

COUGH NOISE SUPPRESSOR FOR VIDEO CONFERENCING  
(AN IMAGE PROCESSING APPROACH)

CHUNG SOON THIEN

This report is submitted in partial fulfillment of the requirements for the award of the  
Bachelor of Electronic Engineering (Communication Wireless) With Honours

Faculty of Electronics and Computer Engineering  
Universiti Teknikal Malaysia Melaka

April 2011

PENEKAN BUNYI BATUK UNTUK SIDANG VIDEO  
( PENDEKATAN PEMROSESSAN IMEJ )

CHUNG SOON THIEN

Laporan ini dikemukakan untuk memenuhi sebahagian daripada syarat penganugerahan Ijazah Sarjana Muda Kejuruteraan Elektronik (Komunikasi Wayarles) Dengan Kepujian

Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer  
Universiti Teknikal Malaysia Melaka

April 2011



**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**  
**FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER**

**BORANG PENGESAHAN STATUS LAPORAN  
 PROJEK SARJANA MUDA II**

**Tajuk Projek** : COUGH NOISE SUPPRESSOR FOR VIDEO CONFERENCING  
 (AN IMAGE PROCESSING APPROACH)

**Sesi Pengajian** : 2010/2011

Saya **CHUNG SOON THIEN**

mengaku membenarkan Laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (  $\checkmark$  ) :

**SULIT\***

\*(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

**TERHAD\*\***

\*\* (Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

**TIDAK TERHAD**

Disahkan oleh:

\_\_\_\_\_  
 (TANDATANGAN PENULIS)

\_\_\_\_\_  
 (COP DAN TANDATANGAN PENYELIA)

Tarikh : April 2011

Tarikh : April 2011

“I hereby declare that this report is result of my own effort except for works that have been cited clearly in the references.”

Signature :.....  
Author :CHUNG SOON THIEN  
Date :.....

“Saya akui laporan ini adalah hasil kerja saya sendiri kecuali ringkasan dan petikan yang tiap-tiap satunya telah saya jelaskan sumbernya”

Tandatangan :.....  
Nama Penulis :CHUNG SOON THIEN  
Tarikh :.....

“I hereby declare that I have read this report and in my opinion this report is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering (Communication Wireless) with Honours”

Signature :.....  
Supervisor's Name : Dr. LIM KIM CHUAN  
Date :.....

“Saya akui bahawa saya telah membaca laporan ini dan pada pandangan saya laporan ini adalah memadai dari segi skop dan kualiti untuk tujuan penganugerahan Ijazah Sarjana Muda Kejuruteraan Elektronik (Komunikasi Wayarles) Dengan Kepujian”

Tandatangan :.....  
Nama Penyelia :Dr. LIM KIM CHUAN  
Tarikh :.....

Special dedicated to my supervisor, my parents, family and fellow friends, who had strongly encouraged and supported me in my entire journey of learning.



## ACKNOWLEDGEMENT

I appreciate everyone that helped me when I faced problem in writing the matlab source code in this project. Finally I finish implement my Bachelor Degree Project 2 (PSM 2) and write my PSM report. I also want to say thanks to advice and support from my supervisor, Dr. LIM KIM CHUAN. He helped me a lot and spare times to teach me in, matlab coding, proposing title, writing report, advise for project presentation and INOTEK preparation. I want to thank to my parents, and my siblings who always support me in the entire journey of my project. Lastly, I appreciate Faculty of Electronics and Computer Engineering (FKEKK) for giving me a chance to implement my project as part of the Bachelor program in Universiti Teknikal Malaysia Melaka (UTeM).

## ABSTRACT

Video conferencing involved real-time video communication between users present at two different locations. The video, consists of both the image frames and voice, are captured and compressed by a computing unit before the transmission. Coughing produced unpleasant sound while communicating through video conferencing and human being tend to cover their mouth with hand to reduce the volume of the produced unpleasant sound. The possibility of detecting the action of covering the mouth and automatically mute the input audio to suppress the noise produced by coughing is explored in this project.

