AN ARTIFICIAL VISION – BASED COMPUTER INTERFACE

INTERFEJS KOMPUTEROWY OPARTY NA SZTUCZNYM WIDZENIU

Ángel Gil¹, Miguel Márquez², Erica Chacón³, Angélica Ramirez⁴

(1,2) Universidad Nacional Experimental del Táchira. Decanato de Investigación.
Laboratorio de Prototipos. Av. Universidad. Paramillo. San Cristóbal – Edo.
Táchira 5001 Venezuela.

E-mails: (1) agil@unet.edu.ve (2) mmarquez@unet.edu.ve

Abstract: An application has been developed to assist in using software applications to individuals that have null or limited mobility of their upper limbs. The system uses a Web camera for movement patterns recognition of the user's face. The image analysis allows emulating the basic functions of the computer mouse. The face detection process is carried out through the implementation of an algorithm that uses a cascade sort key of Haar-Like type. The application was developed in C++ to be used on Windows XP platform. Tryout of the application has been performed showing excellent acceptance and short time training requirements.

Keywords: artificial vision, mouse, disability, handicapped

Streszczenie: Opracowano oprogramowanie aplikacyjne pomagające korzystać z oprogramowania komputerowego osobom, które mają zerową lub ograniczoną ruchomość ich górnych kończyn. Taki system wykorzystuje kamerę internetową do rozpoznawania wzorców przemieszczeń ruchów twarzy użytkownika. Analiza obrazów pozwala na emulację podstawowych funkcji myszy komputerowej. Proces detekcji ruchów twarzy jest realizowany poprzez implementację algorytmu wykorzystującego kaskadowy klucz sortowania typu Haara. Oprogramowanie aplikacyjne zostało napisane w języku C++ i może być używane na platformie Windows XP. Zostały przeprowadzone próby oprogramowania, spotkały się one z doskonałym przyjęciem i akceptacją, gdyż używanie tego oprogramowania wymaga jedynie krótkiego treningu.

Słowa kluczowe: Sztuczne widzenie, myszka, niepełnosprawność, kalectwo

1. Introduction

Nowadays the personal computers have become a tool of daily use in different activities of the society; such they have diverse devices that allow the interaction with the users, among them appears the peripheral used for data input, being the keyboard and the mouse the most popular. The interaction with these peripheral requires of a minimum physical user capability to obtain an efficient use; nevertheless a 10% of the world-wide population present/display at least some kind of motor disability [1], this makes difficult the interaction between them and devices such like the mouse, hence the use of the computers; it is for that reason that arises the necessity to developing alternative means that facilitate the interaction with computers. Artificial vision offers the necessary tools for implementing one of these alternatives by means of analyzing captured images using a camera Web. Through the analysis of the user face gestures in front of the camera the action of the basic operations of the mouse will be emulated by the system.

2. General Description

2.1. General diagram of the System

Figure 1 presents a diagram showing the general operation description of the application, considering the main processes of the developed application.

The application uses a Web camera to capture the scene located in front of the computer, where, in first instance, it searches for a face. Once a face is located in the scene, then it searches for a color pattern, that must be located in the face identified in the scene. This research work uses a green rectangular pattern located in the forehead of the individual. This pattern will serve as target to be followed by the image processing module of the system and allows the user to locate the cursor in the screen by his/her head movements. Another basic function to emulate is the click of the mouse, the application constantly reviews the state of the mouth of the user, that is to say, whether it is open or closed. The system uses the frequency with which the user opens and closes the mouth to emulate the pressing and releasing of the left push button of the mouse. If the open mouth image fulfills certain amount of photogram is assumed that the user wishes to maintain pressed the left push button, to conduct some action such as to drag a folder.

Fig. 1. General diagram of the application

2.2. Description of the Application

Used Tools: The Extreme Programming or XP is used, which offers a set of techniques that conforms a simple methodology for software development ^[2], also C++ Language was used along with the libraries of open code for the treatment of OpenCV images (Open Source Computer Vision) ^[3] and IPL of Intel®, both of them including a series of very useful functions for the development of this kind of applications. The processes of the application were modeled using the Unified Modeled Language (UML).

