VIDEO SURVEILLANCE IMAGE ENHANCEMENT USING DEEP LEARNING

MUHAMAD FARIS CHE AMINUDIN

UNIVERSITI SAINS MALAYSIA

2019

VIDEO SURVEILLANCE IMAGE ENHANCEMENT USING DEEP LEARNING

by

MUHAMAD FARIS CHE AMINUDIN

Thesis submitted in fulfilment of the requirements for the degree of Master of Science

March 2019

ACKNOWLEDGEMENT

First of all, I would like to thank Universiti Sains Malaysia and School of Electrical and Electronic Engineering that had given me the chance to further my study for Master Degree. I would like to express my gratitude to my supervisor, Assoc. Prof. Dr. Hj. Shahrel Azmin for his kindness and inspiration to guide through my Master journey. The dedication and effort that help me to keep in track within my research scope. This thesis would never have been completed without his continuous supervised.

I would also like to thank the Malaysian Ministry of Higher Education for sponsoring a research grant Fundamental Research Grant (FRGS) No.203/PELECT/6091294 that helps to carry out my studies while being Graduate Assistance.

I would like to thank my family that had given me moral support to pursue my Master. Their love and encouragement helped me to keep doing my study. Thousand of thanks to all my friend that had helped and shared their experience on research skill and knowledge. Thank you for all the support and guidance. I also want to thank all the lecturers, staffs and technicians of School of Electrical and Electronic Engineering that helped me. I would also like to thank the staffs of Institute of Postgraduate Studies for their kind assistance when there is problem regarding candidature and thesis format.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
BPNN	Back Propagation Neural Network
CCTV	Closed-Circuit Television
CLAHE	Contrast Limited Adaptive Histogram
CMPNN	Complementary Neural Network
CNN	Convolutional Neural Network
CPU	Computer Processing Unit
DBN	Deep Belief Network
DLB	Deep Learning Block
DLIE	Deep Learning Image Enhancement
DNN	Deep Neural Network
FSRCNN	Fast Super Resolution Convolutional Neural Network
GPU	Graphical Processing Unit
HE	Histogram Equalization
HR	High Resolution
k-NN	K Nearest Neighbor
LR	Low Resolution

ML Machine Learning

MLP Multilayer Perceptron

NN Neural Network

PNN Probabilistic Neural Network

PSNR Peak Signal to Noise Ratio

RAM Random Access Memory

RBM Restricted Boltzmann Machine

SR Super Resolution

SRCNN Super Resolution Convolutional Neural Network

SSIM Structural Similarity Index Metric

SVM Support Vector Machine

VAE Variational Autoencoder

VCR Video Cassette Recorder

VPU Video Processing Unit

VRAM Video Random Access Memory

PENINGKATAN IMEJ PENGAWASAN VIDEO MENGGUNAKAN PEMBELAJARAN MENDALAM

ABSTRAK

Kamera pengawasan telah menjadi suatu kebiasaan untuk meningkatkan keselamatan kerana kegunaannya dalam merakam video atau gambar untuk digunakan dalam analisis. Kepelbagaian model dan spesifikasi kamera pengawasan mempengaruhi kualiti gambar keseluruhan. Kualiti gambar memainkan peranan penting dalam mengekstrak maklumat penting yang terdapat dalam gambar. Dalam sistem pengecaman wajah, gambar yang berkualiti rendah akan mengakibatkan prestasi sistem terjejas. Oleh demikian, memperbaiki kualiti gambar ketika pra-pemprosesan gambar sebelum proses latihan dan ujian akan menangani masalah ini. Gambar yang beresolusi rendah, kadar cahaya rendah, dan gangguan adalah antara beberapa masalah yang kerap berlaku dalam kamera pengawasan. Untuk menyelesaikan masalah ini, meningkatkan kualiti gambar dengan mengunakan kaedah pembelajaran mendalam dicadangkan dengan melatih rangkaian pembelajaran mendalam untuk meningkatkan resolusi gambar, kontras, dan gangguan tanpa mengubah mana-mana parameter. Untuk mencapai matlamat tersebut, "Deep Learning Image Enhancement" DLIE model dicadangkan. Terdapat dua blok pembelajaran mendalam (DLB1 dan DLB2) dan teknik penggabungan gambar dalam model DLIE yang dicadangkan. Kedua-dua "Deep Learning Block 1" (DLB1) dan "Deep Learning Block 2" (DLB2) yang di cadang adalah untuk menyelesaikan masalah resolusi rendah, kontras, dan gangguan dalam gambar kamera pengawasan. Manakala, teknik penggabungan gambar digunakan sebagai cara untuk menggabungkan DLB1 dan DLB2 sebagai satu sistem. DLB1 menggunakan

