

Oponentský posudek disertační práce Ing. Karla Palečka

Doc. Ing. Miloš Železný, Ph.D., katedra kybernetiky FAV ZČU v Plzni

V předložené písemné práci „Audiovizuální rozpoznávání řeči s využitím metod pro automatické odezírání ze rtů“ se autor jako hlavním cílem zabývá návrhem robustní a dostatečně diskriminační vizuální parametrizace vhodné pro audiovizuální rozpoznávání řeči nezávislé na řečníkovi. Jedním z hlavních prvků vizuální parametrizace je využití trojrozměrné informace z hloubkových senzorů. Dalšími dílčími cíli bylo sestavení dostatečně rozsáhlé audiovizuální databáze pro otestování navrhovaných postupů, srovnání nejrozšířenějších parametrizací na více audiovizuálních databázích, systematické vyhodnocení přínosu trojrozměrných dat a srovnání parametrizací a posouzení přínosu vizuální složky v úloze rozpoznávání spojitě řeči s velkým slovníkem. Z předložené práce vyplývá, že stanovené cíle byly splněny.

Práce je napsána v češtině a na cca 120 stranách shrnuje přehled metod používaných pro audiovizuální rozpoznávání řeči, konkrétně metod detekce obličejových částí, vizuálních parametrizací, metod klasifikace a metod kombinace více zdrojů a popisuje navržené metody vizuální parametrizace řeči, přípravu dat a návrh testovacího protokolu a výsledky experimentů s rozpoznáváním izolovaných slov a frází a spojitě řeči. Vše nasvědčuje tomu, že autorem navržená parametrizace metoda může být úspěšně využívána v úlohách audiovizuálního rozpoznávání řeči.

Přínos pro obor: Za největší přínos práce lze považovat nově navržené parametrizace, zahrnující 3D model dat. Se zvýšením dostupnosti levných hloubkových senzorů se zlepšily možnosti zahrnutí 3D dat do modelu a tato práce toho efektivně využívá. Dalším významným přínosem je srovnání nejrozšířenějších typů parametrizací na několika audiovizuálních databázích a vytvoření nové, středně rozsáhlé audiovizuální databáze. Výsledky práce umožní návrhy robustnějších systémů audiovizuálního rozpoznávání řeči v případech, kdy je k dispozici hloubkový senzor.

Postup řešení: Autor se dobře orientuje v aktuálním stavu výzkumu ve světě, svoje metody zakládá na fungujících aktuálních metodách a tyto vylepšuje vhodnými úpravami, nastavením a kombinací. Postup řešení problému je vhodný. Popis navržených metod je však oproti popisu současného stavu velmi krátký. Dále je na základě tabulky 11.3 možno diskutovat o smyslnosti vizuálního rozpoznávání spojitě řeči s velkým slovníkem, když i při použití metodiky přesnosti rozpoznání slov hodnoty s výjimkou minimálního slovníku zůstávají fakticky na nule.

Formální stránka práce: Rozsah práce je dostatečný, vlastní struktura práce je vhodně navržena, nicméně vzájemný poměr rozsahu jednotlivých kapitol by mohl být příznivější ve prospěch navržených metod oproti současnému stavu. Kapitola č. 8 po 70 stranách shrnutí současného stavu zabírá pouze 4 strany. Kladně lze hodnotit rozhodnutí práci vysázet v programu TeX, nicméně v takovém případě je třeba manuálně ošetřit ponechání neslabičných předložek na konci řádku. Hustota této chyby kazí jinak velmi dobrou jazykovou i typografickou úroveň práce.

Publikační aktivita: Autor ve své práci uvádí seznam 9 publikovaných prací. Přestože většina z nich je indexována v databázích Scopus, IEEE Xplore a některé i WoS, jsou to všechno příspěvky z konferencí, čistě časopisecký příspěvek chybí. Nicméně z důvodu počtu publikací a toho, že 4 jsou indexovány v WoS a 5 ve Scopusu, považuji publikační aktivitu za přiměřenou a dostatečnou.

Závěr: Na základě prostudování předložené disertační práce a vzhledem k jejímu výše uvedenému zhodnocení je možno konstatovat, že disertační práce pana ing. Karla Palečka splňuje podmínky samostatné tvůrčí vědecké práce a obsahuje původní a autorem publikované nebo k publikaci přijaté výsledky. **Disertační práci tedy doporučuji k obhajobě.**

V Plzni dne 13. dubna 2016



Doc. Ing. Miloš Železný, Ph.D.
katedra kybernetiky FAV ZČU v Plzni

Prof. Ing. Václav Hlaváč, CSc.

Český institut informatiky, robotiky a kybernetiky ČVUT a
Elektrotechnická fakulta ČVUT, katedra kybernetiky
166 36 Praha 6, Zikova 4, telefon 224 357 465, FAX 224 357 385
email hlavac@ciirc.cvut.cz, <http://people.ciirc.cvut.cz/hlavac>

Praha 21. února 2016

Posudek na disertační práci Ing. Karla Palečka**Úvod**

Text disertace s názvem Audiovizuální rozpoznávání řeči s využitím metod pro automatické odezírání ze rtů je psán česky na 140 stranách. Disertace se mi četla dobře, je typograficky výborně zpracovaná a obsahuje jen velmi málo překlepů (např. jednopísmenné předložky na konci řádků). Školitelem doktoranda byl doc. Ing. Josef Chaloupka, Ph.D. z Fakulty mechatroniky, informatiky a mezioborových studií Technické univerzity v Liberci.

