

HEAD POSITION AND MOVEMENT DETECTION FOR MULTIPLE-USER COMPUTER INTERFACE

Filip Chmiel

Bachelor Degree Programme (3), FIT BUT

E-mail: xchmie02@stud.fit.vutbr.cz

Supervised by: Vítězslav Beran

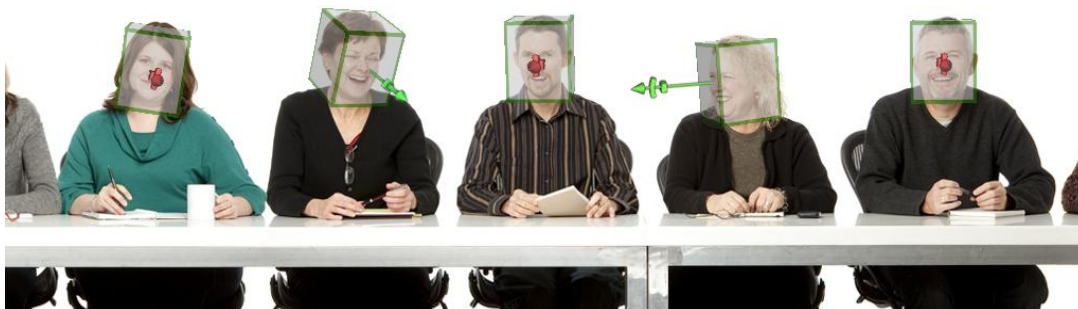
E-mail: beranv@fit.vutbr.cz

Abstract: This paper presents the method for head pose detection based on computer vision techniques. The method is based on the AdaBoost approach for face recognition and the optical flow technique for object tracking. Obtained results are interpreted after linear-differential processing. The method is designed for use in designing and building the user interface.

Keywords: interface, multi-user, computer vision, AdaBoost, optical flow

1. ÚVOD

Cílem tohoto příspěvku je přiblížit způsob návrhu a fungování metody určené pro zpracovávání podnětů v podobě pohybu hlavy získaného od více uživatelů. V posledních letech se začal prosazovat fenomén počítačového vidění, technologie pro to určené poskytují různé možnosti uplatnění v dosud ne příliš prozkoumané oblasti. Vyvinuté algoritmy umožňují snímat polohu uživatele a následovně ji v reálném čase analyzovat. Získané údaje se dají aplikovat v různých oblastech jako například ovládání počítače, interaktivní reklama či multimediálních stanic. Sledovací algoritmy jsou ovšem výpočetně náročné a aby byla dosažena přijatelná odezva je potřeba k tomuto faktu během návrhu přihlídnout.

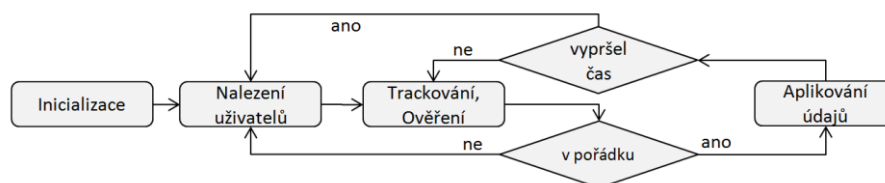


Obrázek 1: Ukázka možné detekce směru pohledu uživatelů

2. STRUKTURA NAVRŽENÉ METODY

Pro správný návrh výsledné metody je nutno brát v úvahu že uživatelé většinou nejsou technicky založení a vyžadují co nejmenší náročnost. Navíc se před kamerou různě pohybují, odcházejí a znovu se vrací. Detekce a sledování uživatelů bylo po testování navrženo tak, že pro počáteční inicializaci používá detektor objektů a pro jejich sledování v následujících snímcích vyhledává korespondující pozice vyznačených bodů v nalezeném obličejí. Výsledkem je metoda (obr. 2) poskytující aplikaci informace o pozici uživatelů nalezených ve video signále. Navíc ověřuje správnost vyznačené pozice pomocí zmíněného detektoru, s tím, že výpočetní náročnost je snížena pomocí vyznačení oblasti pro vyhledávání. Takto lze stanovit, s jakou úspěšností se podařilo nalézt kore-

spondující pozici a případně sledování znovu inicializovat s upřesněnou pozicí nebo v libovolných časových intervalech kontrolovat výskyt nových uživatelů v obraze.



