Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

Automated Robust Facial Expression Recognition using Transfer Learning ResNet50

Habib Ur Rehman[1], Abdul Basit[2]
[1]Department of Computer Science and Information Technology, University of Balochistan, Quetta-Pakistan
[2]Assistant Professor, Department of Computer Science and Information Technology, University of Balochistan, Quetta-Pakistan
habibrehmanuob@gmail.com, drabasit@um.uob.edu.pk

DOI: 10.5281/zenodo.13920706

ABSTRACT

The human face is a convenient, fast, and accurate source of communication. Facial expressions convey internal human emotion developed using different facial traits. Affective computing works on developing systems for Facial Expression Recognition (FER) using machine learning tools and it remains an active research area for the research community. This paper proposes a deep learning-based model ResNet50 for facial expression recognition. We further applied transfer learning and fine-tuning techniques with the proposed model to improve the generalization. The model is trained and validated at the FER2013 dataset and tested with some unseen images from MMA facial expression dataset. The model archives validation accuracy of 86.32% which is comparable with existing research.

Keywords: Facial Expression Recognition, Emotion Detection, FER, Pre-trained Model, ResNet50, Transfer Learning, Fine-Tuning, FER2013.

Cite as: Habib Ur Rehman, & Abdul Basit. (2024). Automated robust Facial Expression Recognition using Transfer Learning ResNet50. *LC International Journal of STEM, 5*(2), 11–19. **https://doi.org/10.5281/zenodo.13920706**

INTRODUCTION

Human communicate their feelings and responses using different ways such as speech, facial expressions and body gestures. Face exhibits massive information regarding one's internal state, mood, responses, and many more using facial muscle movement that develop distinct expressions on face. As a non-verbal communication channel, facial expressions play an important role in daily life, human-human or human-computer communication According to literature the facial expressions has a dominant role that is almost 55% of overall communication, Mehrabian (1967).

In a study, Ekman (1993) stated seven basic facial expressions (Happy, Sad, fear, Surprised, Disgust, Angry and neutral) as being universal. Figure 1. Facial expression has extensive applications in daily life like customer tendency and product reliance detection, quantifying student learning, safe driving monitoring, lie detection, patient health or pain monitoring and mood detection.

The facial expressions are interpreted into relevant class easily and quickly by human beings. However; for computer it is a difficult task to classify these expressions. Machine learning algorithms tends to provide efficient ways for classifying emotions in relevant classes.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

Deep learning has shown promising results in image classification (Naseer et al. (2020), Iqbal et al. (2021), Ge et al (2022), Din et al. (2022)) and is potential tool for improving the accuracy.

The problem of facial expression classification is divided into three stages known as face detection, feature extraction and emotion classification. Viola and jones algorithm is used to detect face in an image. Based on classification procedure adopted, the problem of facial expression recognition is divided into handcrafted features and auto-learned features. Handcrafted features or Human engineered features use Histogram of Oriented Gradient (HOG), Linear Binary Pattern (LBP), Principal Component Analysis (PCA), as feature extractor and the extracted features are provided to a classifier for emotion classification. Auto-learned feature is a unified algorithm that performs the tasks of face detection, features extraction and emotion classification autonomously. Auto learned feature uses Convolutional Neural Network (CNN), Long Short Term Memory network (LSTM), Recurrent Neural Network (RNN), Generative Adversarial Network (GAN) for image label prediction. Deep learning on the path of image classification brought revolution in the field by improving the emotion classification accuracy.

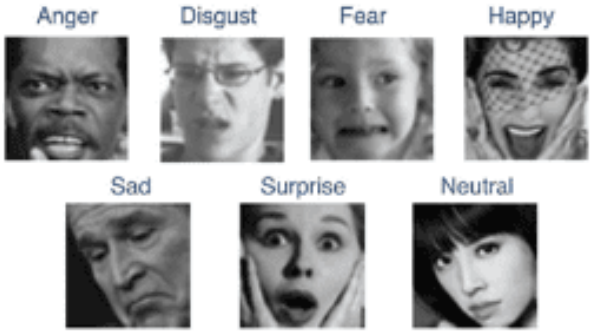

Figure 1: Sample images of seven facial expression from FER2013 dataset

This research proposes an algorithm for emotion recognition using human facial expressions by deep CNN model, ResNet50. The model is trained and validated using publically available dataset FER2013. The research addresses the poor generalization issue of deep learning model. The study aims to contribute the domain of facial expression recognition:

1. Developing and Training a deep learning model ResNet50 using FER2013 dataset that is fast and robust.
2. Improving the generalization of emotion classification by tuning the parameters of the model.
3. Testing the model on unseen images.
4. Comparing the model with state-of-the-art.
5. The trained model provides high level of accuracy that may be used in embedded system and smartphones.

