Volume 4, No. 5
October 2016

Advances in Image And Video Processing

ISSN: 2054-7412

Society for Science and Education - United Kingdom

TABLE OF CONTENTS

EDITORIAL ADVISORY BOARD

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone's views or feelings etc.

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
ISSN 2054-7412

Advanced Technique for Improved Mobile Multimedia Communication Services

U. Ukommi, M. Uko and U. Ekpe
uukommi@yahoo.com

ABSTRACT

The proliferation of smartphones is associated with the development of multimedia applications. Mobile multimedia applications involve audio, data, speech, image, video processing and distribution of over mobile platform. However, compressed media packets are vulnerable to channel errors, thus making it difficult to sustain good perceived video quality performance within limited resources. In this research work, the weight length and impact of different video packets are analyzed. Based on the analysis an advanced technique to enhance mobile multimedia communication services is proposed. The technique involves cross-layer optimization framework, utilizing Network Abstraction Layer Units Length and flexibility of IP-based mobile network in exchanging network information for error protection of sensitive media packets. The simulation results carried out with compatible multiple media streams of different priority levels and Network Abstraction Layer Units Length show overall significant improvement in the received media services.

Keywords: Audio, communications, data, multimedia, network, quality enhancement, speech, video.

1 Introduction

The demand for mobile multimedia services is gradually increasing. The fear is that network level architecture and the air interface do not have sufficient capacity and flexibility to deliver real-time multimedia services at an acceptable quality of service level. Furthermore, the network operators and multimedia users are in need of optimized applications for distribution and consumption of such services including, mobile video services, emergency services for remote consultation, scene of crime work, virtual universities for remote learning, the security industry for telesurveillance, video telephony, business video conferencing, healthcare experts for remote diagnosis and monitoring. In these multimedia applications, video plays significant role. However, these multimedia services exert pressure on the limited mobile network resources due resource constraints and greater demand of improved multimedia services. Moreover, compressed media stream is susceptible to channel distortion due to certain factors including fading, interference, pathloss. These factors affect the performance of these applications. Media encoding algorithm supports error-resilient features such as data partitioning, intra update, slice interleaving for robustness of media stream over error prone channels. However, it is clear that source coding is no longer simply an issue of optimizing rate-distortion characteristics is not enough to combat the impact of channel distortions on received video quality, hence requires advanced protection technique to mitigate impact of channel errors and improve quality performance of received multimedia applications. Traditionally, channel errors can be controlled by existing technologies such as Automatic Re-transmission on Request approach where the corrupted video packets are retransmitted in response to receiver request. However, Automatic

DOI: 10.14738/aivp.45.2009
Publication Date: 20th October, 2016
URL: http://dx.doi.org/10.14738/aivp.45.2009

Re-transmission on Request incurred delays in process of retransmission of loss video packets. Hence, Automatic Re-transmission on Request is not suitable for delay sensitive video applications such as live football match playout and car racing videos. Channel coding such as Forward Error Correction maybe employed in video communication system to enhance the reliability of transmitted video streams over error prone channel. In Forward Error Correction, the additional video packets (redundancy) for protection incur more bandwidth requirement and delays. Advancement in mobile communication system has made it possible to exploit adaptive modulation scheme in improving the quality of video transmission over error prone channel. Several applications of adaptive modulation scheme are found in the literature [2] [3] [4], where the modulation parameters are adapted based on the channel conditions. In addition to the review of existing multimedia distribution technologies, improving the quality performance of mobile multimedia communication services based on the systematic adaptation of media packets is presented in this paper.

2 The Proposed Technique

The proposed advanced technique for improved multimedia communication services is discussed in this section. The aim of is to enhance the quality performance of multimedia services over mobile channel over constrained mobile network resources. In the proposed technique, the adaptation of media packets is based on the sensitivity of the media content to channel errors. The technique adopts cross-layer optimization framework which utilizes Network Abstraction Layer Units Length and flexibility of IP-based mobile network in exchanging network information for error protection of sensitive media packets. The media packets with significant amount of motion are highly prioritized compared to media contents with relative low amount of motion. In the proposed technique, the Network Abstraction Layer Unit (NALU) [1] of media streams with high motion characterization is prioritized and NALU adapted based on the intensity of the motion in the content. The NALU of media streams of low motion characterization are equally adapted differently. However, in order to avoid unbearable transmission overheads the NALU of media streams of highly prioritized packets are made relatively smaller compared to the NALU of media streams that is less sensitive to channel errors.

3 System Design

The system design of the proposed technique is presented in Figure 1. The system design consists of source coding, transmission and receiving chains. The source coding includes capturing of scene using video camera, filtering and encoding process. The source encoding involves removal of redundancies using algorithm. The transmission section involves channel coding and the receiving section consists of decoder and display unit. Figure 1 presents the proposed system architecture.

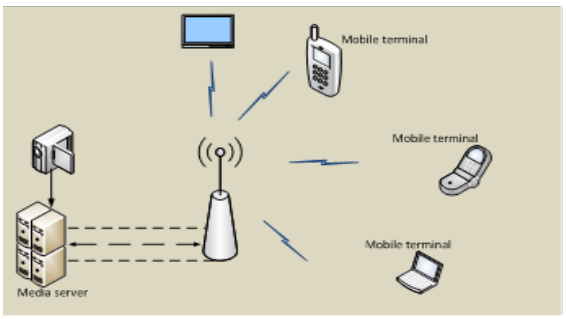


Figure 1: Proposed System

Mobile multimedia applications such as video involves capturing of natural scene by video camera, encoding by media compression algorithms and distribution of the compressed media streams over a wireless channel. The encoding block performs media compression function by exploiting redundancies in video sequence and application of various algorithms to enhance robustness of the media streams. In the proposed technique, the transmission process significantly depends on the media packets sensitivity to channel errors, resource constraints and channel characteristics. Mobile multimedia communication is more challenging due to the limited bandwidth availability and high bit error rates capable of causing quality degradation on the received media performance. The transmitted media streams are processed at the receiver. The reconstructed media stream is processed and displayed on the receiving device. More details on multimedia communication including digital media compression, signal processing and decoding are discussed in the literature [5][6][7]. The model for estimating the sensitivity of media stream in terms of motion characterization is discussed in the literature [8]. However, the spatial and temporal resolutions of the video sequence and the number of frames in the test media stream are also taken into consideration in the simulation process.

4 Experimental System Configuration

The experimental process to measure efficiency of the proposed technique is discussed in this section. The experimental system configuration and simulation are performed to evaluate the proposed technique. In the system configuration, the content dynamic characterization is analyzed using optical flow algorithm of Lucas and Kanade [9]. Two standard test media sequences characterized with high and low motion characterizations are analyzed using the algorithm. The source coding is modeled using H.264/AVC reference software [10]. Simulated wireless channel model is used in the experimental work [11]. In the experiment, the media server stores different media contents of diverse motion characterization. The media encoder algorithm performs encoding and systematic packetization of NALU for improved transportation of media streams. The performance of the proposed technique is tested with two standard test media sequences: Football and Akiyo test media sequences, representing different types of media content services. The media source coding parameter setting include: Group of Picture GOP size of 8, frame rate of 30fps. Common Intermediate Format. Each test media sequence has a total number of 900 frames. The received media streams are processed using H.264/AVC reference software. The channel performance is carried out with pre-simulated error patterns composed of traces of different Signal-to-Noise Ratio for different modulation schemes. The data slot error patterns are obtained by comparing the data bits within original data slot to the transmitted data slot. If there is any bit error within the data slot, it is then declared as an error. More details on path loss, fading and wireless network channel are available in the literature [13] [14].

