat například i s pohybem hlavy.

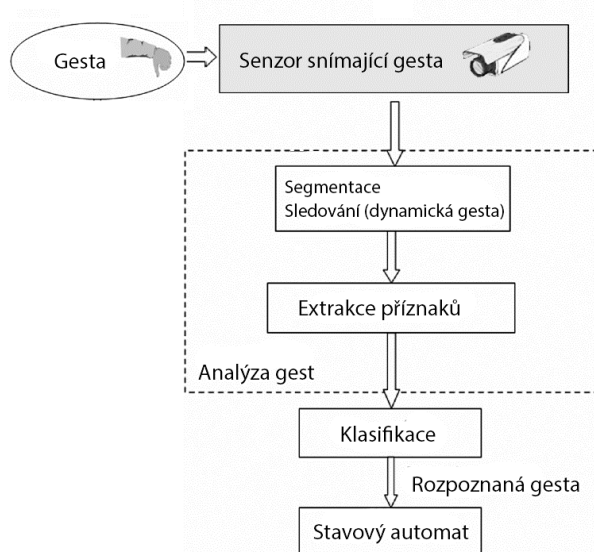
Toho využil Hasanuzzeman, který ovládal robota od Sony pomocí kombinaci osmi gest ruky a gest hlavy (gesta přikývnutí), což mu poskytlo rozsáhlé možnosti ovládání. Gesta v kombinaci s hlasovým vstupem využil Rogall pro robotický asistenční systém. Systém rozpozná gesto, provede kombinaci s hlasovým vstupem a dle toho vykoná příkaz. Slovník gest lze měnit, robot se učí gesta od instruktora s následným použitím pro pohyb či manipulaci. [26][17][42]

1.2.5 Další využití

Systémy krizového řízení pomáhají při řešení krizových situacích, jako jsou přírodní katastrofy (zemětřesení, tornáda) a lidské katastrofy (terorismus, únik jedovatých látek). Reakce na mimořádné události musí být plánovány a koordinovány týmy expertů s přístupem k velkým objemům dat. Pro interface takových systémů je vhodné použít ovládání pomocí gest a to z důvodu rychlosti, intuitivnosti a přirozenosti ovládání, snadného a rychlého dorozumívání a podpory rozhodování mezi řídicími pracovníky a interakci s geoprostorovými informacemi. Další uplatnění gest lze najít při řízení aut kde je cílem snížení rozptylování řidiče. Pro ovládání palubních systému lze využít gesta a řidič se v klidu může soustředit pouze na silnici před sebou. Tyto systémy se prozatím v autech příliš nepoužívají a jsou především ve fázi výzkumu. [54][2]

2 TECHNOLOGIE PRO ROZEZNÁNÍ GEST

Systém sloužící pro rozpoznání gest, který lze sledovat na obrázku (Obr. 2.1), se může skládat z několika částí – prezentace objektu zájmu (gesta), snímacího zařízení (senzoru), analýzy gest (segmentace, sledování gest a extrakce příznaků), samotného rozpoznání gesta (klasifikačního algoritmu) a externí jednotky, která provádí příkazy. Jak bylo zmíněno, gesto je uspořádání nebo pohyb části těla, vyjadřující záměr, příkaz nebo emoci. Sada gest a jejich význam tvoří slovní zásobu. Cílem senzoru je zachytit tyto gesta a převést do digitální podoby. Následuje proces segmentace, což je extrakce objektu zájmu umístěného ve scéně. U dynamických gest lze pak za účelem sledování gesta použít odpovídající algoritmy. Výsledné detekované objekty jsou popsány příznaky. Příznaky jsou dále použity pro klasifikaci gest klasifikačními algoritmy, které zařazují gesta do známých typů tříd. Výsledný stavový automat (externí jednotka) dále používá rozeznané gesto pro změnu elektrického nebo mechanického stavu přístroje. [4]



Obr. 2.1: Systém sloužící pro rozpoznání gest (převzato a upraveno z [4])

Systémy pro rozpoznání gest, umožňující interakci člověka s počítačem, lze rozdělit na dva hlavní typy – kontaktní a bezkontaktní zařízení. Kontaktní zařízení jsou založena na fyzické interakci uživatele se zařízením a uživatel musí být na používání těchto přístrojů navyklý (tedy přístup není uživatelsky přizpůsobivý). Tento typ zařízení může být dále dělen na přístroje mechanické, dotykové, ultrazvukové a magnetické. Bezkontaktní zařízení (optická) jsou založena na počítačovém vidění bez nutnosti fyzické interakce se samotným zařízením. [4][15]

2.1 Kontaktní zařízení pro rozpoznání gest

V poslední době jsou velice rozšířená a populární dotyková zařízení (tablety, mobilní zařízení). Tyto systémy mohou navíc poskytovat i více možností vstupů – dotek, stylus, klávesnici a myš (např. zařízení s Androidem). Pro dotyková zařízení se používá rezistivních a kapacitních dotykových displejů. Rezistivní displeje se využívají v nízkonákladových systémech s malým rozlišením. Dotyková obrazovka má více vrstev, kde jsou vrstvy odděleny malou mezerou. Působením tlaku se vrstvy spojí, vyvolají změnu v odporu a díky tomu je pak určena poloha. Kapacitní displeje se vyznačují větším rozlišením a trvanlivostí. Povrch displeje je pokrytý vodivou vrstvou a při dotyku prstem ruky vznikne mezi okraji displeje a vodivou rukou kapacita, přes kterou se uzavře elektrický obvod. Analýzou kapacit je pak určena poloha prstu. Nevýhodou dotykových obrazovek je nutnost uživatelského kontaktu, náchylnost na poškrábání, snížení citlivosti vlivem nečistot a v některých případech analýza pouze jednoho dotyku.[4][20]

Ultrazvukové systémy pro rozpoznání gest využívají vysokofrekvenční ultrazvukové vlny, měření TOF (Time of Flight) a za pomoci triangulace určují polohu. Společně s Dopplerovým jevem lze určit i rychlost pohybu. Systémy se skládají z ultrazvukových zářičů, emitující ultrazvuk, zvukových disků, odrážejících ultrazvuk a několika senzorů. Systémy mezi komponentami vyžadují nekrytou dráhu. Výhodou je jejich nízká cena a malé rozměry snímače, k nevýhodám patří nízká přesnost, odezva a potřeba synchronizačního pulsu. [4][15]

Gesta lze rozeznávat také pomocí magnetických zařízení připevněných na ruku. Zařízení jsou malá a jejich rychlost daleko vyšší než u ultrazvukových systémů. Problém nastává s ohledem na možné interference s kovovými předměty, nebo elektromagnetickými zářiči. Také prostor je vzhledem k potřebné intenzitě magnetického pole omezený. Systémy využívají prvek, který je zdrojem magnetického pole spolu s malými senzory, umístěnými na například na ruce, které detekují pozici a orientaci v prostoru vzhledem ke zdroji. [4]

Mechanické systémy obsahují jednoosé a víceosé akcelometry, které lze použít v reálném čase pro rozpoznání směrů pohybu ruky. Pokud je systém doplněn o gyroskop, je schopný rychle a přesně určit pozici a orientaci ruky. Kývání ruky generuje u akcelometrického systému sinusovou vlnu, kde první pík je vysoký (největší akcelerace) a druhý nízký (největší deakcelerace). Tyto vrcholy se následně používají u rozpoznávání vzorů pro klasifikaci gest (například u zařízení Nintendo Wii). [4]

2.2 Optická zařízení pro rozpoznání gest

Bezkontaktní zařízení pro rozpoznání gest využívají počítačového vidění. Jedná se o optické metody, kde specifikace použitých analogových kamer (s analogově digitálním převodníkem) a digitálních kamer (většina aplikací), mají vliv na charakteristiku celých systémů a rozpoznání jednotlivých gest. V následujícím textu je tato problematika stručně popsána. [4]

Detektory v digitálních kamerách mohou detekovat ve spektrálním rozsahu od ultrafialového do infračerveného spektra světla. Absorpce fotonů snímacím prvkem může způsobit kvantovou nebo tepelnou odezvu (dle toho dělení detektorů na kvantové a tepelné). Nejvíce jsou v oblasti optických metod rozpoznání rozšířeny kvantové detektory, které pracují ve viditelném spektru světla (Visible Spectrum – VIS: vlnová délka 390 – 750 nm). Lze také použít kvantové detektory pracující v blízké infračervené oblasti (Near Infrared Spectrum – NIR: vlnová délka 0,75 – 1,4 μm) a tepelné detektory snímající v oblasti vzdáleného spektra infračerveného záření (Infrared Spectrum – IR: vlnová délka 8–15 μm). Většina objektů (včetně lidského těla) nevyzařují elektromagnetické vlny ve viditelném světle ani v blízkosti infračerveného světla. Tyto detektory tedy detekují světlo, které se od objektu odráží. Naproti tomu u teplených detektorů není třeba žádný externí zdroj světla, jelikož stejně jako lidské tělo, většina objektů v tomto spektrálním rozsahu energii vyzařují. [4]

Dalšími faktory ovlivňujícími rozpoznání gesta jsou rozlišení a parametry použitého senzoru. Digitální kamery obsahují pole senzorů, které je charakterizováno typem detektoru: CCD (Charge – Coupled device), CMOS (Complementary Metal – Oxide – Semiconductor), mikrobolometr (tepelný detektor), velikostí každého elementu v poli (větší element, větší kvantová účinnost), počtem elementů (rozlišení) a barevnou hloubkou. Nejnižším dnešním standardem pro kamery pracující ve viditelné oblasti pro využití rozeznání gest je rozlišení 640x480, s barevnou hloubkou 24 bpp (bits per pixel) s 16 miliony barev. Nejčastěji je používán barevný model RGB a snímkovací frekvence 30 (NTSC), 25 (PAL) nebo 15 snímků za vteřinu (frames per second - fps). Snímkovací frekvence, rozlišení, barevná hloubka a použitá komprese je úzce spojena s možností přenosu dat z přístroje do počítače. Minimem pro přenos a real time zpracování je rozhraní USB 2.0 s rychlostí 480 Mb/s, případně USB 3.0 (4,8 Gb/s) pro přenos velkých objemů dat. [4]

Kvalitu a zkreslení obrazu může značně ovlivňovat i použitý typ objektivu. Čočky jsou charakterizovány ohniskovou vzdáleností (v mm) a clonou (clonové číslo f). Ohnisková vzdálenost určuje spolu s velikostí snímače zorný úhel zachycené scény. Dle

toho lze objektivy rozlišit na širokoúhlé (malá ohnisková vzdálenost), střední ohniska a teleobjektivy. Velikost clony určuje hloubku ostrosti a množství záření, které dopadne na snímač. Vysoce světelné objektivy, používané při nízkém osvětlení a pro malou hloubku ostrosti, se vyznačují nízkým clonovým číslem (např. $f/1,4$). Při nízkém osvětlení je možnost ovlivnit hloubku ostrosti z důvodu velkého šumu vznikajícího na snímači malá. Kamery se také vyznačují buďto manuálním ostřením nebo automatickým. Automatické ostření může být vzhledem k detekci gest problematické. [4]

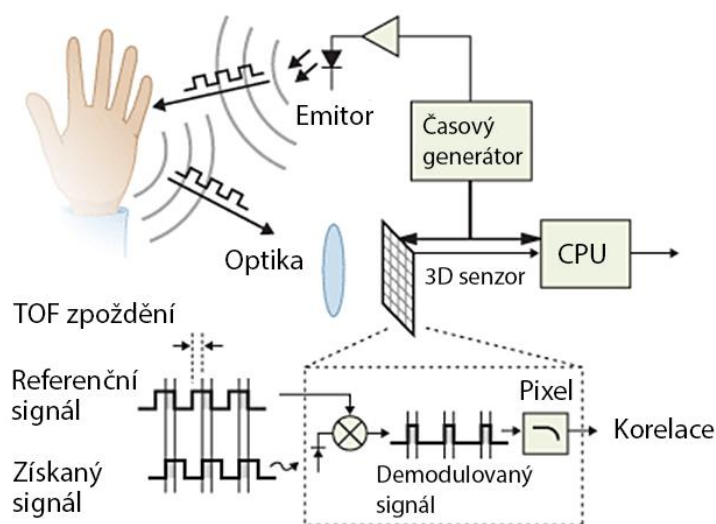
Většina popsaných systémů rozpoznávání gest se zaměřuje na spektrum viditelného světla. Systémů, které pro rozpoznání gest využívají oblast NIR a IR je v literatuře popsáno velice málo, i když nabízí slibnou alternativu. NIR systémy přináší schopnost fungovat ve všech světelných podmínkách, včetně úplné tmy a jsou relativně necitlivé ke změnám osvětlení. Termovizní systémy jsou necitlivé na okolní osvětlení, jelikož snímají vyzařovanou energii z objektu, což zvyšuje schopnost rozpoznání částí těla v podmínkách nevhodného osvětlení nebo ve tmě. Příkladem použití IR systému je začlenění termálního zobrazování Changem do multispektrálního systému pro rozpoznání obličeje, kde se vylepšila výkonnost rozpoznání tváře v měnících se světelných podmínkách. Největším problémem v současné době je vysoká cena těchto systémů. [4][57][45][18]

2.2.1 3D systémy

Většina systémů pracuje v současnosti s 2D zobrazením, ale je zde určitý trend pro vytváření systémů využívající 3D. Nese to sebou zvýšení ceny u komponent, vyšší výpočetní náročnost, ale také zde narůstá robustnost a správná identifikace gest. Získaný 3D poziční vektor je buďto použit přímo k identifikaci gest a nebo použit pro určení vzdáleností ve frontální poloze. 3D metody lze rozdělit do dvou hlavních tříd: pasivních a aktivních [23]. Nejrozšířenější metodou z pasivních metod je stereoskopie [51]. U aktivních metod převažují skenovací systémy, TOF systémy a metody strukturovaného světla (obrazce). [4]

Stereoskopie se zabývá odhadem hloubky scény na základě změn poloh ve dvou či více stereo obrazech. Základem pro snímání stereo snímků může být vhodně umístěná dvojice dvou kamer nebo jedné kamery se dvěma čočkami. Použití několika kamer (každá umístěna na jiné pozici) umožňuje získat přesnější informace, vytvoření 3D modelu ruky a lepší popis ruky. Metoda požaduje dostatečné světlo a vysoký kontrast mezi objekty v blízkém a vzdáleném poli. Z toho důvodu se vývojáři zaměřují na TOF systémy a metody strukturovaného světla, které umí vytvořit 3D profil lidí a objektů ve světle nízké výkonných NIR laserových diod. Filtry blokují okolní světlo, tedy sensory detekují pouze laserovou linii. [4]

System TOF, jehož schéma je zobrazeno na obrázku (Obr. 2.2), pracuje podobně jako laserový radar. Skládá se z IR laseru, který vytváří krátké pulsy a kamery měřící čas návratu z bodů celého zorného pole. Pro komerční použití existuje několik verzí těchto systémů, jako jsou například systémy 3DV, Canesta patřící pod Microsoft a Optrima patřící do hardwarové divize Sofkinetic-Optrima. System měří vzdálenost objektů porovnáním fáze odražených modulovaných pulsů, které byly emitovány laserem. Samostatné senzory tedy měří dobu letu laserových paprsků a zaznamenávají vlnové délky pro další analýzu. Tento systém na rozdíl od metody strukturovaného světla nepotřebuje kalibraci. TOF senzory umožňují v reálném čase odhad hloubky v podmínkách, kde stereoskopie selhává (např. v případě bílých zdí ve scéně). Na druhou stranu TOF senzory mají vyšší šum, horší rozlišení a často selhávají u špatně texturované scény, kde stereoskopie exceluje. [40]

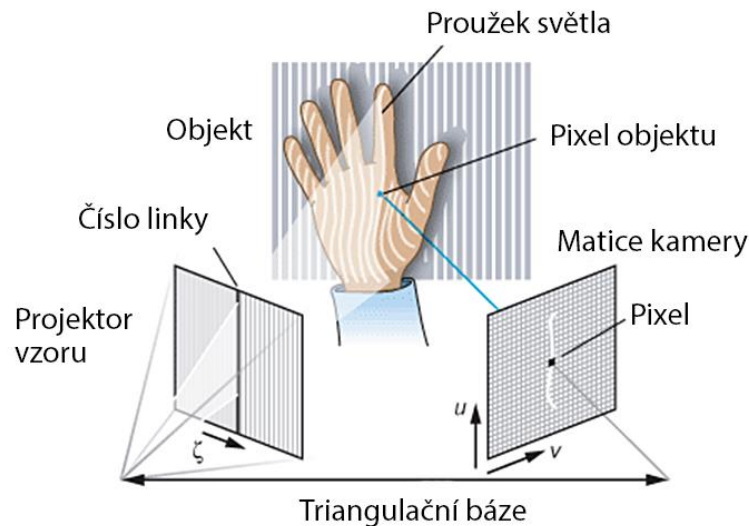


Obr. 2.2: TOF systém měří čas návratu pozorováním fázového posunu signálu vyslaného a signálu přijatého (převzato a upraveno z [40])

System TOF Optrima dokáže pracovat s jakoukoliv 3D kamerou, identifikovat a odstraňovat objekty mimo člověka (židle, stoly, květiny), počítat pozici rukou, noh, ramen a využít tyto informace pro rozpoznání gest a rekonstrukce skeletu postavy. [40]

V případě metody strukturovaného světla se osvětluje scéna vzorem (pole linek) a je snímána pod určitým úhlem. Pokud je vzor promítán na rovnou zeď, kamera zaznamená rovné linie. V případě že je vzorem osvětlena složitější scéna, jako je například osoba stojící před zdí, je možné pozorovat komplexnější profil. Pomocí digitálního zpracování

se analyzuje profil napříč polem ke zmapování topografie obličeje a těla. Pro vytvoření mřížky s vysokým poměrem signálu k šumu (SNR) se s výhodou používá silných laserů. Metodu strukturovaného světla vysvětluje obrázek (Obr. 2.3). [28]



Obr. 2.3: Metoda strukturovaného světla – Strukturované světlo je promítáno na objekt, na kterém se objeví při pohledu ze strany kontury. (převzato a upraveno z [40])

Pro získání dobrého výkonu současně bezpečného pro zrak, lze použít patentovanou techniku, která se jmenuje „světelné kódování“ a je používána pro Kinect od Microsoftu (využívá systému PrimeSense). Místo mřížky se používá speciální kód bohatý na informaci (s téměř nulovou repeticí ve scéně), což dává celému systému velkou spolehlivost. Osvětlovací laser emituje záření v rozsahu neviditelného pouhým okem (800 – 900 nm). V tomto rozsahu ale mají křemíkové CMOS detektory vysokou kvantovou účinnost. Samostatná kamera pak zaznamenává barevné obrazy. Rozlišení kamery je 640 x 480 pixelů s hloubkou 0,8 až 3,5 metrů. Rozlišení je daleko lepší než u konkurenčních TOF systémů a to s cenou pouze několika desítek dolarů. Pro systém PrimeSense jsou vývojářům volně přístupné open-source ovladače. [40]

3 METODY POUŽÍVANÉ PRO ROZPOZNÁNÍ GEST RUKY

Po získání obrazu vstupního zařízení (např. z VIS, IR kamery) následuje zpracování gesta v několika krocích. Jak bylo popsáno v předešlé kapitole, používají se nejčastěji po sobě následující tři kroky: extrakční metody (segmentace), získání příznaků a následné rozpoznání a klasifikace gesta.

3.1 Extrakční metody a předzpracování obrazu

Segmentace je prvním krokem při rozpoznání gest ruky. Cílem je rozdělit vstupní obraz do několika částí a oddělit význačné objekty od pozadí. Jedná se o stěžejní krok, jelikož kvalita segmentace úzce souvisí s následnou extrakcí příznaků a klasifikací. V literatuře je popsáno několik přístupů a případně i jejich kombinací. Patří k nim například segmentace na základě barvy kůže, subtrakční metody nebo také segmentace založené na hranové detekci. Segmentace se provádí u statických gest ale také i u dynamických gest, kde je po detekci ruky sledován i její pohyb. Před samotnou segmentací je dle potřeb nutné provést potlačení šumu (filtry založené na průměrování, mediánový filtr). [21][25]

3.1.1 Segmentace ruky na základě barvy kůže

Přístup segmentace ruky na základě barvy kůže je v literatuře často používaný pro svoji jednoduchost a výpočetní nenáročnost. Detekce barvy kůže může být ale nelehký úkol, jelikož samotná barva je závislá na několika faktorech. Častým problémem je měnící se osvětlení (barva, intenzita, pozice zdroje a s tím související stíny), vhodné nastavení expozice u kamer (zvláště problematická je automatická expozice), citlivost a parametry senzorů, rozdílnost barvy kůže mezi etniky, individuální charakteristiky jedinců ovlivňující barvu kůže (věk, pohlaví, emoční ladění) a barva pozadí (například dřevo) či jiné části těla (hlava). Mnohé tyto problémy, které nastávají při snímání ve viditelné části spektra, je možné překonat použitím kamer pro IR oblast (necitlivost na osvětlení, změny barev, pozadí). Problémem volby těchto systémů pro běžné použití bývá ale především jejich cenová nedostupnost. Z hlediska segmentace lze přistupovat k procesu tak, že obrázek dělíme na pixely, které patří barvě kůže a ostatní, které jí nepatří. Kroky pro detekci kůže v obraze jsou – reprezentace obrazových bodů ve vhodném barevném prostoru, rozdělení pixelů na ty které náleží kůži pomocí vhodného distribučního modelu a klasifikace distribuce. [25][44][29]

Barevný prostor RGB bývá nejpoužívanějším pro uchovávání a reprezentaci digitálních obrázků. Data získaná kamerou tedy bývají nejčastěji v barevném prostoru RGB, který je vyjádřen třemi barvami – červenou, zelenou a modrou. Tento prostor není pro segmentaci příliš vhodný, jelikož mezi jednotlivými složkami je velká závislost a není oddělena jasová složka. Pro snížení závislosti modelu na osvětlení, se provádí normalizace RGB komponent. Dále suma normalizovaných složek je rovna jedné ($r + g + b = 1$), proto lze třetí složku (která nenese žádnou informaci) vypustit a snížit tak dimenzionalitu. Bylo zjištěno, že touto úpravou byla závislost na světelných podmínkách a drobných barevných odlišnostech snížena. [25][52]

Vhodnou volbou může být barevný prostor HSI (HSV, HSL), který se skládá ze složek – barevného tónu (Hue), sytosti barvy (Saturation) a hodnoty jasu (intensity, Lightness, Value). Výhodou pro segmentaci je oddělení barevné složky od jasové. Transformace z RGB prostoru do HSV je invariantní na vysokou intenzitu bílého světla, okolního světla a orientaci povrchu ke zdroji světla. [25][52]

Velice oblíbeným a využívaným prostorem v detekci gest je ortogonální barevný prostor $YCbCr$, který řeší redundanci kanálů RGB barevného prostoru a reprezentuje barvu jako statisticky nezávislou komponentu. Z toho důvodu je vhodným barevným prostorem pro segmentaci barvy kůže. Luminance (Y) zde reprezentuje jas a další dvě složky červenou a modrou chrominanční komponentu (C_r, C_b). [25][52]

