Computer Vision and Digital Inclusion of Persons with Special Needs: Overview and State of Art

Hemerson Pistori
UCDB - Dom Bosco Catholic University, Campo Grande, Brazil
GPEC - Research Group in Engineering and Computing

ABSTRACT: This survey paper addresses some issues related to the application of computer vision techniques to improve the welfare of people with special needs. The main problems and current work on topics like sign language processing and wheelchair control will be presented. The paper also introduces an ongoing project that aims at creating a free software environment that will include implementations of a large amount of computer vision, pattern recognition and machine learning techniques, tuned to the problems related to the digital inclusion of people with special needs. The software will also serve as an experimental environment, where new techniques will be implemented, tested and compared

1 INTRODUCTION

Unfortunately, worldwide, up to date statistics on the quality of life of people with disabilities are not available. However, a recent comprehensive survey on disabled people situation in the United States revealed some very important issues. According to this survey, only 35 percent of adults with special needs are employed, 26 percent live in poverty and 21 percent do not complete high school studies. All these numbers are two or three times worse than for people without disabilities (Krane and Hanson 2004), and probably represent an optimistic upper bound on the situation for most of the countries, particularly the poorest ones.

The United States census also estimated, in 2002, that 11.5 percent of their citizens had severe disability, while 18.1 percent presented some level of disability (Steinmetz 2006). The projection of these numbers to the world population would give more than one billion of people with some level of disability. Computer based assistive technologies are emerging as an important mean to improve communication, mobility and self care abilities for people with special needs. This paper focuses on computer vision based assistive technologies and present current work with high relevance to the problem of deafs and quadriplegics social inclusion.

Besides the social justification for the researches on computer vision based assistive technologies, the problems faced in this area are specially challenging and serve as testbeds for many computer vision techniques. Usually, the systems must implement multiple deformable objects tracking, in complex and moving background, with noise and occlusions. In this paper, two prominent problems will be surveyed: (1) the automatic translation of sign language to written language and (2) the control of a wheelchair using face expressions and movements. Works on each of this problems are presented in the next two sections. A software specially devised to serve as a complete environment for the development of human-machine interaction systems based on computer vision is described in Section 4. The last section is reserved for discussions, conclusions and comments on emerging research topics.

2 SIGN LANGUAGE RECOGNITION

Two important misconceptions are usually associated to sign languages. First, sign languages are not universal. Estimates on the number of sign languages around the world vary from 4000 to 20000 (Woll, Sutton-Spence, and Elton 2001), and as it happens with oral languages, they also present regional dialects and accents. Signs in sign languages are not simple mappings from words or letters in oral languages. Sign languages have their own grammatical structures, many of them without an oral language analog, and which evolve naturally in a very sophisticated, and not completely controlled way, in order to optimize the use of body parts shape, movement and spatial relation. Signs that correspond to alphabetical digits do exist, however, they are only used in some special cases, like in proper names finger-spelling.

As the mechanical movements of hands, face and torso are much slower than the vocal tract control, sign languages are highly dependent on context sensitivity and a sort of multi-channel information transmission in order to achieve the same speed, in communication, as oral languages. Long sentences, in spoken language, can be translated to short sentences in sign language, by the parallel use of facial expressions and different shapes and movements from the right and left hand. Even the written representation of sign languages is an open and challenging problem, as current formal language theory is not appropriate in this context (Huenerfauth 2005).

A sign language recognition system presents almost all of the most difficult and interesting problems in computer vision, as it requires real-time segmentation, tracking and classification of multiple, deformable, self-occluding and non-linear moving objects. The first works on sign language recognition focused on static hand shape recognition (Pistori and Neto 2004), however, in the last five year, many important results on the recognition of complete gestures and face expressions have been reported.

As it happens in speech recognition, hidden Markov models are the most cited technique in gesture recognition. In (Holden, Lee, and Owens 2005), correct classification rates of 97% at the sentence level, and of 99% at the word level have been reported. The proposed system is based on hidden Markov models, trained with 379 utterances, and tested on 163 sentences from the Australian Sign Language. A two stage classifier, combining hidden Markov models and support vector machines were tested on a very large vocabulary, with 4942 different signs and 59304 samples of Chinese Sign Language. An 89.40% correct classification rate was achieved, with the help of two cyber gloves and 3 space-position tracker, to circumvent the hands tracking problem (Ye, Yao, and Jiang 2004).

