Optical Character Recognition Based Intelligent Database Management System for Examination Process Control

Mehdi Rizvi*, Hasnain Raza*, and Shahab Tahzeeb
Department of Computer and Information Systems Engineering
NED University of Engineering and Technology
Karachi, Pakistan.
*Authors contributed equally
Mohammadmehdi.cis@gmail.com,
cis.hasnain@gmail.com, and stahzeeb@neduet.edu.pk

Shan Jaffry
School of Engineering and Advance Studies
Massey University
Palmerston North, New Zealand.
s.jaffry@massey.ac.nz

Abstract*-Computer Vision techniques have recently opened several new avenues of research owing to advancement in hardware and software technologies. This technique relies on extracting relevant information from images to take intelligent decisions. In this project, we are using two subclasses of computer vision techniques namely optical character recognition (OCR) and object recognition to develop an intelligent database management system for post examination process control. Using MNIST database, in which we have 10,000 samples of test set and 60,000 samples of training set, we have applied a deep learning algorithm for handwritten digits recognition on an examination sheet. Images are adjusted using pre-processing techniques. After extraction of information from the first sheet of examination copy, we have classified the signatures from invigilator and examiner using feature extraction algorithm where signatures are treated as objects. The extracted information is stored in database and individual and collective scores are then computed. Our model also generates graphs of result for better understanding. This initial work reduces the workload of post examination data entry process and makes it time efficient. In future, it is expected to be integrated in an android based application for automatic post examination control expected to be used at author's university.

***Keywords*-Handwritten digits; OCR; Signature Recognition; SIFT; Features Extraction.**

1. INTRODUCTION:

Lately artificial intelligence and machine learning techniques have gained much prominence by minimizing external input and optimizing processing in many labor-intensive tasks, both mechanically and computationally. Out of several important applications of machine learning, image recognitions and pattern recognitions enable identification of images or patterns for different use cases, respectively. Image recognition technique focuses on extracting features from images to categorize images. On the other hand, pattern recognition deals with recognizing patterns either in images or data-sets. As an application of above-mentioned techniques, optical character recognition (OCR) has gained much importance as it enables to recognize texts from any given image. The OCR technique can even identify characters from handwritten scripts.

In this research, we have done extraction of enrolment number of students for their identification, marks of students for their result generation and signature recognition of invigilator and checker by object detection technique to automate the post examination work. We have used actual examination sheet from NED University of Engineering and Technology. This project will help in reduction of time and effort to generate result by eliminating manual data entry mechanisms.

Recently, researchers have emphasized on image preprocessing methodologies for optical character recognition and object recognition using dedicated libraries. For instance, a convolution constructed method for static-pattern noise exclusion in OCR as a preprocessing step has been presented in a research paper [1]. On the other hand, convolutional neural networks-based OCR technique has been presented in OCR for Sanskrit

978-1-5386-7729-2/19/$31.00©2019 IEEE

using CNN [2]. In the domain of object recognition through feature mining, scale invariant feature transform algorithm has been implemented in Research on image detection and matching based on SIFT features [3]. The similar technique has been used for facial recognition in a literature we have reviewed [4]. Similarly, signature grouping is accomplished through one-class support vector machine [5]. SIFT also helps in recognition of offline signatures [6].

In this project, we use OpenCV library to accomplish preprocessing tasks. These tasks include image alignment, extraction of region of interests, noise reduction, threshold management and segmentation. Which help us in transforming the image of segmented digits into trained digits of Modified National Institute of Standards and Technology databases. After handwritten digit recognition, we have to recognize signatures. We treated them as an object for classification as real or fake signature and recognize it, using feature recognition algorithm. This research has made contribution in the domain of academics by combining image processing with OCR for post examination process control. The remaining paper is ordered as follows. Unit 2 debate details of system design. Section 3 and 4 presents preprocessing steps and digit recognition technique, respectively. Section 5 demonstrates signature recognition with results presented in Segment 6. Lastly, final remarks and future extensions are discussed in Unit 7.