The performance of the proposed scheme is measured using Peak-Signal-to-Noise-Ratio model. Peak Signal-to-Noise Ratio (PSNR) measures video quality by correlating the maximum possible value of the luminance and the mean squared error (MSE). The overall media quality performance is obtained by averaging the PSNR values throughout the video sequence. Higher PSNR values indicate better quality. Although, PSNR is not the most reliable metric of video quality assessment, it is employed in the research due to its less complexity, ease in calculation and widely usage for video quality assessment.

5 Results and Discussions

The quality performance of the proposed technique was tested with two standard media sequences in Common Intermediate Format. The tested media sequences include standard Football, and Akiyo test sequences. In the experiment, pre-encoded media streams are transmitted to the mobile terminal

through the wireless simulator. The simulations were repeated 15 times to obtain stable results. The results are obtained by averaging the PSNR video quality performance values. Figure 2 presents the quality performance carried out with soccer test media sequence in terms of PSNR(dB).

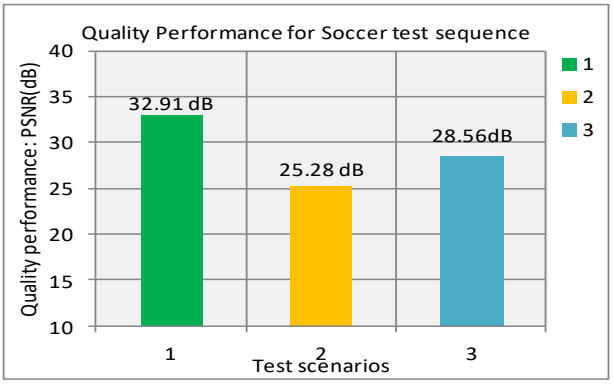


Figure 2: Quality Performance for Soccer test sequence.

Figure 2 presents the test results of the proposed technique performed carried out with soccer test media sequence, under three different scenarios. As shown in Figure 2, the test scenario 1 was performed under error free channel. This was necessary in order to verify the efficiency of the system at various test conditions. It is shown that the quality performance in test scenario 1 recorded the best performance in terms of quality improvement. However, it is noted that test scenario 3 outperformed the quality performance recorded in test scenario 2 with a variation of 3.25dB gain. This significant performance is achieved within equal limited resource network constraints. However, the improvement in the quality performance is a result of systematic adaptation of the NALU base on the sensitivity of the media packets. Figure 3 presents the assessment of test results carried out with Akiyo test media sequence.

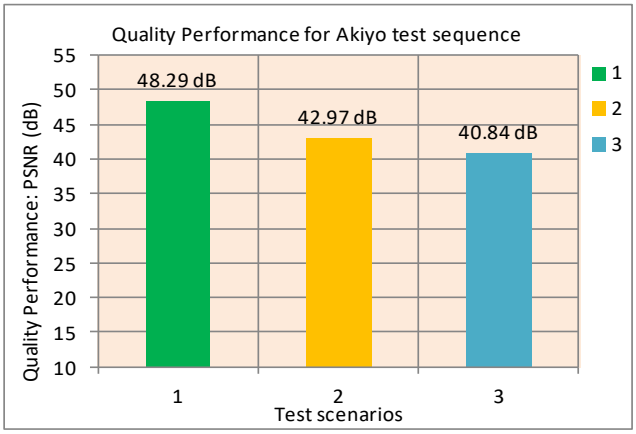


Figure 3: Quality Performance for Akiyo test sequence

Comparing the results obtained under three different test conditions. It has been observed that media quality performance obtained under scenario-1 outperforms that of test scenario-2 and scenario-3 respectively. This is because the test scenario 1 was performed under error free channel conditions. Further observation from Figure 3, shows that test scenario 2 performs better compared to test scenario 3 with Akiyo test media sequence. Based on the observations from Figure 3, at the same source bitrates, media quality performance at scenario 2 shows significant improvement compared to the quality performance scenario 3. Thus, media packets with high sensitivity to channel errors

transmitted through the wireless channel were delivered without much distortion much corrupted packets. It has also been observed that the quality performance of Akiyo test sequence with low priority performed better at the three test conditions. This is due to the fact the low sensitive media streams to channel errors experience negligible distortion at the network level. However, it has been observed that media distribution using the proposed technique improves the overall received quality performance of the tested media streams. Thus, the technique is capable of improving the quality of mobile multimedia communication services.

6 Conclusion and Future Work

Advanced technique for improved mobile multimedia communication services is discussed. The paper investigated the existing technologies for multimedia communication and proposed a technique capable of improving the quality of multimedia communication services over mobile network. The advance technique systematically adapt the NALU of the media packets based on the sensitivity of the media content. The contents with high sensitivity to channel errors are packetized uniquely compared to the media packets of low sensitivity to channel errors. The proposed advance technique saves the limited wireless network resources through intelligent adaptation of the NALU based on media stream error sensitivity. Test results recorded improvement in the overall received media quality performance compared to the conventional approach. Future work investigates further more advanced techniques to improve the quality performance of multimedia communication services.

ACKNOWLEDGMENTS

The authors are grateful to the participants who contributed to the successful conclusion of the research work. The contribution of Akwa Ibom State University to the research work is also acknowledged. However, the role of the publisher in reviewing process is appreciated. Thank you for the support.

REFERENCES

[1] S. Kumar, L. Xu, M.K. Mandal and S. Panchanathan, “Error Resiliency Schemes in H.264/AVC Standard” Journal of Visual and Image Representation, Elsevier, August 2005.

[2] A. J. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *Communications, IEEE Transactions on,* vol. 46, pp. 595-602, 1998.

[3] W. T. Webb, "The modulation scheme for future mobile radio communications," *Electronics and Communication Engineering Journal,* pp. 167-176, August 1992.

[4] S. Sampei, "Rayleigh Fading Compensation for QAM in Land Mobile Radio Communications," *IEEE Transactions on Veh. Tech.,* vol. 42, pp. 137-147, 1993.

[5] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 13, pp. 560-576, 2003.

[6] A.H.Sadka, "Compressed Video Communications," *John Wiley & Sons, Limited, England,* 2002.

[7] I. E. G. Richardson, "H.264 and MPEG-4 Video Compression," *John Wiley and Sons Limited. West Sussex, England,* 2003 2003.

[8] G. Nur, S. Dogan, H. Kodikara Arachchi and A.M. Kondoz, “Impact of Depth Map Spatial Resolution on 3D Video Quality and Depth Perception”, Processings of the 4th IEEE 3DTV Conference, Tampere, Finland, 7-9 June 2010.

[9] D. Fleet and Y. Wiess, "Optical Flow Estimation" *Handbook of Mathematical Models in Computer Vision, Springer,* 2006.

[10] ITU-T and ISO/IEC, "H.264/AVC JM reference Software, 2004.

[11] L. Hanzo, P. Cherriman, and J. Streit, "Wireless Video Communications," *Second Edition, IEEE Press, New York, United States of America,* 2001.

[12] M. Vranjes, S. Rimac-Drlje, and K. Grgic, "Locally averaged PSNR as a simple objective Video Quality Metric," *ELMAR, 2008. 50th International Symposium*, pp. 17-20.

[13] ROHDE and SCHWARZ, "Mobile WiMAX MIMO Multipath Performance Measurements," *WiMAX Forum,* 2010.

