VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

POČÍTÁNÍ PRŮCHODŮ OSOB DVEŘMI S VYUŽITÍM STACIONÁRNÍ KAMERY

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE ĽUBOŠ TICHÝ
AUTHOR

BRNO 2015

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

POČÍTÁNÍ PRŮCHODŮ OSOB DVEŘMI S VYUŽITÍM STACIONÁRNÍ KAMERY

COUNTING OF PEOPLE PASSING THROUGH A DOOR USING A STATIONARY CAMERA

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE ĽUBOŠ TICHÝ
AUTHOR

VEDOUCÍ PRÁCE Ing. JAKUB SOCHOR
SUPERVISOR

BRNO 2015

Abstrakt

Tato bakalářská práce se zabývá počítáním lidí, který procházejí dveřmi. To může být využito v případe, když chceme vědět, kolik osob se nachází ve sledované místnosti. První část je opis existujících metod počítání lidí a druhá se zabývá využitými metodami počítačového vidění. Navrhnutý systém dokáže počítat lidi, a to včetně určeni jejich směru pohybu. Poslední část obsahuje vyhodnocení použitých metod počítačového vidění.

Abstract

This bachelor thesis deals with counting of people passing through a door. It may be used in case, when we need to know, how many people there are in monitored room. The first part is a description of existing method of counting people and the second part deals with used computer vision methods. Proposed system can count people, including distinguishing their motion direction. The last part contains an evaluation of used computer vision methods.

Klíčová slova

počítaní osob, detekce osob, kaskáda klasifikátorů, sledování osob

Keywords

counting of people, person detection, cascade of classifiers, tracking

Citace

Ľuboš Tichý: Počítání průchodů osob dveřmi s využitím stacionární kamery, bakalářská práce, Brno, FIT VUT v Brně, 2015

Počítání průchodů osob dveřmi s využitím stacionární kamery

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Jakuba Sochora. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

……
Ľuboš Tichý
20. května 2015

Poděkování

Tímto bych moc rád poděkoval panu Ing. Jakubovi Sochorovi za odbornou pomoc, profesionalitu a za půjčení kamery. Dále bych chtěl poděkovat své rodině a přátelům, kteří mne povzbuzovali a podporovali při psaní této práce.

© Ľuboš Tichý, 2015
Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

Obsah 1
1. Úvod 2
2. Metódy počítania ľudí 3
2.1. Zhlukovaním výstupov z detektora osôb 3
2.2. Poznávaním vzhľadu z viacerých uhľov pohľadu 6
3. Použité metódy počítačového videnia 13
3.1. Detekcia objektov využitím Houghovej transformácie 13
3.2. Metóda sledovania, poznávania a detekcie 16
3.3. Metóda on-line boosting 18
3.4. Subtrakcia pozadia 21
4. Návrh systému 22
4.1. Detekcia na základe pohybu objektu na obraze 22
4.2. Detekcia osôb detektorom 22
4.3. Kombinácia detekcií 22
4.4. Sledovanie a počítanie osôb 23
5. Implementácia 24
6. Vyhodnotenie 25
6.1. Testovanie detekcie osôb 25
6.2. Testovanie sledovania osôb 26
6.3. Vyhodnotenie počítania ľudí 27
6.4. Záver vyhodnotenia 29
7. Záver 30

1. Úvod

Počítanie je s nami späté už od mala. Už ako malí sme vedeli ukázať príbuzným, koľko máme rokov. Vtedy sme naše malé prsty hrdo vystreli a ukázali, ako sme už vyrástli. Neskôr s počítaním sme sa stretli na hodinách matematiky. V dospelosti sa už počítaniu nedá ani vyhnúť. Počítame peniaze, počítame sekundy, ktoré pomaly plynú, keď sme v práci a počítame mnoho iných vecí.

S pribúdaním dní taktiež pribúda čoraz väčší a väčší počet ľudí na svete. Dnešná doba je veľmi rýchla a divoká a je dobré mať prehľad o počte ľudí. Počítanie ľudí je neľahká úloha. Bola možnosť brigády, ktorej cieľom bolo spočítať v nákupných centrách počet ľudí. Brigádnik motivovaný zárobkom drží v ruke počítadlo a pri prechode osoby počítadlo stlačí. V dnešnom modernom svete, kde už je všetko ovládané počítačom, takéhoto brigádnika už nie je potrebné mať. Ľudí a ich pohyb je možné zachytiť kamerou a špecializovaným softvérom je možné z videa kamery získať počet ľudí. Jednoduché, nie?

Moja práca opisuje už existujúce metódy počítania ľudí. Tie sú opísané v nasledujúcej kapitole. Kapitola 3 opisuje metódy, ktoré sa podieľajú na počítaní a to sú metódy detekcie osôb a sledovania pohybu osôb. Ja som tieto metódy využil a vytvoril som vlastný spôsob počítania ľudí, ktorý je bližšie popísaný v kapitole 4. Počítanie ľudí prebieha kombináciou použitých metód a pracujú s videom, ktoré bolo zhotovené stacionárnou kamerou. Kamera bola postavená pri vchod do miestností a snímala prechádzajúcich ľudí. Samotná implementácia je opísaná v kapitole 5 a v kapitole 6 je obsiahnuté vyhodnotenie, ktoré ukázalo dosiahnuté výsledky.

2. Metódy počítania ľudí

Počítanie ľudí patrí medzi výzvy počítačového videnia. Na počítanie ľudí sa používajú rôzne metódy. Vo väčšine prípadov je využitá detekcia osôb spojená so sledovaním pohybu osôb. V nasledujúcich podkapitolách sú predstavené dve metódy na počítanie ľudí.

2.1. Zhlukovaním výstupov z detektora osôb

Metóda počítania ľudí [1] ako vstup namiesto pohybu osôb vezme zhluky výstupov z detektora osôb. Keďže aj najlepší všeobecný detektor má nepresné výstupy, tak jedna osoba môže byť detekovaná viackrát. Spôsobujú to nepresné vyhľadávania osôb. Pre väčšiu presnosť sú zohľadnené rôzne typy farieb, rôzne typy priestoru a je pridaná dočasná informácia o vlastnostiach jednotlivých zhlukov.

2.1.1. Použitý model s Dirichletovým princípom

Metóda je založená na modeli zmesi využívajúceho Dirichletov princíp (angl. *Dirichlet Process Mixture Model*) [2]. Venuje sa mu zvýšená pozornosť v aplikáciách počítačového videnia, ktoré sú určené pre sledovanie objektu. Jeho výhodou je implicitne bezparametrová podstata. Táto výhoda umožňuje určiť odchýlky zhlukov, keď ich počet nie je vopred známy a je použitá pre odhad skupiny osôb, v ktorých jedna osoba sa môže nachádzať vo viacerých zhlukoch.

Model umožňuje modelovať pozorovania ako modely zmesí, ktoré majú neznámy počet komponentov zmesí [2]. V tejto koncepcií každý komponent zodpovedá zhluku. Ak počet komponentov zmesí smeruje k nekonečnu, tak rozdelenie nadobudne Dirichletov princíp. Model rozhodne, že nekonečný počet komponentov zmesí existuje iba vtedy, ak existuje konečný počet týchto komponentov, ktoré majú k nim priradené pozorovania. Modelovanie dát pozostáva z hľadania parametrov týchto konečných a nekonečných zmesí komponentov. K odhadu parametrov sa použije Markov reťazec vzorkovania, ktorý je podobný Gibssovmu vzorkovaniu. Iteruje nad všetkými pozorovaniami a vzorkuje priradenie k existujúcim alebo k novým zhlukom každého pozorovania. Vzorkuje také priradenia, pre ktoré to je možné. Pravdepodobnosť priradenia je kontrolovaná jednoduchým parametrom α. Jeho vyššie hodnoty výsledku vo viacerých zhlukoch sú vyjadrené rovnicami

$$p(c_n;\alpha) = \begin{cases} \dfrac{N_k}{N+\alpha-1} p(X_n|\theta_k) & \text{pre existujúce } k \\ \dfrac{\alpha}{N+\alpha-1} \displaystyle\int_\theta p(X_n|\theta)\, d\theta & \text{pre nové } k{+}1 \end{cases} \tag{2-1}$$

kde N_k je počet priradení k zhlukom k a N je počet pozorovaní.

Detekcie jednej osoby sa prirodzene zoskupia okolo oblasti práve detekovanej osoby. Ostatné vizuálne vlastnosti, medzi ktoré patrí napríklad farba, sa pohybujú okolo priemernej hodnoty.

2.1.2. Skvalitnenie výsledkov detekcie

Detektor osôb je aplikovaný na každú snímku. Je založený na histograme orientovaných gradientov (angl. *histogram of oriented gradients*), ktorý vypočíta gradienty v obraze v horizontálnych a vo vertikálnych smeroch. Zhromaždí vytvorené histogramy, a tým vzniknú malé bunky po celom obraze. Normalizované hodnoty týchto buniek predstavujú HOG vlastnosti. HOG vlastnosti sú vyňaté pre veľký počet pozitívov (predstavujúce osoby) a niekoľko negatívov (nepredstavujúce osoby). Klasifikátor kaskád ich klasifikuje. Pátracie okienko krokuje počas detekcie po obraze. Klasifikátor rozhoduje podľa vyňatých HOG vlastností pre danú pozíciu pátracieho okienka, či sa na obraze osoba nachádza alebo nenachádza. Veľkosť rozmerov pátracieho okienka je každou iteráciou opakovane zväčšená alebo vo väčšine prípadov zmenšená, a teda sa vykoná viacrozmerné prehľadávanie.

Aby sa rozhodlo o oblastiach detekcie, tak sa vykoná súhrn výsledkov detekcie nad tromi po sebe idúcimi snímkami. Tieto výsledky sú doplnené výpočtom, ktorým sa získajú dve sady máp optických tokov (angl. *optical flow maps*) [3]. Výpočet jednej sady prebehne medzi predchádzajúcou a súčasnou snímkou. Výpočet druhej sady prebehne medzi súčasnou a nasledujúcou snímkou. Tieto výpočty slúžia pre lepší odhad oblastí detekcie vzhľadom k súčasnej snímke. Vektor posunutia oblasti detekcie je priemerom vektorov optických tokov. Vektory optických tokov sú tvorené kľúčovými bodmi, ktorými je pokrytá oblasť detekcie. Využitie detekcie z viacerých snímok kompenzuje potenciálnu stratu pozitívov. V nasledujúcich krokoch sú tiež uplatnené vektory optických tokov s kľúčovými bodmi.

Hodnota pravdepodobnosti popredia je vypočítaná Gaussovým modelom zmesi na základe reprezentácie pozadia [4]. Pri modelovaní využíva predchádzajúcu zmenu farby pixelu. Výsledky detekcie sú vyfiltrované od takých, kde pixely majú menšiu priemernú hodnotu pravdepodobnosti popredia než je preddefinovaný prah. Sú odstránené plochy popredia, ktoré sú väčšie než je preddefinovaná veľkosť. Aplikovaním týchto dvoch heuristík sa redukujú detekcie falošných pozitívov.

2.1.3. Pozorovania a modely zhlukov

Po získaní oblastí detekcie oddelených od snímok sa extrahujú sady vlastností pre každé okienko detekcie. Počas fázy, kedy model zmesi s Direchletovým princípom vzorkuje, sa priradia pozorovania k sadám vlastností. Pri modelovaní pozorovaní sa využijú x a y súradnice pixelu a stredná hodnota farby pixelu, ktorý je súčasťou popredia komponentu. K informáciám o farbe a o priestore sa pridá doplňujúca informácia o smere pohybu ľudí. Táto doplňujúca informácia je získaná zo sady vlastností, ktoré patria mapám optických tokov so susedných snímok. Histogram orientovaných optických tokov (angl. *Histogram of oriented optical flows*) [5], kde každý vektor optického toku prislúcha k určitému

kontajneru histogramu. Kontajner histogramu zodpovedá orientácii ohodnotenej podľa jeho miery. Sú uvedené histogramy so štyrmi kontajnermi a sú doplnené o vlastnosti každej oblasti detekcie.