Rest of the paper is organized as section II describes existing research development, Section III details the Methodology adopted, Section IV and V presents Experimental Setup and Result and Discussion, the last Section VI presents the conclusion of the paper.

LITERATURE REVIEW

The research by Ahmed, et al. (2019) proposes a model that increases validation accuracy meanwhile removing the inconsistency in classification rate and addresses the issue of low recognition rate in

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

classes like fear and disgust. The study uses CNN for feature extraction and emotion classification. The data augmentation and collecting images from multiple datasets increases the accuracy to 95.87%.

In a study, Nour et al. (2020) proposed a method that used three pretrained models (AlexNet, VGG16, ResNet) as feature extractors and SVM for emotion recognition. The CK+ dataset is used to train the model that resulted accuracy as 88.2%, 84.2%, 81.6% for AlexNet, VGG16 and ResNet respectively.

A research by Tripathi (2021) uses Deep CNN models for facial expression recognition and the best model among three is further tuned for higher accuracy gain. Pretrained models AlexNet, VGG19 and ResNet are trained and validated using FER2013 dataset where VGG19 show higher accuracy and was further analyzed using different optimizers (SGD, AdaDelta, AdaGrad RMSProb, ADAM). The resampling technique of K-Fold Cross validation with AdaGrad optimizer increased the validation accuracy to 91.89%.

Taking various approaches for accuracy improvement, Khanzada, et al. (2020) in a study develops a model capable of real world application. Transfer learning is applied on three pretrained networks (ResNet50, SeNet, VGG16). Approaches adopted are auxiliary data preparation, data augmentation, transfer learning, class weighting and ensembling that improves the accuracy to 75.8%. Finally a mobile web app is developed to apply the work in real world.

The work by Gaddam et al. (2022), proposes ResNet50 model for emotion classification. The procedure includes preprocessing, face detection, feature extraction and emotion label prediction. Model training and experimental analysis is carried out on FER2013 dataset. Finally the model achieves a validation accuracy of 55.6%.

The work by Minaee et al. (2021) proposes an attentional Neural Network that focuses on expression relevant parts of face using deep network. The visualization technique used, finds expression relevant portion of face. The method is experimented using 4 popular datasets (FER2013, CK+, FERG and JAFFE) showing 99.3% best results on JAFFE dataset.

The research by Siam et al. (2022) used a method to extract key points from facial images using Mediapipe face mesh algorithm. Extracted key points are encoded using angular encoding modules. To enhance the accuracy, feature decomposition using PCA is performed. The feature are then provided to Support vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Logistic Regression (LR), Random Forest(RF) for classification and Multi-Layer Perceptron (MLP). The model achieves 97% best performing accuracy. The models are evaluated on CK+, JAFFEE and RAF-DB datasets.

The student attention in distance learning in an online teaching environment evaluated by a deep learning framework developed by Hou et al. (2022). In this system MTCNN algorithm detects a face in online video lecture. The face image is then provided to system composed by VGG16 and ECANet for facial expression classification. The model is trained on FER2013 dataset and validated on CK+ dataset for verifying student concentration in online class lecture.

In the work by Akhand et al. (2021), deep CNN (DCNN) models are trained and evaluated for facial expression recognition. The system applied transfer learning on pretrained models (VGG16, VGG19, ResNet18, ResNet34, ResNet50, ResNet152, Inception-v3, DensNet161) using KDEF and JAFFE facial expression dataset. DenseNet161 archived best validation accuracy of 96.51% and 99.52% on KDEF and JAFFE datasets respectively.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

METHODOLOGY

The framework of proposed study comprised of image acquisition, preprocessing image, feature extraction and predicting the emotion in image. The model is trained and validated on publically available FER2013 dataset. The model is tested on unseen images to know the strength of model. The flow of proposed model is presented in Figure 1. Where detailed description of framework is given in next section.

