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

ROZPOZNÁNÍ JEDNODUCHÝCH GEST RUKY

BAKALÁŘSKÁ PRÁCE
BACHELOR´S THESIS

AUTOR PRÁCE Jan Verner
AUTHOR

BRNO 2007

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

ROZPOZNÁNÍ JEDNODUCHÝCH GEST RUKY

SIMPLE HAND GESTURE RECOGNITION

BAKALÁŘSKÁ PRÁCE
BACHELOR´S THESIS

AUTOR PRÁCE Jan Verner
AUTHOR

VEDOUCÍ PRÁCE Ing. Igor Potúček, Ph.D.
SUPERVISOR

BRNO 2007

LICENČNÍ SMLOUVA
POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

1. Pan/paní

Jméno a příjmení: Jan Verner

Bytem: A. Slavíčka 5, 796 04 Prostějov 4

Narozen/a (datum a místo): 20.11.1984, Olomouc

(dále jen „autor“)

a

2. Vysoké učení technické v Brně

Fakulta ……..……

se sídlem …….……..

jejímž jménem jedná na základě písemného pověření děkanem fakulty:

.....

(dále jen „nabyvatel“)

Čl. 1
Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):
   - □ disertační práce
   - □ diplomová práce
   - □ bakalářská práce
   - □ jiná práce, jejíž druh je specifikován jako .....

   (dále jen VŠKP nebo dílo)

Název VŠKP: .....

Vedoucí/ školitel VŠKP: .....

Ústav: .....

Datum obhajoby VŠKP: .....

VŠKP odevzdal autor nabyvateli v*:

□ tištěné formě – počet exemplářů ……..

□ elektronické formě – počet exemplářů ……..

* hodící se zaškrtněte

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.
3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.
4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

Článek 2
Udělení licenčního oprávnění

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užít, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizovaní výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti
   - □ ihned po uzavření této smlouvy
   - □ 1 rok po uzavření této smlouvy
   - □ 3 roky po uzavření této smlouvy
   - □ 5 let po uzavření této smlouvy
   - □ 10 let po uzavření této smlouvy

   (z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/ 1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

Článek 3
Závěrečná ustanovení

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísni a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne: …….

…….. ……
Nabyvatel Autor

Abstrakt

Tato bakalářská práce pojednává o vhodnosti použití algoritmu DTW a jeho variant pro rozpoznávání jednoduchých gest ruky. Shrnuje dosavadní poznatky z oboru počítačového vidění a zaměřuje se na metody detekce ruky a obecný popis algoritmu DTW.

Klíčová slova

gesta ruky, rozpoznávání, DTW, DDTW, počítačové vidění, detekce obličeje, detekce ruky, zpracování obrazu, sledování pohybu

Abstract

This bachelor thesis describes usability of DTW algorithm and it’s variants for simple hand gesture detection. It recapitulates actual situation and findings in branch of computer vision and aims on methods for hand gesture recognition and common description of DTW algorithm.

Keywords

hand gesture, recognition, DTW, DDTW, computer vision, face detection, hand detection, image processing, motion tracking

Citace

Jan Verner: Rozpoznání jednoduchých gest ruky, bakalářská práce, Brno, FIT VUT v Brně, 2007

Rozpoznání jednoduchých gest ruky

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením
Ing. Igora Potúčka, Ph.D.
Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

……..……
Jméno Příjmení
Datum

Poděkování

Na tomto místě bych velmi rád poděkoval vedoucímu mé bakalářské práce za jeho přístup, poskytnuté znalosti a zkušenosti a za velkou ochotu při realizaci tohoto projektu.
Také bych rád poděkoval svým rodičům, Markétě Daňkové a Petrovi Zatloukalovi za trpělivost a pomoc při testování systému a za připomínky k této práci.
Bez podpory žádného z nich by tento projekt nebyl úspěšný. Všem výše jmenovaným z celého srdce

Děkuji!

© Jan Verner, 2007.
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

1 Úvod 2
2 Úvod do problematiky 3
2.1 Cíl práce 3
2.2 Zpracování obrazu 3
2.3 Vektor příznaků 5
2.4 Rozpoznání gest 7
3 Popis řešení 13
3.1 Celkový pohled na systém 13
3.2 Záznam obrazu 14
3.3 Zpracování obrazu 16
3.4 Zpracování vektorů příznaků 17
3.5 Určení gesta 20
4 Testování a vyhodnocení 21
4.1 Úspěšnost detekce částí těla 21
4.2 Algoritmus DTW 22
4.3 Algoritmus DDTW 23
5 Závěr 28

1 Úvod

Během posledních deseti let zaznamenala oblast počítačového vidění nebývalý rozvoj. Výsledky můžeme vidět již nyní. Velkého pokroku bylo dosaženo v oboru lékařství. Přístroje pro diagnostiku a včasné odhalování chorob, založené na poznatcích o zpracování obrazu, denně pomáhají zachraňovat stovky životů. Znalostí z oblasti počítačového vidění jsou využívány také pro vojenské účely. Systémy pro detekci nepřítele nebo navádění raket se staly běžnou součástí výzbroje každé moderní armády. Jednou z nejnovějších oblastí aplikace dosavadních znalostí se stal vesmírný výzkum. Počítačem řízení roboti zkoumající planety, někdy vzdálené i stovky milionů kilometrů, umožňují poznávat blíže historii celého vesmíru.

Se stále zvyšujícím se výkonem běžných stolních počítačů se tyto systémy dostávají blíže běžnému životu. Jmenujme například aplikace pro rozpoznávání a určování osob. Zajímavou oblastí je také virtuální realita, která využívá poznatky z oblasti rozpoznávání pohybu. Již v brzké době se tak dočkáme aplikací, které budou jistě usnadňovat život nám všem.

Cílem této bakalářské práce je navázat na dosavadní výzkum a spojit poznatky z několika oblastí. Těmi hlavními jsou detekce částí lidského těla a rozpoznání mluvené řeči. Výsledkem je systém schopný určovat gesta ruky. Ten lze pak jednoduše integrovat do různých informačních a zabezpečovacích systémů, počítačových her a nebo jej použít ve výše zmiňované virtuální realitě.

Celá práce je rozdělena do několika kapitol a podkapitol.

Po krátkém úvodu následuje teoretická část, ve které jsou formulovány jednotlivé cíle práce, zmíněny metody používané pro detekci částí lidského těla, algoritmy pro rozpoznávání gest a detailně popsán jeden z nich. V této kapitole jsou také zmíněny problémy, které se pojí s danou úlohou.

Další kapitola, nazvaná „Popis řešení“, obsahuje celkový pohled na navržený systém. Pomocí obrázků jsou blíže popsány jednotlivé jeho části a to tak, jak spolu navzájem spolupracují. Tato kapitola se odkazuje zejména na předešlou teoretickou část.

Čtvrtá část se zaměřuje na testování, celkové vyhodnocení chování systému v nejrůznějších situacích a srovnává dva základní přístupy k problému. Ve formě grafů a tabulek podává ucelený přehled o vlastnostech navrženého řešení.

Závěrečná kapitola poté shrnuje dosažené výsledky a naznačuje další vývoj aplikace.

2 Úvod do problematiky

Tato kapitola se zaměřuje především na teoretický úvod do celé problematiky. Formuluje cíle práce, problémy, které mohou vzniknout, a jejich možná řešení. Jsou zde popsány některé metody v současnosti používané pro detekci částí lidského těla a rozpoznání gest, jejich výhody a nevýhody. Klíčové metody jsou pak dále podrobněji rozvedeny.

Na tuto kapitolu navazuje kapitola 3 nazvaná „Popis řešení“, která využívá zde zmíněných teoretických poznatků.

2.1 Cíl práce

Cílem práce je dle zadání navrhnout systém pro rozpoznání jednoduchých gest ruky. Tato úloha má z praktického hlediska několik částí.

1. nalézt vhodné metody pro detekci částí lidského těla
2. nalézt vhodné metody pro určení gest ruky
3. navržený systém popsat a implementovat
4. otestovat a zhodnotit celkové chování systému

Tyto 4 body definují postup celé práce a jsou hlavním vodítkem. Bod číslo 1 rozpracovává kapitola 2.2. Bodu číslo 2 se věnuje kapitola 2.4, bod číslo 3 je pak popsán v kapitole 3. Bod číslo 4 je rozveden v kapitole 4.

2.2 Zpracování obrazu

Výzkum v oblasti počítačového zpracování obrazu je v současné době na velmi vysoké úrovni. Mnoho studií takřka denně přináší nové a nové poznatky. Cílem této podkapitoly je shrnout doposud známe metody zpracování obrazu a blíže popsat ty nejdůležitější.

Metody pro detekci ruky

Nejznámější postup nalezení ruky v obraze je založen na rozpoznání barvy kůže [1]. Celý snímek je zpracován bod po bodu a podle určitého vzoru jsou vybrány jen ty body, které odpovídají barvě pleti. Hlavní výhodou této metody je její jednoduchost a intuitivnost. Problém ovšem je, jak vhodně zvolit vzor, podle kterého se bude pleť rozpoznávat. Existuje několik přístupů [1], žádný z nich se však v praxi neukázal jako univerzální a vhodný.

Částečným řešením je převod celého snímku do jiného barevného prostoru, avšak i toto řešení není dokonalé. Nebere totiž v úvahu proměnlivé osvětlení scény, ani to, zda detekujeme ruku bělocha nebo černocha.

Musíme tedy nalézt univerzálnější postup a vycházet budeme právě z metody klasifikace jednotlivých bodů obrazu.

Pro naše účely nalezení a sledování ruky v obraze využijeme tzv. „*mean shift algoritmus*“ [2]. Využívá převodu snímku do HSV (*hue, saturation, value*) barevného prostoru [3] [4] a pravděpodobnostního rozložení [5].

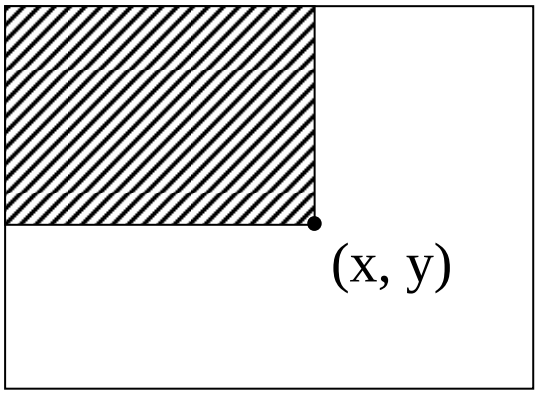
V první fázi je třeba vytvořit za použití histogramu barevný model sledovaného objektu. Poté lze přistoupit k samotnému zpracování. Snímek je nejdříve převeden do HSV barevného prostoru. Za pomoci předem vytvořeného modelu je určena pravděpodobnost, s jakou jednotlivé body obrazu odpovídají danému modelu. Následně je nalezeno místo s nejvyšší pravděpodobností výskytu daného objektu.

Tato metoda má ještě svá vylepšení, která umožňují určit výšku a šířku sledovaného objektu, případně i jeho natočení. Poskytují tak další cenné informace, které lze využít v další fázi – rozpoznávání gest.