Podobne sú modelované aj zhluky. Využíva sa stredná hodnota a zmena tej istej farby a priestorových kontajnerov ako pri kontajneroch histogramov orientovaných optických tokov. Z dôvodu počítania sa nemodelujú Gaussove modely s plnou kovarianciou matíc, ale iba s kovarianciou koeficientov medzi priestorovými komponentmi. Každé pozorovanie je definované 4 parametrami $X{:}(\mu_x, \mu_y, \mu_a, \mu_b, \mu_{h_{1\ldots4}})$ a každý zhluk so 17 parametrami θ: $(\mu_x, \mu_y, \Sigma_{xy}, \mu_a, \sigma_a, \mu_b, \sigma_b, \mu_{h_{1\ldots4}}, \sigma_{h_{1\ldots4}})$. Tento model pravdepodobnosti, že pozorovanie X_n, ktoré je vyjadrené zhlukom k, je vyjadrený rovnicou

$$p(X_n|\theta_k) = N\left(X_{xy}\middle|\mu_{xy}^k, \Sigma_{xy}^k\right) \prod N\left(X|\mu^k, \sigma^k\right) \tag{2-2}$$

kde výsledok pravej strany je vypočítaný pre vlastnosti histogramu orientovaných optických tokov a parametre Gaussovho modelu sú odhadnuté z pozorovaní, ktorým sú priradené zhluky. Rovnice (2-1) a (2-2) priradia pravdepodobnosť pozorovania k existujúcemu alebo k novému zhluku.

2.1.4. Zhlukovanie a parameter α

Využitím sady extrahovaných pozorovaní a odpovedajúcich vlastností k nim pre snímku sa vykoná iteratívne Gibssovo vzorkovanie a vzorka je priradená ku každému pozorovaniu. Pozorovania sú vyhodnotené snímka po snímke a vypočíta sa pravdepodobnosť priradení k existujúcim alebo k novým zhlukom. Je posilnený proces zhlukovania, ktorý vedie k vygenerovaniu sledovateľného počtu zhlukov. Toto posilnenie je implementované do procesu zhlukovania, ktoré je modifikované Gibssovým vzorkovaním pravdepodobnosti s inou pravdepodobnostnou hodnotou, ktorá rešpektuje veľkosť zhluku.

Pre každú snímku sú vypočítané štatistiky pre šírku a výšku pozorovaní. Využitím týchto štatistík sa aktualizuje pravdepodobnosť vzorkovania.

Optimálna hodnota α sa týka hustoty ľudí v scéne a môže sa časom meniť. Jej hodnotu zistíme tak, že sa získajú reprezentatívne video snímky ako predvolený set a navrhnutý algoritmus pobeží s rozdielnymi hodnotami α, aby sa rozhodlo, či ide o kľúčové body optických tokov alebo či ide o chybu.

Ideálny výstup procesu zhlukovania je taký, že každá osoba nachádzajúca sa v obraze je reprezentovaná jedným zreteľným zhlukom, a teda počet osôb v obraze, by sa mal rovnať počtu zhlukom v obraze. To nie je vždy dosiahnuté, pretože osoby sa môžu nachádzať blízko seba a môžu sa na seba podobať, čo vedie k nepresnosti určeniu počtu osôb. Takže počítanie zhlukov sa nesmie brať do úvahy ako počítanie osôb. Je nutné odhadnúť počty osôb N_k v zhlukoch k s využitím vzťahu

$$N_k = \left\lceil \frac{p_k}{\bar{p}_{n\in k}} \right\rceil \tag{2-3}$$

kde p_k je celkový počet kľúčových bodov v zhluku k a $\bar{p}_{n \in k}$ je priemerný počet kľúčových bodov v oblastiach detekcie predstavujúce zhluk k.

2.2. Poznávaním vzhľadu z viacerých uhľov pohľadu

Proces počítania ľudí mnohokrát sťaží situácia, kedy nie je vidieť celú postavu osoby alebo je vidieť veľmi málo z celej postavy osoby. Vznikajú takzvané oklúzie. Spôsobujú ich rôzne prekážky v scéne, ako sú napríklad stĺpy, steny, stromy a iné. Taktiež ich môžu spôsobiť aj samotné osoby, tým že prejdú popri sebe, čím jedna osoba prekryje druhú osobu za ňou. Metóda počítania ľudí [6] rieši tento problém tým, že pracuje s viacerými pohľadmi na scénu.

2.2.1. Neurálne modelovanie na šablónach pohybu

Detekcia popredia v každom pohľade na scénu je základnou stavebnou jednotkou tejto navrhnutej metódy a od jeho presnosti závisí aj presnosť celkového procesu. Preto je použitý samostatne organizovaný model pozadia [7] pre sekvenciu obrázkov, ktorého kvalita detekcie pohybujúcich sa objektov bola overená.

Pre sekvenciu snímok I_t a pre každý pixel x snímkovej domény D je vytvorená nervová mapa pozostávajúca z n váhových vektorov $m_t^i(x), i = 1, \dots, n$. Nervová mapa je nazvaná modelom m pre pixel x. Ak každá snímka sekvencie má P riadkov a Q stĺpcov, tak kompletný set modelov $M_t(x) = (m_t^1(x), \dots, m_t^n(x))$ pre všetky pixely x každej t-tej snímky I_t, je organizovaná ako 3D neurálna mapa M_t s P riadkami, s Q stĺpcami a s n vrstvami, kde každá vrstva L_t^i pozostáva z i-teho váhového vektora $m_t^i(x)$ pre každý pixel x.

Počiatočné modelové pozadie E_0 je odhadnuté z K-tich počiatočných snímok sekvencie prostredníctvom dočasného mediánu. Všetky váhové vektory neurálnej mapy súvisiace k pixelu x sú inicializované hodnotou jasu pixela $E_0(x)$. Vďaka ďalšiemu poznávaniu neurálnej mapy sa pozadie modelu dokáže adaptovať podľa modifikácií v scéne a to bez vplyvu pixelov, ktoré nepatria do scény pozadia. Proces poznávania pozostáva zo selektívneho aktualizovania modelu vďaka zmenám neurálnych váh. Hodnota $I_t(x)$ v čase t, každého prichádzajúceho pixelu x t-tej sekvencie I_t, je porovnávaná so súčasným pixelovým modelom $M_t^n(x) = (m_t^1(x), \dots, m_t^n(x))$. O najlepšej zhode váhového vektora $m_t^n(x)$ sa rozhodne podľa rovnice

$$d\left(m_t^b(x), I_t(x)\right) = \min_{i=1,\dots,n} d\left(m_t^i(x), I_t(x)\right) \tag{2-4}$$

kde metrika d je vhodne vybraná podľa odpovedajúceho špecifického miesta farby.

Ak najlepšia zhoda váhového vektora, ktorá je vypočítaná z rovnice (2-4), je dosť blízko k hodnote pixelu $I_t(x)$, tak môže byť použitá ako aproximácia kódovania pixelu. To nastáva ak platí vzťah

$$d\left(m_t^b(x), I_t(x)\right) \leq \epsilon_d \tag{2-5}$$

kde ϵ_d je prah, podľa ktorého je možné rozoznať medzi pixelmi popredia a medzi pixelmi pozadia. V prípade ak neplatí vzťah (2-5), tak prichádzajúci pixel *x* je klasifikovaný ako popredie. Vďaka tejto voliteľnosti sa model pozadia vyhne objektom, ktoré nepatria do pozadia scény, či už sa tieto objekty pohybujú alebo stoja. Týmto sa zaistí, že osoby budú detekované ako popredie, a teda bude ich možné počítať.

Rovnica (2-4) poskytne pre každý prichádzajúci pixel *x* a pri čase *t* vzdialenosť hodnoty pixelu $I_t(x)$ zo súčasného neurálneho modelu pozadia M_t. Táto vzdialenosť môže byť použitá v prípade aproximácie pravdepodobnosti pixelu *x* v čase *t* súčasťou takzvaného popredia $SOFLIM_t(x)$ (skratka z angl. *Self-Organizing Foreground Likelihood Map*) (Obrázok 2-1), dané rovnicou

$$SOFLIM_t(X) = \frac{d\left(m_t^b(x), I_t(x)\right)}{d_t^{max}} \tag{2-6}$$

kde d_t^{max} je maximálna vzdialenosť od súčasnej snímky I_t po súčasný neurálny model pozadia M_t, dané vzťahom

$$d_t^{max} = \max_{x \in D} d\left(m_t^b(x), I_t(x)\right) \tag{2-7}$$

2.2.2. Lokalizácia osôb

K lokalizácii osôb sú použité informácie z viacerých pohľadov na scénu. Informácie sú odolné voči oklúziam a zabezpečia dostatok viditeľnosti. Typicky vo videách sú taktiež vidieť rovinné štruktúry, napríklad steny a podlaha, ktoré sú tiež využité. Snímkou po snímke je vytvorené popredie SOFLIM, ktoré je špecifické pre každý odlišný pohľad. Popredie synergických pohľadov je vypočítané z výstupov popredí SOFLIM s pomocou klasifikátora maximálnej pravdepodobnosti. Spájanie týchto informácii je založené na metóde *HOC* (skratka z angl. *Homographic Occupancy Constraint*) [8], ktorá pracuje s využitím homografie. Homografia prislúcha k pohľadu na každú scénu a k podlahe každej scény. Navyše je skombinovaná s vizuálnymi vlastnosťami, ktoré sú dostupné z rôznych bodov pohľadu. Pozícia chodidiel osôb na podlahe umožňuje zrekonštruovať „mapu chodidiel“ (angl. *feet map*) (Obrázok 2-1) pomocou homografií z popredí $SOFLIM_t(x)$.

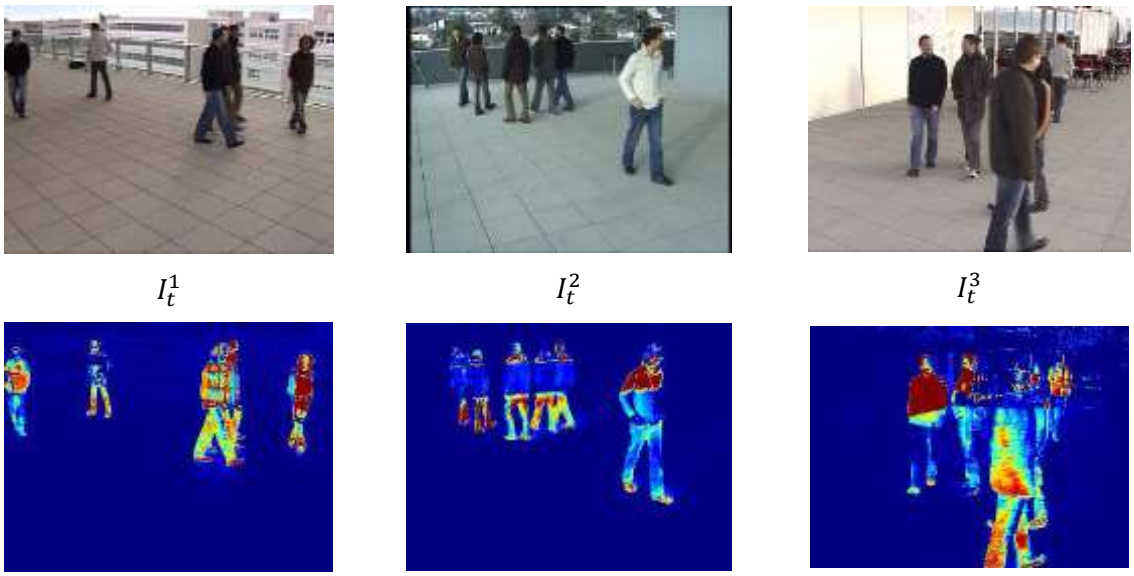
Sú dané popredia $SOFLIM_t^i, i = 1, \dots, V$ v čase *t* získaných zo snímok z rôznych kamier *V*. $SOFLIM_t^r$ sa indikuje popredie pohľadu *r*. Pohľad *r* je vybratý z dostupných pohľadov. Ďalej je daná matica homografie $H_{i_\pi r}$, ktorá je vytvorená rovinou π medzi uvedeným pohľadom *r* a pomocou pohľadu *i*, *i =1,...,V*. Použitím $H_{i_\pi r}$ v každom čase *t* môže byť pixel x^rz popredia pohľadu *r* projektovaný na pixel x^iv popredí $SOFLIM_t^i, i = 1, \dots, V$. Preto na základe homografie je možné definovať sadu pozorovaní $I_t^1, I_t^2, \dots, I_t^V$ „zvlnených“ (angl. *wraped*) popredí pravdepodobnostných máp $WSOFLIM_t^1, WSOFLIM_t^2, \dots, WSOFLIM_t^V$ na pozíciách $x^1, x^2, \dots, x^V$, skrátene zapísané $I_t^i = WSOFLIM_t^i\left(x^i\right), i = 1, \dots, V$.