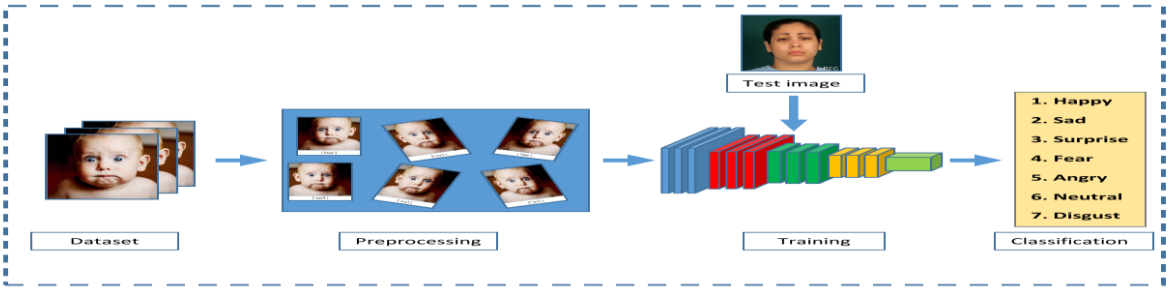

Figure 2: Model flow of facial expression recognition

Dataset

The publically available FER2013 dataset is considered for training and validation of the proposed model. The FER2013 is Comprised of 35,887 grayscale, 48 × 48 images consisting seven facial expression. **Error! Reference source not found.**. The dataset is split into 28,709 training and 7,178 validation and test images. The highest accuracy on FER2013 in published work is 75.2% on individual model. The number of sample images in each class in training, validation and test are given in table 1.

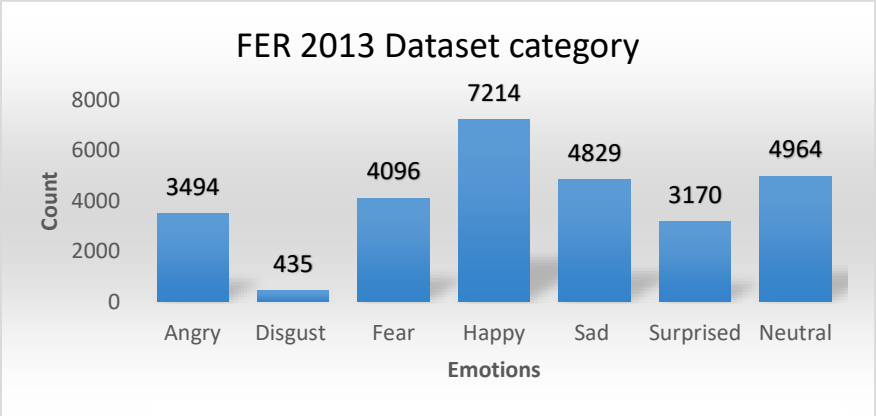

Figure 3: FER2013 dataset class label distribution

Table 1: Distribution of training, validation and test images in FER2013

	Surprise	Fear	Angry	Neutral	Sad	Disgust	Happy	Total
Training samples	3171	4097	3995	4965	4830	436	7251	28709
Validation/Test Samples	831	1024	958	1233	1247	111	1774	7178

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

Preprocessing

The FER2013 is a larger dataset available in csv file that needs some level of preprocessing to prepare it for better results. The FER2013 file contains images, labels and usage as training or test. Images in the dataset are appended and reshaped to 48 × 48. Categories counted, labels appended and an array of the images and labels are created. Dataset is split into train and test. Data augmentation increases number of images in dataset which helps to increase training and validation accuracy.

Model

Deep learning models revolutionized the field of predictive modeling by increasing the classification accuracy meanwhile merging the feature extraction and classification tasks. Existing literature on subject reveals that emotion classification using learned features manifest better result than handcrafted feature that is why we proposed deep learning model ResNet50 for facial expression classification Figure 4.

Architecture of ResNet50

ResNet is a deep learning model developed by He (2016) with an architecture based on theory of more layers more learning. ResNet has variants ResNet18, Renet34, Resnet50, ResNet101, ResNet152, and ResNet164.

Deep learning models like VGG, AlexNet, GoogleNet when going deeper in term of layers, they can solve complex problems. But going too many layers deeper arises vanishing gradient problem that saturates the accuracy.

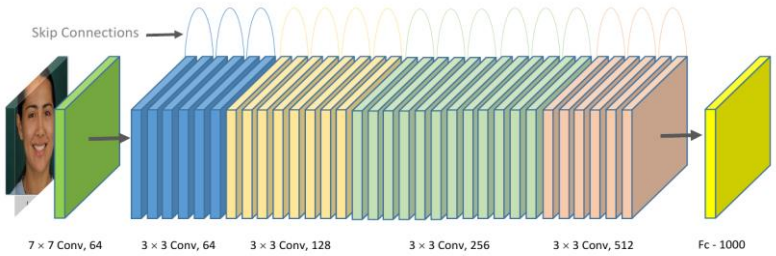

Figure 4:ResNet50 architecture.

