

# Codec with Neuro-Fuzzy Motion Compensation & Multi-Scale Wavelets for Quality Video Frames

**Prakash Jadhav**

K. S. School of Engineering & Management / ECE Department, Bangalore, 560109, India  
Email: pcjadhav12@gmail.com

**G.K.Siddesh**

J.S.S.Academy of Technical Education / ECE Department, Bangalore, 560060, India  
Email: siddeshgundagatti@gmail.com

Received: 18 May 2017; Accepted: 06 June 2017; Published: 08 September 2017

**Abstract**—Virtual Reality or Immersive Multimedia as it is sometimes known, is the realization of real-world environment in terms of video, audio and ambience like smell, airflow, background noise and various ingredients that make up the real world. This environment gives us a sense of reality as if we are living in a real world although the implementation of Virtual Reality is on a laboratory scale. Audio has attained unimaginable clarity by splitting the spectrum into various frequency bands appropriate for rendering on a number of speakers or acoustic wave-guides. The combination and synchronization of audio and video with better clarity has transformed the rendition matched in quality by 3D cinema. Virtual Reality still remains in research and experimental stages. The objective of this research is to explore and innovate the esoteric aspects of the Virtual Reality like stereo vision incorporating depth of scene, rendering of video on a spherical surface, implementing depth-based audio rendering, applying self-modifying wavelets to compress the audio and video payload beyond levels achieved hitherto so that maximum reduction in size of transmitted payload will be achieved. Considering the finer aspects of Virtual Reality we propose to implement like stereo rendering of video and multi-channel rendering of audio with associated back channel activities, the bandwidth requirements increase considerably. Against this backdrop, it becomes necessary to achieve more compression to achieve the real-time rendering of multimedia contents effortlessly.

**Index Terms**—Artificial Neural Networks, Peak Signal to Noise Ratio, Root Mean Square Error, Quantization, Discrete Cosine Transform, Bandwidth.

## I. INTRODUCTION

Virtual Reality known, is as the realization of real-world circumstances in terms of video, audio. This environment gives us a sense of reality as if we are living

in a real world although the implementation of Virtual Reality is on a laboratory scale. Audio has attained unimaginable clarity by splitting the spectrum into various frequency bands appropriate for rendering on a number of speakers or acoustic wave-guides. The combination and synchronization of audio and video with better clarity has transformed the rendition matched in quality by 3D cinema. Multicasting of several channels over a single station, program menu options, parental control of channels and various online activities like gaming, business transactions etc. through back channel activities have made multimedia systems truly entertaining and educative. But then, there are several aspects of virtual reality that are missing from practical implementations even today.

Compression of Video Frames is a heavily researched domain in Digital TV Technology and Multimedia applications and hundreds of researchers have experimented with various techniques of transform coding and quantization mechanisms with varied results. Despite this fact, there is still a tremendous amount of scope for better compression schemes that seek to reduce the utilized band of the spectrum. In real time applications such as Digital TV and streaming Multimedia applications, time plays a crucial role and a successful implementation usually makes a compromise between factors such as (1) quality of reconstructed frames, (2) amount of utilized band owing to compression and (3) ability of the system to cope with increasing frame rates and advanced video profiles that are creeping into Digital TV Standards. It is important to observe that methodologies employed in achieving successful compression based on both inter and intra-frame redundancies rely very heavily on the frequency-domain redundancies generated by transform coding techniques and it is imperative that compression techniques place a great amount of emphasis on the correlation that can be generated on a sub-image basis. Noise in images tends to decrease the correlation and noisy frames tend to compress less. Our objective in this research program is

to focus on ways and means of increasing the redundancies by employing Linear Neural Network techniques with feed-forward mechanisms. Since quantization plays a major role in the amount of compression generated and the quality of the reconstructed frames, adaptive quantization based on a PSNR threshold value is employed so that best compression with a guaranteed reconstructed quality is obtained. Our research program targets to achieve the objectives: (1) better compression ratios with even fairly noisy frames (2) better quality of reconstructed frames in terms of very large PSNR and RMS Error tending towards values extremely small and nearer to 0 and (3) honoring the time constraints imposed by frame rates.