Udalosť, kedy pixel x^{π} na podlahe π reprezentuje polohu na rovinne objektu popredia, je vyjadrená ako $MSOFLIM_t(x)$. Je možné aproximovať hustotu pravdepodobnosti tejto udalosti, tým že sa aplikuje Bayesov teorém [9] k pozorovaniam $I_t^1, I_t^2, \dots, I_t^V$.

Metóda HOC určuje, že ak pixel má bod prieniku na podlahe, ktorý patrí do scény popredia, tak potom má projekciu v popredí každého pohľadu.

Proces lokalizácie osôb je možné zhrnúť do niekoľkých bodov:

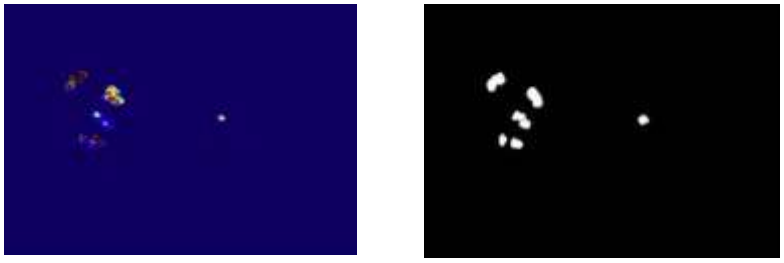
1. Konštrukcia popredia pravdepodobnostných máp $SOFLIM_t^i, i = 1, \dots, V$, pre V pohľadov.
2. Konštrukcia „zvlneného“ popredia pravdepodobnostných máp $WSOFLIM_t^i, i = 1, \dots, V$. Popredie sa zvlní na ploche podlahy aplikovaním homografie na každé popredie pravdepodobnostných máp.
3. Konštrukcia synergickej mapy $MSOFLIM_t(x)$ vzniknutá spojením zvlnených popredí pravdepodobnostných máp.
4. Konštrukcia mapy chodidiel $FM_t(x)$. Mapa chodidiel je mapa pokrytá polohou pohybujúcich sa objektov na ploche podlahy. Je získaná prahovaním synergickej mapy $MSOFLIM_t(x)$. Prahovanie prebieha na základe identifikovania pixelov s najvyššou hodnotou pravdepodobnosti výskytu pixelu vo všetkých pohľadoch scény.



I_t^1 I_t^2 I_t^3

$SOFLIM_t^1$ $SOFLIM_t^2$ $SOFLIM_t^3$

Obrázok 2-1 Sú vidieť rôzne pohľady do scény I_t^i, i = 1,2,3 a pod nimi sú k nim odpovedajúce popredia SOFLIM_t^i, i = 1,2,3.



(a) (b)

Obrázok 2-2 (a) Mapa zo SOFLIM (b) Mapa chodidiel získaná z prahu hodnôt

2.2.3. Sledovanie osôb

Spôsob sledovania je opačný od zvyčajných spôsobov, kedy škvrny (angl. *blobs*), ktoré sú sledované, prichádzajú z popredia scény s jedným pohľadom. Použitý spôsob sledovania sleduje škvrny, ktoré pozostávajú z pozície pohybujúcich sa ľudí pochádzajúcich zo scény podlahy. Avšak, to smeruje k zopár problémom:

- Chromatické časti pozície veľmi nerozlišujú chodidlá rôznych osôb.
- Rozmery takýchto pozícií sú zvyčajne malé, a preto sú ťažšie sledovateľné.
- Homografia podlahy zväčšuje odchýlky pozícií za sebou idúcich snímok a predvídanie pri sledovaní sa stáva zložitejším
- Pozície na mape chodidiel sa zdajú byť fragmentované, pretože ľudské stopy po chodidlách pri chôdzi nemusia byť súvislé a môžu generovať falošné pozitíva pri sledovaní.

Navrhnutá metóda je založená na časovo-priestorovej súdržnosti mapy chodidiel spojených komponentov, ktoré zodpovedajú pohybujúcim sa osobám. Je možné predpokladať, že v podstate pozície, ktoré sú detekované pre niekoľko po sebe idúcich snímok, sú vzájomne závislé, a teda je možné ich sledovať len použitím ich pozície a tvaru. Škvrny sú použité k predpokladaniu nového javu, ktorý použije očakávaný vzhľad modelu Ak sa vezme do úvahy dočasné okienko so šírkou Δt, tak niekoľko hypotéz môže prislúchať pre každú pozíciu a súdržnosť týchto hypotéz môže byť vyhodnotená porovnaním k pozícii mape chodidiel vykonanej v čase t. Základné princípy:

a) Vlastnosti týkajúce sa tvaru každej pozície sú oddelené, vrátane stredu parametrizácie pozície škvrny a vrátane štatistík týkajúcich sa tvaru, ktoré sa počítajú pre súdržnosť pixelov pozícií zahrnutých v škvrne.

b) Súdržnosť je vyhodnotená vzhľadom ku vlastnostiam. Je odhadnutá takzvaná dôveryhodnosť (angl. *confidence*), čo je vzdialenosť medzi vyskúšanými hypotézami a mapou chodidiel škvrny. Ak dôvernosť je väčšia než určitý prah, tak hypotéza je eliminovaná a nasleduje skúšanie ďalších hypotéz. Ak je dôvernosť pod určitým prahom, tak hypotéza je vyhodnotená a pokračuje sa overovaním možných kolízií s inými trajektóriami, napríklad s predchádzajúcimi potvrdenými hypotézami.

c) Analýza kolízie je získaná overovaním možných prienikov medzi predchádzajúcimi hypotézami a inými už existujúcimi trajektóriami. Ak sa nevyskytuje kolízia, tak nová pozícia trajektórie a jej rozmery sú aktualizované. Analýzy sú posilnené vysporiadaním sa s problémami delenia, spájania a so stratou sledovaných objektov. Sú posilnené prevzatým modelom [10], ktorý patrí do skupiny viacnásobného testovania hypotéz.

Algoritmus môže byť načrtnutý pre každú škvrnu nasledujúcim spôsobom:

1. Formulujú sa predvídané hypotézy založené na Kalmanovom filtre [11] pre škvrny predstavujúce pohyb na mape chodidiel,
2. Pre každú hypotézu sa vyhodnotí dôvernosť. Ak je dôvernosť pod určitým prahom, hypotéza je potvrdená, inak je eliminovaná a pokračuje sa ďalšími hypotézami.
3. Analyzujú sa kolízie medzi existujúcimi trajektóriami a potvrdenými hypotézami, kde neskoršia hypotéza je eliminované a vracia sa na krok 2, inak potvrdené hypotézy sa stávajú novými trajektóriami. Ich pozícia a rozmery sú aktualizované Kalmanovým filtrom.
4. Aktualizuje sa list trajektórií.

Keďže sledovanie je založené iba na pozícii a rozmeroch každého spojeného komponentu na mape chodidiel, tak je možne sa vyhnúť analýze chromatických vlastností a to je dosť nápomocné. Avšak využitím časovo-priestorovej súdržnosti na mape chodidiel spojených komponentov umožňuje zachytiť všetky ostatné problémy.

2.2.4. Počítanie osôb

Ako už bolo spomenuté, vzhľad polôh škvŕn na mape chodidiel sa mení vzhľadom na dynamiku scény a to zo snímky na snímku. Pri preplnených scénach ľuďmi v niektorých prípadoch nie je možné jednoducho spočítať pohybujúce sa škvrny, ktoré pozostávajú z osoby. To je prípad, kedy osoby kráčajú blízko seba. Pre vyššiu presnosť je uvedený kontrolný mechanizmus pre dočasnú identitu škvrny. Pred spracovaním b-tej škvrny je prvá identifikovaná, či už bola počítaná v predchádzajúcich snímkach. Pridá sa kontrolná značka (angl. *control flag*) f_t^b indikujúca počet, ktorý udáva koľkokrát bola škvrna spracovaná a nový počet odhadov je aplikovaný, ak už bola škvrna v predchádzajúcich snímkach F-krát spracovaná:

1. ak kontrolná značka je pod hodnotou F, tak sa zvyšuje, až pokým je počet odhadov nezmenený
2. ak je kontrolná značka väčšia ako F, tak jej hodnota sa zresetuje na hodnotu 1, až pokým sa počet odhadov nevypočíta
3. ak skúmané škvrny nie sú identifikované, kontrolná značka je inicializovaná na hodnotu 1, to znamená, že je škvrna sa považuje za novú škvrnu a počet odhadov sa vypočítava

Hodnota F by mala byť starostlivo vybraná, aby zodpovedala dynamickosti scény. Ak sa vyberie hodnota F príliš vysoká, tak algoritmus je menej citlivý na dynamickosť objektu v scéne. Ak dve osoby, ktoré boli rozdelené, sa stretnú a ich škvrny sa spoja do jednej, tak táto zmena je vnímaná algoritmom

iba ak je *F* vyresetované na hodnotu 1. Ak sa vyberie hodnota *F* príliš malá, tak počet odhadov každej škvrny sa postupne aktualizuje. To môže nastať, ak škvrna iba trochu zmení svoju polohu od jednej snímky k druhej, tak potom to môže rozdielne počítať.

Pre každú škvrnu *b* na mape chodidiel v čase *t* je počet odhadov pre celú snímku daný sumou počtu odhadov C_t pre všetky škvrny B_t, dané rovnicou

$$C_t = \sum_{b=1}^{B_t} C_t^b \tag{2-8}$$

K spočítaniu počtu odhadov škvŕn C_t^b v rovnici (2-8) je navrhnutý postup založený na dohliadaní klasifikácie, kde každá škvrna je považovaná za jeden element, ktorý patrí k jednej preddefinovanej triede polôh. Každá trieda popisuje prototyp, ktorý zahŕňa určitý počet osôb. V navrhnutom postupe počty odhadov štatisticky odhadujú hustotu povrchu každej škvrny mapy chodidiel na základe jej informácie o popredí.

Presnosť analýzy škvŕn mapy chodidiel pre presnosť počítania ľudí pochádza z jeho hustoty. To znamená, že objekty môžu byť klasifikované, tým že patria do určitého polohového prototypu. Škvrny mapy chodidiel sú klasifikované na základe sady preddefinovanej polohy prototypov a na analýze vlastností týkajúcich sa plochy škvrny. Analyzuje sa iba tvar škvrny mapy chodidiel a nie jej chromatické vlastnosti. Sú vybraté vlastnosti ako je plocha škvrny a obvod.

Nech *S* je sada známych pozorovaní na vhodnom počte vzoriek, kde škvrny majú rozdielne dimenzie a tvary. Možné triedy škvŕn k detekcie sú definované ako podmnožina prirodzených čísel *k*. $C \equiv \{c_1, c_2, \dots, c_k\} \subset \mathrm{N}$. Prípravná sada majúca kardinalitu *V* pozostáva z páru hodnôt *V* - pozorovanie a trieda. Je definovaná ako $TS \equiv \{(X_1, c_1), (X_2, c_2), \dots, (X_V, c_V)\}$, kde $X_i \in S$ a $c_i \in C, \forall i$. Počas procesu, kedy vyhodnotenie nad vlastnosťami škvŕn vyhodnotí dostatočne vysoký počet vzoriek pre vytvorenie prípravnej sady *TS*, v každom čase *t* algoritmus pozostáva z konštrukcie vektora vlastností (plocha škvrny a obvod) pre každú škvrnu *b* v súčasnej mape chodidiel. Sada je klasifikovaná porovnávaním vektora vlastností s týmito vzorkami prototypov. Odpovedajúca popisujúca hodnota počtu odhadov C_t^b. Klasifikácia je získaná *k*-tým *NN* (skratka z angl. *nearest neighbors*) klasifikátorom, kde nová vzorka prislúcha je klasifikovaná väčšinovým hlasom jeho susedov. Vzorka je priradená ku triede k jeho *k* najbližším susedom. Presnosť takéhoto postupu je dosť vysoká, hoci prípravný set musí byť vytvorený z reprezentatívnych prototypov.

