

# WEBCAM AS A POINTING DEVICE

**Pavel Mynář**

Master Degree Programme (3), FIT BUT

E-mail: xmynar01@stud.fit.vutbr.cz

Supervised by: Vítězslav Beran

E-mail: beranv@fit.vutbr.cz

**Abstract:** This paper presents an application of computer vision as an assistive technology intended for people unable to use standard pointing device (e.g. mouse, trackball). Instead, input is calculated in real time by tracking user's head motion. The presented solution is platform independent and requires only common webcam as an input device.

**Keywords:** assistive technology, user interface, webcam, tracking, OpenCV, pointing device

## 1. ÚVOD

V oblasti asistenčních technologií se stále používá nákladný a specializovaný hardware a rychlost vývoje nových zařízení výrazně zaostává za pokrokem, který vidíme u aplikací pro zábavu a trávení volného času. To neznamená, že by neexistovala kvalitní řešení, umožňující postiženým ovládat počítač za pomoci počítačového vidění (obr. 1.). V podstatě bez výjimky jsou však založena na velmi kvalitních kamerách, pracujících často s přísvitem v infračerveném spektru, a jejich cena se pohybuje v desítkách tisíc korun. Většina z nich vyžaduje doplňky či přídatná zařízení, která je potřeba kalibrovat na konkrétního uživatele, a jejich užití je tedy složité. Řešení prezentované v tomto článku se naopak zaměřuje na usnadnění přístupu k veřejně dostupným počítačům, jako jsou ty v knihovnách a internetových kavárnách, případně nenákladnou alternativu k výše uvedeným nástrojům pro jednotlivce.



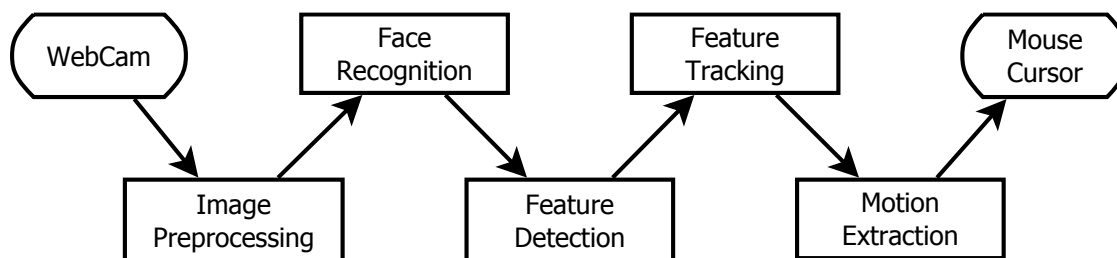
**Obrázek 1:** Příklady existujících systémů. Zleva - SmartNav<sup>®</sup>, Tracker Pro<sup>®</sup>, I4Control<sup>®</sup>

## 2. NÁVRH A REALIZACE

Cílem této práce je navrhnout, implementovat a do testovacího provozu uvést prototyp softwaru, využívajícího pouze vstupu z jedné běžné webkamery k nahrazení funkcí myši. Aplikace je zamýšlena pro uživatele s postižením rukou a sleduje proto pohyby hlavy. Vzhledem k zaměření aplikace není potřeba dosahovat preciznosti moderních myší v náročných aplikacích (akční hry, CAD/CAM aplikace, jiné profesionální nástroje), nicméně musí být rozumně použitelná pro běžné úkony (webové prohlížeče, komunikační programy, zadávání textu na virtuální klávesnici, atd.).

## 2.1. EXTRAKCE POHYBOVÉ INFORMACE

O extrakci informace z pohybu hlavy uživatele se stará funkční knihovna pro komunikaci s webkamerou, která dále předává data samotnému prvku uživatelského rozhraní ovládajícímu chování kurzoru. Knihovna je modulární s důrazem na čistá rozhraní a znovupoužitelnost kódu v jiných projektech. Je strukturována jako kaskáda sériově zapojených modulů. Snímek z kamery je nejdříve upraven a předán detektoru obličeje Viola-Jones. V takto získané oblasti s obličejem jsou nalezeny významné body Harrisovým detektorem rohů. Tyto body jsou následně sledovány metodou optického toku Lucas-Kanade a ze změny jejich polohy je pak vypočtena změna polohy sledovaného objektu (v našem případě obličeje). Pro více informací o těchto metodách, viz např. [1][2]. Obecná struktura výpočtu je znázorněna na obr. 2.



Obrázek 2: Struktura výpočtu.

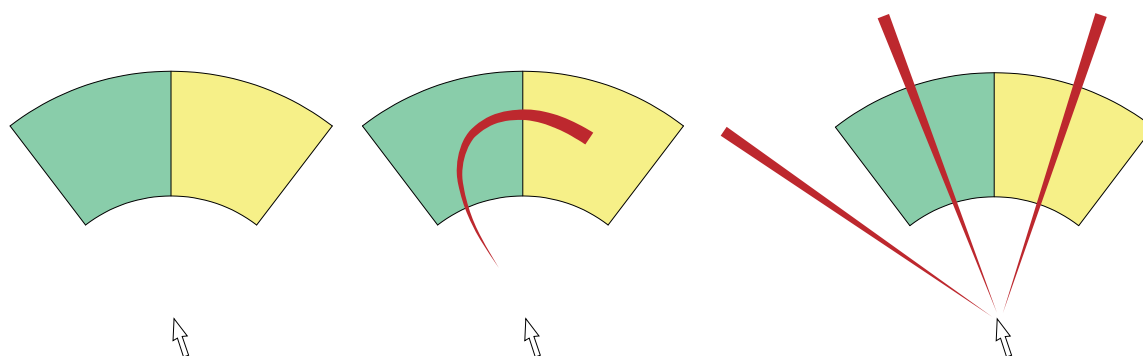
Použitá kombinace metod je dostatečně rychlá a za rozumných světelných podmínek, případně v regulovaném prostředí (např. možnost přísvitů obličeje stolní lampou), též velmi robustní. Její největší slabinou jsou scény s nízkou hladinou osvětlení, případně scény s protisvětlem, nebo silným bočním osvětlením. Levné webkamery v takovém případě předávají velmi kontrastně nevyvážený obraz a přesnost metody se zhoršuje. V současné době probíhá testování různých sad algoritmů, které by odstranily či alespoň do jisté míry potlačily tento nedostatek.

