Development of model for automatic tutor in e-learning environment based on student reactions extraction using facial recognition

Abstract:

In e-learning environment, the main function of the tutor is to provide assistance to one or a group of learners. We distinguish between two types of tutors: The reactive tutor who responds simply to the explicit requests of the learner and the proactive tutor who takes the initiative to provide support to the learner and offer him advice at the appropriate time. Our goal is to create an environment where the proactive role of tutor is to assist and monitor a single learner during the learning process. In our case tutor should have his own problem-solving analysis, his own ability to model the learner and his own expertise in generating explanations. In this paper, an automatic tutor based on face recognition and emotion recognition system is proposed for a distant learning platform, the first step for face recognition system is face detection for this we are using the viola jones object detect algorithm that can provide the detection in real time. Features extraction and tracking is the base of emotion detection for that the KLT algorithm will be applied after face detection. These algorithms will be performed in Matlab as programming tool focused in image processing and facial recognition. This proposed method is used to create an automatic tutor that helps assisting the learner during an educational process in e-learning. The name of this Facial and Audio Recognition Automatic Tutoring Model is FARAT.

SECTION I.

Introduction

The human face represents an important challenge faced in the last few years in the field of machine recognition and image processing. The facial recognition and identification becomes the main research fields as it's applied in many areas such as security, teleconferencing, etc.

In 2013 J. Joseph and K.P. Zacharia used the face recognition to create an Automatic Attendance Management System using PCA for the identification and a PIC card to control the camera [1]. For A. Kumar and M. Bansal facial and voice recognition system is created based on biometrics techniques, the facial recognition are configured by MATLAB as a programming tool that has a toolbox to process images [2].

Recognition algorithm can be divided into two main approaches, local approaches, based on the face geometry, the local features are extracted and the face is represented by vectors [3]. Global approaches identify the face using the integral image of the face as input for the recognition system [4].

Audio-visual recognition systems fall into three broad categories according to their differences in speech processing and analysis [5]:

- Follow the learner's region of the learner and extract visual attributes;
- Produce a decision when the word pronounced by the assembly of the auditory and visual attributes;
- Recognition of single words or continuous words.

The model we propose comes in the framework of cooperation in the automation of tasks of the tutor. In this case our problem is how to help this tool to make the right decision to a precise learning situation and an instantaneous way to guide and assist the learner. Our main goal is to use facial recognition in the e-learning domain by creating an automatic tutor that can help and assist the Learner in the learning process by guiding him depending of his grimaces during the course. For that we are studying the facial features extracted and comparing it to the emotion database. The name of this Facial and Audio Recognition for Automatic Tutoring Model is FARAT.

SECTION II.

The Developpment of The Model for Automatic Tutor in e-Learning Envirement Proposed To Use The Facial and Vocal Recognition of Student's Emotions and Reactions

A. The e-Learning Platform

This distance tutoring system connects five main actors: the Tutor, the Teacher, the Learner, the Administrator and multimedia creator [6]. These actors are connected through a distance learning action whose model is based on the following components:

- Technical choice: These resources help to assist the learner in his distance training, to accomplish their mission they use tools and technical resources such as:
    - Communication tools: Both synchronous (chat, telephony via the net, audio-conference or videoconference, etc.) as well as asynchronous (e-mail, discussion forum, etc…).
    - Remote collaboration tools: Although some can also be used in synchronous (interactive whiteboard), remote collaboration tools are essentially asynchronous (shared workspace and editing tools, such as blog, vlog and other wiki, portfolio, interactive whiteboard, Etc.). Communication tools can also enable or facilitate collaborative work.
    - Tools for filing or sharing documents: The tools of deposit or sharing are essentially asynchronous (space of exchanges and portfolio). However, most communication tools (forum, chat) and collaboration (wiki, blog) also allow sharing documents.
    - Tools for building a learning community: Building a learning community is one of the major learning objectives in e-Learning. It improves the quality of learning, the effectiveness of training and the performance of learners.
- Resources: In order to ensure that the educational task of the e-Learning system functions properly and to facilitate teaching for all the players in this learning system, several technical and educational resources as well as multimedia communication tools are used. For hat the system traits with several educational resources such as:
    - Courses: The tutor can write a course directly on the platform, add categories, edit or delete them. It also has permission to manage the accessibility of students to this course.