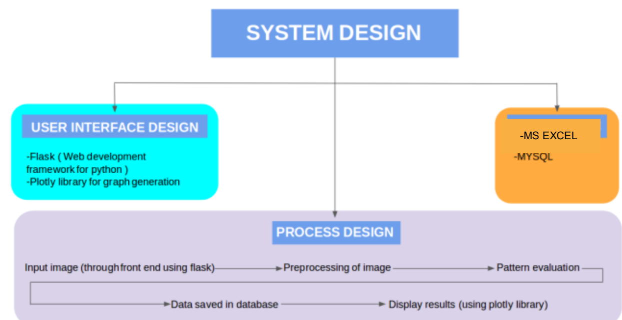
2. System Design:

The main components for our system design are UI design, data design, and process design which are explained in this section.

To develop a user-friendly screen layout, we use Flask. It is a micro web framework written in Python. It lacks the functionalities of full-fledged web frameworks but supports extensions that can add up features as if they were implemented in Flask itself. In our scenario, we take input image through a web page from the user, and display student's individual marks and overall class result graphs once it is extracted from the image. For representation of graphs we use Plotly python library. Plotly is, basically a data visualization toolbox. For data management, we use MS Excel as it the most commonly used management platform for data entry in academics.

This research begins with the input of an image, which is then passed through preprocessing techniques such as image alignment, extraction of region of interest (RoI), noise reduction and threshold management of an image. The RoI extracted includes enrolment number, marks and signatures of invigilator and checker which are constituents of our input image. These constituents are then passed into segmentation algorithm which separates each digit. These techniques are discussed in detail in Section 3. Separated digits are then sent to the digit recognition algorithm which has already been trained by using MNIST dataset. This algorithm translates the digits in the image to textual format. Besides, signatures are classified using a feature matching algorithm. This information extracted out from an image is then sent to database. Plotly graphing library of python then imports this information from database and generates graphs representing results of students on webpage.

Fig. 1. System Design



3. Pre-Processing Techniques:

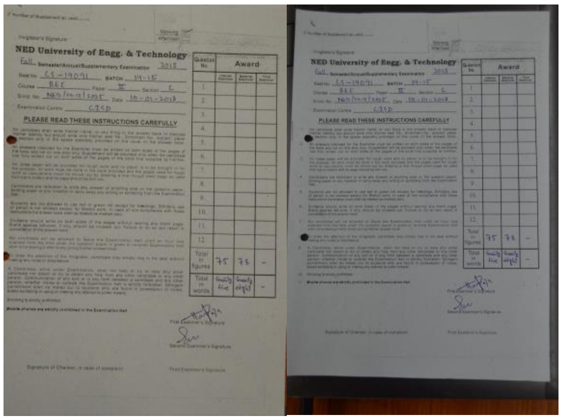
The objective of pre-processing is to enhance picture quality by reducing unfavorable misrepresentations, enhance relevant image features and to make the input image according to the requirement of our algorithm. So that image features and quality don't affect our algorithm's efficiency.

In beginning of the process, we input the initial page of the examination copy. Later preprocessing is done to fit the image according to the requirement of our algorithm. We are using four preprocessing steps. These steps are (i) image alignment, (ii) extraction of region of interests (iii) noise reduction, (iv) threshold management and (v) segmentation. Following is the detailed discussion for individual steps.

3.1. Image Alignment:

In a different images of same scenario the features of one image are different from the other image with respect to the coordinates. Image alignment is a technique of warping one image(s) so that the relevant features align perfectly. In order to extract features like roll number, marks and signatures perfectly, an input image must be aligned with respect to the reference image so that coordinates of interest do not change.

Fig. 2a. Reference image and Fig. 2b. Input image



The base of image alignment techniques is a 3 × 3 matrix called Homograph.

$$H = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix}$$

Perspective transformation of a plane is termed as Homography. It re-projects a plane from one camera into a different camera view, with respect to the change in translation and orientation of the camera. Perspective transformation means mapping of 3D points onto 2D image planes using the transformation matrix that subsumes the camera features like focal length, optical Centre, and the extrinsic parameters like orientation and translation.