Obrázek 2: Základní schéma metody

Potřebné funkce pro detekci objektů a jejich sledování jsou obsaženy ve volně dostupné knihovně OpenCV. Knihovna se nadále vyvíjí a optimalizuje s důrazem na výkon.

2.1. DETEKTOR OBJEKTŮ

V současnosti nejpopulárnější metodou pro detekci objektů v obraze jsou metody založené na principech AdaBoostu [1], která z velkého množství slabých klasifikátorů vybere ty nejvhodnější a jejich lineární kombinací (1) vytvoří výsledný silný klasifikátor.

$$H(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i h(x_i)\right) \quad (1)$$

Slabé klasifikátory $h(x_i)$ mají malou přesnost, ale vysokou rychlost. Pokud se vhodně zkombinují, s přihlédnutím k váze α_i kterou potřebujeme nastavit trénováním, získáme silný klasifikátor $H(x)$ vyznačující se nejenom rychlostí, ale také i požadovanou přesností.

2.2. OPTICKÝ TOK

Optický tok [2] je v počítačovém vidění základním pojmem pro stanovení pohybu jednotlivých objektů ve scéně. Stanovuje se obvykle porovnáním dvou nebo více snímků zobrazujících scénu v různých časových okamžicích. Klasický algoritmus optického toku zformulovaný Lucas a Kanade vychází z předpokladu, že intenzita jasu oblastí vybraných bodů se v čase příliš nemění, a že blízké body často patří povrchu stejného objektu. Koncept spočívá v tom, že se hledá vektor pohybu (V_x, V_y) znázorňující změnu pozice. Výpočet se provádí podle vztahu:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix}^{-1} \times \begin{bmatrix} -\sum I_x I_t \\ -\sum I_y I_t \end{bmatrix} \quad (2)$$

Kde I_x a I_y jsou parciální derivace intenzity obrazu I ve směrech x respektive y .

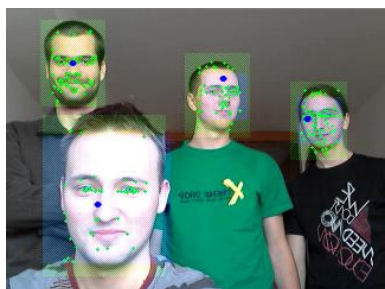
2.3. VYUŽITÍ VÝSLEDKŮ PRO GUI

Pokud tedy zkombinujeme princip detektoru objektů s optickým tokem, získáme rychlý a relativně stabilní metodu sledování. Technologie optical flow je ovšem náchylná k občasnému mylnému určení korespondující pozice bodu způsobené ať už změnami jasu, nekvalitním vstupem či příliš rychlými pohyby v obraze. Tento problém lze řešit ověřením předpokládané pozice detektorem použitým na vyznačenou oblast. Korekce pomocí detekce prováděné nad celým snímkem není nezbytně nutná. Její frekvence záleží na zaměření cílové aplikace, u interaktivní reklamy je vhodnější častější (ideálně každou sekundu) obnova informací, oproti tomu pro ovládání počítače teoreticky postačují minutové intervaly. Pokud známe pozici obličeje, můžeme získané informace doplnit o pozici obličejových rysů a získat tak doplňující popis pozice uživatele. Další filtrování údajů například za využitím Kalmanova filtru se ukázalo, i přes jeho predikční podstatu, jako ne příliš použitelné, kvůli zpoždění které do výstupu zanáší. Místo filtrování výsledků se jako vhodnější metoda stabilizace osvědčila diferenciální interpretace výsledků, která potlačuje šum ve výstupu a dává větší váhu rychlým a výrazným pohybům. Pokud tedy máme kompletní informaci o pozici uživatele, ať už filtrovanou nebo bez úprav, je třeba řešit problém s jejich využitím. Jelikož se při detekci probíhající v reálném čase pracuje se snímky v rozlišení 320 na 240 pixelů, jen občas s většími