Po výběru vhodného barevného prostoru následuje rozdělení pixelů do dvou skupin na pixely, které náleží barvě kůže a na ty, které ne. Nejjednodušším a často používaným způsobem je definovat pevně hranice (prahy) pro rozdílné barevné komponenty zvoleného modelu. Jsou definovány jeden nebo více rozsahů prahových hodnot pro každou barevnou složku a pixely spadající do těchto nastavených rozsahů jsou brány jako pixely náležící barvě kůže. Příkladem může být použitý barevný prostor $YCbCr$ a nastavené hodnoty rozsahu pro složku $C_b = [77, 127]$ a složku $C_r = [133, 173]$ definující pixely náležící barvě obličeje [49]. Výhodou tohoto přístupu je nízká výpočetní náročnost, nevýhodou pak nepříliš vysoká robustnost. [25]

Dalším způsobem klasifikace pixelů do skupin je neparametrický model založený na histogramu. Histogramy jednotlivých barev mohou být použity pro odlišení velkého množství objektů. Z vhodného setu vzorků je vytvořen histogram (také vyhledávací tabulka - LUT), který je normalizován konverzí hodnot histogramu na diskrétní rozdělení pravděpodobnosti. Hodnoty normalizovaného histogramu (normalizované vyhledávací tabulky) udávají pravděpodobnost, zda daný pixel náleží do skupiny pixelů pokožky. Tedy všechny hodnoty pixelů, pro které je odpovídající pravděpodobnost větší, než předem stanovený práh jsou zařazeny do skupiny pixelů náležících pokožce. Alternativou

rozhodovacího pravidla je použití Bayesova klasifikátoru, u něhož se pravděpodobnost pro daný pixel stanovuje pouze pro určitou konkrétní barvu. Další neparametrickou metodou může být například SOM (Self-Organizing Map). Výhodou modelu založeného na histogramu je rychlost, nevýhodou pak závislost na vybraných vzorcích pokožky, ze kterých se sestavují pravděpodobnosti. [25][52]

Pro definici barvy pokožky se často používají i parametrické modely jako je jednoduchý Gaussovský model (SGM) nebo smíšený Gaussovský model. Při stálých podmínkách osvětlení náleží barva kůže jednotlivých jedinců malé oblasti v barevném prostoru. Tedy za určitých světelných podmínek lze využít pro distribuci barvy kůže eliptickou Gaussovskou distribuční funkci pravděpodobnosti v normalizovaném barevném prostoru. Při měnícím se osvětlení jednoduchý Gaussovský model neplatí a bývá nahrazen smíšeným Gaussovským modelem. K dalším parametrickým metodám patří například eliptický model. [25][52]

3.1.2 Subtrakční metody

V případě, že nedochází ke změně polohy kamery a pozadí, lze pro detekci ruky použít subtrakční metody. První možností je nasnímat pozadí (statickou scénu) bez přítomnosti objektu (ruky). Pro detekci ruky lze pak takto získanou scénu odečíst od scény s objektem. Řešení nemusí být vhodné z důvodu možnosti výskytu šumu v obraze nebo mírných změn osvětlení. Pokud se ruka ve scéně dostatečně pohybuje, lze použít model prostředí, který se získá z průměru n po sobě jdoucích snímků a zmíněné nedostatky předchozího přístupu se nemusí projevit. Existuje i řada modifikací této metody, například metoda, kdy je model prostředí aktualizován podle vztahu (3.1)[19]:

$$b_n = \alpha \cdot b_{n-1} + (1 - \alpha) \cdot f_n, \quad (3.1)$$

kde b_n značí model prostředí, f_n je aktuální snímek a α je koeficient, který nabývá hodnot 0 až 1 a má vliv na aktualizaci modelu. Problém může nastat při rychlém pohybu bodů v pozadí, nebo při zastavení objektu v popředí. [19]

3.1.3 Hranová detekce

Hranová reprezentace může nést důležitou informaci o hledaných objektech a bývá využívána pro následující extrakci příznaků. Hranovou detekcí je rozuměno transformování šedotónového obrazu na binární, ve kterém bílé pixely reprezentují pozice hran a černé pixely pozice, kde hrana není. Hrany jsou následně použity pro vytvoření obrysu objektu. Existuje velké množství filtrů pro detekci hran. Filtry aproximující první

derivaci diferencemi jsou například operátory Prewittové a Sobelův operátor, které jsou směrové (detekují hrany v určitém směru). Pro použití detekce ve všech směrech se obraz filtruje dvakrát, jednou maskou pro horizontální a druhou pro vertikální směr hrany. [1][56]

Pro detekci hran lze použít i lokální operátory aproximující druhou derivaci. Jedná se především o Laplacián a Laplacián Gaussianu, který má navíc vlastnost snižovat senzitivitu na šum. Dalším krokem po aplikaci operátorů je získání hrubé a následně výsledné hranové reprezentace, což je binární obraz, kde hodnota jedna vyjadřuje přítomnost hrany a nula opak, dále neobsahuje falešné hrany a tloušťka čar je rovna jedné. [1][56]

Mezi pokročilejší přístupy detekce hran lze řadit například Cannyho detektor, kde se provede filtrace šumu v originálním obraze pomocí Gaussianu, detekují se hrany kompasovým detektorem ve všech směrech a vypočítá se absolutní hodnota gradientu v obraze. Následně se vypočítá směr gradientu, určí se směr hran v každém bodu a provede se normalizace. Na závěr se sleduje hrana, potlačí se veškeré pixely, které nejsou vyhodnoceny jako hranové a provede se prahování s hysterezí. [1][56]

3.1.4 Úpravy po segmentaci

Po segmentaci ruky je dle úspěšnosti nutné provádět další matematické operace jako jsou morfologické operace případně mediánovou filtraci. [1]

V rámci morfologických operací lze jako vhodnou úpravu zvolit erozi, dilataci, otevření nebo uzavření [1].

Binární erozi lze provádět operátorem, který testuje, zda jsou u binárního obrazu pod všemi aktivními prvky definovaného strukturního elementu jedničky. V případě že tomu tak je, zapíše se na pozici referenčního prvku strukturního elementu do výstupního obrazu hodnota jedna. Tuto operaci lze matematicky popsat rovnicí (3.2) [56]

$$Y = E_H(X) = \{x\}H_x \subseteq X \quad (3.2)$$

kde x představuje prostorové souřadnice obrazu, na kterých se vyskytuje referenční prvek strukturního elementu, H_x strukturní element posunutý na souřadnici x a $X(Y)$ samotný vstupní (výstupní) obraz. Eroze objekty odstraňuje, zmenšuje a rozpojuje. [56]

Binární operace dilatace se provádí operátorem, který ověřuje, zda je kdekoliv pod definovaným strukturním elementem jednička obrazu. Pokud ano bude do výstupního obrazu vložena jednička na pozici, která odpovídá pozici centrálního prvku. Operátor lze

matematicky vyjádřit rovnicí (3.3) [56]

$$Y = D_H(X) = \{x | Hx \cap X \neq \emptyset\}. \quad (3.3)$$

Rovnice vyjadřuje vložení hodnoty jedna do výstupního obrazu na pozici centrálního prvku strukturního elementu x na pozicích, na kterých průnik aktivních prvků strukturního elementu a úseku obrazu, který je jím překrytý, není roven prázdné množině. H_x vyjadřuje opět strukturní element posunutý na souřadnici x a $X(Y)$ samotný vstupní (výstupní) obraz. Dilatace objekty zvětšuje, má tendence k vyplňování prázdných míst a propojování objektů. [56]

Operace eroze a dilatace mění velikost objektů v obraze, což je v některých aplikacích nevýhodné. Řešením může být použití operátorů otevření nebo uzavření, které jsou tvořeny zřetěžením operací eroze a dilatace, přičemž zásadní filtrační vliv má první operace. Druhá operace slouží pouze k přibližné restauraci velkých objektů. Účelem použití těchto operátorů je ovlivnění malých objektů při současném přibližném zachování tvarů a velikostí velkých objektů. [56]

Operace otevření je kombinací eroze a následné dilatace. Vzhledem k časové náročnosti je ale výhodnější použít pouze jednu operaci. Operátor funguje následovně. Pokud se pod všemi aktivními prvky strukturního elementu, který je položený na obraz, nacházejí jedničky, jsou pod tyto prvky do výstupního obrazu také zapsány jedničky. Matematický zápis této operace lze vyjádřit rovnicí (3.4) [56]

$$Y = O_H(X) = \cup \{Hx | Hx \subseteq X\} \quad (3.4)$$

kde H_x vyjadřuje strukturní element posunutý na souřadnici x a $X(Y)$. Otevření eliminuje malé výběžky objektu do pozadí obrazu a má tendenci k rozpojování propojených objektů. [56]

Uzavření je kombinací dilatace a následné eroze, přičemž je možné opět použít pouze jednoho operátoru. Pomocí operátoru je filtrováno pozadí, které objekty obklopuje. Zjišťuje se, zda jsou pod všemi aktivními prvky strukturního elementu položeného na obraz prvky patřící pozadí obrazu. V případě že tomu tak je, jsou prvky nacházející se pod aktivními prvky strukturního elementu nastaveny na hodnotu nula (na rozdíl od všech předchozích operací je výstupní obraz inicializován jako pole jedniček). Pokud tomu tak není, výstupní obraz zůstává nezměněn. Matematické vyjádření má podobu rovnice (3.5) [56]

$$Y = U_H(X) = [\cup \{Hx | Hx \subseteq X^c\}]^c. \quad (3.5)$$

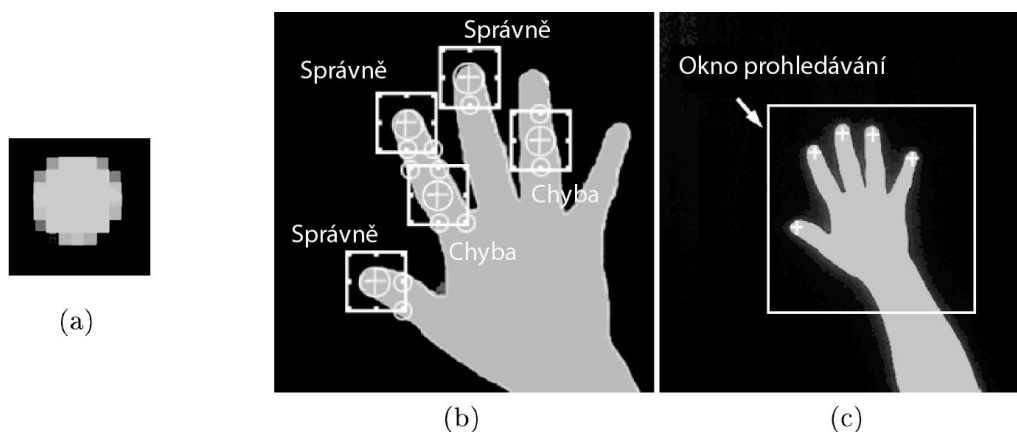
Vstupní obraz X se převede na množinový doplněk, na němž je proveden sled operací jako při otevření, tedy průnik všech aktivních poloh strukturního elementu, a výpočtem

doplňku je získán výsledný „uzavřený“ obraz. Binární uzavření eliminuje díry, výběžky pozadí do objektů a propojuje objekty ležící blízko sebe. [56]

3.2 Extrakce příznaků

Po segmentaci objektu, která je klíčová pro následující zpracování, lze dále provádět extrakci příznaků. Cílem extrakce příznaků je nalezení parametrů, které co možná nejlépe popisují detekovaný objekt. Ze segmentovaného obrazu lze získat n -rozměrný vektor příznaků rozdílnými metodami a to v závislosti na konkrétní aplikaci. Pro statická gesta je možné použít metody, které využívají geometrických příznaků jako jsou polohy a směry prstů, střed dlaně, kontury ruky nebo ne-geometrické příznaky jako je například barva, hodnota jasu pixelů, textura nebo silueta.

Pro získání některých geometrických příznaků je vhodné nalézt jednotlivé prsty a centrum dlaně. Prsty lze hledat metodou vyhledávání vzorů (template-matching) pomocí vzoru tvaru kruhu (viz Obr. 3.1). Algoritmus používá normalizované korelace se vzorem fixní velikosti. Ideálně by se vzor měl lišit pro rozdílné prsty a uživatele. Bylo ale zjištěno, že fixní velikost funguje u různých uživatelů a prstů poměrně dobře. Pro omezení falešných detekcí byla navržena metoda potlačení nalezených prstů okolo kandidáta s největším skórem (získaným z algoritmu) a hledání detekcí nacházejících se uprostřed prstu. Ty jsou odstraněny testováním okolních pixelů v diagonálních směrech. Po odstranění zůstávají správné detekce ležící na koncích prstů. [28]

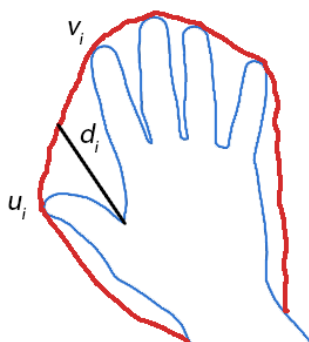


Obr. 3.1: Vyhledávání vzorů: a) zvolený vzor; b) správné a falešné detekce prstů; c) zvolené okno prohledávání [28]

Nalezení středu dlaně je důležité pro rozpoznání různých typů gest, jako je například odhad natažení jednotlivých prstů a jejich poloha. Používané metody počítají střed dlaně jako těžiště ruky, avšak to se významně mění například při otevírání nebo zatínání ruky.

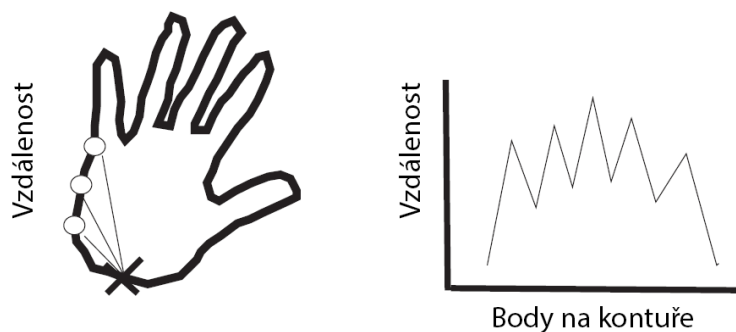
Řešením může být metoda, při níž je centrem ruky uživatele brán bod, kde vzdálenost k nejbližší hranici ruky je maximální. Takovéto umístění středu se vypočítá pomocí morfologické operace eroze detekované ruky. Morfologický operátor je aplikován do té doby, dokud není plocha dostatečně malá a z ní je pak vypočteno těžiště, které je bráno jako střed dlaně. Příznak orientace osy ruky může být dále počítán z momentů obrazu [14]. [28]

Konturu nebo hranovou reprezentaci ruky lze společně s konvexní obálkou použít také pro nalezení specifického vektoru příznaků. Kontura je použita k hledání bodů na konvexní obálce a výpočtu maximální vzdálenosti konvexní obálky (d_i) od kontury mezi dvěma sousedními body (u_i a v_i) konvexní obálky. Následně zprůměrováním vzdáleností získáme příznak d , podle kterého lze gesto klasifikovat. Metoda je vyobrazena na obrázku (Obr. 3.2). [36]



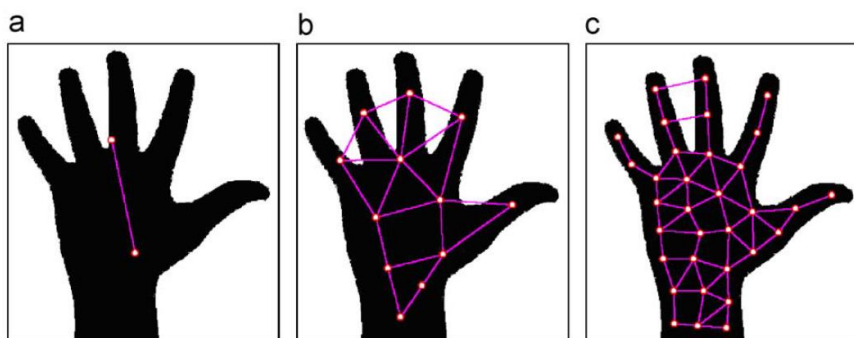
Obr. 3.2: Použití kontury a konvexní obálky pro hledání příznaku d (převzato a upraveno z [36])

Z kontury lze získat i příznaky popisující obvod objektu, poměr hlavní a vedlejší osy opsané elipsy a její natočení. Konturu lze také využít pro hledání prstů pomocí vypočítání vzdáleností například od středu zápěstí, k bodům na kontuře a následným vytvořením grafu a sledováním jednotlivých vrcholů na křivce, jako patrné z obrázku (Obr. 3.3). [48]



Obr. 3.3: Získání příznaků z kontury pro hledání prstů [48]

Další možností získání příznaků je využití neuronových sítí (základy neuronových sítí jsou popsány v další kapitole vysvětlující klasifikaci) [22]. Neuronových sítí využívá například metoda SGONG (Self-Growing and Self-Organized Neural Gas). Algoritmus SGONG je aplikován po segmentaci využívající $YCbCr$ barevném prostoru. SGONG začíná pouze se dvěma neurony a roste až do doby konvergence. Výsledkem je mřížka výstupných neuronů, které mají tvar ruky a slouží pro získání tvaru ruky a příznaků – oblasti dlaně, středu dlaně a náklonu dlaně. Pomocí těchto příznaků lze určit polohu jednotlivých prstů a zda je prst natažený či pokrčený. Ukázka použití algoritmu SGONG je na obrázku (Obr. 3.4). [47]



Obr. 3.4: Rostoucí SGONG síť: a) počáteční stav, b) rostoucí síť, c) konečný výstup sítě neuronů [47]

Poměrně jednoduchá a rychlá metoda extrakce příznaků byla použita v práci [35], kde bylo vypočítáno těžiště ze segmentované ruky a následně vzdálenost od těžiště k nejdelšímu prstu (největšímu extrému). Dále byla sestrojena kružnice o průměru 70 % zjištěné vzdálenosti a byl získán 1D signál pomocí sledování kružnice a zjištění přechodů mezi bílou černou barvou. Zjištěním počtu přechodů a zanedbáním přechodu pro zápěstí byl zjištěn počet prstů. Dalším přístupem může být rozdělení segmentovaného obrazu na jednotlivé bloky různých velikostí, kde příznak je počítán z bloku, který je reprezentován hodnotou jasů v obraze. Velikost zvoleného bloku ovlivňuje rychlost algoritmu a jeho optimální velikost byla diskutována u řady experimentů. Za zmínku stojí i použití Gaussovy funkce hustoty pravděpodobnosti pro extrakci geometrických centrálních momentů jako globálních a lokálních funkcí. [41]

3.3 Klasifikace gest

Po získání vektoru příznaků následuje klasifikace a správné rozpoznání gesta ruky. Klasifikační algoritmy je nutné volit s ohledem na získané příznaky a také typ gesta (statické nebo dynamické). Pro detekci statických gest lze využít obecné klasifikátory

jako je například shluková analýza a do ní patřící algoritmus k-means, metoda k-nejblížších sousedů (k-nearest neighbor) a mean shift, nebo lze použít analýzy hlavních komponent (PCA), neuronových sítí, případně i genetických algoritmů. Pro klasifikaci dynamických gest se jako vhodná jeví metoda skrytých Markovových modelů nebo technika dynamického borcení času (dynamic time warping). [41]

3.3.1 Algoritmus k-means

Algoritmus k-means je nehierarchickým algoritmem, který třídí data na základě jejich vlastností do jednotlivých shluků. Před začátkem algoritmu je nutné definovat celkový počet shluků, který je menší než počet objektů a maximální počet iterací. Algoritmus přiřazuje každý objekt do shluku, jemuž středu je nejbližší. Středů shluků se při každé iteraci přepočítávají jako aritmetické průměry všech bodů shluku. Cílem je dosáhnout co možná nejmenších rozdílů uvnitř shluků. Lze postupovat například následovně. Vytvoří se náhodný počet shluků, který je definován uživatelem. Vypočítají se středů shluků. Objekty se přemísťují do shluků, které jsou k nim nejbližší a tím se mění středů shluků. Zmenšuje se hodnota chybové funkce E dle rovnice (3.6) [35]

$$E = \sum_{k=1}^k \sum_{x \in C_k} d(\bar{x}, \bar{V}_k), \quad (3.6)$$

kde C_k je k-tý shluk, V_k je střed (centroid) shluku C_k , a $d(\bar{x}, \bar{V}_k)$ je vzdálenost \bar{x} od centroidu. Po určitém počtu iterací se složení shluků nemění a algoritmus je ukončen. [50]

Metoda je rychlá a jednoduchá a dokáže pracovat s velkým počtem vstupních dat a to v konečném počtu iterací. Nevýhodou je nutnost stanovení počtu shluků, náchylnost na izolované a vzdálené objekty a závislost na počátečním umístění centroidů shluků. V případě rozdílných jednotek je nutné provádět standardizaci. [50]

3.3.2 Neuronové sítě

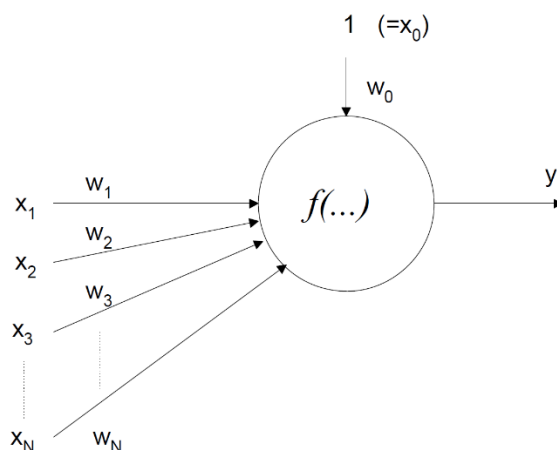
Neuronová síť je systém, který je inspirován biologickými systémy. Skládá se z velkého počtu navzájem propojených elementů (neuronů). K vlastnostem neuronových sítí patří schopnost učit se (měnit své parametry) a to za pomoci zpětné vazby (výstupu). Učení neuronové sítě probíhá s učitelem (síť je učena pomocí příkladů) nebo bez učitele (dle chyb a zadaných kritérií). Další vlastností je robustnost vůči chybám a paralelní

zpracování informací, což zvyšuje jejich rychlost. Neuronovou síť lze charakterizovat typem neuronů (dle přenosových funkcí), topologií (uspořádání a propojení neuronů), zmíněným způsobem učení a vybavování. Nevýhodou použití neuronových sítí je problematická zpětná interpretace způsobu dosažení výsledků. [31]

Jeden neuron (jinak nazýván perceptron) jako procesní prvek s N vstupy a jedním výstupem je popsán rovnicí (3.7) [31]