## II. LITERATURE SURVEY

The existing implementations of Compressor and Decompressor follow the ISO Standard 13818. The standards are evolved from MPEG - 1, MPEG - 2, MPEG - 4, MPEG - 7 and MPEG - 21. There are implementations of CODEC based on H.264 Standard, which uses Wavelets instead of DCT. MPEG2 implements the CODEC based on DCT. The shortcomings of existing CODECs are not adaptive to the patterns of the pixel residents in the video frames. MPEG2 uses a flat and uniform quantization while other implementation uses Vector Quantization techniques. It is important to observe that Vector Quantization, while being marginally superior in terms of compression ratio, has substantial computing overheads which counterbalance the gain resulting from better compression ratio for a given visual quality metric. There is a authentic need for better mechanisms of compression which will seek to achieve (1) decoding of pictures faster at the receiver end (2) compression ratio is better and (3) frames are constructed in better quality factor in terms of Peak Signal to Noise Ratio and Mean Square Error

## III. IMMERSIVE-VIRTUAL REALITY

The proposed framework is based on adopting an adaptive quantization technique and deploying a linear Neural Network in order to perform the compression. It is a well known fact that quantization controls the amount of information that is being discarded as being non-contributive to the human visual system. A flat quantization as adopted by MPEG2 CODEC does not take into account the significance of the transform coefficients. Owing to vast structural differences in varied types of images, a flat quantizer cannot discriminate between contributing and minimal-contributing coefficients. This being the case, sections of images with excessive details (with significant coefficients spread across the frequency matrix) may lose useful information in the quantization stage and compression may not be the maximum. Therefore, an adaptive quantizer that suits itself to the local characteristics of the blocks performs best. Virtual reality

it defines about 'virtual' and 'reality'. The meaning of 'virtual' is near and meaning of 'reality' is as human beings what we feel or experience. Therefore, the 'virtual reality' term means 'near-reality'. This is the classical definition of virtual reality. The components that constitute virtual reality are: audio, video, images, ambience etc.

To make Virtual Reality nearer to reality, the video or images could be immersed in a 3D world with stereo/ 3D vision. The rendering display unit could be on a spherical surface depicting the real world scenario. When we stand at a seashore and observe the horizon, it is spherical in nature. Natural scenes are not flat but embedded in a 3D spherical world. Rendering these on a 2D flat screen removes the effect of depth of scene. Then again, the human visual system is stereo in nature. The scene is captured by both the left and right human eye giving a true sense of depth of objects within. Therefore, stereo rendering of scenes on a spherical manifold is nearer to reality.

Audio rendering has so far reached the level of DOLBY 5.1 with the use of 6 different bands. Although DOLBY 5.1 is considered the most sophisticated media of audio rendering, there is still a lot left in its implementation. DOLBY standard does not generate high frequency sounds around 20Hz with much fidelity. Further, the depth of sound is not perceptible unless the speakers are arranged to synchronize with the acoustics of the room. This being the case, there is sufficient room for research in improving the quality with addition of adaptive noise cancellation, faithful reproduction of high frequency sounds (especially percussion instruments and piano), selective and mild echoing to generate pleasant aural effects.

Virtual (Immersive) Reality seeks to combine the essential ingredients of audio, video and back channel activities in unobtrusive ways to generate a real "virtual world". Adding of environment aspects will make Virtual Reality truly awesome. Although Virtual Reality has been around since a decade, we do not see any true implementation that is worthy of being called realistic. Current implementations concentrate only on specialized implementation targeting specific areas such as education, medicine, stock markets etc.