User Face Detection: The application constantly captures the images of the scene that is in front of the computer screen and in each images it search for a face. This process of face detection is made through the implementation of an algorithm that uses a cascade sort key of Haar-Like type ^[4], this sort key looks for the characteristics of the face in a horizontal way, reason why its single use does not guarantee the detection of the user face, due to the fact that the face can be inclined. Therefore, if the algorithm does not detect a face in first instance, automatically rotates the image 10° degrees clockwise and search the new image, until completing 90°. If the face is not even detected, is come to take the original image, and applies the same procedure, but this time counter clockwise.

The algorithm is able to detect more than one face; reason why it is recommended that within the viewing angle of the camera only remains the application user. Once the face is detected, a circular bit map trimmed image of the zone where the face was detected is stored for later analyses (Figure 2).

Fig. 2. Face Detection

Detection of the pattern to be followed: After the procedure of detection of the face is finished, the system searches in the face image stored to identify the colored pattern that will serve as target to move the cursor through the computer screen. A green rectangular pattern is used, which is recommended to be located in the user's forehead. The green color was selected due to its little presence within the range of colors that can be usually found in the human face (Figure 3).

An algorithm to analyze the G component in the RGB space of color was developed to carry out the detection of this pattern. The algorithm crosses the image pixel by pixel extracting of each one its components RGB with the intention of verifying if they fulfill the following conditions: $G > R$, $G > B$, $(G - R) = 20$ and $(G - B) = 20$.

In case a pixel satisfies the mentioned conditions, then that pixel is considered as part of the target, this way it is extracted all pixels belonging to the rectangular pattern, and then to obtain its space coordinates (Figure 4).

Fig. 3. Pattern to follow

Fig. 4. Extraction of the pattern

Next step is to identify the coordinates of the central point of the pattern, which will be used to translate the position of the image to any position in the computer screen. If the individual's face is rotated an α angle during the process of face identification, then a rotation adjustment must be done as follows. Assuming that the initial coordinates of the pattern central point are $P(X1, Y1)$, and α the rotation angle of the original image, then the original position of the central point is given by:

$$X1 = r * \cos(\alpha) \quad (1)$$

$$Y1 = r * \sin(\alpha) \quad (2)$$

Movement of the Cursor: The movement of the cursor is given by the translation of the central point of the pattern using screen coordinates. Previous to this calculation a mirror transformation is applied to the captured image so there is not confusion between right and left to the user.

In order to manage an efficient movement of the cursor, and allow the user to position elements in the screen, a weighted average of the last five (5) movements carried out by the cursor was used. This procedure takes the last 5 locations from the cursor and calculates a weighted position. The weight assignment gives preference to the most recent position. The calculation of the cursor weighted position is based on the use of a vector of 5 weights generated as follows ^[5]:

For all $0 \leq i \leq N$, where N is the amount of positions to be considered,

$$Weight_i = \frac{1}{1+i} \quad (3)$$

1. It is necessary standardize the components of the vector “ $Weight_i$ ” to the interval [0,1]

$$\forall 0 \leq i \leq N, Weight_i = \frac{Weight_i}{\sum_{j=0}^{j=N-1} Weight_j} \quad (4)$$

2. The positions are stored in a vector, and proceed to calculate the next position where the cursor will be located.

$$Final_Position_{(x,y)} = \sum_{i=0}^{i=N-1} movement_{i(x,y)} * Weight_i \quad (5)$$

This way it is managed to optimize the movement of the cursor given a smooth movement in the screen.

Execution of the left click: The actions to press and to release the left click are carried out by the user through the mouth opening or mouth closing action. Therefore, an algorithm was developed that allows detecting this action, as well as the time the mouth remains open. Then by using API functions of Windows® is possible to execute the actions required by the user in the system. This process is carried out by means of the detection of two ellipses (Figure 5), which represent the outside border and the inside border of the lips, which are searched in inferior half of the image previously filtered using a Gaussian smoothed operation ^[6].

