Computer Vision in Education

Frendy Lio
College of Engineering
University of Hawaii at Mānoa
Hawaii, USA
frendy@hawaii.edu

Abstract—This paper is a survey about computer vision in education. We discuss different situations where computer vision can assist education, as well as why, how, and the pros and cons of each method.

Index Terms—Education, Computer Vision

I. Introduction

This paper will give a brief introduction to different scenarios of how computer vision can be used in education. In each section, we will give the reason why we need it, how is the algorithm implement; and finally the pros and cons of each scenario.

This paper is divided into background, approaches and a conclusion sections.

The background section will talk about the reason why we need computer vision in education. It will also give a motivation and background information about computer vision in education.

The approaches section will discuss the different types of approaches taken. First, we will talk about Computer Vision for Attendance and Emotion Analysis in School Settings. Following this, we will talk about the Automated Online Exam Proctoring. Next, Controlling Mouse Motions Using Eye Tracking UsingComputer Vision. After this, we will discuss Detection & Identification of the Human Body using FA based Algorithms. Besides this, we will discuss Computer Vision-Based Object Tracking as a Teaching Aid for High School Physics Experiments. Finally, we will talk about Using Video to Automatically Detect Learner Effect in Computer-Enabled Classrooms.

The conclusion section will have a summary of this report.

II. Background

A. Problem

With the increase of online lectures, the need to understand the effectiveness of lectures is needed to improve teaching methods or optimize student motivation and desire to learn. Therefore, the problem I would like to address is how to assess and facilitate learning in a non-obstructive manner, without continuous human interaction.

B. Motivation

As in-person classes have been transitioning to online classes because of the current pandemic that the world is facing, as a student I feel my academic performance has not been as optimal as desire. Therefore, this motivates me to address this problem since this is an ongoing issue to not only online lectures but also in-person lectures. I believe that with computer vision, we will be able to assist students with academic performance.

C. Background

Computer vision is concerned with extracting information about a scene by analyzing images of that scene [1]. Computer vision can be used in education to enable the instructor and students benefit not only from online education but also, normal education via enhancement learning, attendance detection, and so on.

Education is a fundamental pillar for society, as education gets better, we need to create tools that will help professors teach more efficiently. This is where computer vision comes to play. Computer Vision is a new approach to help to learn.

III. Approaches

A. Computer Vision for Attendance and Emotion Analysis in School Settings

1) Why?

A way that computer vision can be used in education is by taking attendance and emotion analysis in school settings. By using computer vision for taking attendance, an instructor can save 4 days yearly on attendance alone; it can also be used to observe how engaged is a student during class time [2].

2) How?

First, we will detect a face using a cascade classifier, following his we will recognize the face by using Local Binary Patterns Histograms (LBPH); after this, we will use Query Microsoft Face API to do an emotion analysis. Figure. 1 shows a visual representation of this method.

A Cascade classifier is a classifier that is trained with thousands of positive and negative samples that are used to detect objects in other images [3]. In this situation, the positive example will be imaged with faces and the negative sample will be imaged without faces.

A Local Binary Patterns (LBP) combined with a Histogram of Oriented Gradients (HOG) creates LBPH which can be used for face recognition tasks. LBP is a simple yet very efficient texture operator that labels the pixels of an image by thresholding the neighborhood of each pixel and
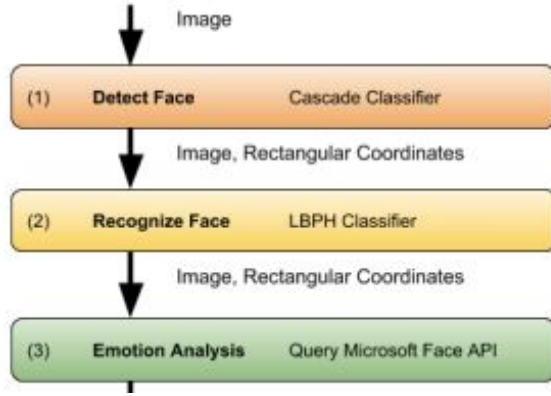
Fig. 1. Visual representation of this method.

considers the result as a binary number. HOG is used to improve the detection performance considerably on some datasets.

