

## Biological Subtle Motion Magnification Based on 3D Wavelet Transform

Yucai WEI, Shuhua XIONG, Liang LIU

Electronic Information Institute, Sichuan University,  
Chengdu 610065, China

E-mail: [yucai.lw@gmail.com](mailto:yucai.lw@gmail.com)

Received: 7 October 2013 /Accepted: 22 November 2013 /Published: 30 December 2013

**Abstract:** This paper presents an amplification algorithm of subtle motion which reveals important aspects of the world. Our analysis is based on spatial and temporal projection of decomposition of the viewpoint video by using 3D wavelet transform, which is used to estimate the optimal scale and threshold function in order to extract the interested region. We discuss parameters of effect and estimate the threshold function by employing integral energy and maximum values of pixels. The results of the subtle motion magnification by using the optimal scale and threshold function have a good effect. *Copyright © 2013 IFSA.*

**Keywords:** 3D wavelet transform, Spatial projection, Temporal projection, Threshold function, Subtle motion magnification.

### 1. Introduction

The ability to monitor a patient's physiological signals by a remote, non-contact means is a tantalizing prospect that would enhance the delivery of primary healthcare. For example, the idea of performing physiological measurements on the face was first postulated by Pavlidis and associates [1] and later demonstrated through analysis of facial thermal videos [2, 3].

The use of photo plethysmography (PPG), a low cost and non-invasive means of sensing the cardiovascular pulse wave (also called the blood volume pulse) through variations in transmitted or reflected light, for non-contact physiological measurements has been investigated recently [4–8].

The variation that human skin color varies slightly with blood circulation, while hard or impossible for humans to see, can be magnified to reveal interesting mechanical behavior and exploited to extract pulse rate [9, 10]. Subtle motion magnification analyzes and redisplay a motion signal, and thus relates to research on manipulating and redisplaying motion capture data, such as modifications to create new actions from others [12-14], and methods to alter style [15].

By recording a video of the motion region, the pixel value of certain position changes periodically along with fluctuations in motion and changes in

ambient lighting conditions. Temporal processing has been used previously to extract invisible signals [9] and to smooth motions [16]. The algorithm of Lagrangian approaches [10], works well to enhance motions of fine point features and support larger amplification factors, however it is sensitive to increase of spatial noise. The algorithm of Eulerian Video Magnification [11] combines spatial and temporal processing of videos to amplify subtle variations with good balance between performance and efficiency, while it can't satisfy with real-time processing require.

Wavelet transform recently has encountered widespread popularity in several areas of object recognition and human facial expressions analysis [17-20].

In this paper, we present a 3D wavelet's methodology for motion magnification, in which we propose to select the optimal scale which not only improves performance of results, but also enhances efficiency of motion magnification.

The characteristics of our algorithm lie in the following two aspects:

- Decomposing and reconstructing the video by the optimal scale of wavelet is used to do instead of Gaussian pyramid.
- Finding the interested domain by using the threshold function of combining integral energy and maximum values of pixels.

In addition, we describe our algorithm and demonstrate how it can compute heart rate measurements from video images of the human face recorded using a simple webcam.

The rest of this paper is organized as following. Section II gives the methodology of our work. Section III describes the wavelet transform analysis in detail and gives experimental results to validate the effectiveness of our ways. The last section presents our concluding remarks.

## 2. Methodology

### 2.1. Eulerian Motion Magnification

In this section, to explain the relationship between temporal processing and motion magnification, we consider the case of a 2D signal undergoing translational motion [11].

Let  $I(x, y, t)$  denote the image intensity at position  $(x, y)$  and time  $t$  [11]. Since the image undergoes translational motion, we can express the observed intensities with respect to a displacement function  $\phi(t)$ , such that  $I(x, y, t) = f(x + \phi(t), y + \phi(t))$  and  $I(x, y, 0) = f(x, y)$ .

