Encoding and Decoding Techniques for Medical Video Signal Transmission and Viewing

Manoranjan Paul and Golam Sorwar
School of Commerce and Management
Southern Cross University, NSW 2457, Australia
Email: {Manoranjan@ieee.org, Golam.Sorwar@scu.edu.au}

Abstract

Healthcare is a critical social issue in all over the world nowadays, especially in small town or rural areas where expert doctors, nurses, electronics equipments and other infrastructures are not adequate. As a promising approach to address this issue, telemedicine has received much attention. One of the major critical issues in telemedicine applications is transmitting data-hungry video stream in real-time through the band-limited wired or wireless channels, especially through the latter one which accommodates only very low-bit transmission. This paper provides an overview of the technical achievements in the area of encoding and decoding techniques for video data transmission and viewing applications. Furthermore, based on the state-of-the-art technology for very low bit rate video coding, few promising future research issues are also suggested, particularly for real-time telemedicine applications.

Keywords: Low-bit rate video coding, telemedicine application, video signal transmission.

1 Introduction

Telemedicine is a multimedia application where real time video, audio, text, graphics, and vital signs are transmitted through the exiting telecommunication facilities. One survey [1] in Israel shows that by using telemedicine technologies and well-planned timeline, the Army Medical Department had successfully reduced its hospital complement from 165 hospitals to just 62 hospitals in the force structure. The same survey also reveals that about 30% of physicians could not have accessed their patient's files in existing non-telemedicine systems and 70% of the records were incomplete. Normally the telemedicine objectives [1] are i) improve patients care, ii) give physicians access to conduct remote examinations, consultations, and monitoring, iii) reduce health-care cost, iv) improve access to health care for rural areas and underserved areas, and v) reduce patient transfers to secondary and tertiary care centres.

Video data is the most dominant part in terms of information contribution of Telemedicine electronic signal. However, it requires huge bandwidth among the other components of signals. Reducing the transmission bit rates concomitantly obtaining the acceptable video quality is still a big challenge for researchers in telemedicine applications in real-time through the band-limited wired or wireless channels, especially through the latter one which accommodates only very low-bit transmission.

Over the last decade, a number of popular video compression standards developed by the International Organization for Standardization (ISO), Moving Picture Expert Group and was known as MPEG-1 [2]. This was designated to provide video and audio compression for CD-ROM storage by operating at typical bit rates of 1.5Mbps. It also targeted transmission over communication channels including integrated-services digital networks (ISDN) and local area networks (LAN). The next MPEG family member was MPEG-2 [3] which operated at typical bit rates of 10Mbps and specifically focused upon compression of higher resolution video signals enhancing the scope of applications for high-quality digital television (DTV) and video, including Standard Definition TV (SDTV), Digital Versatile Disk (DVD) and High Definition TV (HDTV). The most recent coding standard is MPEG-4 [4] and has the explicit aim of extending the capabilities of the earlier standards, particularly in low bit rate video coding applications, tool-kit and content-based coding. One efficient strategy used by MPEG-4 involves sprite technology, which enables high-quality video distribution via the Internet and mobile networks. While ISO MPEG video/audio standards focused upon generic coding applications, namely the storage and asymmetric distribution of media, the International Telecommunication Union (ITU) series of video coding standards H.26X (including H.261, H.263 and H.264) [5-7] targeted fully symmetric, real-time, point-to-point or multi-point communications. Recently, the ITU-T Video Coding Expert Group (VCEG) together with experts from MPEG has collaborated to develop the H.26L standard for low bit-rate visual communications. This standard known as Advanced Video Coding (AVC)/ ISO MPEG-4 Part 10/ H.26L/ ITU-T H.264 [5] is now embedded in the MPEG-4 video compression standard. H.264/AVC affords a number of advances in coding efficiency enhancement, by employing an improved motion-prediction capability, a smaller block-sized integer-based transform, a special deblocking filter and content-based entropy coding, which collectively provides a 50% bit-rate savings for equivalent perceptual quality relative to the performance of earlier standards (especially in higher-latency applications) [8].

Most of the modern telemedicine technologies are evolved based on the MPEG-2/4 and H.263/H.264 video standard. The main target of those technologies is to reduce the bit rate without degradation of image quality or within limited bit rate bandwidth improve the perceptual image quality. A number of recent schemes are trying to reach this target within the standard video coding frameworks. In this paper, we try to describe those schemes with their advantages and disadvantages, and finally try to give a future trend of research in low bit rate video coding area.