V každej snímke I_i sekvencie v čase *t* počet C_t^b osôb, ktorý môže patriť do *b*-tej škvrny súčasnej mapy chodidiel *FM*, môže byť štatisticky vypočítaný na základe hustoty plochy škvrny. Taký počet môže byť odhadnutý zo vzťahu

$$C_t^b = \frac{A_t^b}{HAA} \tag{2-9}$$

kde A_t^b indikuje *b*-tu plochu škvrny a *HAA* (skratka z angl. *Human Average Area*) reprezentuje priemernú plochu obývanú jednou osobou na scéne podlahy. Výber hodnoty pre *HAA* je veľmi dôležitý

pre počítanie. Aby sa štatisticky definoval dôveryhodný interval pre počet odhadov zo vzťahu (2-9), tak sa vhodný počet jednoduchých škvŕn získaných z rozdielnych pozícií a časov generalizuje toľkokrát, koľkokrát je to možné. Predpokladá sa Gaussova distribúcia vybraných vzoriek, kde hodnota pre *HAA* sa berie ako priemerná hodnota vypočítaná pre jednoduché vzorky škvŕn získaných pre počet odhadov vo vzťahu (2-9) a potom dôveryhodný interval je

$$\left[\frac{A_t^b}{(HAA + 2\sigma)}, \frac{A_t^b}{(HAA - 2\sigma)}\right] \tag{2-10}$$

kde σ štandardná odchýlka vypočítaná pre vybrané vzorky.

Spomenutý popis je trochu jednoduchý, ale sotva spoľahlivý, pretože

- presnosť hodnoty *HAA* je silne podmienená výberom vzoriek jednoduchých škvŕn
- výpočet hustoty plochy nie je vhodná informácia pri starostlivom počítaní koherencie medzi počítaniami škvŕn, pretože mapa chodidiel zvykne neočakávane meniť svoj aspekt
- v prípade ak škvrny zahŕňajú viacerých ľudí, tak vzdialenosť medzi nimi môže mať za následok nesprávny výpočet hustoty

3. Použité metódy počítačového videnia

Pri návrhu systému sú prebrané niektoré už vymyslené metódy detekcie a sledovania. V nasledujúcich kapitolách sú bližšie popísané.

3.1. Detekcia objektov využitím Houghovej transformácie

Metóda detekcie s upravenou Houghovou transformáciou [12] slúži pre detekciu viacerých objektov. Táto navrhnutá metóda je spomenutá v článku, v ktorom je zhrnutá práca detekcie behom desiatich rokov [13]. Jej princíp je v nasledujúcich častiach bližšie popísaný, ale najskôr je spomenuté niečo o Houghovej transformácii.

3.1.1. Houghova transformácia v detekcii objektu

Houghova transformácia [12] je jedna z klasických techník počítačového videnia. Bola určená pre identifikáciu priamok v obraze, neskôr bola rozšírená o určenie pozície ľubovoľných tvarov, najčastejšie kruhov alebo elíps. Klasická Houghova transformácia a jej moderné varianty prevádzajú vstupný obraz na novú reprezentáciu nazvanú Houghov obraz, ktorý sa nachádza v doméne nazvanej Houghov priestor. Každý bod v Houghovom priestore zodpovedá hypotéze o objekte záujmu, ktorý sa nachádza v pôvodnom obraze na konkrétnom mieste a s konkrétnou konfiguráciou. Metóda , ktorá využíva Houghovu transformáciu, pracuje na rozdelení vstupného obrazu na sadu takzvaných hlasovacích elementov. Každý hlasovací element hlasuje pre hypotézy, ktoré by mohli generovať tento element. Samozrejme hlasovacie elementy nedokazujú presnú polohu, a teda ich hlasy môžu byť narušené ďalšími inými hypotézami z Houghovho priestoru. Vysoké hodnoty hlasov prislúchajúce hypotézam môžu generovať hlasovacie elementy s vysokou pravdepodobnosťou. Hlasy z pixelov hlasovacích elementov sú pridané do Houghovho obrazu. Objekty záujmu sú detekcie vrcholov v Houghovom obraze. Výška vrcholu určuje presnosť detekcie.

Houghova transformácia je odolná voči deformáciám, voči šumom obrazu a voči niektorým oklúziám. Napriek týmto výhodám Houghovej transformácii chýba konzistentný model pravdepodobnosti. Vedie to k teoretickým a k praktickým problémom. Z teoretického hľadiska to vedie k neuspokojivým výsledkom hlasovacích elementov. Napríklad, ak nastane situácia, kedy maximálne hodnoty v Houghovom obraze zodpovedajú správnej detekcii objektu a hlasovacie elementy boli generované týmto objektom, tak tieto elementy budú pravdepodobne dávať vysoké hodnoty hlasov pre objekt detekcie, ale taktiež budú hlasovať pre iné hypotézy a sila týchto falošných hodnôt hlasov žiadnym spôsobom nezabráni hypotézam, ktoré potvrdzujú existenciu už existujúcich hlasovacích elementov. Z praktického hľadiska sa heuristiky, ktoré nepotlačujú maximá (angl. *non-maximum*

suppression), používajú v detekcií, aby lokalizovali vrcholy v Houghovom obraze. Tieto heuristiky typicky zahrňujú špecifikáciu a optimalizáciu niekoľkých parametrov.

Navrhnutá metóda detekcie s upravenou Houghovou transformáciou má zopár výhod oproti klasickej Houghovej transformácie:

- vykoná detekciu viacerých objektov vďaka optimalizácii energie
- výsledky experimentov ukazujú vyššiu presnosť pre obrazy, na ktorých sa nachádza viac objektov záujmu
- jeho pravdepodobnostná podstata umožňuje ľahšiu integráciu s inými metódami

Nevýhodou navrhnutej metódy je dlhšia doba výpočtu, avšak stále môže konkurovať iným moderným technikám.

3.1.2. Analýza Houghovej transformácie

Pozorovania obrazu sú v tvare N hlasovacích elementov. Tieto elementy zodpovedajú napríklad pixelom mapy hrán alebo bodom objektu záujmu. Houghov priestor H, kde každý bod $h \in H$ zodpovedá hypotéze o prítomnosti objektu záujmu na konkrétnej polohe. Potom súčasťou detekcie je hľadanie konečných podmnožín v Houghovom priestore zodpovedajúce objektom v obraze. Formálne povedané, pre každú hypotézu $h \in H$ je zavedená binárna hodnota y_h, ktorej hodnota nadobudne 1, ak hypotéza zodpovedá objektu, inak nadobudne hodnotu 0.

Houghova transformácia vykoná detekciu pomocou každého hlasovacieho elementu i a generovaním objektu h. Formálne povedané, že je uvedené hodnota x_i, ktorá rozšíri Houghov priestor $H' = H \cup \{0\}$. Priradenie $x_i = h \in H$ implikuje, že hlasovací element i je generovaný objektom h, pokiaľ $x_i = 0$ implikuje element i, ktorý nie je súčasťou žiadneho objektu záujmu. Je možné považovať hlasy ako pseudohustoty $V(x_i = h|I_i)$ v Houghovom priestore. Pseudohustoty sú podmienené deskriptorom I_i hlasujúceho elementu i. Deskriptor zahrňuje pozíciu v priestore a lokálnu informáciu o vzhľade obrazu. Tieto podmienené hustoty sú pridané výpočtu a vrcholy výsledného výpočtu sa považujú za správne hypotézy.

Pridávanie hustôt nemá žiadny pravdepodobnostný význam. Pokiaľ niektorá skúsi interpretovať hlasy ako logaritmus pravdepodobností $V(x_i = h|I_i) = \log p(x_i = h|I_i)$. Pri logaritme pravdepodobnosti hustôt sú hlasy typicky pozitívne s limitovaným rozsahom a v metódach, založených na poznávaní obrazu, sú hlasy brané ako neparametrická pravdepodobnosť hustôt a nie ako logaritmy pravdepodobností.

Ak je ignorovaný predchádzajúci odsek, je možné vidieť Houghove hlasy ako logaritmus $p(x_i = h|I_i)$. Potom ich súčet odpovedá vyhodnoteniam, ktorých rozdelenia nad hypotézami generujú nezávislé hlasovacie elementy. Tieto nezávislé domnienky nie sú úplne spracované. Ak hlasovacie elementy i a j susedia v obraze, potom samozrejme ich hypotézy silne pôsobia na seba a je veľmi pravdepodobné že sú generované rovnakým objektom. Nepotlačenia

maxima (angl. *non-maximum suppression*), ktoré sa bežne vykonávajú v Houghovej transformácii, môžu byť považované za trik ako kompenzovať hranice nezávislosti pravdepodobností. Neskôr sa ukáže, že nepotlačovanie maxima nebude potrebné.

3.1.3. Pravdepodobnosť v tejto metóde

Houghove hlasovanie stavia na fakte, že hlasy môžu byť ľahko definované alebo získané z dát. Najistejšia možnosť je, keď Houghova transformácia vezme najľahšiu a najrýchlejšiu cestu hlasov a jednoducho ju vypočíta.

Táto metóda obchádza spomenutú najistejšiu možnosť. Radšej než sú vypočítané hlasy, je modelované spojité rozdelenie nad všetkými náhodnými hodnotami x a y pravdepodobnostnou metódou. Je možné rozhodnúť o ich hodnotách procesom, ktorý nepotlačuje maximum. Je modelované spájanie x a y, ktoré sa nachádzajú v obraze, kde obraz znamená kolekciu hlasujúcich elementov. Je aplikovaný Bayesov teorém. Teraz je upriamená pozornosť na pravdepodobnostné a priorné výrazy.

3.1.4. Pravdepodobnostné výrazy

Sú vytvorené rozdielne nezávislé predpoklady, aby sa manipulovalo s pravdepodobnostným výrazom. Predpokladá sa, že existujú objekty y a hypotézy priradení x, potom rozdelenia vzhľadov jednotlivých hlasovacích elementov sú nezávislé.

Predpokladá sa, že deskriptor I_i hlasujúceho elementu i závisí iba na objekte priradenia x_i a je podmienečne nezávislý na priradeniach zvyšných hlasovacích elementov a na existencii všetkých ostatných objektov v obraze.

Na prvý pohľad tieto predpoklady vyzerajú trochu nedokončene, pretože vzhľad elementov sa predpokladá byť závislý iba na hypotézach x_i a podmienene nezávislé z ostatných hlasovacích elementov a hypotéz. Avšak v tejto závislosti stále môže byť zahrnutá relatívna pozícia elementu i a objektu zodpovedajúceho hypotéze x_i. Napríklad výraz v prípade detekcie auta môže modelovať vzhľad hlasovacích elementov ako náhodnú hodnotu závislú na časti auta, z ktorej pochádza. Na koniec pri derivácií časti pravdepodobnosti sa ešte jeden krát použije Bayesov teorém a výrazy, ktoré sú konštantné, sa preskočia.

Výraz pre pravdepodobnosť je veľmi podobný k Houghovej transformácii ako dátovo-závislému výrazu, ktorý produkuje výrazy súvisiace s Houghovmi hlasmi.

3.1.5. Priorné výrazy

Pred formuláciou priorneho rozdelenia je nutné dodať, že nie všetky konfigurácie sú správne. Ak hlasujúci element i je priradený k hypotéze, ktorá nie je súčasťou pozadia obrazu, potom hypotéza h zodpovedá existujúcemu objektu. To znamená, že y_h musí byť 1 a konfigurácia sa považuje za správnu

iba ak $y_{x_i} = 1$, pre všetky x_i. Aby sa vyhlo k priradeniam pozadia k obrazu, kde $x_i = 0$, tak je ako špeciálny prípad uvedená hypotéza hodnoty y_0, ktorá je vždy nastavená na 1.