To understand Homography, let (x_1, y_1) and (x_2, y_2) be the point in first image and coordinate of same physical second image's point respectively. Homography narrates them as,

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$$

For finding key points in both the images of top sheet we implemented Oriented FAST and Rotated BRIEF (ORB) feature finder which has two parts: (i) locator (ii) descriptor.

Locator: It specifies those points on top sheet images that are not changing when image is subjected to transformations like translation, scaling and rotation. It finds the coordinates of such points in (x, y) form.

Descriptor: To differentiate the points of interest highlighted by locator from each other, we use descriptors. It encodes the appearance of the point to make one feature point look different from the other. It is a list of real integers and similar physical point in two pictures must have the exact descriptor array. ORB uses an enhanced edition of the descriptor known as Binary Robust Invariant Scalable Key point (BRISK), as BRISK slightly outperforms BRIEF in viewpoint changes. It is also equipped with orientation compensation; to compute the positioning of the main point, BRISK uses local slopes among the sampling couples.

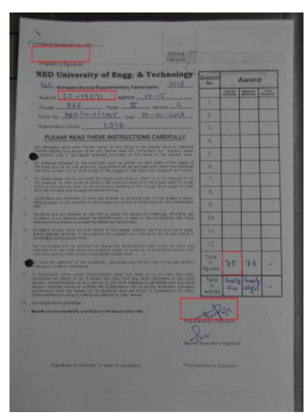
$$G\,(p_i\,,\,p_j) = (p_j - p_i)\;.\;\frac{f\,(p_j\,,\sigma_j) - I\,(p_i,\sigma_i)}{\|p_j - p_i\|\text{^}2}$$

Here g(pi, pj) is the local slope among the sampling couple (pi, pj), I is the smoothed intensity in the corresponding sampling point by the appropriate standard deviation. To compute direction, the slopes among all the long pairs are added and take arctan(gy/gx) – the arctangent of the y component of the slope divided by the x component of the slope. This provides the angle of the key point. By using the angle obtained, image is rotated with respect to the original image.

3.2. Extraction of region of interest (RoI):

RoI can be extracted using two methods. Either by layout analysis algorithms if we have variable layout inputs or by specifying the coordinates of RoI if we have static layout input. In our scenario, we have static layout, which is not changing so we have extracted the RoI by specifying their coordinates. That is why random input image is first aligned with respect to the reference image and then further processing is done.

Fig. 3. Region of interest



3.3. Noise removal:

Noise is an unwanted variation of pixels in an image that reduces the efficiency of image processing mechanism. Suppose we have a noisy pixel, $p = p0+n$ here $p0$ is the original value of pixel and the noise in that pixel is represented by n. We can now take same pixels in large numbers from various images and compute their average. Ideally, the value of p should be equal to $p0$ since mean of noise is zero. We use Gaussian Blurring a smoothing technique to remove minor quantities of noise. We designated a small area around a pixel known as kernel and do operations like Gaussian weighted average, to replace the central element. Odd and positive is the requirement for the length and width of kernel. Gaussian blurring is most effective in eliminating Gaussian noise from the picture.

3.4. Thresholding:

In thresholding, if pixel value is not lesser than a threshold value, it is allocated white value, otherwise it is allocated black value in our scenario. Its vice versa is also possible.

We have used thresholding for two purposes,

3.4.1 Highlighting noise:

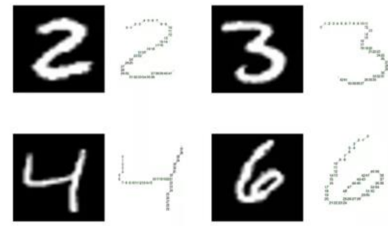
After we received the RoI, we threshold it. Which helps us in differentiating the foreground with background in an image. It converts all the higher value pixels into white and lower value in to black. Noise in form of dots is also highlighted, which is then filtered out by calculating the area of contours present in image and comparing them with a threshold area value indicating foul contours below it.

