

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2018

Bc. Veronika Hříbková



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

**MODELOVÁNÍ POLOHY HLAVY POMOCÍ
STEREOSKOPICKÉ REKONSTRUKCE**

HEAD POSE ESTIMATION VIA STEREOSCOPIC RECONSTRUCTION

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Veronika Hříbková

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Ing. Radim Kolář, Ph.D.

BRNO 2018

Diplomová práce

magisterský navazující studijní obor **Biomedicínské inženýrství a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Bc. Veronika Hříbková

ID: 165012

Ročník: 2

Akademický rok: 2017/18

NÁZEV TÉMATU:

Modelování polohy hlavy pomocí stereoskopické rekonstrukce

POKYNY PRO VYPRACOVÁNÍ:

1) Prostudujte principy stereoskopického snímání obrazů se zaměřením na aplikace trojrozměrné rekonstrukce polohy hlavy. 2) Navrhněte a realizujte systém dvou kamer vhodný pro tato stereoskopická měření a navrhněte způsob akvizice vhodných sekvencí pro další zpracování. 3) Prostudujte a popište metody pro rekonstrukci disparitní mapy. 4) Zvolenou metodu rekonstrukce implementujte a otestujte na získaných datech. 5) Navrhněte a implementujte jednoduchý geometrický model založený na detekci zvolených oblastí (např. očí, nosu, úst) pro odhad polohy hlavy v prostoru. 6) Tento model následně otestujte a dosažené výsledky diskutujte.

DOPORUČENÁ LITERATURA:

[1] NIXON, M. a A. AGUADO. Feature Extraction and Image Processing, Elsevier, 2005.

[2] BREITENSTEIN, M. D. et al. Head Pose Estimation from Passive Stereo Images, Scandinavian Conference on Image Analysis, Lecture Notes in Computer Science, pp. 219-228, 2009.

Termín zadání: 5.2.2018

Termín odevzdání: 18.5.2018

Vedoucí práce: doc. Ing. Radim Kolář, Ph.D.

Konzultant:

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 detekcí polohy lidské hlavy ve stereo datech. Teoretická část poskytuje podklady pro porozumění geometrii kamery, jejím parametrům a způsobu kalibrace. Dále jsou popsány principy stereo analýzy a tvorby disparitních map. V řešeršní části jsou představeny metody používané pro modelování polohy hlavy a je uveden rozbor vybraných publikovaných článků.

V rámci diplomové práce byl navržen systém dvou kamer pro stereoskopická snímání pohybu hlavy a bylo realizováno několik měření. Získaná data byla upravena pro tvorbu disparitních map a další zpracování. Na základě detekce obličejových příznaků, konkrétně vnitřních a vnějších koutků očí a koutků úst, a jejich korespondencí byl vytvořen jednoduchý geometrický model ve tvaru trojúhelníku, který znázorňuje sklon obličejové roviny v prostoru. Výpočtem úhlů sklonu ve třech osách je získána aktuální poloha hlavy. Pohyb je modelován sledováním detekovaných bodů v průběhu video sekvencí.

KLÍČOVÁ SLOVA

Počítačové vidění, stereovize, korespondence bodů, disparitní mapa, poloha hlavy

ABSTRACT

The thesis deals with head pose estimation in stereo data. The theoretical part provides the basis for understanding the geometry of the camera, its parameters and the method of calibration. The following describes the principles of stereo analysis and creating of disparity maps. In the research section, the methods used for head pose modelling are presented and an analysis of selected published articles is given.

In the course of the master's thesis, a system of two cameras for stereoscopic acquisition of motion of the head was designed and several measurements were carried out. The obtained data was prepared for creation of disparity maps and further processing. Based on the detection of facial features, in particular the inner and outer corners of the eyes and corners of the mouth, and their correspondences, a simple geometric model in shape of triangle was created to illustrate the inclination of the facial plane in space. By computing the angle of inclination in three axes, the current head pose is obtained. Motion is modelled by tracking detected points during video sequences.

KEYWORDS

Computer vision, stereovision, point correspondence, disparity map, head pose

HŘÍBKOVÁ, Veronika. *Modelování polohy hlavy pomocí stereoskopické rekonstrukce*. Brno, 2018, 67 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství. Vedoucí práce: doc. Ing. Radim Kolář, PhD.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Modelování polohy hlavy pomocí stereoskopické rekonstrukce“ jsem vypracovala 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 autorka uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského 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.

Brno

.....

podpis autorky

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu diplomové práce panu doc. Ing. Radimu Kolářovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno

.....

podpis autorky

OBSAH

Úvod	9
1 Geometrie kamery	10
1.1 Projektivní model kamery	10
1.2 Vnitřní a vnější parametry kamery	11
1.3 Modifikace snímku a jejich korekce	14
1.3.1 Radiální zkreslení	14
1.3.2 Tangenciální zkreslení	15
1.4 Kalibrace kamery	15
2 Stereo analýza	18
2.1 Epipolární geometrie	18
2.1.1 Fundamentální matice	19
2.1.2 Esenciální matice	21
2.2 Stereorekonstrukce	21
2.2.1 Rektifikace	21
2.2.2 Korespondence bodů	22
2.2.3 Triangulace	23
2.3 Disparitní analýza	24
2.3.1 Disparita a hloubka bodu	24
2.3.2 Principy rekonstrukce disparitní mapy	24
3 Rešerše publikovaných metod	26
3.1 Rozdělení přístupů k detekci polohy hlavy	26
3.2 Popis vybraných metod	28
3.3 Dostupné databáze	29
4 Realizace	30
4.1 Kalibrace kamer	30
4.1.1 Postup kalibrace	30
4.1.2 Výsledky kalibrace	31
4.2 Návrh konfigurace systému kamer	31
4.3 Realizace stereoskopického snímání	32
4.3.1 Popis získaných dat	32
5 Navržená metoda	34
5.1 Rektifikace video sekvencí	34
5.2 Tvorba disparitní mapy, 3D vizualizace	35

5.2.1	Úprava disparitní mapy	37
5.3	Detekce obličejových příznaků a nalezení korespondencí	37
5.3.1	Detekce zorniček a očních koutků	38
5.3.2	Detekce koutků úst	39
5.4	Převod do prostorových souřadnic	40
5.5	Tvorba 3D modelu a výpočet polohy	41
5.6	Trackování detekovaných bodů	42
6	Diskuze dosažených výsledků	46
6.1	Výsledky jednotlivých částí algoritmu	46
6.1.1	Segmentace oblasti pleti a tvorba disparitní mapy	46
6.1.2	Detekce obličejových příznaků	46
6.1.3	Trackování	48
6.1.4	Modelování polohy hlavy	50
6.2	Testování při ztížených podmínkách	51
6.3	Úskalí geometrických metod	52
6.4	Výpočetní nároky	53
7	Závěr	54
	Literatura	56
	Seznam příloh	60
A	Návrh konfigurace kamer	61
B	Ukázky výsledků algoritmu	62
C	Obsah přiloženého DVD	67

SEZNAM OBRÁZKŮ

1.1	Geometrie dírkové kamery	10
1.2	Obrazový a kamerový systém souřadnic	12
1.3	Radiální zkreslení	14
1.4	Změna polohy bodu v důsledku radiálního a tangenciálního zkreslení	15
2.1	Geometrie korespondence bodů	19
2.2	Epipolární geometrie	20
2.3	Problémy v detekci korespondence bodů	23
3.1	Stupně volnosti měřeného pohybu hlavy	26
4.1	Detekce rohů čtverců kalibrační šablony	31
4.2	Chyba vzniklá zpětnou projekcí detekovaných bodů	32
4.3	Ukázka rektifikace stereopáru	33
5.1	Ukázka stereo páru snímků získaných dat	34
5.2	Anaglyf původních a rektifikovaných snímků stereo páru.	35
5.3	Segmentace oblasti podle odstínu pleti	36
5.4	Disparitní mapa a 3D vizualizace	36
5.5	Upravená disparitní mapa a 3D vizualizace	37
5.6	Jasový profil oblasti očí	38
5.7	Detekce zorniček	39
5.8	Výsledek detekce očních koutků	39
5.9	Výsledek detekce koutků úst	40
5.10	Schéma trojúhelníkového modelu	41
5.11	Trojúhelníkový model	42
5.12	Ukázka průběhu trackování	44
5.13	Úprava polohy trackovaných bodů pomocí geometrické transformace	45
5.14	Průběh počítaných úhlů sklonu trojúhelníkového modelu	45
6.1	Ukázka nepřesné detekce očních koutků	48
6.2	Přínos využití vzájemné korelace ke zpřesnění detekce koutků úst	48
6.3	Ukázka průběhu modelování polohy hlavy	50
6.4	Průběh	51
6.5	Vliv nerovnoměrného osvětlení na detekci příznaků a segmentaci pleti	51
6.6	Testování s brýlemi	52
A.1	Konfigurace systému kamer	61
A.2	Parametry kamer určené kalibrací v Matlabu	61
B.1	Segmentace uzavřené oblasti o pletovém odstínu	62
B.2	Ukázka vytvořených disparitních map a 3D vizualizací	63
B.3	Příklady detekce obličejových příznaků	65
B.4	Příklady grafů průběhů úhlů sklonu	66

ÚVOD

Schopnost odhadnout orientaci hlavy druhé osoby nám v běžném životě umožňuje snadno a efektivně vyhodnotit vzájemné postavení, zamýšlené pohyby a je také důležitou součástí neverbální komunikace. Tato lidem přirozená dovednost představuje v posledních letech jedinečnou výzvu pro systémy počítačového vidění. Proces modelování polohy hlavy vyžaduje přechod z dvojrozměrného prostoru (2D) snímků do trojrozměrného (3D) konceptu popisujícího směr.

Ideální systém se musí vyznačovat robustností vůči celé řadě faktorů, mezi které patří jevy fyzikální, například zkreslení kamery, nerovnoměrné osvětlení nebo důsledky projektivní geometrie, stejně tak jevy biologické, jako rozmanitost tváří a výrazů.

Určení polohy hlavy má uplatnění v řadě aplikací. Je důležitým předstupněm detekce, rozpoznávání a modelování lidské tváře a určování směru pohledu, které se následně uplatňují při tvorbě bezpečnostních systémů, sledování aktivity řidičů, řízení robotů a dalších [1], [2].

V rámci diplomové práce byl navržen systém dvou kamer vhodný pro stereoskopická snímání pohybu hlavy člověka sedícího přibližně 70 cm před systémem a vykonávajícího pomalé rotační pohyby kolem hlavních os. Nasnímané sekvence jsou rektifikovány a upraveny pro tvorbu disparitních map. Pro stereoskopickou rekonstrukci polohy hlavy v prostoru je zvolena geometrická metoda, založená na detekci obličejových příznaků a jejich korespondencí. Pomocí této stereorekonstrukce je pohyb v záznamech detekován a vyhodnocován ve třech stupních volnosti – jeden pro každou z hlavních os (dále znázorněno na obr. 3.1). Na základě sledování pozic prostorových souřadnic detekovaných bodů je vytvořen jednoduchý model simulující změnu polohy hlavy jedince v nasnímaných video sekvencích.

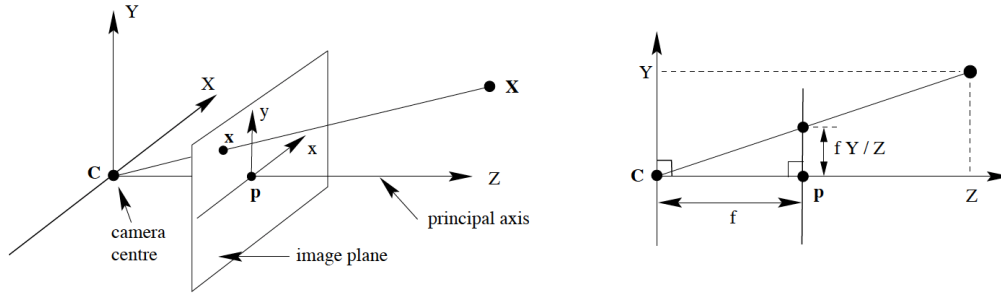
V rámci teoretických podkladů jsou v první části práce popsány geometrické vlastnosti kamery a jejich důsledky, uvedeny vnitřní a vnější parametry kamery a princip jejich získání kalibrací (kapitola 1). Dále budou zmíněna nejvýznamnější zkreslení vznikající vlivem konstrukčních nedokonalostí kamer a možnosti jejich korekce. Kapitola 2 dává teoretické podklady k stereorekonstrukci 3D scény ze dvou pohledů. Pomocí disparitní analýzy získaných sekvencí lze rekonstruovat disparitní mapy, ze kterých je možné přímo vyvozovat informace o hloubce bodů ve scéně. Kapitola 3 poskytuje přehled publikovaných metod týkajících se detekce polohy hlavy a popis postupu práce ve vybraných článcích. Realizace stereoskopického snímání, návrh parametrů systému a kalibrace použitých kamer je popsána v kapitole 4. Kapitola 5 popisuje jednotlivé kroky navržené metody od předzpracování získaných dat a detekce obličejových příznaků, po tvorbu 3D modelu a výpočtů polohy hlavy v průběhu video sekvencí. Dosažené výsledky jsou diskutovány v kapitole 6.

1 GEOMETRIE KAMERY

Pomocí kamery je mapována 3D scéna (předmětový prostor) do 2D obrazu. Obraz je při tom formován jejím optickým a elektronickým zařízením. Jde o komplexní proces, který je matematicky popsán geometrickým modelem kamery. Různé modely zahrnují v závislosti na odlišných předpokladech nebo zjednodušeních různé aspekty formování obrazu. V následující části bude uveden princip středového promítání. Modely středové projekce jsou specializacemi obecné projektivní kamery. Geometrické charakteristiky kamery, jako střed projekce nebo obrazová rovina, mohou být spočítány přímo z její maticové reprezentace. Dále budou popsány vnitřní a vnější parametry kamery a princip její kalibrace.

1.1 Projektivní model kamery

Vztahy pro popis projekce prostorové scény do roviny obrazu podle principu středového promítání lze odvodit z velmi jednoduchého modelu dírkové kamery (z angl. *pinhole camera*) na obr. 1.1.



Obr. 1.1: Geometrie dírkové kamery [3].

Optický střed kamery C je středem promítání a leží v počátku Euklidovského souřadného systému. Obrazová rovina je kolmá na osu z , tzv. hlavní osu kamery, a leží v ohniskové vzdálenosti f od kamerového středu. Bod p , průsečík hlavní osy a obrazové roviny, se nazývá hlavní bod. Poloha objektů je určena vzhledem k referenčnímu souřadnému systému spojenému se scénou, označovaným jako souřadnice světové. Bod X je v prostorové scéně vyjádřen trojicí souřadnic $X = (X, Y, Z)^T$. Do obrazové roviny se zobrazí jako bod x , vyjádřený $x = (x, y, z)^T$ vzhledem k světovým souřadnicím, v místě průsečíku spojnice bodu X se středem kamery C s obrazovou rovinou. Na základě podobnosti trojúhelníků lze souřadnice bodů vyjádřit

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z}. \quad (1.1)$$

Pokud neuvažujeme souřadnici výsledného obrazu v ose z , vztah

$$(X, Y, Z)^T \mapsto (fX/Z, fY/Z)^T \quad (1.2)$$

popisuje mapování světových souřadnic do souřadnic obrazových pomocí středového promítání. Jde o mapování z trojrozměrného Euklidovského prostoru $\mathbb{R}^3$ do dvojrozměrného Euklidovského prostoru $\mathbb{R}^2$.

Pokud jsou světové a obrazové souřadnice určeny homogenními vektory, středové promítání lze popsat jako lineární mapování mezi jejich homogenními souřadnicemi. Uvedené zobrazení matematicky vyjadřuje násobení matice s vektorem homogenních souřadnic

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (1.3)$$

Matice v tomto výrazu může být zapsána jako $\text{diag}(f, f, 1)[\mathbf{I} \mid 0]$, kde $\text{diag}(f, f, 1)$ je diagonální matice a $[\mathbf{I} \mid 0]$ reprezentuje matici o rozměru 3×3 (matici identity) a v tomto případě nulový sloupcový vektor.

Prostorový bod $\mathbf{X}$ tedy může být v homogenních souřadnicích vyjádřen vektorem $\mathbf{X}_h = (X, Y, Z, 1)^T$, podobně bod obrazu $\mathbf{x}_h = (fX, fY, Z)^T$. Potom homogenní matice o rozměru 3×4 je projekční maticí kamery $\mathbf{P}$. Rovnice 1.3 může být zapsána

$$\mathbf{x} = \mathbf{P}\mathbf{X}, \quad (1.4)$$

čímž je definována matice kamery dírkového modelu [3]

$$\mathbf{P} = \text{diag}(f, f, 1)[\mathbf{I} \mid 0]. \quad (1.5)$$

1.2 Vnitřní a vnější parametry kamery

Kamery jsou konstruovány tak, aby hlavní bod ležel v počátku souřadnicového systému obrazové roviny. V reálných aplikacích to vždy platit nemusí, proto se zavádí obecné mapování

$$(X, Y, Z)^T \mapsto (fX/Z + p_x, fY/Z + p_y)^T, \quad (1.6)$$

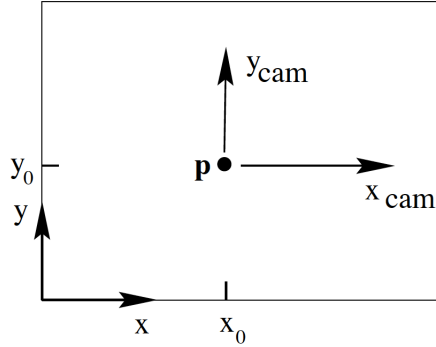
kde $(p_x, p_y)^T$ jsou souřadnice hlavního bodu. Potom je vztah 1.3 na základě obr. 1.2 upraven

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \quad (1.7)$$

$$\mathbf{x} = K[\mathbf{I} \mid 0]\mathbf{X}, \quad (1.8)$$

kde $\mathbf{K}$ je kalibrační matice kamery

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (1.9)$$



Obr. 1.2: Obrazový (x, y) a kamerový (x_{cam}, y_{cam}) systém souřadnic [3].

Souřadnice v obraze bývají vyjadřovány v počtu obrazových bodů – pixelů (z angl. *picture element*). V digitálních kamerách, využívajících například CCD detektory (z angl. *charge-coupled device*), má však pixel často tvar obdélníku, nikoli čtverce. Z tohoto důvodu je třeba zavést parametry zvětšení ve vodorovném s_x a svislém s_y směru obrazu. Pokud m_x a m_y jsou počty pixelů na jednotku vzdálenosti v obrazových souřadnicích v příslušných osách, potom transformace ze světových do obrazových souřadnic je provedena násobením matice $\mathbf{K}$ (výrazu 1.9) zleva faktorem $\text{diag}(m_x, m_y, 1)$. Tím je $\mathbf{K}$ rozšířena na obecný tvar kalibrační matice CCD kamery

$$\mathbf{K} = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1.10)$$

kde $\alpha_x = fm_x$ a $\alpha_y = fm_y$ jsou vyjádřením ohniskové vzdálenosti kamery v počtu pixelů ve směru osy x a osy y , obdobně souřadnice hlavního bodu jsou vyjádřeny $x_0 = m_x p_x$ a $y_0 = m_y p_y$.

Pro zobecnění uvažujme také možnost, že by na sebe osy v obraze nebyly kolmé. Tento jev popisuje parametr zešíkmení pixelů s . S jeho uvažováním má kalibrační

matice tvar

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1.11)$$

U běžně používaných CCD kamer bývá parametr s blízký nebo roven nule.

Uvedené vlastnosti, tedy relativní ohnisková vzdálenost f vyjádřená v pixelech, poměr stran pixelů, zešikmení pixelů a souřadnice hlavního bodu, patří mezi tzv. **vnitřní parametry** kamery. Dále budou uvedeny **parametry vnější**, které vyjadřují polohu kamery vůči referenčnímu souřadnému systému.

Body jsou v prostoru vyjádřeny v souřadnicích světového systému. Kamera, a tedy i počátek jejího souřadného systému, je v obecném případě ve scéně umístěna v pozici, kterou lze popsat posunutím a rotací vůči referenčním světovým souřadnicím.