V tejto metóde sa sústredilo na veľmi všeobecný prior hodnoty y, ktorá jednoducho penalizuje počet aktívnych hypotéz, ktorá preferuje vysvetlenie scény s malými objektami. Je možné použitie rôznych prístupov k určeniu priorneho rozdelenia. Jeden z prístupov je aj Pottovo priorne rozdelenie, ktoré je nepriaznivé pre tenké objekty, čiary.

Je definovaný prior, ktorému je priradená hodnota pravdepodobnosti rovnej 0 ku všetkým konfiguráciám y, kde existuje pár možných hypotéz, ktoré príliš prekrývajú ohraničené boxy. Avšak takémuto prioru sa vyhne. To vytvorí kontrast nášho prístupu voči klasickému nepotlačeniu maxima a umožní detekciu aj silne prekrývajúcich sa objektov.

3.2. Metóda sledovania, poznávania a detekcie

Metóda *TLD* (skratka z angl. *tracking-learning-detection*) [14] je metóda navrhnutá pre dlhodobé sledovanie neznámeho objektu vo videu. Komponenty tvoria sledovač (angl. *tracker*), detektor a poznávanie (angl. *learning*). Sledovač odhaduje pohyb objektu medzi po sebe idúcimi snímkami za predpokladu, že pohyb medzi snímkami je konečný a objekt je viditeľný. Sledovač pravdepodobne zlyhá, ak sa objekt stratí z dohľadu. Detektor nezávisle spracúva každú snímku a skenuje celý obraz, aby lokalizoval všetky vzhľadové vlastnosti, ktoré boli objavené a poznané predtým. Ako každý detektor vytvára falošné pozitíva a falošné negatíva. Poznávanie si všíma prácu sledovača a detektora a odhaduje chyby detektora. Pri odhaľovaní chýb generuje príklady, podľa ktorých sa potom vyhne chybám pri odhaľovaní v budúcnosti. Komponent poznávania predpokladá, že detektor i sledovač môžu zlyhať. K posilneniu poznávania generuje detektor viac vzhľadových vlastností objektu a rozlišuje ich od pozadia.

3.2.1. Sledovanie objektu

Sledovanie objektov je o dobrom odhade pohybu objektu. Počas sledovania sa vyhodnocuje viditeľnosť objektu. Sústreďuje sa na metódy, ktoré reprezentujú objekty ako geometrické tvary a ich pohyb odhadnutý zo snímok, ktoré idú za sebou. Sledovanie pomocou šablón je najlepšou voľbou v tomto prípade. Objekt je vyjadrený ako cieľ šablóny, medzi ktoré patrí napríklad cesta obrazu alebo farba histogramu. Pohybom je definovaná transformácia, ktorá minimalizuje nezhody medzi cieľovou šablónou a kandidátom na opravu. Šablóna sledovania môže byť zrealizovaná ako statická, keď cieľ šablóny sa nemení, a ako adaptívna, keď cieľ šablóny je oddelený od predchádzajúcej snímky. Tieto metódy je možné skombinovať. Šablóny majú obmedzené možnosti modelovania, keď reprezentujú jednu vzhľadovú vlastnosť objektu. K modelovaniu viacerých vzhľadových variácií vzniká generatívny model. Generatívne sledovače modelujú iba vzhľadové vlastnosti objektu a zlyhávajú na preplnenom pozadí. Aby sa vyhlo tomuto problému, tak sledovače modelujú taktiež prostredie, v ktorom sa objekt

pohybuje. Často sa používajú dva návrhy k modelovaniu prostredia. Prvý návrh pozostáva z prostredia, v ktorom sa vyhľadáva podpora pre pohybujúci sa objekt, ktorý je vo vzťahu s objektom záujmu. Druhý návrh pozostáva z prostredia, ktoré sa považuje za negatívnu triedu, voči ktorej je sledovač rozdielny. Obecným prístupom selektívnych sledovačov je vytvorenie binárneho klasifikátora. Binárny klasifikátor reprezentuje hranicu, ktorá rozhoduje medzi objektom a jeho pozadím. Statické selektívne sledovače vytvárajú klasifikátor objektu pred sledovaním. To ich obmedzuje len pre sledovanie známych objektov. Adaptívne selektívne sledovače vytvárajú klasifikátor počas sledovania. Zásadnou fázou adaptívnych selektívnych sledovačov je aktualizácia. Blízke susedné oblasti súčasnej pozície sú použité k vytvoreniu pozitívnych príkladov a vzdialené body sú použité k vytvoreniu negatívnych príkladov. Tieto príklady sa používajú k aktualizácií každej snímky. Táto stratégia zvláda významne zmeny vzhľadových vlastností a taktiež zvláda krátkodobé oklúzie a preplnené pozadie. Avšak stále je možné zlyhanie sledovania, ak sa objekt stratí na dlhšiu dobu ako sa predpokladalo.

3.2.2. Detekcia objektu

Detekcia objektu je o lokalizovaní objektu vo vstupnom obraze. Pojem objekt má rôzne definície. Môže to byť jedna instancia alebo celá trieda objektov. Metódy detekcie objektu sú typicky založené na lokálnych vlastnostiach obrazu alebo sú založené na pohybujúcom sa okienku. Metódy založené na lokálnych vlastnostiach obrazu pracujú podľa nasledujúcich krokov: detekcia vlastnosti, rozoznanie vlastnosti a nasadenie vlastnosti do modelu. Metódy založené na pohybujúcom sa okienku skenujú vstupný obraz s rozdielnymi veľkosťami okienka a pre každé okienko sa rozhodne, či zistené zásadné zmeny objavili objekt záujmu alebo nie. K prispetiu k výkonu prevzali tieto detektory takzvanú kaskádovú architektúru. Pozadie je omnoho viac frekventované než objekt a klasifikátor je rozdelený do viacerých stupňov, ktoré dovoľujú skoré odmietanie opráv pozadia, a teda redukujú počet stupňov. Skúšanie takýchto detektorov vyžaduje veľké množstvo príkladov a intenzívne počítanie, aby bolo zabezpečená presnosť v rozhodovaní medzi objektom a pozadím.

3.2.3. Proces poznávania

Detektory objektov sú dobre pripravené detekovať, ak všetky prípravné príklady boli pred detekciou dobre popísané. Takúto situáciu si nie je možné priať, ak detektor sa pripravil z jedného popísaného príkladu alebo z videa. Je nutné nájsť riešenie. Typicky sa predpokladajú nezávislé a identicky rozdelené dáta s určitými vlastnosťami, akými sú nepopísané formy príkladov z pôvodných zhlukov z priestoru vlastností. Sú využité algoritmy očakávania maximalizácie (angl. *Expectation-Maximalization*), sebapoznávania (angl. *Self-learning*) a spoločnej prípravy (angl. *Co-training*).

3.2.4. P-N poznávanie

P-N poznávanie (angl. *P-N Learning*) je súčasťou poznávacieho komponentu. Cieľom tohto komponentu je zlepšiť výkon detektora. V každej snímke, ktorú detektor bude ohodnocovať, sa identifikujú jej chyby a v budúcnosti sa im vyhne. Kľúčom P-N poznávania je, tie chyby detektoru sú identifikované dvoma typmi „expertov“. P-expert identifikuje iba falošné negatíva a N-expert identifikuje iba falošné negatíva. Oba typy expertov získajú chyby, avšak sú od seba nezávislé a to môže spôsobiť vzájomný vplyv ich chýb.

Nech x je príklad z priestoru vlastností X a y je popisom z priestoru vlastností $Y = \{-1,1\}$. Sada príkladov X sa považuje za nepopísanú sadu, Y sa považuje za sadu popisov a $L = \{(x, y)\}$ je popísaná sada. Vstup P-N poznávania je popísaná sada L_l a nepopísaná sada X_u, kde $l \ll u$. Úlohou P-N poznávania je, aby sa klasifikátor $f: X \rightarrow Y$ naučil z popísanej sady L_l a z nepopísanej sady X_u.

P-N poznávanie pozostáva zo štyroch blokov, medzi ktoré patrí:

- klasifikátor, ktorý bude poznávať
- prípravná sada, čo je kolekcia popísaných prípravných príkladov
- metóda, ktorá pripravuje klasifikátor z prípravnej sady
- P-N experti, čo sú funkcie, ktoré generujú pozitívne a negatívne prípravné príklady počas poznávania

Prípravný proces je inicializovaný vložením popísanej sady L do prípravnej sady. Prípravná sada prejde metódou, ktorá pripraví klasifikátor. Proces poznávania postupuje iteratívnym zavádzaním. Každá iterácia pripravuje klasifikátor. Klasifikácia je analyzovaná P-N expertami, ktorí odhadnú príklady, ktoré boli nesprávne klasifikované. Tieto príklady sú pridané so zmenou popisu do prípravnej sady. Na konci iterácie sa pozmení príprava klasifikátora. Proces iteruje pokiaľ konverguje alebo pokiaľ nedôjde k inému kritickému bodu.

Zásadným elementom P-N poznávania je odhadnutie chýb klasifikátora. Kľúčom je oddeliť odhad falošných pozitívov od odhadu falošných negatívov. Preto nepopísaná sada je rozdelená na dve časti založených na súčasnej klasifikácie a každá časť je analyzovaná nezávislým expertom. P-expert analyzuje príklady klasifikované ako negatíva, odhaduje falošné negatíva a pridáva ich do prípravnej sady s pozitívnym popisom. V iterácií k, P-expert produkuje $n^+(k)$ pozitívne príklady. N-expert analyzuje príklady klasifikované ako pozitíva, odhaduje falošné pozitíva a pridá im negatívny popis do prípravnej sady. V iterácií k, N-expert produkuje $n^-(k)$ negatívne príklady. P-expert zvyšuje všeobecnosť klasifikátora. N-expert zvyšuje rozpoznateľnosť klasifikátora.

3.3. Metóda on-line boosting

Metóda boosting má široké uplatnenie pri úlohách poznávania obrazu a našla uplatnenie pri riešení problémov počítačového videnia. Je uvedené on-line boosting sledovanie [15], ktoré využíva špecifický

boosting algoritmus: diskrétny algoritmus AdaBoost (angl. *adaptive boosting)* pre klasifikáciu [16]. Najprv je predstavené off-line boosting sledovanie, ktoré sa využíva pre výber vlastností.

3.3.1. Off-line boosting

Boosting metóda sa vo všeobecnosti využíva pre zlepšenie presnosti akéhokoľvek poznávacieho algoritmu. To je dosiahnuté kombináciou *N* hypotéz, ktoré sú vygenerované tak, že sa opakujú prípravné sady s tým, že za každým sa podmnožina prípravných dát zmení. Boosting metóda transformuje algoritmus slabého poznávania do silného.

Termíny, ktoré sú v článku použité sú slabý klasifikátor a silný klasifikátor. Slabý klasifikátor pracuje o trochu lepšie ako náhodné hádanie. Napríklad, pre binárnu úlohu rozhodovania musí byť stupeň chyby menši než 50%. Hypotézy h^{weak} sú generované slabým klasifikátorom, ktorý bol vytvorený aplikovaním poznávacieho algoritmu. Ak je daných *N* slabých klasifikátorov, tak silný klasifikátor h^{strong} je vypočítaný ako lineárna kombinácia týchto slabých klasifikátorov.

