

**Semestrálny projekt**

**Detekcia objektov v obraze zo sférickej kamery**

Michal Falát, FAL0045

Akademický rok 2018/2019

**Obsah**

[1. Úvod 3](#_Toc8512724)

[2. Motivácia 3](#_Toc8512725)

[3. Použitie 3](#_Toc8512726)

[4. Existujúce riešenia 4](#_Toc8512727)

[5. Jednoduchá detekcia objektov objektov 4](#_Toc8512728)

[5.1. Detekcia pohybu 4](#_Toc8512729)

[5.2. Detekcia ľudí 5](#_Toc8512730)

[5.3. Detekcia tváre 7](#_Toc8512731)

[5.4. Detekcia osôb v interiéri vozidla 7](#_Toc8512732)

[6. Pokročilá detekcia tváre pomocou landmarkov 9](#_Toc8512733)

[7. Možnosti vylepšenia detekcie 11](#_Toc8512734)

[8. Zhrnutie 12](#_Toc8512735)

[9. Bibliografia 13](#_Toc8512736)

# **Úvod**

V tomto dokumente budem rozoberať problematiku analýzy videa a detekcie objektov zo sférickej (360- stupňovej) kamery. Počas práce som mal k dispozícii kameru *Ricoh Theta V*. Bližšie parametre kamery sú umiestnené v *tabuľke 1*. Táto práca sa zaoberá formátom snímok zachytenej touto kamerou, ich spracovanie, porovnanie detekcie objektov pri použití rôznych rozlíšení a svetelných podmienok. V práci tiež sú zhrnuté ďalšie možnosti vylepšenia detekcie a porovnanie oproti klasickým kamerám.

|  |  |
| --- | --- |
| Senzor | FishEye CMOS 2×12MPix |
| Maximálne rozlíšenie (Video) | 4K / 30fps |
| Maximálne rozlíšenie (Fotografia) | 14.5MP (5376×2688px) |
| Svetelnosť | f/2.0 |
| Vnútorná pamäť | 19GB |

*Tabuľka 1: Parametre kamery Ricoh Theta V*

# **Motivácia**

Túto problematiku som sa rozhodol riešiť najmä kvôli tomu, že popularita detekcie objektov v obraze rastie a ma veľký potenciál do budúcnosti. Analýza obrazu má veľké využitie napríklad v priemysle, kde dokáže slúžiť na kontrolu kvality výroby a s rastúcou popularitou IoT 4.0 sa stále nachádzajú nové možnosti využitia a ich zdokonaľovanie na čo najvyššiu úspešnosť. Táto oblasť ma zaujíma a preto som sa rozhodol pre túto tému. V súčasnosti som taktiež nenašiel veľa ukážok na spracovanie videa zo sférickej kamery a preto by som sa chcel zamerať hlavne na túto oblasť.

# **Použitie**

Použitie detekcie objektov v obraze má prakticky nekonečné využitie. V súčasnosti sa používa v priemysle, automobilovej oblasti, medicíne, logistike a podobne. Medzi hlavné použitie patrí najmä vizuálna kontrola kvality, rozpoznávanie objektov alebo defektov, klasifikácia objektov. Vieme ho však čoraz viac nájsť aj na bežných zariadeniach používaných denne napríklad odomykanie telefónu pomocou tváre, parkovacie asistenty, automatická úprava fotiek a podobne. Použitie sférickej kamery v podobnej oblasti je zatial veľmi zriedkavé. Na väčšinu úloh postačuje aj použitie snímky zachytenej klasickou kamerou. Mnoho riešení sa zameriava aj na rozpoznávanie chodcov, značiek a čiar v premávke. V tejto oblasti je veľmi efektívne použitie sférickej kamery, pretože sníma priestor v celom okolí a nie je potrebné použitie viacerých kamier. Väčšina moderných sférických kamier dokáže nahrávať video v 4K rozlíšení, čo plne postačuje na zachytenie objektov. Veľké rozdiely nastávajú pri znížených svetelných podmienkach, alebo pri prudších pohyboch kamery. Kamera Ricoh Theta V zvláda aj mierne znížené svetelné podmienky, napríklad západ slnka, interiér vozidla, interiér izby a podobne. Obraz za tmy a v noci je však nepoužiteľný na akúkoľvek aplikáciu a detekciu. Kamera zároveň nemá žiadny druh stabilizácie oproti ostatným sférickým kamerám, takže je obraz najmä pri chôdzi často roztrasený rozmazaný. Výstup z kamery je reprezentovaný 2D snímkou, zachytávajúci 3D obraz. Medzi najznámejšie projekcie patrí *equirectangular panorama projection*, ktorá je definovaná dvoma uhlami: *latitude* ϕ ∈ [−90◦ , +90◦ ] a *longitude* λ ∈ [−180◦ , +180◦ ]. V tejto práci sa budem zameriavať na detekciu objektov na snímkach s *equirectangular* projekciou obrazu*.*