Metody pro detekci obličeje

Pro detekci obličeje lze opět použít výše zmíněné algoritmy, zpracovávající jednotlivé body obrazu. Existuje však mnohem sofistikovanější vizuální metoda [6]. Využívá množinu předem natrénovaných vzorů sledovaného objektu, která uchovává informaci o rozložení kontrastu ve snímku. Poskytuje velmi přesné výsledky v relativně krátkém čase.

Před zahájením detekce je třeba natrénovat klasifikátor pomocí stovek vzorových obrázků, které obsahují obličeje. Klasifikátor uchovává speciální reprezentaci obrazu zvanou „integrální snímek“. Ten může být z původního obrazu spočítán velmi jednoduše, pomocí několika operací pro každý bod.



obrázek 1: hodnota integrálního snímku v bodě x, y

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y') \tag{2.1}$$

$ii(x,y)$ je hodnota integrálního snímku v bode x, y a $i(x',y')$ je původní snímek. Na rozdíl od většiny detekčních systémů, které pracují přímo s intenzitami ve snímku, toto řešení nabízí především zvýšení výkonu. Při rozhodovacím procesu lze v *konstantním* čase určit, zda se v dané části snímku nachází obličej či nikoliv.

Klasifikátor je také navržen tak, že je možné rychle upravit jeho velikost a aplikovat jej tak na vetší či menší část snímku. Pokud tedy chceme určit pozici a velikost sledovaného objektu v neznámém obraze, klasifikační proces opakujeme na jednotlivé části obrazu a přitom zvětšujeme postupně velikost klasifikačního okna.

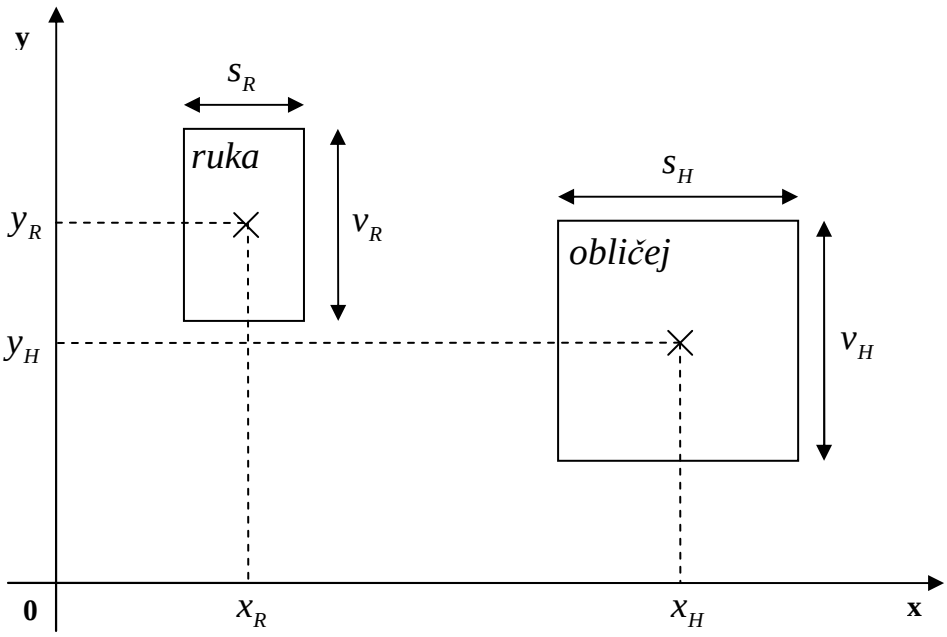
2.3 Vektor příznaků

Pro další výklad je nutné zavést nový pojem „vektor příznaků“. Jedná se o vektor čísel, který nese informace o pozici a velikosti hlavy a ruky v obraze.

Pro naše účely může mít vektor příznaků následující podobu:

$$\vec{v}(x_R, y_R, v_R, s_R, a_R, x_H, y_H, v_H, s_H) \tag{2.2}$$

kde jednotlivá čísla x_R, y_R, v_R, s_R, a_R nesou informace postupně o pozici ruky na ose x a y, a dále o její výšce, šířce a natočení. Podobně je tomu u hlavy – čísla x_H, y_H, v_H, s_H.



obrázek 2: zachycuje význam jednotlivých čísel ve vektoru příznaků,
v obrázku není pro zjednodušení zachycen příznak natočení ruky a_R ,
počátek souřadnicového systému se vztahuje ***k levému dolnímu okraji snímku***

Hodnoty jednotlivých složek vektoru příznaků však silně závisejí na několika faktorech, které musíme vzít v úvahu. Těmi nejvýznamnějšími jsou:

- rozlišení obrazu
- proměnlivá vzdálenost sledované osoby od kamery snímající obraz
- různě zvolené záběry (výřezy) a umístění kamery
- ohnisková vzdálenost kamery

Je nutné proto provést takové úpravy vektoru příznaků, abychom vliv těchto faktorů odstranili. K tomu musíme dané vektory *normalizovat* – převést je do jednotné soustavy souřadnic.

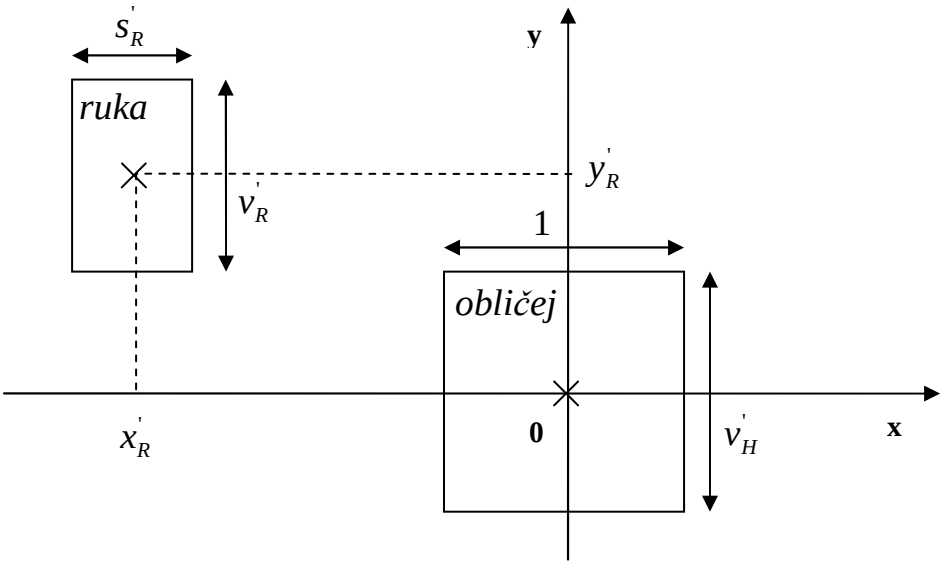
V prvním kroku normalizace prohlásíme střed obličeje (x_H, y_H) za střed nového souřadnicového systému. Poté vektor upravíme následujícím způsobem:

$$\vec{v}_{norm}\left(\frac{x_R - x_H}{s_H}, \frac{y_R - y_H}{s_H}, \frac{v_R}{s_H}, \frac{s_R}{s_H}, a_R, 0, 0, \frac{v_H}{s_H}, \frac{s_H}{s_H}\right). \qquad (2.3)$$

Získáme tak normalizovaný vektor příznaků, který má následující podobu:

$$\vec{v}_{norm}(x_R^{'}, y_R^{'}, v_R^{'}, s_R^{'}, a_R, 0, 0, v_H^{'}, 1). \qquad (2.4)$$

Celou situaci zachycuje obrázek 3.



*obrázek 3: význam jednotlivých čísel v **normalizovaném** vektoru příznaků, v obrázku není pro zjednodušení zachycen příznak natočení ruky a_R , počátek souřadnicového systému se vztahuje **ke středu obličeje***

Normalizací tedy docílíme toho, že vektor příznaků již nepopisuje pozici a rozměry hlavy a ruky v obraze, ale pouze pozici a rozměry ruky. Od toho okamžiku lze hovořit o tom, že vektor

příznaků nese pouze informaci o pozici ruky vzhledem k hlavě. Je nutné mít na paměti, že ve skutečnosti jsou v něm zakódovány i informace o hlavě.

S takto normalizovaným vektorem příznaků pak dále pracujeme a na základě jeho hodnot určujeme jednotlivá gesta. Více v kapitole 2.4.

2.4 Rozpoznání gest

Problém rozpoznání gest souvisí velice úzce s problematikou rozpoznávání mluvené řeči a lze také z této oblasti získat některé důležité poznatky. Budeme vycházet z předpokladu, že gesto, jehož význam chceme určit, máme izolované a popsané *n* vektory normalizovaných příznaků $\vec{v}_{norm}(n)$, které v dalším textu budeme nazývat jednoduše „*vektory příznaků*“ $\vec{a}(n)$. Pro další výklad tato záměna nemá vliv.

Při rozpoznávání gest se, stejně jako u rozpoznávání mluvené řeči, setkáváme s několika problémy, které je nutné brát v úvahu.

Mezi ty nejdůležitější patří samotný člověk. Ten nevykoná žádné gesto 2-krát za sebou přesně stejně. Zároveň gesta různých osob se navzájem velmi liší, lidé mohou dělat gesta levou i pravou rukou, což rozpoznání komplikuje. Zároveň také požadujeme, aby celý proces probíhal v reálném čase. Další faktory, které musíme při výběru metody zohlednit, jsou například velikost slovníku gest, prostředí, ve kterém bude klasifikátor pracovat, atp. Musíme proto hledat takové postupy, které všechny výše zmiňované problémy buďto přímo řeší nebo je přinejmenším zohledňují.

První velkou skupinu klasifikátorů tvoří metody založené na statistickém modelování pomocí tzv. Skrytých Markovových modelů [7]. Pro každou třídu stejných slov ve slovníku jsou v procesu trénování stanoveny parametry modelu a neznámé gesto je zařazeno do té třídy, jejíž model jej generuje s největší pravděpodobností. Trénování modelu je velice časově náročná činnost. Proto tyto metody ponecháme stranou.

Druhou velkou skupinu tvoří klasifikátory, které pracují s gestem jako celkem. Snaží se zařadit dané gesto do té třídy, k jejímuž obrazu má „nejmenší vzdálenost“. Klíčovou otázkou je zde právě určení oné vzdálenosti. Tato je obvykle vypočítána na základě aplikace metody dynamického programování, která je blíže popsána v následující podkapitole. Tato metoda je pro náš účel velmi vhodná. Natrénování slovníku gest je proces relativně krátký a především velmi snadný. Umožňuje tak velkou variabilitu a široký rozsah použití.

Metoda dynamického programování

Tato metoda je aplikovatelná u vektorových klasifikátorů a jejím cílem je určení vzdálenosti dvou obrazů gest.

Při popisu metody budeme vycházet z poznatků z oblasti rozpoznávání mluvené řeči. Zde se při důkladném rozboru signálů získaného vyslovením slova týmž řečníkem několikrát po sobě zjistilo, že hlavní rozdíly signálů jsou v časové oblasti. Jednalo se zejména o nestejnou délku a především o kolísání rychlosti uvnitř slova. A stejný problém nalezneme, budeme-li se bavit o rozpoznávání gest.

