

Oponentní posudek disertační práce

Název práce: Automatické rozpoznávání znakového jazyka z obrazových dat

Student: Ing. Pavel Campr

Výzkumná práce studenta, jejíž průběh a závěry jsou popsány v disertační práci, je zaměřena na vytvoření systému automatického rozpoznávání znakového jazyka z obrazových dat. Téma práce je aktuální a zajímavé, neboť v této oblasti výzkumu existuje mnoho nezodpovězených výzkumných otázek. Aplikační potenciál výzkumné práce je vysoký, především v komunikaci sluchově postižených osob s počítačem či osobami, které neovládají znakový jazyk. Význam disertační práce pro daný obor je vysoký, především z pohledu množství provedených experimentů, které porovnávají různé přístupy parametrizace data pro rozpoznávání izolovaných slov a spojitě řeči.

Vytyčené cíle disertační práce obsahují sedm bodů, přičemž tři z nich jsou zaměřeny na vlastní přínos studenta k danému oboru. Jedná se především o vylepšení metody detekce a sledování rukou a hlavy ve video záznamu, prozkoumání možnosti využití pod-jednotek pro rozpoznávání a vytvoření systému pro vyhledávání na základě příkladu. Ostatní cíle jsou zaměřeny na vytvoření korpusů využitelných pro automatický systém rozpoznávání či porovnání různých parametrizací pro rozpoznávání znakového jazyka.

Všechny vytyčené cíle byly splněny a jsou dostatečně v disertační práci popsány. Student prokázal schopnost vědecké práce především velmi kvalitní volbou metod jak pro samotné řešení problému, tak pro zhodnocení a vyhodnocení získaných poznatků. Metody použité pro detekci a sledování rukou a obličeje, metody parametrizací a metody pro rozpoznávání izolovaných slov a spojitě řeči odrážejí aktuální stav poznání. Student vytvořil systém pro automatické rozpoznávání izolovaných znaků pro laboratorní podmínky, který dosahuje úspěšnosti rozpoznávání závislém na řečnickovy přes 93% a přes 34% pro rozpoznávání nezávislé na řečnickovi pro korpus SLR-P.

Hlavním výsledkem disertační práce je vytvoření dvou databází izolovaných znaků, využitelných pro porovnání různých přístupů automatického rozpoznávání znakové řeči a samotné srovnání a důkladného zhodnocení výsledků a velkého množství parametrizací obrazových dat pro rozpoznávací proces. Všechny výsledky práce jsou dostatečně popsány a analyzovány. Vyzdvihují především systematičnost a preciznost při zpracování velkého množství experimentů a jejich přehledné vyhodnocení s diskusí. Vlastní přínos studenta spatřuji ve zdokonalení metody detekce a sledování rukou a hlavy ve video záznamu, kdy student vytvořil vlastní metodu, která dosahuje lepších výsledků než metody stávající. Z vyhodnocení experimentů ovšem plyne, že přínos vytvořené metody je pouze u rozpoznávání provedeném na korpusu SLR-A nezávislém na řečnickovi. Druhým hlavním vlastním přínosem je vytvoření metody pro identifikaci pod-jednotek a jejich využití pro rozpoznávání. Vytvořený postup prokazuje vylepšení oproti základnímu systému, ovšem zdaleka nedosahuje výsledků rozpoznávání založeném na modelování celých znaků.

Formální a jazyková úprava je velmi kvalitní. Test práce obsahuje minimální množství chyb. Členění práce je logické a přehledné.

Student předkládá v práci přehled 22 publikací. U pěti publikací je student uveden jako hlavní autor. Většina publikací je publikace ve sborníku z konference, kdy některé z nich patří k významným v oboru. Pouze část publikací je přímo vztažena k řešení a cílům disertační práce a proto je disertační práce hlavním zdrojem informací.