### A. *Need for a more efficient Transport Protocol for Payload*

The Fig 1, of a DTTB (Digital Terrestrial Television Broadcasting) system is shown. The video, audio and data are compressed and multiplexed to get streams.

To form MPEG-2 Transport Stream these streams are multiplexed with data source from the programs. Transport streams consists packets of 188 bytes length. To protect transport streams from interruption and noise in transmission channel the encoder FEC takes precautionary measurement. It uses outer interleaving, Convolutional and Reed Solomon coding. To transmit the suitable digital symbols the modulator converts FEC protected transport packet into digital symbol, which are preferred for transmission in the global channel. This

involves QAM & OFDM in DVB-T & ISDB-T systems, or VSB in ATSC-T and PAM. Upper converter is used to convert digital symbols into required appropriate RF Channel. The operation is in reverse order in the receiver section.

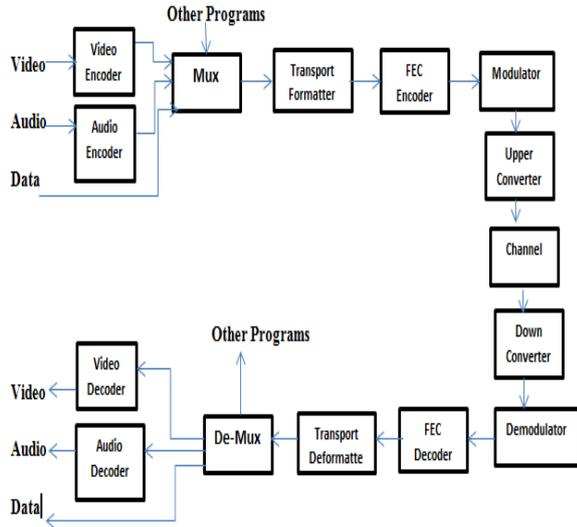


Fig.1. A typical Digital TV Schema

QAM & OFDM in DVB-T & ISDB-T systems, or VSB in ATSC-T and PAM. Upper converter is used to convert digital symbols into required appropriate RF Channel. The operation is in reverse order in the receiver section

**B. Algorithms**

Our research work so far has been related to: Stereo Video Rendering using 3D curvilinear coordinate system with Left and Right video frames, projecting the space-shifted frames on to the video display. Algorithms have been developed by us to give the effect of depth and perspective projection techniques have been used.

Novel compression techniques based on multi-scale wavelets have been invented to generate a balance between video quality and compression levels.

The latest video quality assessment techniques based on CIDIED2000 color differences have been employed to ensure quality of reconstructed frames.

1. The inputs to the rendering codec are the Left and Right frames of the stereo vision camera. The frames are in RGB (Red/Green/Blue) pixel format. Since manipulation of color images is done in the Y, C<sub>b</sub>, C<sub>r</sub> color space, the RGB pixels are converted to Y, C<sub>b</sub>, C<sub>r</sub> using the following linear equations:

$$\begin{aligned}
 Y' &= 16 + (65.481.R' + 128.553.G' + 24.966.B') \\
 C_B &= 128 + (37.797.R' - 74.203.G' + 112.0B') \\
 C_R &= 128 + (112.0.R' - 93.786.G' - 18.214B') \quad (1)
 \end{aligned}$$

The Y,C<sub>b</sub>,C<sub>r</sub>, image is split into 3 planes Y, C<sub>b</sub>, C<sub>r</sub>. These three planes become input to the multi-scale wavelet algorithm.