Nestejnou délku gesta lze odbourat lineární transformací signálu. Ta ovšem nedokáže postihnout časové členění uvnitř gesta. Algoritmus, který tento problém řeší, pracuje s nelineární časovou transformací. Odtud jeho název DTW – *dynamic time warping* [8] (dynamické borcení času). Snaží se eliminovat rozdíly pomocí „borcení“ časové osy a dosáhnout co největší shody se vzorovým obrazem.

V následujícím textu bude zmíněn základní princip algoritmu DTW, popsána metoda měření vzdálenosti obrazů gest a diskutována některá jeho vylepšení.

Základní algoritmus DTW

Pro další výklad si zavedeme označení obrazu testovaného gesta jako:

$$A = \{\vec{a}(1), \vec{a}(2), \ldots, \vec{a}(n), \ldots, \vec{a}(I)\} \tag{2.5}$$

a obrazu referenčního gesta jako:

$$B = \{\vec{b}(1), \vec{b}(2), \ldots, \vec{b}(m), \ldots, \vec{b}(J)\}. \tag{2.6}$$

Přitom $\vec{a}(n)$ je *n*-tý vektor příznaků testovaného slova a $\vec{b}(m)$ je *m*-tý vektor příznaků referenčního slova. Algoritmus DTW přitom hledá v rovině (*m*,*n*) optimální cestu (*m* je funkcí *n*)

$$m = \psi(n), \tag{2.7}$$

která minimalizuje funkci *D* celkové vzdálenosti mezi obrazy *A* a *B*

$$D(A,B) = \sum_{n=1}^{I} \hat{d}\left(\vec{a}(n), \vec{b}(\psi(n))\right), \tag{2.8}$$

přičemž $\hat{d}\left[\vec{a}(n), \vec{b}(\psi(n))\right]$ je lokální vzdálenost mezi *n*-tým vektorem příznaků testovaného slova a *m*-tým $m = \psi(n)$ vektorem příznaků referenčního slova.

Výpočet lokální vzdálenosti $\hat{d}$ je silně závislý na konkrétním typu měřených příznaků. Pro dva vektory o délce *V*, které mají tvar

$$\begin{aligned} &\vec{a}\left(a_1, a_2, ..., a_V\right) \\ &\vec{b}\left(b_1, b_2, ..., b_V\right) \end{aligned} \tag{2.9}$$

lze určit lokální vzdálenost jako:

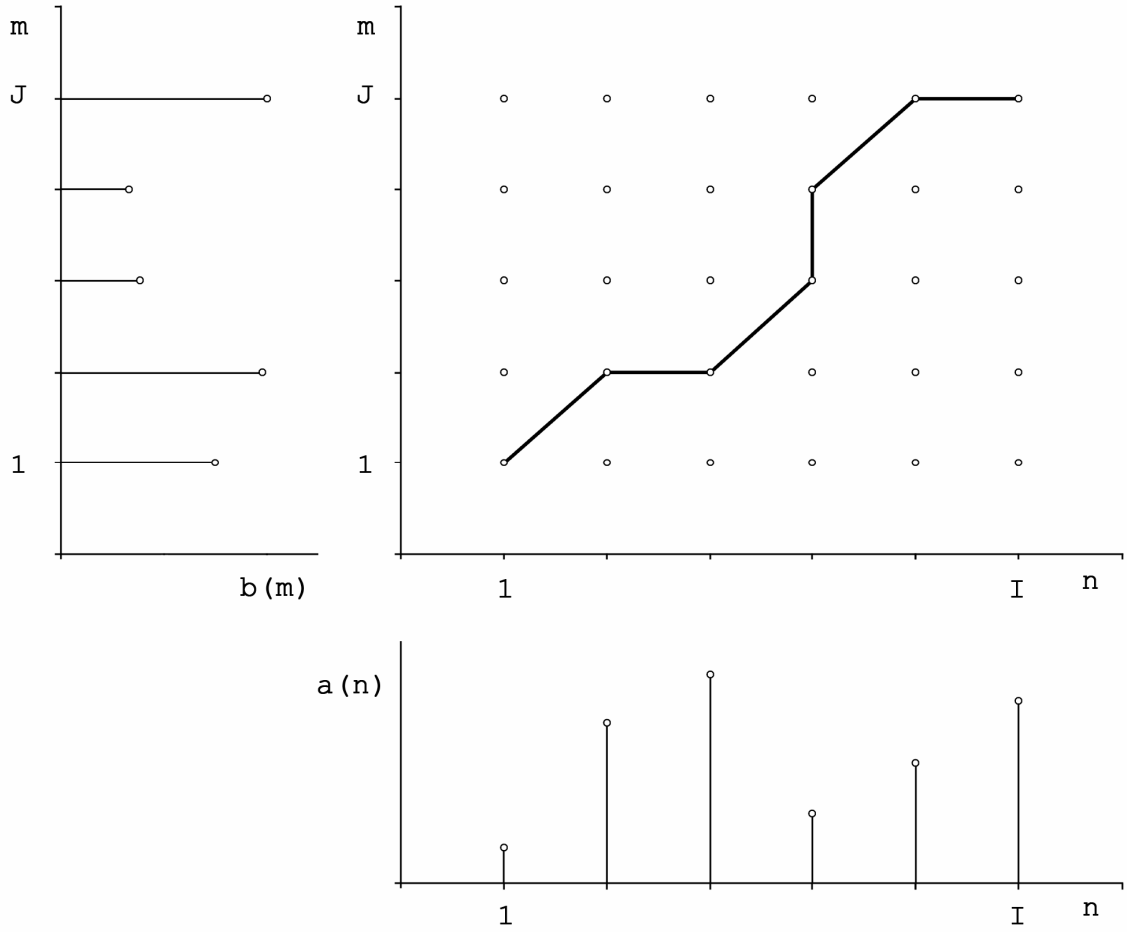
$$\hat{d} = \sqrt{\sum_{v=1}^{V}\left(a_v - b_v\right)^2} \tag{2.10}$$

Dále zavedeme časovou proměnnou *k* a proměnné *m* a *n* vyjádříme jako funkci *k*:

$$\begin{aligned} &n = i(k) \\ &m = j(k) \quad , \\ &k = 1, ..., K \end{aligned} \tag{2.11}$$

kde *K* je délka obecné časové osy pro dané porovnání obrazů A a B.

Celou situaci přehledně znázorňuje obrázek 4.



obrázek 4: způsob mapování jednotlivých bodů testovaného a referenčního slova v rovině (m,n)

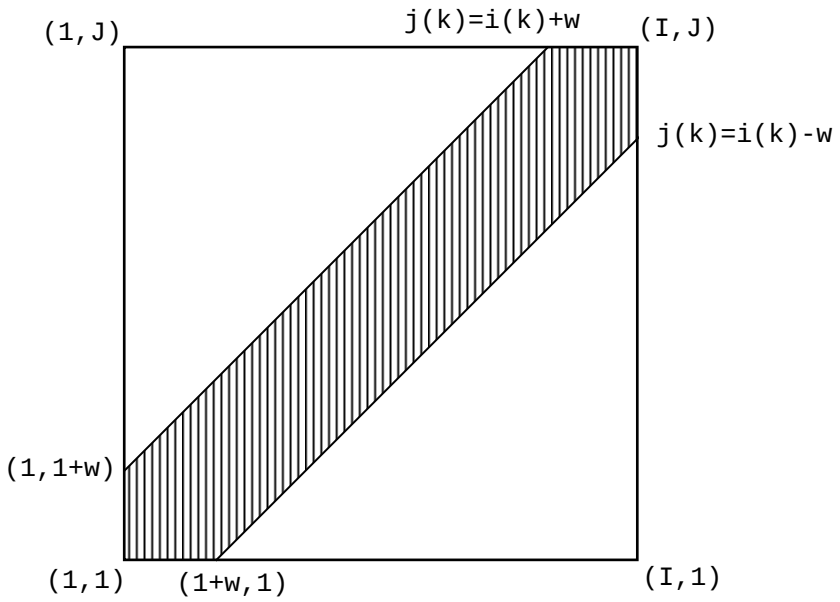
Z výše uvedeného vyplývá jedna podmínka. První a poslední vektor příznaků testovaného a referenčního slova musí být mapovány na sebe

$$\begin{aligned} i(1) = 1,\ j(1) = 1 \\ i(K) = I,\ j(K) = J \end{aligned} . \tag{2.12}$$

Dále se pro omezení pohybu funkce DTW zavádí okénková proměnná *w*. Ta omezuje prostor pohybu na malou část celého prostoru roviny *(m,n)*. Odráží tak skutečnost, že kolísání tempa uvnitř gesta může být jen v určitých mezích. Pro proměnnou *w* platí následující vztah:

$$\left| \frac{J}{I} i(k) - j(k) \right| \leq w . \tag{2.13}$$

Celou situaci znázorňuje obrázek 5.



obrázek 5: vymezení globálního pohybu funkce DTW v rovině (m,n) pomocí proměnné w

Existují i další omezení pohybu funkce DTW. Ty zde však již uvádět nebudeme, jelikož nemají pro další výklad význam.

Určení vzdálenosti

Obecný vztah pro určení skutečné minimální celkové vzdálenosti mezi obrazy A a B lze tedy vyjádřit vztahem

$$D(A,B) = [I+J]^{-1} \left\{ \min_{\{i(k),j(k),K\}} \sum_{k=1}^{K} \hat{d}\left[i(k), j(k)\right] \hat{W}(k) \right\}, \qquad (2.14)$$

kde $\hat{d}\left[i(k), j(k)\right]$ je lokální vzdálenost mezi *n = i(k)-tým* mikrosegmentem testovaného obrazu A a *m=j(k)-tým* mikrosegmentem referenčního obrazu B a $\hat{W}(k)$ je váhová funkce pro *k-tý* úsek funkce DTW

$$\hat{W}(k) = \left[i(k) - i(k-1)\right] + \left[j(k) - j(k-1)\right] \qquad (2.15)$$

a její hodnota závisí na lokální cestě.

Vylepšení algoritmu - DDTW

Jedním z problémů, se kterými se potýká algoritmus DTW, je nesprávné mapování jednotlivých vektorů příznaků navzájem na sebe. K tomuto jevu dochází jen za určitých okolností a snižuje přesnost algoritmu. Z tohoto důvodu byla zavedena varianta algoritmu DTW, která má název DDTW [9] (*Derivated dynamic time warping*). Rozdíl je pouze v tom, že algoritmus DDTW počítá s derivacemi vektorů příznaků.