3) Pros and Cons

One of the advantages of this method is that it is relatively easy to learn for new coders. A beginner can learn the basic skills of face detection, face recognition, and emotional analysis [2]. Another advantage of this method is that it has a lot of potentials to grow; an example will be expanding this method to detect an unauthorized person.

One of the cons is that the accuracy of detecting a correct face is relatively low; there is a 40 percent accuracy; this might be caused by the distance that a camera is used when the picture is taken. Figure. 2 shows an average score relate to this. Another disadvantage of this Computer Vision for Attendance and Emotion Analysis in School Settings paper is that it has a small amount of dataset [2].

TABLE I
ERROR SCORE AND TRAINING DATA-SET SIZE

Number of Training Images	Avg Score For Correct Face Recognition	Avg Score For Incorrect Face Recognition
20	77.5	71.25
180	40.0	43.0

TABLE II
ERROR SCORE AND DISTANCE FROM CAMERA

Distance from Camera	20 Image Data-set	180 Image Data-set
60	77.5	40.0
120	115.0	61.8
180	135.0	81.50

Fig. 2. Image of the Tables from this method.

B. Automated Online Exam Proctoring

1) Why?

As we move to online learning, we need a way to automate online exams to avoid cheating as well as to minimize personnel resources.

2) How?

The Automated Online Exam Proctoring paper talks about its multimedia analytics system to perform automatic and continuous online exam proctoring [4].

This paper's methods architecture has 3 main sections, the OEP Hardware, Basic Components, and Cheat Detection. In this survey, I will be only focusing on a portion of the cheat detection which uses Support Vector Machine (SVM) to detect the cheating behaviors. Figure. 3 shows the architecture of this paper.
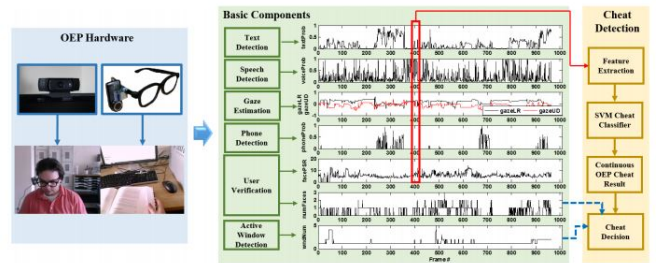

Fig. 3. Architecture of the Automated Online Exam Proctoring Paper.

SVM is a set of supervised learning methods most commonly used in classification, regression, and outliers detections [5]. The main idea is to represent a dataset as a set of points in space and separate them by categories that have a clear gap/distance from each category.

The Automated Online Exam Proctoring paper uses an SVM classifier to make continuous decisions on cheat behaviors. They divided their dataset into a positive class, cheating behaviors, and, negative class, non-cheating behaviors. They divided the positive class into 3 categories, text related cheating, speech cheating, and cheating using a phone [4].

3) Pros and Cons

One of the advantages of this method is that is affordable and convenient to use from the test taker's perspective as it only requires a couple of inexpensive camera and microphone, besides this, they have an accuracy of 87% to detect cheating behaviors [4].

One of the cons of this method is that is a little bit too complex; it consisted of various techniques besides computer vision.

C. Controlling Mouse Motions Using Eye Tracking Using Computer Vision

1) Why?

This paper talks about how to control/mimic a mouse motion using your face. I believe that this will be useful

for educational purposes as students with disabilities; for example, not having hands; to use a computer in an online course.

2) How?

The algorithm of this paper is made of three types of detection, Face detection, Eye Detection, and Mouth Detection [6].

For the face detection of this algorithm, the paper uses a pre-trained dataset iBug 300-Wand openCV to detect a face. For the eye detection section, the paper used the Eye-Aspect-Ratio (EAR) to determine whenever a person's eye is flickering or not. Finally, for the mouth detection, they used the mouth-aspect-ratio (MAR) to determine whenever a mouth was opened or not.

