HUMAN–COMPUTER INTERACTION BASED ON FACIAL EXPRESSION RECOGNITION: A CASE STUDY IN DEGENERATIVE NEUROMUSCULAR DISEASE

Andreia Matos[1], Vítor Filipe[1,2], Pedro Couto[1,3]

[1]School of Science and Technology, University of Trás-os-Montes e Alto Douro, 5001–801 Vila Real, Portugal

[2]INESC TEC, Rua Campo Alegre 687, 4169-007 Porto, Portugal

[3]CITAB, University of Trás-os-Montes e Alto Douro, Vila Real, Portugal.

andreiaj@sapo.pt, vfilipe@utad.pt, pcouto@utad.pt

ABSTRACT

Physical disability can, in certain cases, be a barrier for traditional human-computer interaction based on keyboard and mouse devices. Alternative ways of interaction based on computer vision may be successfully adapted in particular cases of disability. This paper purposes a vision-based assistive technology to help a child with a degenerative neuromuscular disease to interact with the computer through facial expression recognition. The proposed algorithm was evaluated in images extracted from videos of the child and the preliminary results indicate that computer-interaction via facial expression recognition can break down barriers for people with reduced mobility regarding their relation with computers.

CCS Concepts

• **Human-centered computing ~ User studies** • Human-centered computing ~ Empirical studies in HCI • **Human-centered computing ~ Accessibility technologies** • **Computing methodologies ~ Object recognition** • Computing methodologies ~ Image processing

Keywords

Physical disability; Human-computer interaction; Face detection; Facial expression recognition.

1. INTRODUCTION

According to the International Classification of Functioning, Disability and Health, disabilities are "problems in body function or structure as a significant deviation or a loss" [1]. There are several types of disabilities, but the physical disability is the one that presents more difficulties in the use of hardware to access the computer. Physical disability is "complete or partial change of one or more segments of the human body, causing the impairment of physical function, appearing in the form of paraplegia, paraparesis, monoplegia, monoparesis, quadriplegia, tetraparesis, triplegia, triparesia, hemiplegia, hemiparesis, amputation or absence of the member, cerebral palsy, members with congenital or acquired deformities, except cosmetic deformities and those which do not cause difficulties for the performance of functions" [2].

With notorious development over the last few years, technology is increasingly present in our quotidian. There is the need to guarantee equal access to technology for all people, however there are people with specific disabilities to whom this interaction is impossible.

Human-computer interaction (HCI) is a discipline concerned on the communication between users and computers. It has the objective in improving safety, effectiveness and usability of computer based-products. One important attribute of an interactive system is accessibility meaning that it can be used by people with disabilities [3].

Vision-based interaction, computer input techniques based on a camera and computer vision techniques, is a flexible approach with great potential as computer input method for people with physical disabilities that make standard input devices difficult or impossible to use [4]. To combat social exclusion of people with disability in interaction with technology, the idea to use the facial expression recognition as interaction type arises. The use of computer vision and digital image processing methods, namely, facial expression recognition appears as an alternative to human-computer interaction. The facial expressions are pre-programmed, being attributed a specific function in the computer for each expression made by user. Normally, people that use some system of facial expression recognition for control of computer do not have any kind of mobility. Some diseases that affect total mobility are: neuromuscular degenerative diseases (like as multiple sclerosis and amyotrophic lateral sclerosis for example), Parkinson disease, cerebral palsy, among others [4] [5] .

In the case of people with severe motor impairment, as the case study presented in this work, only some facial muscles can be activated, reducing the communication abilities to "yes" or "no" responses which could be recognized through two different facial expressions.

This paper presents a computer-interaction system particularly conceived for a child with a degenerative neuromuscular disease. The child, in fact, is unable to move the members or the head and he can only use some facial muscles to make two discernible expressions: "smile" and "kiss". Moreover, due to the severity of the disability, what should present itself as a simple expression recognition task is, in fact, a very difficult one. The system uses a standard webcam and computer vision techniques to recognize the user's facial expression and execute the corresponding computer command. It is intended that this system will serve as a support to learning so as to become a help in scholar development for the

DSAI 2016, December 01 - 03, 2016, Vila Real, Portugal

ACM 978-1-4503-4748-8/16/12 $15.00

DOI: http://dx.doi.org/10.1145/3019943.3019944

child. The facial recognition algorithm has been tested and the preliminary results are very positives and promising for the development of a software tool to help the child.

