

SEGMENTACE OBRAZU PRO IDENTIFIKACI OSOB POMOCÍ BIPEDÁLNÍ LOKOMOCE

K. Horák, M. Richter

Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně

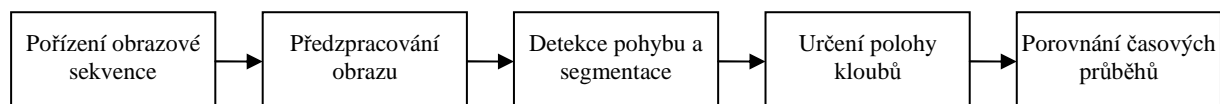
Abstrakt

Článek se zabývá segmentační metodou pro zpracování obrazů reálného videa potřebnou pro identifikaci osob pomocí bipedálních charakteristik. Předzpracování snímků je založeno na základních morfologických operacích a operacích pro filtraci šumu v obrazu. Samotná segmentace osob je realizována pomocí detekce pohybu a metody středního posunu pro shlukování bodů.

1 Úvod

Identifikace osob je myšlenka stará jako lidstvo samo. Intuitivní rozpoznání známé osoby na vzdálenost několika desítek, někdy až stovek metrů nečinilo velké potíže ani lidem, jejichž nejmodernější technologií byl dobře opracovaný kyj. To je důvod, proč většina dnešních identifikačních systémů kopíruje původní lidské intuitivní rozpoznávání počínaje rozpoznáním tváře, očí, hlasu, charakteristického pohybu chůze a otiskem prstu konče. Právě unikátnost lidské chůze popř. běhu lze využít pro identifikaci osob vedle již zmíněných metod otisku prstu, obrazu oční duhovky nebo parametrů hlasového ústrojí. Tato oblast zahrnující studium lidské chůze se odborně nazývá bipedální lokomoce, tedy obecný pohyb dolních končetin.

Kromě znalostí anatomického ústrojí člověka však automatická identifikace osob z video sekvence vyžaduje také technické znalosti z oblasti počítačového vidění pro spolehlivou analýzu a segmentaci pohybujících se objektů v obrazu. K tomuto účelu bylo z hlediska základního výzkumu použito výpočetní prostředí Matlab s knihovnami Image Acquisition Toolbox (dále jen IAT) pro pořízení obrazu a Image Processing Toolbox (dále jen IPT) pro jeho zpracování. Schematický řetězec činností vedoucích k bipedální identifikaci lze znázornit blokovým schématem na obrázku 1.

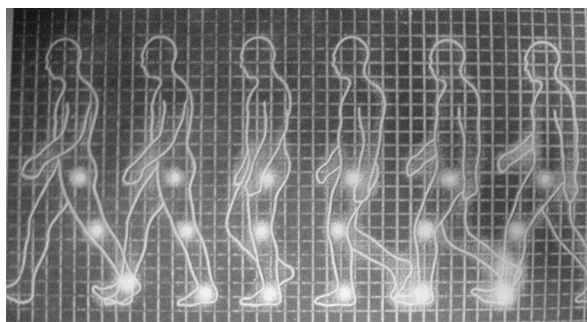


Obrázek 1: Postup zpracování obrazu pro identifikaci osob pomocí bipedální lokomoce

Tento článek, jak je uvedeno i v jeho závěru se kromě obecného vymezení oblasti bipedální lokomoce zabývá zejména prvními třemi bloky výše uvedeného schématu. Jen částečně je v šesté kapitole nastíněn také čtvrtý blok schématu, tedy určení časových průběhů poloh kloubů, který závisí na existenci algoritmu pro extrémně přesnou detekci polohy každého kloubu bez ohledu na vnější okolnosti např. oblečení osoby nebo stíny. Tento algoritmus je postupně zpřesňován, čímž se zvyšuje počet identifikovatelných osob.