**C. Wavelet Transform Schema**

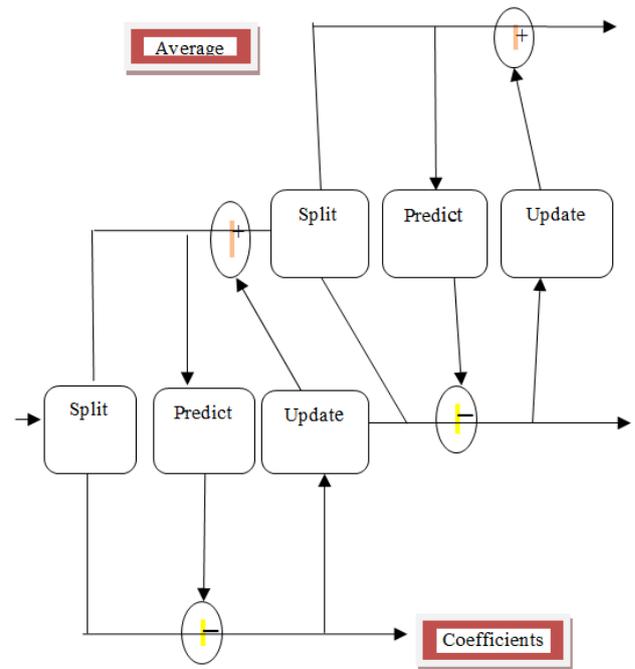


Fig.2. Wavelet Transform Schema

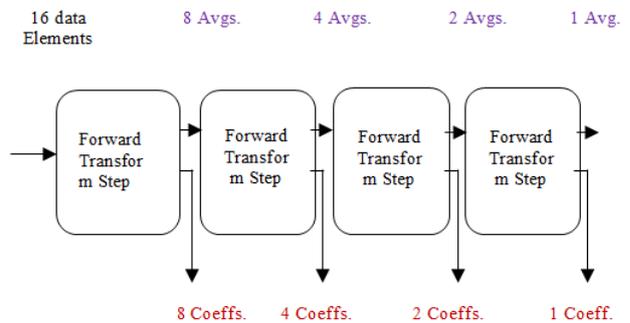


Fig.3. Steps in the Wavelet Lifting Scheme (Forward Transform)

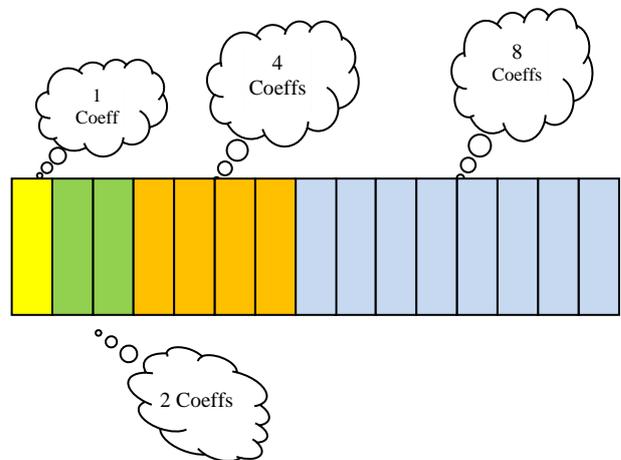


Fig.4. Results of Wavelet Transform with coefficients

Results of Wavelet Transform with coefficients ordered in increasing frequency order

- After application of the transform, we have 3 frequency planes corresponding to Y, C<sub>b</sub>, C<sub>r</sub>. The frequency planes of both Left and Right Images are digitally composited to get a single plane. For 3D to 2D rendering, we transform every X,Y,Z coordinate for rendering as follows:

Apply x-axis rotation to transform coordinates (x<sub>1</sub>, y<sub>1</sub>, z<sub>1</sub>).

$$\begin{aligned} x_1 &= x \\ y_1 &= y * \cos(Rx) + z * \sin(Rx) \\ z_1 &= z * \cos(Rx) - y * \sin(Rx) \end{aligned} \quad (2)$$

Apply y-axis rotation to (x<sub>1</sub>, y<sub>1</sub>, z<sub>1</sub>) to get (x<sub>2</sub>, y<sub>2</sub>, z<sub>2</sub>).

$$\begin{aligned} x_2 &= x_1 * \cos(Ry) - z_1 * \sin(Ry) \\ y_2 &= y_1 \\ z_2 &= z_1 * \cos(Ry) + x_1 * \sin(Ry) \end{aligned} \quad (3)$$