Addressing the vanishing gradient issue, ResNet model provides skip connection approach that converts

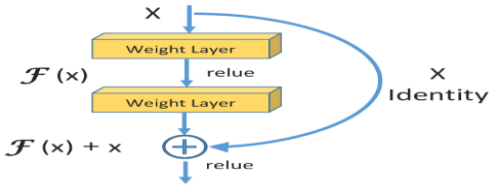

Figure 5: ResNet Skip connection.

a network into residual network Figure 5. This approach skips few layers and creates direct connection to next layers. The input X is multiplied by the weights of the layer and some bias is added.

Transfer Learning and Fine-tuning

Transfer learning is a machine learning method that reuses a pretrained deep learning model trained for one task, at another different task. Deep learning models are trained on ImageNet dataset for prediction on 1000 classes. Transfer learning reuses these models by customizing the architecture of the model.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

The first step to create a base model is to detach the old classifier and attach new one to be trained on dataset of the problem in hand. The second step freezes weights of convolutional base of model which prevents the layers being updated during training. The model is then trained for classification on new dataset applying best parameters by try and error method. Fine-tuning is an optional step that is used to increase the accuracy of model. In fine-tuning, the weights of the base model are unfreeze and the model is retrained with very little learning rate.
In our study, we apply transfer learning at ResNet50 for emotion classification problem that shows satisfactory result. The convolutional layer are freeze, the classifier is replaced with new one and the number of prediction in dens layer are set to 7 classes. In addition we apply fine-tuning to improve training and validation accuracy.

Experimental setup

Resnet50 network with more than 24 million trainable parameters learns better features for facial expression classification. The experiments were carried out on Google Colaboratory that provide NVIDIA K80 (0.82 GHZ / 12 GB) which makes training process faster. Hyperparameters tuning is try and error process that effects the accuracy of model. The model with best possible experimental parameters are found as batch size 64, initially at 10 epochs with fine-tuning and learning rate of lr=0.001 for transfer learning. In fine-tuning step the weights of the convolutional layer are unfreeze and the model is trained with a little learning rate of lr=0.0001. Fine-tuning improves the validation accuracy to 86.32%. The training parameters are updated using Adam optimizer. We experimented with different activation functions (ReLu, Sigmoid, and Softmax) but Softmax activation function for fully connected layer with seven (7) outputs works better.

Deep learning model with millions of learnable parameters has incomparable generalization capability when trained at a larger dataset. Images in FER2013 dataset using data augmentation are increased. The images with rotation rate 10, height and width range 0.1, and zoom range 0.1 are rotated and then flipped horizontally.

RESULT AND DISCUSSION

This section presents the results achieved in experiments of ResNet50 on FER2013 dataset for facial expressions recognition. The ResNet50 model with optimal parameters work better for facial expression recognition. We experimented with data augmentation and without data augmentation techniques which shows a remarkable increase in accuracy for augmented data. Table 2. The model is compared with existing research that shows remarkable surpass. Table 3.

Table 2: Results of data augmentation

Data Augmentation	Accuracy
No	78.57%
Yes	86.32%

The accuracy of ResNet50 on validation set of FER2013 dataset is 75.79%, which is comparable with the existing state-of-the-art. The model also show satisfactory generalization results when evaluated on some unseen images from MMA facial expression recognition dataset.

Table 3: Comparison of model with existing research

Work	Year	Method	Dataset	Accuracy
[10]	2022	VGG16	FER2013	67.40%
[2]	2020	Ensemble	FER2013	75.80%

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

[8]	2021	Attentional CNN	FER2013	70.02%
[7]	2022	Deep CNN	FER2013	55.60%
[12]	2021	CNN	FER2013	66.70%
Proposed Model	**2023**	**ResNet50**	**FER2013**	**86.32%**

Evaluation of test images show actual image label and predicted label with prediction accuracy. Fig.6. Images were taken randomly from MMA dataset.