2 Bipedální lokomoce

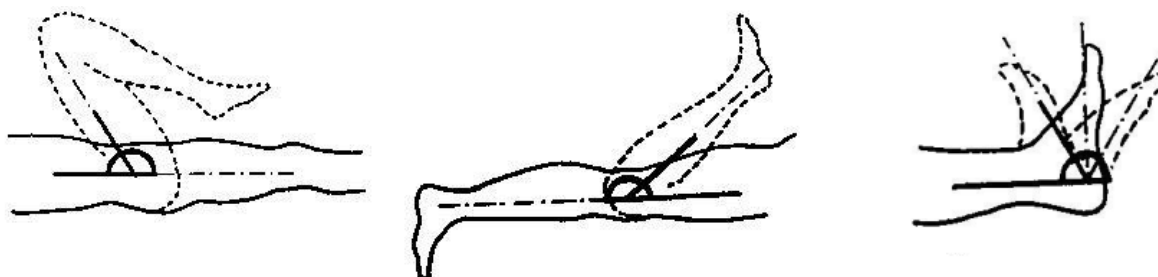
Chůze je nejčastěji definována jako rytmický pohyb dolních končetin doprovázený souhybem ostatních částí těla. Pro potřeby bipedální identifikace osob je důležitá fáze kroku, kdy obě nohy spočívají na zemi (dvojí opora) a fáze, kdy tělo spočívá na jedné končetině a druhá osciluje (jednostranná opora). Přejechod mezi těmito fázemi je znázorněn na obrázku 2.



Obrázek 2: Bipedální přechod mezi fázemi dvojí opory

Z anatomických výzkumů bylo zjištěno, že dynamický stereotyp chůze se stabilizuje až kolem sedmého roku života jedince, od kterého jsou v pohybu upevněny individuální bipedální charakteristiky. Ačkoliv je chůze obecně vrozenou vlastností každého jedince druhu homo sapiens sapiens, učí se individuum chodit samo a tak si vytváří individuální pohybový projev a držení těla. Z uvedeného vyplývá, že spolehlivou bipedální identifikaci lze použít zejména u dospělých jedinců a zejména při chůzi. Problematická by byla identifikace v případě běhu, kdy se vůbec nevyskytují okamžiky dvojí opory a kdy se navíc od určitého věku může po tréninku objevit i tzv. letová fáze, v níž až 85 % celého krokového cyklu není žádná končetina v kontaktu se zemí.

Pro základní identifikaci pomocí bipedální lokomoce je nejdůležitější detekce poloh kloubu kyčelního, kloubu kolenního a kloubu hlezenního, jejichž rozsahy pohybů mezi flexí a extenzí jsou spolu částečně svázány (obrázek 3).



Obrázek 3: Zdroje pohybových informací pro bipedální identifikaci

Časové průběhy horizontální polohy každého ze tří uvedených kloubů jsou unikátní informací o jedinci, které ale nemůže (na rozdíl od sebe destrukce otisku prstu nebo nasazení zkreslující oční čočky) svým vědomým chováním nijak ovlivnit, nepočítáme-li nestandardní situace jako běh při útěku z vyloupené banky apod. Následující kapitoly se již věnují konkrétnímu technickému řešení pro získání uvedených identifikačních dat.

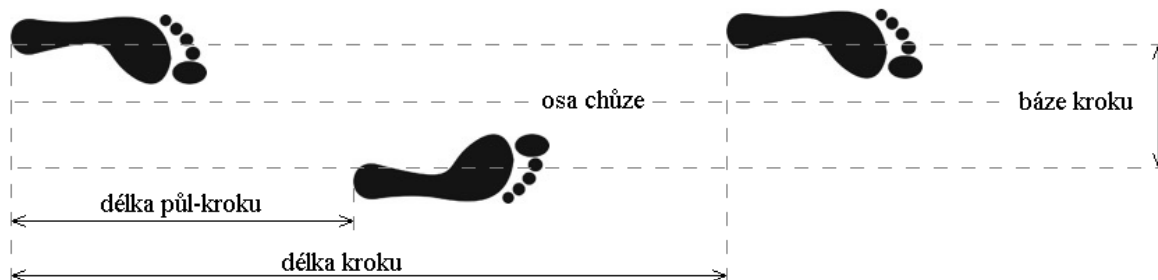
3 Pořízení obrazu

Fáze pořízení obrazu dostatečné informační hodnoty sestává ze tří částí:

- výběr vhodného snímacího zařízení
- určení geometrických parametrů snímání
- parametry expozice kamery a trvání snímání

Pro snímání chůze člověka ve standardních definovaných podmínkách (např. letištní hala) jsou důležité dva základní parametry – snímkovací frekvence a rozlišení čipu. Ostatní parametry jako citlivost čipu, typ čipu CCD/CMOS, rozhraní kamery apod. nejsou ve vývojové fázi podstatné. Požadovaná snímkovací frekvence kamery je dána rychlostí snímaného pohybu. Průměrná rychlost