Pokud $\widetilde{\mathbf{X}}$ je vektor nehomogenních souřadnic bodu ve světových souřadnicích a $\widetilde{\mathbf{X}}_{\text{cam}}$ reprezentuje stejný bod v souřadném systému kamery, platí

$$\widetilde{\mathbf{X}}_{\text{cam}} = \mathbf{R}(\widetilde{\mathbf{X}} - \widetilde{\mathbf{C}}), \quad (1.12)$$

kde $\widetilde{\mathbf{C}}$ je posun kamerového středu vůči počátku systému světových souřadnic a $\mathbf{R}$ je rotační matice o rozměru 3×3 popisující orientaci kamerového systému. S využitím homogenních souřadnic může být vztah přepsán do tvaru

$$\mathbf{X}_{\text{cam}} = \begin{bmatrix} \mathbf{R} & -\mathbf{R}\widetilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (1.13)$$

Doplněním do rov. 1.8 s uvážením matice vnitřních parametrů kamery lze zobrazení z prostrových souřadnic scény do souřadnic roviny obrazu vyjádřit vztahem

$$\mathbf{x} = \mathbf{K}\mathbf{R}[\mathbf{I} \mid -\widetilde{\mathbf{C}}]\mathbf{X}. \quad (1.14)$$

Projekční matice $\mathbf{P}$ obecné dírkové kamery je vyjádřena $\mathbf{P} = \mathbf{K}\mathbf{R}[\mathbf{I} \mid -\widetilde{\mathbf{C}}]$, kde $\mathbf{K}$ je maticí vnitřních parametrů kamery a matice $\mathbf{R}$ a $\widetilde{\mathbf{C}}$ popisují vnější parametry kamery [3][4].

Alternativně může být výpočet $\mathbf{P}$ proveden $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$, kde $\mathbf{t} = -\mathbf{R}\widetilde{\mathbf{C}}$. V tomto případě střed kamery není explicitně vyjádřen a transformace světových na obrazové souřadnice bude provedena jako

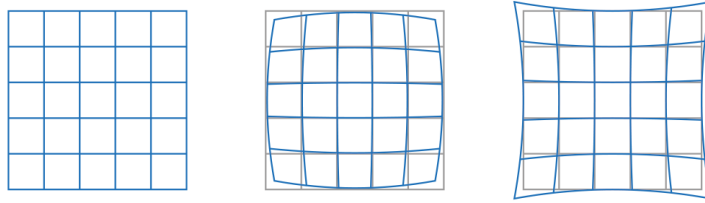
$$\widetilde{\mathbf{X}}_{\text{cam}} = \mathbf{R}\widetilde{\mathbf{X}} + \mathbf{t}. \quad (1.15)$$

1.3 Modifikace snímku a jejich korekce

Model projektivní kamery je popsán lineárními vztahy. V reálných kamerách však dochází vlivem nedokonalého tvaru čoček a dalších konstrukčních nedokonalostí ke zkreslením, která je možno aproximovat vztahy nelineárními. V obrazech se nejvýrazněji uplatňuje radiální a tangenciální zkreslení.

1.3.1 Radiální zkreslení

Doposud odvozený lineární model kamery předpokládá, že bod v prostoru $\mathbf{X}$, obrazový bod $\mathbf{x}$ a střed kamery $\mathbf{C}$ leží na společné přímce. Ve skutečných kamerách však tento předpoklad neplatí a obraz podléhá zkreslením, z nichž nejvýznamnější bývá zkreslení radiální. Vyskytuje se zvláště u levnějších nebo širokoúhlých kamer a je výraznější při použití kratší ohniskové vzdálenosti. Způsobuje nerovnoměrný posun souřadnic v obraze směrem od jeho okraje ke středu v případě tzv. zkreslení typu soudek, nebo naopak od středu k okraji – tzv. poduškové zkreslení, viz obr. 1.3. To se v obraze projeví zobrazením přímek do oblých křivek, nejvýrazněji u jeho okrajů. Radiální zkreslení souřadnic vzniká ještě před samotným zachycením snímku. Ve snaze docílit obrazu, který by zachytila lineární kamera, se provádí jeho korekce.



Obr. 1.3: Radiální zkreslení. Ukázka nezkreslených souřadnic (vlevo), zkreslení typu soudek (uprostřed) a zkreslení typu poduška (vpravo).

Skutečný zobrazený bod souvisí s ideálním nezkresleným bodem podle vztahu

$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = L(\tilde{r}) \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}, \quad (1.16)$$

kde $(x_d, y_d)^T$ je zachycená pozice bodu zkreslená radiální distorzí, $(\tilde{x}, \tilde{y})^T$ je ideální pozice bodu zobrazeného lineární projekcí, $\tilde{r} = \sqrt{\tilde{x}^2 + \tilde{y}^2}$ je vzdálenost od středu radiálního zkreslení a $L(\tilde{r})$ je distorzní faktor, funkce poloměru $\tilde{r}$.

Korekce se provádí podle rovnic

$$\begin{aligned} \hat{x} &= x_c + L(r)(x - x_c), \\ \hat{y} &= y_c + L(r)(y - y_c), \end{aligned} \quad (1.17)$$

kde $(x, y)^T$ jsou naměřené souřadnice v pixelech, $(\hat{x}, \hat{y})^T$ jsou souřadnice po korekci a $(x_c, y_c)^T$ jsou souřadnice středu radiálního zkreslení s $r^2 = (x - x_c)^2 + (y - y_c)^2$.

Funkce $L(\tilde{r})$ bývá aproximována Taylorovým polynomem

$$L(\tilde{r}) = 1 + \kappa_1 \tilde{r} + \kappa_2 \tilde{r}^2 + \kappa_3 \tilde{r}^3 + \dots \quad (1.18)$$

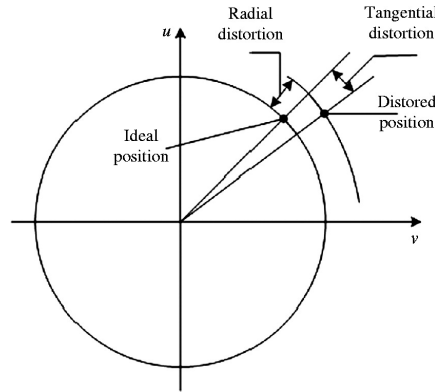
Koeficienty radiální korekce $\{\kappa_1, \kappa_2, \kappa_3, \dots, x_c, y_c\}$ jsou součástí vnitřní kalibrace kamery. Za souřadnice středu radiálního zkreslení jsou obvykle voleny souřadnice hlavního bodu, ačkoli se nutně nemusí zcela shodovat [3][4].

1.3.2 Tangenciální zkreslení

Optický senzor v kameře není vlivem tolerancí při výrobě umístěn zcela rovnoběžně s optickou rovinou objektivu. Tangenciální zkreslení se zpravidla ve snímku neprojevuje v takové míře, jako zkreslení radiální. Vyjadřuje se dvěma parametry p_1 a p_2 a jeho korekce se provádí podle rovnic

$$\begin{aligned} \hat{x} &= x + [2p_1 y + p_2(r^2 + 2x^2)], \\ \hat{y} &= y + [p_1(r^2 + 2y^2) + 2p_2 x]. \end{aligned} \quad (1.19)$$

Vliv obou zmíněných typů zkreslení je znázorněn na obr. 1.4 ze kterého je patrné, že posuny bodu vlivem tangenciálního a radiálního zkreslení jsou na sebe kolmé.



Obr. 1.4: Změna polohy bodu v důsledku radiálního a tangenciálního zkreslení [5].

1.4 Kalibrace kamery

Při použití nekalibrovaných kamer, tedy pokud matice kamer nejsou známy a jedinou dostupnou informací jsou korespondence bodů, lze získat model scény ze souboru snímků pomocí projektivní rekonstrukce. Tato úloha je mnohem jednodušší, pokud

známe souřadnice bodů v obrazech a matice kamer, pomocí nichž lze určit polohu bodů v prostoru tzv. triangulací.

Princip geometrické kalibrace spočívá v nalezení vnitřních a vnějších parametrů kamery, popsanych v kap. 1.2. V minulosti bylo navrženo mnoho metod kalibrace kamer lišících se přesností, složitostí výpočtu a nároky na technické vybavení. Technika fotogrammetrické kalibrace využívá snímání kalibračního objektu s přesně známou geometrií v 3D prostoru. Objekt se obvykle skládá ze dvou nebo tří navzájem kolmých rovin. Naopak techniky auto-kalibrace žádný kalibrační objekt nevyžadují. Parametry kamery lze určit ze snímků pořízených fotoaparátem pohybujícím se ve statické scéně pouze z korespondencí bodů. V současnosti je běžně používána metoda využívající rovinnou kalibrační šablonu, nejčastěji černobílou šachovnici o známých rozměrech čtverců, kterou navrhl Zhengyou Zhang [4]. Výhodou této techniky je především jednoduchost provedení. Vyžaduje pořízení několika (nejméně dvou) snímků kalibračního vzoru v různé orientaci vzhledem ke kameře. Umístění ani posunutí kalibračního vzoru při tom není třeba znát.

V první fázi je určena homografie mezi body šablony a jejich obrazy přímou lineární transformací. Beze ztráty na obecnosti metoda předpokládá, že kalibrační šablona leží v rovině $Z = 0$ světových souřadnic. Označíme-li i -tý sloupec rotační matice $\mathbf{R}$ jako $\mathbf{r}_i$, můžeme vztah 1.14 zjednodušit na

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \quad (1.20)$$

kde $(x, y, 1)^T$ je rozšířený vektor souřadnic bodu obrazu $\mathbf{x}$, $(X, Y, Z, 1)^T$ jsou souřadnice prostorového bodu $\mathbf{X}$, $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$ a $\mathbf{t}$ představují vnější parametry kamery – rotaci a translaci. Parametr s je libovolný nenulový faktor zvětšení.

Mezi bodem $\mathbf{X}$ a jeho obrazem $\mathbf{x}$ platí homografie $\mathbf{H}$:

$$\begin{aligned} s\mathbf{x} &= \mathbf{H}\mathbf{X}, \\ \mathbf{H} &= \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}. \end{aligned} \quad (1.21)$$

Matice $\mathbf{H}$ je rozměru 3×3 . Homografie mezi body kalibračního vzoru a jejich obrazy je určena s využitím kritéria maximální věrohodnosti, podrobněji v [4].

S označením $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3]$ a využitím znalosti, že $\mathbf{r}_1$ a $\mathbf{r}_2$ jsou ortonormální¹, je možné odvodit dvě základní podmínky kladené na vnitřní parametry kamery

$$\begin{aligned} \mathbf{h}_1^T K^{-T} K^{-1} \mathbf{h}_2 &= 0, \\ \mathbf{h}_1^T K^{-T} K^{-1} \mathbf{h}_1 &= \mathbf{h}_2^T K^{-T} K^{-1} \mathbf{h}_2. \end{aligned} \quad (1.22)$$

¹Ortonormální vektory – navzájem kolmé vektory o jednotkové délce.

V metodě je uvažována symetrická matice $\mathbf{B}$ definovaná vektorem

$$\begin{aligned}\mathbf{b} &= [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T, \\ \mathbf{B} &= \mathbf{K}^{-T} \mathbf{K}^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}\end{aligned}\quad (1.23)$$

Ve spojení s rovnicemi 1.22 lze sestavit homogenní soustavu rovnic, po jejímž vyřešení je získán vektor $\mathbf{b}$, který je zpětně přerovnan do matice $\mathbf{B}$. Z matice $\mathbf{B}$ již lze přímo vypočítat pět vnitřních parametrů kamery.

Vnější parametry kamery, rotační matici $\mathbf{R}$ a translační vektor $\mathbf{t}$, je možné počítat dosazením do rovnice 1.21.

Protože reálná data bývají zkreslena šumem, využívá se ke zlepšení daného řešení metoda maximální věrohodnosti. Mějme n obrazů kalibračního vzoru obsahujícího m významných bodů a předpokládejme zkreslení bodů identickým nezávislým šumem. Odhad maximální věrohodnosti bude získán minimalizací funkce

$$\sum_{i=1}^n \sum_{j=1}^m \|\mathbf{x}_{ij} - \hat{\mathbf{x}}(\mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j)\|^2, \quad (1.24)$$

kde $\hat{\mathbf{x}}(\mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j)$ je projekce bodu $\mathbf{X}_j$ v obraze i podle rovnice 1.21. Metoda vyžaduje počáteční odhad $\mathbf{X}$, $\{\mathbf{R}_i, \mathbf{t}_i | i = 1 \dots n\}$, který je získán výše popsanou technikou.

S uvážením koeficientů radiální distorze k_1 a k_2 lze rovnici 1.24 rozšířit na tvar

$$\sum_{i=1}^n \sum_{j=1}^m \|\mathbf{x}_{ij} - \hat{\mathbf{x}}(\mathbf{K}, k_1, k_2, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j)\|^2, \quad (1.25)$$

kde $\hat{\mathbf{x}}(\mathbf{K}, k_1, k_2, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_j)$ je projekce bodu $\mathbf{X}_j$ v obraze i zkresleném distorzí. Počáteční odhad koeficientů k_1 a k_2 lze dopočítat tzv. metodou alternace [4] nebo jednoduše nastavit na 0.

2 STEREO ANALÝZA

Stereo (z řec. *stereos*) znamená prostorový. Stereopsie je binokulární vidění, při němž vzniká prostorový vjem. Geometrií prostorových útvarů se zabývá stereometrie.

V předchozí kapitole byl popsán princip akvizice obrazu s využitím jedné kamery. Se dvěma nebo více snímky téže scény pořízenými z různých pozorovacích úhlů je možné pomocí geometrie zařízení odvozovat informace o hloubce bodů ve scéně a rekonstruovat její 3D model. Pokud jsou obrazy scény zachyceny současně dvěma kamerami, jsou nazvány stereo párem. Vlastnosti tímto způsobem konfigurovaného systému kamer jsou určeny epipolární geometrií, která popisuje vztah mezi body v prostoru zorného pole a jejich obrazy promítanými na příslušné promítací roviny. Obrazy jednoho bodu scény v jednotlivých promítacích rovinách se nazývají korespondující body a jejich vztah je popsán fundamentální maticí. Problém hledání korespondujících bodů patří mezi základní úlohy počítačového vidění. V praxi je hledání zatíženo mnoha faktory, které mohou proces zobrazení porušit. Jedná se o omezení obrazového rozlišení a kvantizace, zkreslení zavedené optickými systémy, šum a další. S využitím principů epipolární geometrie lze však hledání korespondujících bodů omezit do 1D prostoru epipolárních přímek, jak bude ukázáno v kap. 2.1 [3], [4].

2.1 Epipolární geometrie

Získání prostorové informace ze dvou, nebo více rovinných průmětů je základní úlohou počítačového vidění a umožňuje 3D rekonstrukci scény. Epipolární geometrie je geometrií dvou středových promítání. Slouží jako teoretický základ pro určení vztahu mezi snímky různých pohledů na téže scénu. Je nezávislá na obsahu scény, závisí pouze na vnitřních parametrech kamer a jejich vzájemné pozici. S jejím využitím lze mapovat určitý bod v jednom ze snímků stereo páru na přímku ve snímku druhém.

Jak již bylo zmíněno v kap. 1.1, základními prvky jednoduchého matematického modelu jsou střed kamery, projekční rovina, resp. plocha, a ohnisková vzdálenost. Projekční plochu si lze představit jako nasnímaný obraz.

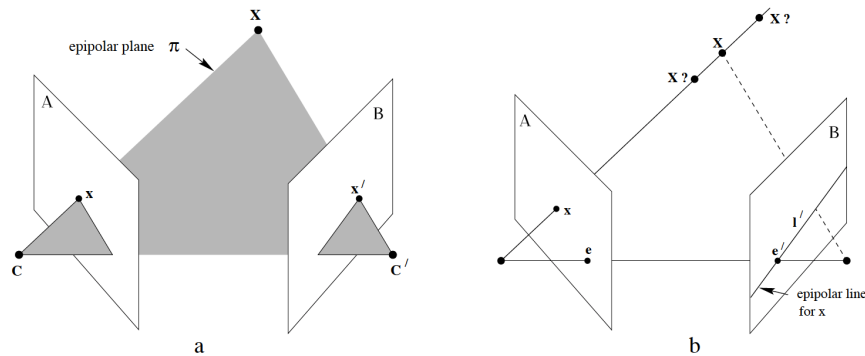
Na obr. 2.1 jsou znázorněny dvě kamery definované svými středy $\mathbf{C}$ a $\mathbf{C}'$ a obrazovými rovinami A a B . Bod $\mathbf{X}$ ležící v prostoru je promítán jako $\mathbf{x}$ do prvního pohledu a $\mathbf{x}'$ do druhého pohledu. Body $\mathbf{x}$, $\mathbf{x}'$ a $\mathbf{X}$ jsou koplanární, ležící v rovině π , a paprsky zpětně promítané ze středů kamer $\mathbf{C}$ a $\mathbf{C}'$ přes body $\mathbf{x}$ a $\mathbf{x}'$ leží v téže rovině a protínají se v bodě $\mathbf{X}$. Paprsek definovaný středem $\mathbf{C}$ a bodem $\mathbf{x}$ se ve druhém pohledu zobrazí jako přímka l' . Protože bod $\mathbf{X}$ leží na tomto paprsku, bude se bod $\mathbf{x}'$ nacházet na právě na přímce l' . Známe-li tedy oba promítnuté body, lze

z definovaného uspořádání kamer určit polohu prostorového bodu $\mathbf{X}$. Této vlastnosti se využívá při zpracování stereo páru na základě korespondencí, kdy prohledávání obrazu B za účelem nalezení bodu $\mathbf{x}'$ korespondujícího s bodem $\mathbf{x}$ obrazu A nezahrnuje celou jeho plochu, ale může být omezeno právě na úsečku l' (viz kap. 2.2.2) [3].

Základní terminologie:

- **Báze** je přímka procházející oběma středy kamer.
- **Epipól** je bod, ve kterém se protíná báze s obrazovou rovinou (v obr. 2.1 označen $\mathbf{e}$, resp. $\mathbf{e}'$), v závislosti na postavení a parametrech kamer však nemusí ležet v zobrazovaném výřezu projekční roviny. Ve speciálním případě, kdy jsou osy kamer vzájemně rovnoběžné, leží epipól v nekonečnu.
- **Epipolární rovina** prochází bází, existuje tedy jednoparametrový soubor epipolárních rovin, tzv. epipolární svazek. Konkrétní epipolární rovina je potom určena bází a příslušným analyzovaným bodem v prostoru.
- **Epipolární přímka** (epipolára) leží v průniku epipolární roviny s projekční rovinou. V obraze B je tedy projekcí přímky procházející středem $\mathbf{C}$ a bodem $\mathbf{X}$. Všechny epipolární přímky se protínají v epipólu.

Algebraickou reprezentací epipolární geometrie je fundamentální matice.



Obr. 2.1: Geometrie korespondence bodů. (a) Dvě kamery definované svými středy $\mathbf{C}$ a $\mathbf{C}'$ a obrazovými rovinami A a B . Prostorový bod $\mathbf{X}$, jeho obrazy $\mathbf{x}$ a $\mathbf{x}'$ a oba středy kamer leží ve společné rovině π . (b) Zpětná projekce bodu $\mathbf{x}$. Paprsek definovaný středem $\mathbf{C}$ a bodem $\mathbf{x}$ se do obrazu B promítne jako přímka l' [3].

2.1.1 Fundamentální matice

Fundamentální matice $\mathbf{F}$ určuje mapování bodu $\mathbf{x}$ v jednom obraze na korespondující epipolární přímku l' v obraze druhém $\mathbf{x} \mapsto l'$. Fundamentální matice je nezávislá na obsahu scény, může být vypočítána pouze z korespondencí bodů v obrazech

bez znalosti vnitřních parametrů kamer nebo jejich vzájemného postavení. Jedná se o matici velikosti 3×3 s hodnotí 2. Pokud je prostorový bod $\mathbf{X}$ zobrazen jako $\mathbf{x}$ v prvním obraze a $\mathbf{x}'$ ve druhém obraze, platí $\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0$.