Sústreďuje sa na diskrétny algoritmus AdaBoost, ktorý preváži prípravné vzorky namiesto toho, aby ich prevzorkoval. Algoritmus pracuje nasledovne:

1. je daná prípravná sada $X = \{\langle x_1, y_1\rangle, \dots, \langle x_L, y_L\rangle \mid x_i \in \mathbb{R}^m, y_i \in \{-1, +1\}\}$ s pozitívnymi a negatívnymi vzorkami a počiatočnou uniformnou distribúciou $p(x_i) = \frac{1}{L}$ nad príkladmi
2. slabý klasifikátor h^{weak} je pripravovaný na cvičnej sade *X* a na počiatočnej uniformnej distribúcií *p(x).*
3. od chyby $e_n$slabého klasifikátora h_n^{weak} záleží priradenie váhy $a_n = \frac{1}{2} * \ln\left(\frac{1-e_n}{e_n}\right)$
4. počiatočná uniformná distribúcia *p(x)* je aktualizovaná tak, že sa pravdepodobnosť zvýši pre vzorky, ktoré sú nesprávne kvalifikované a ak je vzorka klasifikovaná správne, tak odpovedajúca váha je znížená, a preto sa algoritmus sústreďuje na zložité vzorky
5. proces sa opakuje a pri každej boosting iterácií sa pridá nová slabá hypotéza, pokiaľ sa nenarazí na určitú podmienku (napríklad, počet slabých hypotéz už je dostačujúci)

3.3.2. Off-line boosting pre výber vlastností

Základom je to, že každá vlastnosť zodpovedá jednotlivému slabému klasifikátoru a boosting metóda vyberie z týchto vlastností. Ak je daná veľká množina vlastností, tak z dôvodu dlhého výpočtu sa algoritmus zameriava len na podmnožinu vlastností.

Príprava sa deje podobne ako pri štandardnej boosting metóde. V každej iterácií *n* algoritmus vyberie jednu novú vlastnosť a pridá ju (s odpovedajúcim hlasovacím faktorom) k množine. Všetky vlastnosti sú vyhodnotené a najlepšia z nich je vybraná pre formovanie hypotézy h_n^{weak}. Váha α_n je nastavená podľa prislúchajúcej chyby hypotézy h_n^{weak}. Nakoniec je silný klasifikátor h^{strong}

vypočítaný ako váhová lineárna kombinácia slabých klasifikátorov. Ako už bolo spomenuté, cvičná chyba exponenciálne klesá nad boosting iteráciami, čo je ekvivalentné k počtu vybraných vlastností.

3.3.3. On-line boosting

Predchádzajúce opísané kroky je nutné vykonať on-line. On-line aktualizácia slabých klasifikátorov nie je problematická, pretože môžu byť použité on-line boosting algoritmy pre generovanie hypotéz. Takisto sa aj odhad zodpovedajúci hlasovacím váham neodchyľuje, ak je známy odhad chyby slabých klasifikátorov. Kľúčových krokom je výpočet váhovej distribúcie pre vzorky, pretože nie je známa priorna zložitosť vzorky.

Základnou myšlienkou je, že dôležitosť(rozdielnosť) zložky sa odhadne aplikáciou na sadu slabých klasifikátorov. Je to modelovanie získaním informácie, ktorá rešpektuje prvých *n* klasifikátorov a kóduje s dôležitosťou váhy λ pre aktualizáciu *n+1*-tého slabého klasifikátora.

On-line algoritmus vyžaduje, aby bol na začiatku pevný počet slabých klasifikátorov. V prípade off-line algoritmu sú využité k aktualizácií jedného slabého klasifikátoru, pokým v prípade on-line algoritmu jedna vzorka je využitá k aktualizácií všetkých slabých klasifikátorov a zodpovedajúcej hlasovacej váhe.

3.3.4. On-line boosting pre výber vlastností

Je daná sada *M* slabých klasifikátorov s hypotézami $H^{weak} = \{h_1^{weak}, \dots, h_M^{weak}\}$. Selektor vyberie jednu z nich, čo je vyjadrené vzťahom

$$h^{sel}(x) = h_1^{weak}(x) \tag{3-1}$$

kde *m* sa vyberie podľa kritéria optimalizácie. Odhadnutá chyba každého slabého klasifikátora $h_i^{weak} \in H^{weak}$ je vyjadrená vzťahom $m = \arg min_i e_i$.

Selektor môže byť interpretovaný ako klasifikátor. Príprava selektoru je, že každý slabý klasifikátor je aktualizovaný a najlepší je vybraný. Podobne u off-line prípadu, kedy slabé klasifikátory H^{weak} zodpovedajú vlastnostiam.

Hlavnou myšlienkou je neaplikovať on-line boosting metódu priamo na slabé klasifikátory, ale na selektory. Pevná sada *N* selektorov $h_1^{sel}, \dots, h_N^{sel}$ je náhodne inicializovaná. Keď príde nová prípravná vzorka do selektorov, tak je aktualizovaná. Táto aktualizácia rešpektuje dôležitosť váhy λ v súčasnej vzorke. Pre aktualizáciu slabých klasifikátorov je zavedená štandardná EM technika, aby sa odhadlo pravdepodobnostné rozloženie pozitívnych a negatívnych vzoriek a generovanie hypotéz. Slabý klasifikátor s najmenšou chybou je vybraný selektorom rovnicou

$$\arg\min_{\mathrm{m}}(e_{n,m}), e_{n,m} = \frac{\lambda_{n,m}^{wrong}}{\lambda_{n,m}^{corr} + \lambda_{n,m}^{wrong}} \tag{3-2}$$

kde $e_{n,m}$ je chyba m-tého slabého klasifikátora $h_{n,m}^{weak}$ v n-tom selektore odhadnutého z váh správne $\lambda_{n,m}^{corr}$ a nesprávne $\lambda_{n,m}^{wrong}$ klasifikovaných príkladov. Nakoniec, odpovedajúca hlasovacia váha α_n a dôležitosť váhy λ vzorky sú aktualizované a prenechané ďalšiemu selektoru h_{n+1}^{sel}.

Silný klasifikátor je získaný lineárnou kombináciou selektorov, čo je dané rovnicou

$$h^{strong}(x) = sign\left(\sum_{n=1}^{N} \alpha_n * h_n^{sel}(x)\right) \tag{3-3}$$

V porovnaní s off-line verziou je klasifikátor dostupný v ktoromkoľvek čase.

3.4. Subtrakcia pozadia

Pre detekciu objektov na obraze bolo nutné zistiť miesto pohybu na obraze a preto bola využitá subtrakcia pozadia (angl. *Background Subtraction*) [17]. Detekcia rušivých objektov je podstatným krokom pri analýze scény. Rušivé objekty sú objekty, bez ktorých majú obrazy v scéne určité zvyčajné správanie, ktoré je popísané štatistickým modelom. Ak je vytvorený štatistický model, tak rušivé objekty je možné detekovať tým, že v obraze sa vyskytnú lokality týchto objektov, ktoré nezapadajú do modelu. Tento proces sa nazýva subtrakcia pozadia.

Je aplikovaný prístup zdola-nahor a funkcia počíta hustotu pravdepodobnosti zvlášť pre každý pixel. Pixel z nového obrazu je považovaný za pixel pozadia, ak jeho nová hodnota je popísaná funkciou hustoty. Pre statickú scénu by mohol byť najjednoduchším modelom obraz scény bez rušivých objektov. Ďalším krokom by bol napríklad odhad vhodných hodnôt pre zmeny intenzity úrovní pixelov z obrazu, pretože zmeny sa môžu diať z pixela na pixel. Je použitý Gaussov model zmesi (angl. *Gassian Model Mixture*).Je vylepšený tým, že konštante sú adaptované pre každý pixel nielen parametre, ale aj počet komponentov zmesí. Tým, že je vybraný počet komponent pre každý pixel v on-line procese, sa algoritmus môže automaticky plne adaptovať do scény.

4. Návrh systému

Počítanie ľudí pozostáva z ich detekcie na obraze a následne z ich sledovania na obraze. Detekcia osôb vychádzala z metódy popísanej v podkapitole 3.1 a z metód, ktoré popisujú nasledujúce časti podkapitol. Detekcie osôb je možné zhrnúť do nasledujúcich 3 spôsobov:

- detekcia na základe pohybu objektov na obraze
- detekcia osôb s využitím detektoru
- kombinácia predchádzajúcich dvoch spôsobov

Po detekcií sú využité metódy sledovania popísané v podkapitolách 3.2 a 3.3.. Konkrétny návrh pre môj systém je popísaný v tejto kapitole.

Nasledujú podkapitoly venujúce sa metódam detekcie. U každej platí, že nájdené detekcie uložia číslo snímky, na ktorej detekcia prebehla a taktiež uložia veľkosť a polohu detekcie.

4.1. Detekcia na základe pohybu objektu na obraze

Po aplikovaní subtrakcie pozadia sa na obraze vyskytnú rušivé objekty. Tieto rušivé objekty predstavujú potenciálne osoby. Potenciálna osoba je ohraničená boxom, ktorý sa uloží. Taktiež rušivé objekty môže spôsobiť vietor, ktorý hýbe predmetmi, dážď alebo tieň, respektíve vniknuté svetlo do scény. Tieto javy môžu vzniknúť pri otvorení alebo zatvorení dvier a podobne. Pre ošetrenie týchto javov je nutné ohraničujúce boxy obmedziť veľkosťou. Veľkosť musí zodpovedať veľkosti osoby, teda výška boxu musí byť dostatočná. Boxy s malými veľkosťami sú ignorované.

4.2. Detekcia osôb detektorom

V systéme je použitý detektor z OpenCV[1], ktorý je založený na detekcií od P.Viola a M.J.Jones [18] a detektor využívajúci upravenú Houghovu transformáciu [12], ktorý je bližšie popísaný v podkapitole 3.1. Detektor je aplikovaný na snímku po snímke a sú získané detekcie osôb, ktoré sa uložia. Opäť sú ignorované detekcie s malými veľkosťami.

4.3. Kombinácia detekcií

Kombináciou detekcie na základe pohybu na obraze a následnej detekcie osôb je efektívnou metódou. Ako bolo spomenuté v predchádzajúcej podkapitole 4.1, tak aj tu prebehla detekcia na základe pohybu na obraze. Nájdené rušivé objekty s optimálnou veľkosťou boli ohraničené boxami. Ohraničené boxy sa rozšíria, pre lepšiu ďalšiu detekciu detektorom. Rozšíri sa šírka a výška boxu, ale takisto sa x súradnica boxu posunie vľavo a y súradnica boxu sa posunie zasa nahor. V takomto širšom ohraničenom

[1] Dostupné z opencv.org

boxe má lepšie podmienky detektor detekovať osoby. Po aplikácií detektora vzniknú opäť ohraničené boxy, ktoré predstavujú nájdené osoby. Systém uloží súradnice x a *y* ohraničených boxov nájdených osôb s posunom ohraničených boxov, ktoré vytvorili detekcie na základe pohybu. Tým sa dosiahne posun na správne miesto na obraze.

4.4. Sledovanie a počítanie osôb

Po získaní detekcií nasleduje na rad sledovanie ohraničených boxov na obraze. Použité sledovače sú popísané v podkapitolách 3.2 a 3.3. Jednotlivý sledovač má svoju konfiguráciu, ktorá je pred sledovaním nastavená. Nastaví sa počiatočná snímka, rozmery počiatočného ohraničeného boxu, názov videa a jeho umiestnenie a nastavenie špecifické pre konkrétny typ sledovača. Sledovač sa spustí a začne sledovanie počiatočného boxu, ktorý bol uložený po detekcií. Sleduje sa pohyb boxu, čo vytvára cestu pohybu a táto cesta sa uloží. Je uložená podobne ako pri detekciách a to vo forme čísla snímky videa a pozície a vo forme veľkosti ohraničeného boxu. Sleduje sa objekt až pokým ho sledovač nestratí. Avšak treba brať do úvahy pár snímok a počkať pred určením, že objekt zmizol z obrazu, pretože sa môže objekt skryť za predmet na scéne alebo za druhý objekt. Ďalším problémom môže nastať, ak sa objekt zastaví. Pri osobách to trvá pár snímok, ale ak detekcia prebehla na základe pohybu, môže sa jednať aj o objekt nepredstavujúci osobu (napríklad tieň alebo svetlo vzniknuté pri otvorení dverí). Tu treba po pár snímkach ukončiť sledovanie.

Sledovač si zapamätá polohu počiatočného ohraničujúceho boxu a najmä hodnotu súradnice *x*. Ak súradnica *x* pri zmiznutí objektu z obrazu má väčšiu hodnotu ako pri počiatočnom boxe, tak to znamená, že objekt prešiel z ľavej strany do pravej. Ak je hodnota menšia, tak prešiel sprava doľava. Táto informácia sa uloží spolu so snímkou, na ktorej objekt zmizol.