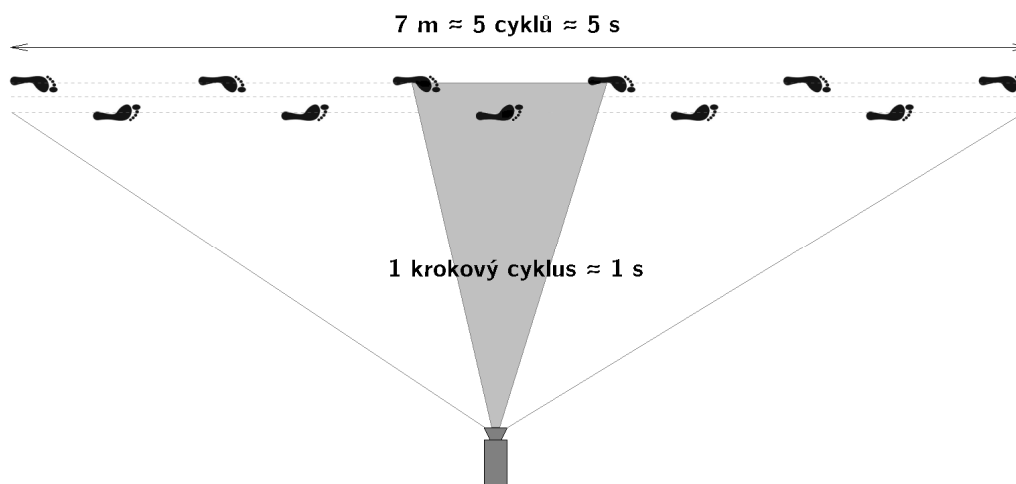
běžné chůze je kolem 5 km/h, průměrná délka půl-kroku (výměna stejné nohy např. levé za pravou) přesahuje vzdálenost 70 cm. Jednoduchým výpočtem lze odvodit, že jeden půl-krok průměrně rychlé chůze trvá přibližně 0,5 s, celý krokový cyklus (viz obrázek 4) tedy jednu sekundu.



Obrázek 4: Krokový cyklus

Pro získání porovnatelných časových průběhů poloh kloubů je zapotřebí pořídít řádově několik desítek vzorků (snímků) jednoho krokového cyklu. Experimentálně zjištěná snímkovací frekvence kamery musí být alespoň 20 snímků/s.

Rozlišení čipu souvisí s geometrickým nastavením snímané scény, které je znázorněno na následujícím obrázku 5. Pro statisticky spolehlivé porovnání průběhů poloh kloubů je potřeba zaznamenat alespoň pět krokových cyklů, což při průměrné délce jednoho půl-kroku 70 cm odpovídá sedmi metrům zaznamenané chůze.



Obrázek 5: Půdorys geometrického rozmístění scény

Horizontální rozlišení pořízeného obrazu musí být v jednom krokovém cyklu alespoň 150 pxl, což odpovídá celkovému horizontálnímu rozlišení čipu 750 pxl pro záběr pěti krokových cyklů¹. Pokud tato podmínka nebude splněna, nemohou být polohy kloubů nalezeny dostatečně přesně, časové průběhy nebudou odpovídat skutečnosti a jejich následné porovnání s průběhy v databázi bude vykazovat vysokou chybu.

Z hlediska dalšího zpracování obrazu je podstatné také nastavit pevně parametry snímání, zejména expoziční dobu, automatické vyvážení bílé a jiné automaticky se měnící parametry, jejichž změny během pořizování obrazové sekvence by musely být později algoritmicky kompenzovány už před detekcí pohybu pomocí rozdílové metody (viz kapitola 5.1). Obdobně musí být zajištěno také

¹ Pro jednoduchost je uvažován stejný počet pixelů pro všechny krokové cykly bez ohledu na pozici cyklu na obrazovém snímači (změna úhlu snímání)

fixní zaostření na snímací vzdálenost, proto nejsou jako snímací zařízení vhodné běžné fotoaparáty nebo kamery s automatickým přeostréváním. Na obrázku 6 je fotografie navržené mobilní sestavy pro experimentální pořízení obrazové sekvence chůze.