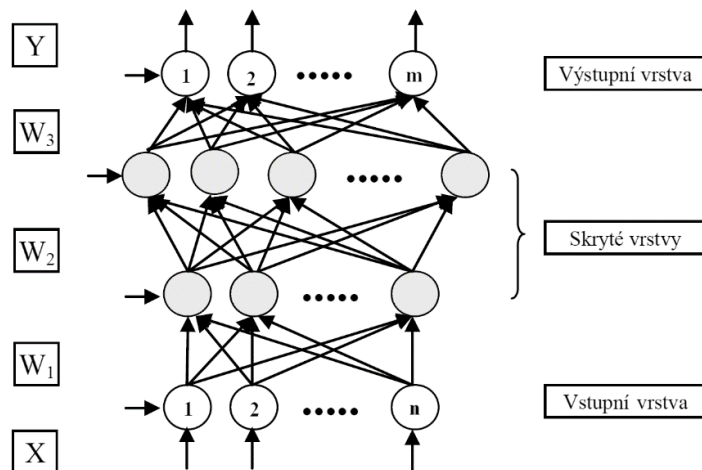
$$y = f \left[\sum_{i=1}^N w_i x_i - \theta \right] = f \left[\sum_{i=0}^N w_i x_i \right], \quad (3.7)$$

kde $\bar{x} = [x_1, x_2, \dots, x_N]^T$ je vstupní vektor, $\bar{w} = [w_1, w_2, \dots, w_N]^T$ je vektor aktuálních vah a θ je aktuální práh neuronu. Váha w_0 představuje práh $-\theta$ při fiktivním vstupu $x_0 = 1$. Funkce $f(\alpha)$ je charakteristika neuronu s argumentem α (aktivace neuronu). Běžně se používají charakteristiky v podobě lineární, skokové a sigmoidní funkce. Na obrázku (Obr. 3.5) je uvedeno schéma jednoho neuronu. [31]



Obr. 3.5: Schéma neuronu [31]

Vzájemné uspořádání jednotlivých neuronů nám určuje topologii sítě. Existuje mnoho typů uspořádání. K nejběžnějším sloužícím ke klasifikaci ze získaných příznaků patří vícevrstvá perceptronová síť skládající se ze vstupní vrstvy, několika skrytých vrstev a vrstvy výstupní. Z každého neuronu v dané vrstvě vedou spoje do všech neuronů v následující vrstvě a to v dopředném směru mezi dvěma sousedními vrstvami. Síť je vyobrazena na obrázku (Obr. 3.6). [31]



Obr. 3.6: Perceptronová vícevrstvá síť (převzato a upraveno z [31])

Učení sítě probíhá pomocí algoritmu zpětného šíření chyby. Jedná se o iterativní proces, který zjednodušeně probíhá tak, že se provede výpočet výstupu sítě (dopředné šíření), po kterém následuje výpočet chyby všech neuronů (zpětné šíření) a na závěr se váhy jednotlivých neuronů upraví (dopředné šíření). Učení končí dosažením maximálního počtu epoch (iterací) nebo dosažením požadované chyby (z výpočtu kvadratické chyby pro celou tréninkovou množinu). [31]

Pro rozpoznání obrazů (objektů) bez nutnosti popisu objektů pomocí jejich příznaků lze využít Hammingovu neuronovou síť. Hammingova síť slouží ke stanovení podobnosti vstupního N -rozměrného binárního vektoru (vstupní binární obraz převedený na vektor) se zapamatovatelnými M vzory (binární vzory převedené na vektor). Síť využívá Hammingovy metriky pro zjištění vzdálenosti vstupního vektoru od jednotlivých tříd. Všechny neurony mají u jednotkového vstupu váhu N a charakteristiku $y = \alpha/2$. Binární vstupy i váhy nabývají hodnot 1 nebo -1 . Učení sítě spočívá v nastavení vah ${}^m w_n$ jednotlivých neuronů pro m -tý vstupní vzor podle složek tohoto vstupu ${}^m x_n$ dle vztahu (3.8)[31]

$${}^m w_n = {}^m x_n, \quad n = 1, 2, \dots, N. \quad (3.8)$$

Výpočet jednotlivých složek p -tého výstupu (odpovídajícímu p -tému vstupnímu vektoru) je dán rovnicí (3.9)[31]

$${}^p y_m = \frac{1}{2} \left(\sum_{n=1}^N {}^m w_n {}^p x_n + N \right), \quad m = 1, 2, \dots, N. \quad (3.9)$$

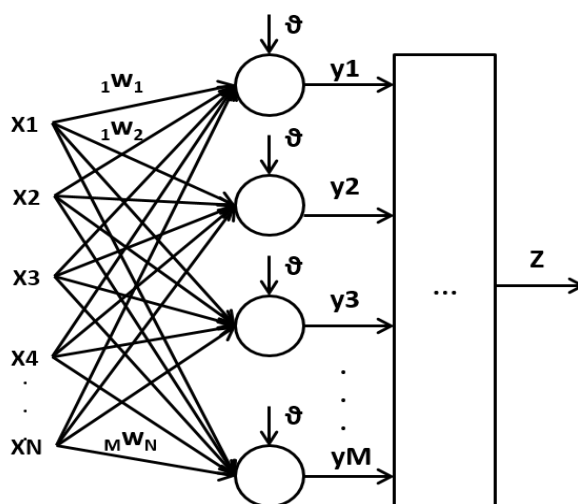
Na výstupu m -tého neuronu je počet souhlasných bitů vzájemně korespondujících vstupů a vah, který vyjadřuje podobnost neznámého vstupu s m -tým vzorem. Při souhlasu všech

vah s jednotlivými složkami vstupního vektoru pak platí vztah (3.10)[31]

$$N = \sum_{n=1}^N m w_n \quad {}^p x_n \quad (3.10)$$

a odtud $N = {}^p y_m$. Na výstupy jednotlivých neuronů navazuje výběr maxima z těchto výstupů, nebo výběr minima Hammingové vzdálenosti (vyjadřuje míru nepodobnosti, která je definována jako $h=N - {}^p y$). [31][43]

Topologii sítě s popisem lze sledovat na obrázku (Obr. 3.7).



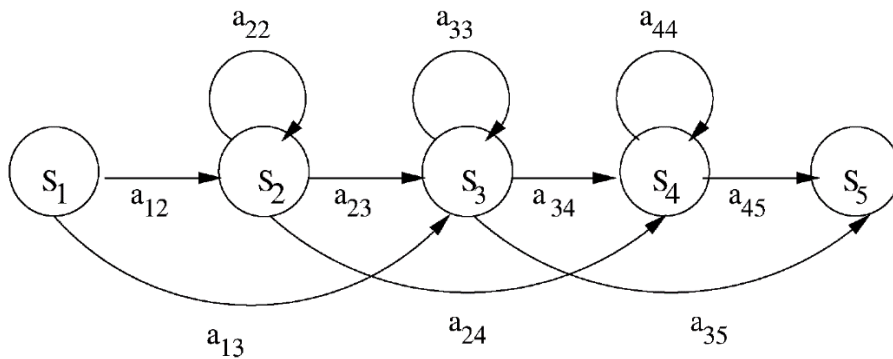
Obr. 3.7: Topologie Hammingovy sítě; $x_1 \dots x_n$ – binární vstupní vektor $\{-1,1\}$, ${}_m w_n$ – váha pro spojení mezi neuronem m a vstupem n (váhy odpovídají zapamatovaným binárním vzorům $\{-1,1\}$), $y_1 \dots y_m$ – výstupy jednotlivých neuronů, z – výsledný výstup ze sítě (maximum z jednotlivých výstupů neuronů, minimum Hammingovy vzdálenosti) [43]

3.3.3 Skryté Markovovy modely

Skryté Markovovy modely (Hidden Markov Models – HMM) jsou nejčastěji používány pro klasifikaci dynamických gest. Markovovy řetězce jsou konečné stavové automaty s pravděpodobnostními přechody mezi stavy. Obvykle popisují diskrétní náhodný proces, pro který platí, že pravděpodobnosti přechodu do následujícího stavu závisí pouze na současném stavu a ne na předchozích stavech. HMM jsou považovány za skryté v případě, že je stochastická proměnná závislá na stavech, které nejsou pozorovatelné ale výstup, který je na skrytém stavu závislý, viditelný je. HMM jsou jako konečné stavové automaty charakterizovány jako (3.11)[9]

$$\lambda = (N, M, A, B, \pi), \quad (3.11)$$

kde N značí vektor skrytých stavů, M je vektor pozorovatelných stavů (maximální počet rozpoznatelných gest), A matice pravděpodobnosti přechodu mezi skrytými stavy, B matice pravděpodobnosti přechodu k pozorovatelným stavům a vektor počátečních stavů je značen π . Markovovy modely se dělí na diskrétní a spojité. Spojité modely bývají pro klasifikaci gest využívány nejčastěji a je pro ně vstupem právě získaný vektor příznaků. Pro klasifikaci gest bývá využívána topologie levopravého (dopředného) modelu uvedeného na obrázku (Obr. 3.8), kde se postupuje od počátečního stavu ke koncovému. Pravděpodobnost přechodu ze současného stavu do stavu předchozího je rovna nule. [38][9]



Obr. 3.8: Pětistavový levopravý HMM model, s – jednotlivé stavy, a – přechody mezi stavy [38]

V případě použití HMM jako klasifikačního modelu řešíme s jakou pravděpodobností generuje daný model sekvenci (evaluace modelu) a to pomocí Forward-Backward algoritmu, jaký je nejpravděpodobnější průchod skrytými stavy při generování dané sekvence (dekódování modelu) Vitebriho algoritmem a jak změnit parametry modelu tak, aby generoval danou sekvenci s větší pravděpodobností (učení modelu) Baum-Welchovým algoritmem. Forward-Backward algoritmus se sestává z výpočtu dopředných pravděpodobností, výpočtu zpětných pravděpodobností a vyhlazení (pomocí Bayesova pravidla a nezávislosti jevů). U Vitebriho algoritmu se dopředně počítá matice cesty skrytými stavy a zpětně se vybírá trasa s nejvyšší pravděpodobností. Baum-Welchův algoritmus hledá neznámé parametry HMM (učení) a je založen na principu EM (Expectation – Maximization), kde může nastat problém se zasekáváním v lokálních extrémech. [38][9]

4 HARDWARE PRO ROZPOZNÁNÍ GEST

Tato kapitola popisuje hardwarové vybavení pro rozeznání gest a to kamery snímající v infračervené oblasti (dostupné na Ústavu biomedicínského inženýrství a Ústavu automatizace a měřicí techniky VUT) a ve viditelné oblasti.

4.1 Hardwarové možnosti pro optické rozpoznání gest

Pro rozpoznání gest jsou k dispozici kamery FLIR i7 a FLIR A655sc, které snímají v infračervené oblasti. Je možné využít i kompaktní termovizní senzor FLIR Lepton připojený k systému Raspberry Pi 2. Pro snímání ve viditelné oblasti lze zvolit například webovou kameru Logitech S7500.

4.1.1 Infračervená kamera FLIR i7

Jedná se o malou termovizní kameru primárně určenou pro rychlé vyhledávání případných závad na elektrických rozvodech, mechanických zařízení a zaměřování míst s chybějící či sníženou izolací spojenou s úniky tepla. Kameru lze použít pouze pro získání jednotlivých snímků, bez možnosti nahrávání sekvence snímků (videa). Možnosti použití pro rozpoznání gest jsou tedy omezené. Lze ji ale využít například pro porovnání navržených algoritmů pro snímky v IR oblasti se snímky pořízenými ve viditelné oblasti. Parametry kamery včetně ceny jsou uvedeny v tabulce (Tab. 4.1) a příklady pořízených snímků gest na (Obr. 4.1). [7]

Tab. 4.1: Parametry infračervené kamery FLIR i7 [7]

Spektrální rozsah; teplotní citlivost:	7,5 - 13 μ m; <0,1 $^{\circ}$ C při 25 $^{\circ}$ C
Typ detektoru:	120 x 120 pixelů (FPA) nechlazený mikrobolometr
Zorné pole:	25 $^{\circ}$ x 25 $^{\circ}$
Způsob ostření:	automatické
Přibližná cena:	63 000 Kč



Obr. 4.1: Pořízené snímky gest infračervenou kamerou FLIR i7

4.1.2 Infračervená kamera FLIR A655sc

Jedná se o infračervenou kameru s vysokým rozlišením určenou spíše pro vědeckou oblast pro měření přesné teploty a úniku tepla u různých zařízení a procesů v reálném čase. Kamera se vyznačuje vysokou kvalitou obrazu a rychlostí snímání. Pro rozpoznání gest se jeví jako velice dobrá volba (pokud nebereme v potaz cenu). Jednotlivé parametry včetně uvedené ceny jsou v tabulce (Tab. 4.2). Příklady snímků získaných gest lze pak sledovat na obrázku (Obr. 4.2). [11]

Tab. 4.2: Parametry infračervené kamery FLIR A655sc [11]

Spektrální rozsah:	7,5 – 14,0 μm	Způsob ostření:	automatické
Typ detektoru:	640 x 480 pixelů nechlazený mikrobolometr	Maximální snímkovací frekvence:	25 fps (640 x 480)
Zorné pole:	25° x 19° (24,6 mm)	Přibližná cena:	140 000 Kč



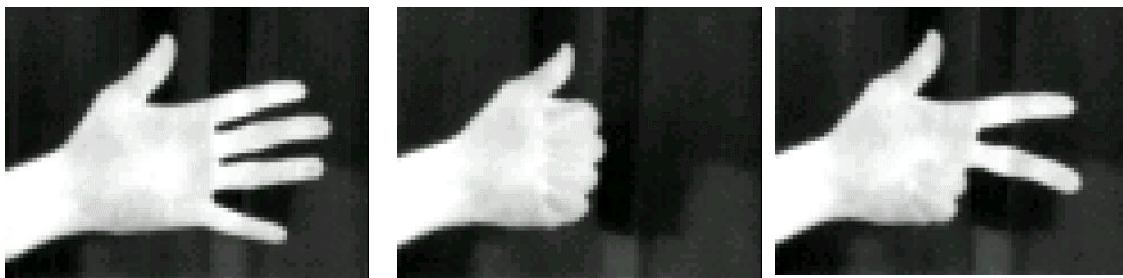
Obr. 4.2: Pořízené snímky gest infračervenou kamerou FLIR A655sc

4.1.3 Modul FLIR Lepton

Lepton je jedním z nejkompaktnějších termovizních modulů s rozměry menšími než pětikoruna. Tento kamerový modul byl navržen tak, aby ho bylo možné snadno použít v mobilních zařízeních a další spotřební elektronice. Lepton obsahuje čočku, vyrobenou plátkovou technologií s mikrobolometrem FPA (Focal Plane Array). Modul navíc zahrnuje pokročilé zpracování infračervených obrazů. Lepton je možné připojit k jednodeskovému počítači Raspberry Pi 2 (s operačním systémem Raspbian) a to prostřednictvím portu GPIO (General-purpose input/output). Pro komunikaci je nutné povolit v terminálu systému komunikační rozhraní SPI a I2C a doinstalovat potřebné balíčky (více o propojení a instalaci v *FLIR Lepton Hookup Guide* [12]). Z modulu lze získat jednak jednotlivé snímky, ale také sekvenci snímků (video). Parametry modulu včetně ceny jsou uvedeny v tabulce (Tab. 4.3) a příklady nasnímaných gest na obrázku (Obr. 4.3). [33]

Tab. 4.3: Parametry modulu FLIR Lepton [33]

Spektrální rozsah:	8 – 14,0 μm	Zorné pole:	25°
Typ detektoru:	80 x 60 pixelů nechlazený VOx mikrobolometr	Snímkovací frekvence:	8,6 fps
Teplotní citlivost:	<50 mK (0.050° C)	Přibližná cena:	Lepton: 4300 Kč Raspberry Pi 2: 1100 Kč



Obr. 4.3: Pořízené snímky gest modulem FLIR Lepton

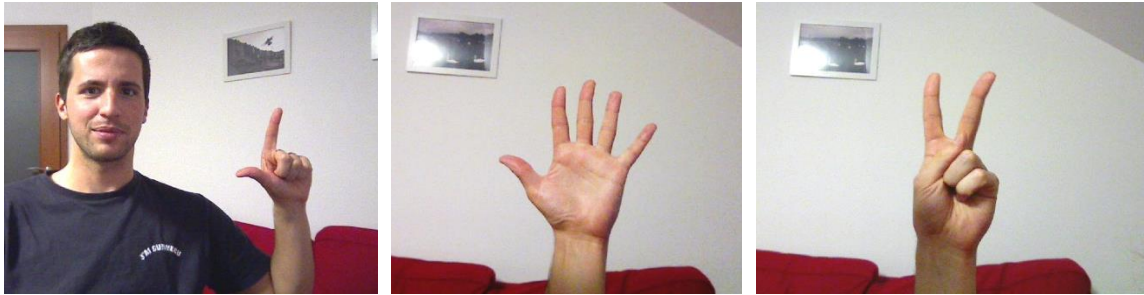
FLIR Lepton se nachází také v komerčním modulu FLIR One, který je určený pro připojení k mobilním zařízením s operačním systémem Android nebo iOS. Modul navíc obsahuje kameru snímající ve viditelné oblasti světla (kamera VGA 640 x 480 pixelů), díky které pomocí technologie MSX blending vylepšuje rozlišení termálního snímku na 160 x 120 pixelů. [13]

4.1.4 Logitech S7500

Logitech S7500 je webová kamera snímající ve viditelné oblasti spektra, poskytující pro hodnocení gest dostatečné rozlišení a snímkovací frekvenci. Její parametry včetně ceny lze nalézt v tabulce (Tab. 4.4) a příklady nasnímaných gest na obrázku (Obr. 4.4). [34]

Tab. 4.4: Parametry kamery Logitech S7500 [34]

Typ detektoru:	1280 x 960 pixelů, CMOS
Zorné pole:	72°
Snímkovací frekvence:	15 fps (1280x960); 30 fps (640x480)
Způsob ostření:	manuální
Přibližná cena:	1500 Kč (před vyřazením z prodeje)



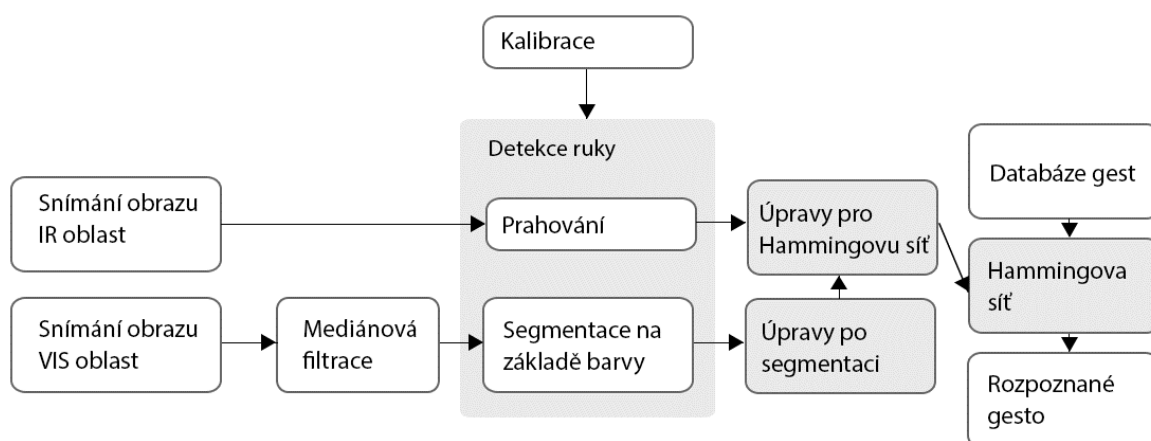
Obr. 4.4: Pořízené snímky gest kamerou Logitech S7500

Jako vhodná optická zařízení pro hodnocení gest snímající v infračervené oblasti se jeví kamera FLIR A655sc a modul FLIR Lepton, jelikož jsou schopny pořizovat nejen jednotlivé snímky, ale také video. Kamera FLIR A655sc umožňuje získávat vysoce kvalitní snímky ve vysokém rozlišení a lze ji tedy použít pro srovnání dalších zařízení. Pro testování je použit modul FLIR Lepton také z toho důvodu, že výrobce sám uvádí možnost využití pro rozpoznání gest. Pro viditelnou oblast byla zvolena cenově dostupná webová kamera Logitech S7500.

5 NAVRŽENÝ ALGORITMUS

V programovém prostředí MATLAB byla vytvořena aplikace pro rozpoznání gest, jejíž podobu a možnosti nastavení lze nalézt v příloze. Aplikace byla vytvořena s ohledem na snadné nastavení algoritmu a za účelem zpracovávat jakákoliv navržená gesta včetně různého natočení ruky a polohy prstů. Cílem byla také možnost jednoduché a rychlé tvorby databáze nových gest.

Algoritmus zahrnuje kalibraci za účelem získání prahů pro segmentaci. Ve viditelné oblasti je vhodné před samotnou segmentací provést odstranění komprimačních artefaktů (vzniklých například u komprese JPEG, AVI) pomocí mediánového filtru. Segmentace je následně v infračervené oblasti prováděna jednoduchým prahováním šedotónového obrázku, ve viditelné oblasti se používá prahování jednotlivých složek barevných prostorů RGB (normalizovaný) a YCbCr. Úpravy po segmentaci ve vhodném prostředí nejsou zpravidla nutné u segmentace získané z infračervených snímků, ale je vhodné je provádět u snímků získaných z viditelné oblasti. Úpravy zahrnují operace vyhledání největšího objektu, mediánovou filtraci, morfologickou operaci otevření, nebo též operaci vyplnění objektu. Takto vysegmentovaný objekt (ruku) je nutné dále zpracovat a provést úpravy pro Hammingovu síť. Úpravy zahrnují vhodný ořez pro odstranění předloktí a rotaci ruky. Po úpravách následuje rozpoznání gest pomocí Hammingové neuronové sítě, která porovná gesto s dostupnou databází gest a správně jej zařadí. Celý postup lze sledovat na obrázku (Obr. 5.1) a bude dále podrobněji popsán.



Obr. 5.1: Algoritmus pro rozpoznání gest

5.1 Kalibrace

Před segmentací, která je z důvodu rychlosti prováděna prahováním jednotlivých barevných složek (viditelná oblast) nebo šedotónového obrázku (infračervená oblast) musí být správně stanoveny jednotlivé prahy. Ty lze získat pro jednotlivé složky z normalizovaných histogramů. Normalizovaný histogram (normalizovaná vyhledávací tabulka) udává pravděpodobnost hodnoty určitého pixelu umístěného v obrázku. Normalizace histogramu se provádí ze standardního histogramu a to podělením každé hodnoty sumou všech hodnot. Suma hodnot normalizovaného histogramu je tedy rovná jedné. Normalizované histogramy lze získat z předem nasnímaných vzorků ruky, avšak lepší variantou je provedení kalibrace před samotným měřením.

Kalibrace využívá pro vymezení ruky metodu narůstání oblastí. Z důvodu odstranění artefaktů vzniklých kompresí je prováděna před segmentací mediánová filtrace s velikostí okna 5. Počáteční bod (semínko) je volen uživatelem přibližně do středu ruky. Prohledávání probíhá v osmiokolí a rozhodnutí, zda budou pixely v okolí zařazeny ke vznikající oblasti, udává kombinace statického a dynamického kritéria a také zadaný práh. Výpočet hodnoty testovaného pixelu pro prahování (také pro přiřazení k narůstající oblasti) je kombinací zmíněného statického a dynamického kritéria a probíhá dle vzorce (5.1)

$$h = h_d * d + h_s * (1 - d) \quad (5.1)$$

kde h značí výslednou hodnotu pixelu vstupující do prahování, h_d původní hodnotu pixelu, d dynamickou kritériální váhu a h_s hodnotu na počátku zvoleného bodu uživatelem do středu dlaně. Hodnota h je tedy na závěr porovnána se zvoleným prahem a dle toho k oblasti přiřazena.