Cesta sledovaného objektu sa ukladá a táto postupnosť boxov sa porovnáva s jednotlivými boxami detekcie uložených pred sledovaním pri detekcií. Porovnanie prebieha na základe rovnakej snímky a stredu boxu. Je braný do úvahy väčší rozsah stredu boxu pre možné odchýlky, ktoré môžu spôsobiť sledovače tým, že „trasú“ sledovaným objektom na obraze. Ak sa nájdu zhody pri porovnávaní, tak dané boxy sa vyradia. Ďalšie sledovanie nastáva ďalším zostávajúcim nevyradeným boxom. Po sledovaní je tiež vyradený, aby sa predišlo jeho opätovnému sledovaniu.

5. Implementácia

Podľa návrhu systému opísanom v kapitole 4 sa dá implementácia rozdeliť na 3 časti:

- detektory
- sledovače
- skript spustenia

Úlohou skriptu je spustenie detektorov a sledovačov a následne spracovanie boxov, čo sú ich výstupy zachytené a uložené v textových súboroch. Z videa datasetu je vytvorená sekvencia snímok. Tento dataset spolu s metódami počítačového videnia sú parametrami môjho skriptu. Podľa parametru metódy detekcie skript spustí danú metódu detekcie, ktorá detekuje osoby na frekvencií snímok. Pri detekcií osoby sa uloží číslo snímky a informácie o boxe do textového súboru *output_detector.txt*. Po skončení detekcie skript spracuje detekované boxy z textového súboru a uloží ich do štruktúry *frames_detector*. Daná štruktúra je vlastne pole informácií o jednotlivých boxoch.

Nasleduje sledovanie získaných boxov. Je nutné pred sledovaním nastaviť konfiguračný súbor typický pre jednotlivý sledovač. V konfiguračnom súbore je nastavený aj počiatočný box, čo predstavuje prvý box v poradí štruktúre detekovaných boxov. Spustí sa sledovač a opäť sa uložia boxy vytvorené pohybom objektu po obraze. Tentokrát sa uložia do textového súboru *output_tracker.txt*. Pri správnom pohybe objektu sa uloží číslo snímky a smer vo forme písmena *L* (predstavujúce smer sprava doľava) alebo písmena *R* (predstavujúce smer zľava doprava) do textového súboru pomenovaným podľa prebehnutej metódy detekcie a práve vykonávaným sledovaním. Správnym pohybom sa myslí, keď objekt prejde z jednej strany na druhú a zmizne. Skript opäť spracuje boxy uložením do štruktúry *frames_tracker*.

Po tejto fáze je potrebné vyradiť zo štruktúry detekované boxy, ktoré sú zhodné s boxami zo sledovača. Vyraďovanie nastáva porovnávaním detekovaných boxov v štruktúre so štruktúrou získanou zo sledovača. Porovnávanie a následné vyradenie boxov prebieha vo funkcií *check_boxes*. Po vykonaní tejto funkcie sa pokračuje ďalej v sledovaní s ďalším boxom v poradí až pokým sa štruktúra nevyprázdni.

6. Vyhodnotenie

V tejto kapitole sa budem zaoberať testovaním a vyhodnotením použitých metód v systéme. V prvej podkapitole sa testuje správna detekcia postáv, potom nasleduje testovanie sledovania osôb a v poslednej, najdôležitejšej, podkapitole je vyhodnotenie samotného počítania ľudí s použitím metód v systéme.

Testy prebiehali na datasetoch (Obrázok 6-1), ktoré boli zhotovené stacionárnou kamerou. Bola postavená pri vchode do budovy školy a menzy a pri vchode v domácnosti. Jeden dataset bol zhotovený vonku, kde kamera snímala ľudí prechádzajúcich sa po chodníku.

vchod do školy

vchod do menzy

chodba

chodník

Obrázok 6-1 Ukážky datasetov

6.1. Testovanie detekcie osôb

V tejto podkapitole budem testovať kvalitu detekcie osôb na obraze.

Pre detekciu na základe pohybu bolo potrebné subtrahovať pozadie, aby sa mohli dať detekovať rušivé objekty vytvorené na obraze. Avšak to nestačí pre hladkú detekciu objektov a je nutné po subtrahovaní pozadia erodovať a dilatovať obraz pre lepšiu detekciu najmä väčších objektov záujmu. Po týchto fázach (Obrázok 6-2) detekcia na základe pohybu zlepšila svoju kvalitu a nevytvárala až také veľké množstvo falošných pozitívov.

(a) (b) (c)

Obrázok 6-2 (a) Pôvodná snímka (b) Snímka po subtrakcií pozadia (c) Vyhladená snímka s detekciou osoby na základe pohybu

Pri detektore z knižnice OpenCV bolo nutné vyriešiť minimálnu a maximálnu veľkosť hľadaného objektu. Najlepšou voľbou pre maximálnu veľkosť bola výška celého obrazu. Pri minimálnej veľkosti to už bolo trochu komplikovanejšie. Najprv bola určená minimálnu veľkosť polovica z výšky obrazu, avšak vznikal veľký počet prehliadnutých výskytov - falošných negatívov. Spôsobili to menšie prehliadnuté objekty. Tento problém vyriešilo to, že minimálna veľkosť bola tretina z výšky obrazu.

Detektor s upravenou Houghovou transformáciou, ktorý je opísaný v podkapitole 3.1, sa ukázal ako veľmi pomalý. Pre zrýchlenie bolo nutné zmeniť parametre výšky a šírky boxu v prípravnom procese detekcie. Pri menších rozmeroch bol detektor rýchlejší, avšak detekoval veľké množstvo falošných pozitívov (Obrázok 6-3). Ich množstvo sa obmedzilo jednoducho tým, že ak ich parameter osi *y* bol pod polovicou obrazu, tak sa daná detekcia nebrala do úvahy.

(a) (b) (c)

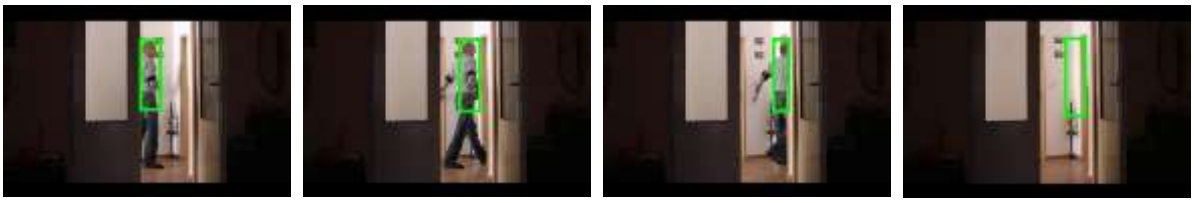
Obrázok 6-3 (a) Ukážka detekcie detektoru s upravenou Houghovou transformáciou s jedným pozitívom a s dvoma negatívami (b) Ukážka detekcie detektoru z knižnice OpenCV s jedným pozitívom a s jedným negatívom (c) Ukážka detekcie detektorov s využitím detekcie pohybu s jedným pozitívom a s oblasťou pohybu.

6.2. Testovanie sledovania osôb

Testovanie sledovania osôb bolo hlavne o dobrom rozhodovaní, či sa objekt už stratil pri sledovaní alebo ešte je stále na obraze. Testovanie prebiehalo pri rôznych počtoch snímok, ktoré určovali dobu čakania na návrat objektu na obraz po strate. Najviac sa osvedčilo čakanie 5 snímok. Pri zvolenom počte 10 dochádzalo už k sledovaní nesprávneho, iného objektu.

Pri obidvoch použitých sledovačoch bolo potrebné zabezpečiť aj správne rozmery počiatočného boxu z detektora. Pri veľkých rozmeroch box nechcel sledovač sledovať, pretože súčasťou boxu zaberalo aj okolie prostredia, ktoré sa menilo. Pri malých rozmeroch zas sledovač nemohol rozoznať postavu z malej plochy a preto sledovač strácal svoj objekt sledovania.

Sledovač pri strate objektu posúval sledovací box v smere osi y a hľadal objekt po obraze. To bolo potrebné ošetriť a boli brané len sledovania, pri ktorých neboli veľké „skoky“ po obraze sledovacieho boxu. Sledovač TLD sa osvedčil a nemuseli byť žiadne odchýlky ošetrené. Obrázok 6-4 zobrazuje sledovanie boxu.



začiatok sledovania | priebeh sledovania | strata objektu | čakanie na možný návrat objektu

Obrázok 6-4 Sledovanie boxu

6.3. Vyhodnotenie počítania ľudí

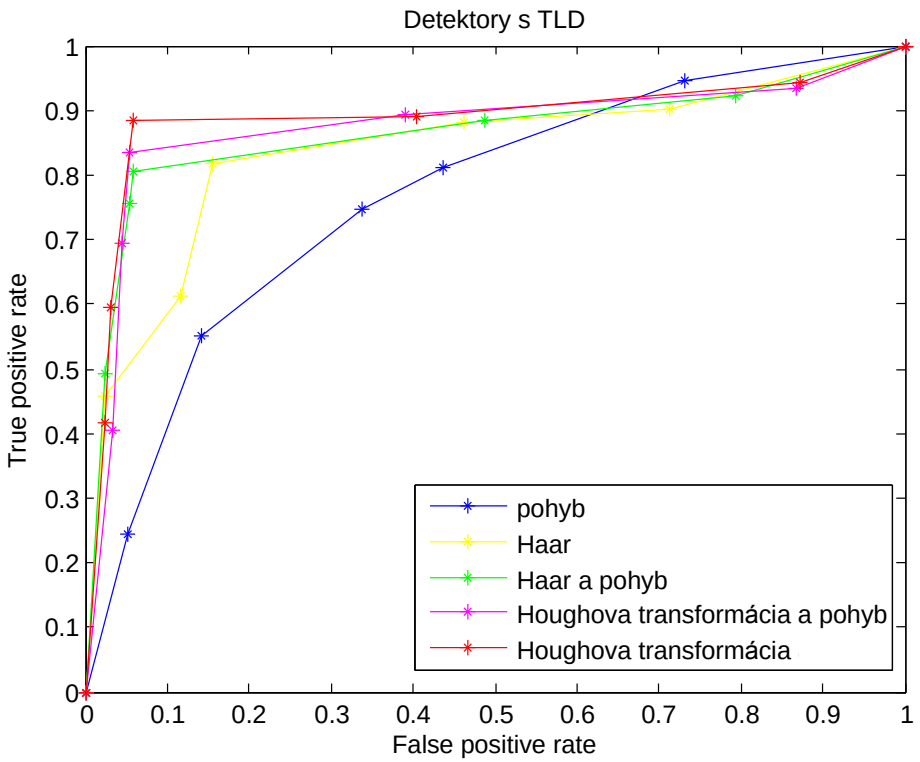
Z video datasetov bola vytvorená sekvencia snímok a anotácia prebehla na základe straty osoby na obraze. Stratou osoby z obrazu sa myslí, ak sa stratí z obrazu hlava a trup postavy.

Spôsob vyhodnotenia je vo forme ROC kriviek [19], ktoré tvoria grafy vyhodnotení (Obrázok 6-5 a Obrázok 6-6). Možno z nich vyčítať výkonnosť jednotlivých metód. Pri tvorbe kriviek určených 5 prahov. Prahy sa dosiahli zmenou hodnôt, ktoré boli opísané v predchádzajúcich podkapitolách a boli kľúčové pri prahovaní.