Aktuálnost tématu

Při rozpoznávání řeči v náročných podmínkách pomáhá sdružování informace z řečového kanálu a obrazové analýzy videosekvence rtů mluvčího podobně jako odezírání u lidí. Toto téma bylo již zkoumáno jinými, ale pro češtinu a jiné slovanské jazyky stále vyřešené není. Téma je zajímavé jak z teoretického, tak i praktického hlediska.

Za významné považuji, že autorovy výsledky posílily jedno z českých pracovišť aktivních v této oblasti. Srovnatelně vysokou kvalitu mají jen týmy na Západočeské univerzitě v Plzni a Vysokém učení technickém v Brně.

Cíle disertace

Disertace deklaruje v sekci 1.2 za cíl návrh robustní a dostatečně rozlišující parametrizace videosekvence rtů mluvčího, která pomůže rozpoznávání nezávislém na řečníkovi. Pro snímání úst je použito zařízení (Kinect 1 od Microsoft), které zaznamenává současně RGB kanály ve viditelné části spektra a pomocí strukturovaného světla poskytuje i informaci o vzdálenosti od kamery. Dále autor chtěl navrhnout nové metody rozpoznávání. Pro jejich ověření mezi cíli stanovil i vytvoření dostatečně rozsáhlé audiovizuální sekvence anotovaných dat pro učení klasifikátorů i jejich validaci. (Této trénovací sekvenci se v práci nesprávně říká audiovizuální databáze, i když databáze je nástroj pro práci s libovolnými daty.) Dalším cílem disertace je rozsáhlé experimentální ověření.

Cíle disertace jsou náročné, vědecky zajímavé a jsou dobře zakotveny ve stavu vědění.

K metodám zpracování disertace

Disertace se využívá osvědčenou metodiku: poskytnout přehled o stavu vědění, popsat nově navržené metody a experimentálně je ověřit. Stav vědění je v kapitolách 2 až 8. Práce navazuje na výsledky více oborů, a tak je popis stavu vědění rozsáhlý. Výhodou je, že v kapitolách lze čtenáři přiblížit látku téměř učebnicovým způsobem, což se v disertaci stalo.

Kapitola 2 se zabývá detekcí obličejových částí. Kapitola 3 se věnuje stavu vědění ve vizuální parametrizaci, vědeckém jádru disertace. Kapitola 4 uvádí metody klasifikace. Kapitola 5 popisuje metody fúze z více zdrojů. Kapitola 6 informuje o audiovizuálních trénovacích sekvencích, a to jak existujících, tak i vlastních. Kapitola 7 znovu uvádí relevantní stav poznání, který je tentokrát vztažen k vlastním cílům a výsledkům autora. Návrh nové vizuální parametrizace řeči je v kapitole 8. Kapitola 9 je servisní o přípravě dat a testovacím protokolu. Kapitola 10 experimentuje s izolovanými slovy a frázemi. Kapitola 11 popisuje totéž pro spojitou řeč.

K výsledkům disertace, splnění cílů

Hlavní výsledky disertace jsou v kapitole 8 (návrh nové metody parametrizace) a v jejích experimentálním ověření v kapitolách 10 a 11. Dosažené výsledky prokazatelně překonávají stav vědění a výsledky jiných, nejlepších v této oblasti.

Stanovené cíle považuji za splněné. Soudím, že disertace je důležitým milníkem pro pracoviště autora. Oceňuji, jak byly předchozí výsledky na stejném pracovišti dobře využity. Ne na každém pracovišti se daří nabytou znalost udržet a stavět na ní dále.

Kritika

Jak jsem již napsal, autorovy výsledky považuji za významné a v disertaci dobře zdokumentované. Moje kritika míří především ke školícímu pracovišti. Považuji za nesprávné, že text disertace je v češtině. Jakkoliv mám svůj mateřský jazyk rád, je potřebné psát disertace v našich technických oborech anglicky. Jinak bude mít text velmi málo čtenářů, a tudíž malý impakt ve vědecké komunitě. Druhá kritika se týká požadovaných publikací kandidáta na vědeckou hodnost Ph.D. Na našem pracovišti se požaduje alespoň jeden článek v časopise zaznamenaném ve Web of Science. Výsledky disertace by jistě na takto kvalitní časopisecký článek postačovaly. Autor disertace žádný svůj časopisecký článek nezmiňuje.

Otázky k obhajobě

1. Kinect má poměrně malé rozlišení. Jaké zlepšení lze očekávat, když budou mít obrazová a hloubková data větší prostorové případně i časové rozlišení? Podnikl autor v tomto směru nějaké experimenty?
2. V disertaci není citovaná práce Ivan Laptev, On Space-Time Interest Points, International Journal on Computer Vision 64 (2/3), 2005, strany 107-123. Zná autor tuto práci? Jaký je její vztah k disertaci?
3. Jsou publikace autora citovány jinými (bez autocitací)?

Závěr

Výsledky doktoranda i disertaci považuji za velmi dobré. Disertace shrnuje obrovské množství práce. Výsledky práce byly publikovány na konferencích. O citacích na výsledky disertace s práce nezmiňuje.

Soudím, že doktorand v disertaci prokázal schopnost samostatné vědecké práce.

Práci doporučuji práci k obhajobě a doporučuji, aby K. Palečkovi byl v případě úspěšné obhajoby udělen titul Ph.D.

Václav Hlaváč