The remainder of this paper is organized as follows. In Section 2, it is presented the proposed algorithm for facial expression recognition. The implementation is present in Section 3. Experiments and results are presented in Section 4. Section 5 provides some final conclusions and directions for future work.

2. RELATED WORK

In recent years many vision-based communication interfaces have been proposed to help users without the ability to use traditional devices for computer interaction, such as the keyboard and the mouse.

The *Camera Mouse* is an interface system that tracks the computer user's movements with a video camera and translates them into the movements of the mouse pointer on the screen. A left-button mouse click can be simulated by hovering with the pointer over the icon to be selected [6]. Manresa-Yee et al. developed a hands-free interface based on computer vision technologies for motor-disabled users who cannot effectively use common input devices. The system works as a pointing device (like a mouse) with a webcam, with nothing attached to the user and with normal background and lighting conditions. The interface transforms the user's nose motion into mouse pointer positions, and a graphical event toolbar handles mouse events [7]. The Facial Mouse is a mouse emulator system based on the facial movement of the user. A webcam is placed in front of the user, focusing on the user's face. A motion extraction algorithm, which is user independent, is used to extract the facial motion from the video. This motion is used to move the mouse pointer that is controlled in a fashion relatively similar to standard mouse devices [3]. FaceMouse is another system that uses a standard webcam and computer vision techniques to track the nose of the person and use this to move the mouse pointer (in accordance with the direction of movement of the nose)[8]. Varona et al. developed a system that allowed people with motor disability to access the computer through user's head movements. The system does not require calibration and automatically detects the face using the Viola and Jones algorithm. Next, the method divides the face into regions: eyes, eyebrows, mouth and nose. 3D Gaussian method is used to detect the skin color region. To determine the eyes and eyebrows regions image threshold is performed. The only facial gesture to take into account is the blink of the eye. The mouse movement is made by the nose position and the eye blink may have different functions. This way, people without movement in the arms or in the trunk can control the computer [9]. In order to develop comfortable and reliable systems for user groups with specific needs, some of them were specially designed and tested in people with cerebral palsy [3] and tetraplegic people [9].

Usually, vision-based interaction systems track some parts of the human body, indeed used to interface with a computer. Most of them assumed that the user can make movements, even if small, with the head, or the hand, or the eyes and this way control the computer. Unfortunately, in some cases of severe physical disability, there is no mobility of the limbs or the head and only certain facial muscles can move, leading that most of existing interface systems are not enough robust to be usable by a seriously disabled user even to take simple actions, such as the choice between "yes" or "no". In cases of extreme movement restrictions, facial expressions are the only alternative to get some interaction with the computer. This paper proposes an algorithm for automatic facial expressions recognition based on the processing of frontal images faces captured by a camera pointed to the user. Figure 1 presents the main blocks of the proposed algorithm.

The case study presented in this work has a severe physical disability, only with mobility in the face muscles. Under these circumstances a specific computer interface based on facial expression recognition was designed to help the child to interact with the computer, taking account his condition and physical restrictions.

3. ALGORITHM PROPOSED FOR FACIAL EXPRESSION RECOGNITION

Facial expression is a demonstration of our affective state, cognitive activity and thought. It transmits a non-verbal communication with other person and it contributes 55% in conversation. It is understood that the mental state of a person can be recognizable by its face expression (in non-verbal communication) [10]. Usually, there are three stages in this type of analysis: face detection, facial feature extraction and expression recognition.

In the first stage of the process, the system detects face image regions in the input image. Face detection can be a difficult task because there are factors such as hair, make up, shave, moustache, glasses or hats, that difficult the location of different facial characteristics. Moreover, scale and the orientation of human face on image, is another major difficulty, which leads to the non-use of fixed models to find the features. The presence of others objects and image noise are also problematic factors that add difficulties to these techniques. Also, in this particular case, as the subject is always with her tongue out of her mouth, another major difficulty, perhaps the biggest, arises. In a second stage, facial characteristics are extracted. These characteristics can be of two types: geometric or permanent characteristics and appearance or transient characteristics. In the last stage of the process, the facial expression is recognized.