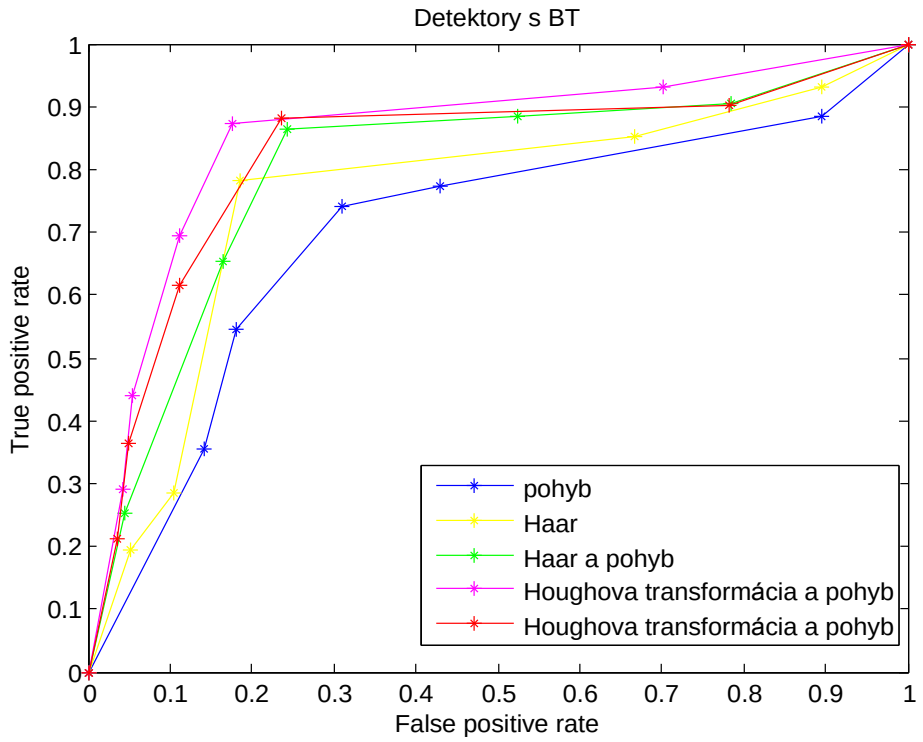
Testovanie prebiehalo na PC (Intel Pentium i3 2.30 GHz so 4 GB RAM) a konkrétne hodnoty a časy trvania jednotlivých metód sa nachádzajú v prílohe B. Tabuľka 6-1 zhodnocuje rýchlosť spracovania snímok za sekundu (angl. *frames per second-FPS*). Taktiež tabuľka vyhodnocuje výkon metód, ktorý je vyjadrený plochou pod ROC krivkou. Najlepšie výsledky sú v tabuľke zvýraznené.

detekcia	sledovanie	FPS	plocha pod ROC krivkou
na základe pohybu	TLD	1,4	0,7665
Haar kaskády	TLD	1,6	0,8396
Haar kaskády + pohyb	TLD	1,3	0,8647
s Houghovou transformáciou	TLD	0,4	0,8747
s Houghovou transformáciou+pohyb	TLD	0,4	0,8890
na základe pohybu pohyb	BT	2,2	0,7013
Haar kaskády	BT	2,1	0,7603
Haar kaskády + pohyb	BT	1,9	0,8033
s Houghovou transformáciou	BT	0,6	0,8556
s Houghovou transformáciou+pohyb	BT	0,4	0,8253

Tabuľka 6-1 Tabuľka vyhodnotenia rýchlostí metód



Obrázok 6-5 Graf vyhodnotenia detektorov s metódou sledovania TLD



Obrázok 6-6 Graf vyhodnotenia detektorov s metódou sledovania BT

6.4. Záver vyhodnotenia

Podľa nadobudnutých vyhodnotení je možné vyvodiť záver, že všetky metódy detekcie sú si podobné. Avšak jediná metóda sa vymyká od ostatných, a tou je metóda detekcie na základe pohybu, ktorá nadobudla najhoršie výsledky zo všetkých. Najrýchlejšie detekcie prebehli s detektorom z knižnice OpenCV, s metódou detekcie na základe pohybu a s ich kombináciou. Ich doba detekcie sa pohybovala okolo 80 sekúnd. V porovnaní s detektorom s upravenou Houghovou transformáciou detekcia trvala až vyše hodiny. Oba použité sledovače sledovali svoje objekty podobnú dobu, avšak metóda on-line boosting bola o niečo rýchlejšia od sledovača z knižnice OpenTLD. Na základe vyhodnotenia možno usúdiť, že najvhodnejšou kombináciou je metóda detekcie z knižnice OpenCV a metóda on-line boosting sledovania, ktoré dosiahli najlepšie výsledky v krátkom čase.

7. Záver

Cieľom práce bolo sa zoznámiť s rôznymi metódami počítačového videnia a vytvoriť systém pre počítanie osôb.

Pomocou získaných znalostí o metódach počítačového videnia som navrhol a implementoval systém, ktorý počíta osoby prechádzajúce dverami. Systém dokáže vrátane počítania osôb určiť aj smer ich pohybu. Poradí si s osobami, ktoré sa pozastavia na obraze, poprípade na malú dobu sa prekryjú s inou osobu. Problém robí osoba, ktorá kráča priamo popri inej osobe, čím je krytý pohľad na ňu a nemožno ju spracovať. Keďže systém najskôr vykoná detekciu osôb a potom vykoná ich sledovanie, nie je možné systém použiť v reálnom čase.

Testovanie a vyhodnocovanie prebehlo na datasetoch, na ktorých sú osoby zachytené stacionárnou kamerou. Datasety sú z rôznych prostredí. Podľa vyhodnotenia je možné vyvodiť záver, že najlepšou metódou detekcie osôb je detektor z knižnice OpenCV a najlepšou metódou sledovania je on-line boosting metóda.

Pre prezentáciu o výsledkoch mojej práce bolo vytvorené video a plagát.

Kto vie, možno raz dnešné činnosti vykonávané človekom budú v budúcnosti úplne nahradené prácou počítača, tak ako sa to udialo s počítaním ľudí.

Literatúra

1. TOPKAYA, I. S. H. ERDOGAN a F. PORIKLI. *Counting people by clustering person detector outputs*. 2014.
2. ANTONIAK, C. E. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. In: ANTONIAK, C. E. *The Annals of Statistics*. 2. 1974, s. 1152–74.
3. GIBSON, J. J. *The Perception of the Visual World*. Boston: MA: Houghton Mifflin, 1950.
4. STAUFFER, C. a E. GRIMSON. Adaptive background mixture models for real-time tracking. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*. 1999, s. 135-44.
5. CHAUDHRY, R. et al. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the …. In: *IEEE Conference on Computer Vision and Pattern*. 2009, s. 1932–39.
6. MADDALENA, L. A. PETROSINO a F. RUSSO. People counting by learning their appearance in a multi-view camera environment. 36. s. 125-34.
7. MADDALENA, L. a A. PETROSINO. Stopped Object Detection by Learning Foreground Model in Videos. In: *IEEE Transactions on Neural Networks and Learning Systems*. s. 723-35.
8. KHAN, S. M. a M. SHAH. A multiview approach to tracking people in crowded. In: *ECCV (4)*. 2006, s. 133–46.
9. DUDA, R. O. P. E. HART a D. G.. STORK. *Pattern Classification*. 2. Wiley, 2000.
10. MADDALENA, L. A. PETROSINO a A. FERONE. Object motion detection and tracking. In: *Int. J. Pattern Recognit. Artif. Intell*. 22. s. 915-28.
11. KALMAN, E. R. A new approach to linear filtering and prediction problems. In: *Trans. ASME–J. Basic Eng.* 82. 1960, s. 43-59.
12. HOUGH, P. Machine analysis of bubble chamber pictures. In: *Int. Conf. High Energy Accelerators and Instrumentation*. 1959.
13. BARINOVA, O. V. LEMPITSKY a P. KHOLI. On Detection of Multiple Object Instances Using Hough Transforms. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 34. 2012, s. 1773-84.
14. KALAL, Z. K. MIKOLAJCZYK a J. MATAS. Tracking-Learning-Detection. In: *Pattern Analysis and Machine Intelligence*.. 2012, s. 1409-22.
15. GRABNER, H. a B. HORST. *On-line Boosting and Vision*. Graz.
16. FREUND, Y. a R. SCHAPIRE. A decision-theoretic generalization of on-line learning and an application to boosting. In: *Journal of Computer and System Sciences*. 1997, s. 119-39.

17. ZIVKOVIC, Z. Improved Adaptive Gaussian Mixture Model for Background Subtraction. In: *Proceedings of the 17th International Conference on Pattern Recognition*. 2004.
18. 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*. 2001.
19. FAWCETT, T. An introduction to ROC analysis. In: *Pattern Recognition Letters*. 2006.

Zoznam príloh

Príloha A DVD
Príloha B Tabuľky vyhodnotenia

Príloha A

Obsah DVD

\src\ - zdrojové súbory, preložené spustiteľné súbory, datasety

\poster\ - prezentačný plagát

\video\ - prezentačné video

\text\ - dokumenty písomnej správy

\README - návod na spustenie

Príloha B

čas detekcie [s]	čas sledovania [s]	TPR	FPR
86	1845	0,2453	0,0514
76	2437	0,5512	0,1408
78	2405	0,7457	0,3372
83	2687	0,8102	0,4374
75	2537	0,9453	0,7312

Tabuľka 7-1 Vyhodnotenie metódy detekcie na základe pohybu a metódy sledovania TLD

čas detekcie [s]	čas sledovania [s]	TPR	FPR
72	2445	0,4581	0,0251
78	3424	0,6134	0,1156
71	2024	0,8166	0,1156
88	3573	0,8815	0,461
83	2477	0,9015	0,7123

Tabuľka 7-2 Vyhodnotenie metódy detekcie s Haar kaskádami a metódy sledovania TLD

čas detekcie [s]	čas sledovania [s]	TPR	FPR
76	2143	0,4924	0,0247
82	2788	0,7544	0,0543
71	2645	0,8045	0,0578
72	3475	0,8843	0,4870
86	3240	0,9214	0,7945

Tabuľka 7-3 Vyhodnotenie metódy detekcie s Haar kaskádami a pohybu a metódy sledovania TLD

čas detekcie [s]	čas sledovania [s]	TPR	FPR
5215	3021	0,4042	0,0345
4587	2917	0,6934	0,0453
4819	2818	0,8354	0,0549
5564	2705	0,8940	0,3907
5402	3102	0,9345	0,8674

Tabuľka 7-4 Vyhodnotenie metódy detekcie s Houghovou transformáciou a pohybu a metódy sledovania TLD

čas detekcie [s]	čas sledovania [s]	TPR	FPR
5452	2457	0,4154	0,0235
5857	2573	0,5944	0,0304
6354	2783	0,8854	0,0590
6750	2796	0,8905	0,4050
6753	3270	0,9433	0,8724

Tabuľka 7-5 Vyhodnotenie metódy detekcie s Houghovou transformáciou a metódy sledovania TLD

čas detekcie [s]	čas sledovania [s]	TPR	FPR
86	1224	0,3541	0,0142
76	1452	0,5460	0,1802
78	1487	0,7424	0,3101
83	1578	0,7742	0,4290
75	1971	0,8840	0,8963

Tabuľka 7-6 Vyhodnotenie metódy detekcie na základe pohybu a metódy sledovania BT

čas detekcie [s]	čas sledovania [s]	TPR	FPR
72	1245	0,1945	0,0512
78	1375	0,2861	0,1044
71	1495	0,7833	0,1845
88	1473	0,8535	0,6667
83	1893	0,9310	0,8946

Tabuľka 7-7 Vyhodnotenie metódy detekcie s Haar kaskádami a metódy sledovania BT

čas detekcie [s]	čas sledovania [s]	TPR	FPR
76	1478	0,2542	0,045
82	1733	0,6521	0,1644
71	1781	0,8645	0,2433
72	1677	0,8847	0,5239
86	1837	0,904	0,7836

Tabuľka 7-8 Vyhodnotenie metódy detekcie s Haar kaskádami a pohybu a metódy sledovania BT

čas detekcie [s]	čas sledovania [s]	TPR	FPR
5215	1576	0,2898	0,0425
4587	1473	0,4414	0,0536
4819	1243	0,6945	0,1124
5564	1837	0,8721	0,1769
5402	2437	0,9317	0,7013

Tabuľka 7-9 Vyhodnotenie metódy detekcie s Houghovou transformáciou a pohybu a metódy sledovania BT

čas detekcie [s]	čas sledovania [s]	TPR	FPR
5452	1437	0,2121	0,0350
5857	1373	0,3634	0,0492
6354	1937	0,6142	0,1125
6750	2373	0,8821	0,2354
6753	2040	0,9013	0,7831

Tabuľka 7-10 Vyhodnotenie metódy detekcie s Houghovou transformáciou a metódy sledovania BT