Fig. 4. Thresholder enrolment number

3.4.2 Digits classification:

Our digit classification algorithm is trained on digits with number written in white having black background. Whereas our input digit is written in black having white background so we used inverted thresh function. Results are given below.

Fig. 5. Conversion using threshold function



3.5. Segmentation:

Our algorithm for digit classification works on single digit. Whereas the RoI we are receiving from image is continuous series of digits like enrolment number and marks. So, to separate digits we use segmentation technique in which we find the contours of the digits and find out their coordinates. Using coordinates, we crop them out and save them in a list, from where they are passed to the digit recognition algorithm.

Fig. 6. Segmentation of Enrollment number

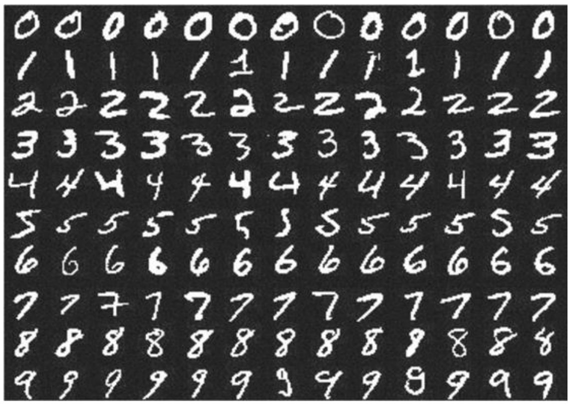
4. <u>HANDWRITTEN DIGITS CLASSIFICATION:</u>

Once the input image is preprocessed and RoI is selected, we apply segmentation on enrolment number and marks and save them to the list. Saved digits in list are then passed to our digit classification algorithm which is a developed deep learning model to get a near advanced performance on the MNIST. We have used Keras with TensorFlow for this purpose.

4.1 MNIST Dataset:

MNIST was built from a numerous scanned document datasets accessible at National Institute of Standards and Technology. These images were then centered and normalized in size. Every image is in a square of 28x28 pixels. 60,000 images were used for training a model and 10,000 were used to test it. Results with 1% inaccuracy were achieved.

Fig. 7. MNIST Dataset digits



4.2 Keras:

Keras is a neural networks API, written in Python. It can also be termed as a wrapper of TensorFlow for python. It was developed for enabling fast experimentation because TensorFlow is complex and requires heavy computations which is impossible for normal CPUs to carry out. We used Keras because we needed a deep learning library that enables easy and fast prototyping, supports both convolutional networks and recurrent networks, as well as combinations of the two and runs on CPU seamlessly. In accordance with the MNIST, Keras library gives a convenient method for loading its dataset. Keras automatically downloads the dataset of MNIST when its mnist_data () function is called and stores it in home directory.

4.3 Convolutional Neural Network:

Convolutional Neural Network (CNN) is very much like normal Neural Network. They are composed of neurons having learnable weights and biases. Every neuron accepts some inputs, performs a dot product and follows it with a non-linearity. CNN makes the assumption that the inputs are images, which enables us to encode certain features into the architecture and make forward function more efficient to implement and also decrease the number of parameters used in the network.

In the layers of a CNN, neurons are arranged in 3 dimensions: width, height, depth. Here three main types of layers are used to build ConvNet architectures: Convolutional Layer, Pooling Layer, and Fully-Connected Layer. CNN adjusts the original image layer by layer from the original pixel values to the ultimate class scores. The Convolution layer is the core building block of a CNN that does most of the heavy computational lifting. It is very common to insert a Pooling layer in-between successive Convolution layers periodically in a CNN architecture. Its function is to reduce the spatial size of the representation to decrease the number of parameters and computation in the network. Neurons which are in fully connected layer have full connections to all activations of previous layer. Their activations can hence be computed with a matrix multiplication followed by a bias offset. In our algorithm, we use multiple layers of convolutional, max pooling and fully connected layers.