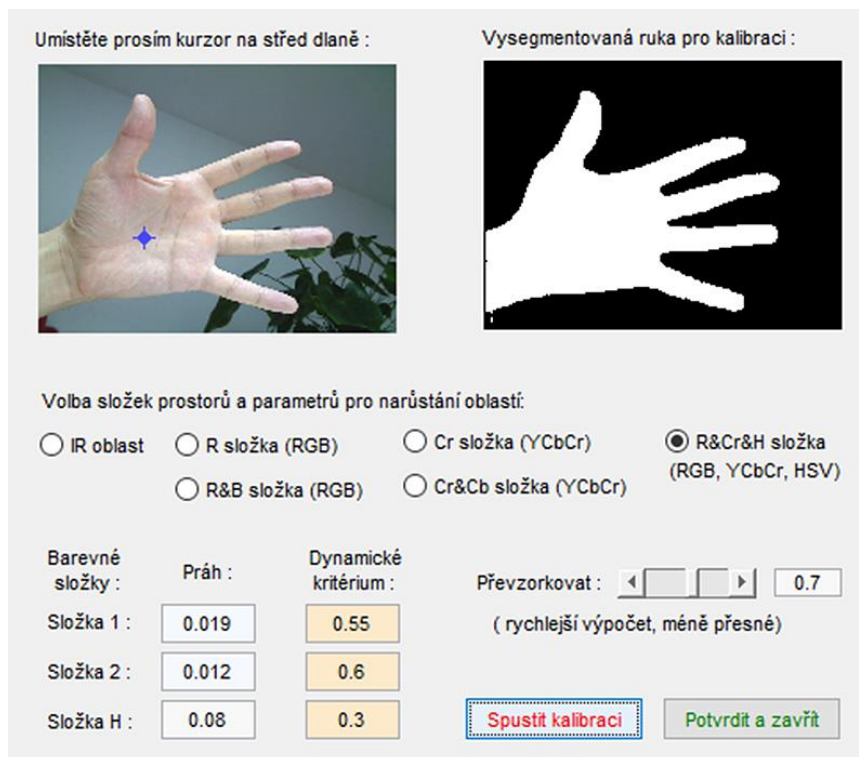
Segmentace metodou narůstání oblastí je provedena v infračervené oblasti na šedotónovém obrázku, v případě viditelné oblasti je možno volit několik variant. Pro omezení závislosti na osvětlení je možné použít normalizovaný RGB barevný prostor, $YCbCr$ barevný prostor nebo barevný prostor HSV. První a nejméně náročnou výpočetní variantou, která není ale příliš robustní, je volba barevného normalizovaného RGB prostoru. V tomto prostoru může segmentace probíhat na normalizované složce R nebo na obrázcích normalizovaných složek R a B, kde se výsledné segmentace mezi sebou vynásobí. Další variantou je barevný prostor $YCbCr$ a využití samostatné složky C_r nebo kombinace složek C_r a C_b . Poslední, výpočetně nejnáročnější a nejrobustnější variantou je kombinace normalizované složky R, složky C_r a složky H prostoru HSV. Doporučené prahy a dynamické kritériální váhy pro jednotlivé složky se nachází v tabulce (Tab. 5.1).

Tab. 5.1: Doporučené hodnoty pro práh a dynamickou kriteriální váhu

Barevný prostor	IR	nRGB			YCbCr			nRGB, YCbCr a HSV		
Zvolená složka		nR	nR	nB	C _r	C _r	C _b	nR	C _r	H
Práh	0,025	0,02	0,022	0,022	0,015	0,012	0,009	0,019	0,012	0,08
Dynamická kriteriální váha	0,4	0,4	0,55	0,55	0,6	0,6	0,6	0,55	0,6	0,3

Po metodě narůstání oblastí následuje vyplnění děr pomocí matlabovské funkce *imfill*. Pro rychlejší výpočet je možné před segmentací snímky převzorkovat (méně přesné). Použití antialiasingového filtru není nutné, jelikož se aliasing na úspěšnosti segmentace příliš nepodílí.

Ze získané vysegmentované ruky jsou pro jednotlivé složky barevných prostorů vypočítány normalizované histogramy. Výstupní hodnoty kalibrace jsou získány porovnáním hodnot normalizovaných histogramů se zvolenými pravděpodobnostními prahy pro jednotlivé složky barevných prostorů. Tím je omezen výskyt chyb způsobených špatnou segmentací. Pravděpodobnostní prahy pro jednotlivé složky normalizovaných histogramů jsou následující: infračervený prostor 0,01; R složka 0,001; G složka 0,002; B složka 0,003; Cr složka 0,003 a Cb složka 0,005. Barevný prostor HSV se dále pro segmentaci nepoužívá z důvodu výpočetní náročnosti převodu z RGB prostoru. Vytvořené aplikační rozhraní určené pro kalibraci lze pozorovat na obrázku (Obr. 5.2).

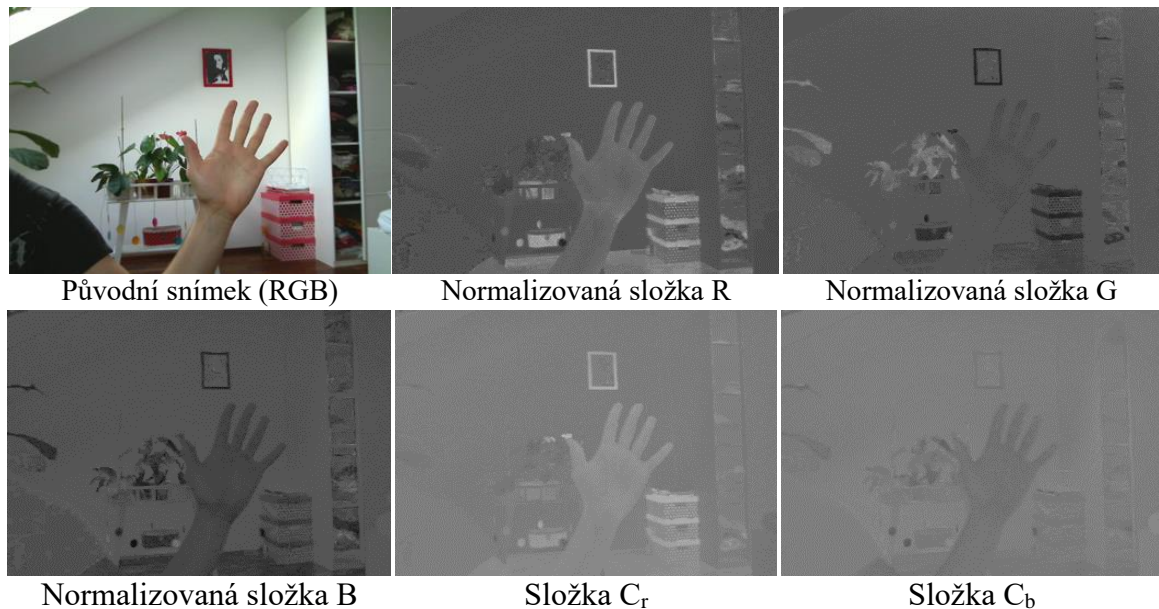


Obr. 5.2: Aplikační rozhraní pro kalibraci

5.2 Detekce ruky

Podobně jako u kalibrace je před samotnou segmentací u snímků získaných kamerou snímající ve viditelné oblasti možné provést mediánovou filtraci pro odstranění kompresních artefaktů. Je možno použít okno velikosti 3 nebo 5, větší hodnoty už mají výrazný vliv na rychlost celého algoritmu a to z důvodu nutnosti provádět mediánovou filtraci pro všechny tři barevné RGB složky zvlášť.

Segmentace v infračervené oblasti zahrnuje prosté prahování a to dle získaných prahových hodnot z kalibrace. Tyto získané prahy je zpravidla nutné upravit, jelikož nemusí zahrnovat veškeré hodnoty pixelů nacházející se v oblasti ruky. Je to dáno posledním krokem kalibrace, kdy méně pravděpodobné hodnoty získané ze segmentace nejsou do výsledných prahů zahrnuty. K získané nižší hodnotě z kalibrace je vhodné v infračervené oblasti odečíst hodnotu 0.08 k vyšší naopak tuto hodnotu přičíst.



Obr. 5.3: Snímek získaný webovou kamerou Logitech S7500 a zobrazení jeho jednotlivých složek v různých barevných prostorech

Ve viditelné oblasti lze provádět prahování (pro omezení závislosti segmentace na osvětlení) v normalizovaném barevném prostoru RGB a prostoru $Y C_b C_r$ (snímek převedený do těchto prostorů a jeho jednotlivé složky lze pozorovat na obrázku (Obr. 5.3)). Před samotným prahováním je opět vhodné upravit prahy získané z kalibrace. Nejrychlejší a nejméně robustní varianta je prahování v normalizovaném prostoru RGB a to pouze normalizované složky R. Z celého normalizovaného RGB prostoru je tato složka pro detekci ruky nejvýznamnější. Pro omezení špatné segmentace je možno pak použít prahování všech tří normalizovaných složek R, G a B. Výsledná segmentace je

získána vynásobením výsledků prahování těchto tří složek. Další možností je prahování v barevném prostoru YC_bC_r . Převod do tohoto prostoru je výpočetně náročnější avšak výsledky prahování mohou být v některých případech lepší (viz kapitola Hodnocení). Nejvýznamnější pro detekci jsou hodnoty složky C_r . Jako doplněk pro omezení chyb v segmentaci lze zvolit složku C_b . V aplikaci je možné zvolit také kombinaci normalizované barevné složky R (z prostoru RGB) a C_r (prostor YC_bC_r). Jedná se o nejvíce výpočetně náročnou úlohu, jelikož je nutné provést převod z prostoru RGB do normalizovaného prostoru RGB a také do prostoru YC_bC_r .

5.3 Úpravy po segmentaci

Úpravy po segmentaci se týkají spíše snímků získaných ve viditelné oblasti. Následující popsané úpravy jsou seřazeny za sebou, jejich pořadí nelze měnit, lze je ale z procesu zpracování vyřadit.

Jelikož ruha je největším a hlavním objektem zájmu na segmentovaném snímku, je vhodnou operací ponechání pouze největšího objektu. Proveďte se hledání všech spojených objektů v osmiokolí pomocí matlabovské funkce *bwconncomp*, nalezně se největší objekt a všechny menší jsou vynulovány. Ze všech úprav segmentace je tato metoda nerychlejší a její použití je téměř nutností (pro viditelnou oblast). Pro odstranění drobných artefaktů algoritmus umožňuje použít mediánovou filtraci s volitelnou velikostí okna. Zde bude velikost okna ovlivňovat velikost artefaktů ale také celkovou rychlost. Následující operací je binární otevření. Jak bylo již zmíněno, otevření má tendenci k eliminaci malých výběžků objektů do pozadí obrazu, zanechává stejně velké výběžky pozadí do objektu a má tendence k rozpojování propojených objektů. Otevření provádí matlabovská funkce *imopen*. Vhodný objekt ve tvaru disku a jeho velikost je definována funkcí *strel*, kde velikost má vliv na filtrační schopnost a rychlost algoritmu. Pro úpravy lze využít i matlabovské funkce *imfill* a to pro vyplnění děr. Popsané operace je možné pozorovat na obrázku (Obr. 5.4).



Obr. 5.4: Snímky s postupnými úpravami po segmentaci (získáno webovou kamerou Logitech S7500)

5.4 Úpravy pro Hammingovu síť

Před vstupem zpracovávaného obrázku do Hammingovy sítě je nutné provést vhodný ořez předloktí a natočení ruky do požadované polohy.

Velikost ořezu předloktí se hledá dvěma způsoby. Prvním způsobem je nalezení středu dlaně, otočení ruky o požadovaný úhel a zjištění vzdálenosti středu dlaně od středu takto rotovaného a ořezaného snímku. Druhým způsobem (méně přesný), který se používá, pokud není nalezen střed dlaně, je zjištění poměru šířky zápěstí a celkové délky správně rotované a částečně erodované ruky. Oba způsoby budou dále podrobně popsány.

Společným bodem obou přístupů je na začátku algoritmu hledání přibližného umístění předloktí. Je to z důvodu nastavení správné rotace objektu a ořezu. Předpokládá se, že předloktí vychází z levého, horního, pravého, nebo spodního kraje obrázku a zbytek ruky se nedotýká okraje. Přibližné umístění je zjištěno tak, že se z každé strany kraje počítá suma okrajů širokých 4 pixely. Největší suma pak udává, z jakého okraje předloktí vychází. V případě že je suma stran stejná, bere se pro umístění první hodnota. Ojedinele může nastat, že suma všech stran bude nulová a to pokud segmentace neproběhne správně, nebo předloktí je zakryto rukávem. V tomto případě se nejdříve provede ořez obrázku na obdélníkový tvar tím způsobem, aby všechny čtyři okraje obsahovaly alespoň jeden pixel segmentovaného objektu (odstraní se tedy okraje s nulovými hodnotami). Na ořezaném snímku je umístění předloktí hledáno stejným již popsaným způsobem. Umístění předloktí ale může být druhým způsobem detekováno chybně.

Pro další zpracování je nutné provést odstranění prstů a případně i vyskytujících se artefaktů. Vhodnou metodou je provedení morfologické operace eroze. Pro tuto operaci je nutné definovat velikost a tvar objektu, kdy velikost objektu musí být dostatečná na to, aby byly odstraněny prsty, ale dlaň a předloktí zůstalo zachováno. Velikost objektu pro erozi lze přibližně zjistit z šířky zápěstí (případně dlaně), jelikož šířka zápěstí bývá u různých gest přibližně stejná a závisí pouze na vzdálenosti od kamery. Přibližná šířka je zjišťována pomocí sumy snímku a to zvláště po sloupcích a řádcích. Jsou získány dva vektory hodnot a v těchto dvou vektorech je hledán medián. Menší medián z obou hodnot (šířka předloktí) slouží pro určení velikosti objektu eroze. V případě, že je ve scéně malá část předloktí, může se jednat o šířku dlaně (místo šířky předloktí) a objekt je nepatrně větší (po řadě testování bylo však zjištěno, že výsledek po ořezání a rotaci není výrazně odlišný).

Objekt je v MATLABU definován dle již zmíněné funkce *strel*, kde vhodným definovaným tvarem pro tento případ může být *disk* nebo *diamond*. Výsledky eroze ruky při volbě obou objektů jsou podobné, avšak v případě použití tvaru *diamond* je eroze o poznání rychlejší a z tohoto důvodu je tento tvar také dále používán.

Po erozi je získán počáteční erodovaný objekt, u kterého je zjištěn úhel rotace pomocí funkce *regionprops* a parametru *orientation*. Funkce vrací úhel mezi osou x a hlavní osou elipsy se stejným druhým centrálním momentem jako má objekt. Hodnoty leží v rozmezí -90° až 90° . Cílem je otočit objekt dle nalezeného předloktí tak, aby byl tento počáteční erodovaný objekt přibližně kolmý k okraji, kde bylo nalezeno předloktí. To lze provést pomocí funkce *imrotate* následujícím způsobem.

Je zjištěno, zda je úhel kladný nebo záporný. V případě umístění předloktí vlevo a vpravo je vstupem do funkce nalezený úhel násobený -1 . Pokud se předloktí nachází nahoře nebo dole, odpovídá tento úhel následující rovnici (5.2)

$$\beta = z * (90 - |\alpha|) \quad (5.2)$$

kde β odpovídá úhlu vstupujícího do funkce *imrotate*, z je znaménko úhlu α , který je výstupem z funkce *regionprops*. Spolu s erodovaným objektem je rotován i původní neerodovaný obrázek. Výsledek těchto operací u vybraného gesta je na obrázku (Obr. 5.5)



Obr. 5.5: Obrázek ruky po počáteční rotaci, ořezu a otočení (bílá barva). Šedá barva zobrazuje rotovanou původní ruku. Snímek před zpracováním byl nasnímán kamerou FLIR A655sc v rozlišení 640 x 480 pixelů

5.4.1 Nalezení středu dlaně

Nalezení středu dlaně probíhá pomocí postupné eroze a nalezení *centroidu* erodovaného objektu.

Navazující postupná eroze se provádí na počátečním erodovaném a otočeném objektu. Pro postupnou erozi je definován objekt o velikosti jedna tvaru *diamond*. Eroze probíhá v nekonečném cyklu, který je ukončen několika podmínkami. Jednou z podmínek je rozdělení objektu na dva. Vznik dvou objektů je způsoben zápěstím ruky, nebo také vzniklými artefakty při erozi. Artefakty jsou filtrovány dle zjištění velikosti jejich plochy. Pokud po odstranění artefaktu zůstávají dva objekty a rozdělení nebylo dosaženo na začátku eroze (zjištěno pomocí poměru plochy počátečního erodovaného objektu a plochy právě erodovaného objektu), jedná se pravděpodobně o rozdělení objektu vlivem zápěstí a eroze končí. Další podmínkou ukončení eroze je získání dostatečně malého objektu. Velikost objektu pro ukončení je dána poměrem plochy objektu počátečního erodovaného objektu a plochy právě erodovaného objektu (výsledkem je jeden objekt). Ukončení eroze popsány podmínkami je vyobrazeno na obrázku (Obr. 5.6).

V průběhu eroze je hledána šířka zápěstí (dlaně) pro zjištění velikosti ořezu dle druhého způsobu (poměru šířky zápěstí a celkové délky správně rotované a částečně erodované ruky) a to dle sumy obrázku po řádcích a po sloupcích (získány dva vektory). Z těchto vektorů jsou vypočteny mediány a je vybrána menší hodnota. Postup je tedy podobný, jako při získání vhodné velikosti objektu pro počáteční erozi. V případě dosažení menšího mediánu o velikosti 20 je uložen částečně erodovaný objekt pro další zpracování.



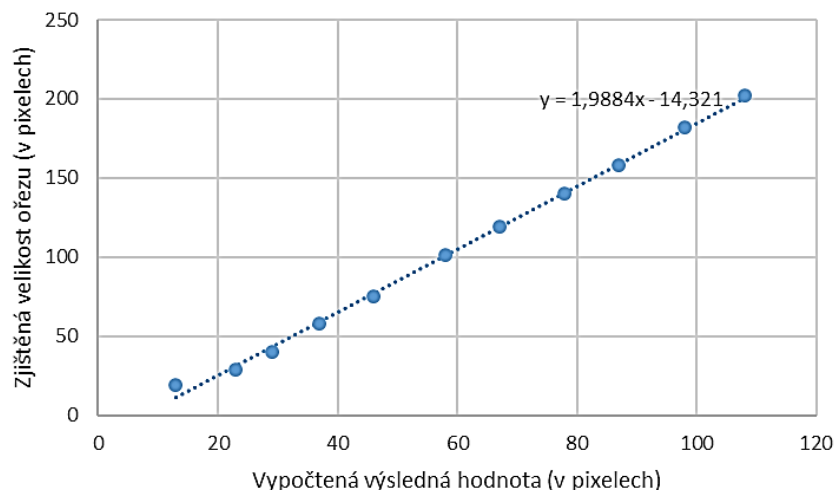
Obr. 5.6: Ukončení eroze – bílá barva označuje výslednou erozi; vlevo je objekt rozdělen na dva vlivem zápěstí, vpravo je eroze ukončena po dosažení dostatečně malého objektu. Původní snímek byl nasnímán kamerou FLIR A655sc v rozlišení 640 x 480 pixelů

Pokud byla ruka rozdělena zápěstím, je nutné definovat, jaký objekt náleží předloktí a jaký dlani. U obou objektů je nalezen střed ve smyslu osy x a y . Tedy je brán průměr indexů jednotlivých objektů na ose x a ose y . Tyto dva středy jsou porovnávány s umístěním dříve zjištěného předloktí a je ponechán objekt, který se nachází od tohoto umístění ve smyslu osy x nebo y dále.

Z výsledného jediného objektu je počítán střed dlaně funkcí *regionprops* s parametrem *centroid*. Dále je nalezen střed obrázku z počátečního erodovaného a otočeného obrázku. Na závěr se vypočítá rozdíl mezi tímto středem a zjištěným středem dlaně. Zjištěnou výslednou hodnotu je nutné převést na velikost ořezu dle vhodné funkce. Za tímto účelem je naměřeno u jednoho konkrétního gesta s různými variantami délky předloktí jedenáct hodnot (algoritmem získané rozdíly), pro které je zjištěna velikost ořezu (manuálně dle vizuální kontroly). Ze získaných hodnot je sestavena rovnice regrese, která má podobu lineární funkce a výsledný tvar (5.3):

$$y = 1,9884 * x - 14,321 \quad (5.3)$$

kde y značí velikost ořezu v pixelech a x hodnotu vypočítanou rozdílem středu obrázku a středu dlaně. Získané hodnoty pro několik variant předloktí s doplněnou rovnicí lze pozorovat na obrázku (Obr. 5.7).



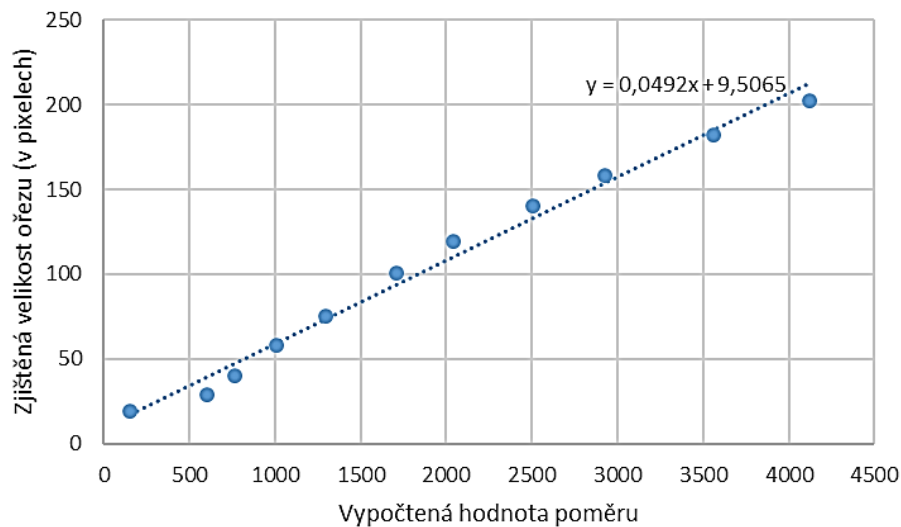
Obr. 5.7 : Získané hodnoty pro několik variant předloktí – vypočtená výsledná hodnota (rozdíl vypočítaný algoritmem) a zjištěná požadovaná velikost ořezu (manuálně) a proložená funkce

5.4.2 Výpočet ořezu dle poměru šířky a délky

Problém může nastat, pokud je ruka pootočená tak, že dlaň po erozi zmizí zcela a zůstane pouze erodované zápěstí. Lze to zjistit jednoduše tak, že po skončené erozi by se objekt neměl dotékat nebo být blízko žádného z okrajů obrázku. Pokud tento případ nastane, počítá se velikost ořezu pomocí druhé varianty z poměru zjištěné šířky částečně erodovaného zápěstí (dlaně) a částečně erodované délky ruky. Délka erodované ruky pro výpočet je získána z předešlého uloženého částečně erodovaného objektu, kde je velikost šířky předloktí (dle mediánu) vždy 20. Poměr se ve výsledku počítá jako délka podělená šířkou (mediánem). Výsledný poměr je opět nutné přepočítat podle vhodné funkce na velikost ořezu. Jako u předešlé varianty bylo naměřeno u různých variant délky předloktí jedenáct hodnot poměru (algoritmem) a zjištěny velikosti ořezu (manuálně dle vizuální kontroly). Ze získaných hodnot byla opět vytvořena lineární funkce výsledného tvaru (5.4)

$$y = 0,0492 * x + 9,5065 \quad (5.4)$$

kde y představuje velikost ořezu a x vypočítanou hodnotu poměru. Funkci s naměřenými a zjištěnými hodnotami lze pak pozorovat na obrázku (Obr. 5.8).