"Convolutional Neural Network" (CNN) untuk meningkatkan resolusi gambar dengan menggunakan kaedah "Super Resolution". Super Resolution adalah salah satu daripada algoritma yang membaiki kualiti gambar dengan membina semula gambar beresolusi rendah kepada gambar beresolusi tinggi. Manakala, DLB2 menggunakan "Denoising Autoencoder" untuk penambahbaikan kontras dan pengurangan ganguan dalam gambar sebelum membina semula gambar tersebut. Oleh yang demikian, gambar gelap dan mempunyai bunyi akan ditambah baik kepada gambar lebih bagus. Hasil kedua-dua rangkaian (DLB1 dan DLB2) yang telah dilatih digabung dengan menggunakan teknik gabungan gambar Wavelet untuk memastikan sistem mendapat kualiti gambar terbaik. Gambar yang dibaikpulih dinilai menggunakan "Peak-to-Signal Noise Ratio" (PSNR) dan "Structural Similarity Index" (SSIM). DLB1 menunjukkan peningkatan di dalam kualiti gambar di antara 0.946 hingga 8 peratus, manakala DLB2 menunjukkan bahawa ia mampu membaiki kontras dan mengurangkan gangguan dalam gambar lebih baik daripada teknik konvensional untuk meningkatkan kualiti gambar. Gambar yang ditingkatkan kualiti oleh model DLIE menunjukkan peningkatan jika dibandingkan dengan gambar yang gelap dan mempunyai gangguan. Peningkatan dengan purata minimum 13.3625 dB hingga ke 22.7728 dB berbanding sebelum peningkatan kualiti gambar iaitu dengan purata 9.3940 dB hingga 12.8398 dB.

VIDEO SURVEILLANCE IMAGE ENHANCEMENT USING DEEP LEARNING

ABSTRACT

Surveillance camera had become common in improving security because of its usefulness to capture video and images for analysis. The variation of surveillance camera model and specification affects the overall image quality. Image quality plays a significant role in extracting the prominent information from an image. In a face-recognition system, a bad quality image will affect the performance of the system. Thus, enhancing the image in image preprocessing before training and testing would deal with this problem. The low-resolution, low-exposure, and noises are several problems that occur in surveillance camera. These problems could be addressed by improving the image resolution and enhancing the contrast and reduce the noise of the image without overexposing it. In conventional image enhancement, each approach could only solve one problem at a time and the parameters need to be changed for each problem. This would cause difficulty in developing an automated system. Therefore, in this research work, image enhancement using deep learning approach is proposed. Image enhancement using deep learning utilizes the deep learning network that could automatically improve the resolution, contrast, and reduce noise of the images without changing any parameter. To achieve the goal, Deep Learning Image Enhancement (DLIE) is proposed. There are two deep learning blocks which are Deep Learning Block 1 and Deep Learning Block2 (DLB1 and DLB2) and image fusion in the proposed DLIE model. Both DLB1 and DLB2 are proposed to solve their respective problems, which is low-resolution, low-contrast, and noise. Whereas, image fusion is used as a method to merge DLB1 and DLB2 outputs into one system. DLB1 utilizes convolutional neural network to enhance the low-resolution image using Super Resolution method. Super resolution is one of the algorithms that could improve the image resolution by reconstructing the low-resolution to high-resolution image. On the other hand, DLB2 utilize denoising autoencoder to obtain contrast enhancement and noise reduction before reconstructing the input image to a good quality image. As a result, dark and noise images can be improved to a cleaner. The outputs of both deep learning techniques (DLB1 and DLB2) are then fused together using Wavelet image fusion to get the best image quality while maintaining the capability of both techniques. The enhanced images are evaluated using image quality assessment such as the peak to signal noise ratio (PSNR) and structural similarity index (SSIM). DLB1 shows an improvement ranging from 0.946 to 8 percent, whereas DLB2 shows that it capable of enhancing image contrast and reduces noise in the image better compared to conventional image enhancement method. The enhanced image from the DLIE shows improvement in terms of PSNR compared to the dark and noisy image with minimum average of 13.3625 dB up to 22.7728 dB, compared to before enhancement which averages of 9.3940 dB up to 12.8398 dB.