Finally, apply z-axis rotation to get the point (x<sub>2</sub>, y<sub>2</sub>).

$$\begin{aligned} x_3 &= x_2 * \cos(Rz) + y_2 * \sin(Rz) \\ y_3 &= y_2 * \cos(Rz) - x_2 * \sin(Rz) \\ z_3 &= z_2 \end{aligned} \quad (4)$$

- Apply mid-rise quantizer to the result of step 1. Each coefficient value is rounded as:

$$\text{Pixel} = \text{floor}(\text{Pixel} + 0.5) \quad (5)$$

- Quantization:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

The wavelet coefficients of step 3 are quantized using the above matrix to yield a coefficient array with smaller dynamic range.

- Rearrange Coefficients in increasing frequency order:

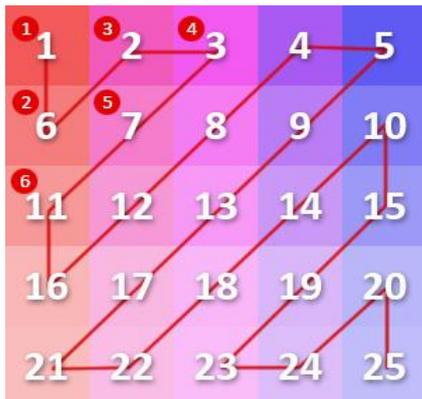


Fig.5. Rearrangement of wavelet coefficients

We use the above scheme to rearrange the wavelet coefficients for subsequent compression.

- Temporal redundancy is achieved by adjacent frame differencing. Each pixel in the given frame is manipulated with the corresponding pixel in the next frame using:

$$\text{Pixel}(x, y) = \text{Pixel.1}(x, y) \text{ XOR } \text{Pixel.2}(x, y) \quad (6)$$

The ANN and Fuzzy Engine is employed at this stage.

- Compression is done by Adaptive Huffman Coding – which is the industry standard method.
- The picture at the receiver end is clocked to the display by retracing steps 7 to 1 in the reverse order.

#### IV. RESULTS

Table 1. Results of MPEG2 Codec with 8x8 DCT

File Name	(O) Size	(C) Size	% Comp	PSNR (dB)
frm2-00.bmp	522296	42801	91.8	1302
frm2-01.bmp	522296	51473	90.1	1290
frm2-02.bmp	522296	55790	89.3	1295
frm2-03.bmp	522296	56495	89.2	1301
frm2-04.bmp	522296	56269	89.2	1292
frm2-05.bmp	522296	58300	88.8	1291
frm2-06.bmp	522296	59766	88.6	1301
frm2-07.bmp	522296	59355	88.6	1289
frm2-08.bmp	522296	59148	88.7	1301



Fig.6. Original Frame



(a)

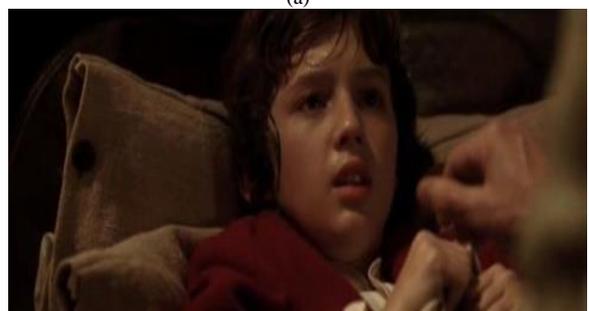


(b)

Fig.7. Reconstructed Frame



(a)

