A Framework for Monitoring the Attention and Detection of Expressions of People with Down syndrome

Rafaela M. de Araujo, Edson D. M. Santos, Lorena R. Lourenço. Juliana S. de Souza Hannum, Talles M. de A. Barbosa, *Member, IEEE*

***Abstract* — This paper proposes a framework that offers support to a user's biofeedback from applications in Android and RemixOS platforms. Its main objective is to enable applications the possibility to provide new functionalities, which are able to monitor the user's attention and expressions. This tool has been developed and evaluated for priority use in children with Down syndrome. They are submitted to rehabilitation activities using computational tools, especially by the use of tablets and personal computers. Through the analysis of data obtained from an embedded camera in a computer system, the framework is able to provide resources that allow the recognition of patterns about behavior and/or emotions captured during the execution of an activity. For this, facial expression and body movement are considered the main variables. At the end, the customization of the proposed activities by psychologists and pedagogues can consider not only the accuracy, but also the behavioral performance captured by the framework.**

I. Introduction

Down syndrome is caused by a chromosomal alteration where instead of a cell having 23 pairs of chromosomes, it has an extra chromosome precisely at pair 21, and hence this anomaly is known as trisomy of chromosome 21. Individuals with Down syndrome experience a delay in the process of intellectual, physical, and mental development having severe to mild intellectual disability [1]. Such limitations end up creating a barrier in their learning. However, it is possible to reach considerable progress with the appropriate stimulation, which can be facilitated by trained professionals and family members.

In PUC Goiás, an extension project called Alfadown [2] is being developed. Its main objective is to assist children and adolescents in the acquisition of language and literacy, with the help of computational tools. This research is authorized by the Research Ethics Committee (CAAE – 32702 114.2.0000.0037). Several hardware and software prototypes have been developed, and they can automatically assist, monitor, and evaluate rehabilitation activities executed using a computer. In addition, the prototypes have purpose-built applications to automate and amplify stimuli in specific activities.

Rafaela M. de Araujo, Edson D. M. Santos, Lorena R. Lourenço. Juliana S. de Souza Hannum, Talles M. de A. Barbosa are with the Department of Computing and Departament of Psychology, Pontifical Catholic University of Goiás, Goiânia, GO, Brazil (phone:+55-6239461381,email:rafalea.campos.araujo@gmail.com,walawbites@gmail.com,talles.ucg@gmal.com) .

II. Related works

Currently, there are educational software that provide resources in assisting activities of pedagogical content that helps the cognitive development and functional abilities of atypical children, such as Livox [3]. Livox allows people with certain disabilities that impair speech to communicate and learn through the use of a tablet.

Including biofeedback resources to aid the analysis of the individual during activities proposed for educational means can become a great differential to these applications. Studies about face recognition and its analysis have an important role in the natural interaction between man and machine, which is one of the many different types of non-verbal communication interpretations. A software, proposed by Jackie Lee [4], is able to recognize faces and monitor their attention through the use of a webcam, providing features that allow graphic designers to quickly create interactive spaces.

Another example of the use of biofeedback is the work developed by Santos [5], which proposes the development of a software for mobile devices that support people with Retinitis Pigmentosa. Retinitis Pigmentosa causes the loss of peripheral vision in several stages, and the software features camera zoom, detection of objects, and monitoring of attention.

III. Materials and Methods

A. Framework Requirements

This project was developed for the Android platform devices using the Android SDK and features available in the Native Development Kit (NDK). Java and C++ were used as programming languages, while NDK allowed the implementation of parts of the framework to run in native code languages such as C and C++.

In order to work with image manipulation, the Open Source Computer Vision (OpenCV) multiplatform library was used. OpenCV is released on a BSD license, free for academic and commercial use. OpenCV implements a variety of image interpretation tools, ranging from simple operations, like a noise filter, to more complex operations, such as motion analysis, pattern recognition, and 3D reconstruction.