An attempt to achieve good generalization, with fewer samples, has been reported in (Bowden, Windridge, Kadir, Zisserman, and Brady 2004). In Bowden work, Independent Component Analysis, hidden Markov models and high-level linguist knowledge were combined to build a system that correctly recognised 96.7% of 43 words from the British sign language, using just 1 sample of each word for training and 200 for testing. This work also builds on the assumption that the dominant hand convey most of the information necessary to recognize the 43 words tested.

Another dominant topic in sign language recognition is skin color based segmentation. Several different color spaces, including HSV, normalized RGB, YCbCr have been investigated in the construction of both parametric and non-parametric skin color models. Real time performance, with a position error from 0.8 to 1.8 pixels, for tracking hands in a Thai Sign Language recognition system, has been recently reported in (Soontranon, Aramvith, and Chalidabhongse 2005). The combination of image features and information captured by accelerometers and data gloves is also being explored, with increments of up to 94.26% in tracking precision (Culver 2004).

In order to deal with accents from different signers, without recurrent to very large training sets, a recent work has evaluated the application of some adaptation techniques already in use in speech recognition. In (von Agris, Schneider, Zieren, and Kraiss 2006), Maximum Likelihood Linear Regression has been used to adapt the parameters of the underlying hidden Markov model. Experiments using a test corpus consisting of 153 isolated signs from British Sign Language, performed by four signers, in a control environment, showed a relative improvement of up to 41.6% compared to the independent user baseline (no adaptation).

Two of the most cited annotated test corpus, for sign language investigation, are the Purdue's RVL-SLLL (Martnez, Wilbur, Shay, and Kak 2002) and *signstream*, one of the most comprehensive collection of annotated videos of American sign language utterances, including images captured from different angles and signers. An extensive survey of 156 papers on sign language analysis can be found in (Ong and Ranganath 2005).

3 WHEELCHAIR CONTROL

Many people, including paraplegics, quadriplegics and elders, depend, permanently or temporarily, on wheelchair for locomotion. The control of an electric wheelchair when joysticks are not an option, includes devices that can be driven by tongue (Struijk 2006) and forehead muscle contractions (Felzer and Freisleben 2002). Computer vision, however, is being explored in the development of wheelchair control mechanisms based on gesture, facial expression, head movements and eye gaze. In this case, the users are free of any mechanical contact and images captured through cameras attached to the wheelchair are interpreted and used to produce the driving commands. Computer vision techniques, specially robotic navigation, are also being explored in the construction of the so called intelligent wheelchairs, which can recognize the environment to avoid collision and even drive, autonomously, the user to some predefined places (Ono, Uchiyama, and Potter 2004). This survey is focused on the works that use images from the user face and hand to generate commands to a wheelchair.

In a recent paper, the control of a wheelchair by face movements has been explored. Multilinear discriminant classifiers were trained to achieve 100% accuracy in face detection. Head pose (turn left and turn right) were estimated by support vector machines, with linear kernels, acchieving a 95.7% correct pose identification rate (Bauckhage, Kaster, and Rotenstein 2006). Matsumoto combined head motion, gaze direction, blinks and lip motion information, captured from a standard PC camera pair, to build a wheelchair controller that has been tested in real situations. Face detection has been implemented by template matching of 3d facial model that included eyebrows, eyes and mouth (Matsumoto, Ido, Takemura, Koeda, and Ogasawara 2004).

The system proposed by Jia, in (P. Jia and Yuan 2006), uses nose pose estimation, by template matching, to control the moving direction of the wheelchair. Five face expressions, discriminated using an algorithm that combines Adaboost and Camshift, working on Haar-like features extracted from the face image, were used to supply additional controls to the wheelchair, like stop and start moving. Tracking was implemented using the mean shift algorithm in the hue space. Only one, out-of-the-shell webcamera, were used as the image capture device. The problem of wrong command interpretation, when the user makes fast head movements in response to some external stimulus, has been tackled in (Kuno, Murakami, and Shimada 2001). The system distinguishes among slow and fast head movements and includes pedestrian recognition to avoid collision.

4 SIGUS

No general method can solve all the problems discussed in the last sections and the success of some algorithms and techniques are usually task dependent and demands a hard work in parameter fine tuning. The SIGUS platform [1] is a software environment intended to aid computer vision investigators in building and testing systems tuned to the problem of machine-human interaction using corporal expressions, including hand gestures, face expression and eye-gaze.