This article is organized is as follows: Section 2 describes the procedure for video data compression. Section 3 provides an overview of the existing algorithms for low bit rate video coding. Based on the state-of-the-art low bit rate coding technology, few promising research directions are also presented for telemedicine applications in this section while Section 4 concludes the paper.

2 Video Data Compression

There are two ways to reduce the video data. One is trivial and simple way which can be applied with any other modern video coding technology with sacrificing video quality and another is standard way which must be used by any professional or commercial purpose. The simple ways to reduce the bit rate for a video sequence in generic coding paradigm are by extending the group of picture, down sampling the image size, skipping frames, non-motion compensated blocks, and residual-error-compensation, and by increasing quantization values.

2.1 Standard Coding

The above mentioned procedures are the basic ways to reduce the video data. Besides this there are encoder technologies which enable the video data reduction significantly. The most video encoders comprise the basic functionality of prediction, transformation, quantisation, and entropy encoding, there still exists considerable variations in the structure of the encoder and decoder (CODEC) arrangement. An arbitrary input frame is firstly sub-divided into MacroBlocks (MBs), which generally correspond to a group of 16×16 non-overlapping pixels in the original image. Each MB is then coded in either intra- or inter-mode determined by the block predictor with the help of the previously coded frames (i.e., Frame Memory). In intra-mode, no motion vector is generated, and thus the original (sometimes there differences between the original and the neighbouring MB in the same frame) MB is transformed by Discrete Cosine Transformation (DCT) [9]. The DCT coefficients are then quantized (Q), re-ordered (zigzag scanned), and entropy coded using any efficient Variable Length Coding (VLC) technique. Sometimes a coding control mechanism is used to control the bit rate by adjusting the quantization value. In inter-mode, an MB is formed by Motion Compensation (MC) prediction from one or more reference frames using Motion Estimation (ME); however, the prediction for each MB may be formed from one or two previous or forward frames (in time order) that have already been encoded and reconstructed. The prediction MB is subtracted from the current MB to produce a residual error MB which is then transformed using DCT and quantized (Q) to give a set of coefficients which are re-ordered (zigzag scanned) and entropy-encoded, using any VLC algorithm. The entropy-encoded coefficients, together with the side information required to decode the MB (such as the MB prediction mode, motion vector and quantization step size) form the compressed bit stream. Inverse quantization (Q^{-1}) and Inverse DCT (IDCT) are used to form the reference frames which are stored in the frame memory for the encoder.

While for special applications, some functional elements are modified or additional blocks are included, the basic structure of a video CODEC remains the same. Examples of some of the additional components include a pre-processing filter– to reduce the noise introduced in capturing images from low-quality sources, or camera shake, and a post-processing filter– to reduce the blocking and/or ringing effects [10]. These additional components enhance the performance in certain cases at the expense of increased hardware complexity.

2.2 H.264

The most recently advanced video coding standard, H.264/AVC [5], has been recently finalized to support a wide range of applications by including new flexible features to merge the concepts of MPEG-X and H.26X. The applications of H.264/AVC include: i) broadcast over cable, satellite, and cable modem, ii) interactive or serial storage on optical or magnetic storage, iii) conversational services over ISDN, LAN, and wireless mobile networks, and iv) video-on-demand. Variations of H.264 compared to the previous standards will be discussed in the rest of the section.

6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007)
0-7695-2841-4/07 $25.00 © 2007 IEEE

A profile is defined as a set of coding functions and specifications that is required by an encoder or decoder which complies with that profile. H.264 supports three profiles, baseline, main, and extended. The baseline profile is used in low bit rate video coding applications such as telemedicine, video telephony, videoconferencing, and wireless communication. The baseline profile of H.264 supports two types of frame, namely Intra and Predicted frames.