To summarize, this paper uses a pre-trained dataset that maps facial coordinate points to a face; following this, it uses the EAR and MAR mathematical equations to determine the eyes and mouth movements. The EAR and MAR equations are shown in Figure. 4 and Figure. 5
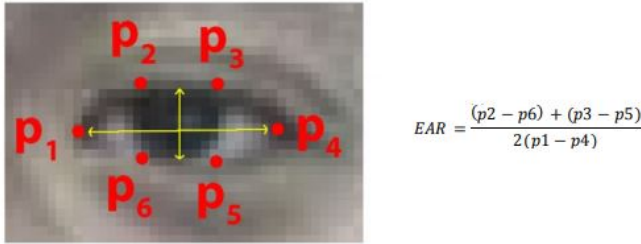

$$EAR = \frac{(p2 - p6) + (p3 - p5)}{2(p1 - p4)}$$

Fig. 4. EAR Equation and Image of Eye Points.

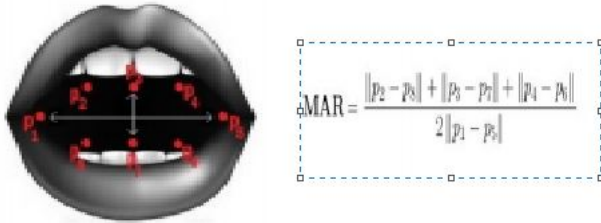$$MAR = \frac{\|p_2 - p_8\| + \|p_3 - p_7\| + \|p_4 - p_6\|}{2\|p_1 - p_5\|}$$

Fig. 5. MAR Equation and Image of Mouths Points.

The algorithm used can be shown in Figure. 6. The basic steps of the algorithm are the following.

- You have a camera that looks at a picture to detect a face.
- If it detects a face, it will decide what kind of eye or mouth movement the user is making.
- If a user wants to scroll a page, they will squint their eyes and move their face up and down.
- If they want to mimic a left or right-click of a mouse, they will squint the corresponding eye.
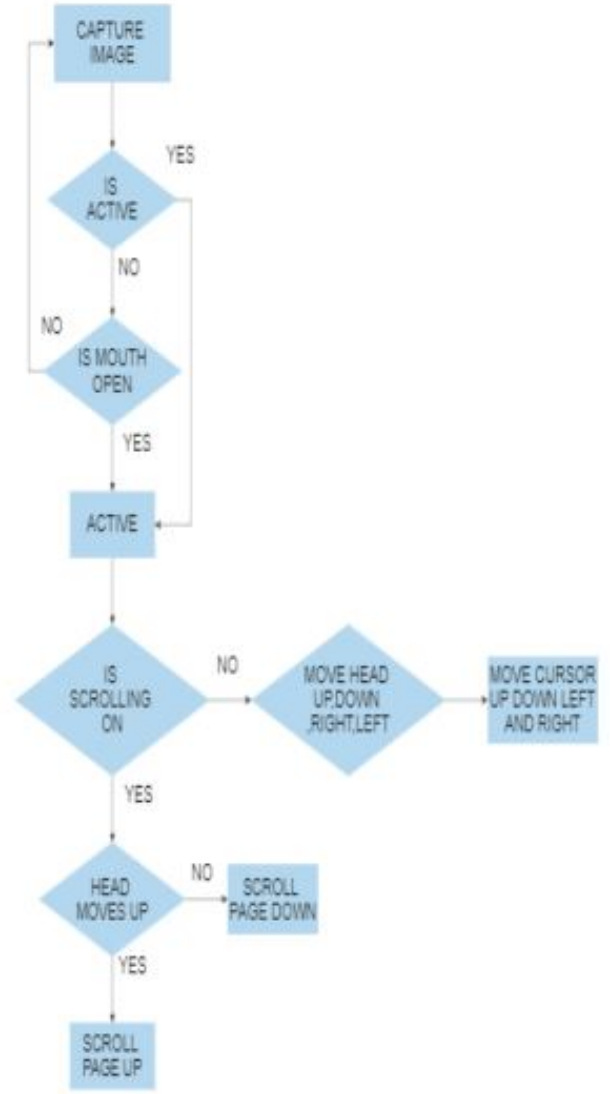- They will open or close their mouth if they want to enable mouse control.