Assuming the image can be approximated by a second-order Taylor series expansion, we write the image at time  $t$ ,  $f(x + \phi(t), y + \phi(t))$  in a second-order Taylor expansion about  $(x, y)$ , as

$$f(x + \phi(t), y + \phi(t)) \approx f(x, y) + \phi(t) \left( \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} \right). \quad (1)$$

Let  $B(x, y, t)$  be the result of applying the threshold function to  $I(x, y, t)$  at every position  $(x, y)$ . For now, let us assume the motion signal,  $\phi(t)$ , is within the threshold function. Then we have

$$B(x, y, t) = \phi(t) \left( \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} \right). \quad (2)$$

In our process, we then amplify that interested domain of threshold function by  $\alpha$  and add it back to  $I(x, y, t)$ , resulting in the processed signal

$$\bar{I}(x, y, t) \approx I(x, y, t) + \alpha B(x, y, t). \quad (3)$$

Combining (1), (2), and (3), we have

$$\bar{I}(x, y, t) = f(x, y) + (1 + \alpha)\phi(t) \left( \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} \right). \quad (4)$$

Assuming the second-order Taylor expansion holds for the amplified larger perturbation,  $(1 + \alpha)\phi(t)$ , we can relate the amplification of the

temporal signal to motion magnification. The processed output is simply

$$\bar{I}(x, y, t) \approx f(x + (1 + \alpha)\phi(t), y + (1 + \alpha)\phi(t)). \quad (1)$$

This effect is demonstrated on (5). The input signal is  $I(x, y, t)$ , which is magnified to be  $\bar{I}(x, y, t)$ . The interested domain of choosing by the threshold function is amplified and added to the original signal to generate a larger translation.

### 2.2. 3D Wavelet Transform

#### 2.2.1. Decomposition of 3D Wavelet Transform

3D Wavelets present a division of video spectrum into multi-scale sub-bands for temporal and spatial information and oriented sub-bands for spatial information. This transform is separable and the decomposition is done by passing through a 3D-filter channel bank. In each scale 8 sub-bands are created and the next scale of decomposition is done only on the lowest sub-band [21]. The implementation of one scale 3D wavelet transform is shown in Fig. 1.

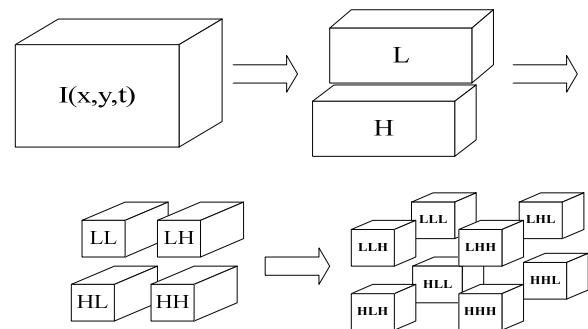


Fig. 1. Decomposition of one scale 3D Wavelet Transform.

#### 2.2.2. Reconstruction of 3D Wavelet Transform

The 3D sub-band reconstruction by the proposed three scale 3D wavelet transform is depicted by Fig. 2. We can reconstruct the third scale by combining 8 blocks noted L3. Then reconstructing the second scale is easy by the same way.

Firstly, we perform wavelet transform in the time axis to find interested frequency motion. The resulting low frequency sub-band in time contains motion components such as gentle head motions, slow changes in lighting, and local movements of relatively large objects. After the temporal wavelet transform is completed we then perform 2D wavelet decomposition in the xy-plane. We want to use the spatial scale to improve the efficiency of processing.

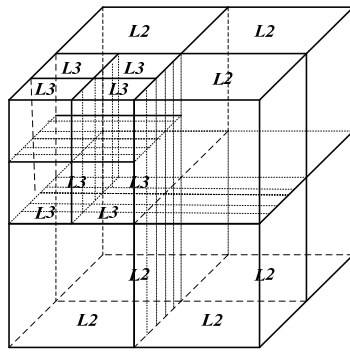


Fig. 2. Reconstruction of three scale 3D Wavelet Transform.