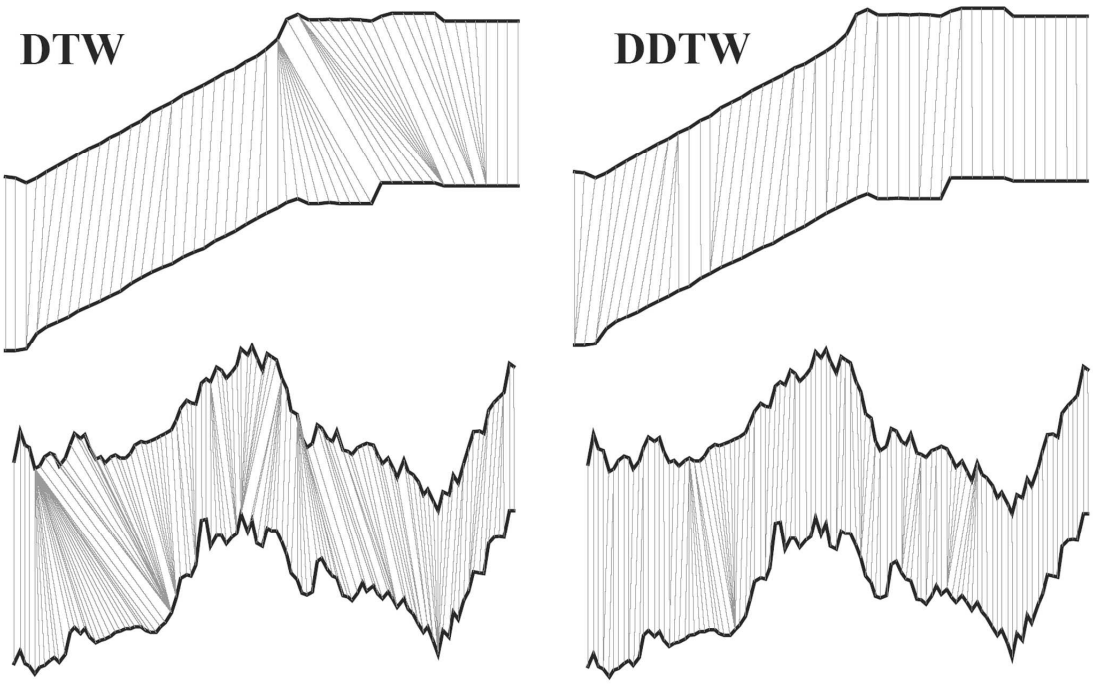
Pro každou *v*-tou složku obecného *n*-tého vektoru příznaků

$$\vec{a}_n\left(a_{1n}, a_{2n}, ..., a_{vn}, ..., a_{Vn}\right), n = 2...N-1, v = 1...V \tag{2.16}$$

lze její derivaci spočítat jako

$$a'_{vn} = \frac{\left(a_{vn} - a_{vn-1}\right) + \left(\left(a_{vn+1} - a_{vn-1}\right)/2\right)}{2}, n = 2...N-1, v = 1...V \tag{2.17}$$

Takto upravený algoritmus pracuje mnohem přesněji, jak to dokládá obrázek 6.



obrázek 6: rozdíl v mapování 2 signálů na sebe, vlevo funkce DTW, vpravo funkce DDTW

3 Popis řešení

Teorie nezbytná pro další výklad byla shrnuta v kapitole 2. S odkazem na ni se dále budeme věnovat popisu samotného systému. V úvodu bude prezentována celková struktura řešení, její jednotlivé části a to jak spolu navzájem spolupracují. Klíčovým bodům systému pak bude věnována větší pozornost a budou popsány blíže. Výsledkem je systém, který je schopný určovat gesta ruky.

3.1 Celkový pohled na systém

Tato podkapitola se věnuje obecnému pohledu na systém. Jsou zde nastíněny základní principy jeho fungování a obecně popsány 4 fáze procesu. Ty jsou zachyceny na obrázku 7.

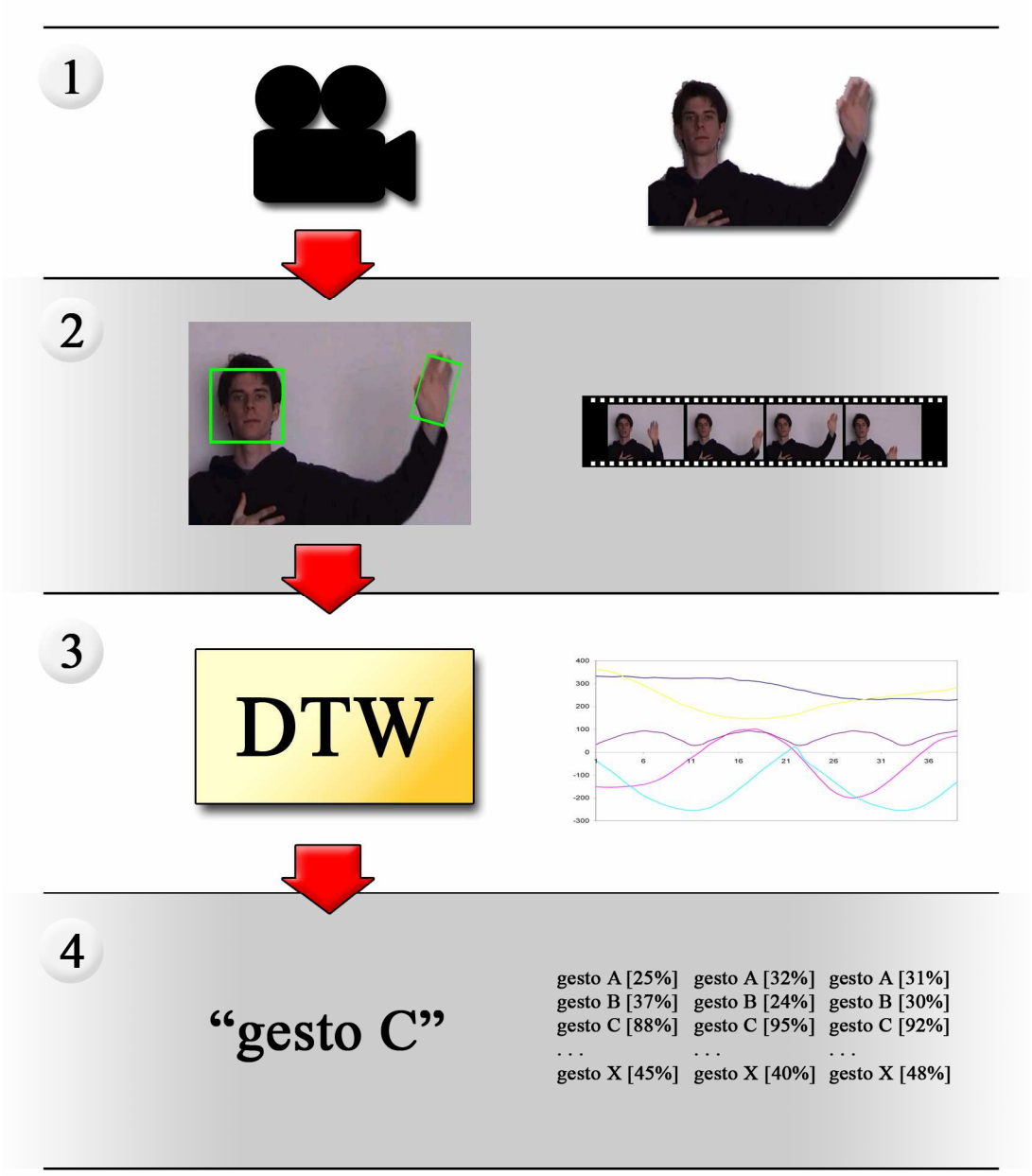
V první fázi procesu určování gest stojí daná osoba před kamerou. Cílem je pořídit sekvenci snímků pro druhou fázi. Osoba může gesta vykonávat pravou i levou rukou, nahoře nebo dole, rychle či pomalu. Všechny tyto rozdíly řeší následující dvě fáze.

Ve druhém kroku, je každý snímek sekvence zpracováván zvlášť. Jsou v něm detekovány části těla – hlava a ruka (kapitola 2.2) a na základě jejich pozice je získán vektor příznaků (kapitola 2.3). Ten je následně normalizován. Vzorkování vektorů příznaků pochopitelně probíhá v diskrétním čase. Každému snímku je tak přiřazen jeden normalizovaný vektor příznaků, který popisuje určitý stav ruky v obraze vzhledem k pozici hlavy.

Třetí fáze pracuje se sérií normalizovaných vektorů příznaků a využívá algoritmus DTW (kapitola 2.4). Jejím cílem je porovnat testovanou sérii se všemi referenčními sériemi, které jsou uloženy v tzv. „slovníku gest“, a pro každou z nich určit vzdálenost. Jedná se o číselný údaj, který vyjadřuje míru podobnosti testovaného a referenčního gesta. Tvoří jádro celého systému a je na ni kladen velký důraz.

Úkolem čtvrté fáze je samotné určení gesta. Přebírá výsledky analýz ze 3. fáze a na základě jejich historie jednoduchým algoritmem vyhodnocuje, zda bylo či nebylo provedeno nějaké gesto a určuje jeho význam.

Systém se navenek chová jako celek a jeho jednotlivé části navzájem spolupracují. Jakákoliv chyba v libovolné fází se promítá dále. Snahou je tedy tyto chyby eliminovat, popřípadě vytvořit natolik robustní systém, který bude k těmto chybám tolerantní.

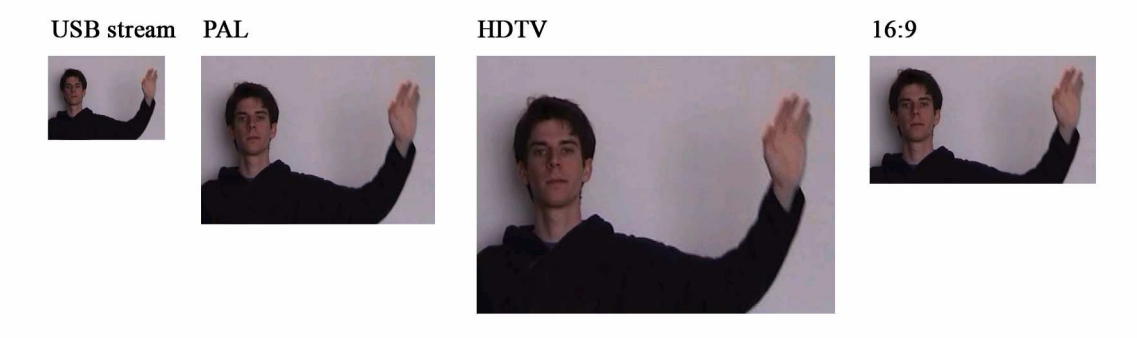


obrázek 7: celkový pohled na systém a na jeho 4 základní fáze:
1. záznam obrazu, 2 zpracování obrazu, 3. zpracování vektorů příznaků, 4. určení gesta

3.2 Záznam obrazu

První fází v procesu určování gest je záznam obrazu. Jejím cílem je pořídit sekvenci snímků, které budou dále zpracovávány.

Vstupním předpokladem této fáze je osoba, která před kamerou vykonává gesta. Rozlišení kamery a poměry stran obrazu jsou libovolné (obrázek 8), avšak pokud požadujeme, aby systém pracoval v reálném čase, musí být zajištěna základní podmínka, aby při zobrazování na daném počítači snímková frekvence neklesla pod hranici 15 snímků za sekundu. V tomto případě by rozpoznávání neprobíhalo s dostatečnou přesností.



obrázek 8: rozlišení kamery a poměr stran obrazu může být libovolné

Při volbě kamery je kladen hlavní důraz na kvalitu obrazu. Snímky by měly být ostré, bez šumu a vyvážení barev by mělo odpovídat snímané scéně. Jen tak lze zaručit bezproblémový průběh druhé fáze (kapitola 3.3).