Obrázek 6: Mobilní sestava pro pořízení obrazu chůze osob

V sestavě byla použita desková kamera Micron Imaging MT9P031 s fyzickým rozlišením čipu 5 Mpxl, pro snímání nastaveno rozlišení 800x600 pxl při snímkové frekvenci 22 fps. Pořízená data byla ukládána bez komprese jako bitmapové obrázky přímo na disk připojeného notebooku.

4 Předzpracování obrazu

Předzpracováním obrazu se obvykle rozumí výřez relevantní části obrazu, filtrace šumu, transformace pro kompenzaci různých geometrických zkreslení nebo bodové a lokální jasové transformace. V případě obrazového záznamu chůze osob za definovaných a do jisté míry i zvolených podmínek (osvětlení, přizpůsobení pozadí scény apod.) je potřeba odstranit pouze případný šum (viz obrázek 7).



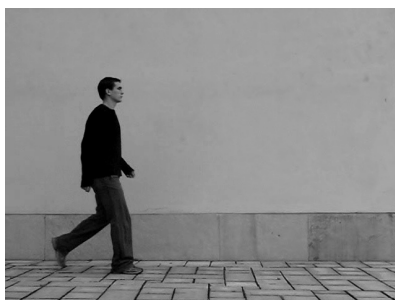
Obrázek 7: Snímek sekvence chůze filtrovaný Gaussovským jádrem 3x3

Pro filtraci bylo použito jádro normálního Gaussova rozdělení o velikosti matice 3x3 bodů se směrodatnou odchylkou rovnou jedné (IPT funkce `imfilter` a `fspecial`).

Pro potřeby detekce pohybu v sekvenci není v tomto případě důležitá informace o barevných složkách, proto jsou ještě v rámci předzpracování obrazu převedena barevná obrazová data do odstínů šedi pomocí funkce `rgb2gray`, která využívá standardního transformačního vztahu (1).

$$Y(i, j) = 0.299 \cdot R(i, j) + 0.587 \cdot G(i, j) + 0.114 \cdot B(i, j) \quad (1)$$

Výsledný šedo-tónový obraz Y s informací jen o jasové složce (obrázek 8) je dále využit pouze pro detekci pohybu a samotná segmentace objektů (kapitola 5.2) se kvůli zvýšení rychlosti algoritmu provádí jen na příslušné části původního barevného obrazu.



Obrázek 8: Šedo-tónový obraz tvoří vstupní data detekce pohybu

5 Rozpoznání osob

Rozpoznání objektů na snímcích, které odpovídají jdoucím osobám je rozděleno na dvě fáze. V první fázi (kapitola 5.1) jsou pouze detekována místa, kde došlo ke změně obrazové funkce, tedy viditelnému pohybu². Druhá fáze (kapitola 5.2) se zabývá segmentací osob (nalezením kontur) a to už jen v místech obrazu, kde došlo k pohybu.

5.1 Detekce pohybu

Základní metodou detekce pohybu je diferenční snímek. Diferenční, někdy také rozdílový snímek vznikne odečtením dvou ne nezbytně po sobě jdoucích obrazů³ (IPT funkce `imsubtract`). Výsledná mapa jednoduše řečeno udává, co je na aktuálním snímku nového oproti snímku předcházejícímu. Avšak podstatnou nevýhodou výsledku této jednoduché metody je to, že neurčuje celý pohybující se objekt, ale jen jeho příslušnou část určenou rychlostí pohybu a velikostí objektu. Názorně je tento jev ilustrován na následujícím obrázku 9.



Obrázek 9: Diferenční snímek (vpravo) vstupních obrazů (vlevo a uprostřed) s časovým rozestupem 70 ms

Na výsledném snímku je jednoznačně patrná oblast, v níž došlo k pohybu, nelze ji však ani po úpravách použít jako binární masku pro přesnou segmentaci kontury osoby. V našem případě, kdy je kamera nehybná a kdy lze alespoň přibližně zaručit konstantní světelné podmínky, je výhodnější použít rozdílovou metodu na jeden snímek sekvence a snímek statického modelu prostředí (estimace modelu pozadí). V okamžicích nulového pohybu ve scéně, které lze snadno detekovat výše uvedeným

² Za určitých podmínek může dojít ve scéně k reálnému pohybu, aniž by se jasová funkce obrazu změnila.

³ I když jen v takovém případě má rozdílový snímek smysl pro detekci pohybu.

rozdílovým snímkem, je aktualizován tzv. model prostředí. Je to v podstatě průměrný snímek čistého pozadí bez pohybujících se objektů. Na obrázku 10 je ilustrován výsledek této metody se stejnými vstupními daty, jako v předešlém případě.



