oponentní posudek na dizertační práci

Automatické rozpoznání výrazu tváře s libovolným natočením v prostoru

Ing. Jana Trojanová

1 Význam práce

Rozpoznávání výrazu tváře z obrazu je jedním z uznávaných aplikačních problémů počítačového vidění. Související problém je detekce tváří v obrazu, která byla populární po roce 2001, kdy vzrostly požadavky na systémy vizuálního sledování. Detekce tváře se dnes považuje za výzkumně vyčerpané téma. Statické rozpoznávání výrazu tváře je téma, které se blíží saturaci a dynamické rozpoznávání výrazu (gesta) v reálném čase je aktuální výzkumné téma vysoce relevantní pro HCI systémy. Dizertantka přispěla k řešení problému statického rozpoznávání výrazu tváře. Dizertační práce se zabývá otázkou, jak velký pozitivní vliv na výsledky rozpoznávání výrazu tváře (kategorické emoce a standardní jednotky svalové aktivity) má korekce natočení tváře a normalizace vstupního obrazu do kanonického úhlu pohledu, přičemž potřebné zobrazení je realizováno promítnutím vstupního obrazu na generický fixní 3D model tváře a následnou reprojekcí do obrazové roviny kanonické kamery. Pro vyřešení tohoto výzkumného projektu sestavila dizertantka vlastní *facial expression recognition* (FER) systém, který registruje klíčové body vstupního obrazu s klíčovými body na referenčním 3D modelu, interpoluje mezilehlou informaci a potom registrovaný model promítne do kanonické obrazové roviny, čímž vznikne normalizovaný obraz. Následuje standardní popis obrazovými deskriptory a proces rozpoznávání. Výsledky na standardní databázi ukazují zhruba desetiprocentní zlepšení rozpoznávání jak kategorických emocí, tak jednotek svalové aktivity. Navržený systém je konceptuálně velmi jednoduchý a snadno realizovatelný.

Podle mého názoru je otázka vlivu 3D korekce na úspěšnost klasifikace statického obrazu spíše dílčí v době, kdy se komunita zabývá rozpoznáváním dynamických složek výrazu.

Z předloženého seznamu publikací dizertantky plyne, že výsledky předložené dizertační práce nebyly publikovány, takže nelze zjistit, jak byly přijaty specialisty v oboru. Nutnost výtek uvedených v následující recenzi implikuje, že práce patrně neprošla žádným nebo prošla jen nedostatečným recenzním řízením, protože obsahuje příliš mnoho nedostatků, které by takovým řízením neprošly.

2 Předložený dizertační spis

Spis o rozsahu 64 stran (bez příloh a seznamu bibliografických referencí) je napsán v českém jazyce. Zhruba polovina textu práce (kap. 1–3, 30 stran) se věnuje popisu stavu problematiky v oblasti vývoje systémů pro automatické rozpoznávání emoce z obrazu. Popis je dobře strukturovaný a přehledný, místy méně přesný, zejména co se týká matematických vzorců. Text je doprovázen

dostatečným množstvím ilustrací, které jsou bez výjimky převzaty z publikovaných článků cizích autorů.

Implementační část je popsána v kapitole 4.3 a výsledky testování vlivu korekce natočení jsou popsány v kap. 5, celkem na 28 stranách, včetně prezentace velmi podrobných experimentálních výsledků.

Spis je zakončen dvoustránkovým závěrem a doplněn 135 bibliografickými referencemi a třemi přílohami. Přílohy podávají přehled vybraných FER systémů, přehled databází pro FER a alternativní reprezentace výrazu tváře. Podle mého názoru mohla být většina obsahu příloh součástí úvodních kapitol, zejména část přílohy B diskutující požadavky na dobrá zdrojová data pro učení a hodnocení FER.

Detailní komentáře

1. Podle mého názoru popis klasifikátoru SVM na úrovni stručného seznámení s podstatou metody do dizertace nepatří (kap. 3.3.1).

2. Matematické vzorce nejsou jednotně číslovány a chybí v nich interpunkce v případech, kdy jsou součástí věty (např. str. 17, 18, 19 atd, atd).

3. Matematická notace není dostatečně vysvětlena. Například, co jsou f, F, $\overline{(\cdot)}$ ve vzorcích (3.7) a (3.8)? Co je l ve vzorci (3.13)? Neplatí $l = N$?