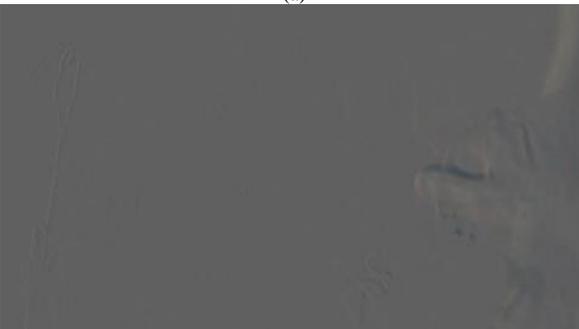


(b)

Fig.9. Original Frame



(a)



(b)

Fig.8. Compensated Frame



(a)



(b)

Fig.10. Reconstructed Frame

Figure 6, Figure 7 & Figure 8 Represents the Averaged Frame 1 with MPEG2 Codec

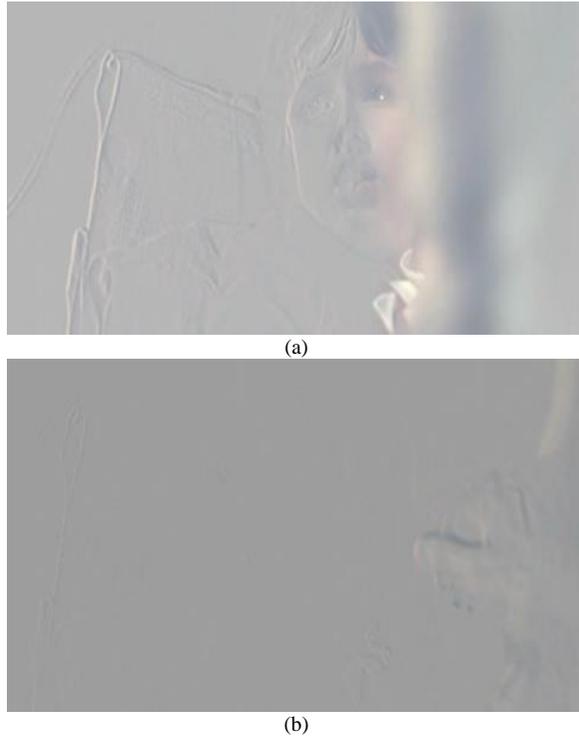


Fig.11. Compensated Frame

Figure 9, Figure 10 & Figure 11 represents the Neuro-Fuzzy Motion Compensation

Table 2. Codec with Neuro-Fuzzy Motion Compensation and Multi-Scale Wavelets

File Name	(O) Size	(C) Size	% Comp	PSNR (dB)
frm2-00.bmp	522296	7686	98.5	1603
frm2-01.bmp	522296	8071	98.5	1591
frm2-02.bmp	522296	8837	98.3	1596
frm2-03.bmp	522296	8781	98.3	1602
frm2-04.bmp	522296	9107	98.3	1593
frm2-05.bmp	522296	9165	98.2	1592
frm2-06.bmp	522296	9592	98.2	1602
frm2-07.bmp	522296	9388	98.2	1590
frm2-08.bmp	522296	9605	98.2	1602

## V. CONCLUSION

On all the image files without exception, Neuro-Fuzzy Motion Compensation produced better compression ratios for a given quality factor. For evaluation of performance in compression time, we performed the compression of several images both by Neuro-Fuzzy Motion Compensation and MPEG2 Codec 500 times in a loop to determine the average time taken. While time for compression understandably varies across different images, the performance of Neuro-Fuzzy Motion Compensation was much better than that of MPEG2

video codec. The algorithm presented here lends itself very easily for implementation both in hardware and software. Future enhancements and extensions to this research work could include deployment of an additional layer clustering based on Markov's Chains. The structure of the implementation is such that any type of transform can be employed in the Neural Network, potential candidates being *wavelets* (in all its variations), *Hadamard* and *Walsh Transforms*.

## ACKNOWLEDGMENT

We thank anonymous referees for their constructive comments. This research is supported by the Department of Electronics and Communication Engineering at K.S. School of Engineering & Management, VTU University, India with special focus on research. Special thanks to Dr. G.K Siddesh Research Guide at J S S Academy of Technical Education, India for his technical support in writing this paper.