Obr. 5.8 : Získané hodnoty pro několik variant předloktí – Vypočtené hodnoty poměru (algoritmem), zjištěné velikosti ořezu (manuálně) a proložená lineární funkce

Aby mohlo být provedeno porovnání výsledného obrazu s databází gest Hammingovou sítí, je nutné, aby měl obrázek vždy stejné rozměry. Z důvodu možného různého natočení ruky je vhodné dodržet rozměry čtverce a zvolit takové rozlišení obrázku, pro které bude Hammingova síť dostatečně rychlá a schopná rozlišit jednotlivé prsty a jejich polohu. Zpracování na výsledné rozlišení 220 x 220 pixelů, které předchozí podmínky splňuje, probíhá následovně. Je zjištěn rozměr obrázku, kde je vybrána větší hodnota. Zjistí se poměr převzorkování jako podíl hodnoty výsledného rozlišení a většího rozměru. Tento poměr je následně použit pro převzorkování ve funkci *imresize*. Na závěr se provede doplnění do čtverce nulami. Vzniklý aliasing není pro rozpoznání významný.

Popsaná eroze je výpočetně nejnáročnější částí celého algoritmu. Bez větší ztráty spolehlivosti, lze vstupní obrázek, sloužící pouze pro erozi, převzorkovat na polovinu původní velikosti a tím výrazně zvýšit rychlost. Aliasing opět není významný. Operaci je mírně ovlivněna pouze velikost ořezu, obrázek, který je ořezáván, se nemění.

Vytvořená aplikace umožňuje doladění míry ořezu a rotace objektu (zadaná hodnota násobí vypočítanou hodnotu algoritmem), což může být prospěšné především při vytváření vhodné databáze gest.

5.5 Hammingova síť

Závěrečné rozpoznání gest je provedeno pomocí Hammingové neuronové sítě, pro kterou je vytvořena databáze gest (vzorů). I když se předpokládá, že vstupní gesto do sítě je předchozím způsobem upraveno tak, že by pro porovnání měla stačit pouze jedna jeho varianta, používá se pro zvýšení robustnosti variant více (6 variant). Každé gesto má různé varianty natočení, posunu a velikosti.

Algoritmus začíná postupným načtením databáze gest pomocí vytvořené funkce *nacti_databazi*. Databáze se sestává z gest, které představují jednotlivé složky a ve složkách se nachází jednotlivé varianty gest ve formátu *BMP (Windows Bitmap)*. Formát *BMP* byl vybrán z důvodu, že je podporován většinou aplikací, nepoužívá žádnou kompresi a načítání tedy probíhá velice rychle bez vzniku zkreslení. Databázi lze pomocí vytvořené aplikace velice rychle a jednoduše upravovat, přidávat nová gesta, k současným gestům přidávat více jejich variant a tím zvyšovat celkovou robustnost. Funkce *nacti_databazi* načítá všechna gesta a jejich varianty a přitom jednotlivé hodnoty pixelů (rozsah 0 až 1) obrázku převádí pro Hammingovu síť na hodnoty -1 (hodnota pixelu je menší nebo rovna 0.5) a 1 (větší jak 0.5). Takto upravené obrázky ukládá jako proměnnou *databaze_gest* datového typu *struct* do souboru *vzory.mat* a složky *Databaze*. Vytvořený soubor je velice malý a v případě, že uživatel nepotřebuje databázi nadále upravovat, je možné jej samostatně používat (bez nutnosti obrazové databáze).

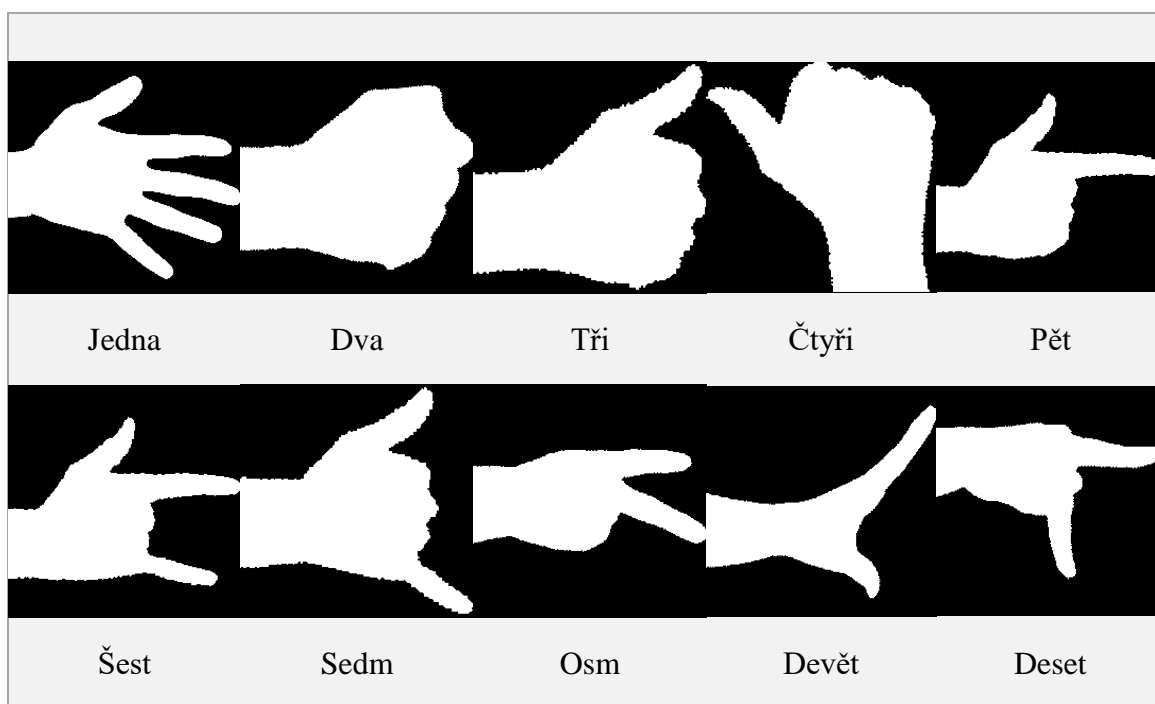
Poslední část algoritmu provádějící rozpoznání gest převádí gesto pro rozpoznání na sloupcový vektor, kde každou hodnotu pixelu binarizuje stejným způsobem jako je prováděno u zpracování databáze a to na hodnoty -1 a 1 (opět dle prahu 0.5). Následně je načten soubor *vzory.mat* obsahující databázi gest. Jednotlivá gesta z databáze (jednotlivé obrazy), jsou před samotným porovnáním převedeny na vektory. Hammingova síť pak stanovuje podobnosti vstupního N -rozměrného vektoru (zpracované gesto získané z kamery) s vektorem vzoru z databáze. Výpočet výstupu jednoho neuronu, který představuje jeden vzor z databáze, probíhá dle rovnice (3.9). Jednotlivé váhy w_n odpovídají jednotlivým hodnotám aktuálního vektoru (jednotlivého obrazu) z databáze gest. Výsledná Hammingova vzdálenost (vyjadřující míru nepodobnosti mezi gesty) pro jednotlivé vzory se určí jako $N - y_m$, kde y_m je výstup aktuálního neuronu představujícího vzor z databáze. Nejbližší nalezené gesto je na závěr určeno jako minimální Hammingova vzdálenost z takto vypočítaných hodnot. Tedy gesto, které je k testovanému vzoru z databáze nejbližší, má nejmenší Hammingovu vzdálenost.

Samotná aplikace také umožňuje označit gesto za nerozpoznané a to dle nastaveného prahu pro zjištěnou minimální Hammingovu vzdálenost (v aplikaci označeno jako

senzitivita). Snížením hodnoty se dosáhne větší specifity a menší senzitivity v rozpoznání, což znamená, že gesto nemusí být vždy rozpoznáno, ale pokud rozpoznáno je, s vysokou pravděpodobností odpovídá správně rozpoznanému gestu. Zvýšením hodnoty roste senzitivita a klesá specifita, tedy gesto s větší pravděpodobností bude rozpoznáno, ale může být označeno chybně.

5.6 Sestavení databáze gest

Databáze gest byla vytvořena s ohledem na jednoduché a rychlé provedení gest a s důrazem na dostatečnou rozdílnost mezi gesty (kvůli omezení záměn). Gesta jsou navržena pro levou ruku, protože se předpokládá, že pravá se bude věnovat jiné činnosti (pohybu myši, držení nástroje apod.). Jelikož vybraná zařízení používají pro snímání různou oblast a mají jiné parametry, může se výsledná segmentace odlišovat. Z toho důvodu bylo nutné pro každé zařízení vytvořit jedinečnou databázi. Z důvodu zvýšení robustnosti algoritmu bylo nutné také pro každé gesto sestavit několik jeho variant. Základní sestava gest je vyobrazena na obrázku (Obr. 5.9).



Obr. 5.9: Navržená databáze gest s pojmenováním gest dle pořadí sloužící pro klasifikaci (získané snímky jsou z kamery FLIR A655sc)

6 HODNOCENÍ

Navržený algoritmus bylo nutné nejdříve ověřit na databázi gest z hlediska funkčnosti a robustnosti. Dále bylo třeba provést jeho srovnání s ostatními přístupy popisovanými v literatuře a shrnout výhody a nevýhody, které nabízí.

Po této části následovalo samotné hodnocení, které zahrnuje testy na použitelný rozsah vzdáleností rozpoznání gest u jednotlivých zařízení a také rychlost a úspěšnost algoritmu pro jednotlivá nastavení a různá prostředí.

6.1 Ověření algoritmu na testovaných gestech

Pro ověření realizovaného algoritmu byla vybrána kamera FLIR A655sc, která z obou zařízení snímajících infračervené oblasti poskytuje nejkvalitnější snímky a rozlišení. Pro testování bylo získáno osm verzí videí, ve kterých se za sebou provádí deset jednotlivých gest. Získaná videa měla rozlišení 640x480 pixelů se snímkovací frekvencí 25 fps.

Ze šesti videí byla sestavena databáze deseti gest, kde každé gesto mělo šest variant. Varianty gest byly pečlivě vybrány s ohledem na co možná největší robustnost algoritmu. Bylo tedy nutné vybrat takové varianty, aby gesta byla správně rozpoznána i když jsou prováděna mírně odlišně (například odlišnosti mezi vzájemnou polohou prstů), nebo je výsledná rotace a velikost předloktí nepatrně odlišná z důvodu zpracování a malé části předloktí na scéně.

Ověření algoritmu probíhalo ve verzi MATLABU *R2014a* na starším notebooku Asus M50VN s parametry uvedenými v tabulce (Tab. 6.1).

Tab. 6.1: Parametry notebooku

Model počítače:	Procesor:	Grafická karta:	RAM:	Rok výroby:
Asus M50VN	Core 2 Duo T9400 (2.5 GHz)	GeForce 9650M GT (bez akcelerace pro MATLAB)	4 GB	2008

Testování algoritmu bylo provedeno pro dvě vybraná videa při třech různých nastaveních, která jsou stěžejní pro ovlivnění rychlosti algoritmu. Ověření bylo zaměřeno na úspěšnost a správné rozpoznání gest, rychlost byla testována až v kapitole Hodnocení.

Při první variantě nastavení byl kladen důraz na výslednou maximální kvalitu zpracovávaných snímků (a tedy i úspěšnost rozpoznání). Rozlišení získaného snímku z kamery nebylo převzorkováno, hodnota pro doladění prahu byla 0,12, před a po segmentaci nebyly prováděny žádné úpravy, míra ořezu byla nastavena na hodnotu 1 a míra rotace na hodnotu 0,8. Převzorkování erodovaného objektu bylo vypnuto. Senzitivita pro Hammingovu síť byla zvolena na 5000.

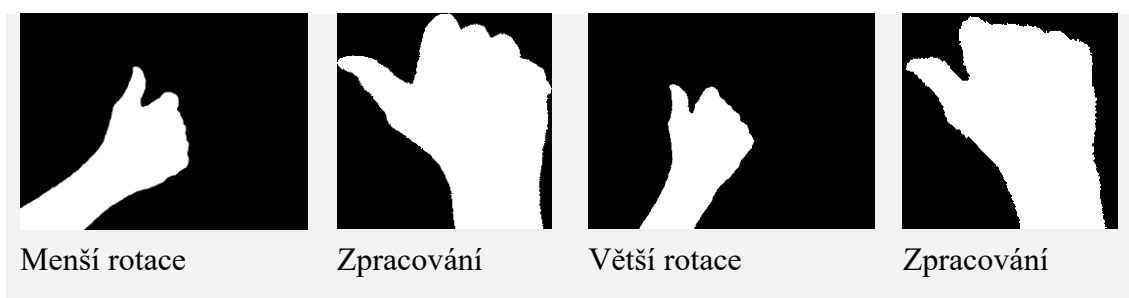
Druhá varianta nastavení se lišila pouze v počátečním převzorkování získaného snímku a to na maximální velikost obrázku 350 px. Třetí varianta zahrnovala počáteční převzorkování a převzorkování objektu pro erozi na poloviční velikost se zachováním nastavení jako u předchozí varianty. Aliasing vzniklý převzorkováním se na výsledku projevoval minimálně, proto nebylo nutné řešit jeho odstranění antialiasingovým filtrem.

Z postupně prováděných gest na videu byly pro každé gesto (mimo čtvrté) vyhrazeny dvě vteřiny (50 snímků), kde měření úspěšnosti rozpoznání začalo vždy na začátku daného gesta.

První video bylo získáno ve vzdálenosti 80 cm dlaně od kamery. To je také nejvhodnější vzdálenost pro zvolenou kameru, kdy lze bez většího soustředění udržet ruku a gesta na snímané scéně a také získat po úpravách dobré rozlišení ruky. První video testovalo rozpoznání gest při dodržené vzdálenosti a délky předloktí na scéně. Robustnost na rotaci gest byla ověřena u čtvrtého gesta, což je gesto tři, otočené o 90 stupňů proti směru hodinových ručiček.

Úspěšnost rozpoznání gest pro všechny tři varianty nastavení byla kromě čtvrtého gesta 100%. Čtvrté gesto bylo rozděleno na první a druhou část gestem, které není v databázi (otočená zařatá pěst). V každé části se opět gesto měřilo dvě vteřiny. Při prvním nastavení (žádné převzorkování) byla úspěšnost rozpoznání v první části 92% (nerozpoznány 4 snímky) v druhé pak 100%. Při druhé variantě nastavení (převzorkování snímku) byla úspěšnost v první části 100% v druhé pak pouze 68% (nerozpoznáno 16 snímků). Třetí varianta měla úspěšnost pro první část 100%, pro druhou část 92%.

Nižší úspěšnost je dána především nároky na správné provedení gesta a vhodně vytvořenou databází. Gesto totiž není zpravidla jednoduché provést tak, aby předloktí vycházelo přibližně kolmo ze spodního okraje snímku, a tudíž velkou část rotace provádí algoritmus. Ideální provedení gesta samozřejmě možné je, ale pak lze diskutovat o jeho použitelnosti, jelikož nemůže být provedeno jednoduše a rychle (což je cílem všech navržených gest). Možností zlepšení úspěšnosti rozpoznání je vytvoření více variant gesta pro databázi. Rozdíly v provedení rotace gesta a výsledky zpracovaného gesta pro Hammingovu síť lze pozorovat na obrázku (Obr. 6.1).



Obr. 6.1: Vliv rotace gesta na výsledné zpracování obrázku pro čtvrté gesto. Lze si povšimnout i mírně jiného úhlu palce mezi zbytkem ruky, což má v součtu nepříznivý dopad na výsledné rozpoznání

Při vyšší senzitivě (Hammingově nejnižší vzdálenosti 6000) by gesto bylo rozpoznáno správně ve 100 % případech, avšak s přidáním nových gesty by mohlo docházet ke špatnému zařazení.

Cílem druhého videa bylo ověřit robustnost celého algoritmu. Během gest se ruka přibližovala a oddalovala od kamery, výrazně posouvalo předloktí (ve smyslu do scény a ven) a zároveň byla prováděna mírná rotace. Každé gesto bylo měřeno první tři vteřiny (75 snímků) od počátku jeho provedení. Výsledky se nacházejí v tabulce (Tab. 6.2)

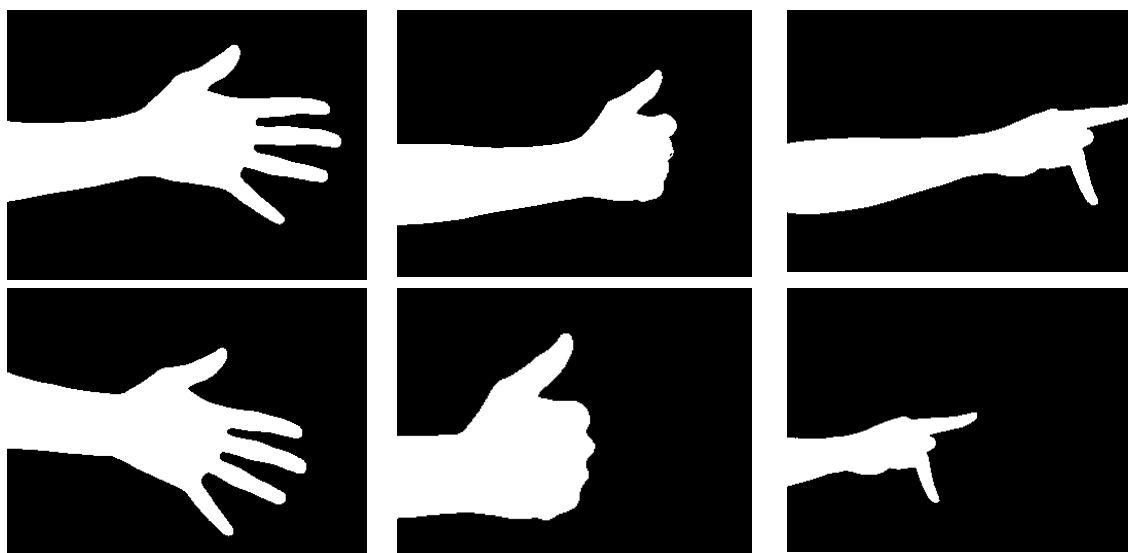
Tab. 6.2: Úspěšnosti rozpoznávaných a správně zařazených gest (vyjádřené v procentech) pro jednotlivé varianty nastavení

Nastavení	Úspěšnost (%)	Gesto									
		1	2	3	4	5	6	7	8	9	10
Varianta 1	Rozpoznáno	100	100	98,7	64	94,6	90,7	100	100	88	100
	Správně zařazeno	100	100	100	100	100	92,6	100	100	100	100
Varianta 2	Rozpoznáno	100	100	96	64	100	88	100	100	100	100
	Správně zařazeno	100	100	100	100	100	90,9	100	100	100	100
Varianta 3	Rozpoznáno	100	100	100	64,3	100	89,3	100	100	100	100
	Správně zařazeno	100	100	100	100	100	97,0	100	100	100	100

Úspěšnost byla nejmenší u gesta čtyři (64% pro první a druhou variantu, 64,3% pro třetí variantu nastavení), což bylo zapříčiněno opět nedostatečně správně provedeným gestem ve smyslu natočení. Menší úspěšnost byla také při rozpoznání gesta šest, kdy navíc docházelo k záměně s gestem sedm. Procentuální správné zařazení gest bylo

počítáno z rozpoznaných gest a nejmenší úspěšnost gesta šest byla u varianty nastavení 2, kde nesprávně zařazených snímků bylo 5 z 68 rozpoznaných gest. Špatné zařazení gesta je způsobeno především tím, že jsou si obě gesta dosti podobná. Pro tato gesta by tedy bylo vhodné vložit do databáze více variant.

Jednotlivé ukázky segmentovaných objektů (ruk) pro ověření robustnosti algoritmu z druhého videa jsou na obrázku (Obr. 6.2).



Obr. 6.2: Znáznornění provedených gest s různým natočením, vzdáleností od kamery a s rozdílnou délkou předloktí zasahující do scény. Všechna zobrazená gesta byla rozpoznána a správně zařazena

6.2 Srovnání algoritmu s ostatními přístupy a hodnocení zvoleného přístupu

Realizovaná metoda je s ostatními přístupy shodná v části segmentace, kde je navíc rozšířena o možnost vytvoření vlastního vzorku ruky pomocí kalibrace. V dalších částech je ale vytvořený algoritmus poněkud odlišný.

Algoritmus vyžaduje pro správný běh pouze nalezení správné rotace ruky a středu dlaně, případně šířky předloktí nebo dlaně. Určité omezení lze nalézt v tom, že na scéně se musí vyskytovat pouze předloktí s dlaní a prsty, v případě loktu, paže nebo i jiných částí těla by nemuselo docházet ke správné rotaci (pokud je loket více ohnutý) nebo nalezení středu dlaně a tím by mohl být nesprávně ovlivněn ořez. Nicméně tento nedostatek ve většině případů neřeší ani jiné studie.

V jiných přístupech je často hledán střed dlaně a dále pak umístění prstů například pomocí metody vyhledávání vzorů (vzorem bývá vhodně zvolený kruh), případně dále i zápěstí či jiné části ruky [16]. Z těchto detekovaných částí ruky jsou získávány příznaky, jako je například počet natažených prstů, úhly mezi nimi, vzdálenosti od středu dlaně, nebo zjišťování různých vzdáleností od kontury ruky, jako bylo popsáno v kapitole Metody používané pro rozpoznání gest ruky.

Navazujícím krokem je získané příznaky správně klasifikovat, což je nejčastěji prováděno vícevrstvou perceptronovou sítí. Tuto síť je ale třeba pro správnou klasifikaci a tedy rozpoznání gesta natrénovat, což při větším počtu gest a příznaků trvá velice dlouho. Navíc při přidání jakéhokoliv dalšího gesta do databáze je nutné kompletní přeučení sítě.