[14] M. Wittmann, J. Marti, and T. Kurner, "Impact of the power delay profile shape on the bit error rate in mobile radio systems," *IEEE Transactions on Vehicular Technology,* vol. 46, pp. 329-339, 1997.

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
ISSN 2054-7412

Visual Interface to Speech-Cue Representation Coding

[1]Ibrahim Patel, [2]Raghavendra Kulkarni, and [3]Y Srinivasa Rao
[1]Dept. of ECE B.V. Raju Inst. of Tech., Narsapur Medak,(T. S) India;
[2]Dept. of ECE K.M.C.E&T, Devarkonda under JNT University, Hyderabad, T. S
[3]Department of Instrument Technology, Andhra University, Visakhapatnam, A. P,
ptlibrahim@gmail.com; srinniwasarau@gmail.com; raghavendrakulkarni444@gmail.com

ABSTRACT

There have being great efforts made in the development of automated Instrumentation system for speech recognition (AISR) to provide a two-way communication between deaf and vocal people. This system performance achievable with the output of current real-time speech recognition systems would be extremely poor relative to normal speech reception. An alternate application of AISR technology to aid the hearing impaired would derive cues from the acoustical speech signal that could be used to supplement speechreading. We propose a study of highly trained receivers of speech signal that indicates that nearly perfect reception of everyday connected speech materials can be achieved at near normal speaking rates. To understand the accuracy that might be achieved with automatically generated cue symbols for visual representation. The system uses (HMM) for recognition of voiced data & Euclidian distance approach for sign language. The proposed task is a complementary work to the ongoing research work for recognizing the finger movement of a vocally disabled person to speech signal called. A New communication Paradigm: "Action-To-Speech"

Keywords: AISR, Speech recognition, HMM, vocally disabled, communication gap, speech-processing, cue-symbol.

1 Introduction

Humans know each other by conveying their ideas, thoughts, and experiences to the people around them. There are numerous ways to achieve this and the best one among the rest is the gift of "Speech". Through speech everyone can very convincingly transfer their thoughts and understands each other. It will be injustice if we ignore those who are deprived of this invaluable gift. The only means of communication available to the vocally disabled is the use of "Sign Language". There are approximately 10 million (8.487%) deaf people in India and nearly 1.25 billion persons with hearing impairments and close to a million who are functionally deaf in the United States. Without Assistive Technologies, there is no possibility for the hearing impaired to recognize sounds efficiently. Medical or surgical solutions such as cochlear implants may not always be possible. Using sign language they are limited to their own world. This limitation prevents them from interacting with the outer world to share their feelings, creative ideas and Potentials.

Another problem is that very few people who are not themselves deaf ever learn to Sign language. This further increases the isolation of deaf and dumb people from the common society. Technology is one way to remove this hindrance and benefit these people. Several researchers have explored these possibilities and have successfully achieved finger spelling recognition with high levels of accuracy. But progress in the recognition of sign language, as a whole has various limitations in today's applications.

DOI: 10.14738/aivp.45.2010
Publication Date: 20th October, 2016
URL: http://dx.doi.org/10.14738/aivp.45.2010

Various systems and algorithms were proposed for the recognition of sign language. A system called "Boltay Haath" is developed to recognize "Pakistan Sign Language" (PSL) for vocally disabled peoples at Sir Syed university of Engineering and Technology. The Boltay Haath project aims to produce sound matching the accent and pronunciation of the people from the sign symbol passed. A wearing Data Glove for vocally disabled is designed, to transform the signed symbols to audible speech signals using gesture recognition. They use the movements of the hand and fingers with sensors to interface with the computer. The system able to eliminate a major communication gap between the vocally disable with common community.

2 State-of-The-Art

Humans know each other by conveying their ideas, thoughts, and experiences to the people around them. There are numerous ways to achieve this and the best one among the rest is the gift of "Speech". Through speech everyone can very convincingly transfer their thoughts and understands each other. It will be injustice if we ignore those who are deprived of this invaluable gift. The only means of communication available to the vocally disabled is the use of "Sign Language". Using sign language they are limited to their own world. This limitation prevents them from interacting with the outer world to share their feelings, creative ideas and Potentials. Another problem is that very few people who are not themselves deaf ever learn to Sign language. This further increases the isolation of deaf and dumb people from the common society. Technology is one way to remove this hindrance and benefit these people. Several researchers have explored these possibilities and have successfully achieved finger spelling recognition with high levels of accuracy. But progress in the recognition of sign language, as a whole has various limitations in today's applications. Various systems and algorithms were proposed for the recognition of sign language. A system called "Boltay Haath" [1] is developed to recognize "Pakistan Sign Language"(PSL) for vocally disabled peoples at Sir Syed university of Engineering and Technology. The Boltay Haath project aims to produce sound matching the accent and pronunciation of the people from the sign symbol passed. A wearing Data Glove for vocally disabled is designed, to transform the signed symbols to audible speech signals using gesture recognition. They use the movements of the hand and fingers with sensors to interface with the computer. The system able to eliminate a major communication gap between the vocally disable with common community. But Boltay Haath has the limitation of reading only the hand or finger movements neglecting the body action, which is also used to convey message. This gives a limitation to only transform the finger and palm movements for speech transformation. The other limitation that can be seen with Boltay Haath system is the signer could be able to communicate with a normal person but the vice versa is not possible with it. This gives the limitation of one-way communication between the listeners and vocally disabled. A similar system is proposed by Kodous and Waleed [2] where they propose a Recognition system for Australian sign language using Instrumented gloves. This proposal also gives the same limitations as seen with Boltay Haath. Don Pearson in his paper "Visual Communication Systems for the Deaf" [6] presented a two way communication approach, where he proposed the practicality of switched television for both deaf-to-hearing and deaf-to-deaf Communication. In his paper attention is given to the requirements of picture communication systems, which enable the deaf to communicate over distances using telephone lines. Extensions of such systems using the public switched telephone network may be possible if the images can be coded into low data rates [13].

3 Methodology

Speech recognition is motivated by the need to improve the performance of voice communications systems in noisy conditions. The applications range from front-ends for speech recognition systems, to enhancement of telecommunications in aviation, military, teleconferencing, cellular and biomedical applications. The goal is either to improve the perceived quality of the speech, or to increase its intelligibility. Speech enhancement is concerned with the processing of noisy and corrupted speech to improve the quality or intelligibility of the signal. Improving quality can be important for reducing listener accuracy in high stress and high noise environments. The precision for a speech recognition system can be measured in terms of speech recognition performance. Various rang of application were found for speech recognition in which one major application is the speech reading.

4 System Approach

An automated speech recognition system is proposed for the recognition of speech signal and transforms it to a cue symbol recognizable by vocally disabled people. Fig. 1 shows the proposed architecture for automated recognition system.