## REFERENCES

- [1] Pinar Sarisaray Boluk and Sebnem Baydere "Robust Image Transmission Over Wireless Sensor Networks," International Journal on Mobile Network Application, Springer, pp. 149-170, 2011.
- [2] Lingyun Lu, Haifeng Du, and Ren Ping Liu "CHOKeR: A Novel AQM Algorithm With Proportional Bandwidth Allocation and TCP Protection," IEEE Transactions On Industrial Informatics, Vol. 10, No. 1, pp.637-644, February 2014.
- [3] Andreas Panayides, Zinonas C. Antoniou "High-Resolution, Low-Delay, and Error-Resilient Medical Ultrasound Video Communication Using H.264/AVC Over Mobile WiMAX Networks," IEEE Journal Of Biomedical And Health Informatics, Vol. 17, No. 3, pp 619-628, May 2013.
- [4] Syed Mahfuzul Aziz, and Duc Minh Pham "Energy Efficient Image Transmission in Wireless Multimedia Sensor Networks," IEEE Communications Letters, Vol. 17, No. 6, pp 1084-1087, June 2013.
- [5] Andrey Norkin, Gisle Bjøntegaard "HEVC Deblocking Filter," IEEE Transactions On Circuits And Systems For Video Technology, VOL. 22, NO. 12, pp 1746-1754, December 2012.
- [6] G. Nur Yilmaz, H.K. Arachchi, S. Dogan, A. Kondo "3D video bit rate adaptation decision taking using ambient illumination context," Engineering Science and Technology, an International Journal, Elsevier, pp 01-11, April 2014.
- [7] Dionisis Kandris, Michail Tsagkaropoulos, Ilias Politis, Anthony Tzes "Energy efficient and perceived QoS aware video routing over Wireless Multimedia Sensor Networks," Ad Hoc Networks Journal, Elsevier, pp 591-607, 2011.
- [8] Tao Ma, Michael Hempel, Dongming Peng and Hamid Sharif "A Survey of Energy-Efficient Compression and Communication Techniques for Multimedia in Resource Constrained Systems," IEEE Communications, pp. 01-10, 2012
- [9] Rickard Sjöberg, Ying Chen, Akira Fujibayashi "Overview of HEVC High-Level Syntax and Reference Picture Management," IEEE Transactions On Circuits And Systems For Video Technology, Vol. 22, No. 12, pp 1858-1870, December 2012.

