Journal of Theoretical and Applied Information Technology
31st March 2013. Vol. 49 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MOTION ESTIMATION BASED VIDEO STABILIZATION FOR OLD ANIMATION VIDEO USING WAVELET TRANSFORM

DWI RATNA SULISTYANINGRUM[1], MOCHAMAD HARIADI[2], MAURIDHI HERY PURNOMO[3]
[1]Department of Mathematics Sepuluh Nopember Institute of Technology, Surabaya – Indonesia
[2,3]Department of Electrical Engineering Sepuluh Nopember Institute of Technology, Surabaya – Indonesia
E-mail: [1]dwratna@gmail.com

ABSTRACT

Video stabilization is one of the important techniques in video restoration researches. Restoration of visual quality in old animation video is necessary due to some artefacts occur since the making of the animation is involving hand-drawn pictures followed by capturing step by camera. These artefacts lead to the unsteadiness and image vibration of the animation scene even in steady condition of capturing step. Video stabilization method aims to reduce those artefacts. This paper proposes video stabilization algorithm in old animation video based on motion estimation. The motion estimation technique employs coarse to fine method based on wavelet transform. The proposed video stabilization algorithm includes local motion estimation, global motion estimation, motion smoothing, and motion compensation. The superior performance of our algorithm over other method for video stabilization in old animation video is demonstrated through simulated experiment. The phase-correlation based algorithm presents ITF increments of about 0.74 -4.56 %, while our proposed algorithm improvements in the ITF of about 4.37-18.53 %.

Keywords: *Motion Estimation, Video Stabilization, Wavelet Transform, Motion Smoothing,Video Restoration.*

1. INTRODUCTION

Animated film is an interesting spectacle for children or adults. It is considered to create a real new animated video; it will be more efficient if we use old animated film. Although the old films have some kind of defect such as intensity flicker, noise, blotches, film unsteadiness and image vibration [1][3]. In order the old animated film still preferred by viewers, visual and audio quality of the old films must satisfy the quality standards which desired by modern audiences.

Film unsteadiness occurs when each frame is displaced with a different offset compared to the neighboring frames [2]. Image vibration can be defined as a rapid spatial random drift and/or rotation between consecutive frames [3]. Film unsteadiness and image vibration are caused by damage at hole punched film, the lack of film transport mechanism or the unstable camera [2][3]. Those defects will lead to the vibrated video.

The vibration disturbance of the old animation video is due to the creation of animation employ human hand-draw and the image capture by camera (rostrum camera). The rostrum camera is a specific type of camera used to film 2D animation, one frame at a time.

Video stabilization generally consists of three main steps [3][4][5][6]: motion estimation, motion filtering/smoothing and image compensation. Motion estimation represents a process to determine the frame displacement of the video sequence. The results of this process will be used for extract the global motion. Motion filtering aims to separate the wanted and unwanted motion. Undesired motion is generally regarded as a high-frequency motion, so the motion filtering is called motion smoothing. Image compensation aims to build new frames from an old video based on refined motion parameters. In this paper, we focus on motion estimation which is resistant to noise and flicker.

There are several motion estimation methods such as block matching algorithm (BMA), gradient and phase correlation. BMA is the most popular, because it is easy to be implemented in software or hardware.

De Juan and Bodenheim [7] use a gradient-based method for motion estimation. This method requires low computation, but it can not handle the displacements greater than 5 pixels [1].

The motion estimation technique employs Phase-Only Correlation (POC) for estimating frame displacement in old film[8]. POC approach

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

can estimate the frame displacement with high accuracy if successive frames have a highly degree of similarity. However the old animation video is drawn by hand, so it has low degree of similarity.

The multiresolution block matching algorithm (RBMA) able to maintain good performance with very high computational speed even for large search area [9]. The multiresolution concept is structure of the wavelet representation

The wavelet-based motion estimation method has been applied for video coding and compression [10][11][12]. This method has a better performance than the estimation motion in the spatial domain. Zhang [11] proposed a video compression scheme based on the wavelet representation and multi-resolution motion estimation. Their method used wavelet transform to decompose a frame into a set of sub-frames with different resolutions.

The research on video stabilization in old video has been performed by few researchers. The histogram of motion vectors method has been proposed for global motion estimation [13]. The global motion between two frames is defined as the most frequent apparent motion. The obtained Global motion is not accurate when object area larger than background area. The adopted motion model in [13] is translational.

Liscar et al [3] proposed video stabilization method using a phase-based motion estimation in multiscale by automatically determination of Region of Fixation (ROF). The adopted global motion model is rotational and translational.