# **Existujúce riešenia**

V súčasnosti je používanie sférických kamier nové odvetvie, ktoré sa pomaly rozrastá hlavne v priemysle. Medzi najväčšie využitie môžeme zaradiť službu Google Street View, ktorá funguje od roku 2007. Služba je v súčasnosti dostupná v takmer 200 krajinách. Hlavnou podstatou tejto služby je snímanie krajín sférickou kamerou a tieto zábery nechať voľne dostupné verejnosti. Užívateľ má dojem, priestorového pohybu po uliciach, rôznych múzeí, katedrál a iných záujmových miest. Aby služba spĺňala všetky legislatívne zákony a zákony GDPR nariadené Európskou úniou, museli jej tvorcovia myslieť aj na vytvorenie anonymity pre ľudí, ktorí boli odfotení, ale aj zakrytie licenčných čísiel na vozidlách. [1] Toto vylepšenie sa začalo hromadne diať až v roku 2008. Tím Google street view vytvoril vlastný algoritmus, ktorý pomocou neurónových sietí a datasetu s počtom 30 000 tvári začali hromadne aplikovať rozmazanie na každú tvár a licenčné číslo vozidla detekovaných kamerou. V súčasnosti sa táto služba stále zdokonaľuje a vytvára priestor pre inovácie v tomto odvetví. V súčasnosti je taktiež nezanedbateľné využitie sférických kamier aj v automobilovom priemysle. Kamera sa zvyčajne nachádza v prednej časti vozidla a slúži na viacero úloh. Primárnou úlohou je detekcia okolia vozidla a detekcia chodcov, značiek alebo prekážok na ceste. Toto má veľké využitie pri vývoji autonómnych vozidiel, kedy tieto kamery zastupujú funkciu namiesto očí šoféra vozidla. Za pomocí takýchto kamier riadiaca jednotka dokáže reagovať na zmeny pred vozidloma zabrániť tak prípadnej kolízii. V takýchto riešeniach sa kladie veľmi vysoký dôraz na bezpečnosť a aj mala chyba môže spôsobiť vážne zranenie nielen posádke vozidla, ale aj ľudom, ktorí sa nachádzajú v blízkosti vozidla. Veľa výrobcov vozidiel používa na takúto detekciu používa Lidar [2], alebo hĺbkovú kameru. Tieto riešenia sú spoľahlivé, ale nedokážu cenovo konkurovať jednoduchej sférickej kamere. Práve preto niektorí výrobcovia stále používajú detekciu obrazu zo sférickej kamerya snažia sa ho vylepšiť pomocou rôznych algoritmov a neurónových sietí. Na snímanie okolia vozidla je použitá zvyčajne jedna alebo dve sférické kamery, ktoré dokážu snímať priestor omnoho efektívnejšie ako použitie viacerých kamier súčasne.