- Documents: There are several types of learning materials that can be found in an e-learning learning platform, such as PDF, Word, PowerPoint or EXEL files, which can be downloaded by the student.
- Digital resources: The tutor can use several types of digital resources such as video or image to help students or learners better understand the courses.
- Links: We can also find quick links or hyperlinks in distance learning platforms.
- Documents available on other media (CD, paper) if needed.
- Technological infrastructure (bandwidth, Video reception equipment).
- Pedagogical choice: Choice of the pedagogical model, the course support (Face-to-face or remote). To read educational resources, the e-Learning system uses communication tools that are the tools that combine sound and image by enabling the creation and development of animations and the optimization of design.
- Multimedia reading tools are the tools that associate the sound and the image allowing the creation and the development of the animations as well as the optimization of design. Mainly a broadcast channel is needed for online learning such as Mozilla, Opera, and Google Chrome, etc. There are several open source tools on the internet that help create, present and disseminate multimedia content such as Macromedia Flash (presentation of animations on websites), Wideo (creation of videos), WebEx (videoconference), etc.

B. The Facial and Audio Recognition Automatic Tutor : Farat

The platform environment used in this system is MOODLE. In this e-learning environment, the Tutor is a main player in the learning process. In this paper we will show the ability to design an intelligent and automatic tutor designed to control and monitor the learner as per a workflow of tasks given by inputs such as video and audio.

For that our main components are the camera for the video stream and the microphone for the voice recording. The recognition of the face, the Lips reading and the reaction study is processed by image processing toolbox on MATLAB. For the voice recording and recognition the speech and audio processing toolbox is used in the MATLAB.

A workflow is designed to be a management tool improving learning process by coordinating resources and actors and supervising the progress of the various tasks especially the tutor [7]. The decision made by the tutor to guide the learner is a result of the process of all the tasks given by the *Workflow E-learning Process Action* WEPA. Thereafter we give some tasks performed by WEPA:

- Analysis and Treatment of data returned by the *Graphical Interactive Student Monitoring* Tool for Moodle or GISMO [8].
- Analysis and Treatment of the data returned by MATLAB;
- Capture of reaction of learner;
- Video processing analyzing of captured transactions;

- Definition of the output layer and send it to the tutor as a result of the process.

On table 1 we represent examples of the inputs needed for the tractability of the student during the learning process.

Table I Examples of input data in wepa

Input data	Form of data
The result of the entrance test	BOOLEAN
The access time to the same course	TIME
The number of learner requests	NUMBER
The number of messages received	NUMBER
The number of access to one resource	NUMBER
The number of resources accessed	NUMBER
The number of access to the same course	NUMBER

The process processes the data of the workflow given by GISMO on the one hand and by MATLAB on the other hand. The result of this process is decision-making by the tutor who is consulted by a learning situation demanded by the learner [9].

On Fig. 1 we represent the proposed model for automatic tutor in e-learning environment using the facial recognition, emotion and reaction of the student. In order to create an automatic tutor that assists the student in the e-learning process, an image of the learner is sent from a camera that is configured by a Matlab image toolbox.