The framework provides support for Android platform devices, such as tablets, mobile phones, and computers using RemixOS [6]. RemixOS is a software capable of running the native Android operating system on home computers, desktop and notebooks. The devices in use need to have an Android version newer than or equal to 4.0.0, also known as Ice Cream Sandwich.

The input is mainly the image stream obtained in real time through the camera sensor of the mobile device. The camera must be properly positioned so that the targeted individual can almost directly look at it. The framework is also capable of analyzing videos or static images provided to the software.

B. Structural Description

A framework is a set of classes that constitutes an abstract project for the solution to a family of problems [7]. The Framework Structure can be seen in the class diagram shown in Fig. 1, which contains information about the methods, attributes, function names, and how they were integrated into the system.

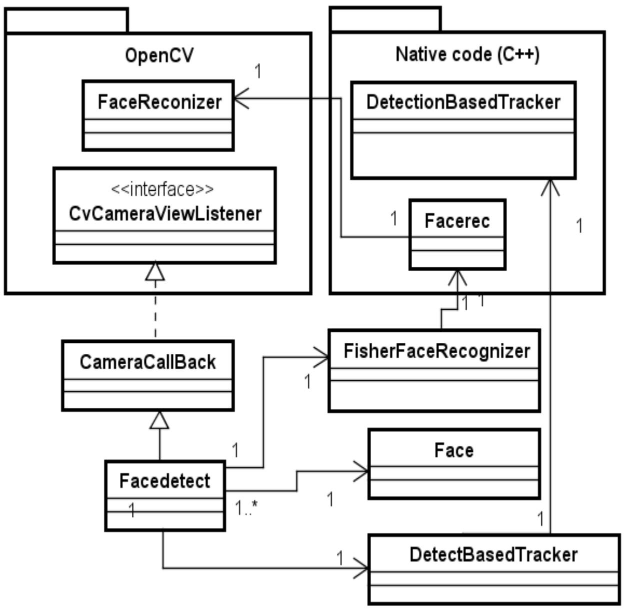


Figure 1. Class Diagram that describes the structure of proposed Framework.

The abstract class CameraCallBack defines an application template enforcing the classes that extend it to implement their methods and the contracts defined in the CvCameraViewListener interface. These methods are responsible for configuring and retrieving the data from the device's camera. Likewise, if it is necessary to create a new class to receive data from the camera, the class CameraCallBack just need to be extended, the same way it is done in the class FaceDetect, which is responsible to identify the face, set the attention level and identify the user's emotion.

In this relationship, the concepts of inheritance and polymorphism from object-oriented programming are applied. These concepts allow attributes and methods to be shared by classes and implemented in their own way, making it easier to insert new features that also analyze the data obtained from the camera.

C. Attention Monitoring

The attention meter developed at the MIT Media Lab [3] aims to measure a user's attention level. It has been adapted for the android platform and included in the composition of the framework. The software is able to identify a person and starts to measure the attention level of the individual, as long as they look towards the camera. Each frame obtained either by using a saved video or in real time through the camera, a face detection algorithm described below is applied in order to determine the level of attention.

In order to determine the attention level, the software needs to first identify a face present in each frame. The face detection method proposed by Paul Viola and Michael Jones [8] present in the OpenCV library is used. This method utilizes a detection technique based on the appearance of the object that when used together with some key features allows a fast and robust implementation of facial detection. These methods tend to learn features from training with positive images (images with faces) and negative images (images without faces), to later detect objects in other images.

Using facial knowledge and looking for distinctive brightness gradients of the eye the attention meter can quickly find, detect eyes, and measure the intermittent rate of a face over several frames. Each identified face is given a rating for attention that varies over time. It starts at 0 and increases according to the attention exhibited by the individual.

Combining affection, emotion, movement patterns, small head gestures, and the score of attention in various ways, will allow us to determine high-level activities about the relationship between individuals and the target of attention. In this way, the attention meter can go beyond simply measuring attention, it can describe the relationships that individuals have with the target and other individuals.