# **Jednoduchá detekcia objektov objektov**

Na detekciu objektov v obraze bola použitá knižnica OpenCV a jazyk Python. Knižnica OpenCV obsahuje široké spektrum metód, ktoré slúžia k spracovávaniu a analýze obrazu. V tomto projekte boli vyskúšané viaceré možnosti využitia a  vytvorený všeobecný prehľad o možnostiach použitia tejto kamery. Tieto možnosti pozostávajú zo samostatných  jednoduchých programov, ktoré sú implementované pomocou knižnice OpenCV. Väčšina týchto programov používa datasety, ktoré sú voľne dostupné na internete. Na meranie časovej náročnosti je do programu pridané meranie spracovania jednotlivých snímok videa, ktoré sa na konci spriemeruje. Tým vieme dostať prehľad o jednotlivých metódach, ktoré boli použité. Meranie úspešnosti bolo náročnejšie. Na správne meranie úspešnosti som vo videu počítal počet správne a nesprávne zdetekovaných objektov. Takýto náročný a zdĺhavý proces je aplikovateľný na kratšie videá. V mojom prípade testovacie videa mali dĺžku od 10 do 26 sekúnd. Pri analýze dlhších videí by bolo vhodné vytvoriť merací nástroj, ktorý by autoamticky vyhodnocoval úspešnosť použitého algoritmu.

## **5.1. Detekcia pohybu**

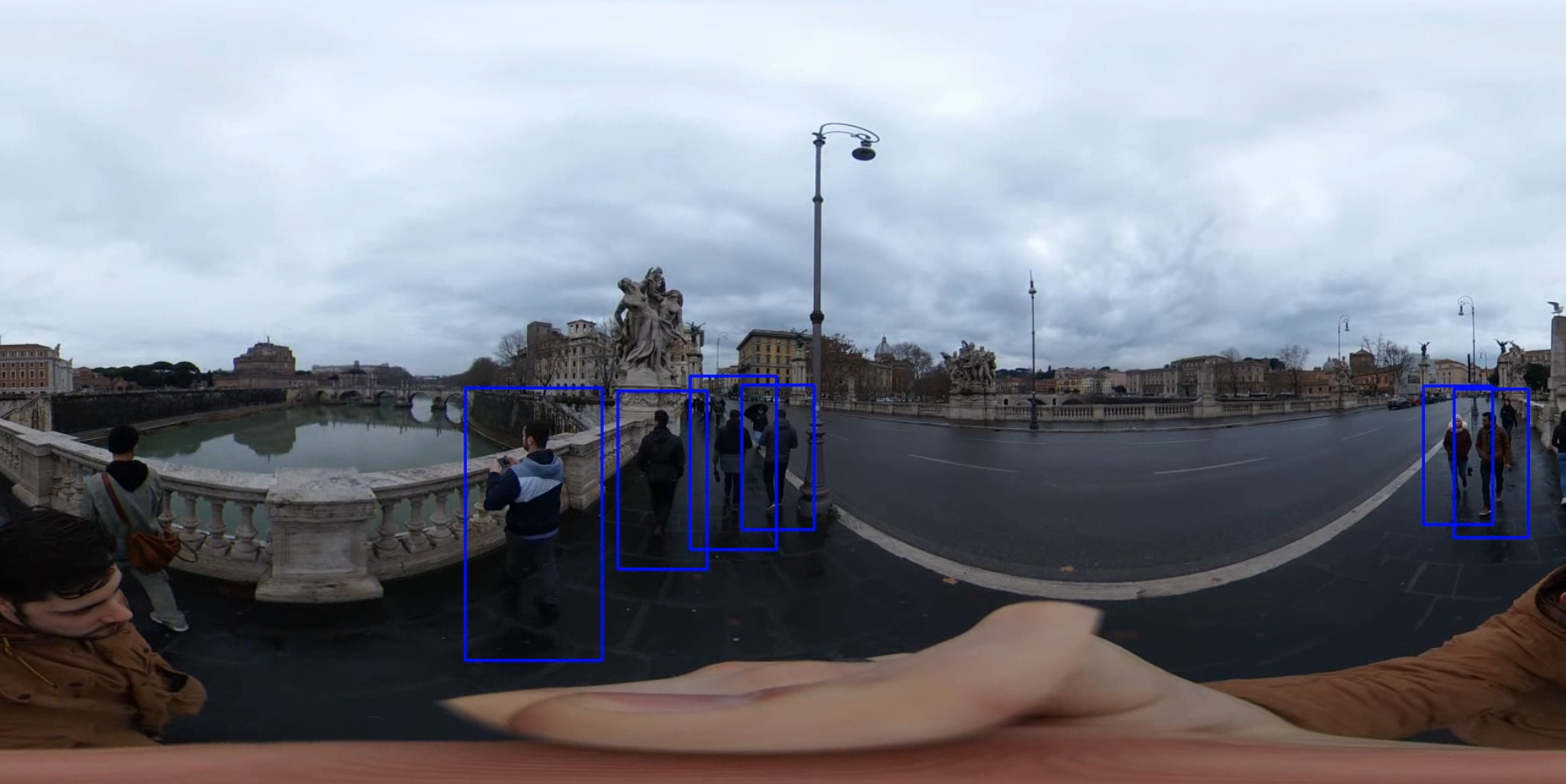
Prvá úloha, ktorú som sa rozhodol riešiť, bola jednoduchá detekcia pohybujúcich sa objektov. Na to bol vytvorený jednoduchý program, ktorý porovnáva každú snímku s predchádzajúcou snímkou z kamery. Výsledkom tohto programu je grafické znázornenie pohybujúceho sa objektu( Obrázok 1.). Na tejto ukážke si môžeme všimnúť zakrivenie obrazu v dolnej oblasti. Toto zakrivenie je spôsobené formátom, akým je reprezentovaný snímok zo sférickej kamery. Na tejto detekcii však toto zakrivenie nemá žiadny vplyv a je plné funkčné aj s týmto zakrivením. Video som spracovával v rozlíšení 2000x1000 pixelov. Na takúto triviálnu vec nieje potreba overovať vyš