Fig. 5. Detection of the open mouth

The found ellipses must fulfill certain conditions as far as their dimensions, with the intention of not confusing the mouth with any other shade or form that can be wrongly detected by the algorithm.

Table 1. Verification of states of the user's mouth

Present State	Previous State	Mouse Action
Open Mouth	Closed Mouth	Press Left Button
Open Mouth	Open Mouth	To position Cursor
Open Mouth	Closed Mouth	To position Cursor
Open Mouth	Open Mouth	Release Left Button

With the intention of optimizing the execution of the click operation, before making any decision regarding the action to make, the algorithm allows verifying certain states of the user's mouth and according to this result it conducts the action. (Table 1)

Additionally, a timer for the action of the click is used that works of the following way:

1. When an opened mouth is detected, the movement of the cursor stops during 5 photogram, this way it is possible to be determined if the user wishes to conduct the clicking action or additionally he wishes to drag an object.
2. If when passing the 5 photogram, the mouth is closed, it is come to release the click, in the original position where it was pressed.
3. If when finalizing the 5 photogram the mouth follows open, it stays pressed the click, to execute the dragging action of the object selected by the user.

3. Results

The developed application is of type Windows GUI ^[7], based on a *Callback* type, that supports the management of messages sent by the operating system, which allows to take care of the requirements of the user. The application offers the possibility of moving the cursor all over the computer screen, through movements of the head, as well as pressing or releasing the left click, with a single mouth opening or closing. This way action such as opening a program or dragging a file can be made. The application gives the possibility of two levels of movement sensitivity, depending on the user's capabilities. Also it allows the constant monitoring of the captured images and processing by a window located in the screen, where the activity of the user is shown in real time (Figure 6).

Fig. 6. Window of monitoring

Windows XP® offers different options to improve the accessibility to the system including; for example, a keyboard in screen. These tools altogether with the application developed in this work allow the user to expand the list of tasks that can perform, with no need of using traditional input data peripherals.

4. Conclusion

An alternative to traditional input data peripherals was developed using artificial vision. The application is directed specially to individuals that present/display some motor disability that affects the mobility of their upper members.

The application counts with a simple interface, which does not demand prolonged training for its use. Also, the system requires only a camera Web for its operation, reason why it represents an economic alternative.

The developed algorithm is able to process up to 20 images per second, which offers to the user a speed of use near to a conventional mouse.

The use of the method of weighted position average, allowed a smooth movement of the cursor on the screen making a more exact movement of the cursor, as well as offering the possibility of establishing two levels of movement sensitivity, which facilitates the use of the application.

References

1. Anzola, Rosario. *Fundación Paso a Paso: "Un mundo con acceso para todos"*. Venezuela. 2000.
2. Dapena P., José. *Metodologías para Desarrollo en Comunidad. IV Seminary the Operating System Linux*, 2004.
3. Intel. *Open Source Computer Vision Library Reference Manual*. Intel Corporation. U.S.A. 2001.
4. Viola P. y Jones M. Robust Real-Time Face Detection. *International Journal of Computer Vision*, pp. 137-154, 2003.
5. Coianiz T. y Torresani L. 2D Deformable Models for Visual Speech Analysis. *Speechreading by Humans and Machines*, pp. 391-398, 1996.
6. Pajares G., de la Cruz J., Molina J., Cuadrado J. y López A. *Imágenes Digitales. Procesamiento práctico con Java*. Alfaomega, México. 2003.
7. Microsoft Corporation. *Windows API*. U.S.A. 2007.

Ángel Gil is Engineer in Computer Science. Investigator assigned to Universidad Nacional Experimental del Táchira's Prototyping Research Laboratory. Second year doctorate student in Mecatronics Engineering of the University of Malaga – Spain. He has special research interest in Robotics.

Doctor Miguel Márquez is a Mechanical Engineer, Head of the Universidad Nacional Experimental del Táchira's Prototyping Research Laboratory. He is Lecturer in Ergonomics, Design and Robotics. Also, is Founder member of the Venezuelan Society for Ergonomics and Occupational Health Research (UVIERSO) and Board member of the International Society for Productivity Enhancement (ISPE).