- [10] GU Junli, SUN Yihe “Optimizing a Parallel Video Encoder with Message Passing and a Shared Memory Architecture,” Tsinghua Science And Technology, ISSN11007- 0214/109/1511Vol. 16, No. 4, pp393-398,2August 2011.
- [11] Payman Moallem, Sayed Mohammad Mostafavi Isfahani, “Facial Image Super Resolution Using Weighted Patch Pairs”, *I.J. Image, Graphics and Signal Processing*, Vol. 5, pp 1-9, march 2013.
- [12] Ashish M. Kothari, Ved Vyas Dwivedi, “ Video Watermarking – Combination of Discrete Wavelet & Cosine Transform to Achieve Extra Robustness”, *I.J. Image, Graphics and Signal Processing*, Vol. 5, pp 36-41, march 2013.
- [13] Kethepalli Mallikarjuna, Kodati Satya Prasad, “ Image Compression and Reconstruction using Discrete Rajan Transform Based Spectral Sparsing”, *I.J. Image, Graphics and Signal Processing*, Vol 8, pp. 59-67, Jan 2016.
- [14] Alireza Tofighi, Nima Khairdoost, “ A Robust Face Recognition System in Image and Video”, *I.J. Image, Graphics and Signal Processing*, Vol 7, pp. 1-11 Jul 2014.
- [15] Khaled Abid, Abdelkrim Mebarki, “ Modified Streaming Format for Direct Access Triangular Data Structures”, *I.J. Image, Graphics and Signal Processing*, Vol 6 , pp14-22, Jan 2014.
- [16] Ying Chu, Hua Mi, Zhen Ji, “BFA based neural network for image compression”, *Journal of Electronics (China)*,May 2008, Volume 25, Issue 3, Pages 405–408,2008
- [17] Robert Cierniak, Michal Knop,“Video Compression Algorithm Based on Neural Networks”,*Artificial Intelligence and Soft Computing*,Volume 7894 of the series Lecture Notes in Computer SciencePages 524-531,2013
- [18] Enrique Pelayo,, David Buldain,, Carlos Orrite, “Selective Image Compression Using MSIC Algorithm”,*Computational Intelligence*,Volume 613,Pages 419-436, 20 November 2015
- [19] FarhanHussain, “Exploiting deep neural networks for digital image compression”,*Web Applications and Networking (WSWAN)*, IEEE, 2015.
- [20] Sicheng Zhao; Hongxun Yao; YueGao; Guiguang Ding; Tat-Seng Chua,“Predicting Personalized Image Emotion Perceptions in Social Networks”,*IEEE Transactions on Affective Computing*Year: 2016, Volume: PP, Issue: 99 ,Pages: 1 - 1, October 2016.
- [21] Mingyi He; Xiaohui Li; Yifan Zhang; Jing Zhang; Weigang Wang, “Hyperspectralimage classification based on deep stacking network” *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*,Pages: 3286 - 3289, 2016
- [22] Shaima I. Jabbar; Charles R. Day; Nicholas Heinz; Edward K. Chadwick,“Using Convolutional Neural Network for edge detection in musculoskeletal ultrasound images”, *International Joint Conference on Neural Networks (IJCNN)*,Pages: 4619 - 4626, 2016.
- [23] Min Yan; Qian Yin; Ping Guo, “Image stitching with single-hidden layer feedforward neural networks”, *International Joint Conference on Neural Networks (IJCNN)* ,Pages: 4162 - 4169, 2016.
- [24] Gong Cheng; Chengcheng Ma; Peicheng Zhou; Xiwen Yao; Junwei Han, “Scene classification of high resolution remote sensing images using convolutional neural networks”,*IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*,Pages: 767 - 770, 2016.

**Authors’ Profiles**



**Prof. Prakash Jadhav** is an Assistant professor in K. S. School of Engineering and Management Bangalore, received his B.E. degree in S.T.J.Institute of Technology. Ranebennur in 1999 and M.Tech degree in A.M.C. Engineering College, Bangalore, in 2006, and pursuing Ph.D in Visvesvaraya Technological University Belgaum, Karnataka, India. He has two international publications, two international conference publications & Six National Conference publications to his credit. Research work is carried out in Multimedia Networking and Communication, and has IEEE & MISTE membership, achieved best paper presentation in “KNOWLEDGE UTSAV” conducted in S.B.M.J.C.E., Bangalore..



**Dr. Siddesh. G. K.** is an Associate Professor in JSS Academy of Technical Education Bangalore, obtained his Bachelors and Masters Degree in Electronics and Communication Engineering from Bangalore University and Manipal Academy of Higher Education , Karnataka respectively. He also obtained his doctrol degree from Visvesvaraya Technological University Belgaum, Karnataka, India. He has 5 international journal publications and one international conference publication to his credit. His research area and area of specialization includes Wireless communication, Computer Networks, Image and Data compression and Processing.

**How to cite this paper:** Prakash Jadhav, G.K.Siddesh,“Codec with Neuro-Fuzzy Motion Compensation & Multi-Scale Wavelets for Quality Video Frames”, *International Journal of Image, Graphics and Signal Processing(IJIGSP)*, Vol.9, No.9, pp.11-17, 2017.DOI: 10.5815/ijigsp.2017.09.02