The platform is built upon two mature open source libraries, the ImageJ (Abramoff, Magelhaes, and Ram 2004; Rusband 2006), for digital image processing and the Weka (Witten and Frank 2005), for machine learning. These two packages were extended with new features and integrated to some other libraries, like randomj, for random number generation (used in Monte Carlo based approaches), mical, for grammar induction and jama, for linear algebra routines. Integration of the SIGUS platform to the SciLab package (Pires and Rogers 2002; Gomez 2006) is also under development.

Training datasets with thousands of images and tools to help annotation and ground truth generation are also part of SIGUS platform. The datasets will include images from hundreds of different gestures and sentences of Brazilian sign language, taken from persons with different skin colors, ages, signaling skill levels, in different environments (office, classroom, outdoor, etc) and illumination setup (natural, incandescent lamps, fluorescent lamps, etc). A face expressions and head movements dataset is also being created, with images taken from cameras attached to wheelchairs, in a large variety of environments and with different persons.

Ground truth sets are being generated to test segmentation, feature extraction, dimension reduction, tracking and classification algorithms in real situations. Different metrics and dissimilarity measures to be used both in performance comparison and by distance based algorithms, are being implemented, including Hausdorff, Chebyshev, Bhattacharrya, Chi-squared, Kullback-Leibler Divergence, Mahalanobis and Earth mover distance.

Several skin segmentation techniques are already available in the platform, including the ones based on mixture models, Gaussian models and some non-parametrics models. The platform also implements adaptive background segmentation, particle filter based tracking and via Weka, a great variety of supervised learning techniques, including state-of-art support vector machines and radial basis networks. Hidden Markov models, for sign language recognition, are also available. Most of the implemented algorithms presents graphical counterparts that can be used to monitor its execution in real time and to adjust parameters. For instance, in particle filters tracking, it is possible to monitor, in a intuitive graphical manner, the number and weights of the particle set.

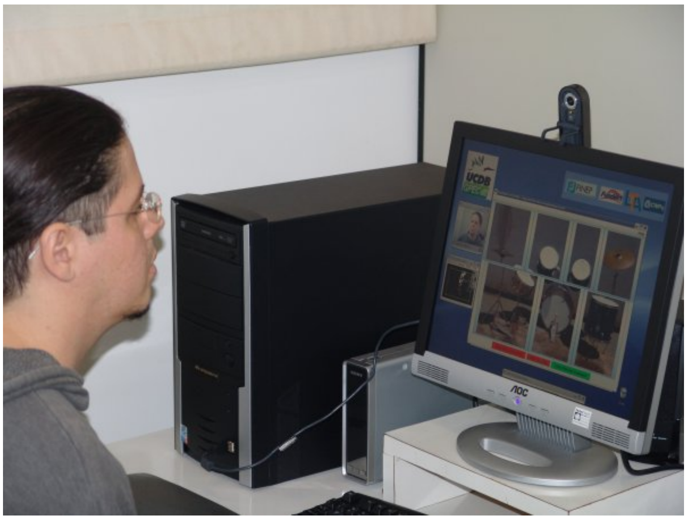

Figure 1: Example of an application created using SIGUS platform: a drum, for quadriplegic people, that can be played by face movements

[1] http://www.gpec.ucdb.br/sigus

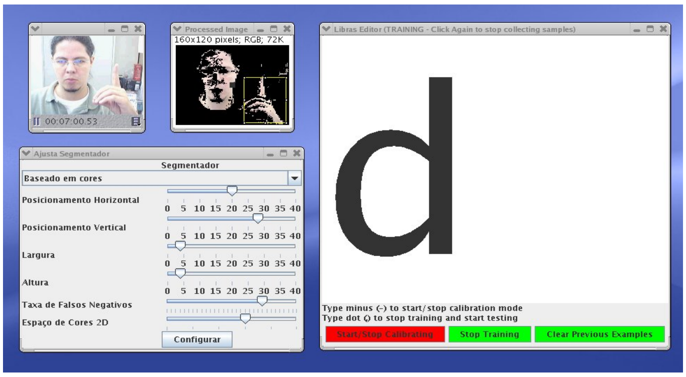

Figure 2: Example of an application created using SIGUS platform: a drum, for quadriplegic persons, that can be played by face movements

Figure 1 illustrates a software created using SIGUS platform. In this software, the user can associate different visual sign (face expressions, face movements or hands configuration, for instance), to each piece of a virtual drum. After a calibration and training phase, the virtual drum can be played without any direct contact to a mouse or keyboard, just by reproducing face or hand configurations in front of a standard webcam, positioned in front of the user. An alphabetical editor, that translates sign language hand configurations to letters and digits, is another example of a software prototype available with the SIGUS platform (Figure 2).