The process of detecting the action of using the hand to cover the mouth, with an image processing approach, can be divided into three steps, namely the human skin pixel classification, skin region extraction and the mouth detection within the extracted skin region. Gaussian Mixture Model (GMM), Bayesian method and Multi Layer Perceptrons (MLP) are three well known methods for skin classification. GMM has been chosen due to the requirement of less memory space compare to Bayesian method and lesser computational cost compared to MLP. Experiment should that reliable human skin classification is needed to produce good result in mouth detection within the classified skin region.

Connected component labeling, one of the general region extracting algorithm provided by most of the image processing toolkit, was used to extract the skin region formed by the classified skin pixels. Making use of the extracted region, the region with seven hollows (two eye browns, two eyes, two nostrils, one mouth) is subsequently detected as “face with mouth detected”. The region with only six hollows is detected as “face with mouth not found” and a signal is subsequently send to the system to mute the

input audio. This allows the coughing noise to be automatically suppressed. The system will send a signal to unmute the input audio when “face with mouth detected” is sensed in the future image frame.

Making use of the three identified component, a software prototype has been designed and implemented. Result show that the implemented system is only able to detect both the “face with mouth detected” and “face with mouth covered by hand” in an ideal situation with around 5 frames per second.

## ABSTRAK

Sidang video melibatkan penghantaran data video secara berterusan diantara dua pengguna atau lebih yang berada di tempat berlainan. Video mengandungi bingkai imej dan bingkai suara, ditangkap dan dimampatkan oleh unit pengkomputeran sebelum proses penghantaran. Batuk menghasilkan bunyi yang tidak menyenangkan semasa berbual melalui sidang video dan manusia mempunyai tabiat untuk menutup mulut dengan tangan supaya mengurangkan nade yang tidak menyenangkan itu. Kebolehan mengesan tingkah laku untuk menutup mulut dan secara automatik menghalang kemasukan audio untuk mengurangkan kesan hingar yang dihasilkan oleh batuk telah dikaji dalam projek ini.

Proses yang melibatkan mengesan tingkah laku menutup mulut dengan pendekatan pemrosesan imej boleh dibahagikan kepada tiga bahagian, iaitu pengelasan piksel kulit manusia, pengekstrakkan kawasan kulit dan juga pengesan mulut dalam pengekstrakkan kawasan kulit. Gaussian mixture model (GMM), Bayesian model, dan Multi Layer Perceptrons (MLP) adalah tiga kaedah yang diketahui untuk pengelasan kulit. GMM dipilih berdasarkan keperluan memori yang sedikit berbanding dengan kaedah Bayesian dan kos pengkomputeran yang rendah berbanding dengan kaedah MLP. Eksperimen telah menunjukkan pengelasan kulit manusia yang tepat amat diperlukan untuk menghasilkan keputusan yang baik pada pengesan mulut dalam pengelasan kawasan kulit.

Penglabelan sambungan komponen adalah satu algoritma pengekstrakkan kawasan yang selalu digunakan dan disediakan dalam kesemua set perisian pemrosesan imej digunakan untuk pengekstrakkan kawasan kulit yang telah dibentuk oleh pengelasan piksel kulit. Dengan menggunakan pengekstrakkan kawasan, dalam kawasan terdapat

tujuh rongga (dua rongga untuk alis mata, dua rongga untuk mata, dua rongga untuk lubang hitung, satu rongga untuk mulut) dikesan sebagai “muka dan mulut dikesan” kawasan yang hanya terdapat dua rongga akan dikesan sebagai “muka dengan mulut tidak dapat dikesan” dan isyarat akan dihantar kepada sistem untuk menghalang audio masukan secara berterusan, ini membolehkan bunyi batuk dituhankan secara automatik. Sistem akan menghantar isyarat ke audio masukan untuk membenarkan audio masukan semula apabila mengesan “muka dan mulut dikesan”.

Dengan menggunakan tiga komponen yang telah dikenalpasti, satu prototaip perisian telah direka dan dibangunkan. Keputusan menunjukkan pelaksanaan sistem hanya boleh mengesan “muka dan mulut dikesan” dan “muka dengan mulut tidak dapat dikesan” dalam keadaan unggul dengan tahap kelajuan 5 bingkai imej sesaat.