Extrakce příznaků v realizovaném algoritmu této diplomové práce není třeba. Pozornost je zaměřena především na správné zpracování gesta tak, aby bylo možné ořezané a otočené gesto dále srovnávat pomocí Hammingové neuronové sítě s dostupnou databází gest. Hammingovu neuronovou síť není nutné trénovat, jelikož informace o vzorech v databázi je zapamatována ve váhových hodnotách a počet vzorů se v této síti rovná počtu neuronů. Díky tomuto faktu je velice jednoduché a rychlé gesta do databáze přidávat, nebo je měnit. Navíc gesto není limitováno detekcí prstů, zápěstí či jiných částí ruky a ruka může být libovolně rotována (ve smyslu pronace a supinace). V případě že není gesto správně rozpoznáno, lze do databáze přidat jeho novou variantu a jeho rozpoznání po této operaci je okamžité a správné.

Nevýhodou je velká závislost na správném zpracování ruky (rotace a ořez) a s tím souvisí i nutnost vytvořit velice pečlivě i více variant gest (a k tomu i varianty vztahující se k samotnému provedení gesta). S přidáním gesty a variantami ale klesá rychlost rozpoznání (je nutné násobit navzájem několik vektorů). Rychlost lze ovlivnit velikostí vzorů a z několika testů také vyplynulo, že by mohlo být rozlišení i daleko menší než použité 220 x 220 pixelů (například i 120 x 120 pixelů).

Nicméně i přes zmíněné nedostatky je navržená metoda poměrně dobře funkční a co se rychlosti týče, splňuje i požadavky na zpracování v reálném čase (viz kapitola Hodnocení).

6.3 Volba hardwaru, gest a parametrů snímků pro hodnocení

Pro hodnocení rozpoznání gest ve viditelné oblasti a infračervené oblasti z hlediska hardwarové a softwarové náročnosti a úspěšnosti byly zvoleny již popsané kamery Logitech S7500, FLIR A655sc a modul FLIR Lepton, verze programového prostředí MATLAB R2014a, notebook ASUS M50VN s grafickou kartou nepodporující akceleraci výpočtů v programovém prostředí MATLAB (parametry viz tabulka (Tab. 6.1)). Pro hodnocení byla vybrána čtyři gesta zobrazená na obrázku (Obr. 6.3), která měla v ověření algoritmu 100 % úspěšnost.

Snahou výběru gest pro hodnocení bylo zamezení špatně rozpoznávaného gesta způsobeného podobností gest a nedokonalostmi algoritmu (ořez, rotace). Databáze čtyř gest, byla pro všechna zařízení vytvořena zvlášť a to podobným způsobem, jako tomu bylo při ověření algoritmu na snímcích z kamery FLIR A655sc (nové tři databáze byly sestaveny z 5 videí, které se nepoužívaly dále k testování). Každé gesto obsahovalo 6 variant provedení.



Obr. 6.3 : Gesta vybraná pro hodnocení algoritmu – gesto jedna, dva, sedm a deset

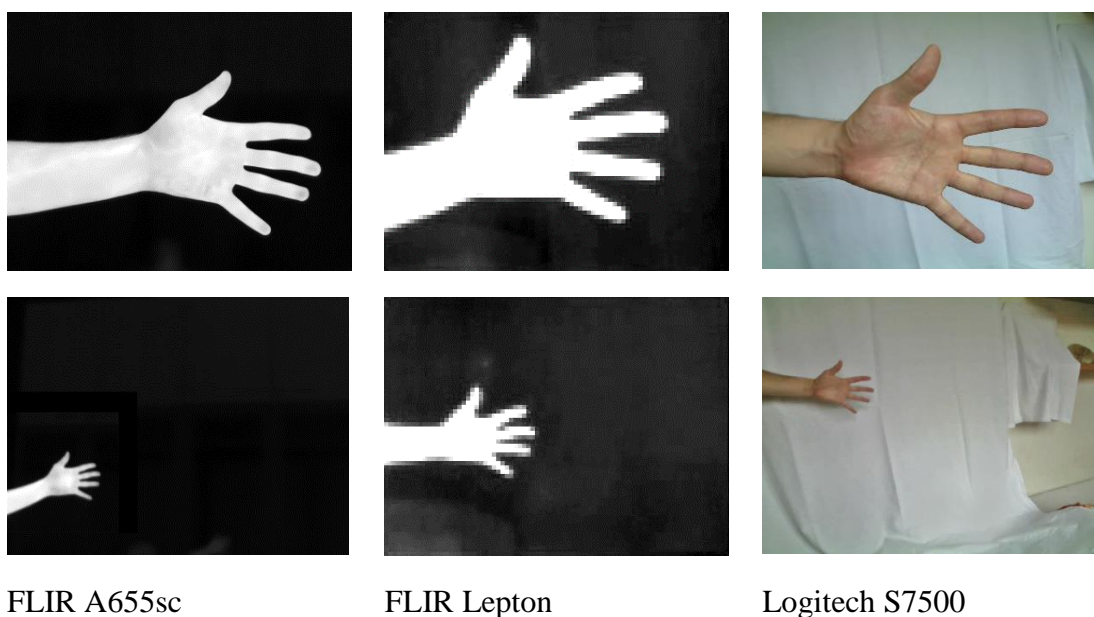
Parametry získaného videa pro kameru FLIR A655sc byly stejné jako v předešlém testování ověření algoritmu. Video získané z kamery Logitech S7500 mělo rozlišení 960 x 720 pixelů se snímkovací frekvencí 15 fps (vyššího rozlišení nebylo možno dosáhnout kvůli absenci USB 3). Z FLIR Leptonu bylo video pořízeno komplikovanějším způsobem, jelikož prozatím neexistuje software, který by uměl v systému Raspbian získané snímky nahrávat. Bylo provedeno nahrání plochy obrazovky programem *RecordMyDesktop* a video bylo získáno ořezem v programovém prostředí MATLAB. Ve výsledku tedy není rozlišení videa 80x60 ale 320x240 pixelů a snímkovací frekvence 15 fps místo původní 8,3 fps. Zkreslení pořízených snímků tedy nastává převzorkováním a kompresí získaného videa (dle vizuální kontroly ale není významné). Z důvodu dalšího zkreslení nebylo video na původní rozlišení převzorkováno, což ale může na druhou stranu mírně změnit výsledky rychlosti zpracování (ve smyslu menší rychlosti). Pro

hodnocení rychlosti bude bráno dále hardwarové omezení původní snímkovací frekvence 8,3 fps.

Účelem testování bylo ověřit použitelný rozsah vzdáleností, na kterých lze u konkrétních zařízení gesta vykonávat, dále pak rychlost a úspěšnost algoritmu pro jednotlivá nastavení a různá prostředí.

6.4 Rozsah vzdáleností ruky od objektivu u jednotlivých zařízení

Pro testování použitelného rozsahu vzdáleností ruky od objektivu u jednotlivých zařízení, bylo vybráno prostředí, ve kterém jsou vlivy popředí, pozadí a osvětlení na segmentované gesto minimální. Ve výsledku tedy hlavně záleží na parametrech zařízení a vytvořeném algoritmu. Pro infračervené prostředí musela být splněna podmínka, kdy má ruka výrazně odlišnou teplotu než okolní objekty. Pro viditelnou oblast musela být zase splněna rovnoměrnost osvětlení scény a pozadí, které se výrazně barevně odlišuje od ruky (například bílá barva). Jednotlivá zvolená prostředí v odpovídající oblasti (IR a VIS) pro zařízení jsou zobrazena na obrázku (Obr. 6.4) v jeho horní části.



Obr. 6.4: Zvolená prostředí pro jednotlivá zařízení, kde probíhalo rozpoznání gest (snímky nahoře). Spodní snímky zobrazují maximální vzdálenost prováděných gest pro jednotlivá zařízení.

Pro testování použitelných rozsahů vzdáleností ruky od objektivu bylo zvoleno následující nastavení. U kamer snímajících v infračervené oblasti bylo nastavení obdobné

jako při ověřování algoritmu a to: rozlišení získaného snímku z kamery nebylo převzorkováno, hodnota pro doladění prahu byla 0,12, nebyly prováděny žádné úpravy před a po segmentaci, míra ořezu byla nastavena na hodnotu 1 a míra rotace na hodnotu 0,8. Převzorkování objektu pro erozi bylo vypnuto. V nastavení pro kameru Logitech S7500 bylo vypnuté opět převzorkování snímku i objektu pro erozi, úpravy před segmentací nebyly žádné, úpravy po segmentaci zahrnovaly ponechání největšího objektu a vyplnění objektu. Pro segmentaci byl vybrán prostor YC_bC_r a složky C_b a C_r (pro toto prostředí to byla vhodnější volba než normalizovaný barevný prostor RGB), doladění prahových hodnot bylo nastaveno na 0,022, přičemž před segmentací byla provedena kalibrace. Senzitivita pro Hammingovu síť ve všech případech byla zvolena na 5000. Úspěšnost rozpoznání gesta byla zvolena na minimální hodnotu 90 %, kde rozpoznání probíhalo na prvních 50 snímcích od začátku provedení gesta. Naměřené maximální a minimální vzdálenosti gest pro jednotlivá zařízení jsou v tabulce (Tab. 6.3). Vybraná gesta prováděná v maximální vzdálenosti od zařízení jsou vyobrazena na obrázku (Obr. 6.4) ve spodní části.

Tab. 6.3: Rozsah vzdáleností použitelných pro rozpoznání gest u jednotlivých zařízení

Kamera	FLIR A655sc	FLIR Lepton	Logitech S7500
Minimální vzdálenost	65 cm	70 cm	25 cm
Maximální vzdálenost	205 cm	150 cm	100 cm
Výsledný rozsah	140 cm	80 cm	75 cm

Určitou limitací kamery FLIR A655sc byla neschopnost automaticky zaostřit na ruku ve vzdálenosti větší než 160 cm, tedy vše vzdálenější mohlo být nepatrně rozostřeno a tím i snížen výsledný použitý rozsah. U kamery Logitech S7500 by se dalo očekávat, že díky vyššímu rozlišení bude výsledný rozsah větší. Limitací je ale použitý širokoúhlý objektiv a čip, který vykazuje poměrně velký šum. Rozsah pro všechna zařízení lze hodnotit jako uspokojivý a uživatel není tedy nucen provádět gesta pouze v určité jedné definované vzdálenosti.

6.5 Rychlost a úspěšnost rozpoznání gest v ideálním prostředí při různých nastaveních

Prostředí, kde byla testována rychlost a úspěšnost algoritmu, bylo zvoleno stejné jako v předchozím testování (gesto není rušeno žádnými vlivy popředí nebo pozadí a scéna je rovnoměrně osvětlena denním světlem).

Pro toto prostředí byly zvoleny tři varianty nastavení. Nastavení pro infračervenou oblast obsahovalo varianty již použité při ověřování algoritmu u kamery FLIR A655c, přičemž pro kameru FLIR Lepton nebylo nutné převzorkování snímků (díky již malému rozlišení snímků). Pro každé zařízení byla použita odpovídající databáze gest. Bylo tedy nutné pro vybrané zařízení vždy databázi přepnout.

U kamery Logitech S7500 bylo nastavení následovné. První variantou bylo zachování původního rozlišení snímků bez převzorkování objektu pro erozi, žádné úpravy před segmentací, zvolený barevný prostor pro segmentaci YC_bC_r se složkami C_b a C_r a úpravou prahových hodnot po kalibraci na 0,022. Jedinou úpravou po segmentaci bylo ponechání největšího objektu. Druhá varianta zahrnovala veškerá předchozí nastavení, pouze byl snímek na počátku převzorkován. Třetí varianta byla totožná s druhou a bylo přidáno převzorkování objektu pro erozi.

U každého gesta probíhalo měření prvních 50 snímků od jeho začátku. Gesta byla prováděna tak, že se ruka celou dobu mírně pohybovala, rotovala a byly prováděny mírně odlišné varianty gest. Vzdálenost ruky od objektivu byla v případě kamery FLIR A655c 80cm, u kamery FLIR Lepton 85 cm a 35 cm v případě kamery Logitech S7500.

Úspěšnost při všech třech nastaveních a u všech zařízení se lišila jen nepatrně (dosahovala více jak 95%). Odchytky byly dány především rozdílnými databázemi a verzemi gest. Rozdíly v rychlosti algoritmu už byly významné.

Nejmenších rychlostí bylo dosaženo při první variantě nastavení, kde nebylo prováděno žádné převzorkování. Výsledky pro jednotlivá zařízení, zvolený notebook a gesta jsou uvedeny v tabulce (Tab. 6.4). Získané časy jsou uvedeny v sekundách.

Tab. 6.4. Změřené časy algoritmu (v sekundách) a jeho částí pro různá zařízení a zvolená čtyři gesta při první variantě nastavení

	Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem
FLIR A655c	1	0,0115	0	0,1540	0,0071	0,1726
	2	0,0113	0	0,1240	0,0069	0,1422
	3	0,0115	0	0,1301	0,0069	0,1485
	4	0,0115	0	0,0940	0,0070	0,1125
FLIR Lepton	1	0,0040	0	0,0483	0,0075	0,0598
	2	0,0042	0	0,0429	0,0069	0,0540
	3	0,0042	0	0,0464	0,0071	0,0577
	4	0,0040	0	0,0364	0,0068	0,0472
Logitech S7500	1	0,1541	0,0299	0,5165	0,0071	0,7076
	2	0,1664	0,0297	0,4746	0,0072	0,6779
	3	0,1645	0,0302	0,5604	0,0075	0,7626
	4	0,1630	0,0276	0,2919	0,0072	0,4897

Z tabulky je patrné, že nejvyšší rychlosti (nejkratší čas) je dosahováno při zpracování snímků z FLIR Leptonu a to díky tomu, že mají snímky výrazněji nižší rozlišení. Již při tomto nastavení je rychlost pro Lepton na starším notebooku dostatečně rychlá na to, aby hlavním omezením byla snímkovací frekvence modulu (8,6 fps).

Rychlost segmentace snímků je výrazně odlišná pro infračervenou oblast a pro viditelnou, což je dáno z velké části převodem RGB prostoru do $YCbCr$ a prahováním obou složek C_b a C_r a také vyšším rozlišením kamery Logitech (720 x 960 pixelů) než má FLIR A655c (480 x 640 pixelů). Úpravy před a po segmentaci nejsou v infračervené oblasti nutné a v rychlosti se projevují především ve viditelné oblasti, kde změřené časy nejsou zanedbatelné. Rychlost zpracování snímků (ořez a rotace) je opět dána rozlišením a projevuje se na celkovém čase nejvíce. Za povšimnutí stojí kratší čas u čtvrtého gesta, kde není třeba pro dosažení rozdělení ruky zápěstím tolik kroků eroze, jako je tomu u jiných gest. Rychlost rozpoznání je pak pro všechna zařízení stejná.

Při převzorkování vstupních snímků při druhé variantě nastavení došlo k výraznému urychlení výpočtů pro kameru FLIR A655sc a Logitech S7500. U FLIR Leptonu převzorkování postrádá význam, jelikož vstupní rozlišení je menší než výsledné požadované. Z toho důvodu není v následující tabulce FLIR Lepton uveden a dosažené změřené časy Leptonu jsou totožné jako v tabulce (Tab. 6.4). Dosažené časy u FLIR A655sc a Logitech S7500 jsou v tabulce (Tab. 6.5).

Tab. 6.5: Změřené časy algoritmu (v sekundách) a jeho částí pro FLIR A655c a Logitech S7500 a zvolená čtyři gesta při druhé variantě nastavení

	Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem
FLIR A655c	1	0,0042	0	0,0541	0,0072	0,0655
	2	0,0041	0	0,0495	0,0076	0,0612
	3	0,0041	0	0,0486	0,0077	0,0604
	4	0,0040	0	0,0384	0,0078	0,0502
Logitech S7500	1	0,0153	0,0025	0,0604	0,0079	0,0861
	2	0,0172	0,0030	0,0724	0,0078	0,1004
	3	0,0155	0,0024	0,0787	0,0076	0,1042
	4	0,0154	0,0023	0,0510	0,0073	0,0760

FLIR Lepton opět drží v rychlosti zpracování prvenství, což je dáno nižším rozlišením snímků. Rozdíly mezi FLIR A655sc a Logitech S7500 jsou především v rychlosti segmentace a ve zpracování. Dosažené rychlosti úprav po segmentaci jsou zanedbatelné. V rychlosti zpracování hraje roli mírně odlišný výstup segmentace a procentuální velikost ruky vyskytující se na scéně (což není jednoduché předem při snímání odhadnout). Jelikož je rozlišení pro zpracování stejné, výsledky v časech by se lišit příliš neměly. I kdyby byl tento fakt brán v potaz, díky rychlosti segmentace bude celý algoritmus stále nejpomalejší u kamery Logitech S7500. Při rychlostech snímání obou kamer je algoritmus běžící na starším notebooku stále určitou limitací, ale již ne tolik výraznou.

Třetí varianta nastavení zahrnuje převzorkování snímků a převzorkování objektu pro erozi a je ze všech variant nejrychlejší. Při dodržených popsanych vzdálenostech od kamery navíc nedochází ke snížení úspěšnosti rozpoznávání, avšak při limitních maximálních vzdálenostech by pravděpodobně byla úspěšnost menší. Výsledky lze opět pozorovat v tabulce (Tab. 6.6).

Tab. 6.6: Změřené časy algoritmu (v sekundách) a jeho částí pro různá zařízení a zvolená čtyři gesta při třetí variantě nastavení

	Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem
FLIR A655sc	1	0,0042	0	0,0315	0,0073	0,0430
	2	0,0041	0	0,0299	0,0064	0,0404
	3	0,0042	0	0,0301	0,0069	0,0412
	4	0,0039	0	0,0256	0,0070	0,0365
FLIR Lepton	1	0,0042	0	0,0304	0,0067	0,0413
	2	0,0039	0	0,0285	0,0069	0,0393
	3	0,0042	0	0,0296	0,0067	0,0405
	4	0,0041	0	0,0254	0,0068	0,0363
Logitech S7500	1	0,0159	0,0027	0,0345	0,0076	0,0607
	2	0,0153	0,0024	0,0387	0,0072	0,0636
	3	0,0158	0,0024	0,0395	0,0075	0,0652
	4	0,0153	0,0025	0,0295	0,0072	0,0545

Celý algoritmus běží nejrychleji na snímcích pořízených z kamer v infračervené oblasti, kde jsou rozdíly v rychlostech obou kamer již nevýznamné. Zpracování ve viditelné oblasti je nejpomalejší opět díky segmentaci. Při získaných videích (FLIR A655sc 25fps, FLIR Lepton 8,6fps, Logitech S7500 15 fps) byla dosažena podmínka dostatečně rychlého běhu algoritmu a na celkovém čase se začínají projevovat hardwarové limity.

Při zběžném testování algoritmu na nejnovějších procesorech Intel Core i5 a i7 (rok vydání procesorů 2015) bez grafické akcelerace výpočtů pro programové prostředí MATLAB byla dosažena rychlost dvakrát až čtyřikrát vyšší než naměřená na testovaném notebooku s procesorem Core 2 Duo T9400. Algoritmus by tedy bylo teoreticky možné na nových počítačích použít bez limitací (rychlosti algoritmu) i s jiným nastavením.

Co se týče úspěšnosti správného zařazení gest, tak pro všechna tři nastavení byla všechna gesta (snímky) správně zařazena.

6.6 Problematická prostředí pro rozpoznání gest pro infračervenou oblast

Schopnost rozpoznání gest pro infračervenou a viditelnou oblast byla testována pro prostředí, která jsou pro segmentaci nějakým způsobem problematická. V infračervené oblasti se jedná především o scénu, kde jsou v pozadí objekty, které mají podobnou teplotu jako má ruka (například zdroje napětí, výpočetní technika, zahřátý povrch od slunečního svitu či jiná osoba). Ve viditelné oblasti to může být barva objektů, které jsou podobné kůži (dřevo), intenzita a rovnoměrnost osvětlení.

Pro kameru FLIR A655sc byla vybrána scéna, obsahující v pozadí zdroje notebooku a nabíječky na mobil. Snahou bylo tyto rušivé elementy vyfiltrovat takovým způsobem, aby byla co největší úspěšnost rozpoznání a zároveň tím nebyla příliš ovlivněna rychlost. Převzorkování snímků a objektu pro erozi bylo zapnuto. Úpravou po segmentaci bylo zanechání největšího objektu. Po kalibraci bylo doladění prahových hodnot nastaveno na 0,12. Senzitivita byla nastavena na 5000. Kamera měla nastavenou automatickou kalibraci na dynamický rozsah scény (nejvyšší teplota byla na snímku jako hodnota 1, nejnižší 0). Vzdálenost ruky od kamery byla 70 cm. Testované prostředí spolu s výslednou segmentací a úpravami lze pozorovat na obrázku (Obr. 6.5).



Obr. 6.5: Problematické prostředí pro IR oblast obsahující v pozadí napájecí zdroje. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po segmentaci prahováním a na třetím je vyobrazen snímek po úpravách

Pro každé gesto bylo vybráno opět 50 snímků od začátku jeho provedení, přičemž ruka se mírně pohybovala a rotovala po scéně. Výsledky se nacházejí v tabulce (Tab. 6.7).

Tab. 6.7: Naměřené časy (v sekundách) a úspěšnost algoritmu

Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem	Úspěšnost rozpoznání
1	0,0042	0,0025	0,0412	0,0078	0,0557	90 %
2	0,0039	0,0026	0,0398	0,0071	0,0535	92 %
3	0,0042	0,0024	0,0423	0,0075	0,0565	94 %
4	0,0039	0,0026	0,0389	0,0073	0,0528	90 %

Rychlost algoritmu se změnila od předchozí varianty v ideálním prostředí jen nepatrně (díky úpravám po segmentaci) a úspěšnost mírně poklesla, což bylo dáno při segmentaci přiřazením zdrojů napájení k ruce. Chyby ve špatném zařazení gesta nenastávaly.

Posledním prostředím, kde byla testována kamera FLIR A655sc, bylo umístění ruky před 3D tiskárnou (výrazněji teplejší, než samotná ruka). Vyobrazené prostředí je spolu s výsledky segmentace a úprav na obrázku (Obr. 6.6).



Obr. 6.6: Problematické prostředí pro IR oblast obsahující v pozadí 3D tiskárnu. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po segmentaci prahováním a třetím je vyobrazen snímek po úpravách

Nastavení bylo stejné jako v předchozím testování, pouze se lišilo v úpravách po segmentaci, které zahrnovaly mediánovou filtraci (okno velikosti 5), operaci otevření (velikost objektu 5) a vyplnění objektu. Vzdálenost ruky od kamery byla 80 cm. Jelikož by pro senzitivitu 5000 vycházela úspěšnost rozpoznání blízka nule, byla zvýšena na 6000. Výsledky jsou v tabulce (Tab. 6.8).