### 3. Results and Analysis

#### 3.1. Overview of the 3D Wavelet Transform Video Magnification Framework

The system first decomposes the input video sequence into different frequency bands, and applies the analysis of 3D wavelet spatial projection to all bands to choose the best wavelet scale. Then we use the threshold function in the analysis of 3D wavelet temporal projection to find the interested bands, as shown in Fig. 3.

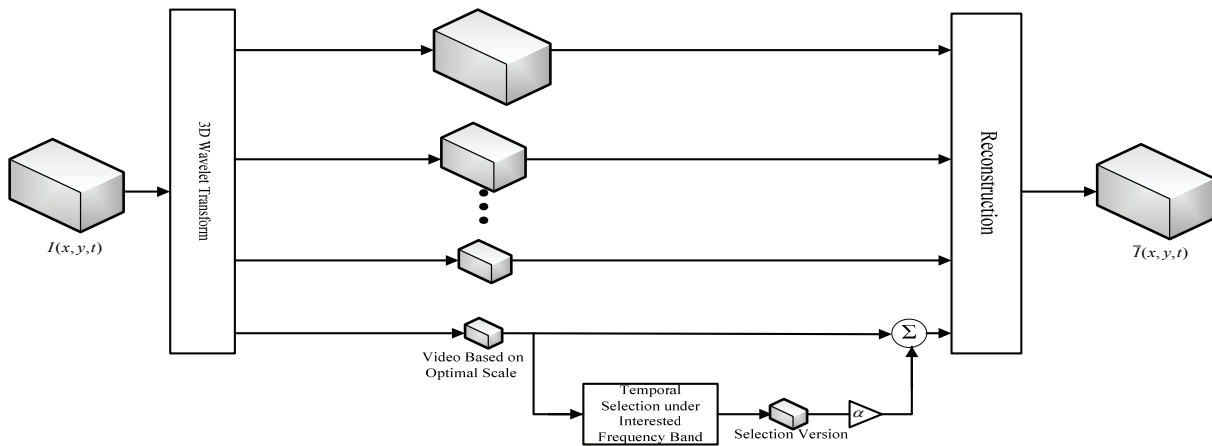


Fig. 3. Overview of the 3D wavelet transform video magnification framework.

The bands are then amplified by a given factor, added back to the original signal, and collapsed to generate the output video. The choice of threshold function and amplification factors can be tuned to support different applications.

#### 3.2. The Analysis of 3D Wavelet Spatial Projection

After spatial projection of decomposition of the viewpoint video using the 3D wavelet transform, the resulting scale we need is darkened, as shown in Fig. 4.

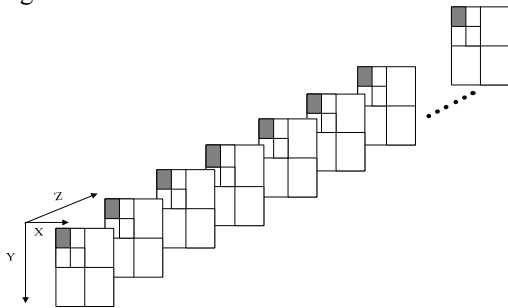


Fig. 4. Spatial projection of 2-levels 3D-DWT on viewpoint images.

The lowest frequency sub-band is a coarse scale approximation of the image of original viewpoint video and the rest of the frequency bands are detail signals. The wavelet transform can be applied recursively to the lowest frequency sub-band to obtain decomposition at coarser scales. After decomposition of the image of viewpoint video using the wavelet transform, the resulting lowest frequency sub-bands are assembled, as shown in Fig. 4 which is darkened and. Then, we reconstruct the certain scale by wavelet coefficients which can help us get the best effect.