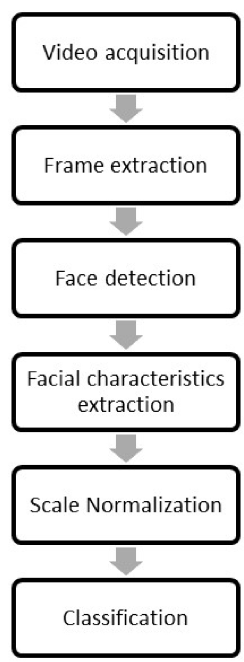

Figure 1 : Diagram of facial expression recognition algorithm

The camera continuously acquires video frames, represented in RGB color space, which are individual processed to recognize the facial expression. Firstly, a face detector should tell whether the image contains a human face and if so, where it is. The face image region is detected using the Viola and Jones algorithm [11].

This algorithm is one of the methods most used to detect human faces and is suitable to process images quickly and efficiently. The algorithm has three major steps:

- **Creation of integral image:** A new image representation is created in a feature space, based on Haar basis functions.
- **Classifier training:** A variant of AdaBoost learning algorithm is used both to select the "best" features and to train the classifier.
- **Construction of a cascade of classifiers:** Early stages of the cascade use simpler classifiers to reject many of the negative sub-windows while more complex classifiers are called to focus on promising regions.

After face detection, the mouth and eyes are extracted, also using Viola and Jones algorithm. Only the left eye is detected to enhance system performance. The eye and mouth are the two facial characteristics used to recognize the expression.

Usually, the dimensions of mouth and eye images extracted are different from the images stored in the training set. A scale normalization procedure is applied in order to obtain images always with the same dimension. The reference eye image dimension has been established in 92x112 pixels and the reference image mouth dimension has been established 145x92 pixels. These dimensions were defined after several tests with different values to obtain the best results.

The mouth and eye appearance varies among images depending on the viewpoint, illumination conditions and the person itself, which difficult the facial expression classification step. In this work, PCA (principal component analysis) method is used to retain the variations in the mouth and eye images at the same time it reduces the dimensionality of the images. The PCA tries to construct a small group of components that summarizes original data, reducing the dimension thereof, only preserving the components more significant. The following steps are needed to perform PCA [12]:

- Calculate the average of input images;
- For each image in the set subtract the average image;
- Calculate the covariance matrix.
- Calculate eigenvectors and eigenvalues of the covariance matrix;
- Retain just the eigenvectors with the biggest eigenvalues (e.g. principal components).

The *EigenMouths* and the *EigenEyes* are obtained by applying the PCA on the training sets of mouth and eye images, respectively.

Finally, given an input facial expression image to be classified, the mouth and eye subimages are projected on the corresponding eigenspaces. The facial expression is recognized by choosing the closest facial expression in the training set, based on a metric of distance. The distance between facial expressions is calculated as the sum of the Euclidian distances for the mouth and the eye projected in the PCA space.

4. IMPLEMENTATION

The main goal of this work is to develop a software tool for Human Computer Interaction based on the recognition of facial expressions, intended to be used by a person with reduced mobility.

The software was specially designed and implemented for a particular disable person that is the case study of this work. The person is a female with 15 years old that has a degenerative neuromuscular disease affecting her muscles in such way that she is incapable of having any kind of mobility in her members. The face muscles are the only ones with mobility. She also has a ventilator to breath. In her quotidian, she frequents classes of special education, speech therapy, occupational therapy, physiotherapy and psychomotricity. In classes and therapies, she has access to computer. In classes she plays didactic games to learn concepts about several themes. In speech therapy she trains communication using the computer and interacting with the gaze. However this interaction isn't viable because the therapist cannot accurately acknowledge the direction of her look, unabling them to correctly determine her response. In occupational therapy some games are conducted in order to promote the child interaction. However, since she is unable to interact with the computer, all these activities become monotonous, stressing and hard to realize, unabling therapists to correctly conduct the therapy plane. This was the major motivation for this project. Additionally, the child rejected commercial products, available for expression recognition, such as the Eye Tracker, because they present themselves as something external to her understanding. For these reasons, facial expression recognition via webcam (without her understanding) can be the ideal solution. In the lessons with the teacher or therapist, she can answer questions with facial expressions, which will cause an action in the computer. The only facial expressions that she does in a simple way are the *smile* and the *kiss*. The software should detect these two expressions to activate two mouse functions, being left mouse click the principal. This decision was taken according with classes activities where the child plays some didactic games, being the left mouse click necessary for the interaction. The system doesn't need calibration what is crucial because calibration operations get the child tired and causes rejection of the technology.

The algorithm for facial expression recognition has been implemented in Matlab environment (2015 version). The face, eyes and mouth regions are extracted from the videos frames using the Viola Jones algorithm. The facial expressions are classified with PCA method by processing mouth and eye sub-images.