K práci mám následující připomínky:

- Kapitola 5.1.5 obsahuje popis metody „Active appearance model“, která je použita i pro následný popis výrazu tváře. Některá implementace této metody, řeší i problém s detekcí obličeje v případě překryvů. Tato implementace by byla velmi vhodná právě pro rozpoznávání znakového jazyka.
- Kapitola 6.1 – V kapitole je zmiňováno, že všechny metody použité pro rozpoznávání s korpusem SLR-A a SLR-P je možné použít i pro ostatní korpusy, které zaznamenávají jednoho řečníka z čelního pohledu. Toto tvrzení je odvážné, neboť ostatní korpusy mohou být zaznamenány za značně odlišných podmínek a metody detekce a sledování rukou, či parametrizace mohou selhávat.
- Kapitola 6.2.4 popisuje problému použití metody AAM při částečném překryvu obličeje. Je zde zmiňováno, že metoda vytvořená studentem pro odlišení rukou a obličeje při vzájemném překryvu tyto problémy řeší. Metoda ovšem nahrazuje skutečný tvar obličeje při překryvu průměrným tvarem obličejů z oblastí bez překryvu. Při překryvu je tedy zbytečné využívat metodu AAM, neboť nedochází k získání žádné nové informace. Dále je v kapitole uvedeno, že popisovač aam-ext eliminuje použití parametrů závislých na řečnickovi. Popisovač aam-ext ovšem obsahuje parametry jako výška a šířka vnější hranice rtů, která je bez normalizace závislá na řečnickovi.
- Kapitola 6.3.3 – V kapitole je diskutováno zhoršení výsledků rozpoznávání při použití nemanuální složky (výraz obličeje). Problém je vysvětlován tím, že popis výrazu obličeje obsahuje více informací o pohybech hlavy, které nemají vztah k znakovému slovu a tím pádem se jedná o šum. Variace úspěšnosti rozpoznávání z jednotlivých opakování experimentů by mohla být vodítkem, zda je toto vysvětlení správné.
- Kapitola 6.4 má název rozpoznávání znakového jazyka pro spojitou řeč. V kapitole je ovšem popisován případ pospojování izolovaných znaků, na kterém jsou prováděny rozpoznávací experimenty. Tento simulovaný případ může být ovšem velmi vzdálen od reálné spojitě řeči, neboť nejen okrajové části znaku, ale i samotná centrální část znaku může u spojitě řeči značně odlišná z důvodu návaznosti znaků a rychlosti jejich znakování.

Celkově hodnotím práci jako velmi zdařilou. Srovnání mnoha metod parametrizace přináší nové poznání pro danou oblast. Je škoda, že se student nesoustředil na vylepšení metod pro parametrizaci dat pro případ rozpoznávání nezávislém na řečnickovy, kde metoda selhává a věnoval se dalším částem jako například vyhledávání podle vzoru. V práci také postrádám porovnání úspěšnosti rozpoznávání při využití 3D pozic rukou a 2D pozic rukou na které je používán korpus připraven. Toto srovnání by objasnilo vhodnost použití nových zařízení jako například Kinect pro reálné nasazení tohoto systému.

Autor ve své disertační práci **prokázal** schopnost samostatné vědecké práce v daném oboru. Z tohoto důvodů **doporučuji** disertační práci k obhajobě.

V Nových Hradech, dne 21.8.2013

.....
Ing. Petr Císař Ph.D.

Review of Dissertation Thesis
Automatic Sign Language Recognition from Image Data
submitted by Pavel Campr
at the Faculty of Applied Sciences, ZČU in Pilsen

The thesis was completed in December, 2012. It is written in English and has 105 pages.

Scientific background of the thesis:

As speech technology is getting more and more advanced, researchers focus on more specialized and more challenging tasks. One of them is sign language recognition. The systems for sign language recognition do not have wide commercial use, but they can help deaf people or people with a hearing handicap mainly in interaction with a computer.