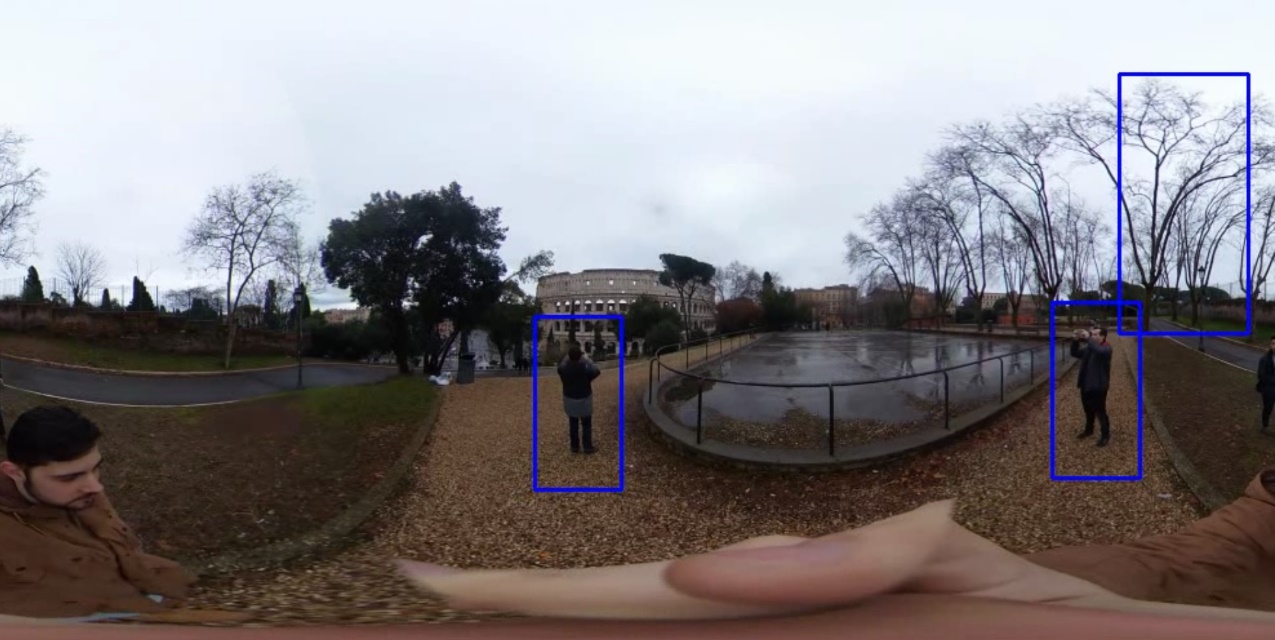
*Obrázok 1: pohybujúce objekty (zelená farba)*