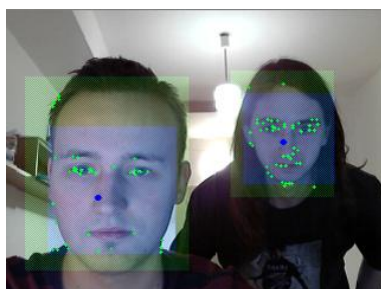
právě kvůli zvětšující se s rozlišením výpočetní náročnosti a narůstající odezvě kamery, získané údaje je nutno vhodně interpretovat pro rozlišení používané v cílových zařízeních. Pokud je použito přímého lineárního mapování (3), kde změna v pozici je interpretována přímo, může být dosažená přesnost a citlivost nedostatečná.

$$\vec{x} = \vec{x} \cdot d \quad (3)$$

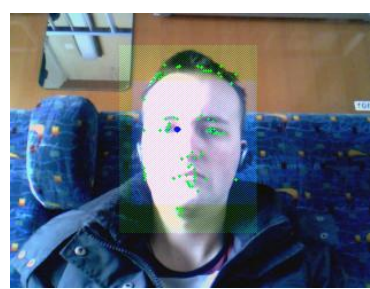
Pro zlepšení lze použít kombinaci lineární a diferenciální (4) interpretace [3]. Princip diferenciální interpretace vychází z toho, že pokud změna pozice uživatele oproti pozici v minulém snímku je malá tak se s vysokou pravděpodobností jedná o šum nebo o drobný pohyb, proto je mu kladena malá váha. Oproti tomu pokud je změna pozice za časový úsek zřetelná, dá se považovat za výrazný podnět, tedy se musí ve výsledku více projevit.



Obrázek 3: Detekce více uživatelů



Obrázek 4: Detail detekce



Obrázek 5: Funkčnost v horších podmínkách

$$\vec{x} = \frac{\Delta x \cdot |\Delta x|}{\Delta t} \cdot d \quad (4)$$

Získané údaje (obr 1. 2. 3.) lze využít v různých aplikacích, například pro inovativní způsob ovládní aplikací.

3. ZÁVĚR

Kombinace detektoru objektů, metody pro výpočet optického toku a lineárně-diferenciální interpretace ukazuje zajímavé výsledky použitelné nejen pro osobní počítače, ale například i pro multimediální studia, domácí přehrávače či interaktivní reklamy. Výsledná metoda běží v reálném čase s důrazem na co nejmenší výpočetní náročnost a je navržena obecně. Po úpravě lze tedy použít pro sledování i jiných objektů.

PODĚKOVÁNÍ

Tato práce vznikla za podpory grantu Centrum excellence IT4Innovations, CZ 1.05/1.1.00/02.0070.

REFERENCE

- [1] VIOLA, Paul ; JONES, Michael. Robust Real-time Object Detection. *IJCV* [online]. 2001, [cit. 2011-02-24]. Dostupný z WWW: <http://research.microsoft.com/~viola/Pubs/Detect/violaJones_IJCV.pdf>.
- [2] D. LUCAS, Bruce; KANADE, Takeo. An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging Understanding Workshop* [online]. 1981, [cit. 2011-02-23]. Dostupný z WWW: <<http://www-cse.ucsd.edu/classes/sp02/cse252/lucaskanade81.pdf>>.
- [3] Zich, L. *Video-based human computer interface*. Master thesis, Cybernetics at the Czech Technical University, 2009.