Obrázek 10: Diferenční snímek (vpravo) vstupního obrazu (vlevo) a modelu prostředí (uprostřed)

Výsledný diferenční snímek nese dostačující informaci o poloze a rozměrech celého pohybujícího se objektu. Z uvedeného výsledku je potřeba získat už jen binární mapu objektu. Toto lze provést pomocí segmentační metody prostého prahování (IPT funkce `im2bw`) a morfologické operace uzavření (IPT funkce `imclose`) tak, jak je uvedeno dále.



Obrázek 11: Diferenční snímek (vlevo), prahovaný (uprostřed) a upravený morfologickým uzavřením (vpravo)

Pro operaci morfologického uzavření je použit strukturální element (IPT funkce `strel`) diskového tvaru s poloměrem pět bodů, jehož tvar je dán maticí SE_{close} (2).

$$SE_{close} = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \quad (2)$$

Při pohledu na snímek zcela vpravo na obrázku 11 vyvstává otázka, proč konturu takto detekovaného objektu nepoužít jako skutečnou konturu osoby pro nalezení kloubů? Důvodem je přesnost a hlavně robustnost předchozího zpracování obrazu. Použitá sekvence i metody samotné jsou tak jednoduché, že mnohdy dochází zejména vlivem prahování ke ztrátě části obrysu osoby nebo naopak připojení části pozadí scény, zejména v oblastech došlapu chodila, kde vznikají velmi tmavé stíny (tento jev lze částečně pozorovat na prostředním snímku obrázku 11 pod chodidlem zadní nohy).

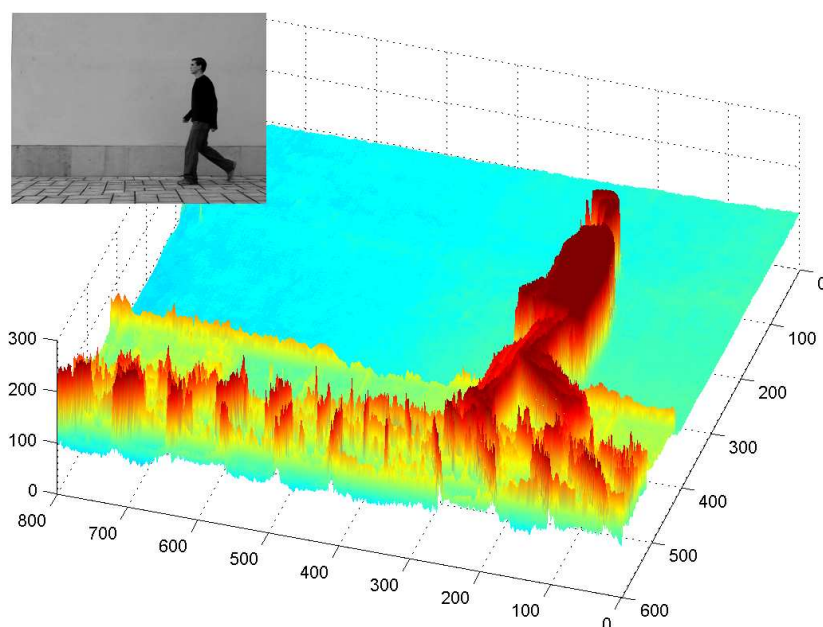
Vymezení oblasti pro přesnou segmentaci kontury osoby lze provést pomocí operace morfologické dilatace, při níž jsou k obrysu objektu zjednodušeně řečeno přidávány okrajové vnější slupky. Forma a rozsah dilatace jsou dány strukturálním elementem SE_{dilate} , který je volen opět ve tvaru disku tentokrát však s větším poloměrem 10 bodů, aby výsledná binární mapa bezpečně obsahovala všechny obrazové body skutečného objektu. Vstupní obraz, výsledná binární mapa objektu i příslušná původní obrazová data jsou uvedena na obrázku 12.