Osoba může gesta vykonávat levou i pravou rukou, může stát blízko i daleko od kamery a gesta může vykonávat v libovolné pozici vzhledem k hlavě a libovolně rychle. Tyto rozdíly řeší další fáze procesu. Taktéž lze zvolit různé výřezy snímané scény. Podmínkou je pouze to, aby na každém snímku sekvence byly zachyceny důležité části těla celé. Všechny rozdíly zmíněné v tomto odstavci ilustruje obrázek 9.

obrázek 9: rozdílné zachycení scény

Pokud sekvence snímků nesplňuje výše uvedené požadavky, na samotný záznam to pochopitelně nemá vliv, ovšem tato skutečnost má dopad na následující fáze procesu. Podle závažnosti komplikuje nebo naprosto znemožňuje další průběh.

3.3 Zpracování obrazu

V pořadí druhou fází v procesu rozpoznávání gest je zpracování obrazu a generování vektorů příznaků. Tato fáze se spoléhá především na dobrý a kvalitní obraz, jehož vlastnosti byly popsány v kapitole 3.2. Vstupem do této fáze je sekvence snímků, která je postupně zpracovávána.

V každém snímku je nejdříve určena pozice a rozměry hlavy a ruky. Jelikož detekce hlavy v obraze je výpočetně náročný úkol, je nutné zavést účinné postupy, které celý proces urychlí. Jedná se zejména o omezení oblasti, ve které je v následujícím snímku obličej vyhledáván a omezení minimální velikosti obličeje. Takto velmi snadno docílíme zpracování obrazu v reálném čase.

Z těchto údajů se následně vytvoří normalizovaný vektor příznaků (kapitola 2.3). Ten charakterizuje stav snímku, ke kterému je přiřazen. Pokud nelze ve snímku detekovat obě části těla, nelze tudíž provést ani normalizaci, snímku je přiřazen *prázdný vektor příznaků*.

obrázek 10: jednotlivé kroky fáze zpracování obrazu

Každý vygenerovaný vektor příznaků, ať již normalizovaný či prázdný, je předán do další fáze pro následné zpracování.

3.4 Zpracování vektorů příznaků

Tato fáze tvoří jádro celého systému a bude proto popsána detailněji. Bude zmíněna její hlavní funkce, popsány jednotlivé části a veškeré situace, které mohou nastat.

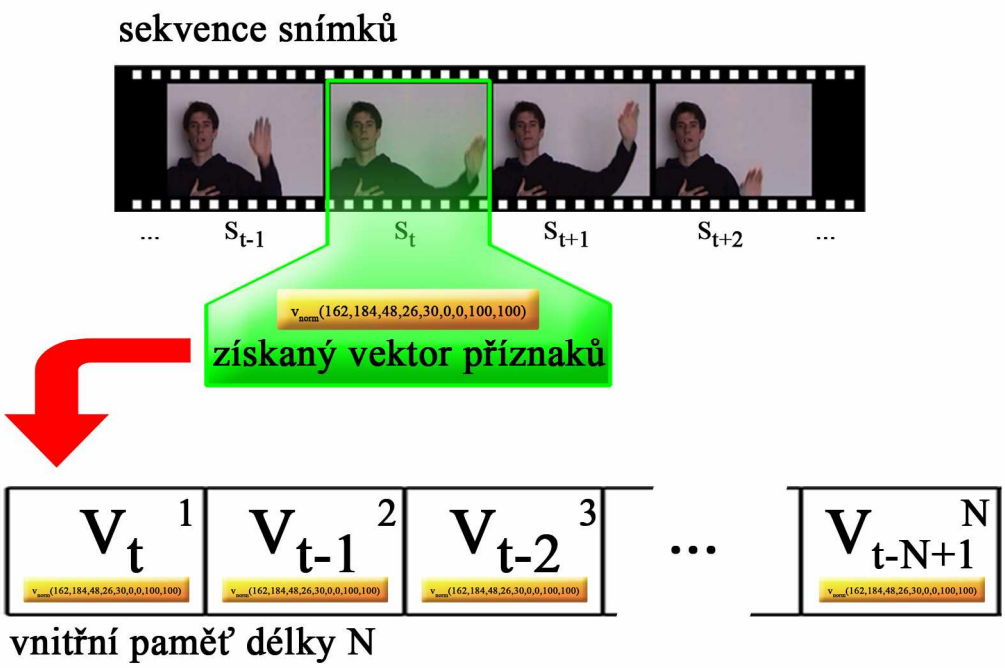
Vstupem do ní jsou vektory příznaků přiřazené jednotlivým snímkům. Za normálních okolností dochází k průběžnému ukládání těchto vektorů do *vnitřní paměti* (obrázek 11), která se za určitých okolností vyhodnocuje. Paměť má obecnou délku *N*. Tento parametr popisuje maximální počet snímků, které systém uchovává pro zpětné vyhodnocení. Parametr *N* lze průběžně měnit a tak zvyšovat či snižovat rozsah detekce a zároveň naopak snižovat nebo zvyšovat výpočetní nároky. Na základě analýzy průměrné délky gest lze stanovit minimální a maximální hodnotu *N*, což může v praxi přinést zvýšení výkonu a přesnosti

$$N \in \langle N_{MIN}, N_{MAX} \rangle . \tag{3.1}$$

Pro praktické účely lze stanovit parametry N_{MIN}, N_{MAX} takto

$$N_{MIN} = 10, N_{MAX} = 100 . \tag{3.2}$$

Při snímkové frekvenci 25 snímků za sekundu lze pak rozpoznávat gesta v rozmezí 0,4 – 4 sekundy.



obrázek 11: uspořádání vektorů příznaků ve vnitřní paměti vzhledem k sekvenci snímků

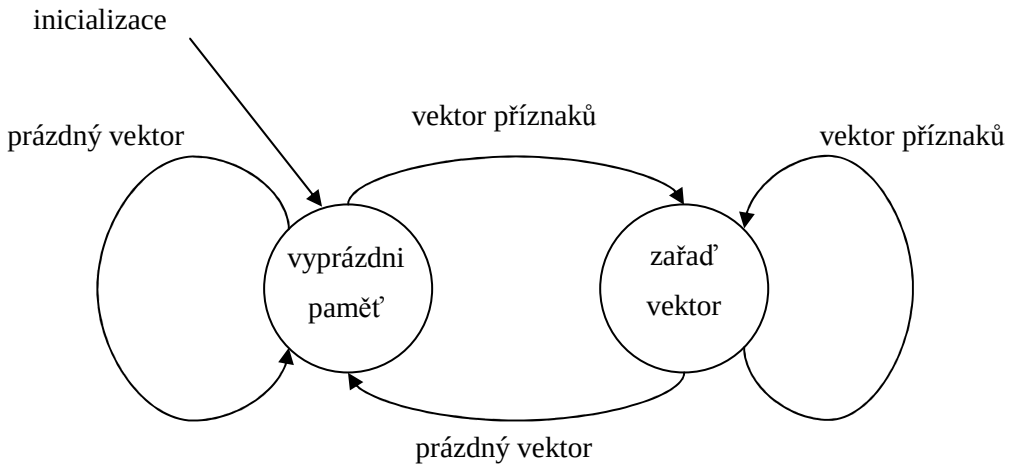
Organizace *vnitřní paměti* je podobná frontě FIFO (first in, first out). Každý nový vektor příznaků vyvolá posun starších položek o jednu pozici dále a zařadí se do fronty na pozici číslo 1.

Na vrcholu je tedy vždy poslední získaný vektor příznaků. Nejstarší záznamy (na poslední pozici *N*) z paměti vypadávají a dále se pro vyhodnocování již nepoužívají.

V daném okamžiku mohou nastat 2 situace. Vnitřní paměť může být buď prázdná nebo může obsahovat 1 – *N* vektorů příznaků.

První situace nastane po samotném spuštění detekce před okamžikem příchodu prvního vektoru příznaků nebo po příchodu prázdného vektoru příznaků. Ten je vygenerován v procesu zpracování obrazu pokud nastane chyba v detekci (viz. kapitola 3.3).

Druhá situace odpovídá běžnému stavu. Postupně přicházející vektory jsou řazeny do paměti a starší záznamy odsouvány, popřípadě z paměti vyřazovány.



obrázek 12: stavový diagram charakterizující práci s vnitřní pamětí, hrany značí příchod vektorů příznaků a uzly reakce na tyto události

Nyní se zaměříme na proces vyhodnocování obsahu vnitřní paměti. Děje se tak vždy po příchodu nového vektoru příznaků a pokud vnitřní paměť obsahuje alespoň N_{MIN} položek. Uplatňuje se zde právě algoritmus DTW.

Nezbytnou součástí je tzv. „slovník gest“. Jedná se o soubor několika předem natrénovaných referenčních obrazů gest. Ta jsou při každém vyhodnocení porovnána s testovaným obrazem gesta a algoritmus DTW vyhodnotí jejich vzdálenost.

Kdy však dochází k vyhodnocování?

Algoritmus DTW je primárně určen pro rozpoznávání *izolovaných* slov, v našem případě gest. Musíme tedy zavést účinnou metodu, která mezi *N* vektory příznaků uložených ve vnitřní paměti nalezne a izoluje gesto od ostatních pohybů ruky.

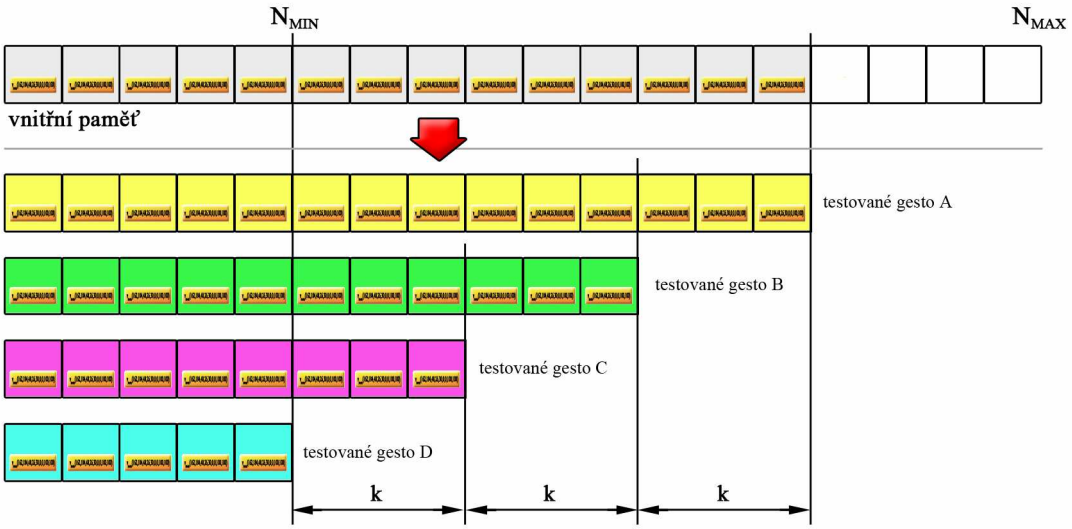
Jedna z možností, jak izolovat gesto, je ohraničit jeho začátek a konec klidovým stavem ruky, využít tedy derivace její polohy. V okamžiku kdy se ruka začne pohybovat zaznamenáme začátek gesta a postupně ukládáme jednotlivé vektory příznaků. Až se ruka zastaví, dojde k ukončení gesta,

proběhne analýza získaných dat a proces se může opakovat. Tato metoda sice šetří výpočetní výkon, avšak neumožňuje detekovat například několik gest na sebe navazujících.