Geometrické odvození

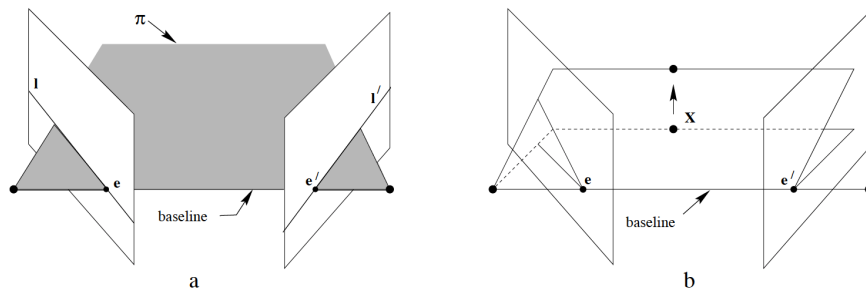
Mapování bodu z jednoho obrazu na korespondující epipoláru ve druhém obraze může být rozděleno do dvou kroků:

1. Přesun bodu prostřednictvím roviny. Uvažujme rovinu π , která neprochází ani jedním z kamerových středů. Paprsek s počátkem ve středu kamery $\mathbf{C}$ procházející bodem $\mathbf{x}$ protíná rovinu π v bodě $\mathbf{X}$. Bod $\mathbf{X}$ je potom promítnut jako $\mathbf{x}'$ v obraze druhé kamery. Protože $\mathbf{X}$ leží na paprsku procházejícím $\mathbf{x}$, promítaný bod $\mathbf{x}'$ musí ležet na epipoláře l' , která je obrazem tohoto paprsku. Body $\mathbf{x}$ a $\mathbf{x}'$ jsou tedy obrazy 3D bodu $\mathbf{X}$ ležícího v rovině π . Soubor bodů $\mathbf{x}_i$ v prvním obraze a korespondující body $\mathbf{x}_i'$ ve druhém obraze jsou projektivně ekvivalentní, protože každý z nich je projektivně ekvivalentní souboru planárních bodů $\mathbf{X}_i$. Existuje tedy 2D homografie H_π mapující každý $\mathbf{x}_i$ na $\mathbf{x}_i'$. Bod $\mathbf{x}'$ tedy může být popsán

$$\mathbf{x}' = H_\pi \mathbf{x}. \quad (2.1)$$

2. Konstrukce epipolární přímky. Pokud je dán bod $\mathbf{x}'$, lze epipoláru l' procházející $\mathbf{x}'$ a epipólem $\mathbf{e}'$ zapsat $l' = \mathbf{e}' \times \mathbf{x}' = [\mathbf{e}']_\times \mathbf{x}'$. S dosazením vztahu 2.1.

Geometricky reprezentuje $\mathbf{F}$ mapování z 2D projekční roviny P^2 prvního obrazu na svazek epipolárních přímek procházejících epipólem $\mathbf{e}'$ druhého obrazu. Reprezentuje tedy mapování z dvourozměrného do jednorozměrného prostoru, a proto má hodnotu 2.



Obr. 2.2: Epipolární geometrie. (a) Epipolární rovina π procházející bází a korespondujícími epipolárními přímkami l a l' . (b) V závislosti na poloze bodu $\mathbf{X}$ se rovina π otáčí podél báze, vytváří epipolární svazek [3].

2.1.2 Esenciální matice

Esenciální matice je speciálním případem fundamentální matice, která splňuje podmínku normalizovaných souřadnic. To znamená, že kamerový střed se nachází v počátku souřadného systému a osa z je hlavní osou kamery. Popisuje tedy geometrii kalibrované kamery. Esenciální matice $\mathbf{E}$ je definována

$$\hat{\mathbf{x}}'^T \mathbf{E} \hat{\mathbf{x}} = 0, \quad (2.2)$$

kde $\hat{\mathbf{x}}$ a $\hat{\mathbf{x}}'$ jsou normalizované obrazové souřadnice korespondujících bodů.

Vztah mezi fundamentální maticí $\mathbf{F}$, kalibrační maticí $\mathbf{K}$ a esenciální maticí $\mathbf{E}$ lze vyjádřit

$$\mathbf{E} = \mathbf{K}'^T \mathbf{F} \mathbf{K}. \quad (2.3)$$

Esenciální matice má 5 stupňů volnosti – 3 parametry pro rotaci a 2 pro translaci [3].

2.2 Stereorekonstrukce

V kap. 2.2.2 bude popsáno, jakými postupy lze nalézt korespondující body ve dvou obrazech stejné scény zachycených pod různými pozorovacími úhly. Je zřejmé, že korespondence mohou být hledány pouze v oblastech, ve kterých se obrazy překrývají. Epipolární geometrie redukuje problém vyhledávání korespondence bodů na hledání podél přímky v obraze procesem nazývaným rektifikace souřadnic (kap. 2.2.1). Pro každý pár korespondujících bodů je triangulací určen 3D bod, čímž lze rekonstruovat scénu v prostoru.

2.2.1 Rektifikace

Při použití speciálního uspořádání kamer, tzv. rektifikované konfigurace, na sebe obrazové roviny kamer navazují a jsou paralelní k bázi. Epipóly tedy leží v nekonečnu a epipolární přímky se shodují s řádky obrazu. Dále se předpokládá, že vnitřní parametry obou kamer jsou stejné.

V obecném případě jsou však používány kamerové systémy, které těmto předpokladům nevyhovují, např. při použití konvergujících kamer. Rektifikace obrazu je geometrická transformace, která umožňuje projekci obrazů, pořízených kamerami v obecné konfiguraci, do společné roviny paralelní k bázi. Spočívá v nalezení k tomuto účelu vhodných homografií H_L a H_R , úpravě obrazů a projekčních matic kamer. Potom

$$\begin{aligned} \mathbf{P}_L^R &= H_L \mathbf{P}_L = H_L \mathbf{K}_L \mathbf{R}_L [\mathbf{I} \mid -\mathbf{C}_L], \\ \mathbf{P}_R^R &= H_R \mathbf{P}_R = H_R \mathbf{K}_R \mathbf{R}_R [\mathbf{I} \mid -\mathbf{C}_R], \end{aligned} \quad (2.4)$$

kde $\mathbf{P}$ jsou projekční matice pravé a levé kamery, H jsou homografie a index R značí matice po provedené rektifikaci [6].

2.2.2 Korespondence bodů

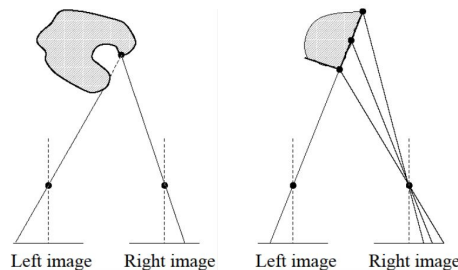
Určení korespondence znamená najít v obrazech stereo páru pixely $\mathbf{p}_L = (x_L, y_L)$ a $\mathbf{p}_R = (x_R, y_R)$, které jsou projekcí stejného bodu $\mathbf{P}$ 3D scény. Dále jsou uvedeny předpoklady, omezení a zjednodušení, kterých může být využito pro zvýšení spolehlivosti nebo zjednodušení a zefektivnění výpočtů bodových korespondencí [6], [7].

- Epipolární (geometrické) omezení: hledání korespondencí lze omezit na 1D problém vyhledávání podél příslušné epipolární přímky (viz kap. 2.2.1).
- Podmínka jedinečnosti: pro většinu bodů obrazu existuje jen jeden korespondující bod v obraze druhém. Výjimka nastává, pokud více bodů leží na paprsku vycházejícím z první kamery, tedy v prvním obraze se všechny zobrazí do jednoho bodu, ale do druhého obrazu se promítají jako samostatné body (viz obr. 2.3).
- Fotometrická kompatibilita: intenzity korespondujících bodů se liší jen velmi málo. Jejich rozdíl je způsoben odlišnými vzájemnými úhly mezi zdrojem světla, povrchovými normálami a senzorem.
- Geometrická podobnost: vyplývá z předpokladu, že geometrické charakteristiky objektů se mezi obrazy liší jen málo v závislosti na úhlu pohledu.
- Kontinualita disparity: disparita má tendenci se měnit pomalu, rozdíl mezi sousedními body může být omezen zadaným maximem.
- Kompatibilita znaků: body jsou považovány za korespondující, pokud si odpovídají i určité znaky obrazu v jejich okolí, např. hranové pixely odpovídají pouze jiným hranovým pixelům, rovněž je možné sledovat směr průběhu hrany. Zjištění náhlých nesouvislostí lze navíc využít pro odhalení okluzních bodů.
- Zachování pořadí: na površích o podobné hloubce korespondující body leží na epipolární přímce ve stejném pořadí. Ke změně pořadí bodů dochází v případě úzkého objektu směřujícího k bázi kamer.

Algoritmy využívané k detekci korespondencí je možné rozdělit na dva hlavní typy [6]:

1. low-level, bottom-up metody využívající korelace,
2. high-level, top-down metody založené na detekci vybraných znaků.

Metody založené na korelaci (*ad* 1) využívají předpokladu, že intenzity korespondujících pixelů jsou podobné. Pro snížení nejednoznačností v detekci využívají zároveň informaci z okolí pixelu, proto bývají označovány též jako algoritmy regionální. Hlavním parametrem určujícím úspěšnost detekce je nastavení velikosti



Obr. 2.3: Problémy v detekci korespondence bodů. Okluzní bod (vlevo), příklad výjimky z pravidla jedinečnosti (vpravo) [6].

porovnávaného okna – použití většího okna umožňuje lepší odlišení za cenu zvýšení výpočetního času. Příkladem může být algoritmus *Block matching*.

Znakové metody (*ad 2*) využívají detekce vybraných charakteristických znaků, např. hran, rohů nebo linií. Korespondence jsou poté určovány na základě jejich vlastností, jako orientace nebo délky. Tento přístup umožňuje lépe se vypořádat s nejednoznačnostmi především snížením počtu možných kandidátů na korespondující body, snížením závislosti na změnách jasu a zpřesněním detekce na subpixelovou úroveň.

K chybnému určení korespondujícího bodu často dochází v oblastech s jednotným jasnem nebo opakujícími se vzory. Dalším problémem jsou tzv. okluzní body, tedy body, které jsou viditelné pouze z jednoho pohledu [6].

2.2.3 Triangulace

Pojem triangulace má svůj původ ve fotogrammetrii a vychází z označení podobnosti trojúhelníků. Známe-li matice kamer $\mathbf{P}$ a $\mathbf{P}'$ a body obrazů stereo páru $\mathbf{x}$ a $\mathbf{x}'$, můžeme s použitím rovnic

$$\begin{aligned}\mathbf{x} &= \mathbf{P}\mathbf{X}, \\ \mathbf{x}' &= \mathbf{P}'\mathbf{X}\end{aligned}\tag{2.5}$$

vypočítat souřadnice prostorového bodu $\mathbf{X}$. Z geometrického hlediska triangulace obnáší nalezení společných průsečíků n paprsků daných zpětnou projekcí bodů obrazu, kde n označuje počet pohledů, v našem případě $n = 2$. Pokud by v obraze nebyl přítomen žádný šum a zkreslení, rovnice 2.5 by měly jedno řešení. V reálných případech se paprsky protínat nemusí a při $n = 2$ je zjednodušeně $\mathbf{X}$ počítáno jako střed nejkratší úsečky mezi dvěma paprsky. Pro minimalizaci chyby způsobené zpětnou projekcí bývá používáno kritérium maximální věrohodnosti [3], [6].

2.3 Disparitní analýza

2.3.1 Disparita a hloubka bodu

V obecném případě nerektifikovaných snímků je vektor disparit $\mathbf{d}$ dán rozdílem pozic korespondujících bodů $\mathbf{p}_L$ a $\mathbf{p}_R$ v obrazech

$$\mathbf{d} = \mathbf{p}_L - \mathbf{p}_R = (d_x, d_y). \quad (2.6)$$

Při použití rektifikovaných snímků, kdy epipolární přímky leží v odpovídajících si řádcích obrazu (viz kap. 2.2.1) lze rov. 2.6 zjednodušit na tvar

$$\mathbf{d} = (x_L - x_R, 0). \quad (2.7)$$

Disparitou d bude nadále míněn pouze posun bodu podél osy x .

Hloubka D pixelu $\mathbf{p} = (x, y)^T$ je definována jako Z -souřadnice příslušného bodu scény $\mathbf{X} = (X, Y, Z)^T$, potom $D(x, y)^T = Z$. Vyjadřuje tedy vzdálenost bodu $\mathbf{X}$ od středu kamery 1 měřenou podél hlavní osy kamery [3]. Hloubka je nepřímo úměrná disparitě. Předpokládáme-li, že kamery mají bázi b a stejné ohniskové vzdálenosti f , může být souřadnice Z bodu $\mathbf{X}$ vyjádřena

$$Z = \frac{bf}{d}. \quad (2.8)$$

Z rov. 2.8 vyplývá, že pokud $d \rightarrow 0$, potom $Z \rightarrow \infty$. Tedy body ve velkých vzdálenostech od kamery mají disparitu malou. Široká báze kamer může omezit relativní chybu v určování hloubky D [6], [8], [9].

Soubor disparit pro pixely celých obrazů stereopáru se nazývá disparitní mapa.

2.3.2 Principy rekonstrukce disparitní mapy

Hlavní úlohou při tvorbě disparitní mapy je tedy co nejpřesnější určení bodových korespondencí, jehož hlavní zásady byly popsány v kap. 2.2.2. Principy založené pouze na fotometrické kompatibilitě pixelů nejsou pro určení jednoznačné korespondence dostatečné, proto se využívá porovnávání okolí zahrnujícího několik sousedních pixelů. K nalezení pixelu $\mathbf{p}_R = (x_R, y_R)$ korespondujícího s pixelem $\mathbf{p}_L = (x_L, y_L)$ je vytvořeno obdélníkové okno se středem v $\mathbf{p}_L$ a v prohledávaném obraze je posouváno podél příslušné epipolární přímky. V každém kroku se přitom porovnává podobnost intenzit bodů uvnitř oken.

Pro srovnávání intenzit uvnitř oken jsou používány různé metriky. Nejjednodušší z nich, v praxi využívaná v mnohých systémech pracujících v reálném čase, je *SAD* (*Sum of Absolute Differences*). Pro každý pixel obrazu je vybrána disparita s nejmenší hodnotou *SAD*. Komplexnější metodou je minimalizace *SSD* (*Sum*

of Squared Differences), která umocněním zvýrazňuje větší rozdíly mezi intenzitami. Je používána také její upravená varianta, kde od každé hodnoty intenzity je odečtena průměrná hodnota daného okna. Další zlepšení výsledků přináší použití korelačních metod, například normalizované křížové korelace *NCC* (*Normalized Cross-Correlation*). Normalizace hodnot se využívá pro minimalizaci vlivu rozdílů jasu způsobených nerovnoměrným osvětlením. Problém určení optimální velikosti porovnávaného okna se už někteří autoři pokoušeli řešit jeho adaptivní velikostí nebo i tvarem.

Pro rekonstrukci modelu scény není třeba znát korespondenci všech bodů, stačí najít spolehlivé korespondence pro výrazné body v obraze. Jsou to body se specifickým okolím, jako hrany, rohy, lokální minima nebo maxima a podobně. Algoritmy pro hledání těchto bodů často využívají hranovou detekci, příkladem může být Moravcův detektor nebo Harris Conner detektor. Další možností je použití tzv. deskriptorů, které v obraze naleznou výrazné body a popíší jeho vlastnosti, na základě kterých potom mohou být hledány korespondence. Příkladem takového deskriptoru může být SIFT (*Scale-Invariant Feature Transform*) nebo SURF (*Speeded Up Robust Features*) [10].

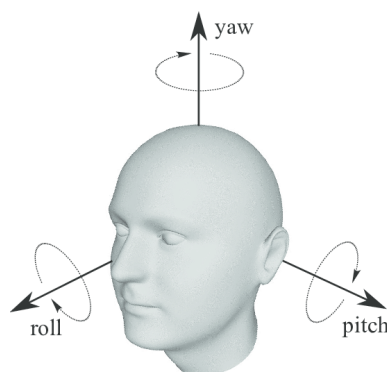
Chybné detekce a tedy i hodnoty disparit lze upravit následným zpracováním disparitních map. K chybnému určení často dochází v oblastech s homogenním jasnem. Takové oblasti mohou být identifikovány a příslušné disparity označeny za nespolehlivé. Spolehlivost detekce je možné hodnotit i na základě jednoznačnosti výběru peaku optimalizované veličiny. Další možností je hodnocení síly textur pomocí entropie. Disparitní mapy mohou být také vyhlazeny filtrací. Okluzní body je možné nahradit například pomocí lineární interpolace [8].

3 REŠERŠE PUBLIKOVANÝCH METOD

V kontextu počítačového vidění představuje proces odhadu polohy lidské hlavy určení její orientace z digitálních snímků. Zpracování vyžaduje řadu kroků, které se dokáží vyrovnat s variabilitou ve vzhledu lidské tváře i v podmínkách měření. Mezi požadavky kladené na detekční systémy patří především nezávislost na prostředí a světelných podmínkách, robustnost vůči vzhledovým odlišnostem, dostatečný rozsah měřitelného pohybu, automatické fungování bez potřeby manuálních zásahů nebo možnost aplikace na běžně dostupném hardwaru.

Algoritmy na nejnižší úrovni identifikují hlavu v některé z diskrétních pozic, například v předním pohledu, v pravém nebo levém profilu. Cílem je dosáhnout algoritmů, které budou schopny provádět kontinuální měření v daném úhlovém rozmezí s využitím několika stupňů volnosti (*Degrees of Freedom*, DOF).

Při záběru hlavy otočené na stranu nebude kvůli práci svalů její profil stejný, jako když bude zabírána v neutrální pozici kamerou z boční strany. Navzdory tomu bývá s hlavou pracováno jako se samostatným tuhým objektem. Na základě toho lze rozsah možných pohybů omezit na 3 DOF – otočení, úklon a náklon (viz obr. 3.1) [11].



Obr. 3.1: Stupně volnosti měřeného pohybu hlavy [11].

3.1 Rozdělení přístupů k detekci polohy hlavy

Postupy určování polohy lidské hlavy se liší v mnoha ohledech. Lze využít jak monokulární akvizici, tak stereo záznamy, zpracovávat snímky pořízené z blízkých vzdáleností i ve vzdálených polích, která jsou charakteristická nižším rozlišením. Dále jsou odlišeny stupněm automatizace, kdy některé přístupy mohou vyžadovat iniciální

interaktivní označení vybraných oblastí. V [11] je představen průzkum metod používaných pro detekci polohy hlavy. Metody jsou dle koncepce rozděleny do několika kategorií.

Metody využívající předlohy porovnávají nový obraz se souborem předloh. Předlohy jsou tvořeny obrazy lidských hlav v definovaných diskrétních pózách. Novému obrazu je přiřazena poloha hlavy definovaná pro nejpodobnější předlohu. K porovnání obrazů se využívá např. vzájemná normalizovaná korelace (*normalized cross-correlation*, NCC) nebo minimalizace kvadratické odchylky (*mean squared error*, MSE). Výhodou tohoto postupu je, že soubor předloh je možné kdykoli rozšířit a adaptovat na nové podmínky. Velmi dobře fungují také na datech s menším rozlišením. Mezi hlavní nevýhody patří fakt, že je možné určovat polohu hlavy jen v diskrétních pozicích dle zvoleného souboru předloh. Metoda navíc předpokládá vstupní data s již detekovanou oblastí hlavy.

Metody využívající detektorová pole trénují sérii detektorů hlavy, každý pro jednu diskrétní polohu. Přístup nevyžaduje předchozí detekci a lokalizaci hlavy. Díky procesu učení s učitelem trénování na mnoha předlohách jsou tyto systémy schopny vypořádat se s vzhledovou variabilitou. Jejich nevýhodou jsou především nároky na velikost souboru trénovacích dat, s tím rostoucí čas potřebný k trénování a nutnost trénovat velké množství detektorů dle počtu měřených póz. Protože detektory jsou použity jako binární klasifikátory, je nutné, aby si ve svých výstupech vzájemně neodporovaly. Učení proto zahrnuje také trénink na negativních snímcích. V praxi bývá detekce omezena na 1 stupeň volnosti a méně než 12 diskrétních póz.

Metody nelineární regrese podle souboru dat se známými polohami hlavy vytváří nelineární funkční mapování z obrazového prostoru do jednoho nebo více směrů pohybu. Vytvořený model umožňuje diskrétní nebo spojitou detekci v neznámých datech. Do této kategorie patří využití neuronových sítí, příkladem může být vícevrstvý perceptron trénovaný zpětným šířením chyby. Vícevrstvý perceptron má jeden výstup pro každý stupeň volnosti. Hodnota na tomto výstupu je přímo úměrná korespondující orientaci. Alternativně lze trénovat více perceptronů zvlášť pro každý stupeň volnosti. Výhodou neuronových sítí je především rychlost výpočtů a schopnost detekovat v blízkém i vzdáleném poli s vysokou přesností.

Pružné modely vyžadují u vstupních dat navíc detekci charakteristických znaků tváře. Na tomto základě je na individuální tvář fitován pružný model, který je tvořen deformovatelným grafem lokálních znaků. Tyto metody se vzhledem k výše uvedeným lépe vypořádávají s variabilitou ve vzhledu. Pro úspěšnou detekci je však nutné, aby se všechny porovnávané znaky stále nacházely na viditelných místech.

Geometrické metody využívají vzájemnou konfiguraci znaků, např. očí, úst nebo špičky nosu, bilaterální symetrii a polohu tváře vzhledem ke kontuře hlavy. Obvykle bývá používáno pět bodů: vnější koutky očí, špička nosu a koutky úst.