In this paper, we present a method of video stabilization in the old animation video using motion estimation based on wavelet to eliminate the effect of vibration that interfere the video quality. In the proposed technique, we use motion estimation method by approach coarse-to-fine based on wavelet transform. The six-parameter affine of global motion model is obtained from the motion estimation.

2. THE PROPOSED VIDEO STABILIZATION ALGORITHM

Figure 1 shows the diagram of the proposed video stabilization algorithm. Step 1, the motion vector between two consecutive frames is estimated by motion estimation based on wavelet transform. Step 2, the six-parameter affine of global motion model is determined using RANSAC. Step 3, the global motion parameters are accumulated and smoothed with a Gaussian filter. The final step, the frames are compensated based on the refined parameters to form a stable video. Details of the procedure are described as follows.

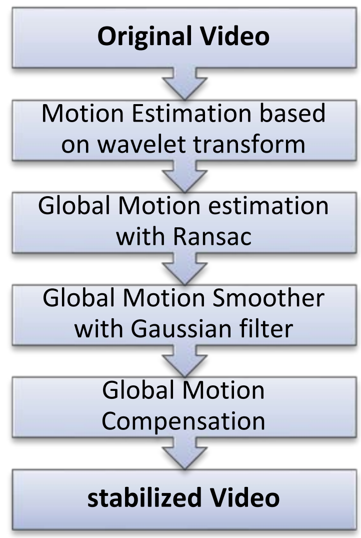

Figure 1 Diagram Of Proposed Video Stabilization Algorithm.

2.1 Motion Estimation Based On Wavelet Transform

Motion estimation in wavelet domain has a lot of attention rather than spatial domain, because it has better performance, especially in video coding field.

Motion estimation with the coarse-to-fine approach reduces the computational. In this paper, we use algorithms coarse-to-fine based on wavelet transform. The proposed scheme similar to that used in [10][11]. The motion estimation in the propose scheme is only work on the low-pass band.

Discrete wavelet transform (DWT) of the two-dimensional image can be expressed as a sequence of sub-image of the following:

$$\left\{S_{2^M}, \left[W^j_{2^1}\right]_{j=V,H,D}, \cdots, \left[W^j_{2^M}\right]_{j=V,H,D}\right\} \qquad (1)$$

where S_{2^k} is the set of approximation $f(x, y)$ on the resolution 2^{-M}. $W^j_{2^M}$ is a detail image at a resolution of 2^{-M} at location j with the *V*, *H* and *D* respectively are the vertical, horizontal and diagonal. S_0 is the original image and S_2 is an approximation of in next resolution 2^{-1}

Motion vector in a given resolution level, can be predicted from the motion vector at lower resolution multiplied by two. A smoothness factor is added to this prediction after carrying out a motion search nearby the motion prediction. Let

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$V_k(x,y)$ represent the motion vectors centered at *(x, y)* for the sub-image at level *k*. Then estimation of $V_k(x,y)$ is given by

$$V_k(x,y) = 2V_{k+1}(x,y) + \delta_k(x,y) \quad (2)$$

$$for \quad k = 0,1,2,\ldots,M$$

where M is the level of decomposition and δ_k is a smoothing factor that is determined by the motion estimation nearby the predicted value.

Figure 2 shows the structure of coarse-to-fine based on wavelet transform algorithm which proposed. First time, motion vector is predicted on lowest resolution. Motion vectors in higher resolution is the smoothing of motion vectors of its lower resolution.

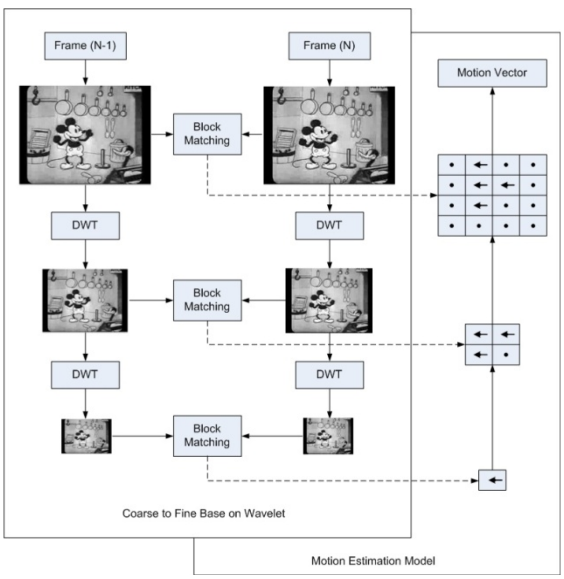

Figure 2 The Structure of The Coarse-to-Fine Based on Wavelet Transform Algorithm