Proto byl zvolen jiný přístup. V každém okamžiku s jistotou víme, že se na první pozici vnitřní paměti nachází vektor příznaků, který nese informaci o poslední poloze ruky v obraze, která může odpovídat konci právě provedeného gesta. V takovém případě je nutné určit jeho začátek.

Začátek gesta může být na poslední pozici ve vnitřní paměti, pak jako první testované gesto zvolíme celý obsah vnitřní paměti. Pokud tomu tak není, neznáme kde přesně gesto ve vnitřní paměti začíná, je nutné paměť rozdělit s krokem *k* na několik částí. Od celkového počtu vektorů v paměti se postupně odečítají násobky *k* a vždy s každým odečtením vznikne nové testované gesto. Tento proces se opakuje, dokud délka nově vzniklého testovaného gesta neklesne pod minimum dané parametrem N_{MIN} (obrázek 13).

Výhodou tohoto přístupu je fakt, že lze detekovat vzájemně na sebe navazující gesta. Dále lze průběžně měnit parametr *k* a zvyšovat či snižovat přesnost detekce v závislosti na výpočetním výkonu.



obrázek 13: příklad rozdělení vnitřní paměti na 4 testovaná gesta A, B, C, D

Pokud tedy shrneme předcházející odstavce. Každý nový příchozí vektor příznaků je zařazen do vnitřní paměti. Ta je následně rozdělena na několik částí, kde každá odpovídá jednomu testovanému obrazu gesta. Následně algoritmus DTW vyhodnotí pro každou dvojici *testovaný obraz* <=> *referenční obraz* jejich vzájemnou míru podobnosti (vzdálenost) a tyto výsledky předává do následující fáze „určení gest“ (kapitola 3.5) kde jsou zpracovány.

3.5 Určení gesta

Toto je poslední fáze celého procesu. Jejím cílem je na základě posledních několika výsledků zpracovaní příznaků určit konečné gesto.

Není problém v daném čase *t* vyhodnotit zpětně pro několik snímků nejpravděpodobnější gesto včetně jeho vzdálenosti ke svému vzorovému obrazu. Pro každý následující snímek však toto gesto může být obecně různé. Pokud však dojde k situaci, kdy v průběhu několika snímků po sobě dojde k vyhodnocení stejného gesta jako nejpravděpodobnějšího, lze s jistotou prohlásit, že dané gesto bylo vykonáno.

Celý proces lze vylepšit zavedením tzv. „tříd gest“, kdy do příslušné třídy zařadíme několik referenčních obrazů stejného gesta. Proces určení gesta je pak obdobný. Tato metoda se v praxi ukázala jako velmi účinná a zvyšuje robustnost celého rozpoznávacího procesu.

4 Testování a vyhodnocení

Tato kapitola obsahuje detailní testování a zhodnocení navrženého systému. V úvodu jsou stručně popsány dvě hlavní etapy, kterými aplikace během svého vývoje procházela. Dále následují testy, které se snaží podat ucelenou zprávu o kvalitách a nedostatcích jednotlivých vlastností řešení. Vždy jsou popsány výchozí podmínky testů, jejich účel a forma. Následuje zhodnocení výsledků a závěr. Na konci této kapitoly jsou jednotlivé dílčí výsledky přehledně shrnuty.

Navržený systém prošel během svého vývoje několika zásadními změnami. Některé z nich měly za účel zvýšit výkonnost a přesnost rozpoznávání gest, jiné si kladly za cíl především rozšířit slovník a tak zvýšit rozmanitost odhalitelných gest.

V první fázi systém pracoval s algoritmem DTW a slovníkem, který obsahoval pouze 9 gest. Tomuto řešení bude věnován druhý test. Jeho cílem je popsat některé nedostatky a důvody pro přechod na algoritmus DDTW. Ten posunul aplikaci do druhé, finální fáze vývoje.

Testování probíhalo na Notebooku Acer TravelMate 4500 s procesorem Intel Centrino 1,5 GHz, 2 MB L2 cache, operační pamětí 512 MB, pevným diskem o kapacitě 40 GB, grafickou kartou ATI Mobility Radeon 9700 a operačním systémem Microsoft Windows XP Professional SP2. Dále pro záznam obrazu byla použita kamera Sony DCR-HC39E v režimu PAL a ponechány veškeré její parametry v režimu „auto“. Obraz byl živě přenášen do počítače přes USB 2.0 v rozlišení 320x240 bodů a snímkové frekvenci odpovídající zhruba třiceti snímkům za sekundu.

Osoby, které se testu účastnily, byly vybrány náhodně. O povaze a vlastnostech systému neměli dopředu žádné informace. Obdrželi pouze papír s gesty, které měli vykonat. Testy se soustředily na určení gest výhradně v reálném čase.

4.1 Úspěšnost detekce částí těla

Druhou fází v procesu určování gest je detekce částí těla. Na úspěšnosti a přesnosti této fáze závisí další části systému. Tento test se snaží krátce popsat vlastnosti a chování navrženého řešení v nejrůznějších modelových situacích.

Testování bylo prováděno za různé denní doby a různého osvětlení. Snahou bylo také vyzkoušet úspěšnost detekce na co největším počtu osob, které měli i různé oblečení a účesy.

Výsledky lze shrnout následujícím způsobem. Detekce ruky je při správném nastavení kamery, vyvážení bílé barvy, expozice a správném nastavení parametrů modelu pro detekci ruky velmi spolehlivá a navržené řešení poskytuje přesné výsledky ve velmi krátkém čase. Nedostatky se objevují pouze při nízké hladině osvětlení nebo pokud má osoba oblečení v barvách pleti.

Detekce obličeje funguje velmi dobře i při nízkém osvětlení, chybném vyvážení bílé barvy, nezávisle na věku, pohlaví nebo barvě účesu sledované osoby. Metoda je však výpočetně velmi náročná. Úspěšnost detekce obličeje klesá, pokud má osoba hlavu natočenou nebo jakkoliv nakloněnou od přímého směru ke kameře.

Z výše uvedeného lze učinit tento závěr: při zajištění dostatečného osvětlení, správném nastavení parametrů modelu pro detekci ruky, lze velmi přesně a spolehlivě sledovat jednotlivé části těla. Detekce obličeje je však časově náročná.

4.2 Algoritmus DTW

Cílem tohoto testu je zhodnotit základní vlastnosti čisté podoby algoritmu DTW, který tvoří základ aplikace. V tomto testu slovník gest obsahuje 9 tříd, v každé 5 variant téhož gesta. Algoritmus pracuje s normalizovanými vektory příznaků. Výsledky by měly odhalit především závislost přesnosti detekce na výchozí poloze a pohybu ruky.

Jelikož slovník byl natrénován na rozpoznávání gest, která jsou provedena pouze levou rukou, musela být daná osoba s tímto faktem dopředu seznámena. V závěru k této skutečnosti bude přihlédnuto.

Každé z 9 natrénovaných gest bylo v testu provedeno 20-krát. Vždy odlišnou rychlostí, s různou počáteční polohou ruky a měnila se také plocha, kterou gesto v prostoru zabíralo. Snahou bylo dosáhnout co největší rozmanitosti gest, při zachování rozumné míry jejich vzájemné odlišnosti. Test měl simulovat reálnou situaci.

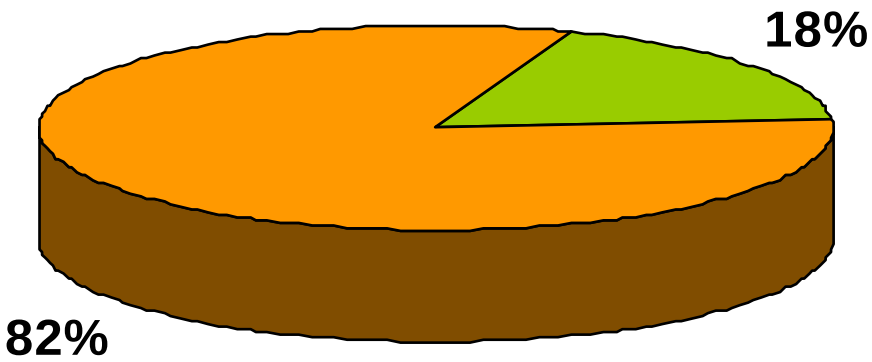
Parametry testu byly zvoleny následujícím způsobem: maximální délka gesta 60 snímků, přesnost detekce 4 (odpovídá kroku rozdělení vnitřní paměti $k = 8$ – viz. obrázek 13).

Následující tabulka shrnuje výsledky testu. V jejím levém sloupci jsou názvy jednotlivých gest a na příslušném řádku 1 - v případě úspěšného rozpoznání nebo 0 - při neúspěchu. Pravý sloupec obsahuje úspěšnost v procentech.

gesto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%
New	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10%**
Rename	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0%**
Zpet	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	**10%**
Vyber	0	0	1	1	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	**30%**
Delta	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0	1	**55%**
Plus	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	**15%**
About	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0%**
Reload	1	1	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	**30%**
Xclose	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	**15%**

tabulka 1: hodnocení algoritmu DTW

Celková úspěšnost rozpoznání gest činí pouze ***18,33%***.