The system implements a speech recognition system based on the speech reading and the cue samples passed to the processing unit. The processing system consists of a speech recognition unit with cue symbol generator, which determines the speech signal and produces an equivalent coded symbol for the recognized speech signal using HMM process. In this work the design of the overall system will be implemented. The system will be operating in close to real-time and will take the speech input from the microphone and will convert it to synthesized speech or finger spelling. Speech recognition will be implemented for the considered languages. Language models will be used to solve ambiguities. Finger spelling synthesis will be implemented

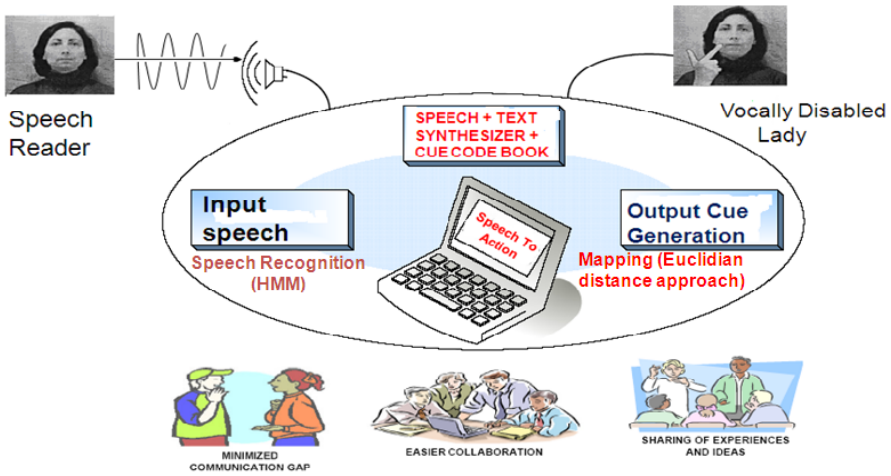


Figure 1: Proposed Automated Instrumentation Speech Recognition System

5 Working Principle

The proposed system perform three principle functions

1) Capture and parameterization of the acoustic speech input.

2) Signal identification via speech recognition and generates an equivalent symbol.

3) Generate an equivalent cue symbol based on the coded symbol obtained from the speech recognition unit. Finger spelling synthesis will be implemented. The system is given in Figure 2

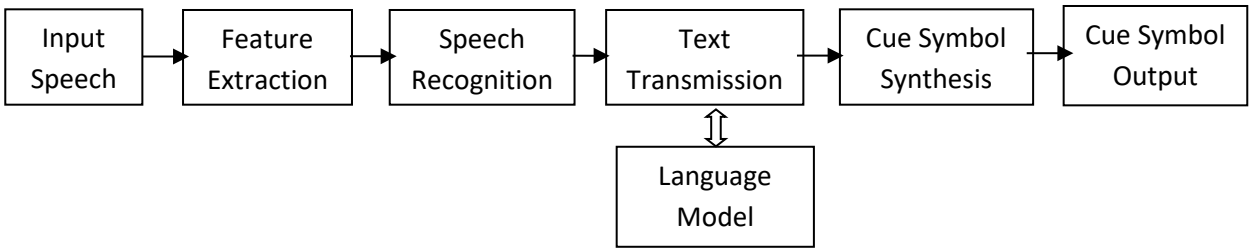


Figure.2: over system implementation

The recognition is performed using Hidden Markov Model (HMM), training the recognition system with speech features. A speech vocabulary for commonly spoken speech signal is maintained and its features are passed to the recognition system. On the recognition of the speech sentence the system generates and equivalent coded symbol in the processing unit. The symbols are then passed to the cue symbol generator unit, where an appropriate cue symbol is generated using the LMSE algorithm. For the generation of cue symbol a cue data base consisting of all the cue symbols are passed to the cue symbol generator. Figure.3 shows the cue symbols passed to the system.

The operational functionality of the HMM modeling is made as; A Hidden Markov Model is a statistical model for an ordered sequence of variables, which can be well characterized as a parametric random process. It is assumed that the speech signal can be well characterized as a parametric random process and the parameters of the stochastic process can be determined in a precise, well-defined manner. Therefore, signal characteristics of a word will change to another basic speech unit as time increase, and it indicates a transition to another state with certain transition probability as defined by HMM.

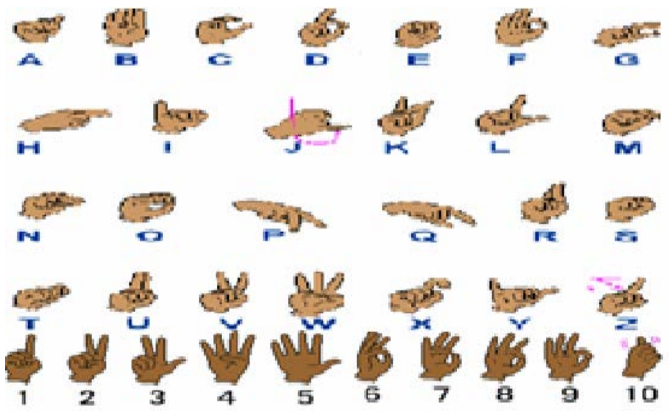


Figure 3 Equivalent English cue symbols for database. The symbols passed are the equivalent English characteristics.

5.1 Mel Spectrum Approach

A block diagram of the structure of an MFCC processor is given in Figure 4 the speech input is typically recorded at a sampling rate above 10000 Hz. This sampling frequency was chosen to minimize the effects of aliasing in the analog-to-digital conversion. These sampled signals can capture all frequencies up to 5 kHz, which cover most energy of sounds that are generated by humans. As been discussed previously, the main purpose of the MFCC processor is to mimic the behavior of the human ears. In addition, rather than the speech waveforms themselves, MFFC's are shown to be less susceptible to mentioned variations.

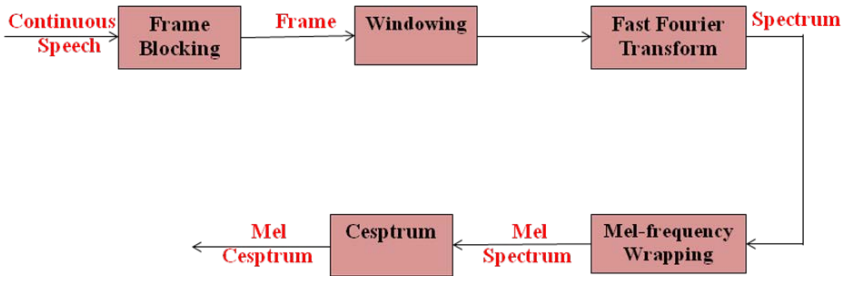


Figure 4: Block diagram of the MFCC Processor.

5.2 Hidden Markow Model Operation (HMM)

A Hidden Markov Model is a statistical model for an ordered sequence of variables, which can be well characterized as a parametric random process. It is assumed that the speech signal can be well characterized as a parametric random process and the parameters of the stochastic process can be determined in a precise, well-defined manner. Therefore, signal characteristics of a word will change to another basic speech unit as time increase, and it indicates a transition to another state with certain transition probability as defined by HMM shown in fig 4. This observed sequence of observation vectors O can be denoted by

$$O = o(1), o(2), \ldots\ldots o(T) \tag{1}$$

where each observation of (t) is an m-dimensional vector, extracted at time t with

$$O(t) = [O_1(t), O_2(t), \ldots\ldots\ldots\ldots O_m(t)]^T \tag{2}$$

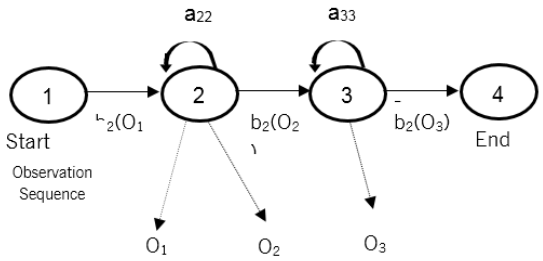


Figure.5 A typical left-right HMM (a_{ij} is the station transition probability from state i to state j; $O(t)$ is the observation vector at time t and bi $O(t)$ is the probability that $O(t)$ is generated by state i).