4. Matematická notace je místy nepřesná (například x, y má dvojí význam v jediném vzorci (3.5)).

5. Text je někdy poměrně nepřesný. Například v kapitole 5.3 se tvrdí, že problémem, kterému se práce věnuje, je rozhodnout, zda snímek obsahuje jednu nebo více AU jednotek a o dva odstavce dále se tvrdí, že cílem rozpoznání je zjištění, zda daná AU jednotka je pozorovatelná v aktuálním snímku či ne. To jsou ovšem naprosto odlišné úlohy. Podle všeho šlo o to, které z dané množiny AU jednotek jsou rozpoznány v obrazu.

6. V textu je velké množství gramatických chyb, zejména ve shodě rodu (například na str. 8, 23, 25, 27, 28, 29, 30). Množství překlepů není velké.

3 Implementace metody FER

Na výslednou výkonnost systému má nepochybně vliv kvalita realizace klíčové součásti navrženého systému, kterou je registrace obrazu a 3D modelu. Základem registrace je nalezení korespondencí mezi klíčovými body ve 2D snímku a 3D modelem. Na str. 41 se píše o „vytvoření jednorázové mapovací matice". V práci není popsán algoritmus. Jak tato matice vznikne?

K husté registraci obrazu a 3D modelu se používá velmi jednoduchá metoda. Je zvolená metoda v nějakém smyslu optimální? Jak se tato metoda vyrovná s chybnými korespondencemi? Není jasné, jaký je model ekvivalentní kamery, do které se výsledný obraz promítne, ani jak byl tento model vybrán.

Detailní komentáře a otázky

1. Za referenční model byl náhodně vybrán jeden model z databáze *Face Recognition Grand Challenge* (zřejmě muž středního až staršího věku). Proč nebyl vybrán střední model? Jak by se výsledky změnily při výběru jiného modelu z databáze? Jak velký vliv na výsledky má pohlaví referenčního 3D modelu?

2. Jakou metodou jsou nalezeny klíčové body (kap. 4.3)? V textu chybí bibliografická reference, případně popis metody, je-li to metoda vlastní. V kap. 5 se dočteme, že to byla vlastní implementace detektoru tváře podle autorů Viola & Jones, spolu s metodou AAM popsanou v kap. 3.1.3, kde ovšem chybí popis metody řešící optimalizační problém. Metoda tedy není reprodukovatelná. Pracuje se s 62 manuálně označenými klíčovými body, specifickými pro každého řečníka. To je výraznou nevýhodou pro praktické aplikace. Další nevýhodou je selhávání AAM modelu v 15% všech případů. Soudě podle textu práce, dizertantka se nepokusila tuto míru neúspěšnosti snížit. Je selhání kompenzováno hlasovacím mechanismem přes více snímků (5.1), jak se stručně konstatuje v kapitole 5.4? Jak efektivně je kompenzováno, když zřejmě jde o systematické selhávání, které vylučuje schopnost pozorovat významný temporální segment výrazu?

3. Jak je reprezentována topologie 3D sítě (str. 39)?

4. Proč je natočený 3D model na obrázku 4.8 (úhel otočení 90° ve srovnání s úhlem otočení −90°) tolik protažený? Není to artefakt metody, kterou se hloubková mapa otáčí?

5. Co se rozumí vektorovým dělením trojúhelníků? V práci chybí bibliografická reference. Proč při dělení vznikají duplikované body (jak plyne z popisu v obr 4.11)? Standardně se postupuje dělením hran triangulované sítě, při čemž žádné nadbytečné vrcholy sítě nevznikají.

6. Jaký je rozdíl mezi stěnou a trojúhelníkem sítě (str. 42)?

4 Experimentální ověření

Pro každou nově navrženou metodu je experimentální validace a ověření výkonnosti nutnou podmínkou toho, aby metoda mohla být přijata výzkumnou komunitou. V práci je použita metodika ze soutěže FERA 2011, což umožňuje nejen odpověď na výzkumnou otázku, která byla cílem dizertace, ale i srovnání s jinými implementacemi (před rokem 2011).

V textu chybí srovnání navržené metody se state of the art metodami anebo alespoň s metodami ze soutěže FERA a příslušná diskuse. Nejsem expert na rozpoznávání výrazu z obrazů, ale letmé srovnání s výsledky ve článku [115] ukazuje, že realizovaný systém zřejmě nedosahuje výkonnosti systémů známých před rokem 2011.