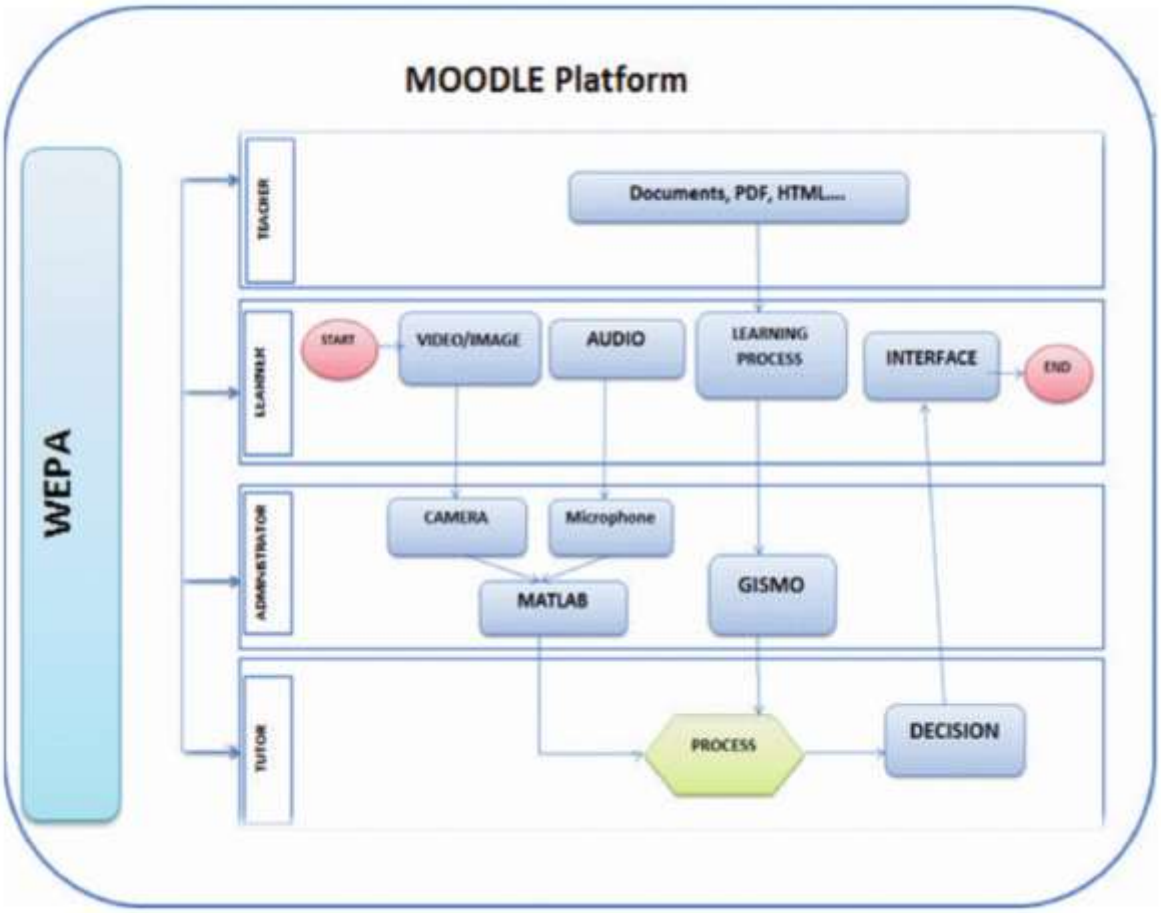

Fig. 1. The FARAT system components.

The traceability tool in the system returns information about the student activity during the learning process and since his enrollment in the system. It returns data to be completed and progresses in the tutoring process. This input layer of WEPA contains the following types of information.

SECTION III.

Facial Recognition Techniques for the Developpement of Farat Model

The researchers have developed a multitude of facial recognition techniques ranging from the simple detection of the face to the precise location of the characteristic regions of the face such as the eyes, the nose, the mouth, etc.

There are two techniques of facial recognition: 2D technique and 3D technique. Several approaches based on the combination of 2D and 3D face data have been developed recently to improve the performance of recognition.

A. Facial Recognition Using 2d Technique

The local approaches for the 2D techniques of facial recognition are based on models and on a separate treatment of the different landmarks of the image of a face. These approaches are based

on an extraction of the local facial features whose faces are represented by vectors, in this type of approach there are several techniques of analysis of the information based on the face geometry, the graphs (Elastic Bunch Graph Matching EBGM) or the appearance local [10].