Fig. 6. FlowChar1 of the Algorithm.

3) Pros and Cons

An advantage of this method is that it uses basic concepts and resources for face detection, besides this, the dataset iBug 300-Wand is quite big as it has been created for competitions [7].

A disadvantage of this paper is that it doesn't tell us how accurate or the success rate of this algorithm; it just tells us that it showed correct results [4]. Another issue might be that the face detection might be inaccurate for some of the points for the EAR and MARS equation

D. Detection & Identification of Human Body using FA based Algorithms

1) Why?

We can use computer vision to detect the human body. This is useful in education as we can use this to detect unusual shapes or behavior by identifying human shapes.

2) How?

This paper's algorithm is divided into two sections. First, we have the Person Detection System which uses a Feature and Appearance (FA) Algorithm. Then we have the person recognition System that uses a Multiclass Support Vector Machine (MSVM) [8]. Figure x and y show the complete system for each section.

An MSVM is a regular SVM scaled to multiclass problems. More information and mathematical equations can be found in On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines paper [9].

3) Pros and Cons

An advantage of this method is that it has a high accuracy for detecting the human body as shown in Figure. 7.

Dataset	Size	Standard deviation	Mean
Hall monitor	4.25 MB	0.014	0.892
Walk	4.00 MB	0.024	0.832
Fight	3.27 MB	0.124	0.823
Skip	2.50 MB	0.105	0.876

Fig. 7. Accuracy for FA algorithm

A disadvantage of this method or paper is that it doesn't specify what type of feature and appearance algorithm it uses. It also uses a complex type of classifier, MSVM.

E. Computer Vision Based Object Tracking as a Teaching Aid for High School Physics Experiments

1) Why?

Computer Vision can also be used to enhance students learning ability by keeping them engaged. An example of this is in science education. As we know, experiments play a vital role in learning the material. In high school physics, especially in mechanics, many experiments are conducted where tracking a single or multiple objects are required. This manual method is time-consuming, generates a higher error, and incapable of producing multiple readings rapidly [?]. This is where Computer Vision can be used to help students gather data easier and faster.

2) How?

This paper's methods have 5 main parts. First, we have image acquisition where we use a camera or any sort of device to take pictures or videos of an experiment.

Following this, we have the Image preprocessing where they use Histogram equalization and they crop the image to include only the region of interest.

Next, we have the feature extraction where they use simple edge detectors such as the canny edge detector or the differential edge detector.

After this we have the object identification system where it uses different types of detectors such as Harris [11], KLT [12], and Moravec [13].

Finally, they export the data, and students are enabled to use such data for their experiments.
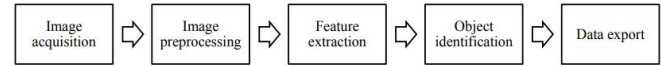

Fig. 8. FlowChart for Computer Vision in Experiments

3) Pros and Cons

Some advantages of this method include that it is flexible and can be changed depending on the experiment or situation. According to this paper, several standard high school physics experiments were conducted where computer vision-based tracking was used to aid the data gathering [10]. This proposed technique can be used to increase accuracy and to create new ways of conducting science experiments.

Figure. 9 and Figure. 10 shows a couple of graphs of data gathering that this method used for some experiments.
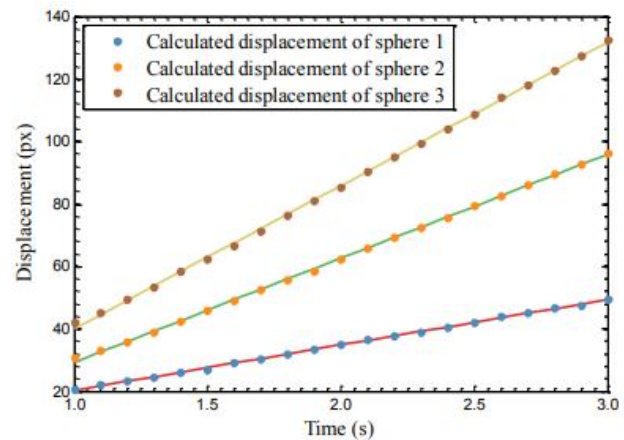

Fig. 9. Tracking a ball location against time

A disadvantage of this method is that it is too broad. As this method doesn't specify what type of image preprocessing or feature detection; we will need to test various types of techniques to find the most optimal result for each experiment.