Hlavním výsledkem dizertačním práce je experimentální ověření hypotézy, že kompenzace vlivu natočení hlavy má pozitivní vliv na výsledek rozpoznání emoce a jednotek svalové aktivity. Bylo pozorováno asi 10% zlepšení v implementaci, která se blíží základnímu systému v evaluační studii [115]. Protože ale výkonnost realizovaného systému nedosahuje výkonnosti publikovaných metod, nebylo prokázáno, že takové zlepšení by bylo dosaženo i u nejlepších známých metod. Diskuse k tomu v práci chybí.

Z názvu práce plyne, že studovaná metoda bude pracovat s „libovolným natočením tváře v prostoru". Experiment ovšem pracuje s natočením v rozsahu ±30°, navíc je 15% snímků vyřazeno pro selhání registrace AAM modelu. Poskytuje tedy práce výsledek slíbený v názvu?

Podle mého názoru je testovací soubor je příliš malý na to, aby experimentální rozdíly mezi metodami byly dostatečně průkazné ve statistickém smyslu.

Detailní komentáře a otázky

1. Není jasné, jaký vliv na výsledek klasifikace má přesnost určení klíčových bodů pro AAM model.

2. Není jasné, jaký vliv mají na výsledek kalibrační parametry kamery pořizující vstupní snímky a simulované kanonické kamery, případně jejich nesoulad, zejména ohnisková vzdálenost a radiální zkreslení.

3. Není jasné, jestli majoritní hlasování (vzorec (5.1)) je vhodná metoda jak určit hlavní výraz (kategorickou emoci). Jak by se takovýto postup realizoval v kontinuálním zpracování videa, kdy není předem zřejmé, v jakém časovém okně se má hlasování provést?

4. Výsledky pro *NS* (neznámý subjekt) a NS_{3D} (neznámý subjekt s použitím navrhované metody korekce) v tab. 5.2 se významně liší od výsledků uvedených v matici záměn v tab. 5.4 (u emocí *vztek*, *radost*, *úleva*). Které výsledky platí? Jak jsou touto chybou ovlivněny závěry předložené práce?

5. Konstatuje se, že u emoce *smutek* 3D rekonstrukce zhoršuje výsledky na polovinu. Ale z tab. 5.4 je rozdíl 5 versus 2 správně rozpoznané emoce. Je takovýto výsledek statisticky významný? Proč nebyl proveden pokus o získání či pořízení databáze, na které by bylo možné dospět ke statisticky významným výsledkům?

5 Shrnutí

Pokud se omezíme na původní výsledky obsažené v dizertačním spisu, má tento zhruba rozsah jednoho monotematického konferenčního článku, nikoliv doktorandského projektu. S výhradami uvedenými výše by pouze na základě předloženého spisu nebylo možné doporučit obhajobu. S přihlédnutím k tomu, že dizertantka přispěla do oboru dalšími publikovanými pracemi, která s tématem dizertační práce blízce souvisí, a že tyto práce byly citovány, tedy výzkumníky v oboru přijaty, obhajobu doporučuji, s tím, že při obhajobě dizertantka také shrne všechny své příspěvky a ukáže, jaký měly vliv na obor.

V Praze, 27. prosince 2013

doc. Dr. Techn. Ing. Radim Šára

katedra kybernetiky
fakulta elektrotechnická
České vysoké učení technické v Praze
Technická 2, 166 27 Praha 6

doc. Ing. Josef Chaloupka, Ph.D., Technická univerzita v Liberci, Fakulta mechatroniky, informatiky a mezioborových studií, Ústav informačních technologií a informatiky, Studentská 2, Liberec, 46117

Oponentský posudek disertační práce

Autorka práce: **Ing. Jana Trojanová**

Název práce: **Automatické rozpoznání výrazu tváře s libovolným natočením v prostoru**

Aktuálnost zvoleného tématu

Předložená disertační práce se zabývá problematikou 3D rekonstrukce objektu z 2D snímku pro rozpoznávání výrazu tváře v podmínkách, kdy je snímaný objekt (lidská hlava) libovolně natočen vzhledem k snímací technice. Jedná se o aktuální problematiku, která je v současné době předmětem zkoumání mnoha předních vědecko-výzkumných pracovišť.

Cíle disertační práce

Doktorandka si za svůj hlavní cíl disertační práce vytyčila navrhnout, vytvořit a funkčně otestovat metodu pro 3D rekonstrukci tváře z 2D snímku. Tato metoda měla být především vhodná pro rozpoznávání výrazu tváře (emocí). Druhotným cílem pak bylo definovat faktory nutné k rozpoznávání výrazu tváře pomocí výpočetní techniky. Navržené cíle považuji za disertabilní.