The Global approaches for the 2D techniques identify a face using the integral image of the face as input to the recognition system. Each face image of dimension (*n, m*) is represented by a simple vector of dimension *n * m*, concatenating the values of the gray level of all the pixels of the face image. The set of image vectors of the face is called “image space”. From this approaches 2 algorithms are the most used for facial recognition:

- Principal Component analysis PCA;
- Linear discriminant analysis LDA.

Hybrid approach: Several approaches have been proposed for visual recognition, except that none of them is capable of adapting to the changing environment (pose, facial expression, luminosity, etc.). The robustness of a recognition system can be increased by merging several approaches. Hybrid approaches are the result of the fusion of local and global approaches. Local and global characteristics have a different sensitivity to variation factors. For example, the change in illumination affects local characteristics while facial expression changes have more impact on overall characteristics.

On Table 2 are shows the sensitivity of local and global characteristics to the variation factors.

Table II: The sensitivity of the features to the factors of variation

Factors of variation	Local Features	Global Features
illumination	Very sensitive	sensitive
Facial expression	Not sensitive	sensitive
Pose	sensitive	Very sensitive
Noise	Very sensitive	sensitive

B. Facial Recognition Using 3d Technique

The simplest method of a 3D representation is the polygon mesh (a list of points connected by polygon edges). The data on which the researchers work exist in two main classes with two types of vision [11].

- Active Vision: A vision system that uses a sensor that illuminates the objects of the scene by projecting structured patterns on their surfaces and observing the deformations from another. This sensor is said to be active. The difficulty for this type of vision is the matching of the points of the initial pattern to the scanned object. In this case calibration may be necessary.
- Passive vision: it is the use of two fixed cameras, a binocular vision or stereoscopic vision could constitute a specific form of binocular vision, based on different system capacities of the fusion.

SECTION IV.

Theorecal Bases for The Developpement of Qalgorithmes for Facial Recognition of Student's Emotions and Reactions in The Proposed Model

For our system the main steps to contribute the automatic tutor are represented in three main levels which are:

- Framing: first level of FRAT system uses the webcam as input and frame as result of the framing process of the live video.
- Recognition: frame from the first level are the input for the second level of FARAT that uses Viola jones algorithm for the face detection and the recognition. As result of this process features of the face detected are sent to the Data base and compared to the existing faces. If the person does not exist in the DB, his features will be added to the base.
- Emotion treatment: The existing feature in the person data base undergoes the tracking action using the KLT algorithm. For the face treatment the system uses an emotion detect algorithm by sending information to the DB and comparing the result to the existing emotion.

The emotion detected by the face treatment process is the input for making the decision by the tutor.