The overall steps of H.264 video coding can be divided into several steps such as, motion estimation and compensation [11], transform, quantization, entropy coding. Inter macroblock prediction creates a predicted macroblock of the current block from one or more previously encoded video pictures through motion estimation. H.264 considers each MB as either skipped or non-skipped MB. The MB with no motion or little motion is considered as skipped MB and no motion vectors or residual errors are needed as it will be copied from the reference frame directly. Each non-skipped MB (16×16) may be divided four ways, and motion estimation and compensation are carried out either as one 16×16 block, two 16×8, two 8×16, or four 8×8 blocks. If the 8×8 mode is selected, each of the four 8×8 sub MBs within the MB may be further divided four ways, either as one 8×8 sub MB, two 8×4 sub MBs, two 4×8 sub MBs, or four 4×4 sub MBs. The motion estimation mode is selected based on the minimum value of the Langragian cost function. The Langrangian cost function is defined by the total bits needed to encode the MB and the sum of square differences between the original MB and the reconstructed MB multiplied by Langrangian Multiplier [12][16].

After motion estimation, DCT is applied on the difference between original MB and best matched MB (based on the motion) in reference frame. Unlike H.263, H.264 used 4×4 integer transformation [17] instead of 8×8 non integer DCT transformation. Transform coefficients are quantized for more compression. H.264 uses 52 levels of quantization where H.263 uses only 31 levels. Then entropy coded using a context-based adaptive variable length coding (CAVLC) [33].

3 Existing Low Bit Rate Coding

The National Aeronautics and Space Administration (NASA) played an important part in the early development of telemedicine [18] when humans began flying in space. Physiological parameters were telemetered from both the spacecraft and the space suits during missions. In this decade, H.26X standard played an important role in low bit rate video coding. The H.261 and H.263 standards are, however, unable to efficiently encode the boundary-adjoined part of a moving object within a 16×16 pixel MB during motion estimation, resulting in all 256 residual error values being transmitted for motion compensation regardless of whether there are moving objects or not. Efficient encoding of these blocks needs to eliminate the intra-block temporal redundancy (ITR), because they are almost static in the successive frames. None of the block-based standards, however, is able to exploit the ITR in the form of static background within the MB. To remove this inefficiency, the MPEG-4 video standard first introduced the concept of content-based coding, by dividing video frames into separate segments (instead of MBs), comprising a background and one or more arbitrary-shaped moving objects that are coded separately. As this process requires expensive segmentation and shape coding, and is also ineffective for real-world video objects, it is not suitable for low processing devices or wireless applications.

One solution in exploiting ITR is to sub-divide the MB and then apply motion estimation and compensation to each sub-block. With sufficient numbers of sub-blocks, the shape of a moving object can be more accurately represented. This solution has been implemented in the recent H.264 standard using the variable block size (VBS) mode. It, however, requires not only bits overhead for the motion vector and VBS mode for each partition, but also higher computational complexity in order to identify the best partitioning. Obviously, the smaller the sub-blocks, the higher these overheads will be and that could potentially offset all the coding efficiency resulting from better moving object shape approximation. As a result, low bit rate coding using H.264 avoids smaller sub-blocks, and thus makes its VBS mode ineffective.

Paul *et al*. [13-15] Wong *et al*.[19], and Fukuhara *et al*. [20] recently exploit the idea of the MPEG-4 in partitioning MBs via a simplified segmentation process that again avoided handling the exact shape of the moving objects, so that popular MB-based motion estimation techniques could be applied. In this approach, fixed number of pre-defined regular shaped 64-pixel patterns is used. Patterns are used as an intermediate processing tool to get the moving regions from a binary matrix created from the current and reference frames based on their relative pixel intensity changes. Each pattern is defined as a 16×16 block where the white region represents '1' (i.e., motion) and the black region represents '0' (i.e., background). If the moving region of an MB is well covered by a particular pattern, then the MB can be coded by considering only the 64 pixels of that pattern with the remaining 192 pixels being skipped as static background. Successful pattern matching, therefore, has a theoretical maximum of four times compression for such an MB. The actual achievable compression, however, will be much

6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007)
0-7695-2841-4/07 $25.00 © 2007 IEEE

lower due to the bit overheads for handling an additional MB type, pattern identification numbering, and pattern matching errors. The existing pattern-based video coding improved the compression ratio by around 20% in very low bit rate ranges.

For telemedicine applications, it is common for the coefficients of a block to be zero after motion estimation/compensation, DCT and quantisation. A method that detects all-zero DCT coefficient blocks before the calculation of DCT and quantization has recently been introduced and has greatly improved the coding speed [21][22].