## CONTENTS

CHAPTER	CONTENT	PAGE
	<b>TITLE</b>	i
	<b>CONFESSION</b>	v
	<b>DEDICATION</b>	ix
	<b>ACKNOWLEDGEMENT</b>	x
	<b>ABSTRACT</b>	xi
	<b>CONTENTS</b>	xv
	<b>LIST OF FIGURES</b>	xviii
	<b>LIST OF TABLES</b>	xx
	<b>LIST OF ABBREVIATIONS</b>	xxi
<b>I</b>	<b>INTRODUCTION</b>	
	1.1 Introduction Project	1
	1.2 Problem Statement	2
	1.3 Objectives	3
	1.4 Scope of Project	3
	1.5 Thesis Outline	4
<b>II</b>	<b>LITERATURE REVIEW</b>	
	2.1 Skin Classification	5

2.2	Data Set for Gaussian Mixture Model (GMM)	7
2.3	Color Space	9
<b>III</b>	<b>THRESHOLD METHOD FOR SKIN CLASSIFICATION</b>	
3.1	General Flow Chart for Cough Noise Suppressor System	17
3.2	Thresholding and Color Pixel Calculation Flow Chart	18
3.2.1	Capture real time Image	19
3.2.2	Thresholding	22
3.2.3	Color Pixel Calculation Algorithm	28
<b>IV</b>	<b>SKIN CLASSIFICATION</b>	
4.1	Skin Segmentation Flow Chart	31
4.1.1	Webcam	33
4.1.2	The value of A Data set	34
4.1.3	Acquired Frames and Calculate The Covariance of RGB color	34
4.1.4	Skin GMM	37
4.1.5	Non-Skin GMM	38
4.1.6	Comparison Between the Probability of Skin and The Probability of non-skin	39
<b>V</b>	<b>MOUTH DETECTION</b>	
5.1	Mouth Detection Flow Chart	43
5.1.1	Label the connected object	45
5.1.2	Define the coordinate for the connected object	47
5.1.3	Define Hollow within The Skin Region	49

5.1.4 Logic Suppressor (Mute Mic)	51
-----------------------------------	----

**VI****CONCLUSION**

6.1 Conclusion	52
----------------	----

6.2 Future Work	54
-----------------	----

<b>REFERENCE</b>	55
------------------	----

<b>BIBLIOGRAPHY</b>	56
---------------------	----



## LIST OF FIGURES

NO	TITLE	PAGE
2.1	Color spectrum seen by passing white light through a prism	9
2.2	Wavelength comprising the visible range of the electromagnetic spectrum	10
2.3	Secondary color	11
2.4	CIE chromaticity diagram	12
2.5	RGB color model	13
2.6	HSV color model	14
3.1	General Flow Chart for Cough Noise Suppressor System	17
3.2	Thresholding and color pixel calculation Flow Chart	18
3.3	Matlab command window result for imaqhwinfa syntax.	19
3.4	Detail of the installed adaptor name winvideo	20
3.5	The RGB image captured in the office	21
3.6	Three dimensions for an RGB image	23
3.7	Manually skin segmentation by using paint tool.	23
3.8	Max and min values for RGB color.	24

3.9	Skin pixel in binary image	24
3.10	RGB values for the skin pixels.	25
3.11	Output image for the skin region	25
3.12	The mask of all pixels in binary „1“ minus the mask of the skin region	26
3.13	The result for 1-skin region	26
3.14	The RGB values for non-skin pixels	27
3.15	The output image for non-skin pixels	27
3.16	Example of acquired the data of mean and covariance by using mlcgmm for one dimension data.	29
4.1	Skin Classification Flow Chart	32
4.2	The matlab window result for syntax size.	35
4.3	Image in $m \times n$ pixels convert to $1 \times mn$ pixels.	36
4.4	Result for skin covariance matrix and non-skin covariance matrix	37
4.5	Image in $1 \times mn$ pixels convert to $m \times n$ pixels	40
4.6	Output for skin segmentation	40
5.1	Mouth Detection Flow Chart	43
5.2	Image for the skin segmentation	45
5.3	Object's connectivity with 4 connected pixels and 8 connected pixels	46
5.4	Matlab command window result for labeling connected pixels.	46
5.5	Matlab command window result for the coordinate of connected object number 2	47
5.6	Matlab command window for the mask binary image with „0“ value	48
5.7	Matlab command window result for the new binary image of component label number 2 and number 3	48
5.8	Sample for the binary face region image and the result for the Eulerno	49
5.9	The result of the connected component in face image is 1.	49
5.10	Matlab command window result for the numofholes	50
5.11	The matlab command window result for Inew1	50
5.12	Matlab command window result when numofholes =2	51