An HMM could be very complicated, but in general they can all be characterized by the following parameters:

a) N, the number of the states in the model. The state is hidden, however, each state within a process usually has some physical significance, like in the case of speech recognition, and each state could represent a basic speech unit. The state were denoted as $S = (s_1, s_2 \ldots\ldots s_N)$ and the state at time t as q_t.

b) M, the number of the Gaussian mixture components per state, i.e., the discrete alphabet size. The individual symbols are denoted as $V = \{v_1, v_2, \ldots\ldots v_M\}$

c) A, the state transition probability distribution $A = \{a_{ij}\}$ where the probability of being in state s_j at time $t+1$ given that we were in state s_j at time t and

$$a_{ij} = p[q_{t+1} = s_j, q_t = s_i], 1 < i, j < N \tag{3}$$

$$\sum_{j=1}^{N} a_{ij} = 1 \quad 1 < j < N, \tag{4}$$

There are many types of HMMs. For the special case such as ergodic model where all states can be reached by any other states, $a_{ij} > O$ for all $i,\ j$,

d) B, for continuous HMMs, it is the matrix of observation probability distribution over all the state and all the observations. $B = \{b_j(k)\}$, where

$$b_j(k) = p[o_t = v_k \ \ q_t = s_i], \quad 1 < j < N \qquad 1 < j < T \tag{5}$$

$$V = \{v_1, v_2, \ldots\ldots v_M\} \quad \text{and}$$

$$\sum_{t=1}^{T} b_j(t) = 1 \quad 1 < j < N \tag{6}$$

e) $\prod$, the initial state distribution $\prod = \{\pi_i\}$, in which

$$\pi_i = p[q_1 = s_j] \qquad 1 < i < N \tag{7}$$

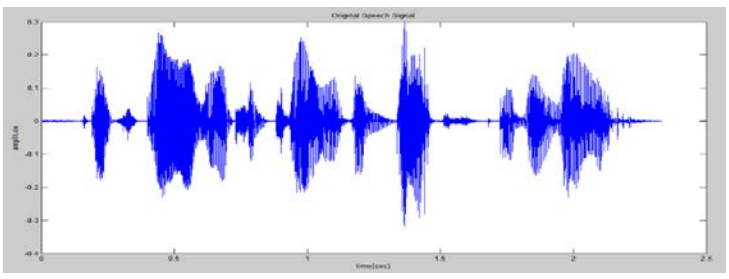
A complete specification of a HMM requires specification of two model parameters, N and M, specification of the observation symbols, and the specification of three sets of probability measures $A,\ B,\ \pi_i$ so an HMM can also be defined as a compact form $\lambda = \{A,\ B,\ \pi\}$.

5.3 Analyzer

The system evaluates the parameter of recognition system for various noises considering MFCC & MFCC with sub band as feature extraction technique. The analyzer model reads the parameter such as computation time the learning rate accuracy level, qualification rate & with respect to time to analyzing efficiency implemented system.

For the training of HMM network for the recognition of speech a vocabulary consist of collection words are maintained. The vocabulary consists of words given as, “DISCRETE”, “FOURIER”, “TRANSFORM”, “WISY”, “EASY”, “TELL”, “FELL”, “THE”, “DEPTH”, “WELL”, “CELL”, “FIVE”, each word in the vocabulary is stored in correspondence to a feature define as a knowledge to each speech word during training of HMM network. The features are extracted on only voice sample for the corresponding word. Test speech utterance: “it’s easy to tell the depth of a well”, taken at 16 KHz shown in figure 6 (a) (b) and (c) and 7. The speech signal are decomposed into a set of sub-bands with a hierarchical coding of speech signal using set of high and low pass filters. The obtained bands are then processed with the mel-frequency (MFCC) estimation, where mel frequencies are extracted for each of the band. This results in extraction of mel feature coefficient at a finer spectral level.

Test sample S1:
File Name: S1.wav
Sentence: “It easy to tell the depth of well”
Duration: 0.04 sec



(a)

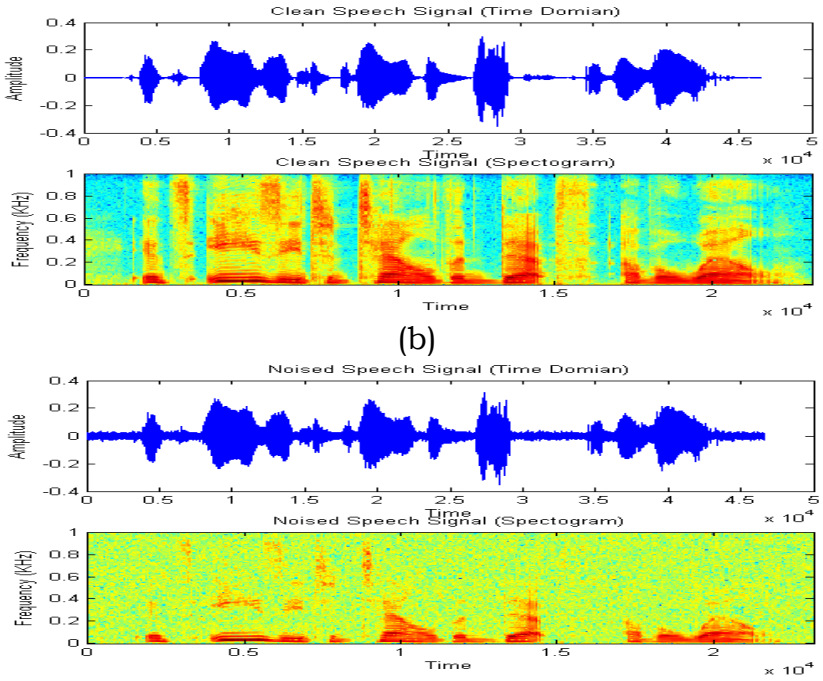


Figure 6 (a): Original Test sample (S1), (b) Spectral plot for clean speech, (c) Noise affected signal, with its spectral plot

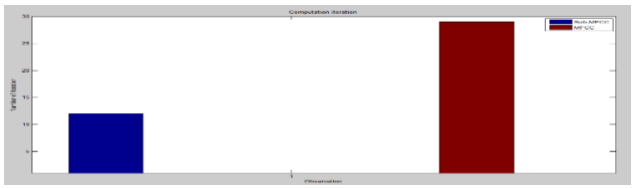


Figure 7: Computation Iteration for the developed methods

$$Retrieval\ Accuracy(\%) = \left(\frac{No.of\ truely\ recognition Words}{Total\ No.of\ Words} \right) \times 100$$

6 Mapping

Mapping of corresponding speech information into equivalent Cue symbols is done using Euclidian distance approach. The classification of the query is carried out using Euclidean distance. The Euclidean distance function measures the query & knowledge distance. The formula for this distance between a point ***X*** (***X1, X2,*** etc.) and a point ***Y*** (***Y1, Y2,*** etc.) is:

$$d = \sqrt{\sum_{i=1}^{n} (x_i - y_j)^2}$$

Deriving the Euclidean distance between two data points involves computing the square root of the sum of the squares of the differences between corresponding values. The system automatically starts searching the database for the words that starts with the specified word. This process continues letter by word until the last word. The system recognizes if the sign exists in the database or not. If it exists, it is called and shown on the monitor of the portable computer, otherwise the sign is finger spelled just like what deaf people do in their daily life.