5 DISCUSSIONS AND CONCLUSIONS

This paper presented an overview of some recent work related to the use of computer vision techniques to improve human-machine interaction for people with special needs. A new free software platform, not intended to be just another general purpose computer vision package, but a development environment tailored to the programming and testing of systems for people with special needs were also presented.

The two problems presented in this paper, namely the sign language translation and wheelchair control have been investigated in other areas than computer vision, in systems based on another kind of sensors and devices. An important future work is the creation and application of new methods to compare the performance of systems based on very different technologies, like computer vision, electroculography (Williams and Kirsch 2005) or electroencefalogram (Kauhanen, Nykopp, Lehtonen, Jylnki, Heikkonen, Rantanen, Alaranta, and Sams 2006) based wheelchair control. There is also a great need for benchmark datasets. Authors usually report correct classification rates that cannot be mutually compared, as the testing assumptions are too different.

An important criticism to the work of many computer scientists that deals with systems for people with special needs is the lack of a more consistent and methodical interaction and participation of the final users during research, development and testing. Fehr (Fehr, Langbein, and Skaar 2000) experiments with some systems prototypes showed that many of the proposed human-machine interfaces, supposed to be aimed at people with special needs, presented serious problems of usability.

Another problem, not explored in this paper, that is being tackled using computer vision techniques, is vision-tactile substitution. The problem consists in translating the images captured from a camera, usually installed in sun-glasses, to a low-resolution matrix of tactile stimulators. In the most common approach, a matrix of electrodes is attached to the body of a blind person, usually in a place with high tactile sensibility, like the tongue, the belly, the fingertips or the forehead (Tachi 2006; Tyler 2000). The processed image or some high level information, extracted from the scene (like a text or a number), is “printed” on the skin of the blind, who can be trained to associate the “printed” signs to visual information.

Acknowledgments

This work has been funded by the Dom Bosco Catholic University, UCDB, the Foundation of Teaching, Science and Technology Development of Mato Grosso do Sul State, FUNDECT, and the Brazilian Studies and Projects Funding Body, FINEP. The author of this paper and some of his advisees hold scholarships from the Brazilian National Counsel of Technological and Scientific Development, CNPQ

REFERENCES

Abramoff, M. D., P. J. Magelhaes, and S. J. Ram (2004). Image processing with ImageJ. *Biophotonics International 11*(7), 36–42.

Bauckhage, C., T. Kaster, and A. M. Rotenstein (2006). Fast learning for customizable head pose recognition in robotic wheelchair control. In *FGR ’06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, Los Alamitos, CA, USA, pp. 311–316. IEEE Computer Society.

Bowden, R., D. Windridge, T. Kadir, A. Zisserman, and M. Brady (2004, May 11-14). A linguistic feature vector for the visual interpretation of sign language. In *ECCV 2004, 8th European Conference on Computer Vision*, pp. 390–401.

Culver, V. R. (2004). A hybrid sign language recognition system. In *Eighth IEEE International Symposium on Wearable Computers (ISWC’04)*, Los Alamitos, CA, USA, pp. 30–33. IEEE Computer Society.

Fehr, L., W. E. Langbein, and S. B. Skaar (2000). Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey. *Journal of Rehabilitation Research and Development 37*(3), 353–360.

Felzer, T. and B. Freisleben (2002). Hawcos: the "hands-free" wheelchair control system. In *Proceedings of the fifth international ACM conference on Assistive technologies (ASSETS)*, pp. 127–134.

Gomez, C. (2006, Junho). Scilab. *Available at http://www.scilab.org/*.

Holden, E.-J., G. Lee, and R. Owens (2005). Automatic recognition of colloquial australian sign language. In *WACV-MOTION '05: Proceedings of the IEEE Workshop on Motion and Video Computing (WACV/MOTION'05) - Volume 2*, Washington, DC, USA, pp. 183–188. IEEE Computer Society.