Tab. 6.8: Naměřené časy (v sekundách) a úspěšnost algoritmu

Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem	Úspěšnost rozpoznání
1	0,0038	0,0679	0,04070	0,0081	0,1205	74 %
2	0,0039	0,0721	0,03882	0,0079	0,1227	70 %
3	0,0041	0,0715	0,04125	0,0076	0,1245	72 %
4	0,0039	0,0693	0,03652	0,0078	0,1175	78 %

Rychlost běhu algoritmu byla výrazně snížena díky nezbytným úpravám po segmentaci. Všechna gesta jsou zařazena správně, úspěšnost rozpoznání zůstává stále na použitelné hodnotě, rychlost je ale značně limitována během algoritmu a to na rozpoznání ruky okolo 8,3 snímků za vteřinu.

Pro FLIR Lepton bylo vybráno první problematické prostředí zobrazené na obrázku (Obr. 6.7), které obsahovalo v pozadí rozsvícenou lampu. Bylo zapnuto převzorkování objektu pro erozi. Jako úprava po segmentaci bylo použito zanechání největšího objektu. Po kalibraci bylo doladění prahových hodnot nastaveno na 0,12. Ruka od objektivu se nacházela vzdálenosti 80 cm. Senzitivita byla nastavena na 5000.



Obr. 6.7: Problematické prostředí pro IR oblast obsahující v pozadí rozsvícenou lampu. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po segmentaci prahováním a třetím je vyobrazen snímek po úpravách

Pro každé gesto bylo vybráno opět 50 snímků od začátku jeho provedení, přičemž ruka se mírně pohybovala a rotovala po scéně. Výsledky se nacházejí v tabulce (Tab. 6.9).

Tab. 6.9: Naměřené časy (v sekundách) a úspěšnost algoritmu

Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem	Úspěšnost rozpoznání
1	0,0040	0,0025	0,0412	0,0071	0,0477	94 %
2	0,0042	0,0023	0,0349	0,0078	0,0492	98 %
3	0,0041	0,0022	0,0356	0,0077	0,0496	94 %
4	0,0045	0,0022	0,0308	0,0079	0,0545	92 %

Dosažená úspěšnost rozpoznání gest je opět vysoká, rychlost je nepatrně větší jak u FLIR A655sc, což je dáno opět menším rozlišením Leptonu. Všechna gesta byla zařazena správně.

Druhým prostředím pro FLIR Lepton, které lze sledovat na obrázku (Obr. 6.8), bylo pozadí s aktivním reproduktorem (studiovým monitorem) a tabletem, kde běžely výpočetní operace.



Obr. 6.8: Problematické prostředí pro IR oblast obsahující v pozadí aktivní reproduktor a tablet. Na prvním snímku se nachází zvolené prostředí, na druhém výsledek po segmentaci prahováním a třetím je vyobrazen snímek po úpravách

Nastavení pro hodnocení bylo naprosto totožné jako u kamery FLIR A655sc u druhého problematického prostředí a zahrnovalo veškeré úpravy po segmentaci a senzitivitu nastavenou na 6000. Výsledky jsou v tabulce (Tab. 6.10).

Tab. 6.10: Naměřené časy (v sekundách) a úspěšnost algoritmu

Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem	Úspěšnost rozpoznání
1	0,0034	0,0585	0,0687	0,0070	0,1376	68 %
2	0,0031	0,0584	0,0599	0,0069	0,1283	70 %
3	0,0034	0,0607	0,0698	0,0070	0,1412	56 %
4	0,0032	0,0585	0,0522	0,0073	0,1212	58 %

Úspěšnost rozpoznání je již výrazně snížena, ale díky zjištěné nulové záměně gesta za jiné, by mohla být s omezeními použita i tato varianta prostředí. Rychlost algoritmu je jen drobnou limitací v porovnání s možnou rychlostí snímání FLIR Leptonem 8,6 fps (v nejhorším případě vychází rychlost algoritmu na 7,13 fps)

Důležitým faktorem, který se v infračervené oblasti projevuje, je prokrvení ruky a prstů (a s tím související teplota). Nedostatečně prokrvenou ruku lze pozorovat na obrázku a ze snímků je patrné, že schopnost rozpoznání gest je za těchto podmínek zcela znemožněna.



Obr. 6.9: Nedostatečně prokrvená ruka a první tři varianty gesta

6.7 Problematická prostředí pro rozpoznání gest pro viditelnou oblast

Výběr tří prostředí pro viditelnou oblast zahrnoval prostředí, kde byla v pozadí červená sedačka, kousek černobílého obrázku a zelený list, dále pak prostředí kde se v pozadí vyskytovaly dřevěné poličky a prostěradlo hnědočervené barvy a v posledním prostředí bylo nerovnoměrné osvětlení ruky (přímé sluneční světlo).

První popsané prostředí spolu s výsledky segmentace a úprav lze pozorovat na obrázku (Obr. 6.10). Ruka byla umístěna 35 cm od kamery.



Obr. 6.10: Problematické prostředí pro VIS oblast obsahuje v pozadí červenou sedačku, list rostliny a obrázek. První snímek zobrazuje zvolené prostředí, snímek napravo pak segmentaci v normalizovaném prostředí RGB a složek R,G,B a poslední snímek pak výsledek po úpravách

Nastavení pro tuto scénu zahrnovalo převzorkování snímků a objektu pro erozi, segmentaci v normalizovaném prostoru RGB u všech tří složek, s úpravou prahů o 0,022. Úpravy po segmentaci se týkaly ponechání největšího objektu a jeho vyplnění. Dosažené výsledky se nacházejí v tabulce (Tab. 6.11).

Tab. 6.11: Naměřené časy (v sekundách) a úspěšnost algoritmu

Gesto	Segmentace	Úpravy před a po segmentaci	Zpracování	Rozpoznání	Celkem	Úspěšnost Rozpoznání
1	0,0164	0,0086	0,0392	0,0075	0,0716	98 %
2	0,0171	0,0121	0,0445	0,0078	0,0816	100 %
3	0,0176	0,0118	0,0359	0,0075	0,0727	100%
4	0,0158	0,0083	0,0290	0,0073	0,0603	100%

Úspěšnost dosahovala téměř sto procent, na rychlosti algoritmu mimo zpracování snímku se projevila rychlost segmentace a úpravy před a po segmentaci. Vzhledem k získané snímkovací frekvenci 15 fps z kamery není dosažená rychlost algoritmu limitující.

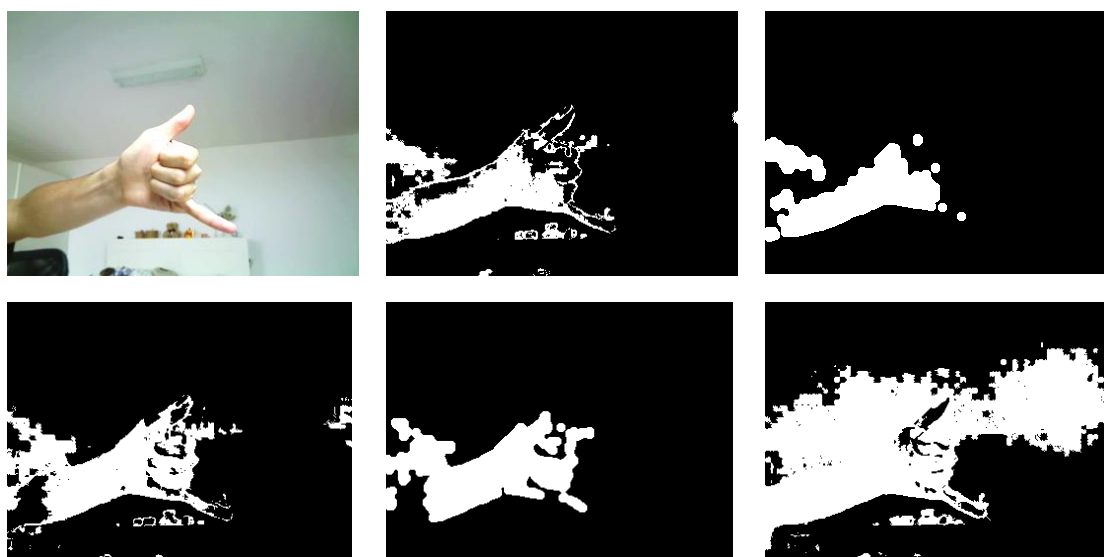
Druhé použité prostředí je vyobrazené na obrázku (Obr. 6.11) spolu se segmentacemi v barevných prostorech RGB, $YCbCr$ a následnými úpravami. Úpravy zahrnovaly mediánovou filtraci (s velikostí okna 5) před segmentací, ponechání největšího objektu, mediánovou filtraci (okno 5), operaci otevření (velikost objektu 5).



Obr. 6.11: Druhé problematické prostředí: První snímek v horním řádku zobrazuje zvolené prostředí, snímky napravo pak segmentaci v normalizovaném RGB všech tří složek a následné úpravy. První snímek ve spodním řádku zobrazuje segmentaci v $YCbCr$ prostoru a to dle složek C_b a C_r , další pak snímek po úpravách a poslední snímek pak segmentaci získanou kombinací složek R a C_r (zde jsou nastavené prahy odlišné jako v přechozí variantě barevných prostorů zvlášť)

I při důsledném nastavování prahových hodnot a úprav se nepodařilo ruku vysegmentovat tak, aby s ní bylo možné nadále pracovat. Úspěšnost rozpoznání takto zpracované ruky se blížilo nule a celkový čas algoritmu byl roven 0,2194 sekund.

Na obrázku (Obr. 6.12) je poslední testované prostředí spolu se segmentacemi v barevných prostorech RGB, YC_bC_r a následnými úpravami. Nastavení algoritmu bylo stejné jako v předchozím případě, lišilo se pouze v přidání operace vyplnění objektu v rámci úprav po segmentaci. Nerovnoměrnost osvětlení ruky bylo natolik velké, že se i při důsledném nastavování v obou barevných prostorech nepodařilo ruku uspokojivě vysegmentovat. Výsledek úspěšnosti rozpoznání gest se opět blížil nule s podobným časem jako v předchozím případě.



Obr. 6.12: Třetí problematické prostředí: První snímek zobrazuje zvolené prostředí, snímky napravo pak segmentaci v normalizovaném RGB všech tří složek a následné úpravy. První snímek ve spodním řádku zobrazuje segmentaci v YC_bC_r prostoru a to složek C_b a C_r , další pak snímek po úpravách a poslední snímek pak segmentaci získanou kombinací složek R a C_r (zde jsou nastavené prahy odlišné jako v přechodí variantě barevných prostorů zvlášť)

Z výsledků je patrné, že rozpoznání gest ve viditelné oblasti je silně závislé na použitém prostředí a pro vysokou úspěšnost rozpoznání je nutné prostředí pečlivě vybírat.

6.8 Srovnání výsledků hodnocení v infračervené a viditelné spektrální oblasti

V případě, že je pro daná zařízení a realizovaný algoritmus zvoleno ideální prostředí, dochází k úspěšnému rozpoznání na přijatelném rozsahu vzdáleností ruky od objektivu. Použitelnost minimálních a maximálních vzdáleností záleží na využití pro konkrétní aplikaci. U infračervených kamer je pro zvolené objektivy nutné mít ruku více vzdálenou od kamery, což může být z hlediska správného umístění dlaně do scény mírně problematické a uživatel je většinou nucen umístění kontrolovat. Nedostatek je možný řešit výměnou objektivů (nejsou na UBMI VUT dostupné).

Pokud je v ideálním prostředí a vhodně zvolené vzdálenosti ruky od kamery testována rychlost, je v původním rozlišení u kamer FLIR A655sc a Logitech S7500 zpracování pomalé. Rychlost pro Logitech S7500 se dostává na nepoužitelnou hranici pod frekvenci snímků 2 fps, u kamery FLIR A655sc na 6 fps, což by s určitou limitací již použitelné být mohlo. Z důvodu malého rozlišení u FLIR Leptonu je zpracování dostatečně rychlé na to, aby hlavním omezením byla hardwarová snímkovací frekvence modulu (8,6 fps).

Při sníženém rozlišení na maximální hodnotu 350 pixelů, je omezení rychlosti rozpoznání algoritmem výrazně menší. U kamery FLIR A655sc se dostáváme na rychlost okolo 16 fps a u kamery Logitech S7500 na 11 fps. Omezení algoritmem je patrné, ale rychlost překonává snímkovací frekvenci modulu Lepton 8,6 fps.

Převzorkování objektu pro erozi způsobí další nárůst teoretické rychlosti a to u FLIR A655sc na hodnotu okolo 25 fps a u kamery Logitech S7500 na 16 fps. Zde jsou hodnoty velice blízké hardwarovým limitům a omezení algoritmem není tedy téměř žádné. Při hodnocení rychlosti je nutné také přihlédnout k faktu, že stáří notebooku (a procesoru) na kterém probíhalo testování je více jak 7 let (rok výroby 2008) a s novými procesory (Intel Core i5 a i7) je možné dle zběžných testů teoreticky dosáhnout dvakrát až čtyřikrát větší rychlosti.

Z hlediska použitelnosti je rychlost rozpoznání gest u FLIR Leptonu dostačující (gesto je při 8,6 fps rozpoznáno za 0,116 s). Pokud by byl kladen požadavek na rozpoznání gesta při zachování pocitu naprosté plynulosti (25 fps), je omezení zřejmé a bylo by nutné uvažovat nad jinou volbou kamerového zařízení s odpovídajícími parametry výpočetní jednotky.

Problematické prostředí pro rozpoznání gest v infračervené a viditelné oblasti sebou přináší snížení rychlosti i úspěšnosti rozpoznání. Výsledky jsou vždy závislé na správné

segmentaci ruky. V infračervené oblasti záleží především na velikosti a umístění objektů, které mají podobnou teplotu jako je teplota ruky. Objekty mohou být samotným zdrojem tepla (žárovka, zdroje napětí, procesory), nebo mohou být zdroji zahřátý na určitou teplotu (zahřátý černý předmět od slunce). Pokud se objekt nachází například mezi prsty ruky, není snadné ho vyfiltrovat a úspěšnost rozpoznání se sníží. Filtrací se také snižuje rychlost algoritmu, což v případě FLIR A655sc přinášelo při maximálních úpravách snížení rychlosti na 8 fps. Výhodou infračervené oblasti je skutečnost, že objekty snižující úspěšnost segmentace se mnohem častěji se scéně nevyskytují a je nutné je v běžných místnostech (domácnost, kanceláře) spíše vyhledávat (pokud se nejedná o jiné osoby).

Daleko výraznějším problémem při snímání v infračervené oblasti je závislost segmentace na aktuálním prokrvení (teplotě) ruky a prstů. V případě, že se uživatel nachází v chladném prostředí, umyl si ruce studenou vodou, nebo má potíže s prokrvením končetin, je rozpoznání gesta velmi ztíženo.

Správná volba prostředí je ve viditelné oblasti daleko důležitější. Je třeba dát důsledný pozor na rovnoměrnost osvětlení a volbu pozadí, jelikož jakýkoliv objekt, který má podobnou barvu jako ruka, způsobí špatnou segmentaci. Tedy místnosti, kde se v pozadí hojně vyskytuje například dřevo, věci jsou zbarvené do hněda, ruka je osvětlena sluncem, nebo je naopak světla nedostatek, jsou pro použití naprosto nevhodné a úspěšnost rozpoznání gest bude blízka nule. Navýšení výpočetní náročnosti v nevhodném prostředí v souvislosti s úpravami po segmentaci je podobné jako u infračervených kamer.

7 ZÁVĚR

Cílem práce bylo seznámit se s různými aplikacemi využití optických metod k rozeznání gest, navrhnout metodu detekce a hodnocení gest ve VIS a IR oblasti s využitím kamery, IR obrazového senzoru a sestavení testovací skupiny gest určených pro klasifikaci. Dále navrženou metodu realizovat, ověřit na testovacích gestech a v závěru také porovnat rozdíly mezi hodnocením gest ve viditelné a infračervené oblasti z hlediska hardwarové a softwarové náročnosti a účinnosti metody.

Pro navržení metody detekce a rozeznání gest ve VIS a IR oblasti bylo třeba nejdříve zjistit, jak lze obecně jednotlivá gesta členit (například pro dynamická a statická gesta se používají rozdílné algoritmy). Dále bylo třeba zabývat se možnostmi doposud používaných technologií a zaměřit se především na optická zařízení a jejich parametry (rozlišení čipu, snímkovací frekvenci, datový přenos, charakteristiky objektivů).

Nejčastěji jsou ve studiích pro rozpoznání gest používány kamery, které snímají ve viditelné oblasti záření. Jedná se o cenově dostupnou technologii, která ale sebou přináší řadu nedostatků. Velkým problémem pro následnou detekci ruky bývá její nerovnoměrné osvětlení, variabilita barvy kůže u jednotlivců a především etnik, barva pozadí, nízké osvětlení scény (vznik šumu, pohybová neostrost, problémy se zaostřením, problematická expozice) a výskyt jiných částí těla. Většinu těchto problémů lze vyřešit použitím infračervených kamer. Hlavní důvod, proč nejsou v rozeznání gest zařízení rozšířená, je především jejich cenová nedostupnost.

Pro systém sloužící k rozeznání gest byla vybrána infračervená kamera FLIR A655sc, která poskytuje výbornou kvalitu obrazu, cenově dostupný modul FLIR Lepton připojený k deskovému počítači Raspberry Pi 2 a kamera Logitech S7500 snímající ve viditelné oblasti spektra.

Navržený a realizovaný způsob řešení zahrnoval segmentaci gest pomocí prahování šedotónového obrázku v případě snímků získaných z infračervené oblasti a prahování složek v barevných prostorech RGB (normalizovaný) a $YCbCr$ pro viditelnou oblast. Za účelem získání vhodných prahů byla navržena kalibrace, zahrnující metodu narůstání oblastí pro detekci hledané ruky. Z této vymezené oblasti byly získány prahy pomocí normalizovaných histogramů. Dále algoritmus obsahoval volbu úprav po segmentaci a zpracování snímků pro Hammingovu síť. Součástí zpracování snímků bylo provedení rotace a správného ořezu předloktí a to pomocí nalezení středu dlaně, nebo poměru předloktí a délky ruky. Takto zpracovaný snímek byl v Hammingově síti porovnáván s vytvořenou databází gest.

Výhodou navržené metody je především univerzálnost použití na jakákoliv gesta bez omezení natočení ruky ve smyslu pronace nebo supinace. Další nemalou výhodou je velice jednoduché přidání nového gesta, či upravení stávajícího. K hlavním nevýhodám pak patří nutnost pro každé gesto pečlivě vytvořit více variant a narůstání výpočetní náročnosti se zvětšující se databází.

Použitelný rozsah vzdáleností ruky od kamery vycházel v ideálních podmínkách pro obě spektrální oblasti poměrně dobře. Při hodnocení rychlosti zpracování gest bylo zjištěno, že pro dosažení dobré snímkovací frekvence je u kamer A655sc a Logitech S7500 nutné snímky převzorkovat. Navýšení rychlosti bylo možné dosáhnout dalším převzorkováním objektu sloužícího k erozi. Rychlost algoritmu po této operaci již nebyla omezením pro hardwarovou snímkovací frekvenci. Při testování bylo nutné brát také v potaz, na jaké výpočetní jednotce algoritmus běží, jelikož u nejnovějších procesorů mohou být výsledky výrazně lepší. Snímkovací frekvence u FLIR Leptonu je ze všech zařízení nejmenší, ale rychlost rozpoznání je stále na dobré úrovni.

Testování v různých podmínkách a prostředích odhalilo nedostatky rozpoznání gest v infračervené i viditelné spektrální oblasti. V infračervené oblasti může být největším problémem samotná ruka, která pokud není dobře prokrvená, může způsobit při rozpoznání gesta selhání. Menším nedostatkem je pak pozadí, na kterém se nacházejí objekty stejné teploty jako má ruka. Tyto objekty je možné v řadě případů s ohledem na narůstající výpočetní náročnost a klesající úspěšnost rozpoznání filtrovat. U viditelné oblasti bylo zjištěno, že je nutné dbát daleko více na pečlivější výběr prostředí a podmínek osvětlení. Jakékoliv běžně se vyskytující objekty podobné barvy jako má ruka způsobí chybnou segmentaci, bez možnosti dále gesto správně rozpoznat. I přes použité barevné prostory, které mají minimalizovat vliv osvětlení, nejsou výsledky segmentace při nerovnoměrném osvětlení nikterak uspokojivé. Přímé osvětlení ruky slunce a vznikající stíny způsobí selhání celého rozpoznávání.

S ohledem na popsané výhody a nevýhody obou spektrálních oblastí a dostupných zařízení je možné konstatovat, že nejvhodnější (v rámci úspěšnosti a rychlosti rozpoznání) je použití kamery FLIR A655sc. Pokud je ale přihlédnuto k ceně, je použití tohoto řešení nereálné. Pokud bychom chtěli zachovat výhody snímání v infračervené oblasti (především s ohledem na prostředí) při zachování cenové dostupnosti, je vhodnou alternativou FLIR Lepton, který spolu se systémem Raspberry Pi2 stojí v přepočtu kolem 5400 korun.