## **5.2. Detekcia ľudí**

Na detekciu ľudí bola použitá vstavaná implementácia z OpenCV. Tá pozostáva z predtrénovaného detektora HOG. Pomocou tejto jednoduchej implementácie je možné detekovať ľudí v obrázkoch aj vo videostreame. Na začiatku program bol inicializovaný HOG detector. Pre nastavenie detekcie ľudí bola použitý predtrénovaný model z OpenCV – support vector machine(SVMDetector). Detekcia bola testovaná na viacerých rozlíšeniach s výškou snímku od 500px do 1920px. Pri rôznych rozlíšeniach sa úspešnosť držala na približne rovnakej hodnote. Z toho vyplýva, že kvôli lepším časom je lepšie použiť nižšie rozlíšenie. Na obrázku 2 môžeme vidieť, detekciu ľudí aj na snímke s horšími svetelnými podmienkami a nižším kontrastom. Problém s detekciou ľudí bol spôsobený napríklad objektami, ktoré pripomínali ľudské telo napríklad lampy, stromy, špecifické zábradlie a podobne. Chybnú detekciu môžeme vidieť na obrázku 3. Táto detekcia však nie je spôsobená formátom sférickej fotografie a podobná chyba by sa prejavila aj na obyčajných kamerách.



*Obrázok 2 Detekcia ľudí (rozlíšenie snímky 500x1000px, čas spracovania 550ms)*



*Obrázok 3: Detekcia ľudí s chybou (rozlíšenie snímky 500x1000px, čas spracovania 472ms)*

## **5.3. Detekcia tváre**

Na detekciu tváre bol použitý Haar Face detector z OpenCV. Tento detektor bol postupne vyvíjaný od roku 2001 Paulom Violom a Michaelom Jonesom [3] a existuje mnoho upravených variácii. Detektor s použitím knižnice *Dlib* a jazyku Python dokáže rozpoznávať tváre pomerne spoľahlivo a rýchlo. V testovacích videách sa spoľahlivosť pohybovala okolo 95%. Problém nastal, ak sa tvár v obraze nachádzala v hornej alebo dolnej časti snímky, kde dochádza k značnému skresleniu a deformácii v obrazu kvôli 2D projekcii.

*Obrázok 4: Detekcia tvárí(rozlíšenie snímky 1400x700px)*

## **Detekcia osôb v interiéri vozidla**

Hlavnou úlohou v tomto prípade bolo umiestnenie kamery na správne miesto. To som vyriešil stojanom na palubnú dosku. Tým je dosiahnutá najlepšia poloha kamery, ktorá dokáže v tomto prípade snímať posádku vozidla, ale aj okolie pred autom. Tým je možné využiť kameru nielen na detekciu tvári, ale napríklad aj na detekciu chodcov a iných prekážok pred vozidlom. Detekcia tvári bola skúšaná na rôznych rozlíšeniach. Po niekoľkých pokusoch som zistil, že chybovosť pri vysokom rozlíšení je výrazne vyššia. Výsledky som sa rozhodol interpretovať na rovnakom videu pri použití rôznych rozlíšení. Tieto výsledky sú zhrnuté v *tabuľke 2*.