F. Using Video to Automatically Detect Learner Affect in Computer-Enabled Classrooms

1) Why?

During online lectures, it is really hard to see if a student is engaged or distracted from the lecture. Computer Vision can help detect students' effect from facial expression and gross body movements [14]. By detecting such effects, we
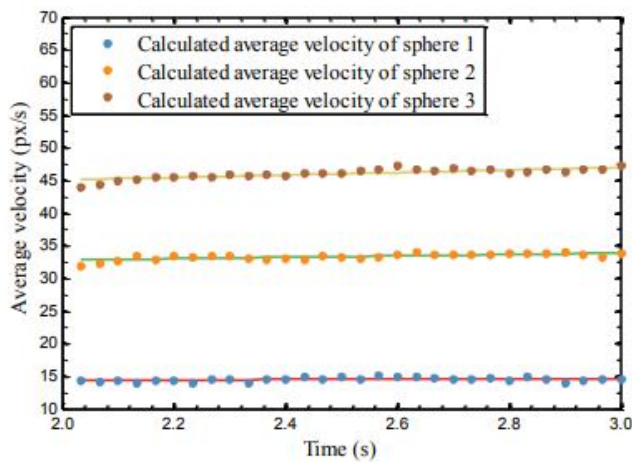
Fig. 10. Tracking an average terminal velocity of different size spheres

can inform the instructor so that the instructor can do something about it.

2) How?

This paper method uses FACET, a commercialized version of the CERT computer-vision software [14]. FACET is a program that is no longer publicly available that is used for facial feature extraction.

FACET gives a likelihood of facial movement such as Inner Brow Raise, Nose Wrinkle, and so on. Figure 11 shows an example of the FACET program.
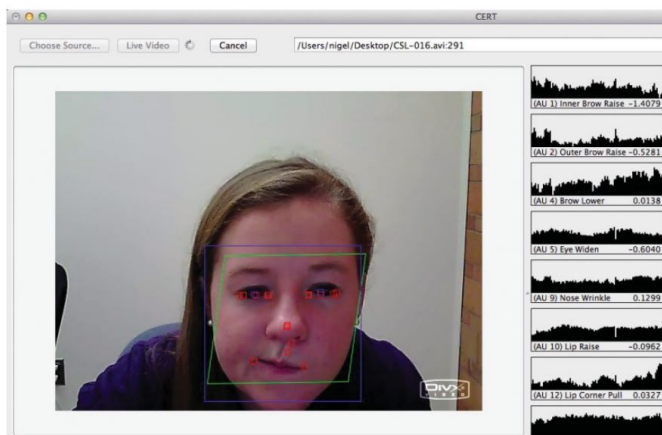

Fig. 11. FACET program

To compute gross body movement, this paper by measuring the proportion of pixels in each video frame and see if there was a difference in each of them by a certain threshold.

Following this, they used a supervised learning method to calculate the emotion of a student, if they were bored, confused, happy, etc. They used 14 different classifiers such as Bayesian classifiers, logistic regression, classification via clustering, and so on [14].

Finally, they used Cross-validation for baseline classification, Figure. 12 shows the result that this paper obtained.

Classification	AUC	Accuracy	Classifier	No. Instances	No. Features	Window Size (s)
Boredom	.610	64%	Classification via Clustering (k-means)	1305	20	12
Confusion	.649	74%	Bayes Net	1305	15	12
Delight	.867	83%	Updateable Naïve Bayes	1003	24	3
Engagement	.679	64%	Bayes Net	1228	51	9
Frustration	.631	62%	Bayes Net	1132	51	6
Off-task	.816	81%	Logistic Regression	1381	15	12

Fig. 12. Details and Results for Baseline Classification with All Data

3) Pros and Cons

An advantage of this method is that is quite promising as it uses FACET for face recognition.