6.1 Simulation Observation

For the simulation of the suggested approach various speech samples are been trained and tested. Speech samples from 'A' to 'Z' were recorded and their corresponding cue symbols are stored onto a database. The training database features are as tabulated,

Training Character	Energy Level, $E = (\sum_{i=1}^{m}\sum_{j=1}^{n} x(i,j))$
A	75
B	120
C	39
D	65
E	46
F	52
G	73
H	78
I	42
J	53
K	71
L	45
M	106
N	110
O	70
P	108
Q	98
R	77
S	57
T	56
U	87
V	59
W	91
X	78
Y	54
Z	61
A	95
B	97
C	62
D	93
E	68
F	70
G	78
H	106
I	50
J	56
K	89
L	62
M	120
N	71
O	84
P	76
Q	94
R	97
S	69
T	60
U	65

V	59
W	104
X	73
Y	61
Z	100

If the word is found in the Dictionary, then the cue clip related to the word is displayed filling the entire page as shown in Figure 8 to 11. However, if the word is found not to be in the database, the window is divided according to the number if words so that the entire word is displayed in the window as clearly as possible as shown in Figure 11. For the training of HMM network for the recognition of speech a vocabulary consist of collection words are maintained. The vocabulary consists of words given as, “BOOK”, “BANK”, “FINISH”, “AND”, “DAWN”, “DUSK”, “FLIES”, “BUGS”, “DEPTH”, “WELL”, “CELL”, “FIVIE”, each word in the vocabulary is stored in correspondence to a feature define as a knowledge to each speech word during training of HMM network. The features are extracted on only speech sample for the corresponding word. Test speech utterance: “it’s easy to tell the depth of a well”, taken at 16 KHz. The recognized of the speech words is processed for first 6 words and the recognized character and there symbol is as shown below

1) Test sample: ‘BOOK’, Obtained cue symbol is,

Figure 8: Obtained cue symbol for speech Sample ‘BOOK’

2) Test sample: ‘AND’ , Obtained cue symbol is,

Figure 9: Obtained cue symbol for speech sample ‘AND’,

2) Test sample: ‘FINISH’, Obtained cue symbol is,

Figure 10: Obtained cue symbol for speech sample ‘FINISH’

4) Test sample: ‘BANK’, Obtained cue symbol is,

Figure.11. Obtained cue symbol for speech sample ‘BANK’

7 Conclusion

This paper presents an approach towards automated recognition of speech signal for vocally disabled people. The system proposed could efficiently recognize the speech signal using HMM and generate an equivalent cue symbol. The proposed AISR system find its application for the vocally disable peoples for providing a communication link between normal and disabled people. The system could be integrated with finger spelling recognition system such as “Action-to-Speech” for a complete communication between the common person and the vocally disable people.

REFERENCES

[1] DONPEARSON "Visual Communication Systems for the Deaf" IEEE transactions on communications, vol. com-29, no. 12, December 1981

[2] Alison Wary, Stephen Cox, Mike Lincoln and Judy Tryggvason "A formulaic Approach to Translation at the Post Office: Reading the Signs", The Journal of Language & Communication, No. 24, pp. 59-75, 2004.

[3] Glenn Lancaster, Karen Alkoby, Jeff Campen, Roymieco Carter, Mary Jo Davidson, Dan Ethridge, Jacob Furst, Damien Hinkle, Bret Kroll, Ryan Layesa, BarbaraLoeding, John McDonald, Nedjla Ougouag, Jerry Schnepp, Lori Smallwood, Prabhakar Srinivasan, Jorge Toro, Rosalee Wolfe, "Voice Activated Display of American Sign Language for Airport Security". Technology and Persons with Disabilities Conference 2003. California State University at Northridge, Los Angeles, CA March 17-22, 2003

[4] Eric Sedgwick, Karen Alkoby, Mary Jo Davidson, Roymieco Carter, Juliet Christopher, Brock Craft, Jacob Furst, Damien Hinkle, Brian Konie, Glenn Lancaster,Steve Luecking, Ashley Morris, John McDonald, Noriko Tomuro, Jorge Toro,Rosalee Wolf, "Toward the Effective Animation of American Sign Language".Proceedings of the 9th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media. Plyen, Czech Republic,February 6 - 9, 2001. 375-378.

[5] Suszczanska, N., Szmal, P., and Francik, J., "Translating Polish Texts into Sign language in the TGT System", the 20th IASTED International Multi-Conference on Allpied Informatics, Innsbruck, Austria, pp. 282-287, 2002.

[6] Scarlatos, T., Scarlatos, L., Gallarotti, F., "iSIGN: Making The Benefits of ReadingAloud Accessible to Families with Deaf Children". The 6th IASTED International Conference on Computers, Graphics, and Imaging CGIM 2003, Hawaii, USA, August 13-15, 2003.

[7] San-Segundo, R., Montero, J.M., Macias-Guarasa, J., Cordoba, R., Ferreiros, J., and Pardo, J.M., "Generating Gestures from Speech", Proc. of the InternationalConference on Spoken Language Processing (ICSLP'2004). Isla Jeju (corea).October 4-8, 2004.

[8] Aleem khalid ,Ali M, M. Usman, S. Mumtaz, Yousuf "Bolthay Haath – Paskistan sign Language Recgnition" CSIDC 2005

[9] Kadous, Waleed "GRASP: Recognition of Australian sign language using Instrumented gloves", Australia, October 1995, pp. 1-2, 4-8.

[10] D. E. Pearson and J. P. Sumner, "An experimental visual telephone system for the deaf," J. Roy. Television Society *vol*. 16, no. 2. pp. 6-10, 1976.

[11] Guitarte Perez, J.F.; Frangi, A.F.; Lleida Solano, E.; Lukas, K. "Lip Reading for Robust Speech Recognition on Embedded Devices" Volume 1, March 18-23, 2005 PP473 – 476

[12] SantoshKumar,S.A.; Ramasubramanian, V." Automatic Language Identification Using Ergodic HMM" Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP'05).IEEE International Conference Vol1,March18-23,2005Page(s):609-612

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
ISSN 2054-7412

Pattern Recognition Based on YIQ Colour Space with Simulated Annealing Algorithm and Optoelectronic Joint Transform Correlation

Chulung Chen, Kaining Gu, Chungcheng Lee, Jianshuen Fang and Yachu Hsieh
Department of Photonics Engineering, Yuan Ze University, Taiwan;
chulung@saturn.yzu.edu.tw

ABSTRACT

For pattern recognition on various views of the interested colour object, we adopt the YIQ colour space when using simulated annealing algorithm to design the template matching function. Joint transform correlation is devoted for recognition of colour targets. Quantized reference functions are designed for the purpose of display on liquid crystal spatial light modulators. Each reference function is trained with true class images rotated in-plane at 2 degrees intervals between -14 degrees and 14 degrees. Numerical result shows that, generally, YIQ space outperforms conventional RGB space.

Keywords: Simulated annealing; Joint transform correlation; Colour pattern recognition; YIQ colour space.

1 Introduction

There are two main families of optical correlator, VanderLugt correlator (VLC) [1] and the joint transform correlator (JTC) [2]. VLC was proposed for comparing two signals by utilising the Fourier transforming properties of a lens. In 1966, Weaver and Goodman introduced the JTC for pattern recognition application. A few years later, LCD based JTC [3] proposed by Yu and Lu became an attractive tool for pattern recognition. Since then, the JTC configuration has received increased attention in the past several years because of the less restrictive alignment requirement in comparison with the VLC. However, the classical JTC suffers from strong zero order term (also called DC term) and broad correlation width. The DC term is the sum of each auto-correlation of the reference image and the target image at the output of correlation plane. The existence of the DC term will influence the performance, therefore the removal of the nonzero-order term is of great importance.