Medical video compression is constrained by the fact that high quality video is necessary due to legal reasons (depending on the corresponding country's laws) and due to the fear of misdiagnosis. Since the data transmission facility in rural areas is poor (dial up internet or expensive mobile network), it is almost impossible to transmit whole video frames with good enough quality. A possible solution to this dilemma is to offer video coding techniques which allow a video frame to be selectively, *region of interest* (ROI) compressed. In this direction, Xu *et al.* [23] and Sun *et al*. [24] proposed content-selective video data processing and compression method based on the Discrete Wavelet Transform (DWT) for transmitting high-quality neurophysiology video via the Internet for the applications of telemedicine. Their algorithm is adaptive to intraoperative monitoring video data and has a great scalability on real-time network bandwidth allocation compared to the general-purpose video compression methods. In this method, low-sub-band of wavelet transform coefficients are emphasied whereas the other sub-band coefficients are cropped. Thus, a large number of bits are allocated for the region of interest and less number of bits are for other regions. The experimental results demonstrated that this method offered higher quality within the critical field of neurosurgical video. Liu *et al.* [25] proposed an automatic detection of region of interest, which is more important, for intraoperative monitoring video data. Lin *et al.* [26] used region-based quantization level for better perceptual image quality. They used relatively course quantization level for less moving regions and fine quantization level for heavily moving regions. In this way they improve the rate-distortion performance in a limited band width. Isechi, *et al.* [27] showed that wireless LANs and video-telephony over Windows PCs could be utilised for telemedicine in rural islands.

Recently, distributed video coding approaches have been proposed to provide potential reverse in computational complexity between encoder and decoder [28][29][32]. Generally, computationally heavy encoder and light decoder are needed for downlink transmission model of TV broadcast; on the other hand, light encoder and heavy decoder are needed for up link wireless video transmission, for instance, low battery-powered cam coder or hand held mobile devices. This new concept is based on the distributed source coding theories introduced by Slepian-Wolf [30] and Wyner-Ziv [31]. In telemedicine applications, it may be needed to switch between these kinds of encoder and decoder architectures depending on the existing facilities. The distributed video coding reduces the computational complexity significantly; however, the performance is far from the conventional coding.

In a nutshell, video data transmission requires very high bit rates. To reduce the bit rate video content needs to be divided into important or non-important parts and then some approaches need to be implemented to allocate more bits for important parts and less bits for non-important parts. Now the critical issues are how to define important and non-important parts, how to automatically/manually detect the two parts, and how to divide the given bits for both parts. These questions are still open research issues as not yet efficiently solved.

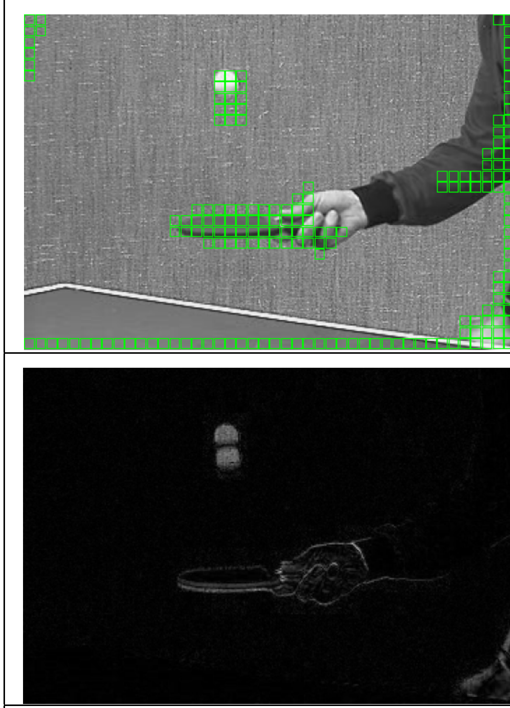

Figure 1: Region of interest detected by phase correlation using motion information between frame # 1 and 2 of Tennis video sequence (top) and pixel intensity differences between both frames (bottom).

A substantial research has been conducted in a range of video/image processing applications using phase correlation; for example, image registration [34], image stabilization [35], motion estimation [36] and video segmentation [37] applications. To the best of our knowledge, no attempts have been made investigating the potential of this technique in identifying the ROI for low bit-rate telemedicine video coding applications. Phase correlation can predict motion information between current and reference blocks. Exploiting this information we can easily define the ROI by assuming that frequently changing moving regions are of our region of interest.

Figure 1 shows an example of ROI detected by phase correlation technique using motion information. From the figure it is easily observed that the moving regions (the ball, the bat and the tennis player himself) are successfully identified by the method. After finding the region of interest we may apply different methods to encode this area more efficiently compared to the other area.