The following diagram represents the task performed by FARAT gradually

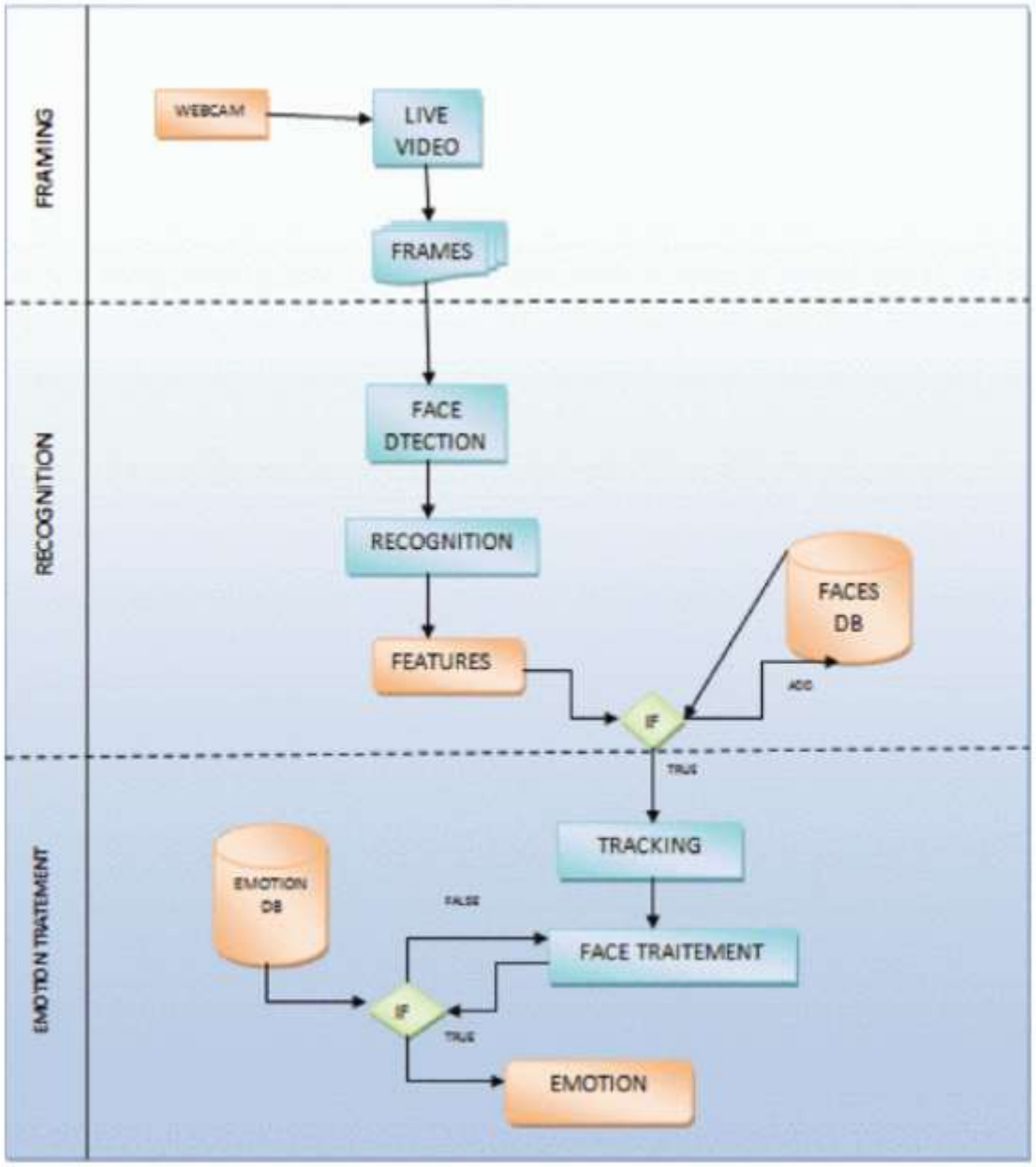

Fig 2. FARAT activity diagram

A. The Viola Jones Algorithm for The Facial Recognition

This is the first real-time face detection algorithm, it must tell if the image contains a human face and specify its location.

A natural framework for the examination of this problem is that of the binary classification, in which a classifier is conducted in order to minimize the risk of classification errors.

The viola jones method consists of scanning an image using an image of size 24px by 24px as introduced in the initial algorithm. It is an approach based on appearance. There are three components that work in harmony to allow a precise and fast detection of the face [12]:

- The integral image as input of the computational function;
- Adaboost for the selection of functions;
- The intentional Cascade for an efficient allocation of computational resources.

This algorithm is used in the second level of the FARAT diagram for the face detection and localization, this algorithm is used as first step of the face recognition and identification.

The choice of this method in our system is that it differs from other methods by:

- The use of an integral image (to calculate the characteristics more quickly);
- Selection by boosting characteristics;
- Faster execution.