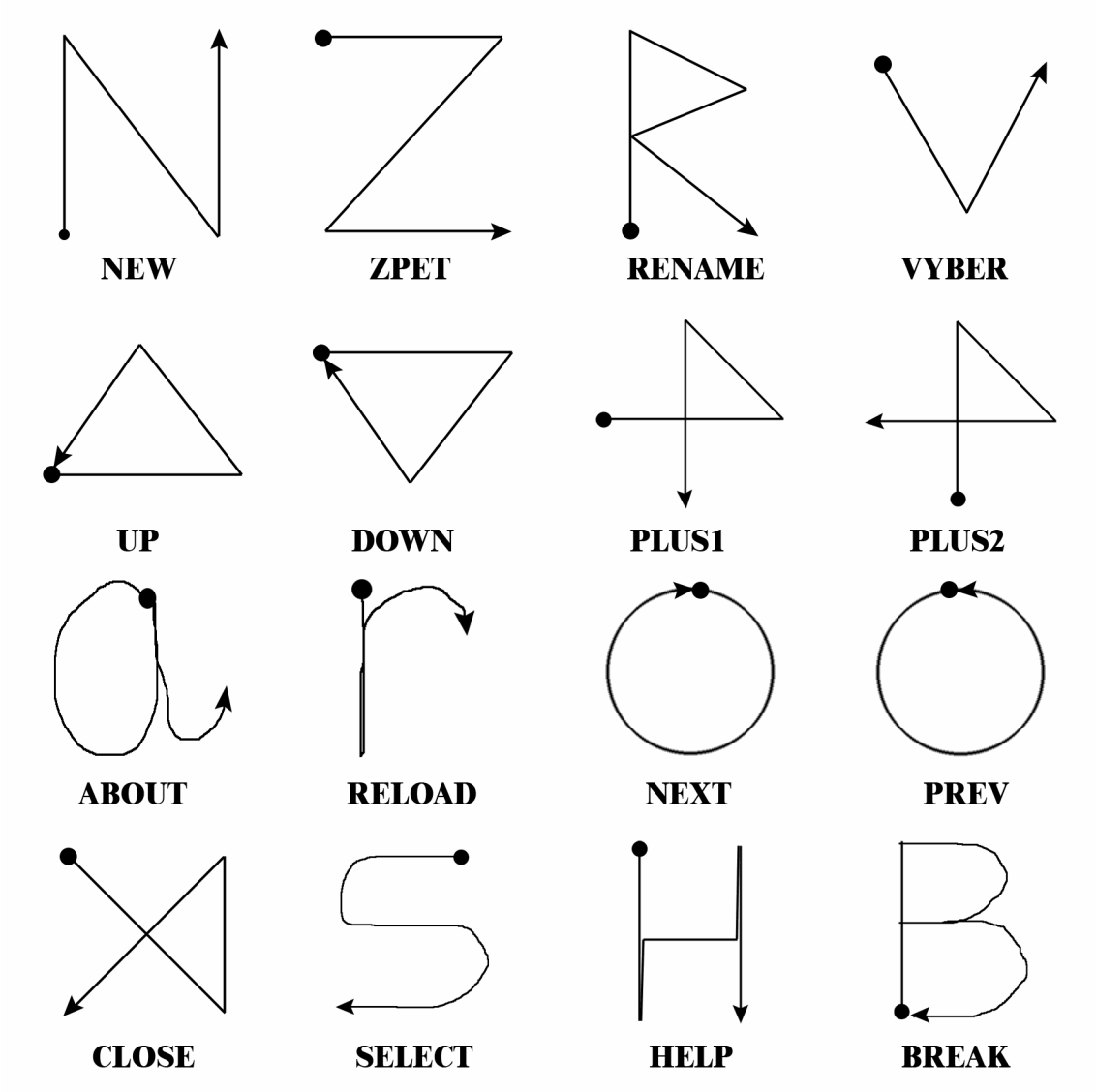
graf 1: celková úspěšnost detekce gest pomocí algoritmu DTW,
zelená oblast – úspěšně rozpoznaná gesta, oranžová oblast – nerozpoznaná gesta

Závěr, který plyne z tohoto testu: úspěšnost rozpoznání gest při použití algoritmu DTW silně závisí na kvalitě slovníku. Gesta musejí být prováděna přesně podle jejich vzoru, což komplikuje rozpoznávání u většiny osob. Tento nedostatek lze odstranit zvětšením počtu obrazů gest v jednotlivých třídách, tím však rostou výpočetní nároky. Na druhou stranu k úspěšnému rozpoznání gesta dochází s velkou přesností a nezávisle na rychlosti jeho provedení.

4.3 Algoritmus DDTW

Následující testy se věnují hodnocení algoritmu DDTW, který se od varianty DTW liší pouze v tom, že počítá s prvními derivacemi vektorů příznaků v čase. Toto řešení s sebou přináší mnoho výhod.

Nejdříve bude vhodné uvést podobu slovníku gest, který bude použit v následujícím testu. Ten obsahuje 16 gest, každé 3-krát natrénované, dohromady tedy 48 referenčních obrazů. Jejich vzhled je na obrázku 14.



obrázek 14: slovník gest pro testování algoritmu DDTW, 16 gest se svými názvy

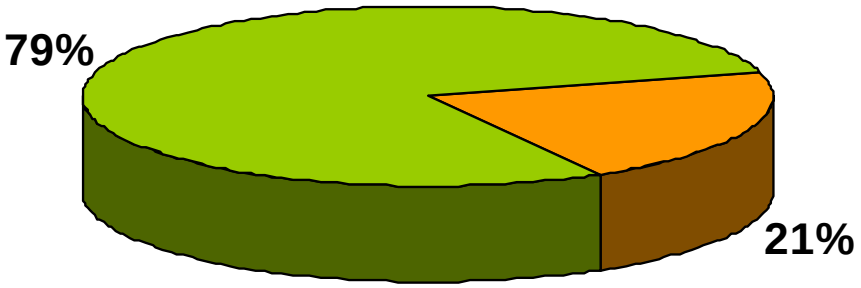
Test 1 – celková úspěšnost určování gest

Prvním kritériem hodnocení navrženého řešení je celková úspěšnost rozpoznávání gest. Cílem je podat obecné charakteristiky navrženého řešení.

Testu se účastnily 4 osoby. Každá z nich vykonala všechna gesta ze slovníku 5-krát, celkem tedy 4x16x5 = 320 gest v libovolném provedení bez předchozích instrukcí. Mezi jednotlivými gesty byla vždy 1 sekundu pauza.

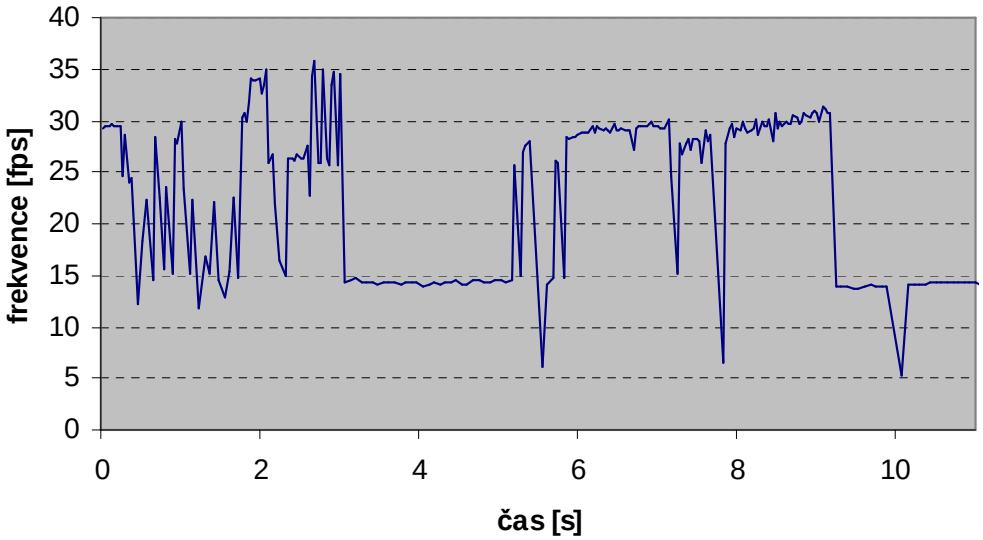
Parametry testu byly zvoleny následujícím způsobem: maximální délka gesta 60 snímků, přesnost detekce 4 (odpovídá kroku rozdělení vnitřní paměti k = 8 – viz. obrázek 13).

Přesnost detekce byla u všech 4 osob velmi podobná. Nezáviselo na tom, zda gesta vykonávají levou či pravou rukou ani na její počáteční poloze. Celková úspěšnost určení gest činí ***78,79%***.



graf 2: celková úspěšnost detekce gest pomocí algoritmu DDTW,
zelená oblast – úspěšně rozpoznaná gesta, oranžová oblast – nerozpoznaná gesta

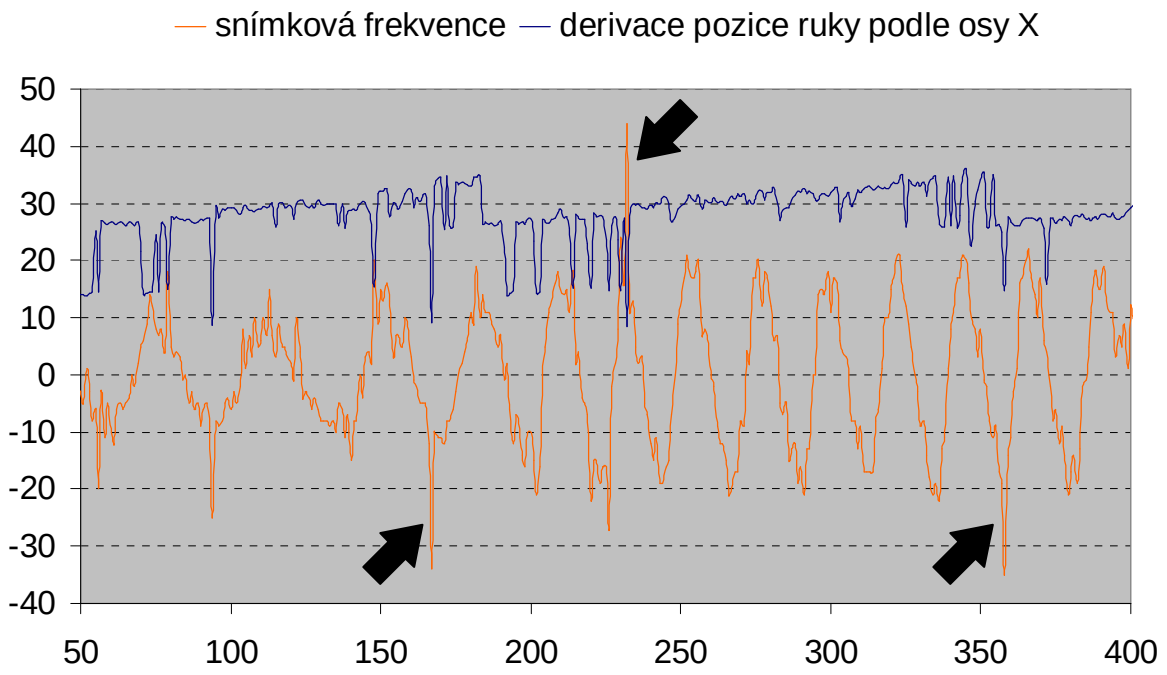
Během testu byl odhalen jeden závažný nedostatek, který se netýká algoritmu DDTW. Problém se týká samotného rozhraní a jeho snímkové frekvence, která se s časem mění (graf 3). V některých okamžicích se dokonce objevují zásadní poklesy výkonu, které mají za následek nerozpoznání gesta.



graf 3: kolísající snímková frekvence v čase

Důvod je jednoduchý. Algoritmus DDTW počítá s derivacemi polohy ruky vůči hlavě. Pokud budeme uvažovat konstantní rychlost vykonávání gest, pak při poklesu snímkové frekvence dojde ke zvětšení derivací. První derivace polohy ruky vyjadřuje její rychlost a při snížení snímkové frekvence dochází k zvětšování hodnot derivací. Situaci znázorňuje graf 4. Šipkou jsou označena kritická místa, ve kterých je rozdíl derivací zvlášť výrazný.

Tento nedostatek lze odstranit pouze použitím záznamu s konstantní snímkovou frekvencí. V případě zpracování obrazu v reálném čase se situace komplikuje.



graf 4: závislost snímkové frekvence a derivace pozice ruky, osa x – číslo snímku, oranžový graf – derivace pozice ruky podle x (v ideálním případe sinusoida), modrý graf – snímková frekvence [fps], zvlášť vyznačená kritická místa, kde lze pozorovat výrazné rozdíly v derivacích při poklesu snímkové frekvence

Závěr testu číslo 1: algoritmus DDTW umožňuje vykonávání natrénovaných gest v libovolné pozici vzhledem k hlavě, což rozšiřuje jeho použití. Porovnává však rychlost pohybu ruky (první derivace) v čase, což znesnadňuje rozpoznávání.