|  |  |  |
| --- | --- | --- |
| Rozlíšenie(px) | Úspešnosť (%) | Čas spracovania snímky(ms) |
| 1000x500 | 85 | 209 |
| 1400x700 | 90.95 | 305 |
| 2000x1000 | 85 |  |
| 3840x1920 |  |  |

*Tabuľka 2: Výsledky detekcie osôb v interiéri vozidla*



*Obrázok 5: Detekcia tvárí vo vozidle ( rozlíšenie 1400x700px)*

Pri vyššom rozlíšení nielen že dochádzalo k vyššej časovej náročnosti, ale aj k zhoršeniu úspešnosti detekcie. Program pri vyššom rozlíšení začal označovať tváre v rôznych častiach snímky. (Obrázok 6). Kamera dokázala pracovať aj pri horších svetelných podmienkach s vysokou mieru úspešnosti. Najväčší problém pre kameru však boli nočné zábery, kde je interiér vozidla úplne tmavý a bez akéhokoľvek svetla. V takom prípade by bolo vhodné použiť napríklad infračervené podsvietenie, ktoré ľudské oko nevidí, ale výslednú snímku dokáže osvetliť na dostatočnú úroveň. S takýmto svetlom by bolo možné vylepšiť detekciu najmä v nočných záberoch bez denného svetla. Ďalší problém, ktorý nastal, boli prudké svetelné prechody. Kamere trvalo približne 1-2s, kým sa úroveň svetla v obraze zastabilizovala na prijateľnú úroveň. Medzi takéto prípady patril napríklad vjazd do tunela, jazda oproti priamemu slnku, alebo prejdenie do priameho svetelného lúča napríklad pri západe slnka. Všetky takéto problémy môžu narušovať detekciu vo veľkej miere. Pre zvýšenie kvality detekcie v takýchto prípadoch by bolo vhodné použiť kvalitnejšiu kameru, ktorá dokáže zmenu svetelných podmienok spracovať efektívnejšie. Cena takejto kamery sa však pohybuje omnoho vyššie ako cena Ricoh Theta V. Pri hľadaní využitia sférickej kamery rozhoduje aj cena, ktorú by bol potenciálny zákaznik zaplatiť za plne funkčnú detekciu. V takejto časti by bolo vhodné vyskúšať viacero kamier a nájsť rozumný kompromis medzi optimálnymi a postačujúcimi zábermi z kamery a cenou.



*Obrázok 6: Detekcia tvári s vyššou chybovosťou ( rozlíšenie 3840x1920px)*

# **Pokročilá detekcia tváre pomocou landmarkov**

Po jednoduchej detekcii tvári som sa rozhodol túto problematiku riešiť komplexnejšie. Pre analýzu vodičovho správania som sa rozhodol využiť landmarky. [4]. Program spočíva v hľadaní bodov záujmu tváre a tie sú následné označené bodmi. Program využíva naučený dataset. Vďaka tomu je možné v tvári analyzovať jej časti. Podľa toho vieme napríklad predikovať, ako často vodič mrká očami, alebo či si nezavrel oči na dlhšiu dobu, napríklad kvôli mikrospánku. Takáto analýza vodičovho správania je zložitejšia a vyžaduje si určitú mieru učenia. Nieje možné spraviť univerzálny program, ktorý by fungoval pre všetkých vodičov rovnako správne. Program fungoval spoľahlivo iba na vyšších rozlíšeniach. Pri menšom rozlíšení tvár nebol vôbec zdetekovaná. To môže byť spôsobené tým že obraz v menšom rozlíšení nieje natoľko kvalitný, aby vyhodnotil časti tváre s dostatočne vysokou mierou úspešnosti. Z dôvodu použitia vyššieho rozlíšenia sa tiež zvýšila časová náročnosť na spracovanie jednej snímky.