B. The KLT Algorithm for the Face Tracking (KLT)

It is an algorithm for the detection and tracking of a face in a video. In the world of the computer, the tracking of vision has long been a formidable task. There are essentially three steps, which must take place during human face tracking [13]:

- Locate the face in a stream of images: using Viola Jones algorithm.
- Select the characteristics to follow: using *detectMinEigenF eatures*.
- Follow the characteristics: using *vision.Point'Iracker*

This algorithm is used to detect and track a single face in a recorded video or live video.

For the detection task one uses *Vision. CasacadeObjectDetector* of the Viola-Jones algorithm to locate the face in the video. This is used in the third level of the FARAT algorithm to initial the face analyses and the emotions recognition.

C. The Mel Frequency Cepstral Coefficient (MFCC) Algorithm for the Audio Recognition

MFCCs are defined as the dominant characteristics used for speech recognition and features extraction of the parametric representation of acoustic signals. The success of this system is based on the ability to represent the voice amplitude spectrum in compact form [14]:

The process to train an MFCC is on 7 steps:

- Pre-emphasis: Signal processing through a filter that focuses on the highest frequencies.
- Framing: Segmentation of speech samples obtained by analog-to-digital conversion.
- Hamming windowing: it's a conical function which takes into account the following sequence of the characteristic extraction processing chain and which integrates the nearest frequencies.
- Fast Fourier Transform: It is used to convert the convolution of the pulse *U[n]* and the impulse response of the vocal tract in the temporal domain *H[n]*.

- Mel Filter Bank Processing: Each amplitude; coefficient FFT is multiplied by the gain of the corresponding filter. Each filter output is the sum of its filtered spectral components (Fig. 3).

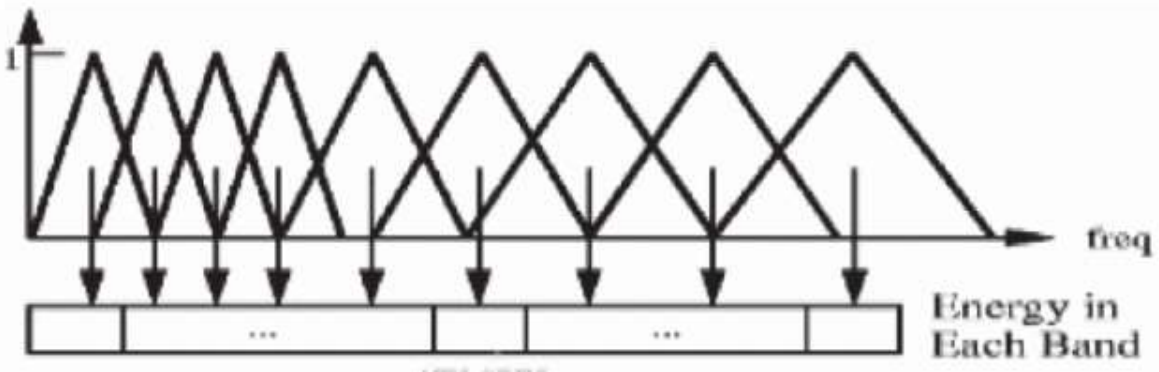

Fig3. Mel scale filter bank

- Discret Cosine Transform: This step consists of converting the Mel spectrum in the time domain using the DCT;
- Delta energy and Delta spectrum: This step consists in adding the characteristics related to the speech signal to a change of the cepstral characteristics in a time domain. To calculate the delta coefficients, the following formula (1) is used:

$$d_t = \frac{\sum_n^N n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^N n^2} \quad \textbf{(1)}$$

In MATLAB, MFCC uses the commands bellows for voice and speech recognition:

- *Audacity:* record the input speech database;
- *Wavread:* read the input wave file and returns its samples.

SECTION V.

Conclusion

To contribute to create an automatic tutor based on facial and audio recognition (FARAT) in a distant learning system, we used several algorithm such Viola jones and KLT helping making decision for the tutor by detecting the features of the learner face and MFCC as an algorithm used with others à command and algorithm to detect the wave in the learner voice. The use of these algorithms helps the system to identify the mood and the reaction of learner in the e-learning process.