Motion vectors of every pixel in the image can be obtained through block matching. In this paper, a matching function which used for local motion estimation is the mean absolute difference (MAD). Given the matching block size N x M then MAD is defined as follows

$$MAD(dx,dy) = \frac{1}{NM}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|I_k(m,n) - I_{k-1}(m+dx,n+dy)\right| \quad (3)$$

where $I_k(m,n)$ is the pixel value at coordinates (m,n) in frame-(k), $I_k(m+dx,n+dy)$ is the pixel value at coordinates $(m+dx,n+dy)$ in frame-(k+1). The motion vector of pixel at coordinates (m,n) is given by:

$$(MV_x, MV_y) = \min_{(dx,dy)\in R^2} MAD(dx.dy)$$

2.2 Global Motion Estimation

Global motion vector represents camera motion of current frame and previous one, that formulated by motion parameter models. The process of determining value of global motion vector is called Global Motion Estimation (GME).

In this paper, RANSAC algorithm [14] is used to extract global motion parameters of motion vector field that produced by a local motion estimation algorithm explained in section 2.1.

This study assumed that objects are not in the corners of image field so that local motion estimation is carried out from all of the corner areas that shown in Figure 3. This assumption is aimed to decrease computation time.

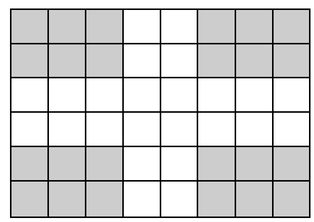

Figure 3 Optimization of Area Local Motion Estimation

Considering the complexity and durability, this study uses simplified affine motion, because of this model can accommodate all of scaling, rotation and translation. Simplified affine motion is expressed by four parameters, named S, θ, t_x and t_y that formulated by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} S\cos\theta & -S\sin\theta \\ S\sin\theta & S\cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (4)$$

with $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} x' \\ y' \end{bmatrix}$ are the two corresponding pixel coordinate in two frame, S is responsible for scaling, θ is responsible for rotation, t_x and t_y is responsible for translation in both horizontal and vertical direction.

Given $a = S\cos\theta$ and $b = S\sin\theta$ then equation (4) can be written

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (5)$$

Assuming there are N motion vector associated with the image of at the position (x_i, y_i) and $(x_i^{'}, y_i^{'})$ to $i = 1,2,3,\cdots,N$ of two consecutive frames. The simplified affine motion between two consecutive frames of N motion vectors is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

estimated by the solution the following linear equation.

$$\cdot\begin{pmatrix} x_1 & -y_1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & -y_n & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_n & x_n & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ t_x \\ t_y \end{pmatrix} = \begin{pmatrix} x_1^{'} \\ \vdots \\ x_n^{'} \\ y_1^{n} \\ \vdots \\ y_n^{'} \end{pmatrix} \quad (6)$$

Linear systems in the form of $Ax = B$ and its solution can be solved by RANSAC method.

2.3 Motion Smoothing

Video has high quality if inter-frame motion is slow and smooth. Vibrated videos are caused by fluctuation camera motion. A steady motion can be obtained by removing the fluctuation motion by using motion smoothing.

Let T_i is a transformation that associated with camera motion from frame-(i) to frame-(i+1). According to equation (4), simplified affine motion between frames-(i) and the frames-(i+1) can defined as,

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_i & -b_i & t_{xi} \\ b_i & a_i & t_{y_i} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

and transformation T_i is defined as

$$T_i = \begin{bmatrix} a_i & -b_i & t_{xi} \\ b_i & a_i & t_{y_i} \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

The low-pass filtering approach can be applied to produce smoothed transformation $\tilde{T}_i$ with following formula:

$$\tilde{T}_i = T_i * G(x) \quad (9)$$

where $*$ is convolution operator and $G(x)$ is the Gaussian kernel given as

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{(-x^2/2*\sigma^2)} \quad (10)$$

where σ is the standard deviation of the distribution. Parameter of Gaussian filter is set as $\sigma = \sqrt{k_t}$ [15], with k_t is the length of the Gaussian window filter. Gaussian window size very affect to the results of the smoothing process. If k_t is set in high number, it will cause problems oversmoothing. Otherwise if k_t regulated small number, it will cause problems undersmoothing.

2.4 Motion Compensation

The last step of video stabilization is motion compensation that creates the steady frame based on current frame and use the smoothed global motion parameter.