Huenerfauth, M. (2005). Representing coordination and non-coordination in an american sign language animation. In *Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, New York, NY, USA, pp. 44–51. ACM Press.

Kauhanen, L., T. Nykopp, J. Lehtonen, P. Jylnki, J. Heikkonen, P. Rantanen, H. Alaranta, and M. Sams (2006). EEG and MEG brain-computer interface for tetraplegic patients. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 14*(2).

Krane, D. and K. Hanson (2004). *National Organization on Disability and Harris 2004 Survey of Americans with Disabilities*. New Work: Harris Interactive.

Kuno, Y., Y. Murakami, and N. Shimada (2001). User and social interfaces by observing human faces for intelligent wheelchairs. In *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*, New York, NY, USA, pp. 1–4. ACM Press.

Martnez, A., R. Wilbur, R. Shay, and A. Kak (2002). Purdue RVL-SLLL ASL database for automatic recognition of american sign language. In *Proc. IEEE International Conference on Multimodal Interfaces*.

Matsumoto, Y., J. Ido, K. Takemura, M. Koeda, and T. Ogasawara (2004, May 17-19). Portable facial information measurement system and its application to human modeling and human interfaces. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2004)*, pp. 475–480.

Ong, S. C. W. and S. Ranganath (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. Pattern Anal. Mach. Intell. 27*(6), 873–891.

Ono, Y., H. Uchiyama, and W. Potter (2004). A mobile robot for corridor navigation: a multi-agent approach. In *ACM-SE 42: Proceedings of the 42nd annual Southeast regional conference*, New York, NY, USA, pp. 379–384. ACM Press.

P. Jia, H. Hu, T. L. and K. Yuan (2006, October). Head gesture recognition for hands-free control of an intelligent wheelchair. *Journal of Industrial Robot 33*(6).

Pires, P. S. M. and D. A. Rogers (2002, November). Free/open source software: An alternative for engineering students. In *Proceed. of FIE - Frontiers in Education*.

Pistori, H. and J. J. Neto (2004, September 29). An experiment on handshape sign recognition using adaptive technology: Preliminary results. *Lecture Notes in Artificial Intelligence. XVII Brazilian Symposium on Artificial Intelligence - SBIA'04 3171*, 464–473.

Rusband, W. (2006, Junho). Imagej. *Available at http://rsb.info.nih.gov/ij/*.

Soontranon, N., S. Aramvith, and T. H. Chalidabhongse (2005). Improved face and hand tracking for sign language recognition. In *ITCC '05: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05)*, Washington, DC, USA, pp. 141–146. IEEE Computer Society.

Steinmetz, E. (2006, May). Americans With Disabilities: 2002 - Household Economic Studies. Technical Report Current Population Reports, Bureau of The Census - U.S. Department of Commerce, Washington, DC.

Struijk, L. N. S. A. (2006). A tongue based control for disabled people. *Lecture Notes in Artificial Intelligence. Computers Helping People with Special Needs. 4061*, 913–918.

Tachi, H. K. Y. K. S. (2006). Forehead electro-tactile display for vision substitution. In *EuroHaptics*.

Tyler, K. A. K. M. E. (2000). Effect of electrode geometry and intensity control method on comfort of electrotactile stimulation on the tongue. In *Proc. of the ASME, Dynamic Systems and Control Division*, Volume 2, pp. 1239–43.

von Agris, U., D. Schneider, J. Zieren, and K.-F. Kraiss (2006, June 17-22). Rapid signer adaptation for isolated sign language recognition.

Los Alamitos, CA, USA, pp. 159–164. IEEE Computer Society.

Williams, M. and R. Kirsch (2005, July 5-8). Feasibility of electroculography as a command interface for a high tetraplegia neural prosthesis. In *10th Annual Conference of the International Functional Electrical Stimulation Society*, Montreal, Canada.

Witten, I. H. and E. Frank (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann.

Woll, B., R. Sutton-Spence, and F. Elton (2001, October). *The Sociolinguistics of Sign Languages*, Chapter Multilingualism - the global approach to sign languages. Cambridge, UK: Cambridge University Press.

Ye, J., H. Yao, and F. Jiang (2004). Based on HMM and SVM multilayer architecture classifier for chinese sign language recognition with large vocabulary. In *Third International Conference on Image and Graphics (ICIG'04)*, pp. 377–380.