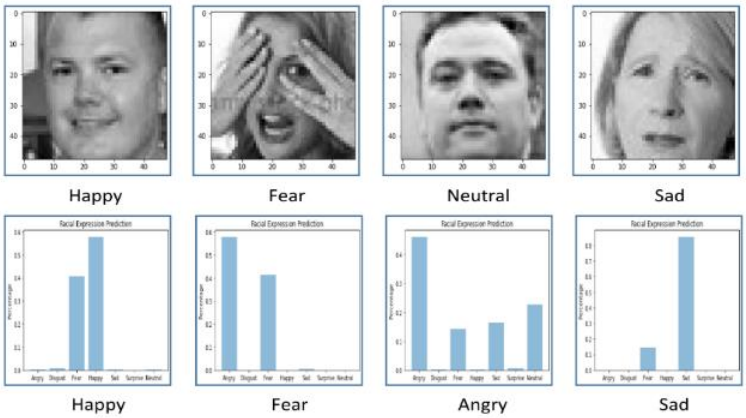

Figure 6: Test results of model on unseen images

CONCLUSION AND RECOMMENDATIONS

This research was conducted to develop a model for Facial expression recognition in frontal images. The method proposed Resnet50 for emotion classification. The model is trained on publically available Dataset FER2013 dataset using transfer learning. Additionally, the model was fine tuned for improving the generalization. In preprocessing, the data augmentation, where images were rotated and flipped, was applied on the dataset. The training with data augmentation and without data augmentation show validation accuracies as 78.57% and 86.32% respectively, improving 7% almost. The study aims to improve the model in future work by cleaning FER2013 and comparing the best deep learning model for facial expression recognition.

REFERENCES

Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. Journal of personality and social psychology, 6(1), 109.

Ekman, P. (1993). Facial expression and emotion. American psychologist, 48(4), 384.

Ahmed, T. U., Hossain, S., Hossain, M. S., ul Islam, R., & Andersson, K. (2019, May). Facial expression recognition using convolutional neural network with data augmentation. In 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 336-341). IEEE.

Naseer, G. J., Basit, A., Ali, I., & Iqbal, A. (2020). Balochi non cursive isolated character recognition using deep neural network. *International Journal of Advanced Computer Science and Applications*, *11*(4).

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

Iqbal, A., Basit, A., Ali, I., Babar, J., & Ullah, I. (2021). Automated Meter Reading Detection Using Inception with Single Shot Multi-Box Detector. *Intelligent Automation & Soft Computing*, *27*(2).

Ge, H., Zhu, Z., Dai, Y., Wang, B., & Wu, X. (2022). Facial expression recognition based on deep learning. *Computer Methods and Programs in Biomedicine*, *215*, 106621.

Naseer-u-Din, & Basit, Abdul & Ullah, Ihsan & Noor, Waheed & Ahmed, Atiq & Sheikh, Naveed. (2022). Brain tumor detection in MRI scans using single shot multibox detector. *Journal of Intelligent & Fuzzy Systems. 43. 1-9. 10.3233/JIFS-219298. Nour, N., Elhebir, M., & Viriri, S. (2020).*

Face expression recognition using convolution neural network (CNN) models. *International Journal of Grid Computing & Applications, 11(4), 1-11.*

Tripathi, M. (2021). FACIAL EMOTION RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK. *Ictact Journal on Image and Video Processing, 12(01).*

Khanzada, A., Bai, C., & Celepcikay, F. T. (2020). *Facial expression recognition with deep learning. arXiv preprint arXiv:2004.11823.*

Gaddam, D. K. R., Ansari, M. D., Vuppala, S., Gunjan, V. K., & Sati, M. M. (2022). Human facial emotion detection using deep learning. In ICDSMLA 2020: *Proceedings of the 2nd International Conference on Data Science, Machine Learning and Applications (pp. 1417-1427). Springer Singapore.*

Minaee, S., Minaei, M., & Abdolrashidi, A. (2021*). Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors, 21(9), 3046.*

Siam, A. I., Soliman, N. F., Algarni, A. D., El-Samie, A., Fathi, E., & Sedik, A. (2022). *Deploying machine learning techniques for human emotion detection. Computational Intelligence and Neuroscience, 2022.*

Hou, C., Ai, J., Lin, Y., Guan, C., Li, J., & Zhu, W. (2022). *Evaluation of Online Teaching Quality Based on Facial Expression Recognition. Future Internet, 14(6), 177.*

Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). *Facial emotion recognition using transfer learning in the deep CNN. Electronics, 10(9), 1036.*

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).*

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-05 | Issue-02 | June-2024

AUTHORS PROFILE (All author profiles are mandatory)

Author-1 Habib Ur Rehman is MS scholar in Computer Science and Information Technology. He did his MCS in 2004 from Department of Computer Science and Information Technology, University of Balochistan, Quetta, Pakistan. He is currently working as I.T Teacher in Government Boys High School Moti Ram Road Quetta, Balochistan Pakistan. His area of interest is predictive modeling using machine learning algorithm. His main focus is on using pertained deep learning models for image recognition and real time video analysis.

Author-2 Dr. Abdul Basit is currently working as Assistant Professor in the Department of Computer Science & IT, University of Balochistan, Quetta, Pakistan. He did his PhD from Asian Institute of Technology, Thailand.
Area of Interest: My research interest includes machine vision and image processing theory and their applications. Some of my project include working on small and portable vehicle such as SUGV (Small Unmanned Aeriel Vehicles) and UAV (Unmanned Aerial Vehicles).