Chang et.al [16] developed the method to generate steady frame based on the difference of global motion value between before and after motion smoothing process. In figure 4, $I_i, i = 1,2,3,\ldots n$ represents the original frames of the video and $\bar{I}_i, i = 1,2,3,\ldots n$ are corresponding steady frames.

A Steady frame can be built by warping the involved original frame. If T_i is an affine transformation that defined by equation (8) then the relation between two consecutive frame is calculated by following equation:

$$I_{i+1}(x', y', 1) = I_i(T_i \cdot (x, y, 1)^T) \quad (11)$$

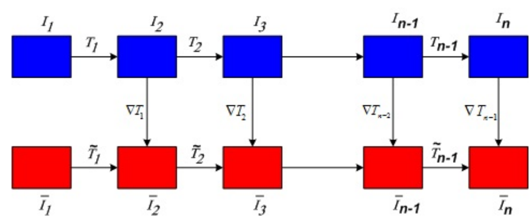

Figure 4 The relation of the original and stabilized frame.

The affine transformation between two frames can be extended to gets accumulated affine transformation between the frame-(n) and the first frame as follows:

$$T_n^A = \prod_{i=1}^{n} T_i \quad (12)$$

Note that the tranformation T_n^A is still a simplified transformation. Similarly, the affine transformation between the stabilized frame-(n) and the first frame can be written as:

$$T_n^{A,S} = \prod_{i=1}^{n} \tilde{T}_i \quad (13)$$

Therefore, the camera motion transformation between the original input frames and the stabilized frame can be derived from the following equation:

$$T_n^A(\nabla T_n) = T_n^{A,S}$$
$$(\nabla T_n) = (T_n^A)^{-1} T_n^{A,S} \quad (14)$$

The stabilized frame can be generated by using the affine transformation ∇T_n. The relationship

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

between the original frame I_i and the stabilized frame $\bar{I}_i$ can be defined as follows:

$$\bar{I}_i(x,y,1) = I_i((\nabla T_n)\cdot(x,y,1)^T)) \quad (15)$$

In those equation, the warping from I_i to $\bar{I}_i$ using image interpolation.

3. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of proposed video stabilization algorithm. In our simulation, we test four old animation video taken from youtube website. They have different features, as illustrated in Fig. 5, where the "Felix" and "Charlie" have stationary background and stationary object, while the "Popeye" and "Mickey" have stationary background and moving object.

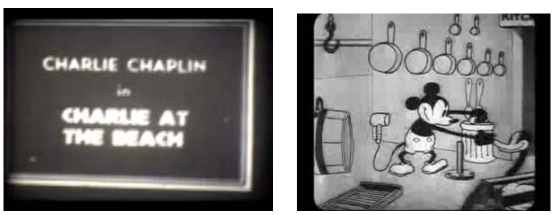

Figure 5 Two Sample frame

Performance of video stabilization is measured by the temporal smoothness. In this paper, the Interframe Transformation Fidelity (ITF) [17] is used as a temporal smoothness measure which is defined as follows:

$$MSE(k) = \frac{1}{NM}\sum_{x=1}^{N}\sum_{y=1}^{M}(I_k(x,y) - I_{k+1}(x,y))^2$$

$$PSNR(k) = 10\times\log_{10}\left(\frac{255^2}{MSE(k)}\right) \quad (16)$$

$$ITF = \frac{1}{Nframe-1}\sum_{k=1}^{Nframe-1}PSNR(k)$$

where MSE is the Mean Square Error and Nframe is the number of frames in the video. Typically, a stabilized video has higher ITF values than original video.

The examples of four smoothed parameters are shown in Figure 6, with the smoothed parameter using Gaussian filter. In the smoothing motion process, the value of k_t be set to 5, 15 and 25. The figure show that the smoothed parameter is almost similar if k_t set in 15 and k_t set in 25. In this experiment, the value of k_t be set to 15.