Fig. 2 shows the flowchart that describes the operation of the attention monitor and Fig. 3 shows the face identification and the attention meter applied in the software to monitor real time attention of those looking towards the camera.

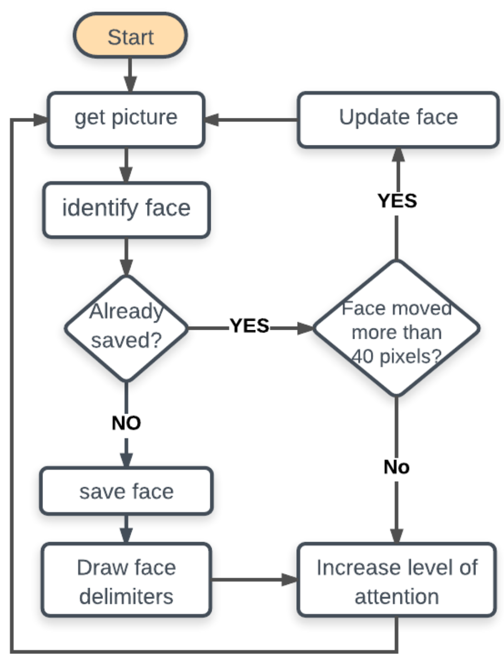


Figure 2. Flowchart that describes each step in the attention meter algorithm.

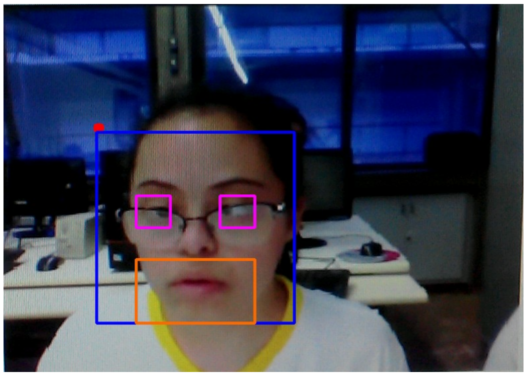


Figure 3. Shows the face delimiters drawn from the data received from the OpenCV classifier. The red bar above the face delimiter represents the current level of attention of the individual.

A. Expression Recognition

The Fisherfaces algorithm available in the OpenCV library was used in the system so it could learn to recognize expression patterns, which uses ideal machine learning for facial recognition. In Fisherfaces, all pixels of a face, or the image that contains a face is utilized as input of the recognition system [9]. Thus, all images information can be considered in this approach despite having a disadvantage of the high dimensional data increasing computational costs. To solve this problem, the statistical method Principal Component Analysis (PCA) is used in order to have a Dimensionality Reduction.

The creation of the dataset is a very important factor in order to be successful in the recognition process. In this step, one or more face images are grouped and encoded for use in the model forming the face-space. After selecting a face and establishing a similarity of it with the dataset, the algorithm looks for features that define the face.

In this project, the dataset of images was obtained from videos collected with the help of the Alfadown researchers. The videos were recorded during the rehabilitation activities performed by the group. From the analysis of these videos, it was possible to obtain images that contained certain expressions of emotions. It was important to guarantee that the videos were recorded while the individuals performed familiar activities and suffered no interruption during it.

Expression recognition is a delicate process that is vulnerable to small changes in lighting condition. To help minimize these problems and increase its accuracy, the images are automatically adjusted before applying the algorithm. The methods applied are conversion to grayscale, histogram equalization, and bilateral filter.

Face detection has a better performance on grayscale images, thus each image obtained is converted to grayscale. Generally, grayscale images are more robust to variations in lighting [10].

An image histogram contains the information on the number of times the hue of a color is repeated. The histogram equalization is used to adjust the intensity of the image improving its brightness and contrast, useful for getting more details in images with backgrounds and foregrounds that can be both bright or both dark [11].