Metody zpracování a přínosy

Disertační práce byla dokončena v roce 2013, obsahuje 102 stran a je psána českým jazykem. Práce je rozdělená do 6 kapitol a 3 příloh. Po krátké úvodní kapitole se doktorandka věnuje teoretické části. Zde jsou osvětleny vztahy mezi emocí a výrazem tváře a jsou zde i představeny některé realizované FER systémy.

V kapitole 3 jsou relativně přehledně popsány metody a algoritmy pro rozpoznávání výrazu tváře, které se v současné době používají. Tato kapitola je podepřena velkou řadou relevantních citačních odkazů. Lze tak konstatovat, že se doktorandka dostatečně seznámila se současným poznáním zkoumání v oblasti rozpoznávání výrazu tváře a že se v této oblasti slušně orientuje.

Kapitola 4 se jmenuje trochu zvláštně „Natočení tváře na čelní pohled", nicméně je zde uvedeno řešení problematiky rozpoznávání výrazu tváře, pokud není obličej snímán z čelního pohledu a jsou zde také uvedeny principy 3D rekonstrukce tváře. V části 4.3 pak doktorandka uvádí navrženou a realizovanou metodologii 3D rekonstrukce tváře vhodnou pro rozpoznávání výrazu. Navržená metodologie se zdá být funkční, což potvrzuje otestování této metodologie v kapitole 5.

Kladně hodnotím, že v rámci testování byla použita databáze FERA 2011, na které již byly testovány jiné systémy pro rozpoznávání výrazu tváře. Lze si tak snadno udělat představu o úspěšnosti využití doktorandkou navržené metody. V této kapitole jsou uvedeny výhody i jisté nevýhody normalizace natočení obličeje, kdy dochází k určitému zkreslení normalizovaného snímku. Pokud jsem si správně všiml, tak testovací sada obsahovala stejné

mluvčí (tj. vizuálně stejné objekty) jako trénovací. Bylo by určitě zajímavé otestovat navržený systém i pro případ kdy v testovací a trénovací sadě budou naprosto odlišné osoby. Mohlo by se ukázat, že emočně-identifikační skóre by bylo výrazně nižší. Každopádně, z popsaných experimentů je patrné, že nově navržený algoritmus s 3D rekonstrukcí (normalizací) dává o přibližně 10% lepší výsledky než běžné systémy bez 3D rekonstrukce, což je výrazné zlepšení vztažené k absolutní výši identifikačního skóre.

Po formální stránce mě poněkud překvapila zvláštní grafická úprava některých obrázků (obr 1.1, obr 4.10...), kde je velikost fontu několikanásobně větší než je velikost fontu psaného textu. Také řazení abstraktu až na konec práce je, z mého hlediska, nestandartní. Přílohy mohly být v této disertační práci regulérními kapitolami. V práci se objevuje nezanedbatelný počet dělených slov (natoče-ní – str. 2, před-stav – str. 8 ...), Chápu, že tato chyba byla pravděpodobně způsobena sázecím programem, přesto mohla být jistě před tiskem finální práce odstraněna.

Doktorandka ve své práci zmiňuje 14 publikací a 7 patentů. U těchto vědeckých výstupů je doktorandka hlavní autorkou nebo spoluautorkou. Toto množství (kvantitu i kvalitu) považuji za dostatečné.

Závěr

Doktorandka se v předložené disertační práci zaobírala aktuálním tématem využití metod a algoritmů 3D vidění pro rozpoznávání výrazu tváře. Jedná se o náročnou problematiku, jak z teoretického hlediska, tak i z pohledu praktické realizace. V práci byl představen a úspěšně otestován algoritmus pro identifikaci výrazu lidské tváře s podporou 3D rekonstrukce. Další zlepšení navržené identifikace výrazu tváře by jistě přinesla identifikace dynamiky pohybu hlavy a dynamika změn výrazu tváře, jak správně doktorandka ve své práci uvádí. K vlastní praktické realizaci nemám žádné výhrady, naopak kladně hodnotím schopnost doktorandky samostatné tvůrčí výzkumné práce. Celkově předložená disertační práce splňuje po formální, teoretické i praktické stránce všechny potřebné náležitosti. Na základě těchto skutečností **doporučuji tuto disertační práci k obhajobě**.

V Liberci dne 19.11.2013

doc. Ing. Josef Chaloupka, Ph.D.
ITE, FM, TUL