Obrázek 12: Přibližná kontura (vlevo), binární mapa oblasti osoby (uprostřed) a původní obrazová data (vpravo)

5.2 Segmentace objektů

Vstupním obrazem pro přesnou segmentaci objektu je snímek zcela vpravo na obrázku 12. Pro segmentaci je použita metoda středního posuvu pro shlukování bodů, častěji označovaná originálním názvem mean-shift. Tato metoda převádí úlohu segmentace na úlohu hledání shluků podobných bodů na základě jejich vzdálenosti v obecně n -rozměrném prostoru. V případě šedo-tónového snímku jej lze reprezentovat trojrozměrnou maticí, jak je uvedeno na obrázku 13, jejíž první dvě dimenze tvoří prostorové souřadnice x a y a třetí jasová úroveň příslušného obrazového bodu. Objekt lze pak segmentovat na základě vzdálenosti⁴ vyšetřovaných bodů od zvoleného středu shluku.

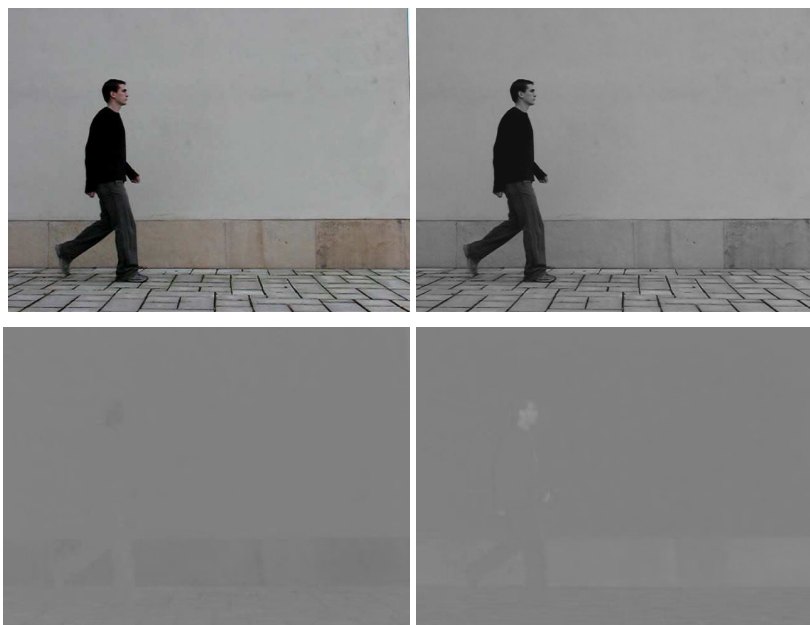


Obrázek 13: Trojrozměrná matice reprezentuje šedo-tónový diskretní obraz

U šedo-tónového obrazu se jedná o tří-rozměrnou matici (jak ukazuje obrázek), u barevného obrazu pak ještě o dvě dimenze více. Algoritmus tedy hledá shluky bodů v pěti-rozměrném prostoru (dvě souřadnice+tři barvy), proto bylo nutné předchozí metodou detekce pohybu eliminovat hluchá místa obrazu, aby byl celý algoritmus použitelný z hlediska výpočetní náročnosti.

⁴ Vzdálenost podle zvolené metriky, v případě obrazů nejčastěji vážená norma L_2

Základní podmínkou pro správný výsledek algoritmu mean-shift je, aby body náležející objektu v realitě (fyzická blízkost) byly podle zvolené normy geometricky blízké i v n-rozměrném prostoru. Tomuto předpokladu však příliš nevyhovuje barevný prostor RGB, který je ale zase téměř výhradně využívaný výrobci snímacích zařízení. Obraz je proto převeden na jinou, vhodnější reprezentaci. Teoreticky lze jako vhodné kandidáty navrhnout intuitivní barevný prostor HSV a barevný prostor doplňkových barev YC_bC_r . Experimentálně byl vybrán druhý jmenovaný, tedy prostor YC_bC_r , jehož dekompozice je znázorněna na obrázku 14.



Obrázek 14: Dekompozice barevného obrazu (vlevo nahoře) na složky Y (vpravo nahoře), C_b (vlevo dole) a C_r (vpravo dole) doplňkového prostoru YC_bC_r