Spojením středů úseček mezi koutky očí a úst je získána osa symetrie tváře. Zároveň lze využít předpokladu, že vzdálenosti mezi těmito body jsou neměnné. Geometrické metody mohou být velmi jednoduché a účinné. Hlavním problémem je však přesnost detekce zvolených oblastí v případech, kdy se osoba pohybuje ve větší vzdálenosti od kamery nebo pokud v rámci rozsahu jejího pohybu tyto body nejsou po celou dobu viditelné. Podobný problém mohou způsobit též brýle, dlouhé vlasy nebo pokrývky hlavy stínící obličej.

Trackování využívá sledování relativního pohybu hlavy mezi po sobě jdoucími snímky video sekvence. Tyto metody dosahují vysoké přesnosti, ale vyžadují počáteční inicializaci. Sekvence tak typicky začíná detekcí v přímé pozici, kterou je někdy třeba opakovaně reinitializovat.

Hybridní metody využívají kombinace zmíněných metod k překonání jejich nedostatků. Obvyklým postupem je využít některé ze statických metod pro počáteční určení polohy a poté pokračovat trackováním. Pokud při trackování začne docházet k nepřesnostem, je možné odhad polohy reinitializovat. Hybridní systémy také umožňují použít dva nebo více navzávislých systémů a jejich odhady složit do jednoho výsledku.

3.2 Popis vybraných metod

Mezi nejpoužívanější techniky akvizice dat patří využití stereo systému kamer, strukturovaného světla nebo laserových senzorů. Použití páru kamer pro získání 3D informace je z uvedených nejméně náročné na technické vybavení. Projekci světelného vzoru na objekt lze získat 3D informaci analýzou zkreslení světelných skvrn na jeho povrchu pouze s využitím jedné kamery. Skenování pomocí laserových senzorů dosahuje nejpřesnějších výsledků, ale nevýhodami jsou vyšší náklady a doba akvizice.

Uvedené metody se v rámci zpracování dat liší v mnoha ohledech. V případě využití stereo systému kamer jsou však některé počáteční kroky společné. Nejprve je nutné kamery kalibrovat. V dalším kroku dochází k předzpracování dat, které spočívá v rektifikaci snímků, případně filtraci pro odstranění šumu nebo převodu do jiné barevné stupnice. Následuje oddělení oblasti hlavy od pozadí. K tomu lze v případě neutrálního pozadí použít prosté prahování [9]. V [1] je využito předpokladu, že lidská tvář je objekt nejbližší kamery, a hlava je segmentována na základě vytvořené disparitní mapy. Pro segmentaci tváře bývá dále využíváno předpokladu oválného tvaru a barvy kůže. Detekci je potom možné provést v HSV (*Hue-Saturation-Value*) barevném prostoru metodou narůstání oblastí s vytvářením propojené oblasti v přibližném tvaru elipsy. V tomto způsobem předzpracovaném obraze nyní může probíhat detekce jednotlivých znaků obličeje. V [12] je pro tuto segmentaci využito předpokladu nižšího kontrastu oblasti očí, nosních dírek a úst vůči okolní pleti,

který je ještě zvýrazněn pomocí morfologických operací. Následně jsou na snímcích s normalizovanou polohou hlavy vytvořeny šedotónové průběhy průměrných intenzit po řádcích a po sloupcích a v obou směrech jsou s využitím dalších omezení detekována minima, ve kterých se předpokládá výskyt těchto klíčových znaků. Dále je použito upravené Houghovy transformace pro segmentaci křivek obličeje hledáním největší, nejlépe aproximující elipsy, uvnitř které leží všechny znaky obličeje a odpovídá dříve vysegmentované oblasti. Na základě symetrie tváře jsou vytvořena další omezení, která například zabraňují označení okraje spodního rtu místo brady. Odhad obrysu je zpřesněn použitím segmentační metody aktivních kontur. Odhad polohy hlavy, který je v tomto případě považován za shodný se směrem pohledu, je určen na základě geometrických vlastností trojúhelníku s vrcholy v zorničkách a středu úst.

U metod využívajících k detekci zvolených oblastí šablony, kdy například pro detekci vnitřního koutku oka je použita šablona obsahující ve střední části koutek oka vzniklý průměrováním různých takových koutků, často dochází k nepřesnostem kvůli nedostatečné univerzálnosti šablon. V [13] je využito šesti význačných bodů: vnitřních a vnějších koutků očí a koutků úst. Úsečka spojující vnější koutky očí a úsečka spojující koutky úst jsou v 3D prostoru považovány za paralelní. K extrakci těchto znaků je použit operátor SUSAN (*smallest univalve segment assimilating nucleus*), popsáný v [14] přizpůsobený k detekci koutků. Nejdříve jsou nalezeny iniciální obdélníkové oblasti zvlášť pro každé oko a ústa, ve kterých bude hledání probíhat. Použitím operátoru SUSAN lze nalézt požadované body až s 90% přesností. Pro určení polohy hlavy je využíván úběžník definovaný kvůli větší robustnosti vůči šumu třemi přímkami procházejícími vnitřním a vnějším koutkem obou očí a koutky úst. Je nalezena rovina společná pro detekované znaky a směr natočení tváře je určen jako její normála.

3.3 Dostupné databáze

Protože detekce a rozpoznávání tváře patří mezi aktuální témata oboru počítačového vidění, bylo již vytvořeno několik databází obsahujících datasety snímků i sekvencí sloužících k testování algoritmů. Mezi používané databáze, které jsou volně nebo na požádání dostupné, patří:

- Color FERET database,
- M2VTS database,
- Texas 3D Face Recognition Database (Texas 3DFRD),
- Denver Intensity of Spontaneous Facial Action (DISFA) Database,
- Biwi Kinect Head Pose Database,
- ICT 3D HeadPose Database.

4 REALIZACE

Pro realizaci stereoskopického měření byl navržen systém dvou kamer umístěných vedle sebe na stativu s možností měnit jejich vzájemnou vzdálenost i natočení. Pro natáčení byly použity dvě kamery Imaging Source DFK 23U445 color camera. Kamery umožňují snímání barevných sekvencí v rozlišení $1280 \times 960 \text{ px}$, s maximální snímkovací frekvencí 30 fps a dynamickým rozsahem 8 bit . Dále byl využit externí zdroj napěťových pulzů pro synchronizaci snímkování.

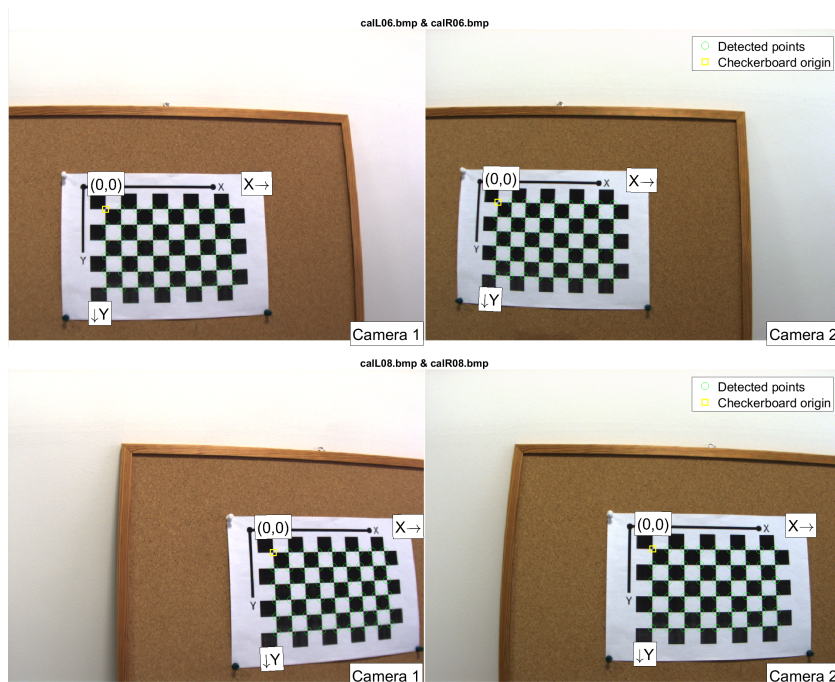
Akvizice sekvencí byla provedena v několika konfiguracích. Před každým novým snímáním bylo tedy nutné provést kalibraci kamer. Získaná videa byla sestříhána a dále zpracovávána v programovém prostředí Matlab. Sekvence byly rektifikovány a u vybraných snímků byly vytvořeny disparitní mapy. Podle kvality získaných dat byla následně konfigurace kamer dle potřeb upravena a provedeno nové měření.

4.1 Kalibrace kamer

Tato kapitola je věnována praktické realizaci kalibrace použitého páru kamer. Kalibrace je provedena v Matlabu s využitím aplikace *Stereo Camera Calibrator* a knihoven *Camera Calibration Toolbox* a *Computer Vision System Toolbox*. Aplikace byla vytvořena Jean-Yves Bouguetem na základě algoritmů popsanych v [4], [5] (kap. 1.4). Kompletní dokumentaci k aplikaci lze nalézt v [15]. Ke kalibraci jsou využity snímky kalibračního vzoru (viz obr. 4.1). Jedna z kamer je prohlášena za referenční (dále jen kamera 1) a k ní bude vztahována poloha druhé kamery (dále kamera 2). Výsledkem kalibrace tedy budou nalezené matice kamer, koeficienty radiálního zkreslení k_1 , k_2 a k_3 , koeficienty tangenciálního zkreslení p_1 a p_2 a přesné určení posunutí a rotace kamery 2 vzhledem ke kameře 1.

4.1.1 Postup kalibrace

Pro kalibraci byla použita kalibrační šablona černobílé šachovnice se čtverci o délce strany 22 mm . Byla pořízena série snímků tak, aby šablona postupně pokryla co největší část záběru. U jednotlivých provedených kalibrací bylo použito vždy 15 až 22 stereo párů. Po načtení snímků do aplikace byly automaticky detekovány rohy čtverců šablony ve všech snímcích (viz obr. 4.1) a vytvořen graf chyb vzniklých zpětnou projekcí bodů (viz obr. 4.2). Kalibraci bylo tedy možné provádět opakovaně, a zadáním počátečního odhadu kalibrační matice a koeficientů radiálního zkreslení, podle již dostupných odhadů, a vynecháním stereopárů s největšími chybami zpětné projekce, zvyšovat její přesnost.



Obr. 4.1: Ukázka detekce rohů čtverců kalibrační šablony u dvou stereo párů.

4.1.2 Výsledky kalibrace

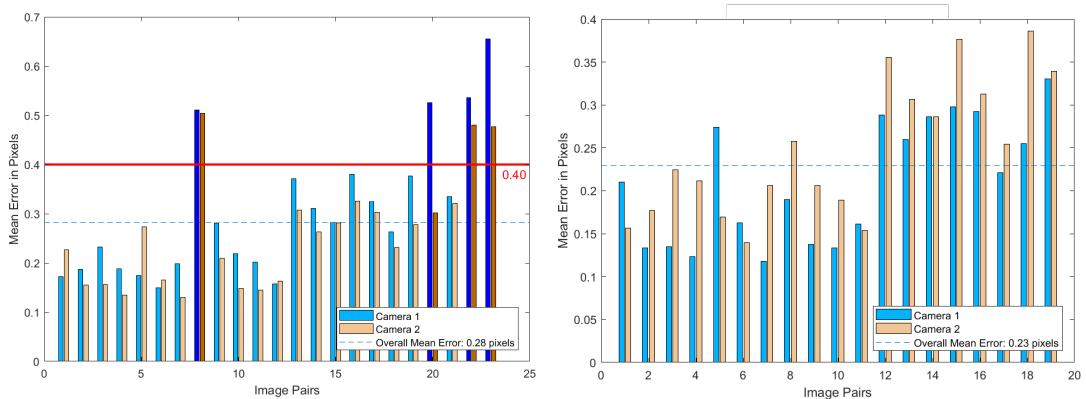
Protože akvizice byla prováděna několikrát při různých konfiguracích kamer, bylo nutné i kalibraci celého systému provádět opakovaně.

Výsledkem stereo kalibrace v Matlabu je objekt `stereoParameters` obsahující matice obou kamer, rotační matice a translační vektor kamery 2, fundamentální a esenciální matice a informace o detekovaných vzorech na jednotlivých kalibračních snímcích stereo párů. Příklad nalezených vnitřních parametrů kamer a jejich vzájemného postavení je na obr. A.2.

Na základě určených parametrů je možné snímky rektifikovat s využitím funkce `rectifyStereoImages`. Jejimi výstupy jsou příslušné nezkreslené a rektifikované snímky stereo páru. Příklad rektifikovaných snímků kalibračních šablon je na obr. 4.3. Lze vidět, že na liniích kalibračních šablon a na hranách nástěnky docházelo k výraznému radiálnímu zkreslení, které bylo rektifikací odstraněno.

4.2 Návrh konfigurace systému kamer

S cílem určit konfiguraci kamer vhodnou pro tato snímání bylo provedeno několik měření s bází b o délce 9 – 15 cm, vzdáleností od objektu přibližně 70 cm a mírnou konvergencí kamer. Pro náhled a ověření vhodnosti dat k další analýze byla vytvo-



Obr. 4.2: Chyba vzniklá zpětnou projekcí detekovaných bodů. Snímky s chybou větší než je červeně vyznačená hranice (vlevo) do konečné kalibrace (vpravo) zařazeny nebudou.

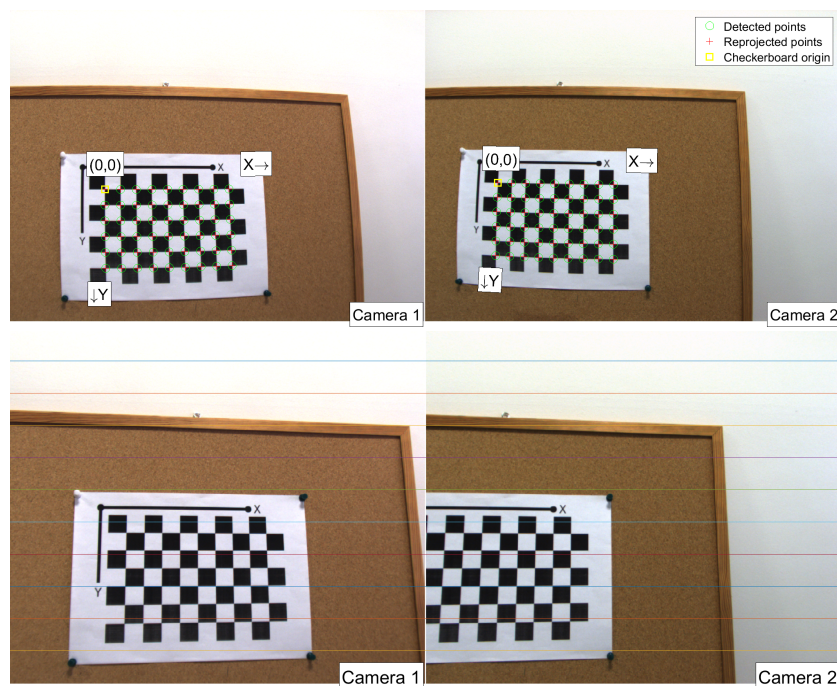
řena disparitní mapa pomocí funkce `disparity`, která využívá metodu *Semi-global Matching*. Příklad konfigurace kamer použité při sběru dat, s bází $14,5\text{ cm}$ a mírnou konvergencí, je v příloze na obr. A.1.

4.3 Realizace stereoskopického snímání

Sekvence byly zaznamenávány s využitím softwaru IC Capture, který je volně dostupný na webových stránkách výrobce použitých kamer. Pro ukládání byl použit formát *avi* (*Audio Video Interleaved*) a kodek MJPEG (*Motion JPEG*). Maximální možná frekvence při synchronizaci kamer pomocí externího zdroje napěťových pulzů byla 14 Hz . Pro první úpravu videí byl použit volně dostupný software VirtualDub. Pomocí něj byla videa sestřihána a byly vybrány zájmové úseky sekvencí. Do Matlabu jsou nejprve uloženy informace o videu do objektu `videoReader`, který obsahuje parametry popisující formát videa, rozlišení, snímkovací frekvenci, počet snímků, počet bitů na jeden pixel obrazových dat a další.

4.3.1 Popis získaných dat

S předpokladem, že cílem dalšího zpracování bude detekce pohybu v rámci tří DOF popsaných v kap. 3.2 a znázorněných na obr. 3.1, byli figuranti požádáni o vykonání pomalých rotačních pohybů kolem hlavních os: otočení hlavou doleva a doprava kolem osy z , předklon a lehký záklon kolem osy x , náklon na strany kolem osy y a nakonec volný pohyb hlavy ve všech třech osách. Snímání začínalo i končilo v neutrální poloze s pohledem do středu kamer. Ukázka jednoho stereo páru získané



Obr. 4.3: Ukázka rektifikace stereo páru.

sekvence je na obr. 5.1.

Celkem bylo nasnímáno 11 osob, 9 žen a 2 muži. Z toho 3 ženy a 1 muž byli natočení také s dioptrickými brýlemi, 4 ženy se sepnutými i rozpuštěnými vlasy a 2 s dodatečným bočním osvětlením.

Záběry byly sestříhány do sekvencí dlouhých 30 až 140 snímků a roztrženy do složek podle druhu pohybu. Při zpracovávání dat vyšlo najevo, že při natáčení delších sekvencí docházelo k problému se synchronizací kamer. Některé záběry tak musely být dodatečně upravovány nebo vynechány z testování.

5 NAVRŽENÁ METODA

Navržená metoda určování polohy hlavy je založena na detekci obličejových příznaků a nalezení jejich korespondencí. V prvních snímcích video sekvencí obou pohledů, u kterých se předpokládá přibližně čelní postavení, je detekováno po osmi bodech: středy zorniček, vnitřní a vnější koutky očí a koutky úst. Příklad takových snímků je na obr. 5.1. Ze znalosti parametrů kamerového systému jsou vypočítány jejich prostorové souřadnice a vzájemné postavení. Body je následně proložen jednoduchý geometrický model ve tvaru trojúhelníku se dvěma vrcholy v očích a jedním vrcholem na středu spojnice ústních koutků. Tento trojúhelník představuje obličejovou rovinu a jeho sklon určuje polohu hlavy. Změny polohy jsou modelovány sledováním detekovaných bodů v průběhu video sekvencí.

Prostorové souřadnice jsou vždy vztahovány k souřadnému systému kamery 1 a také označení „levý“ a „pravý“ u obličejových příznaků je z jejího pohledu.



Obr. 5.1: Ukázka stereo páru snímků získaných dat z pohledu levé kamery (vlevo) a pravé kamery (vpravo).

5.1 Rektifikace video sekvencí

Prvním krokem zpracování pořízených video sekvencí je rektifikace jednotlivých snímků na základě známých parametrů kamerového systému. Rektifikace je geometrická transformace, která umožňuje projekci obrazů, pořízených kamerami v obecné konfiguraci, do roviny paralelní k bázi (viz kap. 2.2.1). Epipolární přímky se shodují s řádky obrazu, což v dalších krocích usnadňuje problém hledání korespondencí, který je tímto zjednodušen na odpovídající si řádky. Na obr. 5.2 je zobrazen anaglyf stereo páru původních dat a odpovídající anaglyf stereo páru po rektifikaci. Během rektifikace jsou také odstraněna zkreslení zavedená optickým systémem kamer.

Na snímcích je patrný především vliv radiálního zkreslení.



Obr. 5.2: Anaglyf původních (vlevo) a rektifikovaných (vpravo) snímků stereo páru. Pro znázornění vlivu radiální distorze je použit režim zobrazení *full*.

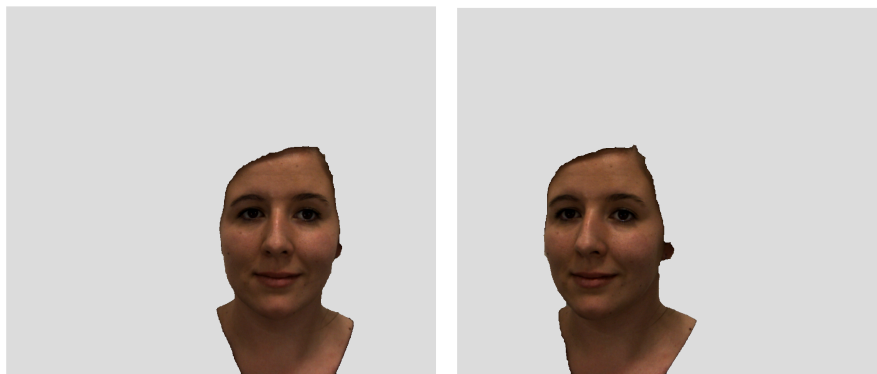
5.2 Tvorba disparitní mapy, 3D vizualizace

Pro lepší náhled na získaná data je pro první stereo pár video sekvencí vytvořena disparitní mapa a 3D vizualizace.