CHAPTER ONE

INTRODUCTION

1.1 Overview

Surveillance cameras became a common technology that is in use for monitoring and security. Video surveillance helps people to know what is happening without being there and can monitor several places at the same time. There are a lot of applications for the surveillance camera such as in traffic monitoring, video surveillance, criminal recognition, crowd monitoring, etc. Nowadays, all the Closed-Circuit Televisions (CCTV) that are commonly installed are usually digital CCTV instead of analog CCTV (Cermeño et al., 2018). This evolution of technology helped in further video processing and analysis instead of old limitation in analog CCTV. The improvement of machine learning in this ever-expanding age of artificial intelligence has strengthened security by implementing it into security surveillance camera system. There are few existing cameras with embedded deep learning technology that are used for people-counting, heat mapping and queue detection that are used in retail stored (Technology, 2018). There are several concerns of implementing deep learning in real-time application. One of the problems is the processing power required to process all the required information into the network. This would load the system capacity that only has a small processing unit but with development technology such as a video processing unit (VPU) has made deep learning embedded system possible (Strom, 2018). VPU was developed by Intel's Movidus group, and it is a dedicated hardware accelerator for deep neural network that could be used to embed the camera for deep learning. Other common hardware used to solve processing power problem is by using Graphical Processing Unit (GPU) due to its capability to do parallel computing and making huge dataset could be computed easily. With this kind of growth with deep neural network, soon the deep learning will become a staple for CCTV that can be applied depending on its usage.

The qualities of visual data produced from CCTV are varied depending on its hardware such as sensors and lenses. The qualities of visual data are important for further analysis to capture the important information from the images. Due to cost limitation, there are a lot of low-end CCTVs installed, and this type of CCTV produces low visual data quality. The low quality images from the CCTV would affect the analysis, especially for recognition because of the lack of information in the images. Low resolution, exposure, contrast, and noises are the frequent problems that affect the image quality to be low (Loza et al., 2013). This problem would cause the performance of detection and recognition such as for face recognition degraded because of the image will be either too dark or even the image is too pixelated, and the person in the image could not be identified. Thus, image enhancement becomes important to cope with these problems by enhancing the image in the preprocessing stage before doing further process. Image enhancement such as histogram equalization, Gaussian filtering, noise removal and many others are the common approach on enhancing the image for specific case and problem.

Histogram Equalization (HE) is able to improve the contrast of the image by equally distributing the histogram of the image (Cheng and Shi, 2004; Hall, 1974). Thus making the image is able to achieve better contrast by making a wider dynamic range of image gray level. The downside of this process is the noise in the image becomes more visible and will affect the image quality. This downside is overcome by a new approach using contrast limited adaptive histogram (CLAHE) proposed by Pisano et al. (1998). CLAHE fixed the contrast on an image by improving the contrast on the region of interest without amplifying the noise within the image. These are several common methods used in improving the contrast of the image (Meena et al., 2017). All of these conventional methods are commonly used in image enhancement to improve the contrast of the image.