*Obrázok 7: Detekcie Landmarkov (rozlíšenie snímky 3840x1920px)*

Pri rozlíšení menšom ako 1800x900 je detekcia landmarkov neúspešná. Úspešnosť bola meraná na rovnakom videu pri rôznych rozlíšeniach. Na testovacom videu, ktoré trvalo 26 sekúnd sa nachádzali 2 ľudia, ktorí boli stále otočení tvárou ku kamere. Toto video sa skladá zo 763 snímkov a z toho vyplýva, že pre maximálnu úspešnosť bolo potrebné detekovať 1526 tvári. Najväčšiu úspešnosť dosiahlo video s najvyšším rozlíšením 3840x1920px. V tomto prípade bolo zdetekovaných 1512 tvárí, čo nám dáva úspšešnosť približne 99.1% (Tabuľka 3.) Pri takomto vysokom rozlíšení však dochádzalo ku veľkej časovej náročnosti, ktorá by v reálnom svete bola pre real-time videu nepoužiteľná. Aby sme mohli spracovávať takéto video v reálnom čase, bolo by vhodné dostať sa pod časovú hranicu aspoň 100ms.

Na vylepšenie detekcie by bolo vhodné robiť z celej fotky iba výrez, kde sa ľudská tvár nachádza. Aby sme mohli urobiť takúto optimalizáciu, musela by byť kamera v každom prípade otočená rovnakým smerom a nemohla by byť umiestnená inak.

|  |  |  |
| --- | --- | --- |
| Rozlíšenie(*px*) | Úspešnosť (*%*) | Čas spracovania snímky(*ms*) |
| 1000x500 | 0 | 185 |
| 1400x700 | 0 | 260 |
| 1800x900 | 55.29 | 339 |
| 2000x1000 | 96.29 | 395 |
| 2880x1440 | 99.6 | 699 |
| 3840x1920 | 99.1 | 1088 |

*Tabuľka 3: Úspešnosť detekcie landmarkov*

# **Možnosti vylepšenia detekcie**

# **Zhrnutie**

V tomto projekte som zhrnul problematiku detekcie objektov sférickou kamerou. Na detekciu objektov som vytvoril programy v jazyku Python, využil funkcionalitu OpenCV a datasety objektov. Zistil som, že za určitých podmienok je možné použiť sférickú kameru na bežnú detekciu objektov. Limity nastávajú pri objektoch, ktoré sa nachádzajú v hornej alebo dolnej oblasti snímku. To je zapríčenené reprezentáciou snímku so sférickej kamery. Najväčší potenciál využitia kamery vidím hlavne v automobile. Kamerou je možné kontrolovať napríklad správanie vodiča, predvídať jeho únavu a predísť tak nehode. Kamerou je zároveň možné súčasne detekovať aj prekážky pred vozidlom, ktoré by mohli ohroziť jazdu alebo spôsobiť kolíziu. V diplomovej práci by som sa chcel zamerať najmä na pokročilejšiu detekciu a analýzu ľudskej tváre vodiča vo vozidle.

# **Bibliografia**

|  |  |
| --- | --- |
| [1] | Anguelov, Dragomir and Dulong, Carole and Filip, Daniel and Frueh, Christian and Lafon, Stephane and Lyon, Richard and Ogale, Abhijit and Vincent, Luc and Weaver, Josh, „Google street view: Capturing the world at street level,“ IEEE, 2010. |
| [2] | Premebida, C., Monteiro, G., Nunes, U., & Peixoto, P., „A lidar and vision-based approach for pedestrian and vehicle detection and tracking.,“ IEEE, 2007. |
| [3] | Viola, Paul and Jones, Michael and others, „Rapid object detection using a boosted cascade of simple features,“ CVPR (1), 2001. |
| [4] | Zhang, Zhanpeng and Luo, Ping and Loy, Chen Change and Tang, Xiaoou, „Facial landmark detection by deep multi-task learning,“ Springer, European conference on computer vision, 2014. |
| [5] | Han, Sungji, Youngjoon Han, and Hernsoo Hahn, „Vehicle detection method using Haar-like feature on real time system.,“ World academy of science, engineering and technology, 2009. |