Nejprve je z obrazů odstraněno pozadí a také další objekty mimo oblast zájmu, jako oblečení a vlasy. To je provedeno nalezením oblasti s odstínem pleti. Pro tuto segmentaci je snímek převeden do barevného modelu YCbCr, který odděluje informaci o jasů do složky Y a informaci o barvě do složek Cb a Cr. Pro samotnou segmentaci je použita pouze složka Cr. Tento obraz je prahován s využitím automatické volby prahu metodou Otsu. Ve vzniklé binární masce je vybrán jeden největší objekt představující obličej, případně část krku a dekoltu. Masky je dále upravena binárním uzavřením s využitím strukturního elementu tvaru disku pro uzavření hranic a dále vyplněním děr, aby do výpočtů vstupovaly také důležité obličejové příznaky.

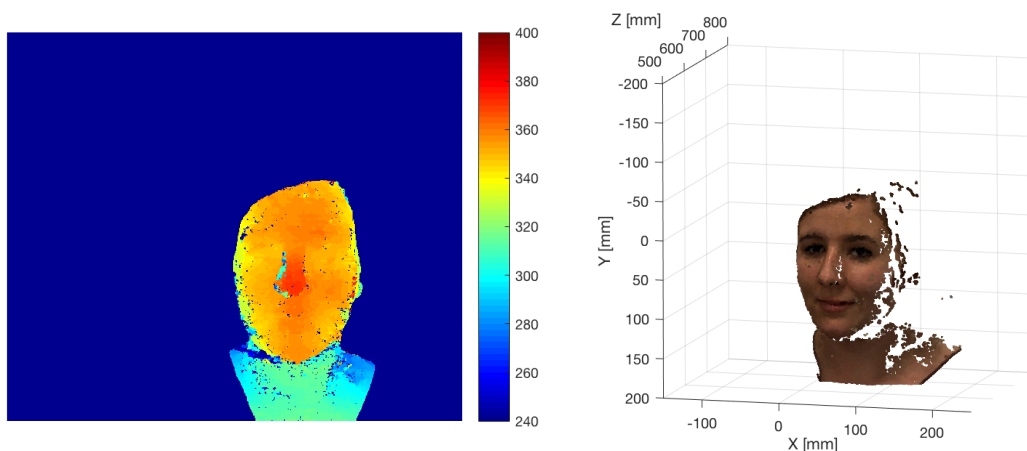
Disparita bodů je hledána metodou *Semi-Global Block Matching*, která pro porovnání podobnosti pixelů ve zvoleném okolí používá metriku *SAD* (*Sum of Absolute Differences*). Funkce zároveň vynucuje podobnou disparitu pro sousední bloky. Mezi parametry algoritmu patří velikost porovnávaného okolí a předpokládaný rozsah disparit, dále například práh pro jedinečnost nebo rozsah kontrastu pixelů, podle kterých je posuzována spolehlivost nalezených korespondencí. Výstupem je matice hodnot disparit pro jednotlivé pixely, tzv. disparitní mapa.

Bylo experimentálně ověřeno, že při tvorbě disparitní mapy vzniká nejméně artefaktů, pokud je jeden ze snímků převeden na obraz s vysegmentovanou oblastí pleti a druhý ponechán v původním stavu.



Obr. 5.3: Segmentace oblasti podle odstínu pleti uzavřené přes obličejové příznaky v levém pohledu (vlevo) a pravém pohledu (vpravo).

Nalezené hodnoty disparity lze podle vztahu 2.8 uvedeného v kap. 2.3 přímo převést na hodnoty hloubky ve světových jednotkách a stejným principem dopočítat i ostatní souřadnice. Vypočítané hodnoty se vztahují k souřadnému systému kamery 1 (viz kap. 5.5). V Matlabu je pro práci s 3D souřadnicemi použit objekt `pointCloud`, který umožňuje snadné vykreslení všech detekovaných bodů do souřadnicového systému, včetně přidělení jejich původní barvy v obraze (obr. 5.4).



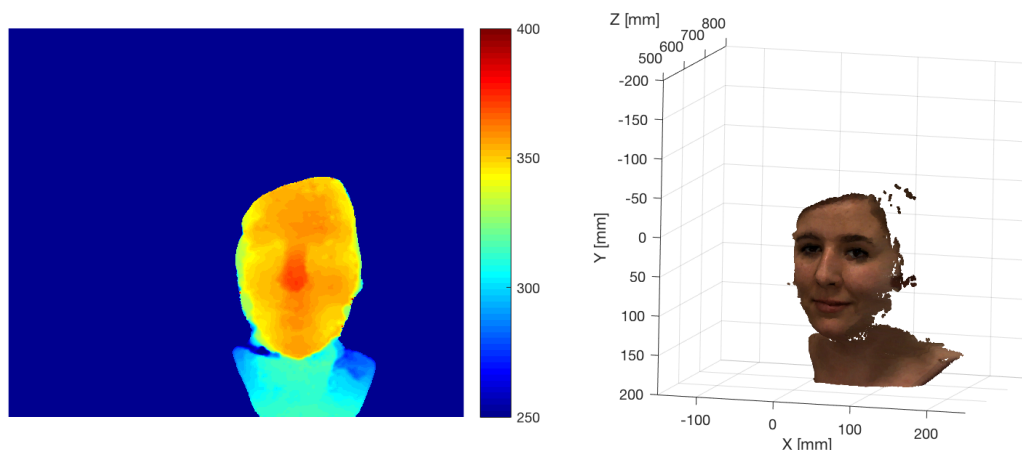
Obr. 5.4: Disparitní mapa (vlevo) a příslušná 3D vizualizace (vpravo). Disparitní mapa obsahuje velké množství chybně detekovaných nebo okluzních bodů, které se projeví jako nespojitosti v 3D vizualizaci.

5.2.1 Úprava disparitní mapy

Na obr. 5.4 je patrné, že vytvořená disparitní mapa obsahuje velké množství chybně detekovaných a okluzních bodů, zvláště po stranách nosu a v okrajových oblastech obličeje. Cílem dalších úprav je tedy tyto body najít a nahradit vhodnými hodnotami.

Nejprve je disparitní mapa filtrována mediánovým filtrem, který dostatečně zachová hrany, o velkém okolí (bylo použito okolí $29 \times 29 px$). Tento vyhlazený obraz je odečten od původní disparitní mapy. Zůstanou tedy jen oblasti s výrazně odlišnými hodnotami od svého okolí. V původní disparitní mapě jsou tyto body nahrazeny výrazy *NaN*. Následně je aplikována metoda *Inpainting*, která označené body nahradí na základě obsahu jejich okolí pomocí lineární interpolace. Posledním krokem je aplikace Wienerova filtru pro konečné vyhlazení.

Výsledky upravené disparitní mapy a příslušná 3D vizualizace jsou na obr. 5.5.



Obr. 5.5: Upravená disparitní mapa (vlevo) a příslušná 3D vizualizace (vpravo). Ve srovnání s obr. 5.4 došlo k vyhlazení disparitní mapy s možností kvalitnější vizualizace.

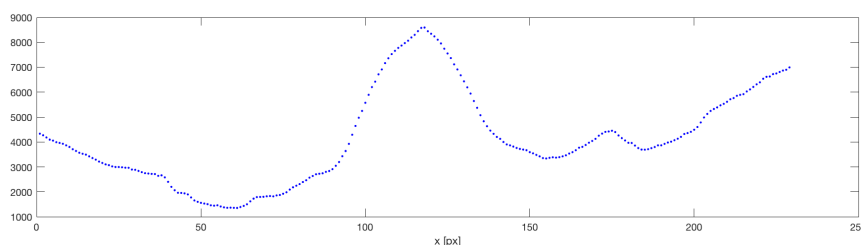
5.3 Detekce obličejových příznaků a nalezení korespondencí

Detekce obličejových příznaků, na jejichž základě bude později vytvářen model, začíná nalezením oblasti obličeje v obou pohledech. K tomu je využit kaskádový detektor představený P. Violou a M. Jonesem [16]. Detektor pracuje s integrálním obrazem, ze kterého získává množství jednoduchých příznaků na principu Haarových vlnek, tzv. *Haar-like features*, které vstupují do klasifikačního algoritmu *AdaBoost*

(*Adaptive Boosting*). Algoritmus implementovaný v Matlabu je naučený například na detekci obličeje, očí, nosu nebo úst. V rámci této práce je používán k určení prvotní oblasti zájmu, ve které budou vyhledávány konkrétní požadované body. Cílem je nalezení přesných korespondencí mezi pohledy pro zmíněných osm obličejových bodů.

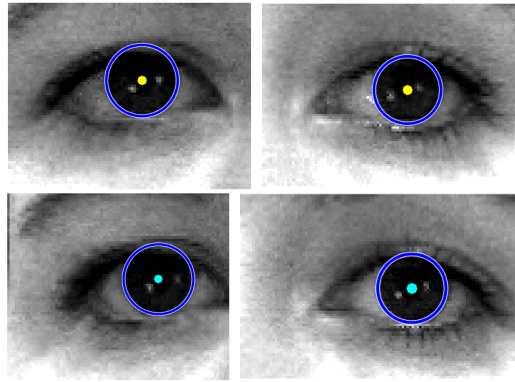
5.3.1 Detekce zorníček a očních koutků

Na obličejovou oblast levého pohledu je aplikován kaskádový detektor pro pár očí, který na získaných datech prokazoval mnohem lepší účinnost než detektor jednotlivých očí. Oblast je z obrazu vyřezána a je vypočítán její jasový profil, kde každá hodnota udává součet hodnot v příslušném sloupci (obr. 5.6). Ve střední části tohoto profilu je výrazný pík, který představuje kořen nosu. Detekcí tohoto píku lze oči oddělit od sebe.



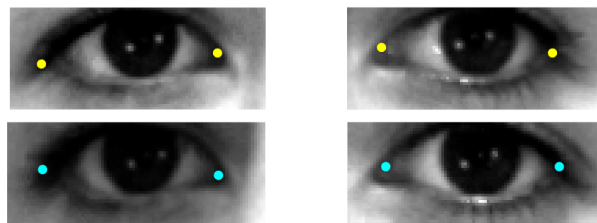
Obr. 5.6: Jasový profil oblasti očí. Výrazný pík ve střední části představuje kořen nosu.

Nejprve jsou detekovány středy zorníček. Z obou zájmových oblastí obsahujících oči jsou použity jen R kanály vykazující nejvyšší kontrast mezi duhovkou, bělimou a pletí. Kontrast je ještě zvýšen konvolucí s lokálním maskovým operátorem pro zvýraznění hran, konkrétně byla použita maska $[0, -1, 0; -1, 6, -1; 0, -1, 0]$. Na tyto zvýrazněné obrazy je aplikována Houghova transformace pro vyhledávání kružnic, v tomto případě tmavších, než je jejich okolí. Vstupní parametry, například rozsah poloměrů vyhledávaných kružnic, senzitivita a práh pro hodnotu gradientu hrany, jsou v případě, že na výstupu funkce nebyly nalezené žádné kružnice, iterativně upravovány. Tímto způsobem je pro levou i pravou duhovku obou pohledů nalezeno několik kružnic – kandidátů. Výběr nejvhodnějších kružnic ohraničujících duhovky je s využitím výhod rektifikace uskutečněn minimalizací v rozdílu řádkových souřadnic a poloměrů zvlášť pro levé a pravé oko mezi pohledy. Za střed zorničky je označen střed detekované kružnice. Výsledek detekce zorníček v upravených oblastech očí je zobrazen na obr. 5.7.



Obr. 5.7: Detekce zorniček v levém (nahore) a pravém (dole) pohledu pomocí Houghovy transformace. K hledání byly použity výřezy snímků se zvýrazněnými hranami a kontrastem.

Hledání očních koutků probíhá v obdélníkových oblastech kolem již detekovaných zorniček. Oblast je vyřezána z originálního obrazu a převedena do šedotónu s ekvalizací histogramu a je na ní aplikován Cannyho detektor hran. Výstupem je binární obraz, který je dále upraven pomocí odstranění objektů o velikosti menší než 10 px za cílem získat přibližnou hranovou reprezentaci obrysu oka. Jako koutky jsou označeny pixely ležící na pravém a levém okraji této reprezentace. Jejich poloha je ještě upravena posunem do nejtmavšího bodu v okolí $11 \times 11\text{ px}$ se středem v původně detekovaném bodu v šedotónovém obrazu [17]. Nalezené oční koutky v obou pohledech jsou zobrazeny na obr. 5.8.



Obr. 5.8: Výsledek detekce očních koutků v levém pohledu (nahore) a nalezené korepondence v pravém pohledu.

5.3.2 Detekce koutků úst

Na spodní část obličejové oblasti (pod detekovaným párem očí) obou pohledů je aplikován detektor naučený na vyhledávání úst. Podmínky pro detekce jsou opět na základě výstupu upravovány za cílem získat v každém pohledu alespoň jednu

kandidátní oblast. V případě více detekovaných oblastí jsou mezi pohledy opět vybrány ty, které nejlépe odpovídají předpokladům řádkových korespondencí.

Koutky úst jsou hledány použitím Sobelova operátoru pro detekci horizontálních hran na upravený šedotónový obraz oblasti úst [18]. Binárním uzavřením s použitím strukturního elementu ve tvaru horizontální čáry o délce $3px$ je upravena hranová reprezentace, která v tomto případě představuje kontury rtů, a především jejich spojnicí. Pozice levého koutku v pravém pohledu je zpřesněna hledáním vzoru v podobě okolí koutku detekovaného v levém pohledu ve větším okolí příslušného koutku detekovaného v pravém pohledu. Stejným způsobem je upravena i pozice pravého koutku v levém pohledu (viz obr. 5.9).



Obr. 5.9: Detekce koutků úst v levém (vlevo) a pravém (vpravo) pohledu. Červeně jsou znázorněny body detekované primárním algoritmem a žlutě/modře jejich upravené pozice.

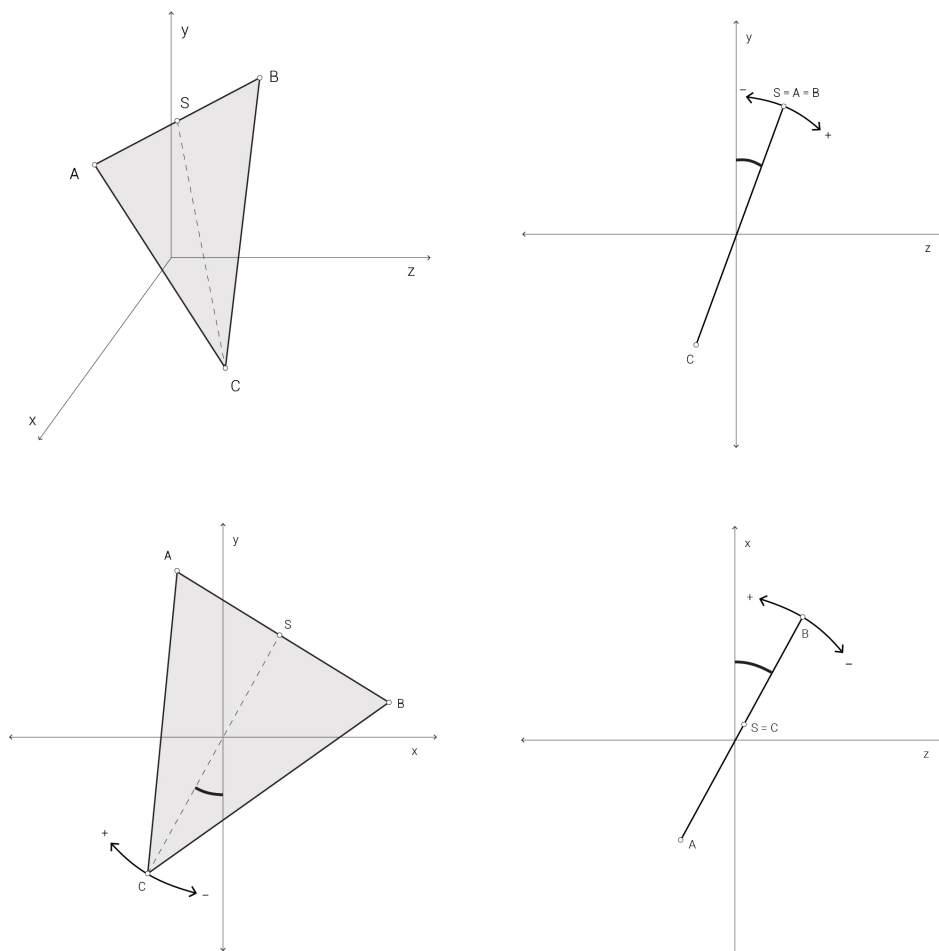
5.4 Převod do prostorových souřadnic

Po nalezení přesných korespondencí pro osm hlavních bodů následuje výpočet jejich prostorových souřadnic $(X, Y, Z)^T$ ve světových jednotkách (jsou použity *mm*). Veškeré hodnoty jsou vztahovány k souřadnému systému s počátkem v optickém středu kamery 1. Do výpočtů vstupují obrazové souřadnice bodů $(x, y)^T$ v obou pohledech, hlavní bod kamery 1 $(p_x, p_y)^T$, dále báze kamer b a ohnisková vzdálenost f . Nejdříve je podle vztahu 2.7 určena disparita d , kterou je podle rovnice

$$Z = \frac{bf}{d} \quad (5.1)$$

možné převést na hloubku bodu v obraze Z . Souřadnice X a Y jsou vypočítány podle rovnic

$$X = \frac{(x - p_x)Z}{f}, \quad Y = \frac{(y - p_y)Z}{f}. \quad (5.2)$$

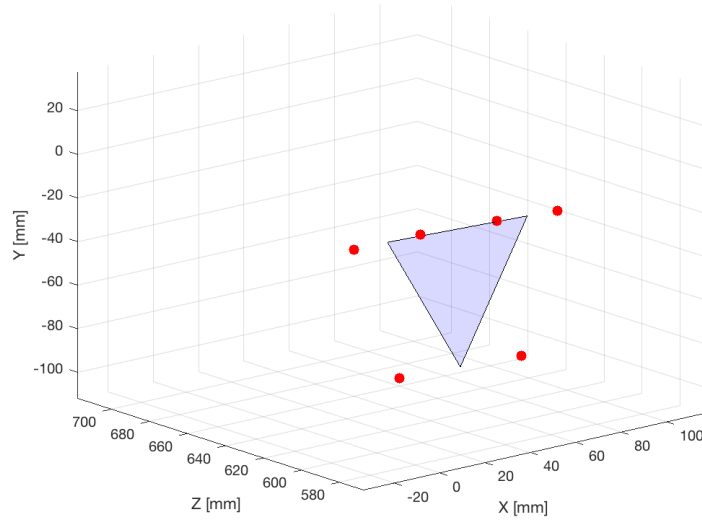


Obr. 5.10: Schéma trojúhelníkového modelu v souřadném systému kamery 1 a jeho 2D projekce se znázorněním měřených úhlů.

5.5 Tvorba 3D modelu a výpočet polohy

Geometrický model znázorňující polohu hlavy je počítán z šesti detekovaných bodů: vnitřních a vnějších koutků očí a koutků úst. Zorničky jsou z modelování vynechány kvůli pohybům, které nesouvisejí s pohybem celé hlavy. Body je proložen trojúhelník s vrcholy **A**, **B**, a **C**. Souřadnice bodu **A** jsou vypočítány jako průměr souřadnic vnitřního a vnějšího koutku levého oka, podobně bod **B** je počítán jako průměr souřadnic vnitřního a vnějšího koutku pravého oka. Bod **C** leží na středu spojnice koutků úst. Pro další výpočty je do středu úsečky **AB** umístěn ještě pomocný bod **S**. Celá situace je znázorněna na schématu na obr. 5.10, reálný vykreslený model včetně detekovaných bodů na obr. 5.11.

Jak již bylo zmíněno, tento trojúhelník přímo simuluje orientaci tváře. Jeho sklon



Obr. 5.11: Trojúhelníkový model znázorňující polohu a sklon tváře vypočítaný z reálných dat. Červeně jsou vykresleny detekované vnitřní a vnější koutky očí a koutky úst.

je počítán ve třech DOF vůči hlavním osám (viz obr. 3.1). DOF tedy představují pohyby úklon (z angl. *pitch*, podle osy x), otočení (z angl. *yaw*, podle osy y) a náklon (z angl. *roll*, podle osy z). Výpočet úhlů probíhá v jednotlivých snímcích na základě aktuálních souřadnic bodů **A**, **B**, **C** a **S** podle rovnic

$$pitch = tg^{-1} \left(\frac{S_z - C_z}{S_y - C_y} \right), \quad (5.3)$$

$$yaw = tg^{-1} \left(\frac{B_z - A_z}{B_x - A_x} \right), \quad (5.4)$$

$$roll = tg^{-1} \left(\frac{S_x - C_x}{S_y - C_y} \right). \quad (5.5)$$

Schémata, která znázorňují geometrický význam těchto rovnic, jsou na obr. 5.10.