The result of the different scale for test videos used here are a multiple of 592\*528\*250, which simplifies the computation of the 3D wavelet transform, as shown in Fig. 5.

The result of the different scale for test videos used here are a multiple of 592\*528\*240, which simplifies the computation of the 3D wavelet transform, as shown in Fig. 5. For this experiment, the scale used in the wavelet transform process has been chosen is 2, 3, 4, 5, 6, 7, 8 and 9. The temporal frame of viewpoint video in the Fig. 5 has been chosen is 0'7'', 01'10'', 02'12'', 03'14'', 04'11'', 05'17'', 06'15'' and 07'21''.



Fig. 5 (a). Results of different scale by 3D wavelet transform  $N = 2$ .

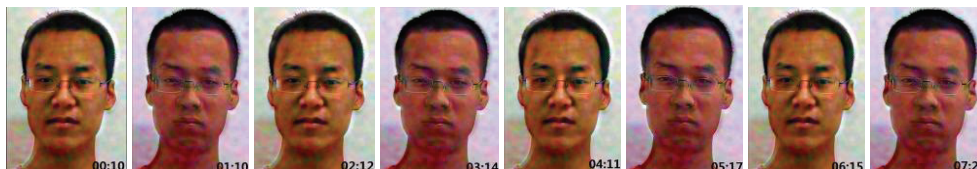


Fig. 5 (b). Results of different scale by 3D wavelet transform  $N = 3$ .



Fig. 5 (c). Results of different scale by 3D wavelet transform  $N = 4$ .



Fig. 5 (d). Results of different scale by 3D wavelet transform  $N = 5$ .



Fig. 5 (e). Results of different scale by 3D wavelet transform  $N = 6$ .



Fig. 5 (f). Results of different scale by 3D wavelet transform  $N = 7$ .



Fig. 5 (g). Results of different scale by 3D wavelet transform  $N = 8$ .

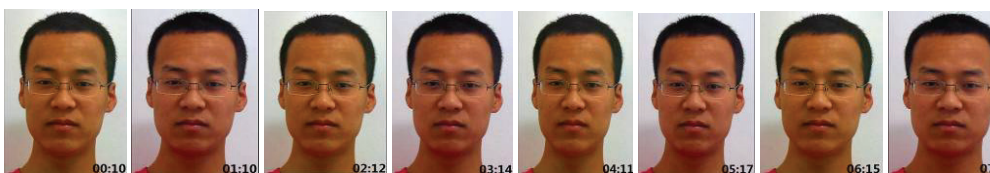


Fig. 5 (h). Results of different scale by 3D wavelet transform  $N = 9$ .

We see different results of the motions or signals that we wish to be amplified in Fig. 5. The result of second scale of reconstruction is generally apparent with the noise. A vague result of second scale of reconstruction for the bands resulting from the application of the spatial wavelet transform can be observed. As the scale of reconstruction increasing, the noise of facial color magnification reduced gradually, while the effect of facial color magnification is gradually unobvious. At the ninth scale of the wavelet transform, almost obvious noise are filtered. Because the domain of magnification focuses on low frequency while the noise concentrate on high frequency.

The results were generated using non-optimized MATLAB code on a machine with a four-core processor and 4 GB RAM.

We used the sym4 of wavelet coefficients to construct the different scale. The video is 592\*528 and 280 frames in total. The computation time is shown in Fig. 6. We choose the fifth scale to be the best reconstruction by combining the effect and the computation time of facial color magnification.

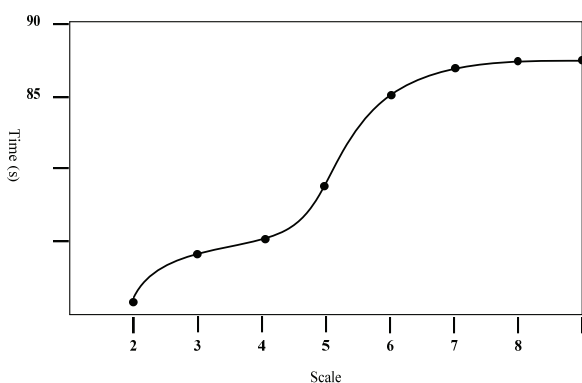


Fig. 6. Computation of different scale.