Encoder and decoder architecture is another vital issue. For a specific case, choice of the encoder and decoder (either heavy or light) is another issue to be solved. Based on the state-of-the-art technology and the demand from real-world applications, we also believe that a hybrid encoder combining region of interest, pattern templates, and distributed video coding concepts will be the future very low bit rate video coding scheme especially for telemedicine applications.

4 Conclusion

One of the difficult problems in telemedicine applications is transmitting video data in real-time through the band-limited wired or wireless channels, where very low-bit transmission is required. This paper has provided an overview of some techniques established in this area based on the region of interest, pattern templates, and distributed video coding concepts. The distributed video coding is limited in its rate-distortion performance. The patter-based video coding is sometimes failed if the object is not aligned with its pre-defined pattern templates. Though these three concepts are prominent in very low bit rate areas, ROI based approach has been increasingly considered to be the most appropriate in the framework of telemedicine application in trading of good quality and various bit rate. Though region of interest (ROI) based approach may provide a better solution for low-bit rate video coding for telemedicine applications, appropriate ROI selection is still a research issue. In this paper, we have proposed a ROI selection technique using phase correlation method. Our experimental results have shown its potential to effectively identify the ROI from unimportant regions for low bit–rate video coding application.

Finally, few other promising research issues have also been identified for very low bit rate video coding scheme especially for telemedicine applications.

5 References

[1] Zach, S., “Telemedicine overview and summary”, 19th Convention of Electrical and Electronics Engineers in Israel, 409-412, 1996.

[2] ISO/IEC 11172. MPEG-1, “Information technology-coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s,” 1993.

[3] ISO/IEC 13818. MPEG-2, “Information technology: generic coding of moving pictures and associated audio information: Video,” 1995.

[4] ISO/IEC 14496. MPEG-4, “Information Technology-coding of audio-visual objects,” 1998.

[5] ITU-T Rec. H.264/ISO/IEC 14496-10 AVC. Joint Video Team (JVT) of ISO MPEG and ITU-T VCEG, JVT-G050, 2003.

[6] ITU-T Recommendation H.261, “Video CODEC for audiovisual services at p□ 64kbits/s,” 1993.

[7] ITU-T Recommendation H.263, “Video coding for low bit-rate communication. Version 3,” 2000. ITU-T Recommendation H.263, “Video coding for low bit-rate communication. Version 3,” 2000.

[8] Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A. (2003). Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol, 13(7), 560-576.

[9] Ahmed, N., T. Nararajan, and K. R. Rao, “Discrete cosine transform,” IEEE Trans. Comput., pp. 90–93, 1974.

[10] Richardson, I. E. G., “H.264/MPEG-4 Part 10: Overview,” Retrieved January 14, 2004 from http://www.vcodex.fsnet.co.uk/h264.html, 2004.

[11] Sorwar, G., M. Murshed, and L. Dooley, “A fully adaptive performance-scalable distance-dependent thresholding search algorithm for video coding ,” IEEE International Conference on Speech, Acoustics, and Signal Processing (ICASSP-03), 3, 649-652, 2003.

[12] Paul, M. M. Frater, and J. Arnold, “Efficient mode selection algorithm using image distortion for H.264 video encoder,” IEEE International Conference on Signal Processing (ICSP-06), Accepted for publication, 2006.

[13] Paul, M., M. Murshed, and L. Dooley, “A real-time pattern selection algorithm for very low bit-rate video coding using relevance and similarity metrics,” IEEE Trans. on Cir. and Sys. on Video Tech., vol. 15, no. 6, pp. 753–761, June, 2005.

[14] Paul, M. and M. Murshed, “An efficient similarity metric for pattern-based very low bit-rate video coding,” Journal of Internet Technology, vol. 6, no. 3, 2005.

6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007)
0-7695-2841-4/07 $25.00 © 2007 IEEE

[15] Paul, M. and M. Murshed, "Advanced Very Low Bit Rate Video Coding Using Preferential Pattern Selection Algorithms," Journal of Research and Practice in Information Technology, vol. 37, no. 2, pp. 89-99, 2005.

[16] Weigrand, T., H. Schwarz, A. Joch, and F. Kossentini, "Rate-contrained coder control and comparison of video coding standards," IEEE CSVT, 13 (7), 688–702, 2003.