**LIST OF TABLES**

<b>NO</b>	<b>TITLE</b>	<b>PAGE</b>
2.1	Means, covariances and weights for mixture of gaussian skin color models.	7
2.2	Means, covariances and weights for mixture of gaussian non-skin color models.	8

## LIST OF ABBREVIATIONS

GMM	-	Gaussian Mixture Model
MLP	-	Multi Layer Perceptrons
RGB	-	Red, Green, Blue color space
HSV	-	Hue, Saturation, and Value color space

**LIST OF APPENDIXES**

<b>NO</b>	<b>TITLE</b>	<b>PAGE</b>
A1	Matlab source code for capture imge	56
A2	Matlab source code for thresholding	57
A3	Matlab source code for skin pixel calculation	58
B1	Matlab source code for skin classification	60
C1	Matlab source code for mouth detection	62
C2	Matlab source code for mouth detection with binary image	63



## CHAPTER I

### INTRODUCTION

#### 1.1 Introduction Project

Coughing is a common action that appears in our daily life. Normally, human did not care so much about coughing, but the noise produced by coughing in a video conferencing will be amplified and disturb the participants of the conference. Coughing is a sudden and often repeatedly occurring reflex action started by stimulation of sensory nerves in the lining of the respiratory passage and there is a short intake of breath and the larynx. This unpleasant sound come from coughing will make human feel unconvinced where human's ear did not suitable to hear this kind of noise and the ear will lose their ability to hear if keep exposed to this noise for a long period.

Coughing produced unpleasant sound while communicating through video conferencing and human being tend to cover their mouth with hand to reduce the volume of the produced unpleasant sound. Making use of both the common reaction produced by the human being and the camera used during the video conferencing, it is possible to process the image frames captured by the camera to detect the coughing and mute the input audio automatically.

When a user covers his mouth with hand to reduce the volume of noise generated by coughing, his mouth will be blocked by the hand and hence cannot be seen in the

image captured by the camera. An image is formed by a two dimensional array of pixel elements. The type of values stored in the pixels will be different depend on the type of images produced by the camera. The type of value stored in each of the pixels of a color image will be the color information representing the captured information. A lip with red color will produce a region of pixels (depend on the distance of the lip to the camera, the nearer the lip to the camera, the larger is the region) with red pixel values in the captured image. A simple thresholding algorithm can be applied to detect the present of mouth (region of red pixels) in the captured image. However, object with similar color within the view point of the camera will be falsely detected as mouth too.

## **1.2 Problem Statement**

To prevent the coughing sound to be heard by other people at other side through the video conferencing, the input audio is suppressed while one of them is coughing. An image processing approach is employed to detect the action of coughing (covering the mouth with hand) and give a signal to the video conferencing system to suppress the input audio.

A simple color thresholding method fails to detect the present of lips if object with similar color is around the view point of the camera. Since the mouth is located within the face region, the spatial information shall be made used to detect the present of mouth.

The process of detecting the face region required firstly classified each of the image pixels as “skin” or “non-skin” and subsequently groups all the nearby similar pixels into regions. Color space is used for skin classification where the human’s skin pixels are in color. A lot of color spaces are widely use in image processing. RGB and HSV are well known color space. Due to the time constraint, the RGB color space is used in this project.

Thresholding is a simple method that can be used for the skin classification. Using a valid skin sample, every pixels of an image can be decided easily as skin pixel or non-skin pixel. However, with the growing of the number of skin sample, the process