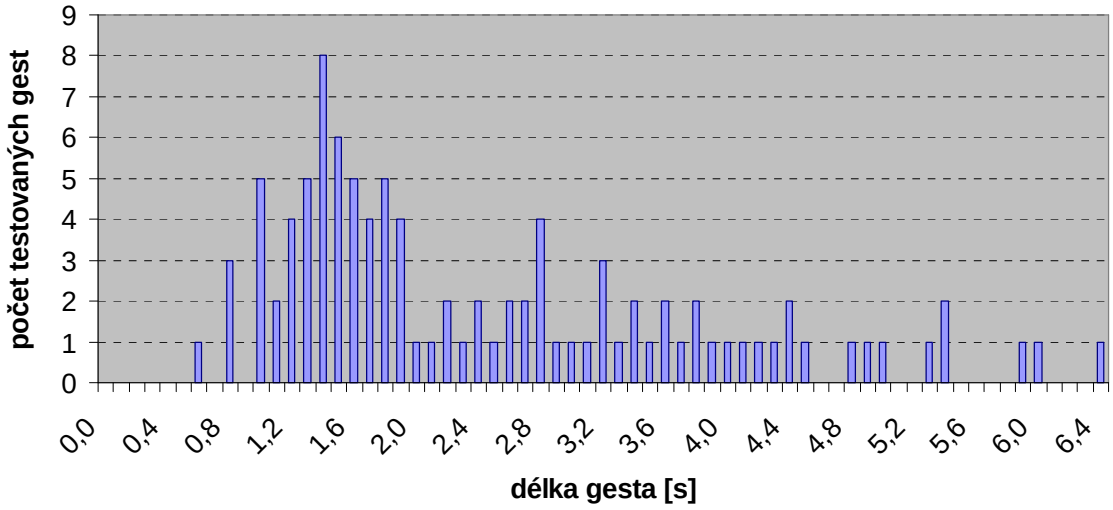
Test 2 - vliv rychlosti provedení gesta

Z předchozího testu vyplynul jeden důležitý poznatek. Algoritmus DDTW porovnává rychlost pohybu ryky v horizontálním a vertikálním směru.

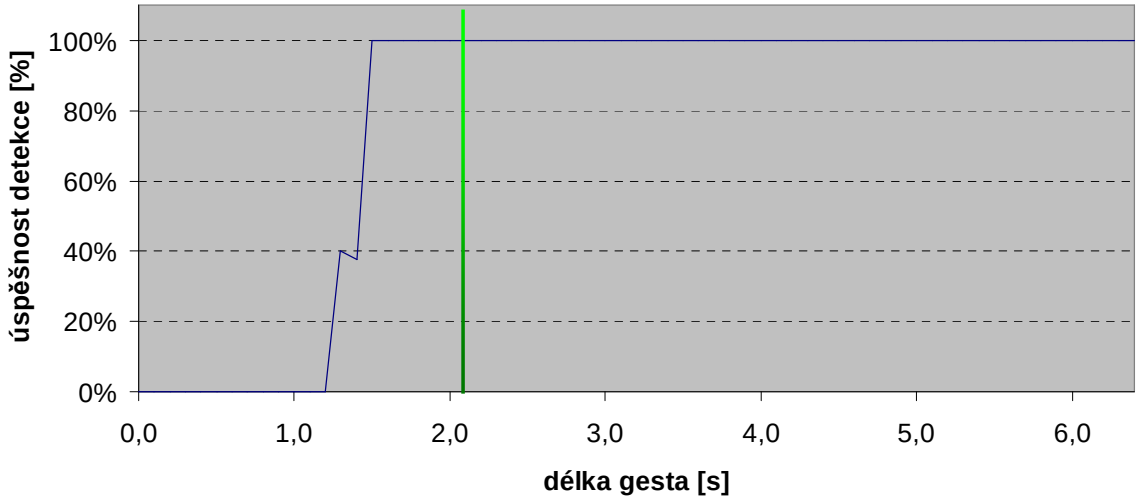
Následující test si klade za cíl odhalit úspěšnost detekce gesta v závislosti na rychlosti jeho provedení. Pro tento účel bylo do slovníku natrénováno 8 referenčních obrazů gesta „Select“ o shodné délce 2,12 sekundy, tj. 53 vektorů pro každý referenční obraz. Testu se účastnila 1 osoba. Byla požádána, aby gesto „Select“ zopakovala několikrát po sobě s různou rychlostí provedení a s krátkými pauzami mezi jednotlivými pokusy. Pro zajištění konstantní snímkové frekvence byla natáčena na kameru v rozlišení PAL 720x576 a snímkové frekvenci 25 snímků za sekundu. Záznam byl následně zpracován a gesta vyhodnocena.

Parametry testu byly zvoleny následujícím způsobem: maximální délka gesta 100 snímků, přesnost detekce 5 (odpovídá kroku rozdělení vnitřní paměti $k = 4$ – viz. obrázek 13).

Celkem bylo vykonáno 100 gest v délkách od 0,6 – 6,4 s. Rozložení délek gest zachycuje graf 5. Úspěšnost rozpoznání gesta v závislosti na jeho délce je promítnuta do grafu 6.



graf 5: histogram počtu testovaných gest vzhledem k příslušné délce gesta



graf 6: úspěšnost rozpoznání gesta v závislosti na jeho délce,
v grafu je zeleně vyznačena délka referenčních obrazů gesta 2,12 s

Z grafu 6 lze učinit dva důležité závěry:

- algoritmus DDTW není vhodný pro detekci velmi rychlých gest
- algoritmus DDTW je schopen detekovat velmi pomalá gesta, jejichž délka přesahuje až trojnásobek délky referenčního gesta, s velmi vysokou přesností

5 Závěr

Cílem této bakalářské práce bylo navrhnout, implementovat a otestovat systém pro rozpoznávání jednoduchých gest ruky. Všechny body se podařilo splnit.

Z výsledků testování plyne důležitý závěr. Algoritmus DDTW lze velmi dobře aplikovat na problém rozpoznávání gest ruky v reálném čase. Důležitou roli hraje zajištění konstantní snímkové frekvence během celého rozpoznávání. Předností algoritmu je snadné natrénování slovníku, který může obsahovat až desítky gest. Ve srovnání s algoritmem DTW nabízí větší variabilitu co do možnosti jejich provedení. Detekci však omezuje na středně rychlá až pomalá gesta. Rychlá gesta lze touto metodou detekovat pouze v případě, kdy do slovníku přímo taková gesta uložíme.

Navržený systém umožňuje širokou škálu použití. Je velmi variabilní a univerzální, nezávislý na rozlišení obrazu a velmi tolerantní k jeho kvalitě. Sledovaná osoba může být umístěna jak blízko tak daleko od kamery. Gesta může provádět v libovolné pozici vzhledem k hlavě, levou i pravou rukou. Jediným požadavkem pro bezproblémové a přesné určení gest je zachování konstantní snímkové frekvence a použití slovníku natrénovaného právě pro danou frekvenci. V opačném případě nemusí k rozpoznání gesta dojít.

Během testování a práce se systémem byl učiněn jeden důležitý poznatek. Přesnost detekce závisí především na kvalitě natrénovaného slovníku gest. Úspěšnost detekce se zvyšuje se zvyšujícím se počtem referenčních obrazů daného gesta. Ideální jsou 3 obrazy. Při větším počtu se již přesnost rozpoznání nezvyšuje, daleko více stoupají výpočetní nároky. V pozdní fázi programu byla také objevena metoda, která umožňuje natrénování velmi kvalitního referenčních slovníku [9]. Její princip je založen na porovnání a zprůměrňování několika obrazů téhož gesta do jednoho výsledného referenčního obrazu. Její vliv však již nebylo možné posoudit. Toto řešení však lze uvažovat jako jednu z variant dalšího stupně vývoje aplikace.

Možnosti nasazení tohoto systému jsou velmi široké. Od jednoduchých počítačových her, přes virtuální realitu až po možnost rozpoznávání znakové řeči. Jednou z variant jeho použití jsou informační systémy ovládané gesty ruky. Například interaktivní plány velkých budov, promítané na obrazovku, před kterou bude stát uživatel. Gestem bude volit například patra, přibližovat, oddalovat a posouvat celou mapu komplexu. Takovéto systémy mohou být výhodné především z pohledu bezpečnosti. Nebude docházet ke styku uživatele s jednotlivými hardwarovými částmi systému a jejich odcizení tak bude značně komplikovanější. Informace se budou přenášet pouze vizuálně pomocí kamery a obrazovky.

Literatura

[1] Vezhnevets V., Sazonov V., Andreeva A. A Survey on Pixel-Based Skin Color Detection Techniques. Proc. Graphicon-2003, pp. 85-92, Moscow, Russia, September 2003, Dostupný také z WWW: <http://graphics.cmc.msu.ru/en/publications/index.html>.

[2] BRADSKI, Gary R. Computer Vision Face Tracking For Use in a Perceptual User. Intel Technology Journal [online]. Q2-1998 [cit. 2007-04-13]. Dostupný z WWW: <http://developer.intel.com/technology/itj/q21998/pdf/camshift.pdf>.

[3] Smith, A. R. Color Gamut Transform Pairs. SIGGRAPH 78, pp. 12-19, 1978.

[4] Foley, J.D.; A. van Dam; Feiner, S. K.; Hughes, J.F. Computer graphics principles and practice. Addison-Wesley, pp. 590-591.

[5] Comaniciu, D.; Meer, P. Robust Analysis of Feature Spaces: Color Image Segmentation CVPR'97, pp. 750-755.

[6] Viola, P.; Jones, M. Robust Real-time Object Detection. July 13, 2001. Vancouver, Canada [cit. 2007-04-13]. Dostupný z WWW: < http://cs223b.stanford.edu/notes/viola01robust.pdf>.

[7] VINTSYUK, T K. Element-wise recognition of continuous speech composed of words from a specified dictionary. Kibernetika, 7, 1971, Mar.-Apr., s. 133-143.

[8] PSUTKA, J. *Komunikace s počítačem mluvenou řečí*, Praha, Academia, 1996, str. 127-137

[9] KEOGH, J. E.; PAZZANI, M. J. Derivative Dynamic Time Warping. in First SIAM International Conference on Data Mining, (Chicago, IL, 2001). Dostupný z WWW: <http://www.cs.rutgers.edu/~mlittman/courses/lightai03/DDTW-2001.pdf>

[10] Abdulla, W. H., Chow, D., Sin, G. Cross-words reference template for DTW based speech recognition systems, IEEE TENCON 2003, October 14-17, Bangalore, India. Dostupný z WWW: <http://www.ele.auckland.ac.nz/~wabd002/DTW_TENCON_final.pdf>

A. Obsah přiloženého CD

- Manuál k programu gestures.exe
- Zdrojové texty programu
- Knihovna OpenCv
- Slovníky gest
- Použitá literatura
- Videosekvence
- Text této bakalářské práce