A disadvantage of this method is the amount of data that we have to train the model as mentioned in the paper [14]. Another disadvantage might be the lighting conditions of the environment as it might affect the detection.

G. Conclusion

In this paper, we talked about different types of computer vision that we can use in education.

First of all, we talked about how to use computer vision for attendance and emotion analysis. We can use LBPH for face recognition and use Microsoft Face API to detect their emotions.

Following this, we talked about how we can automate online exam proctoring that uses SVM to make a continuous decision on cheat behaviors.

Next, we talked about using face recognition to help students with disabilities use a computer for educational purposes. It uses face recognition, EAR, and MAR equations to have a controlling mouse moving program using eye-tracking.

After this, we talked about how we can identify the Human body using feature and appearance algorithms. This algorithm uses a multiclass support vector machine and has a high accuracy for detecting.

Besides this, we talked about how computer vision can be used during classes. It can be used to enhance student data gathering for experiments or labs.

Finally, we talked about how we can use video to automatically detect the learner effect in computer-enabled classrooms. We learn that we can use FACET to help us do this detection.

We learn that we can use existing computer vision algorithms and apply them to education by tracking students' behavior, helping students with disabilities, and more. Computer vision creates a new set of possibilities for education.

References

[1] A. Rosenfeld, "Computer vision: basic principles," in Proceedings of the IEEE, vol. 76, no. 8, pp. 863-868, Aug. 1988, doi: 10.1109/5.5961.

[2] S. Deniz et al., "Computer Vision for Attendance and Emotion Analysis in School Settings," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0134-0139, doi: 10.1109/CCWC.2019.8666488.

[3] OpenCV, "Cascade Classifier Training", Dec 2020. Accessed on: Dec, 2020. [Online]. Available: https://docs.opencv.org/3.4/dc/d88/tutorial_traincascade.html

[4] Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu and X. Liu, "Automated Online Exam Proctoring," in IEEE Transactions on Multimedia, vol. 19, no. 7, pp. 1609-1624, July 2017, doi: 10.1109/TMM.2017.2656064.

[5] Scikit-Learn, "1.4 Support Vector Machine". Accessed on: Dec, 2020. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html

[6] K. Meena, M. Kumar and M. Jangra, "Controlling Mouse Motions Using Eye Tracking Using Computer Vision," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 1001-1005, doi: 10.1109/ICICCS48265.2020.9121137.

[7] iBug, "300 Faces In-The-Wild Challenge (300-W), IMAVIS 2014". Accessed on: Dec, 2020. [Online]. Available: https://ibug.doc.ic.ac.uk/resources/300-W_IMAVIS.

[8] S. Upadhyaya and M. Sujatha, "Detection & identification of human body using FA based algorithms," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, 2017, pp. 469-472, doi: 10.1109/ICECDS.2017.8390210.

[9] K. Crammer and Y. Singer. On the Algorithmic Implementation of Multi-class.

[10] G. D. Illeperuma and D. U. J. Sonnadara, "Computer vision based object tracking as a teaching aid for high school physics experiments," 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, 2017, pp. 1-6, doi: 10.1109/EECSI.2017.8239112.

[11] K. Derpanis, The harris corner detector. 2004.

[12] Jianbo Shi and Tomasi, "Good features to track," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR94, 1994, pp. 593–600.

[13] H. P. Moravec, "Rover visual obstacle avoidance," Proc. 7th Int. Jt. Conf. Artif. Intell. IJCAI, vol. 2, pp. 785–790, 1981.

[14] Nigel Bosch, Sidney K. D'mello, Jaclyn Ocumpaugh, Ryan S. Baker, and Valerie Shute. 2016. Using Video to Automatically Detect Learner Affect in Computer-Enabled Classrooms. <i>ACM Trans. Interact. Intell. Syst.</i> 6, 2, Article 17 (August 2016), 26 pages. DOI:https://doi-org.eres.library.manoa.hawaii.edu/10.1145/2946837