## 2.2. NAHRAZENÍ FUNKCÍ MYŠI

Přestože přizpůsobení uživatelských rozhraní jednotlivých aplikací pro ovládání pomocí webkamery by přineslo lepší výsledky, z důvodu malé univerzálnosti byl tento přístup zavrhnut. Místo toho byla zvolena snaha co nejdříve simulovat funkci klasické myši, neboť pro toto vstupní zařízení je optimalizována většina dnešních aplikací pro osobní počítače. Problematikou prvků uživatelských rozhraní bez přítomnosti klávesnice a myši se zabývá např. [3].

Samotný prvek uživatelského rozhraní pro ovládání počítačových aplikací využívá sledování pohybu hlavy pro polohování kurzoru myši. Nelineární akcelerace kurzoru umožňuje přesné pohyby na malé vzdálenosti a současně pokrývá celou oblast plochy. Po určité době nečinnosti se zobrazí oblast pro simulaci funkcí tlačítek myši. Tvar oblasti je navržen tak, aby jakýkoliv přímočarý pohyb kurzoru nezpůsobil spuštění akce, a tím se zamezilo nechtěným klikům myši (obr. 3). Pro spuštění specifické akce je potřeba navštívit kurzorem více než jeden segment oblasti v daném pořadí. Pokud systém rozezná gesto, provede akci v místě původního zastavení pohybu kurzoru. V opačném případě skryje oblast pro výběr akce a pokračuje v pohybu kurzorem. Velikost, tvar a orientace této oblasti stejně jako citlivost a akcelerační křivku kurzoru lze jednoduše měnit a přizpůsobit tak preferencím uživatele. Experimentálně byl podobný systém generování jednoduchých příkazů implementován např. v práci [4] pro dálkové ovládání televizoru.

V současné době jsou nahrazeny čtyři typy akcí - jednoduché kliknutí levým a pravým tlačítkem, držení levého tlačítka pro funkce jako je drag&drop a přesun okna po ploše a nakonec levý dvojklik. Aktivace těchto akcí je implementována za pomoci knihovnic funkcí správce oken daného operačního systému.



**Obrázek 3:** Oblast pro ovládání tlačítek myši (vlevo), gesto pro pravý klik (uprostřed), přímočarý pohyb myši kterýmkoliv směrem nespustí žádnou akci (vpravo)

### 3. ZÁVĚR

Navržená aplikace v současnosti funguje na platformách Ubuntu Linux a Windows XP a vyšší. Hlavním jazykem projektu je Python ve verzi 2.7 a knihovna OpenCV 2.3. Na počítači Eee PC 901 (Intel Atom N270, 1GB RAM), což je realisticky hluboko pod výkonem systémů, na kterých se předpokládá nasazení aplikace, dosahuje při použití integrované webkamery rychlost zpracování 9-10 snímků za vteřinu za špatných a 27-28 snímků za vteřinu za dobrých světelných podmínek. Teoretické maximum kamery při použití rozlišení snímku 640x480 pixelů je 30 snímků za vteřinu. Prozatímni testy ukazují, že i při tomto malém rozlišení vstupního videa lze bez problémů dosahovat dostatečné přesnosti umístění kurzoru vzhledem k velikosti běžných prvků uživatelských rozhraní (výběr textu, výběr z nabídky, aktivace odkazu na webu). Při zajištění alespoň průměrné kvality osvětlení a při použití softwarové klávesnice operačního systému umožňuje aplikace většinu běžných úkonů pouze pohybem hlavy a to bez neúměrného zatížení uživatele. V současné době je téměř dokončen prototyp, který bude ve spolupráci se Svazem tělesně postižených v České republice, o. s., Městskou organizací Brno rozšířen spolu s dotazníkem mezi cílovou komunitu uživatelů.

### PODĚKOVÁNÍ

Tento projekt vznikl ve spolupráci se Svazem tělesně postižených v České republice, o. s., Městskou organizací Brno.

### REFERENCE

- [1] Viola, P., Jones, M.: Robust Real-time Object Detection. Vancouver, Canada, 2001
- [2] Jianbo Shi, Tomasi, C.: Good features to track. In Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on , vol., no., pp.593-600, 21-23 Jun 1994
- [3] Bailly, G., Walter, R., Müller, J., Ning, T., Lecolinet, E.: Comparing Free Hand Menu Techniques for Distant Displays using Linear, Marking and Finger-Count Menus. IFIP INTERACT'11. 248-262, 2011
- [4] Miyake, T., Asakawa, T., Yoshida, T., Imamura, T., Zhong Zhang: Detection of view direction with a single camera and its application using eye gaze. In Industrial Electronics, 2009. IECON '09. 35th Annual Conference of IEEE, vol., no., pp.2037-2043, 3-5 Nov. 2009