To deal with the DC term, Lu et al. [4] adoped phase-shifting technique to design a nonzero- order JTC (N0JTC) and Li et al. [6] used the joint transform power spectrum (JTPS) subtraction strategy to realize the N0JTC. The Mach-Zehnder JTC (MZJTC) [6] can remove the zero-order term in only one step directly without storing the Fourier spectra of both the reference and target images beforehand. Later, Chen et al. [7,8] adopted constraint optimization based on Lagrangian method to yield a sharp correlation peak.

To write reference function and the input test scene onto a spatial light modulator, quantized versions are necessary. On the other hand, the simulated annealing (SA) method [9,10] have been successfully applied to optimization problems. Annealing is a physical process of decreasing temperature gradually in order to reach the global minimum energy states. SA is an effective and commonly numerical optimization algorithm used to solve non linear optimization problems. SA is a Monte Carlo approach

DOI: 10.14738/aivp.45.2614
Publication Date: 28st,December 2016
URL: http://dx.doi.org/10.14738/aivp.45.2614

to minimize multivariate functions. We will take advantage of this feature for colour pattern recognition.

2 Analysis

RGB color model has been widely used and is easy to understand. It consists of the red, green and blue respectively. However, RGB color model is not the most suitable color model on many applications. In this paper, the color seperation to design the reference function (or template) is based on YIQ color space. Y means luminance that represents the achromatic (black and white) image without any color. I and Q are the two chrominance components. I is deviations from orange-luminance to cyan-luminance and Q is deviations from purple-luminance to chartreuse–luminance. It transform RGB source into one luminance and two chrominance components. On the otherhand, the optoelectronic system is based on a MZJTC structure, as shown in Figure. 1. It is consisted of one laser, one spatial filter, one collimated lens (CL), 3 beam splitters (BS), 3 polarizing beam splitters (PBS), 3 Fourier lenses (FL), 3 reflective liquid spatial light modulators (RLCSLM), 3 charge coupled device (CCD) cameras, 1 electronic subtractor (ES) which is used for removing the zero-order term at the final output, and 1 computer for controlling the whole system. Besides, there are 1 half wave plate (HWP) and 1 quarter wave plate (QWP) in front of each RLCSLM. The MZJTC structure is based on the Mach-Zehnder interferometer technique with Stokes relationships. The difference between conventional N0JTC and MZJTC is that the MZJTC structure needs only one step to remove the zero-order term. The processes are presented as follows.

First, 3 colour components of the test colour image are jointly displayed in grayscale at the RLCSLM1. Similarly, 3 colour component of the test colour image are also displayed in grayscale at the RLCSLM2. The target on the RLCSLM1 is illuminated and Fourier optically transformed by FL1. After passing through the PBS3, the irradiance of transmitted and reflected Fourier spectrum is respectively detected by CCD1 and CCD2 in the Fourier frequency domain. Then, the difference of joint Fourier power spectrum between CCD1 and CCD2 is displayed at the RLCSLM3, such that the zero-order term will be subsequently removed at the output. Finally, CCD3 captures another Fouier transform spectrum of the difference. The output contains the overlapping of each cross-correlation of the reference component and the target component. More detailed analysis of MZJTC can be found in the literature [10-12].

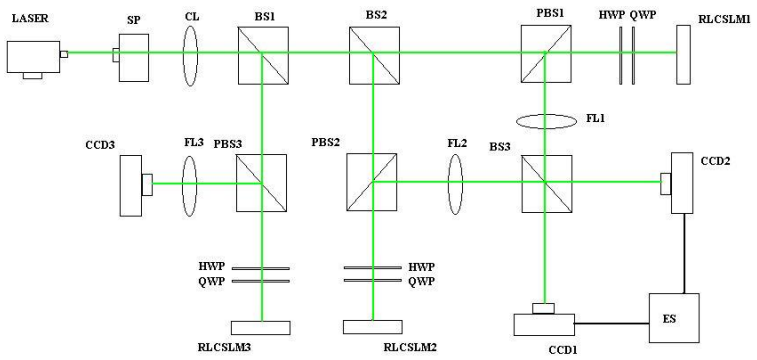


Figure 1. Mach-Zehnder joint transform correlator.

To evaluate the recognition capability, some measurement criteria [13] including correlation peak intensity (CPI) and peak to sidelobe ratio (PSR) are utilized. CPI is the cross-correlation peak intensity

at the correlation output plane. PSR is the primary correlation peak energy versus secodary peak energy in the region of interest.

3 Proposed Algorithm

One colourful insect is selected as the basic pattern of the target, whose size is of 64 × 64 × 3 pixels. It is separated into Y, I and Q channels. For the sake of comparison, another insect is selected as the nontarget. These two images are shown in Figure 2. For simplicity, We rotate these 2 objects in plane from -14° to 14°, and select patterns 2° apart. Totally there are 15 rotationally distorted patterns per object used as the training set for each colour channel. Next, for each training set, we utilize SA algorithm to obtain the reference template. In our study, the CPE (correlation plane energy) is proposed to construct the energy function.

Figure 2: Target (left) and nontarget (right)

SA simulates the cooling process by slowly lowering the system temperature until it converges to a steady, frozen state. The steps of SA algorithm for each channel are similar to those described in our previous paper[14].

- Step 1: Yield the initial reference function randomly.
- Step 2: Calculate CPE and CPI for each training image, then compute the ratio, and add all the ratios together as the energy function E_{old}. It is expressed as.

$$E_{old} = \sum_{i=1}^{N} \frac{\mathrm{CPE_i}}{\mathrm{CPI_i}} \tag{1}$$

  Here *i* is the index of the training image.

- Step 3: Alter the level number just for one pixel of the reference function *h(x, y)*, and then calculate the new energy function E_{new}.
- Step 4: If the minimum peak value of the new cross-correlation energy function over all training images is not higher than, say, 0.85 times of the minimum peak value of the old cross-correlation energy function, the change of the pixel value won't be accepted and the process returns to the step 3.
- Step 5: Calculate the difference of energy functions, which is ΔE and expressed as

$$\Delta E = E_{new} - E_{old} \tag{2}$$

- Step 6: If $\Delta E \le 0$, accept the level number in the new reference function h(x, y), set E_{new} as the next system temperature T, which is the new starting point E_{old}

- Step 7: If not, compute the probability. If it is greater than a randomly generated number within the range between 0 and 1, and then accept the alteration of the pixel value.
- Step 8: Check whether all pixels have been scanned. If they have, move to the next step. Otherwise go back to step 4.
- Step 9: Record the value of energy function in each cycle. If the normalized standard deviation of energy for the last, say, 5 cycles is smaller than, say, 0.03, and then terminate the computaion and exit the algorithm. Otherwise reduce system temperature typically by, say, 10%, and go back to step 3.