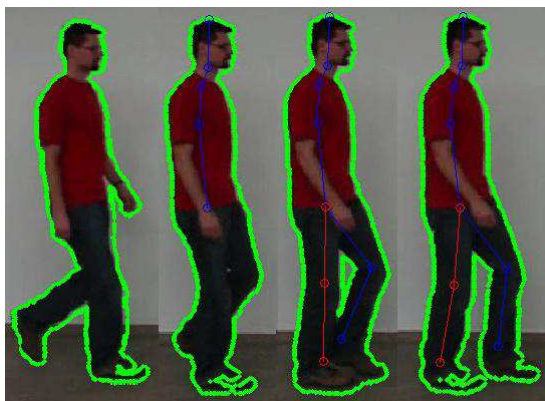
Převod barevného prostoru RGB kamery Micron Imaging (čip CMOS) na barevný prostor doplňkových barev YC_bC_r vhodný pro shlukovací algoritmus mean-shift se za předpokladu 8 bitové barevné hloubky řídí normovaným vztahem (3) a lze jej provést IPT funkcí `rgb2ycbcr`.

$$\begin{bmatrix} Y(i, j) \\ C_b(i, j) \\ C_r(i, j) \end{bmatrix} = \frac{1}{2^8} \cdot \begin{bmatrix} 65.7 & 129.1 & 25.1 \\ -37.9 & -74.5 & 112.4 \\ 112.4 & -94.2 & -18.3 \end{bmatrix} \cdot \begin{bmatrix} R(i, j) \\ G(i, j) \\ B(i, j) \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (3)$$

Na takto transformovaný prostor je aplikován iterativní mean-shift algoritmus pro nalezení lokálního maxima vzorků, jehož princip lze shrnout do následujících sekvence:

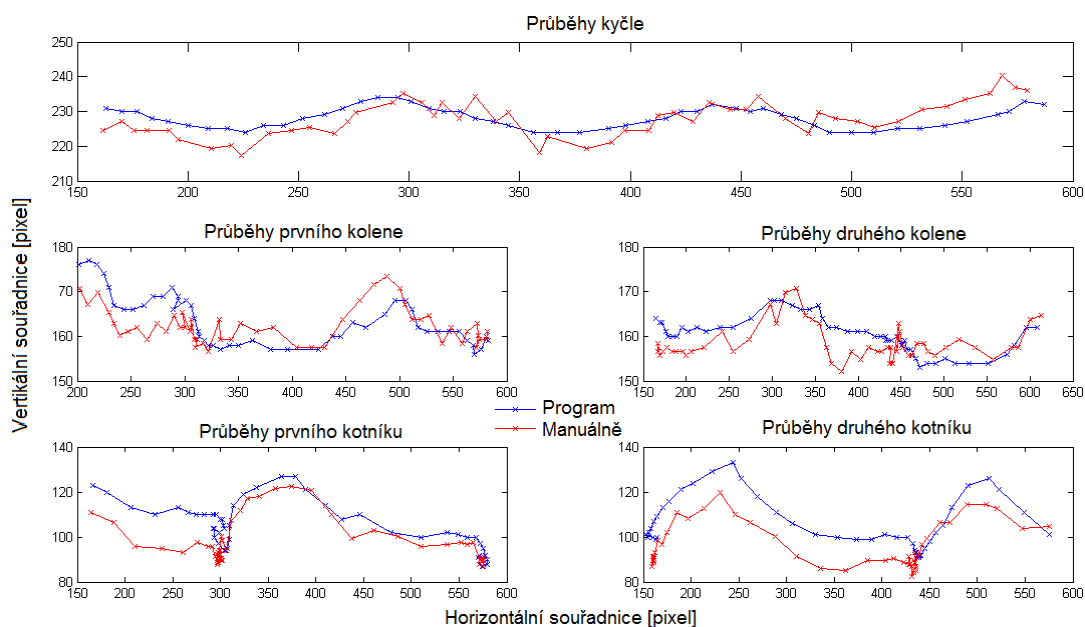
- určení N počátečních středů shluků
- inicializace = přiřazení každého obrazového bodu k jednomu nejbližšímu z N středů podle zvolené metriky
- určení nového středu z průměru všech členů shluku
- přiřazení nových členů shluku podle aktuálních středů
- iterativní návrat k bodu c), dokud nejsou gradienty dvou po sobě jdoucích středů všech shluků rovny nule
- spojování lokálních shluků blízkých hodnot do jednoho objektu

Výsledkem segmentační metody jsou jednoznačně indexované oblasti, jejichž spojením je určena obrazová plocha celého objektu a ze které je pomocí IPT funkce `bwperim` získána kontura kráčející osoby, jak uvádí obrázek 15 v zelené barvě. Tato kontura popř. celá binární plocha kráčející osoby tvoří vstupní data pro detekci polohy kyčelního, kolenního a hlezenního kloubu, jejichž časové průběhy jsou základním kamenem identifikace.