LITERATURA

- [1] A.IBRAHEEM, N., R. Z. KHAN a M. M. HASAN. Comparative Study of Skin Color based Segmentation Techniques. *International Journal of Applied Information Systems*. 2013, 5(10): 24-39. DOI: 10.5120/ijais13-450985.
- [2] ALPERN, M. a K. MINARDO. Developing a car gesture interface for use as a secondary task. *CHI '03 extended abstracts on Human factors in computing systems - CHI '03*. New York, New York, USA: ACM Press, 2003, : 932-. DOI: 10.1145/765891.766078.
- [3] BADI, H. a S. HUSSEIN. Hand posture and gesture recognition technology. *Neural Computing and Applications*. London: Springer London, 0140n. 1., 25(3): 871-878. DOI: 10.1007/s00521-014-1574-4.
- [4] BERMAN, S. a H. STERN. Sensors for Gesture Recognition Systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 2012, 42(3): 277-290. DOI: 10.1109/TSMCC.2011.2161077.
- [5] Binh, N. D. A new approach dedicated to real-time hand gesture recognition. *International conference on image processing, computer vision, and pattern recognition*, Las Vegas, Nevada, USA. 2006, 2:481–488
- [6] BRASHEAR, H., V. HENDERSON, K.-H. PARK, H. HAMILTON, S. LEE a T. STARNER. American sign language recognition in game development for deaf children. *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility - Assets '06*. New York, New York, USA: ACM Press, 2006, : 79-. DOI: 10.1145/1168987.1169002
- [7] Compact InfraRed Camera: Easy to Use Troubleshooting Tool. *FLIR Systems, Inc.* [online]. [cit. 2015-12-28]. Dostupné z: www.flir.com/WorkArea/DownloadAsset.aspx?id=27352
- [8] CORRADINI, Andrea. Real-Time Gesture Recognition by Means of Hybrid Recognizers. *International Gesture Workshop*. London: GW, 2001, s. 34. ISBN 978-3-540-47873-7.
- [9] ČERNOCKÝ, J. *Zpracování řečových signálů – studijní opora*. Brno: Vysoké učení technické v Brně, Fakulta informačních technologií, Speech FIT, 2006. Dostupné z: <http://www.fit.vutbr.cz/study/courses/ZRE/public/opora/zre_opora.pdf>.
- [10] DOHERTY-SNEDDON, G. *Neverbální komunikace dětí*. 1. vyd. Praha: Portál, 2005. 195 s. ISBN 80-73670-43-7.
- [11] FLIR A655sc: High resolution science grade LWIR camera. *FLIR Systems, Inc.* [online]. [cit. 2016-01-02]. Dostupné z: http://www.flirmedia.com/MMC/THG/Brochures/RND_011/RND_011_US.pdf
- [12] FLIR Lepton Hookup Guide. *SparkFun* [online]. [cit. 2016-01-02]. Dostupné z: <https://learn.sparkfun.com/tutorials/flir-lepton-hookup-guide>

- [13] FLIR One Launch Faq. *FLIR Systems, Inc.* [online]. [cit. 2016-05-02]. Dostupné z: <http://www.flir.com/flirone/press/FLIR-ONE-Android-iOS-FAQ.pdf>
- [14] FREEMAN, W.T., D.B. ANDERSON, P. BEARDSLEY, C.N. DODGE, M. ROTH, C.D. WEISSMAN, W.S. YERAZUNIS, H. KAGE, I. KYUMA, et al. Computer vision for interactive computer graphics. *Computer Graphics and Applications, IEEE*. USA: IEEE, 1998, 18(3): 42-53. DOI: 10.1109/38.674971.
- [15] HASAN, H. S. a S. A. KAREEM. Human Computer Interaction for Vision Based Hand Gesture Recognition: A Survey. *2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT)*. IEEE, 2012, : 55-60. DOI: 10.1109/ACSAT.2012.37.
- [16] HASAN, M. M. a P. K. MISHRA. Novel Algorithm for Multi Hand Detection and Geometric Features Extraction and Recognition. *International Journal of Scientific & Engineering Research*. 2012, 3(5), ISSN 2229 - 5518.
- [17] HASANUZZAMAN, M., V. AMPORNARAMVETH, M.A. TAO ZHANG, Y. BHUIYAN, H. SHIRAI a H. UENO. Real-time Vision-based Gesture Recognition for Human Robot Interaction. *Robotics and Biomimetics, 2004. ROBIO 2004. IEEE International Conference on*. IEEE, 2004, : 413-418. DOI: 10.1109/ROBIO.2004.1521814.
- [18] HONG, CH.A., M. KOSCHAN, S.G. ABIDI, S.G. KONG a H.W. CHANG. Multispectral visible and infrared imaging for face recognition. *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*. IEEE, 2008, : 1-6. DOI: 10.1109/CVPRW.2008.4563054.
- [19] HONGBIN, Z., X. GANG, Ch. JIUJUN, G. FEI a Y. XIAOFANG. Real-time fish detection based on improved adaptive background. *7th WSEAS Int. Conf. on applied computer & applied computational science*. 2008. ISSN 1790-5117.
- [20] HWANG, T.-H., W.-H. CUI, I.-S. YANG a O.-K. KWON. A highly area-efficient controller for capacitive touch screen panel systems. *IEEE Transactions on Consumer Electronics*. 2010, 56(2): 1115-1122. DOI: 10.1109/TCE.2010.5506047.
- [21] IBRAHEEM, A.N., Z.R. KHAN, M.M. HASAN, A.N. IBRAHEEM. Comparative Study of Skin Color based Segmentation Techniques. *International Journal of Applied Information Systems*. 2013, 5(10): 24-39. DOI: 10.5120/ijais13-450985.
- [22] IBRAHEEM, N. A., KHAN, R. Z. Vision based gesture recognition using neural networks approaches: A review. *International Journal of human Computer Interaction (IJHCI)*, 2012, 3.1: 1.
- [23] JIEJIE, Z., W. LIANG, Y. RUIGANG, J.E. DAVIS a P. ZHIGENG. Reliability Fusion of Time-of-Flight Depth and Stereo Geometry for High Quality Depth Maps. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. USA: IEEE, 2011, 33(7): 1400-1414. DOI: 10.1109/TPAMI.2010.172.

- [24] KAÂNICHE, M.-B. Gesture recognition from video sequences. *Signal and Image processing*. Universite Nice Sophia Antipolis. 2009. PhD Thesis.
- [25] KAKUMANU, P., S. MAKROGIANNIS a N. BOURBAKIS. A survey of skin-color modeling and detection methods. *Pattern Recognition* [online]. 2007, 40(3): 1106-1122 [cit. 2015-12-29]. DOI: 10.1016/j.patcog.2006.06.010.
- [26] KAWARAZAKI, N., I. HOYA, K. NISHIHARA a T. YOSHIDOME. 7 Cooperative Welfare Robot System Using Hand Gesture Instructions. *Advances in Rehabilitation Robotics*. Springer Berlin Heidelberg, 2004, : 143. DOI: 10.1007/10946978_7.
- [27] KIM, D., J. SONG a D. KIM. Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMs. *Pattern Recognition*. 2007, 40(11): 3012-3026. DOI: 10.1016/j.patcog.2007.02.010. ISSN 00313203.
- [28] KOIKE, H., Y. SATO a Y. KOBAYASHI. Integrating paper and digital information on EnhancedDesk: a method for realtime finger tracking on an augmented desk system. *ACM Transactions on Computer-Human Interaction*. 2001. 8(4): 307-322. DOI: 10.1145/504704.504706
- [29] KONG, S. G., J. HEO, B. R. ABIDI, J. PAIK a M. A. ABIDI. Recent advances in visual and infrared face recognition—a review. *Computer Vision and Image Understanding*. 2005, 97(1): 103-135. DOI: 10.1016/j.cviu.2004.04.001.
- [30] KOVÁŘOVÁ, D. *Gestika a její podoby ve výuce cizího jazyka*. Acta FF ZČU. Plzeň: ZČU 2010, s.41-59.
- [31] KOZUMPLÍK, J., I. PROVAZNÍK *Umělá inteligence v medicíně*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2007.
- [32] LAMAR, M.V. Hand gesture recognition using T-comb net A neural network model dedicated to temporal information processing. Nagoya Institute of Technology, Japan. 2001. Ph.D. thesis.
- [33] LEPTON Longwave Infrared (LWIR) Camera Module: Product Brief. *MDS Technology* [online]. [cit. 2016-01-02]. Dostupné z: <http://www.mds-flir.com/datasheet/FLIR-Lepton-datasheet.pdf>
- [34] Logitech QuickCam S7500: Details. *Ferrika* [online]. [cit. 2016-01-02]. Dostupné z: <http://ferrika.com/id/logitech-quickcam-s7500.html>
- [35] MALIMA, A., E. OZGUR a M. CETIN. A Fast Algorithm for Vision-Based Hand Gesture Recognition for Robot Control. *2006 IEEE 14th Signal Processing and Communications Applications*. IEEE, 2006, : 1-4. DOI: 10.1109/SIU.2006.1659822.

- [36] MANRESA, C., J. VARONA, R. MAS a J.P. FRANCISCO. *Real –Time Hand Tracking and Gesture Recognition for Human-Computer Interact.* *Electronic Letters on Computer Vision and Image Analysis* 0(0):1-7. 2000.
- [37] MCNEILL, D. *Hand and mind: what gestures reveal about thought.* Paperback ed. Chicago: University of Chicago Press, 1995. ISBN 978-022-6561-349.
- [38] MITRA, S. a T. ACHARYA. *Gesture Recognition: A Survey.* *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews).* 2007, 37(3): 311-324. DOI: 10.1109/TSMCC.2007.893280.
- [39] NISHIKAWA, A., T. HOSOI, K. KOARA, D. NEGORO, A. HIKITA, S. ASANO, H. KAKUTANI, F. MIYAZAKI, M. SEKIMOTO, et al. *FAce MOUSE: A novel human-machine interface for controlling the position of a laparoscope.* *Robotics and Automation, IEEE Transactions on.* USA: IEEE. 2003, 0031n. 1., 19(5): 825-841. DOI: 10.1109/TRA.2003.817093.
- [40] *Photonic frontiers: Gesture recognition: Lasers bring gesture recognition to the home.* JEFF, H. *PennWell Corporation* [online]. [cit. 2015-12-02]. Dostupné z: <http://www.laserfocusworld.com/articles/2011/01/lasers-bring-gesture-recognition-to-the-home.html>
- [41] RAFIQUZ ZAMAN KHAN a NOOR ADNAN IBRAHEEM. *Hand Gesture Recognition: A Literature Review.* *International Journal of Artificial Intelligence & Applications* [online]. 2012, 3(4): 161-174 [cit. 2015-12-31]. ISSN 0976-2191.
- [42] ROGALLA, O., M. EHRENMANN, R. ZOLLNER, R. BECHER a R. DILLMANN. *Using gesture and speech control for commanding a robot assistant.* *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on.* USA: IEEE, 2002, : 454-459. DOI: 10.1109/ROMAN.2002.1045664.
- [43] RONZHINA, M. *Hammingova síť: materiály ke cvičení předmětu Umělá inteligence v medicíně.* Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2016.
- [44] SOCOLINSKY, D. A., A. SELINGER a J. D. NEUHEISEL. *Face recognition with visible and thermal infrared imagery.* *Computer Vision and Image Understanding.* 2003, 91(1-2): 72-114. DOI: 10.1016/S1077-3142(03)00075-4.
- [45] SOCOLINSKY, D.A. a A. SELINGER. *A comparative analysis of face recognition performance with visible and thermal infrared imagery.* *Pattern Recognition, 2002. Proceedings. 16th International Conference on.* USA: IEEE, 2002, 4: 217-222. DOI: 10.1109/ICPR.2002.1047436.
- [46] Starner, T., Leibe, B., Singletary, B., and Pair, J. *Mindwarping: Towards creating a compelling collaborative augmented reality game.* *In Proceedings of the Fifth International Conference on Intelligent User Interfaces.* ACM Press, New York, 2000, 256–259.

- [47] STERGIOPOULOU, E., N. PAPAMARKOS. Hand gesture recognition using a neural network shape fitting technique. *Engineering Applications of Artificial Intelligence* [online]. 2009, 22(8): 1141-1158 [cit. 2015-12-29]. DOI: 10.1016/j.engappai.2009.03.008. ISSN 0952-1976.
- [48] TANIBATA, N., SHIMADA, N., SHIRAI, Y. Extraction of hand features for recognition of sign language words. *International conference on vision interface*. 2002. p. 391-398.
- [49] TERRILLON, J.-C., M. DAVID a S. AKAMATSU. Detection of human faces in complex scene images by use of a skin color model and of invariant Fourier-Mellin moments. *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*. IEEE Comput. Soc, 1998, : 1350-1355. DOI: 10.1109/ICPR.1998.711952.
- [50] THEODORIDIS, S. a K. KOUTROUMBAS. *Pattern Recognition* [online]. 4th ed. Burlington: Elsevier, 2008 [cit. 2015-12]. ISBN 978-008-0949-123.
- [51] VERRI, A. a V. TORRE. Absolute depth estimate in stereopsis. *Journal of the Optical Society of America A*. 1986, 3(3): 297. DOI: 10.1364/JOSAA.3.000297.
- [52] VEZHNEVETS, V., SAZONOV, V., ANDREEVA, A., A survey on pixel-based skin color detection techniques. In: *Proc. Graphicon*. 2003. p. 85-92.
- [53] WACHS, J. P., H. I. STERN, Y. EDAN, M. GILLAM, J. HANDLER, C. FEIED a M. SMITH. A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association: JAMIA*. 2008, 15(3): 321. DOI: 10.1197/jamia.M241. ISSN 10675027.
- [54] WACHS, J., M. KÖLSCH, H. STERN a Y. EDAN. Vision-Based Hand-Gesture Applications. *Association for Computing Machinery. Communications of the ACM*. New York, 2011, 54(2): 60. ISSN 00010782.
- [55] WACHS, J., U. KARTOUN, H. STERN a Y. EDAN. Real-time hand gesture telerobotic system using fuzzy c-means clustering. *Automation Congress, 2002 Proceedings of the 5th Biannual World*. USA: IEEE, 2002, 13: 403-409. DOI: 10.1109/WAC.2002.1049576
- [56] WALEK, P., M. LAMOŠ a J. JAN. *Analýza biomedicínských obrazů: Počítačová cvičení*. První. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2013. ISBN 978-80-214-4792-9. Dostupné také z: <http://www.dbme.feec.vutbr.cz/sites/default/files/news/fabo.pdf>
- [57] WILDER, J., P.J. PHILLIPS, J. CUNHONG a S. WIENER. Comparison of visible and infrared imagery for face recognition. *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*. IEEE Comput. Soc. Press, 1996, : 182-187. DOI: 10.1109/AFGR.1996.557262.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

MRI – Magnetic Resonance Imaging

TOF – Time of Flight

VIS – Visible Spectrum

IR – Infrared Spectrum

NIR – Near Infrared Spectrum

CCD – Charge Coupled Device

CMOS – Complementary Metal Oxide Semiconductor

FPS – Frames Per Second

SNR – Signal to Noise Ratio

LUT – Lookup Table

SGONG – Self Growing and Self Organized Neural Gas

HMM – Hidden Markov Models

EM – Expectation Maximization

FPA – Focal Plane Array

GPIO – General Purpose Input/Output

R – Red

G – Green

B – Blue

H – Hue

S – Saturation

I – Intensity

Y – Luminance

C_b – Blue Difference Chroma

C_r – Red Difference Chroma

BMP – Windows Bitmap

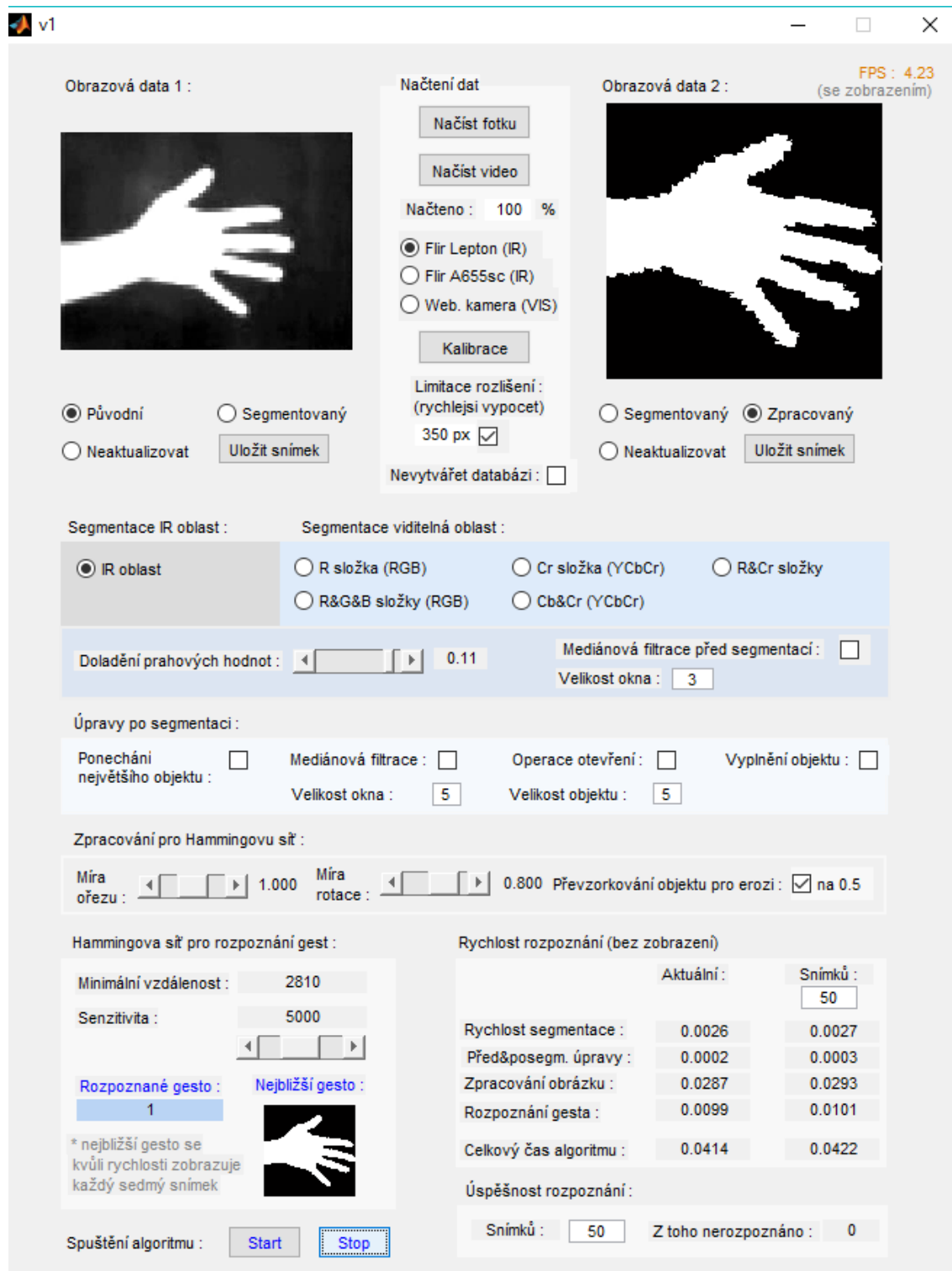
SEZNAM PŘÍLOH

A	Vytvořená aplikace	76
A.1	Popis vytvořené aplikace	76
A.2	Ovládání aplikace	78
B	Obsah CD	79

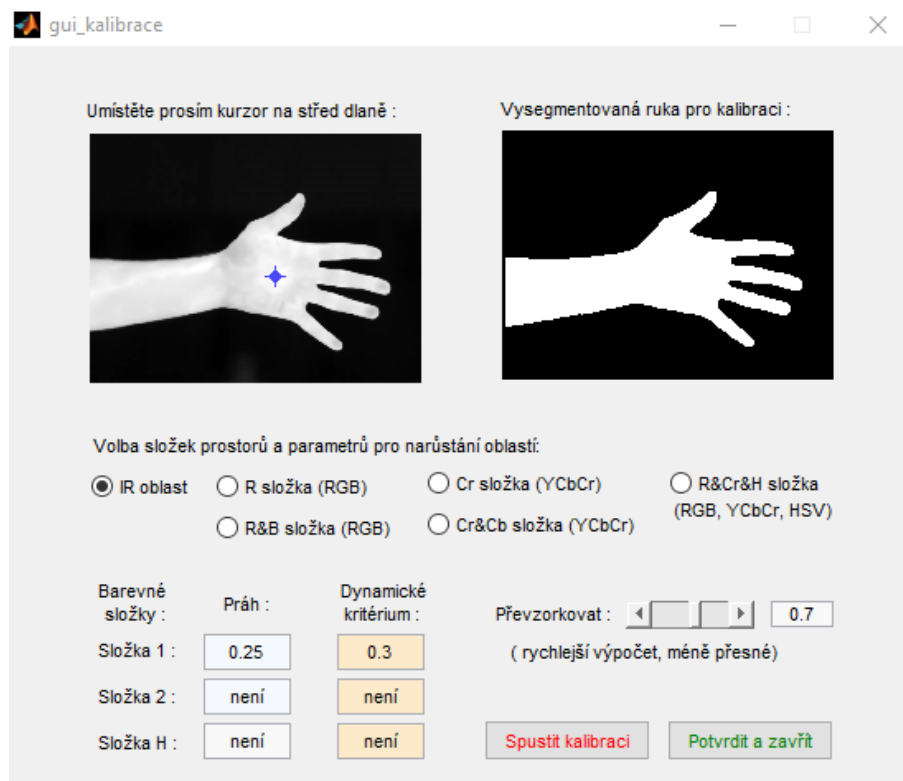
A VYTVOŘENÁ APLIKACE

A.1 Popis vytvořené aplikace

Pro analýzu gest byla v programovém prostředí MATLAB vytvořena aplikace, jejíž podobu lze sledovat na Obrázek 1. Součástí aplikace je i možnost provádět kalibraci (viz Obrázek 2) a to dle pomoci segmentace metodou narůstání oblastí v různých barevných prostorech. Aplikace umožňuje načítat buďto jednotlivé snímky nebo videa. V horní levé části lze zobrazit původní nezpracované snímky, segmentované snímky nebo aktuální obrazová data nezobrazovat (což zrychlí běh celé aplikace bez vlivu na výpočetní rychlost). V pravé části lze pak zobrazit segmentovaný nebo zpracovaný snímek nebo aktuální data opět nezobrazovat. Pro urychlení výpočtů je možné zapnout převzorkování a v případě že není třeba používat obrazovou databázi a tu upravovat, je možné zatrhnout položku *nevytvářet databázi*. Zvolením zařízení dojde k automatickému vyplnění hodnot nastavení. Pro segmentaci lze zvolit různé barevné prostory v případě viditelné oblasti a prahování šedotónového obrázku v případě infračervené oblasti. Získané prahové hodnoty lze doladit jednoduše pomocí posuvníku a obrázek lze před samotnou segmentací také filtrovat mediánovým filtrem s volitelnou velikostí okna. V rámci úprav po segmentaci lze provádět operace ponechání pouze největšího objektu, mediánovou filtraci s volitelnou velikostí okna, operaci otevření s volitelnou velikostí objektu a vyplnění objektu. V rámci zpracování snímku lze upravovat míru ořezu a rotaci (vhodné pro získání databáze gest) a dále je možné provést převzorkování objektu pro erozi na poloviční velikost (pro zrychlení výpočtů). Spodní část okna je věnována zobrazení výsledků. V levé části se nachází výsledky Hammingovy sítě – minimální Hammingova vzdálenost spolu s možností nastavení senzitivity pro rozpoznání, dále pak číslo rozpoznávaného gesta a vyobrazené nejbližší gesto. Pravá spodní část zahrnuje dosažené časy jednotlivých částí algoritmu, přičemž lze zvolit počet snímků, na kterých je počítán průměr těchto časů. Zvolit počet snímků lze také při zjišťování úspěšnosti rozpoznání.



Obrázek 1: Aplikační rozhraní



Obrázek 2: Kalibrační rozhraní

A.2 Ovládání aplikace

Ovládání aplikace je jednoduché, pro základní nastavení a správný běh je nutné provést pouze několik kroků.

- 1) V prvním kroku je nutné načíst fotku nebo video. Načtení je dokončeno při dosažení procentuální hodnoty načítání 100%
- 2) Dále je nutné zvolit správné zařízení, nebo odpovídající oblast, která se zařízením souvisí.
- 3) V případě nedostupné obrazové databáze je nutné zaškrtnout položku nevytvářet databázi.
- 4) V dalším kroku je vhodné provést kalibraci. Po otevření kalibračního okna je třeba umístit kurzor do středu dlaně a zvolit prostor podle toho, kde byl snímek získán. Kalibrace se zavře tlačítkem *potvrdit a zavřít*.
- 5) Posledním krokem je spuštění rozpoznání gest pomocí tlačítka *start*.

B OBSAH CD

Diplomová práce: Jan_Netopil_DP.pdf

Program: hlavní spouštěcí funkce: *v1.m*

další funkce: *gui_kalibrace, kalibrace, segincreasearea, segmentace, uprava_segmentace, analyza_orez, hamming_sit*

Databáze: Databáze obsahující 10 gest pro FLIR A655sc

Jednotlivé databáze obsahující 4 zvolená gesta pro zařízení FLIR A655sc, FLIR Lepton, Logitech S7500

Data: Videa a snímky pro různá prostředí