5.6 Trackování detekovaných bodů

Často používanou metodou pro trackování bodů je algoritmus KLT (*Kanade-Lucas-Tomasi*, 1981 [19]) původně navržený pro lícování obrazů. Principem je hledání pozice předlohy v podobě okolí sledovaného bodu v novém snímku na základě výpočtu optického toku. Je při tom využito obrazových pyramid, ve kterých se v každé

úrovni snižuje rozlišení s faktorem 2. KLT tracker je často používán pro krátkodobé trackování v rámci rozsáhlejších trackovacích algoritmů. Mezi jeho nevýhody patří především to, že při chybné detekci nebo ztrátě sledovaného bodu nedochází k reinicializaci.

V této práci je KLT algoritmus využit jako základ pro stanovení lokace bodu v dalším snímku, současně jsou zavedeny další korekce a kritéria pro posouzení správnosti detekce.

Pro trackování je kromě šesti hlavních detekovaných bodů využito pomocných bodů – příznaků, rohových bodů detekovaných metodou *Minimum Eigenvalue Algorithm*, kterou pro výběr vhodných bodů k trackování navrhli J. Shi a C. Tomasi [20]. Na začátku jsou tyto body inicializovány v prvním snímku sekvencí v oblastech očí, nosu a úst, které jsou hledány pomocí kaskádového objektového detektoru uvnitř obličejové oblasti. Z detekovaných bodů je vybráno 100 s nejsilnější metrikou. Tyto body jsou trackovány zvlášť s nastavením přísných podmínek pro skóre (použita hodnota 0.9993), které udává podobnost okolí bodu v původním a novém snímku pomocí *SSD*. Body, které tyto podmínky nesplní, jsou z trackování odstraněny. Pokud počet pomocných bodů v některé iteraci trackování klesne pod 70, dojde k jejich novému hledání v příslušném snímku. Kaskádový detektor je natrénován na hledání očí, nosu a úst jen v přímém pohledu, proto je nyní použit pouze detektor obličeje. Oblast detekce nových bodů je omezena průnikem čtverce na jeho výstupu a oblasti, ve které barva odpovídá pletovému odstínu. Tímto způsobem je zajištěn stálý dostatečný počet pomocných bodů o vysokém skóre. Trackování probíhá současně pro hlavní i pomocné body v sekvencích obou pohledů. Na obr. 5.12 je ukázán příklad dvou snímků v průběhu sekvence s vykreslenými hlavními a pomocnými trackovanými body.

Také hlavní body mohou být z trackování vyřazeny kvůli nízkému skóre (limitní hodnota nastavena na 0.98). V takovém případě jsou nahrazeny výrazem *NaN*. Z hlavních bodů je na začátku vytvořen trojúhelníkový model pro první snímek sekvencí. Z prostorových souřadnic je vypočítána a uložena referenční matice vzájemných euklidovských vzdáleností šesti bodů. V každé iteraci probíhá výpočet nových prostorových souřadnic a nové vzdálenostní matice. Odečtením této matice od referenční je získána matice rozdílů. Pokud se vzdálenost některého z bodů od ostatních liší v součtu více než o 30 mm, jsou jeho pozice v pravém a levém pohledu označeny za nespolehlivé a nahrazeny výrazem *NaN*. Z pomocných bodů, detekovaných v předchozím a aktuálním snímku, jsou vypočítány parametry 2D afinní geometrické transformace. Nové pozice všech nespolehlivých bodů jsou vypočteny aplikací této transformace na jejich původní pozice. Transformace může být aplikována nejvýše ve 3 po sobě jdoucích snímcích. Příklad vhodné aplikace je uveden na obr. 5.13, kde došlo k okamžitému navrácení chybně detekovaného bodu

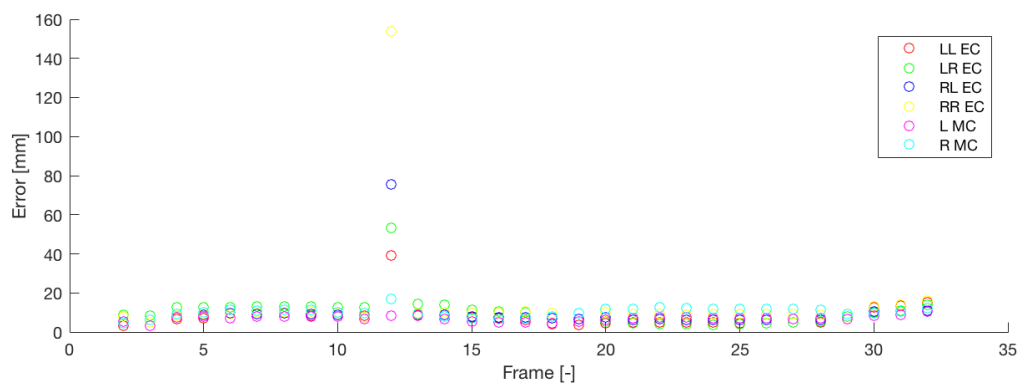


Obr. 5.12: Ukázka průběhu trackování. Nahoře vlevo snímek 2 a vpravo snímek 24 levého pohledu, dole příslušné snímky pravého pohledu. Zeleně jsou vykresleny hlavní trackované body, bíle pomocné body.

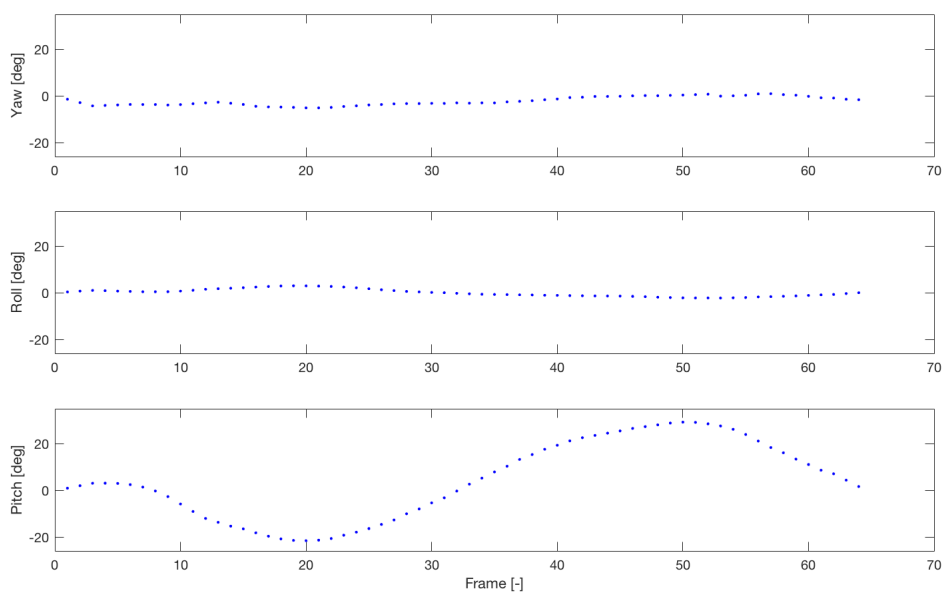
do správné pozice. V grafu jsou vykresleny hodnoty součtů řádků rozdílové vzdálenostní matice. Největší hodnota tedy přísluší chybně detekovanému bodu. Graf sloužil pouze pro hrubou kontrolu průběhu trackování.

V rámci trackování v každé iteraci proběhne zároveň výpočet a uložení prostorových souřadnic hlavních bodů, vrcholů trojúhelníkového modelu pro příslušné snímky i určení jeho sklonu popsáního třemi úhly. Na obr. 5.14 je zobrazen průběh vypočítaných úhlů sklonu v jednotlivých snímcích sekvence. Pro odstranění drobných nepřesností je signál filtrován průměrovacím filtrem o velikosti okna 3.

Vytvořená funkce `visualization.m` umožňuje náhled na simulaci pohybu hlavy postupným vykreslením trojúhelníků do souřadnicového systému. Pro názornější a plynulejší animaci modelu jsou záznamy souřadnic vrcholů trojúhelníku převzorokovány s faktorem 3 a interpolovány.



Obr. 5.13: Úprava polohy trackovaných bodů aplikací geometrické transformace na chybně detekovaný bod ve snímku 12. Vykreslené body představují součty v řádcích rozdílové vzdálenostní matice.



Obr. 5.14: Průběh tří počítaných úhlů sklonu trojúhelníkového modelu v jednotlivých snímcích sekvence.

6 DISKUZE DOSAŽENÝCH VÝSLEDKŮ

V této kapitole jsou popsány úspěšnosti dílčích částí algoritmu. Vzhledem k povaze práce byla většina výsledků hodnocena subjektivně, například posouzením odchylky detekce od ideálně označeného bodu.

6.1 Výsledky jednotlivých částí algoritmu

6.1.1 Segmentace oblasti pleti a tvorba disparitní mapy

Lidská kůže je složena z několika vrstev tkáně, jejíž barva se liší v závislosti na množství přítomných pigmentů, především melaninu, a krevních buněk. Přes tyto odlišnosti však stále odpovídá omezenému rozsahu hodnot odstínu. Pro segmentaci pleti je tedy výhodné použít barevný model, který odděluje jasové hodnoty od barevných složek, například *HSV* nebo *YCbCr* [21]. Použití modelu *YCbCr* v kombinaci s prahováním a dodatečnými úpravami se v rámci této práce ukázalo být efektivní. Ve dvou případech byla do vytvořené masky zahrnuta také oblast vlasů, které byly u těchto osob obarveny na světle nebo do odstínů s nádechem červené. Tento problém by se dal řešit použitím pokročilejší metody pro volbu prahu, na další průběh algoritmu však neměl podstatný vliv. Ukázky segmentací oblastí s odstínem pleti s úpravou masky, aby zahrnovala také oblasti obličejových příznaků, jsou v příloze B.

Pro tvorbu disparitní mapy byla použita metoda *Semi-global Matching* [22]. Vytvořené mapy byly dodatečně upraveny detekcí a náhradou nespojitostí. Tak byly úspěšně vyhlazeny. Na základě nalezených disparit byly body vykresleny do prostorových souřadnic s přiřazením jejich původní barvy. Pozice jednotlivých bodů této vizualizace odpovídají jejich skutečným pozicím, zobrazené obličeje nejsou zkreslené a obsahují jen minimální množství nespojitostí, viz B.

6.1.2 Detekce obličejových příznaků

Detekce probíhala vždy v prvním snímku video sekvence, kde se předpokládal přibližně čelní pohled vzhledem ke kamerovému systému. Přesnosti detekce zorniček, vnitřních a vnějších koutků očí a koutků úst u 11 zúčastněných osob jsou shrnuty v tab. 6.1. Výsledky algoritmu byly porovnávány s ručně označenými pozicemi. Protože se detekované příznaky v obrazech s použitým rozlišením nezobrazí jako konkrétní body, ale spíše oblasti, je zavedena tolerance pro označení bodu, které se považuje za přesné. Pro detekci středu zorniček bylo zvoleno okolí o poloměru 3 px , pro oční koutky 4 px a pro koutky úst 6 px . Ukázky detekovaných příznaků jsou k nahlédnutí v příloze B.

Použití kaskádového objektového detektoru

Na základě testování lze usoudit, že kaskádový detektor implementovaný v Matlabu je dobře naučený na detekci oblasti obličeje. Tato detekce byla úspěšná ve všech případech. Problematictější bylo jeho použití k detekci jednotlivých obličejových částí. Ačkoli byla oblast hledání vhodně upravena, detektor často označil hned několik zájmových oblastí, ze kterých bylo nutné dodatečnými postupy vybrat ty požadované. Také se ukázalo, že k hledání oblasti očí je vhodnější použít detektor páru očí a ty od sebe následně oddělit než detekovat každé oko zvlášť.

Detekce zorníček

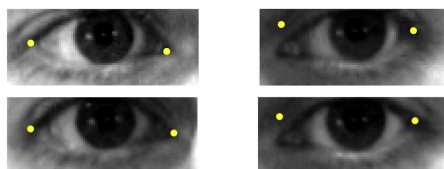
Pro detekci zorníček byla zvolena Houghova transformace, pomocí níž je segmentována kružnice ohraničující duhovku, se středem ve středu zorničky. Jedná se o standardně používanou metodu [18], [23]. Bylo účinné detekovat v oblasti oka raději několik kružnic a z nich následně postupem popsáním v kap. 5.3.1 vybrat tu nejvhodnější. Z barevné reprezentace obrazu byl použit pouze R kanál, který je nejméně citlivý na změny světelných podmínek. To umožňuje v pracích, které se zabývají detekcí a sledováním očí za účelem stanovení směru pohledu, s výhodou použít IR spektrum (z angl. *Infrared*) [24].

Zorničky byly v rámci testování algoritmu detekovány u všech osob s velkou přesností, jak je uvedeno v tab. 6.1. U 3 osob detekovaná kružnice netvořila přesný obrys duhovky, čímž došlo k odchylce od středu zorničky nejvýše o 4 *px*.

Detekce vnitřních a vnějších očních koutků

Zobrazení očních koutků je v reálných podmínkách výrazně ovlivněno stíny, které vrhají nadočnicové oblouky a nos. Pro záznamy sekvencí za účelem jejich detekce by proto bylo vhodné použít osvětlení z více směrů. Při snímání však byly použity pouze stropní zářivky. Proto jsou oblasti hledání upraveny ekvalizací histogramu.

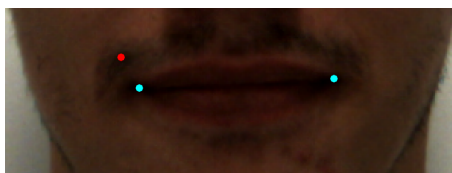
Při testování detekce obličejových příznaků bylo 82 % vnitřních a 70 % vnějších koutků označeno přesně. Menší přesnost detekce vnějších koutků je způsobena přítomností řas, které se mohou objevit také v hranové reprezentaci. K chybné detekci docházelo, pokud byly obrysy očí málo kontrastní vůči pleti, nebo byly oči naopak výrazně nalíčené. V obou případech se často odchylka detekce projevila podobně v obou pohledech, čímž byly zachovány alespoň přibližné korespondence detekovaného bodu, viz 6.1.



Obr. 6.1: Ukázka nepřesné detekce očních koutků. U vnitřního koutku pravého oka došlo v levém (nahore) i pravém (dole) pohledu k podobným odchylkám.

Detekce koutků úst

Pro nalezení koutků úst se použití metod hranové detekce ukázalo být dostačující. Pro určení lepších korespondencí mezi snímky obou pohledů byla použita vzájemná korelace, která významně přispěla ke zpřesnění pozic koutků, viz 6.2. Díky tomu byly koutky úst detekovány se 100% úspěšností, jak je doloženo v tab. 6.1.



Obr. 6.2: Příklad přispění vzájemné korelace ke zpřesnění detekce koutků úst. Červeně je označena původně detekovaná pozice, která byla v tomto případě ovlivněna přítomností vousů, modře upravená pozice.

Tab. 6.1: Úspěšnost detekce obličejových příznaků.

Detekovaný bod	Přesná detekce [%]	Prům. chyba [px]	Sm. odchylka [px]
Zorničky	88,63	2,80	0,75
Vnitřní koutky očí	81,81	6,25	4,68
Vnější koutky očí	70,45	4,63	2,74
Koutky úst	100	0	0

6.1.3 Trackování

Metody, které využívají sledování konkrétních bodů, jsou v modelování pohybu omezeny jejich viditelností. Při sledování obličejových příznaků je rozsah měřitelných úhlů natočení omezen okluzemi očních koutků, vznikajícími ve stínu nosu, a odvrácením koutků úst od pozorovatele. Geometrické metody většinou předpokládají,

Tab. 6.2: Testování trackovacího algoritmu.

Počet hodnocených párů video sekvencí	20
Průměrná délka sekvence [počet snímků]	74
Celkový počet hodnocených snímků	2956
Počet párů sekvencí s úspěšným trackováním	7
Počet snímků s chybně detekovanými body	192
Poměr snímků s chybně detekovanými body [%]	6,43
Průměrný počet chybně detekovaných bodů	1,82

že se hlava chová jako tuhé těleso. Vzájemné postavení detekovaných příznaků se však může výrazně měnit s výrazem tváře. S mimikou se mění i samotný vzhled příznaků. Přesnost detekce je ovlivněna také rychlostí pohybu. Zvláště při použití nízké snímkovací frekvence může docházet k rozmazání obrazu.

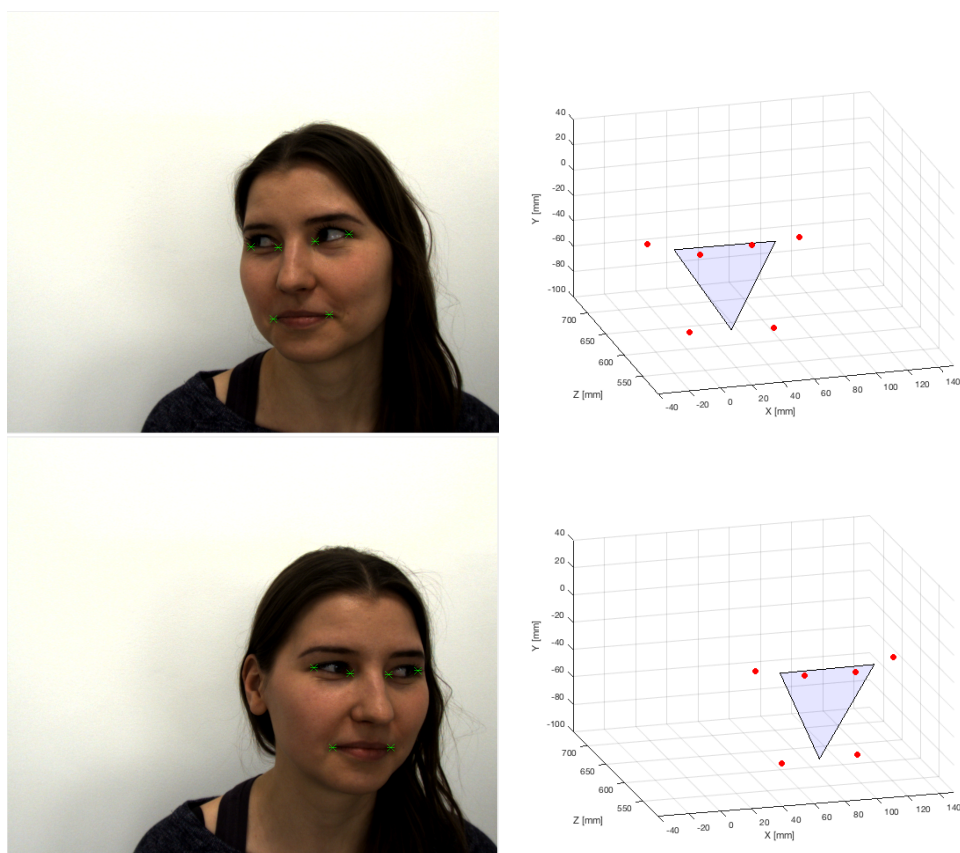
Na základě uvážení výše uvedených faktorů bylo pro hodnocení trackovacího algoritmu vybráno 20 video sekvencí. Sekvence obsahují všechny 3 druhy sledovaného pohybu. Jejich průměrná délka je 74 snímků.

Celkem bylo zhodnoceno 2965 snímků. U 7 sekvencí bylo trackování vyhodnoceno jako úspěšné v celém průběhu v obou pohledech. Při odchýlení některého z bodů z jeho správné pozice byl snímek započítán jako chybný. Počet snímků s alespoň jedním chybně detekovaným bodem tvořil 6,43 % všech kontrolovaných snímků. V průměru na nich bylo nesprávně detekováno 1,82 bodu. Záznamy o testování trackovacího algoritmu jsou shrnuty v tab. 6.2.