Obrázek 15: Kontura kráčející postavy nalezená iterativním algoritmem `mean-shift`

Přesné kontury slouží jako vstupní data pro výpočet poloh kloubů, jejichž časové průběhy jsou použity pro bipedální identifikaci osob. Na obrázku 15 jsou mimo samotnou konturu také vyznačeny základní kosterní linie, jejichž průběhy mají individuální charakter a slouží pro statistické porovnání s jinými průběhy. Ukázka časových průběhů z obrazové sekvence chůze ukazující vertikální změny v průbězích kyčle, obou kolen a obou kotníků je uvedena na následujícím obrázku 16.



Obrázek 16: Časové průběhy poloh kloubů pro identifikaci pomocí bipedální lokomoce

Získávání přesných poloh kloubů, sestavení časových průběhů a statistické srovnání s databázovými údaji již nespadá do tématu tohoto článku, kterým byla přesná segmentace kráčejících osob ve zvoleném prostředí.

6 Závěr

Navržený segmentační algoritmus se vyznačuje svou jednoduchostí a tím i spolehlivostí a lze jej relativně snadno přizpůsobit různým variacím složitosti snímané scény – letiště, školy, kancelářské budovy aj. Pokračování práce identifikace pomocí bipedální lokomoce spočívá ve vývoji algoritmu pro přesné nalezení kyčelního, kolenního a hlezenního kloubu z obrazu kontury osoby, kterou tento algoritmus získává. Ze znalosti časových průběhů těchto kloubů a ze znalosti časových průběhů kloubů hříšníků dříve uložených v centrální databázi lze finálně provést jejich porovnání a určení identity v reálném čase, dříve než podezřelá osoba opustí kontrolovaný prostor.

Poděkování

Článek vznikl za podpory Výzkumného centra aplikované kybernetiky (projekt MŠMT 1M0567) a interního grantu GRAFO Vysokého učení technického v Brně.

Reference

- [1] HALEDOVÁ, E., NECHVÁTALOVÁ, L. *Vyšetřovací metody hybného systému*. Institut pro další vzdělání pracovníků ve zdravotnictví, Brno 1997, 137 s. ISBN 80-7013-237-X.
- [2] STRAUS, J. - JONÁK, J. *Kriminalistická a technická analýza bipedální lokomoce*. Vydalo PA ČR v Praze pro Policejní akademii České republiky, 2007, ISBN – 978-80-7251-268-3
- [3] SONKA, M., HLAVAC, V., BOYLE, R.. *Image Processing, Analysis and Machine Vision*. 3rd edition. Toronto : Thomson, 2008. 829 s. ISBN 978-0-495-08252-1.
- [4] VERNON, David. *Machine Vision : Automated Visual Inspection and Robot Vision*. Hemel Hempstead : Prentice Hall International (UK) Ltd., 1991. 260 s. ISBN 0-13-543398-3.
- [5] TEKALP, M. *Digital Video Processing*. Prentice Hall International Ltd., 1995. 526 pages. ISBN 01-13-190075-7.
- [6] RUSS, C. *The Image Processing Handbook*. CRC Press, 1994. 674 pages. ISBN 0-8493-2516-1.
- [7] BARTSCH, H.J. *Matematické vzorce*. SNTL Praha. 1971. ISBN 04-001-71.
- [8] CROWLEY, J.L., CHRISTENSEN, H.I. *Vision as Process*. Springer-Verlag. 1995. ISBN 3-540-58143.
- [9] DAVIES, E.R. *Machine Vision : Theory, Algorithms, Practicalities*. Morgan Kaufmann. 2005. ISBN 0-12-206093-8.
- [10] JÄHNE, B., HAUBECKER, H. *Computer Vision and Applications, A Guide for Students and Practitioners*. Academic Press. 2000. ISBN 0-13-085198-1.
- [11] KOTEK, Z. *Metody rozpoznávání a jejich aplikace*. Academia Praha, 1993.

Ing. Karel Horák, Ph.D.

Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně
Kolejní 4, 612 00 Brno, tel. +420 541143647, e-mail: horakk@feec.vutbr.cz

Ing. Miloslav Richter, Ph.D.

Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně
Kolejní 4, 612 00 Brno, tel. +420 541141194, e-mail: richter@feec.vutbr.cz