[17] Docef, A., F. Kossentini, K. Nguuyen-Phi, and I. R. Ismail, "The quantized DCT and its application to DCT-based video coding," IEEE Trans. on Image Processing, vol. 11, no. 3, pp. 177–187, 2002.

[18] Bashshur R, Lovett J., "Assessment of telemedicine: Results of the initial experience," Aviation Space and Environmental Medicine, 48(1): 65-70, January, 1977.

[19] Wong, K.-W., K. –M. Lam, and W. –C. Siu, "An efficient low bit-rate video-coding algorithm focusing on moving regions," IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no. 10, pp. 1128–1134, 2001.

[20] Fukuhara, T., K. Asai, and T. Murakami, "Very low bit-rate video coding with block partitioning and adaptive selection of two time-differential frame memories," IEEE Trans. Circuits and Systems for Video Technology, vol. 7, pp. 212–220, 1997.

[21] Xuan, Z., Y. Henghua, and Y. Songyu, "Method for detecting all-zero DCT coefficients ahead of discrete cosine transformation and quantisation," Electronics Letters, vol. 34, no. 19, pp. 1839–1840, 1998.

[22] Jun, S. and S. Yu, "Efficient method for early detection of all-zero DCT coefficients," Electronics Letters, vol. 37, no. 3, pp. 160–161, 2001.

[23] Xu, J., Sclabassi, R.J., Liu, Q., Hu, C., Chaparro, L.F., and Sun, M., "A region-based video coding method for remote monitoring of neurosurgery," Proc. of the IEEE 31st Annual Northeast Bioengineering Conference, 89-91, 2005.

[24] Sun, M., Q. Liu, J. Xu, B. Liu, C. Hu, A. Kassam, and R. J. Sclabassi, "A multimedia system for remote neurosurgical monitoring," Proceedings of Int. Symposium on Intelligent Sig. Proc. and Comm. Sys. (ISPACS-04), 379-383, 2004.

[25] Liu, Bing, "Detection of region of interest in neurosurgical video used for telemedicine," IEEE International Conference of Information Acquisition, 2005.

[26] Lin, C-H., Ja-Ling Wu, "Content-based rate control scheme for very low bit-rate video coding", IEEE Transaction on consumer electronics, 123-133, 43(2), 1997.

[27] Isechi. A., C., Oda, R. Shinkura, S. Akiba, H. Fujikawa, and K. Yamazaki, "Experiment of Internet-based Tele-medicine in Amami Rural Islands," IEEE International Symposium on Applications and the Internet Workshops, 2004.

[28] Puri, R and K. Ramchandran, "PRISM: a new robust video coding based on distributed compression principles," in Proc. Of Conference on Communication, Control, and Computing, Allenton, IL, 2002.

[29] Aron, A and B. Girod, "Towards practical wyner ziv coding of video," in Proc of IEEE International Conference on Image Processing, 2003.

[30] Slepian, J. D., and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Transaction on Information Theory, 19, 471–490, 1973.

[31] Wyner, D. and J. Ziv, "The rate-distortion function for source coding with side information at the decoder,"IEEE Transaction on Information Theory, 22, 1–10, 1976.

[32] Ali, M. and M. M. Murshed, "Lossless compression of correlated images/data with low complexity encoder using distributed source coding Techniques," Lecture Notes in Computer Science, 3656, 302–309, 2005.

[33] Bjontegaard, G., and K. Lillevoid, "Context-adaptive VLC coding of coefficients," JVT document JVT-028, JVT of ISO MPEG & ITU VCEG, Rairfax, Virginia, USA, May, 2002.

[34] Zhongke, L., Y. Xiaohui, and W. lenan, "Image registration based on Hough transform and phase correlation," Int. Con. on Neural Networks and Signal Processing, vol. 2, pp. 956–959, 2003.

[35] Erturks, S., "Digital image stabilization with sub-image phase correlation based on global motion estimation," IEEE transaction on consumer electronics, vol. 49(4), pp. 1320–1325, 2003.

[36] Argyvriou, V. and T. Vlachos, " Motion estimation using quad-tree phase correlation," IEEE International conference on Image Processing, Vol. 1, pp. 1081–4, 2005.

[37] Erturkerturk, A. and S. Erturkerturk, "Unsupervised segmentation of huper special images using modified phase correlation," IEEE geoscience and remote sensing letter, Vol. 3(4), pp.527–531, 2006.