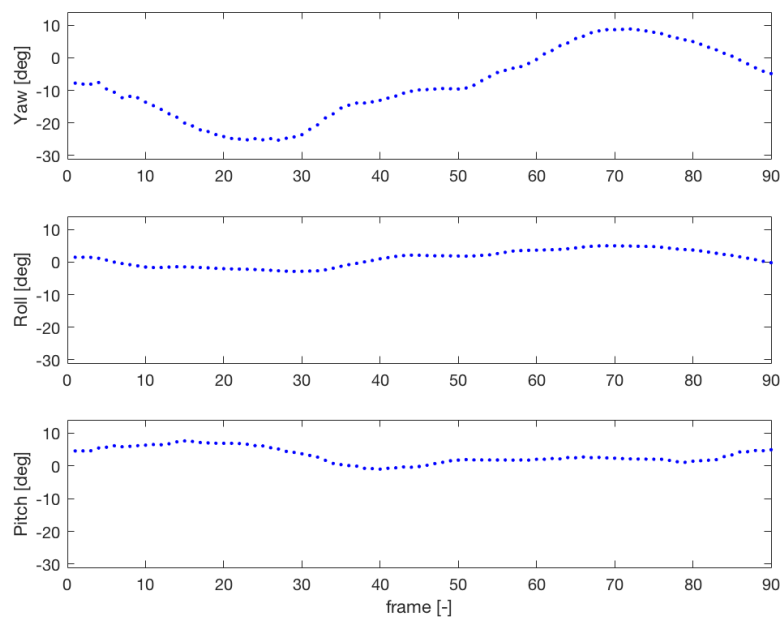
Ukázalo se, že s využitím aplikace geometrické transformace na nespolehlivě detekované body lze napravit krátkodobé nesrovnalosti, způsobené například mrknutím, viz 5.13. Problémem však stále zůstává okluze bodů. Algoritmus sice dokáže body pozastavit na místě, kde příznak zmizel z obrazu, nebo jej posunout ve směru pohybu okolních bodů, výpočet polohy je však v tomto okamžiku zatížen chybou. Také může nastat situace, že se i po návratu sledovaného bodu do obrazu začne trackovat bod jiný. Od té chvíle dochází ke kumulaci chyby, která nemusí být zaznamenána. V [25] je tento problém řešen uložením počátečního okolí, vzoru, sledovaného bodu do paměti. Okolí trackovaného bodu je potom v každém kroku porovnáváno se vzorem pomocí vzájemné korelace. V případě poklesu korelačního koeficientu pod zvolený práh je provedena nová detekce a inicializace trackeru. Tato metoda může pomoci vyrovnat se například se změnami mimiky tváře, reinitializace bodů v případě jiného, než čelního postavení hlavy, však může být problematická a nepřesná. S určováním polohy hlavy v rámci větších rozsahů pohybu se lépe vypořádávají metody založené například na fitování pružných modelů nebo metody využívající tvaru hlavy apod.

6.1.4 Modelování polohy hlavy

Spolehlivost hodnot vypočítaných úhlů sklonu tváře v jednotlivých osách úzce souvisí s přesností trackování detekovaných bodů. Při hodnocení tedy bylo přihlíženo k tomu, jestli je trackování v příslušných snímcích úspěšné. Rozmezí úhlů, ve kterých bylo možné určit sklon, aniž by byl model deformován chybnou detekcí bodů, bylo 73° při otočení hlavou zleva doprava, 80° při náklonu hlavy směrem k ramenům a 52° při úklonu dopředu a dozadu. Příklady grafů průběhů hodnot jednotlivých úhlů jsou k nahlédnutí v příloze B. U grafů se dá také sledovat plynulost průběhů úhlů vzhledem k pohybu na videu nebo zda si číselně odpovídají hodnoty úhlů, když se pozorovaná osoba otočí zpět do výchozí pozice.



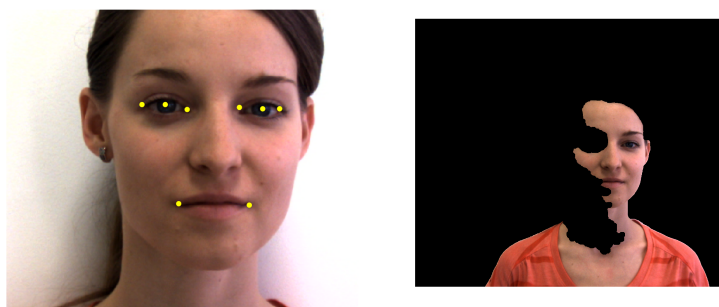
Obr. 6.3: Ukázka průběhu modelování polohy hlavy. Nahoře je snímek 24 z levé kamery s označením trackovaných bodů a příslušný model. Vypočítané aktuální úhly sklonu byly $yaw = -24,82^\circ$, $roll = -2,31^\circ$, $pitch = 6,15^\circ$. Dole je obdobně snímek 72 s hodnotami úhlů $yaw = 8,81^\circ$, $roll = 4,91^\circ$, $pitch = 2,01^\circ$. Celý průběh hodnot vypočítaných úhlů pro danou sekvenci je na obr. 6.4.



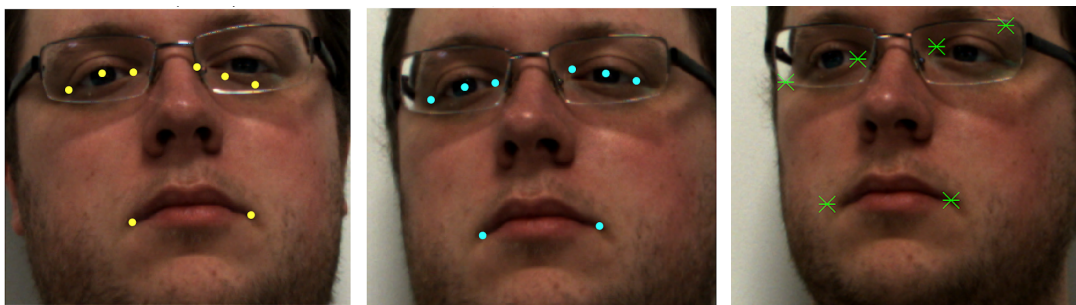
Obr. 6.4: Průběh vypočítaných hodnot úhlů příslušné video sekvence. Hlavní pohyb byl otočení hlavou doprava a doleva.

6.2 Testování při ztížených podmínkách

Algoritmus byl testován také na video záznamech osob s dioptrickými brýlemi a záznamech s dodatečným bočním osvětlením. Nerovnoměrné osvětlení mělo vliv především na segmentaci oblasti pleti, jak je ukázáno na obr. 6.5. Detekce obličejových příznaků, stejně jako trackování, bylo úspěšné. Hledání příznaků naopak selhalo v případě použití dioptrických brýlí, kdy místo zorníček a očních koutků byly detekovány obroučky, viz obr. 6.6.



Obr. 6.5: Vliv nerovnoměrného osvětlení na detekci příznaků a segmentaci pleti.



Obr. 6.6: Testování s brýlemi. V pravém pohledu (uprostřed) došlo ke správné detekci příznaků, které však byly v průběhu trackování zaměněny s obroučkami brýlí, viz snímek 54 trackování (vpravo). V levém pohledu (vlevo) byla detekce chybná už od počátku.

6.3 Úskalí geometrických metod

Geometrické metody určování polohy hlavy nejvíce odpovídají lidskému vnímání. Vycházejí z detekce obličejových příznaků a vyhodnocení jejich vzájemného rozestavení, odchylek od bilaterální symetrie apod. Mohou také uvažovat umístění příznaků v rámci oblasti hlavy v obraze nebo její tvar. Jejich výhodou je, že jsou rychlé a jednoduché, k určení přibližné polohy stačí detekovat jen několik bodů v obraze. Často detekovanými příznaky jsou vnitřní a vnější koutky očí, koutky úst nebo špička nosu. Přesná detekce příznaků je však zároveň největším problémem. Vzhled jednotlivých příznaků je v populaci velmi variabilní. Oči a ústa se liší ve tvaru, barvě nebo kontrastu vůči pleti, která také nabývá různých odstínů. Vzhled celé tváře lze také výrazně ovlivnit použitím make-upu. Algoritmy se dále musí vypořádat s celou řadou faktorů, mezi které patří:

- Pozice obličeje. Vzhled příznaků se s natáčením obličeje vzhledem ke kameře výrazně mění. Rychlé pohyby navíc způsobují rozmazání obrazu.
- Přítomnost nebo absence různých součástí, jako jsou vousy, brýle, vlasy částečně překrývající tvář.
- Výraz tváře, který přímo ovlivňuje vzhled jednotlivých příznaků. Změnou výrazu se zároveň mění jejich vzájemné rozestavení, což může mít vliv také na výpočet polohy hlavy.
- Okluze. I pokud neuvažujeme možnost zakrývání obličeje rukami nebo jinými předměty, mohou být některé části obličeje při pohybu zakryty jinými, například nosem nebo vlasy.
- Vnější podmínky. Na vzhled tváře a viditelnost příznaků má přímý vliv intenzita a druh osvětlení, dále například vlastnosti kamery.

Geometrické metody jsou tedy vhodné k použití pro detekci v blízkém poli, při použití obrazů s dobrým rozlišením, nasnímaných za vhodných světelných podmínek. Právě kvůli detekci malého množství bodů jsou mnohem více citlivé na okluze než například metody využívající předlohy, které pracují s celou obličejovou oblastí [28].

6.4 Výpočetní nároky

Tab. 6.3 uvádí průměrné výpočetní časy na snímek testované sekvence pro jednotlivé části algoritmu. Uvedené časy nezahrnují vykreslování obrázků ani grafů.

Tab. 6.3: Průměrné výpočetní časy.

Úkon	Výpočetní čas [s/snímek]
Rektifikace	0,11
Tvorba disparitní mapy vč. úprav	5,81
Detekce příznaků	6,39
Trackování	0,61

Práce byla vytvářena na notebooku s dvoujádrovým procesorem Intel Core i5 (2,6 GHz) a pamětí 8 GB.

7 ZÁVĚR

V rámci této diplomové práce byla prostudována a popsána tematika týkající se geometrie kamery a jejích parametrů. Také byly uvedeny teoretické podklady pro využití dvou pohledů k vytvoření snímků 3D scény pomocí stereoskopické rekonstrukce. Hlavním cílem práce bylo navrhnout algoritmus detekce a modelování polohy hlavy ve video sekvencích dvou pohledů.

V první části byl sestaven systém dvou kamer vhodný pro stereoskopická snímání pohybu hlavy osoby v blízkém poli. Při návrhu vhodného systému je třeba uvážit několik faktorů. Konfigurace s úzkou bází kamer usnadňuje automatické vyhledávání korespondencí bodů, zato při použití široké báze vzniká velké množství okluzních bodů. 3D data však mohou být rekonstruována s mnohem větší přesností. Použitá konfigurace je znázorněna na obr. A.1 a její parametry na obr. A.2 přílohy. Parametry systému byly získány kalibrací pomocí snímků kalibrační šablony.

Na základě kalibračních parametrů byla získaná data nejprve upravena rektifikací a odstraněním zkreslení zavedeným optickým systémem kamer. Pro první snímky sekvencí byly vytvořeny disparitní mapy s využitím metody *Semi-Global Block Matching* a 3D vizualizace tímto způsobem nalezených bodových korespondencí. Zájmová oblast byla vysegmentována na základě odstínu pleti s převodem obrazů do reprezentace YCbCr.

Následovala detekce obličejových příznaků – vnitřních a vnějších koutků očí a koutků úst. Detekce dosahovala u všech osob na testovacích sekvencích poměrně přesných výsledků s největší průměrnou chybou $6,25\text{ px}$ pro vnitřní koutky očí, přičemž velikost obličejové oblasti odpovídala velikosti okolo $350 \times 320\text{ px}$. Nalezením korespondencí těchto bodů bylo možné vypočítat jejich pozice v prostoru. Body byl následně proložen trojúhelníkový model představující obličejovou rovinu. Sklon a umístění tohoto modelu odpovídá aktuální poloze hlavy a je hodnocen ve třech stupních volnosti.

Pohyb je modelován sledováním detekovaných bodů v průběhu sekvencí. Použitý trackovací algoritmus vykazoval dobrou úspěšnost na sekvencích s menším rozsahem pohybu, kdy byly sledované body viditelné na převážné většině snímků. Pohyby byly úspěšně detekovány v rozmezí 73° při otočení hlavou zleva doprava, 80° při náklonu hlavy směrem k ramenům a 52° při úklonu dopředu a dozadu.

Výhodou použití geometrických metod je jejich poměrně snadná implementace a rychlost výpočtů. K určení polohy stačí detekovat v obrazech jen několik bodů. Přesnost detekce je však omezena na použití sledování osob v blízkém poli kamery, v obrazech s dobrým rozlišením a při vhodných světelných podmínkách. Podstatnou roli hraje také variabilita obličejových příznaků populace. Jejich vzhled se zároveň výrazně mění s mimikou nebo natočením tváře. Přílišným odvrácením tváře od ka-

mery nebo zakrytím některých částí jinými, například očních koutků kořenem nosu, dochází k chybné detekci nebo její ztrátě. Geometrické metody je tedy výhodné použít k hodnocení pohybů o menších rozsazích a v sekvencích, které umožňují opětovnou reinicializaci detekovaných bodů, ideálně v čelní pozici.

LITERATURA

- [1] HOFFKEN, M., E. TARAYAN, Ulrich KRESEL a K. DIETMAYER. *Stereo vision-based driver head pose estimation*. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings [online]. IEEE, 2014, s. 253-260 [cit. 2017-11-30]. DOI: 10.1109/IVS.2014.6856578. ISBN 978-1-4799-3638-0. Dostupné z: <http://ieeexplore.ieee.org/document/6856578/>.
- [2] FERNÁNDEZ, A., R. USAMENTIAGA, J. CARÚS a R. CASADO. *Driver Distraction Using Visual-Based Sensors and Algorithms*. Sensors [online]. 2016, 16(11), 1805- [cit. 2018-04-16]. DOI: 10.3390/s16111805. ISSN 1424-8220. Dostupné z: <http://www.mdpi.com/1424-8220/16/11/1805>.
- [3] HARTLEY, R. a A. ZISSERMAN. *Multiple view geometry in computer vision*. 2nd ed. New York: Cambridge University Press, 2003. ISBN 0-521-54051-8.
- [4] ZHANG, Z. *A flexible new technique for camera calibration*. IEEE Transactions on Pattern Analysis and Machine Intelligence [online]. 2000, 22(11), 1330-1334 [cit. 2017-11-30]. DOI: 10.1109/34.888718. ISSN 01628828. Dostupné z: <http://ieeexplore.ieee.org/document/888718/>.
- [5] ZHAO, J., S. YU a H. CAI. *A precise method for correcting stereo pairs collected by a new binocular device*. Industrial Robot: An International Journal [online]. 2007, 34(6), 512-519 [cit. 2017-11-30]. DOI: 10.1108/01439910710832110. ISSN 0143-991x. Dostupné z: <http://www.emeraldinsight.com/doi/10.1108/01439910710832110>.
- [6] ŠONKA, M., V. HLAVÁČ a R. BOYLE. *Image processing, analysis, and machine vision*. 4th edition. United States of America: Cengage Learning, 2015. International Edition. ISBN 978-1-133-59369-0.
- [7] CYGANÉK, B. a J. P. SIEBERT. *An introduction to 3D computer vision techniques and algorithms*. Chichester, U.K.: J. Wiley, 2009. ISBN 978-0-470-01704-3.
- [8] SCHREER, O., P. KAUFF a T. SIKORA. *3D videocommunication: algorithms, concepts, and real-time systems in human centred communication*. Hoboken, NJ: Wiley, 2005. ISBN 978-0-470-02271-9.
- [9] KOMOROWSKI, J. a P. ROKITA. *Pose Invariant Face Recognition Method Using Stereo Vision*. Computer Recognition Systems 4 [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, s. 13-22 [cit. 2017-11-30]. Advances in

- Intelligent and Soft Computing. DOI: 10.1007/978-3-642-20320-6-2. ISBN 978-3-642-20319-0. Dostupné z:
 <<http://link.springer.com/10.1007/978-3-642-20320-6-2/>>.
- [10] NIXON, M. S. a A. S. AGUADO. *Feature extraction and image processing*. 2nd ed. London: Elsevier, 2008. ISBN 978-0-1237-2538-7.
- [11] MURPHY-CHUTORIAN, E. a M. M. TRIVEDI. *Head Pose Estimation in Computer Vision: A Survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence [online]. 2009, 31(4), 607-626 [cit. 2017-11-30]. DOI: 10.1109/TPAMI.2008.106. ISSN 0162-8828. Dostupné z:
 <<http://ieeexplore.ieee.org/document/4497208/>>.
- [12] NIKOLAIDIS, A., I. PITAS. *Facial feature extraction and pose determination*. Pattern Recognition, 2000, 33(11), 1783–1791.
- [13] WANG, J. G. a E. SUNG. EM enhancement of 3D head pose estimated by point at infinity. *Image and Vision Computing*. [online]. 2007, 25(12), 1864-1874 [cit. 2017-12-12]. DOI: 10.1016/j.imavis.2005.12.017. ISSN 02628856. Dostupné z:
 <<http://linkinghub.elsevier.com/retrieve/pii/S0262885606002848>>.
- [14] WU, H., J. INADA, T. SHIOYAMA, Q. CHEN a T. SIMADA. *Automatic facial feature points detection with SUSAN Operator*. In: Proceedings of Scandinavian Conference on Image Analysis. June 11–14, Bergen, Norway, 2001.
- [15] BOUGUET, J. Y.. *Camera calibration toolbox for Matlab*. [online]. California Institute of Technology, 2008. Poslední změna: 2015-10-14 [cit. 2017-11-30]. Dostupné z:
 <http://www.vision.caltech.edu/bouguetj/calib_doc/index.html>.
- [16] VIOLA, P. a M. JONES. *Rapid object detection using a boosted cascade of simple features*. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 [online]. IEEE Comput. Soc, 2001, I-511-I-518 [cit. 2018-04-13]. DOI: 10.1109/CVPR.2001.990517. ISBN 0-7695-1272-0. Dostupné z:
 <<http://ieeexplore.ieee.org/document/990517/>>.
- [17] DASGUPTA, A., A. MANDLOI, A. GEORGE a A. ROUTRAY. *An improved algorithm for eye corner detection*. In: 2016 International Conference on Signal Processing and Communications (SPCOM)[online]. IEEE, 2016, 2016, s. 1-4 [cit. 2018-04-14]. DOI: 10.1109/SPCOM.2016.7746627. ISBN 978-1-5090-1746-1. Dostupné z:
 <<http://ieeexplore.ieee.org/document/7746627/>>.

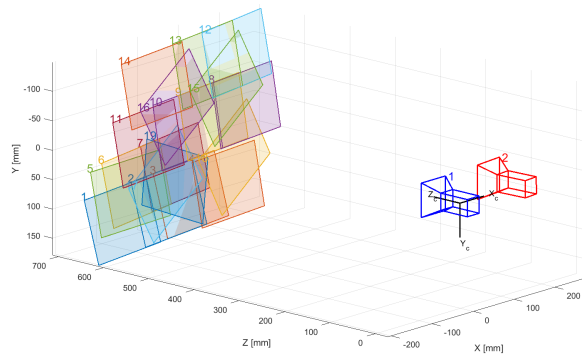
- [18] BERALDIN, J. A., N. ERDOGMUS, J. L. DUGELAY, et al. *Automatic extraction of facial interest points based on 2D and 3D data* [online]. In: . 2011-1-23, 786400- [cit. 2018-04-14]. DOI: 10.1117/12.876675. Dostupné z: <<http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.876675>>.
- [19] LUCAS, B. D. a T. KANADE. *An Iterative Image Registration Technique with an Application to Stereo Vision*. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence. IJCAI, William Kaufmann, 1981, s. 674–679.
- [20] JIANBO S. a C. TOMASI. *Good features to track*. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94. IEEE Comput. Soc. Press, 1994, 1994, s. 593-600. DOI: 10.1109/CVPR.1994.323794. ISBN 0-8186-5825-8. Dostupné z: <<http://ieeexplore.ieee.org/document/323794/>>.
- [21] HEMDAN, I., S. KARUNGARU a K. TERADA. *Video surveillance using facial features-based tracking*. International Journal of Innovative Computing, Information and Control. 2012, 8 (4), s. 2761-2776.
- [22] HIRSCHMULLER, H. *Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information*. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) [online]. IEEE, 2005, s. 807-814 [cit. 2018-04-12]. DOI: 10.1109/CVPR.2005.56. ISBN 0-7695-2372-2. Dostupné z: <<http://ieeexplore.ieee.org/document/1467526/>>.
- [23] TURAN, B. a H. I. ESKİKURT. *Iris and eye corner detection by processing internal webcam images*. Turkish Journal of Electrical Engineering & Computer Sciences. 2016, 24, 3036-3048. DOI: 10.3906/elk-1401-96. ISSN 13000632. Dostupné z: <<http://online.journals.tubitak.gov.tr/openDoiPdf.htm?mKodu=elk-1401-96>>.
- [24] LIM S., D. LEE. *Real-Time Eye Tracking Using IR Stereo Camera for Indoor and Outdoor Environments*. KSII Transactions on Internet and Information Systems [online]. 2017, 11(8), - [cit. 2018-04-15]. DOI: 10.3837/tiis.2017.08.012. ISSN 19767277. Dostupné z: <<http://itiis.org/digital-library/manuscript/1770>>.