Recently, deep neural network has been used in image enhancement and has achieved significant performance in the image-processing task. Stacked denoising autoencoder has been used to learn important feature from the data and filter the useless data (Fan et al., 2017). The learned feature from the trained network is then used when reconstructing back the image. The convolutional network is also utilized for image enhancement by training the network with a reference image either from some algorithms or human adjustment (Gharbi et al., 2017a). Other image enhancement methods using deep learning are colorization (Iizuka et al., 2016), demosaicking and denoising (Gharbi et al., 2016), portrait matting (Shen et al., 2016) and Super resolution (Dong et al., 2016a).

Super resolution (SR) is one of the image enhancement methods that could help in enhancing a low-resolution image. The amount of important information in a low-resolution image is usually very bad. Blockiness in the image can visibly be seen when the image is further zoomed in. Similarity-based (Yoo et al., 2016), dictionary-based (Li et al., 2016) and learning based (Dong et al., 2016c) are some of SR approaches in enhancing the image resolution. SR works by reconstructing a low resolution (LR) image into a high resolution (HR) image on high-resolution planes by smoothing and upsampling the LR images. As a result, the HR image appearance will be visually improved with less noise in the image.

On the other hand, image fusion is one of the processes that could help in improving image quality when there are two or more images that need to be combined together to obtain better image quality. This process helps in storing important information in each image and combining it into one image that contains all the information (Nikolov et al., 2001). There are various image fusion methods such as wavelet image fusion, Laplacian pyramid, simple pixel averaging, principal component analysis and intensity hue saturation.

Image quality problems in surveillance system such as low-resolution, low contrast, and noise are known to affect the performance of image analysis in surveillance system. The conventional image enhancement required parameters need to be changed for each different problems. Most of the existing researches on image quality enhancement has complex algorithm and required many parameters tinkering. Thus, making it hard to create a universal parameter to enhance the surveillance image from various CCTV qualities. Deep learning addresses this problem by training the network with its required parameters and appropriate architecture. Deep learning can understand more complex image structures using its learning capability compared to conventional image enhancement method. After the network is trained, the network can be used as it is without any further parameter tinkering for enhancing their respective problems.

1.2 Problem Statement

There are several common problems within the video surveillance system that making further analysis become more complicated. The problems are:

1. The qualities of the surveillance camera vary depending on its hardware such as sensors and lenses. The manufacturer of the surveillance camera produced different specification of the surveillance camera with price variation. The variation of the surveillance camera cause differences in image quality and would interfere with further image analysis. Problems such as low resolution in the image from the low-end quality camera would really reduce the image quality. Therefore, an image enhancement technique that could improve the image quality regardless of the camera quality to a certain quality is required.

2. The position of a surveillance camera could affect the image quality. If the surveillance camera is installed in a place without sufficient lighting, the image will have noise and would affect the contrast of the image. Without enough light, camera sensor will have a problem capturing a good-quality image. Thus, an image enhancement that could improve the brightness of the image and remove noise that visible on the image is required.

3. The conventional image enhancement method can only solve one problem at a time. This would make the system more complex and could not tackle more than one problem at a time. Therefore, an intelligent image enhancement model that could solve several problems at a time could help in this situation.

1.3 Research Objectives

The main objectives of this research are as follows:

1. To design and develop a deep learning-based image enhancement technique to improve the low-resolution video surveillance image.

2. To develop a deep learning-based image enhancement technique to improve low-contrast and reduce noise in video surveillance image.

3. To solve the three problems simultaneously in one system.

1.4 Scope of Thesis

This thesis covers the following scope:

1. This research only uses still grayscale images from surveillance cameras to reduce training time and complexity in deep learning algorithm. This also reduced other problem such as size of GPU memory that could be trained on.

2. The architecture used in this research is only focusing on autoencoder and convolutional neural network as it fits the criteria for developing image enhancement algorithm in deep learning framework.

3. The database used in this research is SCFace database because the database mimics the real-world surveillance camera. This is the only database that uses real surveillance camera for capturing the image. The database contains five different cameras with three different distance when capturing each subject.