5. SIGNATURE RECOGNITION:

After pre-processing and recognition of enrollment numbers and marks from the top sheet of examination copy. Now we have to recognize the signature of invigilators and checkers. We treated signatures as an object. So, for signature recognition, we extracted the features of signatures to classify the query signatures.

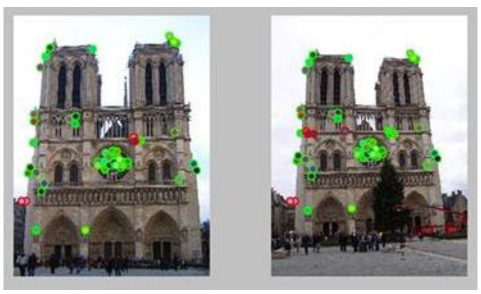
We used Scale Invariant Feature Transform (SIFT) algorithm for features extraction here. SIFT detects key points of the signature to recognize the signature under test as either fake or original. Following is an outline of processes occurring with SIFT to extract the features of a signature.

1. **Construction of a scale space** in this initial step, an internal representation of the original signature image is made to safeguard scale invariance. This is done by generating a "scale space".
2. **Fast Approximation,** the Laplacian of Gaussian is widely used for finding fascinating points but the problem here is computational expensiveness. So here we approximate LoG using the portrayal created earlier of a signature.
3. **Finding key points** after fast approximation, we find key points of the signature, which are maxima and minima in the Difference of Gaussian image we calculate in step 2.
4. **Remove nasty key points,** edges and low contrast regions are bad key points. Removing these makes the algorithm robust and efficient. A technique similar to the Harris Corner Detector is used here.

5. Assigning an orientation to the key points an orientation is computed for every key point. Further computations are done relative to this orientation. This effectively invalids the effect of orientation of a signature, making it rotation invariant which is very important and useful in object recognition.

6. Generation of SIFT features once scale and rotation invariance are in place, one more representation is generated. This helps us in unique identification of features. With this representation, we can easily identify the feature we are looking for i.e. a signature.

Fig. 8. SIFT Extracted Features Matching

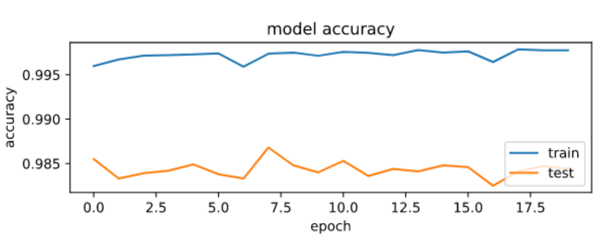


We implemented SIFT function through OpenCV and python. Our algorithm takes signature image as an input then matched input extracted features to the features of original image and if they both matched up to a certain criterion it will gives the name of the signature owner otherwise it will generate error. In our model, we must need a one fine copy of each checker so that we can match the random signature on the first page with the one present in our database so that SIFT can match features and classify the signature.

6. RESULTS:

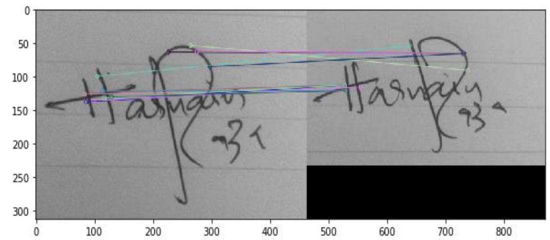
Our CNN based deep learning model recognizes the handwritten digits using Keras library and gives 99% accuracy which is excellent.