### 3.3. The Analysis of 3D Wavelet Temporal Projection

In addition to spatial projection of decomposition of the viewpoint video using the 3D wavelet transform, we can analyze the time projection to obtain the threshold function for interested pixels.

After the time projection of the viewpoint video using the 3D wavelet transform, the value of certain pixel we are interested in is changing with time, as shown in Fig. 7. We can see the intensity of pixel at  $(x_0, y_0)$  is different with frame changing. It is obvious that this variation of intensity of certain position is periodic.

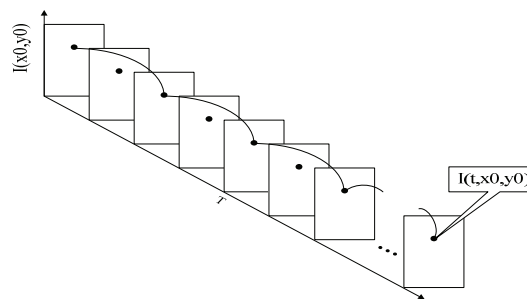


Fig. 7. The intensity of pixel at  $(x_0, y_0)$  with time  $t$ .

We used the pixel' intensity of 3D wavelet transform reconstruction to confirm the accuracy of our interested region and to verify that the color amplification signal extracted from our method matches the photo plethysmogram.

We can see the difference between interested and uninterested region, as shown in Fig. 8. As we can see from solid line which is intensity of interested pixel, the max intensity of pixel is  $4.5 \times 10^{-3}$  and the min value is  $-4.7 \times 10^{-3}$ . While the dotted line is intensity of uninterested pixel, the max value is  $1.8 \times 10^{-3}$  and the min value is  $-1.2 \times 10^{-3}$ . From the data and the periodic, we can use the integral energy of pixels and the sum of maximum values of pixels to design the threshold function model to distinguish between the different interested and uninterested pixels.

The method to find out all pixels we are interested in is how we obtain the threshold function.

Step 1. Domain  $[T_1, T_2]$ .

$$T_1 = N_t / 2 - \min(f_1, f_2) \times F_r$$

$$T_2 = N_t / 2 + \max(f_1, f_2) \times F_r$$

$N_t$  represents the total frames of the video,  $(f_1, f_2)$  means the interested physical frequencies and  $F_r$  is the video frame rate.

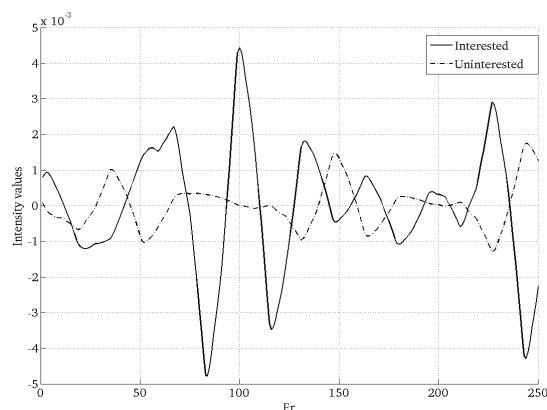


Fig. 8. The intensity of interested and uninterested pixel by DWT reconstruction.

Step 2. Integral energy.

The integral energy of interested pixels is larger than uninterested in the selected band at  $[T_1, T_2]$  [22]. The definition of integral energy is

$$E = \sum_{n=T_1}^{T_2} |f(n)|^2.$$

As shown in Fig 9.(a), the effect of finding interested region is not obvious because of the noise.