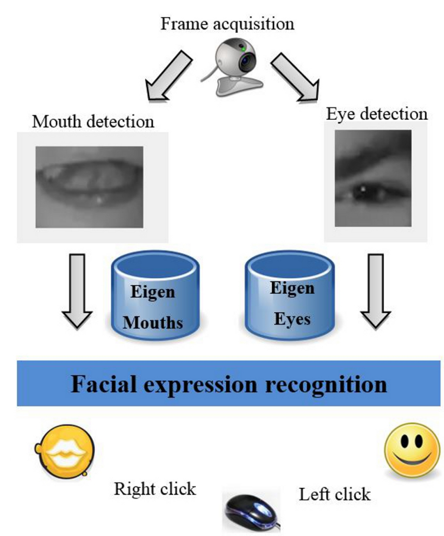

Figure 2: Flowchart of computer program operation

The recognition algorithm was trained to classify facial expression into three categories: *smile*, *kiss* and neutral. For a reliable operation of the application, the classifier must be trained with a training set with a minimum of six images of each expression.

A real time prototype was implemented to enable the interaction with the mouse computer. Figure 2 depicts the principal steps of the computer program operation.

If the facial expression is recognized as smile, a left mouse click is executed, while the right mouse click is performed when kiss expression is recognized.If the algorithm detects the neutral expression or fails to detect the mouth and/or the eye image regions, no command is executed.

5. EXPERIMENTS AND RESULTS

The accuracy and robustness of the expression recognition algorithm was evaluated in a total of 660 frames extracted from 21 videos with facial expressions of the disabled child. The videos were acquired in different days with slight variations in scene illumination conditions using a Smartphone camera pointing to the frontal view of the child. The child was asked to do the three expressions: smile, kiss and neutral. The neutral expression represents all other expressions which are not smile or kiss. Each video has duration of approximately one second, corresponding to 30 frames.

A training set consisting of 30 labelled images of the mouth and the left eye (automatically extracted from the input image with Viola-Jones method and validated by an operator) were used to calculate the EigenEyes and EigenMouths. The testing data set was composed of 630 input images, with 210 samples of each expression. The expression recognition algorithm was applied to the testing data set, counting the number of true positives (TP), true negatives (TN), false positives (TN) and false negatives (FN). These values were used to calculate the algorithm accuracy, sensibility and specificity. The results based on the testing data set of 630 expressions are resumed in Table 1.

Table 1. Results of expression recognition algorithm in the case study.

	#Samples	TP	TN	FP	FN	Accuracy	Sensibility	Specificity
Smile	210	194	397	23	16	89.4%	92.4%	94.5%
Kiss	210	185	396	24	25	88.5%	88.1%	94.3%
Neutral	210	181	418	2	29	98.9%	86.2%	99.5%

The algorithm accuracy is very similar for the smile and kiss expression, being higher (98.90%) in the neutral expression due to the low number of false positives. The smile expression presents the best score for sensibility, with 92.40%, because it registered a small number of false negatives. On the contrary, the neutral expression presents more false negatives, dropping the sensibility to 86.20%. The specificity values in smile and kiss were very similar because both expressions had similar number of true negatives and false positives. The neutral expression attains specificity close to 100% (99.50%) because only two images were wrongly classified as false positive.

The classification performance of the recognition expression algorithm was also described with a confusion matrix, presented in Table 2. The confusion matrix contains information about the real expressions, classified by a human operator, and the predicted expression recognized by the algorithm. The column "Not identified" represents the number of images in which the mouth or the eye were not detected.

Table 2. Confusion matrix in the case study.

Predicted expression / Real expression	Smile	Kiss	Neutral	Not identified
Smile	92.4%	5.2%	0.5%	1.9%
Kiss	11.0%	88.0%	0.5%	0.5%
Neutral	0.0%	6.2%	86.2%	7.6%

The smile expression obtained the best recognition rate (92.4%), then the kiss and lastly the neutral expression. The kiss expression registered the biggest number of errors with 11% of samples wrongly classified as smile expression. This high error may be caused by the fact that the child is always with her tongue out of the mouth, making it difficult to distinguish the two expressions. In 7.6% of neutral expressions the eye or the mouth is not detected, making impossible the expression evaluation.

For performance comparison, the algorithm was also evaluated with images of facial expressions from five persons without disability. Two males and three females with ages between 23 and 53 years old participated in the data collection to build a data set of facial expressions with 2250 input images.