The bilateral filter helps to smooth the lower contrast areas of the image without affecting the areas of higher contrast, reducing the noise present in the image while preserving its contours.

Figure 4 shows an original image taken from the dataset, and the result after multiple adjustments has been made.

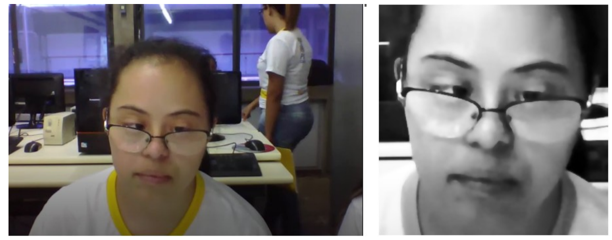


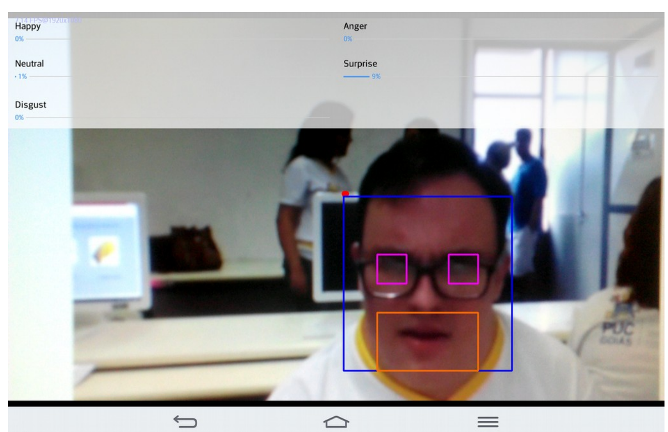
Figure 4. Before and after image of the adjustments made to increase the accuracy of the emotion identification algorithm.

For the algorithm to begin the identification of expressions, it needs to be trained with a set of face images created from the Alfadown videos. The training set helps to establish a knowledge base with information that will be used for comparison between input faces.

The system loads images with a number that represents which emotion it belongs to. Then the input image is passed to the software where the identification is made. The input face is compared to the other images in the database. Fisherfaces creates a low-level representation of the face, result from the linear discriminant analysis, over the projection of the image in the sub-space created by the PCA. Because of this comparison, the method returns a number that represents which emotion the face is associated with.

Figure 5 shows the emotion recognition applied in a video obtained by Alfadown during the rehabilitation activities.

Figure 5. Emotion recognition.



IV. RESULTS

In order to verify if the algorithm is able to recognize emotions correctly in people with Down syndrome besides the database created from the videos of Alfadown, to perform the tests and to compare the obtained results, database was used with typical people Cohn-Kanade (CK +) [12], created with the aim of promoting the investigation of facial expressions.

The Alfadown database has 20 images to represent each of the 4 emotions, they are neutral, happiness, anger and surprise, a total of 80 images. To perform the tests, the same amount of images of the Cohn-Kanede database and the same four emotions were used. In all images of both bases the adjustments described in section 3.4.2.

Initially algorithm was training with only 80% of the images of the database 64 images, of these being 16 of each of the four emotions. After training, the other 20% of the database, 16 images, 4 of each emotion, was used to test the efficiency of the algorithm.

Using the database with people with Down syndrome, the algorithm was able to correctly recognize 10 images and miss other 6, obtaining a success of 62.5%, and with the Cohn-Kanede database of typical people the algorithm was successful of 87.5%, of 16 images erroneous. Table I and Table II show the confusion matrix that summarizes the results of the algorithm tests for this analysis. All correct guesses are located on the main diagonal of the table.