- [25] CHUN, J. a W. KIM. *3D face pose estimation by a robust real time tracking of facial features*. Multimedia Tools and Applications [online]. 2016, 75(23), 15693-15708 [cit. 2018-04-16]. DOI: 10.1007/s11042-014-2356-9. ISSN 1380-7501. Dostupné z:
<<http://link.springer.com/10.1007/s11042-014-2356-9>>.
- [26] ZHAO, W.a R.CHELLAPPA. *Face processing: advanced modeling and methods*. Boston: Elsevier, 2006. ISBN 978-0-12-088452-0.
- [27] MAINO, G. a G. L. FORESTI. *Image analysis and processing–ICIAP 2011: 16th international conference, Ravenna, Italy, September 14-16, 2011 : proceedings*. New York: Springer, 2011. Lecture notes in computer science, Part I, LNCS 6978, s. 227–236. ISBN 3642240879.
- [28] ZHENG, Y., J. CHANG, Z. ZHENG a Z. WANG. *3D Face Reconstruction from Stereo: A Model Based Approach*. In: 2007 IEEE International Conference on Image Processing [online]. IEEE, 2007, III - 65-III - 68 [cit. 2017-11-30]. DOI: 10.1109/ICIP.2007.4379247. ISBN 978-1-4244-1436-9. Dostupné z:
<<http://ieeexplore.ieee.org/document/4379247/>>.

SEZNAM PŘÍLOH

A	Návrh konfigurace kamer	61
B	Ukázky výsledků algoritmu	62
C	Obsah přiloženého DVD	67

A NÁVRH KONFIGURACE KAMER



Obr. A.1: Konfigurace systému kamer. Modře je znázorněna referenční (levá) kamera 1 a červeně kamera 2. Barevně jsou vykresleny jednotlivé plochy, které pokrývala kalibrační šablona při kalibraci kamer.

```
Parameters of Two Cameras:
-----
Camera 1 Intrinsics
  IntrinsicMatrix: [1.488e+03,0,0;-1.909,1.488e+03,0;6.162e+02,4.979e+02,1]
  FocalLength: [1.4883e+03,1.4883e+03]
  PrincipalPoint: [616.2461,497.9101]
  Skew: -1.9095
  RadialDistortion: [-0.4054,0.2203,0.1034]
  TangentialDistortion: [-0.0037,-0.0034]
  ImageSize: [960,1280]

Camera 2 Intrinsics
  IntrinsicMatrix: [1.458e+03,0,0;-2.669,1.456e+03,0;7.126e+02,4.801e+02,1]
  FocalLength: [1.4583e+03,1.4556e+03]
  PrincipalPoint: [712.5582,480.0961]
  Skew: -2.6690
  RadialDistortion: [-0.4165,0.2722,-0.0861]
  TangentialDistortion: [-0.0028,2.5117e-04]
  ImageSize: [960,1280]

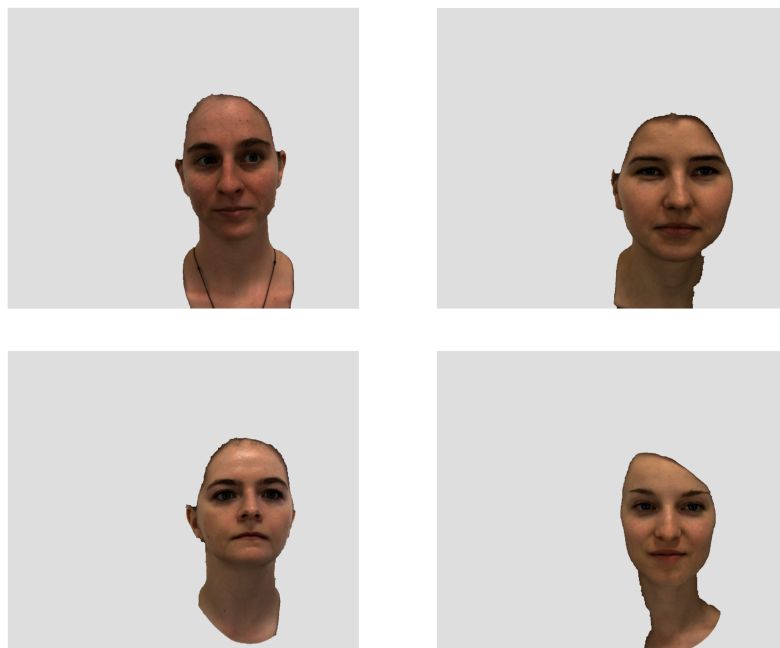
Inter-camera Geometry
  RotationOfCamera2: [0.997,0.012,-0.073;-0.012,0.999,-0.003;0.0729,0.004,0.997]
  TranslationOfCamera2: [-143.0714,0.0258,6.3955]
  FundamentalMatrix: [-3.566e-08,-2.947e-06,0.001;-1.873e-06,-2.268e-07,0.099;-2.278e-04,-0.094,-0.839]
  EssentialMatrix: [-0.077,-6.395,0.002;-4.057,-0.474,1.432e+02;-1.715,-1.431e+02,-0.524]

Accuracy of Estimation
  MeanReprojectionError: 0.2295

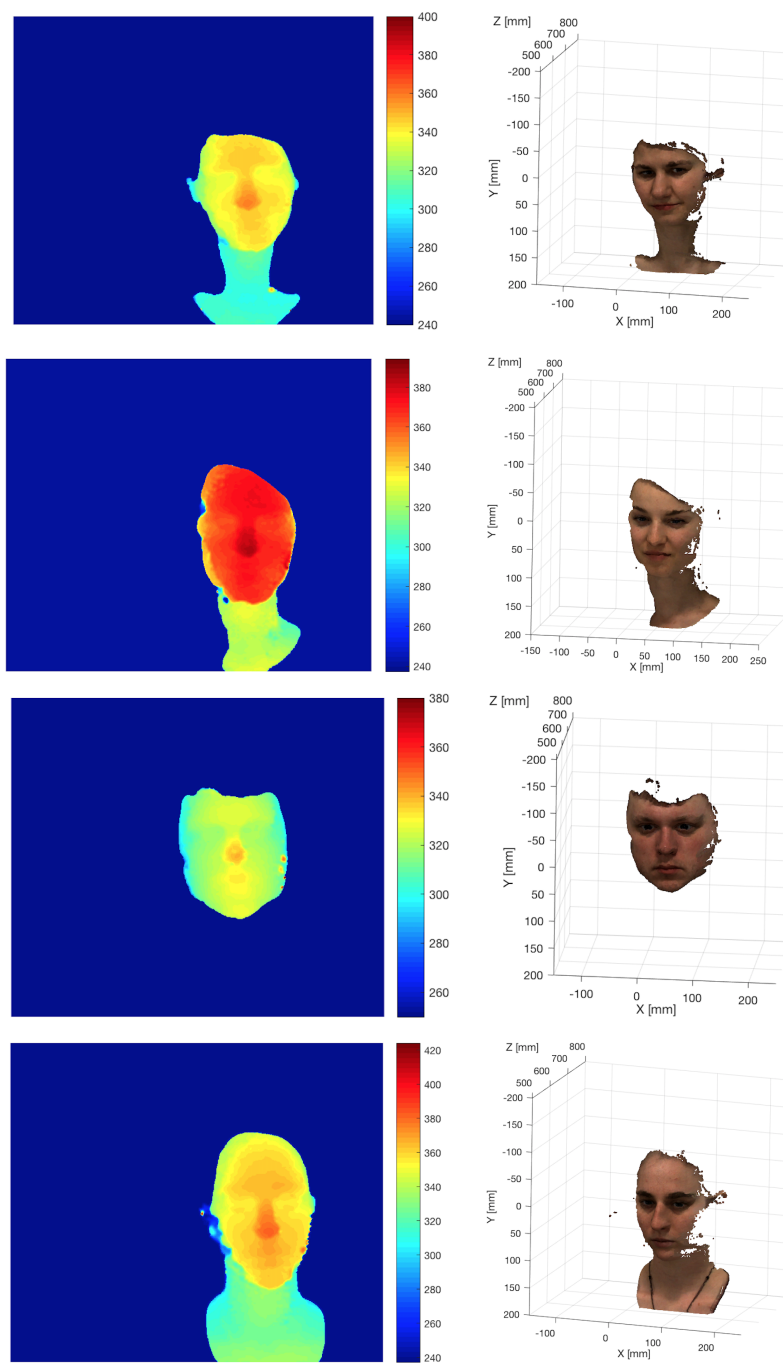
Calibration Settings
  NumPatterns: 19
  WorldPoints: [54x2 double]
  WorldUnits: 'millimeters'
```

Obr. A.2: Vnitřní parametry kamer a jejich vzájemné postavení určené kalibrací v Matlabu.

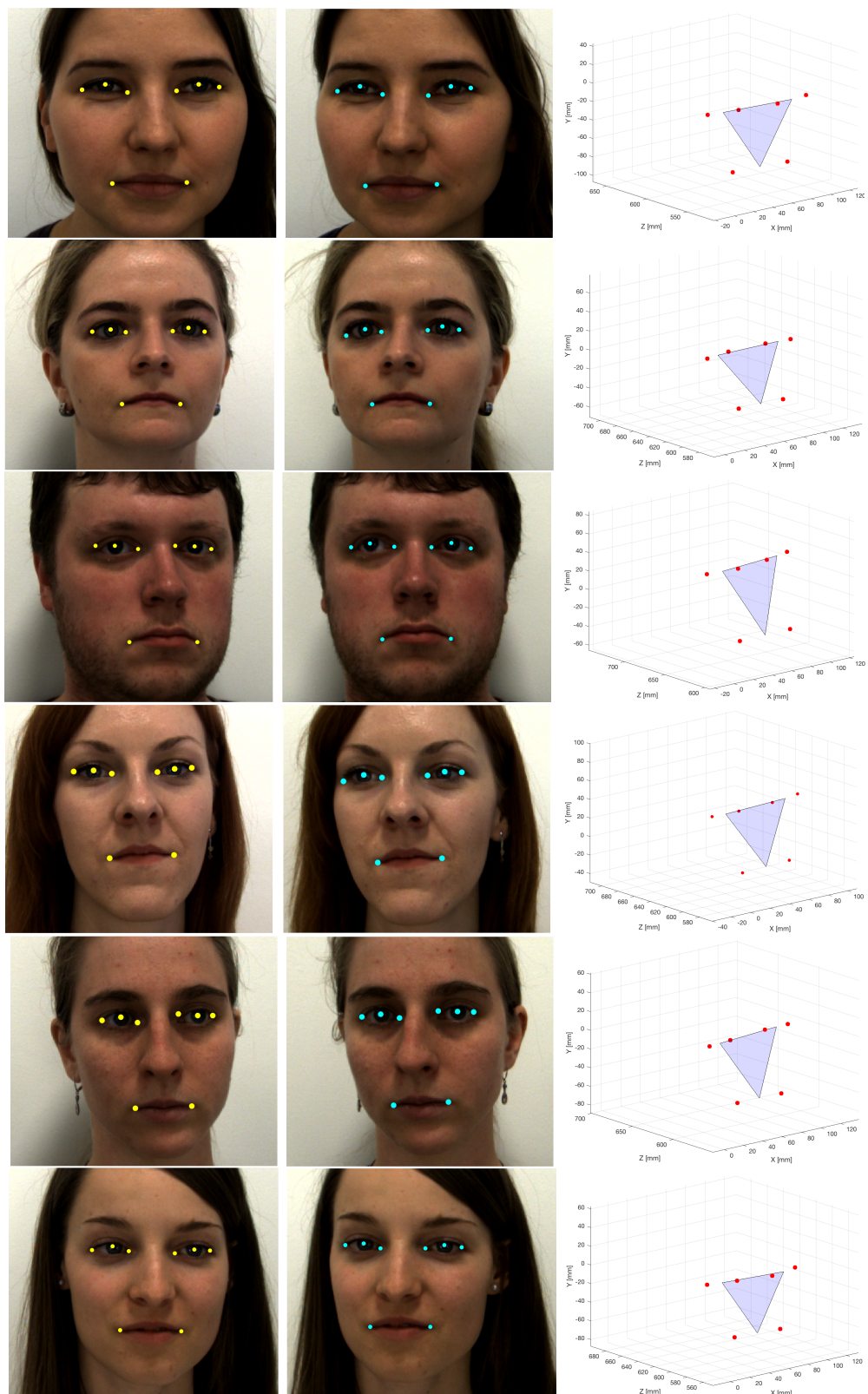
B UKÁZKY VÝSLEDKŮ ALGORITMU

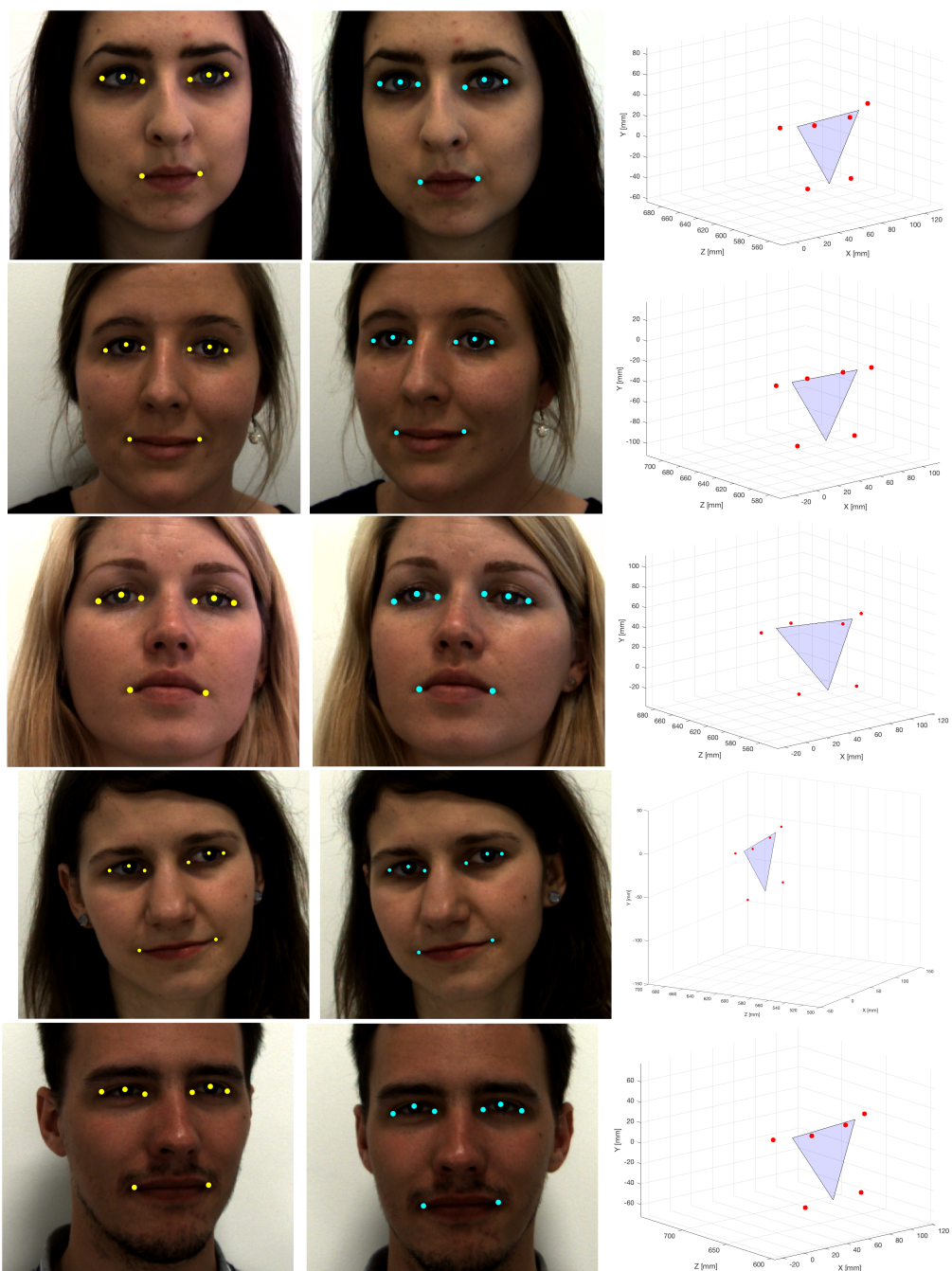


Obr. B.1: Ukázka výsledků segmentace uzavřené oblasti o pleťovém odstínu.

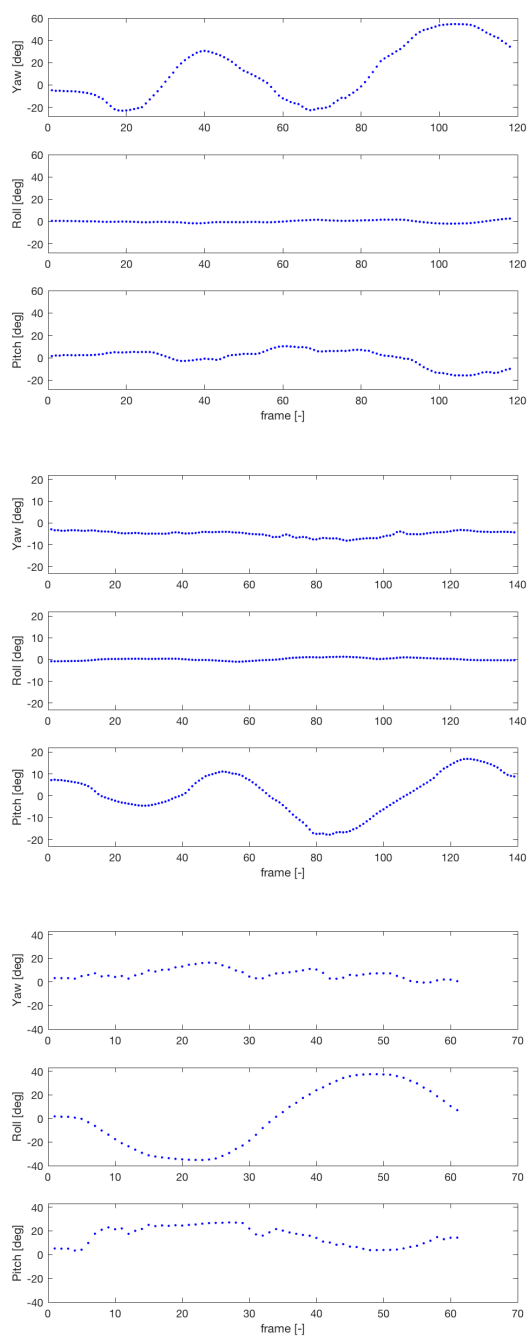


Obr. B.2: Ukázka vytvořených disparitních map a 3D vizualizací.





Obr. B.3: Příklady detekce obličejových příznaků v levém pohledu (levý sloupec) a pravém pohledu (pravý sloupec) a příslušné trojúhelníkové modely.



Obr. B.4: Příklady grafů průběhů úhlů sklonu ve třech hodnocených stupních volnosti.

C OBSAH PŘÍLOŽENÉHO DVD

Na přiložené DVD je vložena elektronická verze této práce *veronika_hribkova_DP.pdf* jsou nahrány složky *calibration* a *data* a použité zdrojové kódy.

Složka *calibration* obsahuje snímky použité pro kalibraci kamer, skript `calibrate.m`, který na jejich základě provede kalibraci, a uložené výsledky kalibrace: proměnné `stereoParams.mat` a `estimationErrors.mat`. Do složky *data* jsou vloženy ukázky použitých video sekvencí.

Seznam příložených funkcí:

- `main.m`
- `count3Dcoord.m`
- `detectEyeCorners.m`
- `detectFaceFeatures.m`
- `detectMouthCorners.m`
- `detectPupils.m`
- `disparityAndPointCloud.m`
- `euclDist2.m`
- `euclDist3.m`
- `findCircles.m`
- `findCorrespondingMouthCorners.m`
- `findMouth.m`
- `findNose.m`
- `inpaint_nans.m`
- `modelAngles.m`
- `nearestPoints.m`
- `pointTracking.m`
- `rectifySequences.m`
- `skinMask.m`
- `visualization.m`

Hlavní algoritmus obsahuje funkce `main.m`.

Dále je přiložen soubor `README.txt` obsahující instrukce k použití metody.

Práce byla vytvořena v programovém prostředí Matlab, verzi R2016b.