Fig. 9. Result of CNN model accuracy



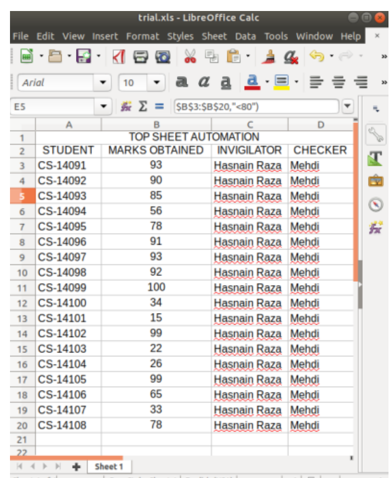
Same as handwritten OCR our algorithm does accurate signature recognition. It can classify the signature as well as it can also recognize them as fake or real.

Fig. 10. Signature Matching



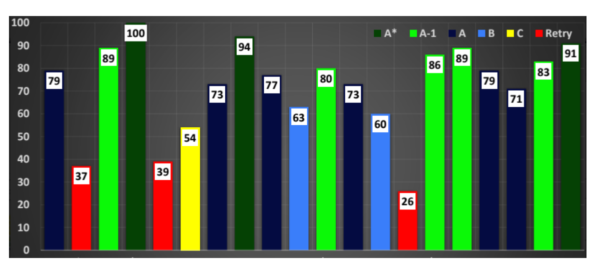
Finally, the most important thing in top sheet automation is to safe data in a digital file for instance excel sheet to easily use, edit and visualize the data. The library used for writing data and formatting information to Excel files is xlwt. Our model saves all the extracted data from top sheet in the excel sheet and hence reduce the manual work. After all it can also generate graphs to visualize and analyze the trend in data for suppose the trend of class in a course.

Fig. 11. Submission of all relevant data in xls file



	A	B	C	D
1	TOP SHEET AUTOMATION			
2	STUDENT	MARKS OBTAINED	INVIGILATOR	CHECKER
3	CS-14091	93	Hasnain Raza	Mehdi
4	CS-14092	90	Hasnain Raza	Mehdi
5	CS-14093	85	Hasnain Raza	Mehdi
6	CS-14094	56	Hasnain Raza	Mehdi
7	CS-14095	78	Hasnain Raza	Mehdi
8	CS-14096	91	Hasnain Raza	Mehdi
9	CS-14097	93	Hasnain Raza	Mehdi
10	CS-14098	92	Hasnain Raza	Mehdi
11	CS-14099	100	Hasnain Raza	Mehdi
12	CS-14100	34	Hasnain Raza	Mehdi
13	CS-14101	15	Hasnain Raza	Mehdi
14	CS-14102	99	Hasnain Raza	Mehdi
15	CS-14103	22	Hasnain Raza	Mehdi
16	CS-14104	26	Hasnain Raza	Mehdi
17	CS-14105	99	Hasnain Raza	Mehdi
18	CS-14106	65	Hasnain Raza	Mehdi
19	CS-14107	33	Hasnain Raza	Mehdi
20	CS-14108	78	Hasnain Raza	Mehdi
21				
22				

Fig. 12. Class Result



In the same way various researches have been carried out to reduce the workload of human by reducing manual work in different aspects of life. The OCR is one of the techniques which is widely being implemented for this purpose of automation under computer vision domain. Recent literature shows that extensive work on recognition of digits and alphabets is being performed by industries and academic institutions. For an instance, a research has been carried out in which process of manual data entry when money transactions are done through cheques is being carried out by implementing CNN on MNIST dataset which significantly minimizes the time consumed by recognizing the digits and alphabets automatically. This research work obtained 95.71% accuracy after testing their model on sample cheques provided by the banks [7]. Likewise, a research for automatic recognition of number plate digits has been carried out for safety and security purposes in which same technique of OCR using CNN model is used [8]. Another work for detection of student learning pedagogy using OCR has recently been carried out in which 'special forms' are filled by students and data is extracted from images of those forms using CNN model and obtained 96.87% accuracy in recognizing correct digits from image [9]. In comparison with these researches, our work also implements the same technique of OCR using CNN model trained by MNIST dataset to recognize the marks and enrolment number of students from the top sheet of their examination copy as a replacement for manual data entry once examination copies are checked. This work has obtained the efficiency of 98.5% in recognition of digits from the images of examination copies. In addition to this, our research also classifies signatures of invigilators and checkers for identification and review purposes, which has also emerged as one of the most concerned area for researchers in recent years. Like the research which has combined SIFT algorithm and LBP for offline verification system which gives better performance than manual signature classification [10].