Main objectives of the work:

The thesis of Pavel Campr deals with the design of a system for automatic sign language recognition. The author focuses mainly on developing and testing methods and algorithms for sign language feature extraction, hand tracking from video recordings of sign language and feature selection and fusion. The resulting set of features seems to be useful for sign language recognition. I appreciate greatly the effort Pavel has put into the creation of special databases of different forms of sign language, a very time-consuming task. He is the co-author of the Czech sign language online dictionary and he has developed a system for data-driven construction of sub-units for sign language recognition.

The state of the art is described briefly in the thesis. In the introductory part (chapter 3), the author briefly describes the sign language and its use in human - computer interaction. Chapter 4 is about sign language video databases. Two such databases were also made at the Department of Cybernetics and Pavel Campr is one of the authors. In chapter 5, he describes several approaches and methods of video analysis, where it is necessary to solve the problem of skin color, face detection, hand and head tracking and extraction of features. Here, the author uses methods and algorithms already known from image and signal processing, e.g. histogram of oriented gradients, local binary patterns, active appearance models, principal component analysis etc. Section 5.2 describes sign modeling by HMM and section 5.3 is about language modeling. The equations in section 5.4 describe sign decoding methods and accuracy evaluation. The last part (5.5) introduces the system search-by-example for searching in sign language corpora. In chapter 6, the author presents two original systems for sign language recognition: the one of isolated signs and the one of continuous speech. The evaluation of these systems using his own databases is presented in this part too. The results (accuracy/correctness) obtained during single experiments are very good, owing to the complexity of the recognition task, but it is necessary to say that both the databases for experiments are relatively small.

Methods and methodology:

The thesis demonstrates that Pavel Campr has become familiar with all the basic concepts and algorithms used in image and signal processing and recognition. All the described methods and algorithms are well designed and it is clear from the results that they are useful and robust enough for sign language recognition of isolated signs or continuous speech.

Results and achievements:

In his thesis, Pavel Campr presents the results of several experiments to demonstrate how his systems for sign language recognition function. The author has done a lot of research work, obtained interesting results and proposed several original methods, e.g. a novel hand tracking with face occlusion handling or data-driven construction of sub-units for sign language recognition. The results shown in tables and graphs are of sufficient informative value and it appears that it will be possible to use these systems in real conditions, e.g. in informative kiosks for deaf people or people with impaired hearing who use sign language. However, the thesis does not include any comparison to other systems for sign language recognition, which is the weak point. Otherwise, the results are presented clearly, with detailed description and discussion, and they are well illustrated with abundance of images, graphs, and tables.

Scientific contribution of the thesis:

The thesis introduces several algorithms for feature extraction, fusion and hand tracking from video images and their use in sign language recognition. These algorithms and methods appear to have good potential for practical application. The text of the thesis is clear, well organized and contains minimum typos. However, strangely, pages 4, 12, 20, 44, 82, 86, 98, 102 are completely blank and they do not have page numbers. I believe this is due to inattention while typing.

As to Pavel Campr's publishing activity: So far, his list of papers contains 23 already published conference or journal contributions relevant to his thesis. Several of them are indexed in the ISI - Web of Knowledge database. This clearly shows that his research has been known to the community.

Conclusion:

The author of the thesis has proved ability to conduct research and to obtain original scientific results. I recommend the thesis for presentation with the aim of receiving the degree of Ph.D.

Questions:

- 1) Have you ever tested any other fusion methods, such as those discussed in part 5.14? Would it be appropriate to use weights for single streams?
- 2) It is presented in chapter 5.1.1 that you used similar rules and thresholds for skin color segmentation to those described in Kovac's paper. How successful is this method with people with darker color of the skin? Did you use any other method for skin color segmentation? With what success?

In Liberec, March 19th, 2013


doc. Ing. Josef Chaloupka, Ph.D.

Institute of Information Technology and Electronics
Technical University of Liberec