Table 3 shows the results with the image dataset of persons without disability. Compared to the case study, the image dataset of persons without disability showed better results in the metrics accuracy, sensibility and specificity, for the three expressions. The difference in the results can be explained by the fact that the child's tongue is always out of her mouth, making difficult to recognize the expression. This difficulty is not observed in people without disability.

Table 3. Results of expression recognition algorithm in persons without disability.

	# Samples	TP	TN	TP	TN	Accuracy	Sensibility	Specificity
Smile	750	731	1479	21	19	97.2%	97.5%	98.6%
Kiss	750	682	1496	4	68	99.4%	90.9%	99.7%
Neutral	750	727	1498	2	23	99.7%	96.9%	99.9%

Table 4 presents the confusion matrix in persons without disability. Comparing the results, it is observed that the recognition rates are better, for the three expressions, in persons without disability. The better results obtained in people without disability also confirmed the complexity in analyzing the child's expressions. Although this conclusion, the recognition rates are all above 86% in the case study. The smile expression is the expression with the best result for both cases.

Table 4. Confusion matrix in persons without disability.

Predicted expression / Real expression	Smile	Kiss	Neutral	Not identified
Smile	97.5%	0.5%	0.3%	1.7%
Kiss	2.8%	90.9%	0.0%	6.3%
Neutral	0.0%	0.0%	96.9%	3.1%

6. CONCLUSIONS AND FUTURE WORK

Despite the large amount of Human-Computer interaction studies using computer vision techniques for facial expressions recognition, few of them are validated in people with disabilities. Moreover, the particularity of the disease in each subject requires that adjustments be made in each case. As an example, in this work it was necessary to develop a vision-based prototype properly adapted to a child with a degenerative neuromuscular disease, which had the particularity of always having the tongue out of mouth, making difficult the mouth detection and the facial expression recognition.

In the evaluation of expressions recognition algorithm good performance results were obtained, confirming that computer vision-based interaction is effective to help the child from this case study to interact with the computer.

As future work, it is necessary to integrate the interface tool with the applications used by the child in her lessons. The recognition of other facial expressions, beside *smile* and *kiss*, can also be useful to increase the level of interaction with the computer.

7. REFERENCES

[1] World Health Organization, International Classification of Functioning, Disability and Health, Classification, Assessment, Surveys and Terminology Team, Geneva , 2001.

[2] R. Bersch e R. Machado, Atendimento educacional especializado, Brasília: SEESP / SEED / MEC, 2007.

[3] C. Mauri, T. Granollers, J. Lorés, M. García, "Computer vision interaction for people with severe movement restrictions", *Interdisciplinary Journal on Humans in ICT Environments*, vol 2, pp. 38–54, 2006.

[4] Rick Kjeldsen "Improvements in vision-based pointer control" Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2006, Portland, Oregon, USA, October 23-25, 2006.

[5] A. Feinstein, J. Freeman, A. C. Lo, "Treatment of progressive multiple sclerosis: what works, what does not, and what is needed", *Lancet Neurol,* vol. 14, pp. 194–207, 2015.

[6] Betke, M., Gips, J., & Fleming, P. "The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities". IEEE Transactions on Neural Systems and Rehabilitation Engineering, 10(1), 1–10, 2002.

[7] Manresa-Yee, C., Ponsa, P., Varona, J. & Perales, F., "User experience to improve the usability of a vision-based interface". Interacting with Computers, vol. 31, nº. 3, pp. 357-605, 2010.

[8] Perini, E., Soria, S., Prati, A., Cucchiara, R., "Facemouse: a human–computer interface for tetraplegic people". In: ECCV Workshop on HCI. Lecture Notes in Computer Science, vol. 3979, pp. 99–108, 2006.

[9] J. Varona, C. Manresa-Yee, F. Perales, "Hands-free vision-based interface for computer accessibility", *Journal of Network and Computer Applications*, 31, pp. 357–374, 2008.

[10] J. Kumari, R. Rajesh, KM Pooja, "Facial expression recognition: A survey", *Second International Symposium on Computer Vision and the Internet*, pp. 486 – 491, 2015.

[11] P. Viola e M. Jones, "Robust Real-time Face Detection," *International Journal of Computer Vision 57*(2), pp.137–154, 2004.

[12] Renkjumnong, W. "SVD and PCA in Image Processing." Thesis, Georgia State University, 2007. http://scholarworks.gsu.edu/math_theses/31