7. CONCLUSION:

In this research, the workload of entering marks manually into system's database and generating results afterwards is being largely decreased. We have countered this problem in order to provide our services to academic institutions. For this purpose, we have used image processing tools like image alignment, thresholding, blurring, noise reduction and segmentation to make image of the first page of examination copy understandable for our digit recognition algorithm. After segmentation of our region of interest, each separate digit is sent to our trained model, it recognizes it and then send it to the database. From database, the information is used and displayed in front of the teachers in form of graphs which are easy to understand and can be quickly analyzed to see the result trends among the students in a subject. Through this work we have automatized the entire process of entering and generating results manually after checking of papers till the very end of result display.

In future, this work can be extended to mobile application, based on android and IOS operating systems. We can also enlarge our scope of using OCR from digits to characters which can be used widely for academic purposes.

8. REFERENCES:

[1] J. Mo, B. Wang, Z. Zhang, Z. Chen, Z. Huang, J. Zhang, and X. Ni "A Convolution-Based Approach for Fixed-Pattern Noise Removal in OCR" International Conference on Artificial Intelligent and Big Data, 2018, pg. 134-138.

[2] M. Avadesh and N. Goyal, "Optical Characters Recognition for Sanskrit using Convolution Neural Networks"13th Int'l Workshop on Doc. Analy. Sys., 2018, pp. 447-452.

[3] F. Guo, J. Yang, Y. Chen, and B. Yao "Research on Image Detection and Matching Based on SIFT Features" 3rd International Conference on Control and Robotics Engineering, 2018, pg. 130-134.

[4] M. Mahamdioua and M. Ben Muhammad "Automatic Adaptation of SIFT for Robust Facial Recognition in Uncontrolled Lighting Conditions" IET Comput. Vis., 2018, Vol. 12 Iss. 5, pg. 623-633.

[5] Y. Guerbai, Y. Chibani, and B. Hadjadji "The effective use of the One-Class SVM classifier for reduced training samples and its application to handwritten signature verification" International Conference on Multimedia Computing and Systems (ICMCS), 2014.

[6] Adesesan B. Adeyemo, Adeyinka O. Abiodun "Adaptive SIFT/SURF Algorithm for Off-line Signature Recognition" Egyptian Computer Science Journal Vol. 39 No. 1, 2015, pg. 50-56.

[7] Srivastava S., Priyadarshini J., Gopal S., Gupta S., Dayal H.S. (2019) Optical Character Recognition on Bank Cheques Using 2D Convolution Neural Network. In: Malik H., Srivastava S., Sood Y., Ahmad A. (eds) Applications of Artificial Intelligence Techniques in Engineering. Advances in Intelligent Systems and Computing, vol 697. Springer, Singapore.

[8] Radzi S.A., Khalil-Hani M. (2011) Character Recognition of License Plate Number Using Convolutional Neural Network. In: Badioze Zaman H. et al. (eds) Visual Informatics: Sustaining Research and Innovations. IVIC 2011. Lecture Notes in Computer Science, vol 7066. Springer, Berlin, Heidelberg.

[9] Arnab Sen Sharma, Marium-E-Jannat, Maruf Ahmed Mridul and Md Saiful Islam. "A Deep CNN Model For Student Learning Pedagogy Detection Data Collection Using OCR". International Conference on Bangla Speech and Language Processing (ICBSLP), 21-22 September 2018.

[10] Bhushan S. Thakare and Hemant R. Deshmukh." A Combined Feature Extraction Model Using SIFT and LBP for Offline Signature Verification System". 2018 3rd International Conference for Convergence in Technology (I2CT). 6-8 April 2018.