Step 3. Sum of maximum values

The sum of maximum values of interested pixels is larger than uninterested in the selected band at  $[T_1, T_2]$  [23]. The definition of integral energy is

$$M = \sum_{n=T_1}^{T_2} (f(n))_p.$$

$( )_p$  represents the symbol of calculating maximum values [22]. As shown in Fig 9.(b), the effect of finding interested region is vague.

Step 4. Threshold function

$$F = aE + bM \quad (a + b = 1)$$

We evaluate effect for motion magnification, we used several different values about  $(a, b)$ . For all values, the best effect to find out the interested region is

$$F = 0.7E + 0.3M.$$

As shown in Fig 9 (c).

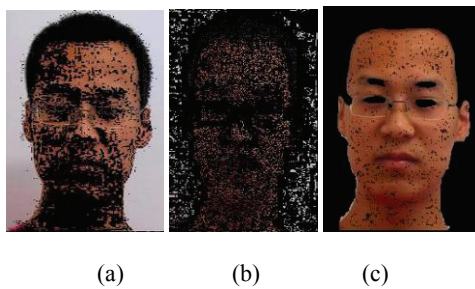
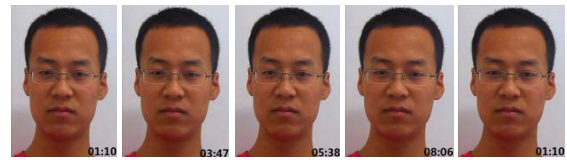


Fig. 9. The interested region based on different threshold function.

## 4. Results

We amplified facial color, human pulse and other subtle motions changes in our experiment. In order to demonstrate the capability of 3D wavelet transform methodology to magnify subtle motions of persons obviously, we showed the facial color magnification. Fig. 10 shows the original and magnification results measured by the proposed technique (wavelet transform magnification), as well as by reference EVM.



(a) Input



(b) Magnified

Fig. 10. An example of using 3D wavelet transform magnification for visualizing the subtle motion of the human face. (a) Eight frames from the original video sequence (face). (b) The same eight frames with the subject's face signal amplified.

## 5. Conclusions

We described a straightforward method that takes a video as input and exaggerates subtle color changes by 3D wavelet transform. This 3D wavelet transform method, which combines processes and spatial processes, successfully reveals informative signals and amplifies small motions in real-world videos. Using 3D wavelet transform, the effect is improved, comparing with EVM. To amplify color motion, the method does not perform feature tracking or optical flow computation by using wavelet transform processing.

## Acknowledgment

We would like to thank to Fund: European Union FP7-PEOPLE-IRSES-S2EuNet (No.247083). Also we would like to thank Xu Sun and Menghan Li for their helpful feedback. We thank Xiaogang Huang for helpful discussions on the 3D wavelet transform analysis. We also thank Yali Ma and Wei Fan for helping us collect videos of experiment. The opinions expressed here are those of the authors and may or may not reflect those of the sponsoring parties.

## References

- [1]. I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey, Interacting with human physiology, *Computer Vision Image Understanding*, Vol. 108, Issue 1-2, 2007, pp. 150–170.
- [2]. M. Garbey, N. Sun, A. Merla, and I. Pavlidis, Contact-free measurement of cardiac pulse based on the analysis of thermal imagery, *IEEE Transactions on Biomedical Engineering*, Vol. 54, Issue 8, 2007, pp. 1418–1426.
- [3]. J. Fei, and I. Pavlidis, Thermistor at a distance: unobtrusive measurement of breathing, *IEEE Transactions on Biomedical Engineering*, Vol. 57, Issue 4, 2010, pp. 988–998.