TABLE I. MATRIX OF CONFUSION OF RESULTS OBTAINED USING THE ALFADOWN DATABASE

		Observed				
		Neutral	Happy	Anger	Surprise	ACC
Predicted	Neutral	2	0	2	0	0,50
	Happy	1	3	0	0	0,75
	Anger	2	0	2	0	0,50
	Surprise	0	0	1	3	0,75
	REL	0,40	1,00	0,40	1,00	

TABLE II. MATRIX OF CONFUSION OF RESULTS OBTAINED USING THE COHN-KANEDE DATABASE

		Observed				
		Neutral	Happy	Anger	Surprise	ACC
Predicted	Neutral	3	0	1	0	0,75
	Happy	0	4	0	0	1,00
	Anger	0	0	4	0	1,00
	Surprise	1	0	0	3	0,75
	REL	0,75	1,00	0,80	1,00	

The numbers in the Accuracy (ACC) column show the accuracy of the classification, the images are correctly classified according to the class that belongs.

The figures in the Reliability (REL) line show the reliability of the classes in the classified image. It is the fraction of correctly classified images in relation to all images classified with this class.

V. CONCLUSION

Children with Down syndrome tend to have an easier time learning things through observation. This means that they learn better when looking at and copying others, finding it easier to understand information with the help of illustrations, gestures, or objects that they can see. The use of the attention feature proved very useful in activities that required greater observation from the individual, helping to analyze their involvement with the activities and present individuals who may be present during the activity, allowing us to observe how it affects their performance.

The specialists will be able to perform analysis of the behavior of the child during pedagogical activities by obtaining the variations of emotion through the analysis of the expressions, combined with the level of attention. This analysis makes it possible to identify behaviors, such as movement that can imply reading, long periods of open mouths, shaking of the head, a pattern of movement that indicates that someone was impressed or even bored, as well as how long a person smiles implying that the context or target was interesting.

ACKNOWLEDGMENT

This study was supported by FAPEG (in Portuguese, Fundação de Amparo à Pesquisa de Goiás) at Pontifical Catholic University of Goiás.

REFERENCES

[1] PUESCHEL, Siegried. Síndrome de Down: guia para pais e educadores. (1995). Tradução Lúcia Helena Reily. São Paulo, Ed. Papirus, pág. 105.

[2] Web/Pucgoias Projeto Alfadown, “http://sites.pucgoias.edu.br/extensao/blog/tag/alfadown/”. October 18, 2016

[3] Carlos Edmar Pereira, Livox. Tecnologias de Inclusão Social Ltda “*http://www.livox.com.br/quemsomos*”, october 7,2016.

[4] LEE, Chia-Hsun Jackie et al. Attention meter: a vision-based input toolkit for interaction designers. In: CHI'06 extended abstracts on Human factors in computing systems. ACM, 2006.

[5] Santos, Edson DM, et al. A Mobile Context-Aware System to Support People with the Retinitis Pigmentosa Disease. 2015

[6] Jide, Remix OS. “https://developer.android.com/studio/index.html” Acesso 16 de agosto de 2016.

[7] Sommerville, Ian. Engenharia de Software. Tradução Selma Shin Shimizu Melnikoff; Reginaldo Arakaki; Edilson de Andrade Barbosa. 8. ed. São Paulo: Addison Wesley

[8] Viola, P., and Jones Michael. "Robust Real-time Object Detection." *Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling*. 2004.

[9] Belhumeur, Peter N.; Hespanha, João P.; KRIEGMAN, David J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, v. 19, n. 7, p. 711-720, 1997.

[10] I Tuytelaars T. Mikolajczyk K. Local Invariant Feature Detectors: A Survey, Foundations and Trends in Computer Graphics and Vision Vol. 3, No. 3. 2007

[11] Figueredo, M. B. and Rodrigues de Souza, J. Face recognition model applied to the missing people problem. In Information Systems and Technologies (CISTI), 2013 *8th Iberian Conference on, pages 1–3. IEEE, 2013*.

[12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the Computer Vision and Pattern Recognition Workshops (CVPRW 2010), pages 94–101, 2010. 4