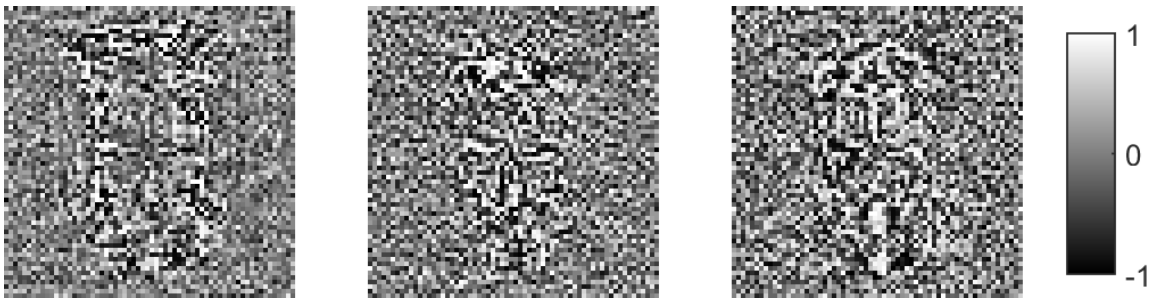


Figure 3: Y, I, Q reference functions calculated by SA algorithm

4 Result

Figure 3 shows respectively the synthesized reference function of Y, I, Q components from left to right, in grayscale of 31 levels using SA algorithm. To meet the optical requirement, the value of the grayscale is confined between -1 to 1, as illustrated by the dynamic range on the right-hand side of the figure. It is worth noting that the CPI is unlikely to be the same for all training targets. Specifically, however, in our proposed technique, the minimum value of the CPI from these training targets is set as the threshold. Furthermore, the correlation intensity has been normalized to a range between 0 and 1, based on the threshold value. Therefore, values above the threshold CPI are set to 1. The CPI curve versus the rotation angle for the target as well as for the nontarget are shown in Figure 4 for the purpose of comparison. These 2 curves are seperated considerably. To determine whether the object under test is the target, we can set a threshold value of correlation peak, above which the input can be treated as a target and below which it is a non-target. Figure 5 shows the intensity distribution of the correlation output in the region of interest, where addition of desired cross correlations between the reference and the colour component from all 3 channels occurs. Both the target and nontarget are 0° rotated. As expected, high correlation peak corresponds to the correct pattern, whereas low correlation profile is observed for the nontarget. We obtain recognition of target and discrimination of nontarget. To see how much YIQ space improves, PSR curve for RGB colour space is also plotted in figure 6. The curve is lower than that for YIQ curve at each rotation angle of the target. The reason is that, in most cases, when compared with RGB components, YIQ components are less correlated with other. This explains why the correlation profile is sharper for YIQ colour space.

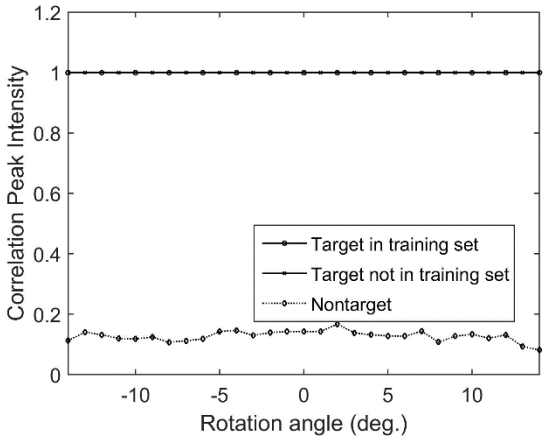


Figure 4: CPI versus rotation angle

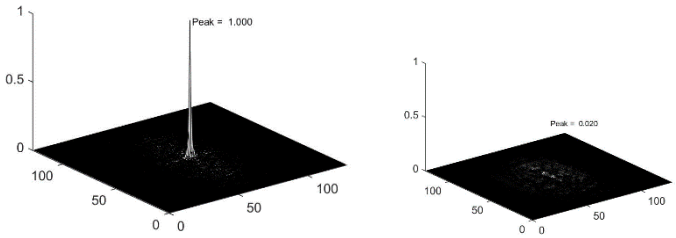


Figure 5: Example of correlation output for target (left) and nontarget (right) without rotation.

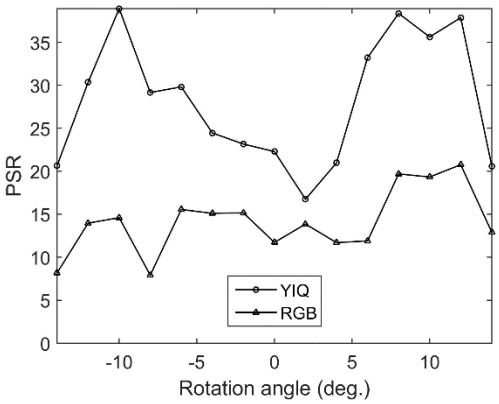


Figure 6: Comparison of the PSR between YIQ and RGB colour spaces as the target rotates.

5 Conclusion

In this paper, we have proposed to utilized YIQ colour space together with SA for pattern recognition on JTC. Comparison between YIQ and RGB colour spaces has been evaluated in terms of PSR. The improvement is remarkable. The result verifies the feasibility of our proposed method. It is exactly what we expected to see. The performance for the optoelectronic pattern recognition is promising. In the future work, we will further improve the algorithm.

REFERENCES

[1]. VanderLugt, A., *Signal detection by complex spatial filtering*. Information Theory, IEEE Transactions on, 1964. 10(2): p. 139-145.

[2]. Weaver, C. S. and J. W. Goodman, *A technique for optically convolving two functions*. Applied Optics, 1966. 5: p. 1248-1249.

[3]. Yu, F. T. S. and X. J. Lu, *A real-time programmable joint-transform correlator*. Optics Communications, 1984. 52: p. 10-16

[4]. Lu, G., et al., *Imentation of a non-zero-order joint-transform correlator by use of phase-shifting techniques*. Applied Optics, 1997. 36: p. 470-483.

[5]. Li, C., S. Yin and F. T. S. Yu, *Nonzero-order joint transform correlator*. Optical Engineering, 1998. 37: p. 58-65.

[6]. Cheng, C. and H. Tu, *Implementation of a nonzero-order joint transform correlator using interferometric technique*. Optical Review, 2002. 9: p. 193-196.

[7]. Wu, C., C. Chen, and J. Fang, *Linearly constrained color pattern recognition with a non-zero order joint transform correlator*. Optics Communications, 2002. 214: p. 65-75.

[8]. Chen C., and J. Fang, *Optimal synthesis of a real-valued template for synthetic aperture radar pattern recognition*. Microwave and Optical Technology Letters, 2002. 32(2): p. 91-95.

[9]. Kirkpatrick, S., et al., *Optimization by simulated annealing*. Science, 1983. 220: p. 671-680,

[10]. Chen, C. and C. Chen, *A Mach-Zehnder joint transform correlator with the simulated annealing algorithm for pattern recognition*. Optics Communications, 2011. 284: p. 3946-3953.

[11]. Fu, S., et al., Application of simulated annealing for color pattern recognition to the optoelectronic correlator with liquid crystal device. The 2012 IAENG International Conference on Imaging Engineering. p.683-688.

[12]. Liu, C., et al., Pattern recognition by Mach-Zehnder joint transform correlator with binary power spectrum," Proceedings SPIE 8559.

[13]. Kumar, B. V. K. V. and L. Hassebrook, *Performance measures for correlation filters*. Applied Optics, 1990. 29: p. 2997-3006.

[14]. Chen, C., et al., Application of simulated annealing for color pattern recognition to hybrid optoelectronic joint transform correlator. Advances in Image and Video Processing, 2015. 3(6): p. 13-17.