- [4]. F. P. Wieringa, F. Mastik, and A. F. van der Steen, Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO<sub>2</sub> camera" technology, *Annals of Biomedical Engineering*, Vol. 33, Issue 8, 2005, pp. 1034–1041.
- [5]. K. Humphreys, T. Ward, and C. Markham, Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry, *Review of Scientific Instruments*, Vol. 78, Issue 4, 2007, 044304.
- [6]. C. Takano, and Y. Ohta, Heart rate measurement based on a time-lapse image, *Medical Engineering & Physics*, Vol. 29, Issue 8, 2007, pp. 853–857.
- [7]. S. Hu, J. Zheng, V. Chouliaras, and R. Summers, Feasibility of imaging photoplethysmography, in *Proceedings of the IEEE Conference on BioMedical Engineering and Informatics*, 2008, pp. 72–75.
- [8]. W. Verkruyse, L. O. Svaasand, and J. S. Nelson, Remote plethysmographic imaging using ambient light, *Optics Express*, Vol. 16, Issue 26, 2008, pp. 21434–21445.
- [9]. M.-Z. Poh, D. J. McDuff, and R. W. Picard, Non-contact, automated cardiac pulse measurements using video imaging and blind source separation, *Optics Express*, Vol. 18, Issue 10, 2010, 10762–10774.
- [10]. C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, Motion magnification, *ACM Transactions on Graphics*, Vol. 24, Issue 3, 2005, pp. 519–526.
- [11]. Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Gutttag, Fredo Durand, and William T. Freeman, Eulerian video magnification for revealing subtle changes in the world, *ACM Transactions on Graphics*, Vol. 31, Issue 4, 2012.
- [12]. O. Arikan, and D. A. Forsyth, Synthesizing constrained motions from examples, *ACM Transactions on Graphics*, Vol. 21, Issue 3, July 2002, pp. 483–490.
- [13]. J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard, Interactive control of avatars animated with human motion data, *ACM Transactions on Graphics*, Vol. 21, Issue 3, July 2002, pp. 491–500.
- [14]. K. Pullen, and C. Bregler, Motion capture assisted animation: texture and synthesis, *ACM Transactions on Graphics*, Vol. 21, Issue 3, July 2002, pp. 501–508.
- [15]. M. Brand and A. Hertzmann, Style machines, in *Proceedings of the ACM 27<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH*, 2000, pp. 183–192.
- [16]. M. Fuchs, T. Chen, O. Wang, R. Raskar, H.-P. Seidel, and H. P. Lensch, Real-time temporal shaping of high-speed video streams, *Computers & Graphics*, Vol. 34, Issue 5, 2010, pp. 575–584.
- [17]. V. Zlokolica, A. Pizurica, W. Philips, Wavelet-domain video denoising based on reliability measures, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 16, Issue 8, August 2006, pp. 993-1007.
- [18]. L. E. Coria, M. R. Pickering, P. Nasiopoulos, R. K. Ward, A video watermarking scheme based on the dual-tree complex wavelet transform, *IEEE Transactions on Information Forensics and Security*, Vol. 3, Issue 3, September 2008, pp. 466-474.
- [19]. M. Sablatash, T. Cooklev, T. Kida, The coding of image sequences by wavelets, wavelet packets and other subband coding schemes, *IEEE Transactions on Broadcasting*, Vol. 40, Issue 4, December 1994, pp. 201-205.
- [20]. M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Transactions on Image Processing*, Vol. 4, Issue 11, November 1995, pp. 1549-1560.
- [21]. S. Lian, J. Sun, and Z. Wang, A secure 3D-SPIHT codec, in *Proceedings of the European Signal Processing Conference*, September 2004, pp. 813-816.
- [22]. P. Midya, P. Wagh, P. Rakers, Quadrature integral noise shaping for generation of modulated RF signals Circuits and Systems, in *Proceeding of the IEEE Congress on MWSCAS*, October 2002, Vol. 2, pp. 537-540.
- [23]. Byeong Seok Ahn, Generation of OWA operator weights based on extreme point approach, in *Proceeding of the IEEE Congress on FSKD*, July 2009, pp. 196-199.