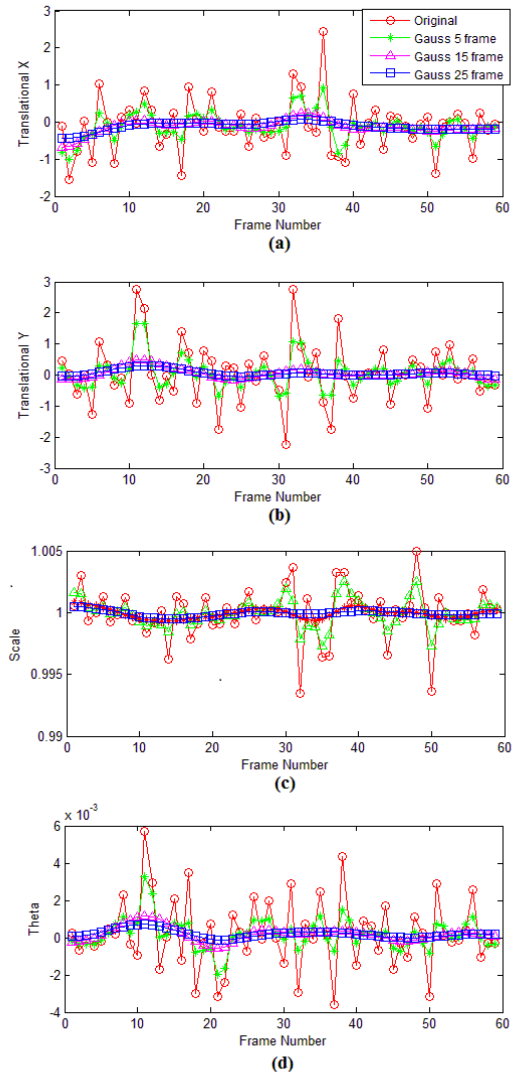

Figure 6 The Global Motion Before and After Smoothing for Different Filter Lenght, (a) Transitional X, (b) Transitional Y, (c) Scale, and (d) Rotational θ

In this experiment, we compare the performance of our proposed algorithm with video stabilization based on phase correlation. In the video stabilization algorithm with phase correlation, the motion estimation using phase correlation based block matching [8][3].

Table 1 shows ITF values for original sequences, stabilized sequences with phase correlation and with wavelet transform. It is observed that the reconstructed video have improved visual quality and the ITF. The ITF improvement for proposed algorithm is more than the corresponding ITF gains of Phase-correlation.

Journal of Theoretical and Applied Information Technology
31st March 2013. Vol. 49 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1. ITF comparison of stabilized video with phase correlation and wavelet transform.

FILE	Original ITF	Phase correlation ITF	Wavelet transform ITF
Felix	26.60	26.82	31.53
Charlie	25.62	26.79	27.37
Popeye	24.25	24.43	25.31
Mickey	19.77	20.13	20.66

Figure 7 show the difference of PSNR between the original frames and the stabilized frames with phase correlation and with wavelet transform. We observe that PSNR of stabilized videos produced by the two methods are higher than original video. Among the two, video stabilization with wavelet transform gives the highest PSNR.

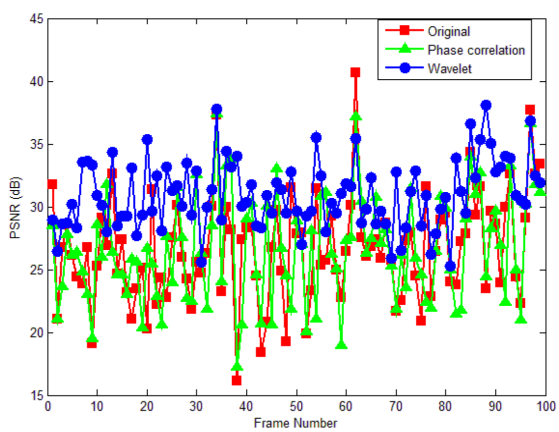

Figure 7 PSNR Comparison of The Phase Correlation and Wavelet Algorithm for The Felix Sequences.

4. CONCLUSION

In this paper, we presented a video stabilization method in the old animation video using motion estimation based on wavelet transform. The simulation results show that performance of the wavelet transform-based algorithm better than the phase-correlation based algorithm. The phase-correlation based algorithm presents ITF values increments of about 0.74-4.56%, while our proposed algorithm improvements in the ITF of about 4.37-18.53 %. The experiment results indicate that the proposed wavelet based video stabilization algorithm has the ITF 31.53 dB for Felix sequence, 27.37 for Charlie sequence, 25.31 for Popeye sequence and 20.66 for Mickey sequence.

REFERENCES

[1] A.C. Kokaram, *Motion Picture Restoration. Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*, Springer-Verlag London, 1998.

[2] A. Rares and J. Biemond, "Archived Film Analysis and Restoration", TU Delft, 2004.

[3] A. Licsár, L. Czúni, and T. Szirányi, "Adaptive Stabilization of Vibration in Archive Films", *Proc. CAIP*, 2003, pp.230-237.

[4] S. Gomez, V. Naranjo, A. Albiol, and J. M. Mossi, "Extracting Random Vibration Components from Global Motion Vectors", *Proceedings of the The 5th IASTED International Conference on Visualisation, Imaging and Image Processing*, Benidorm, Spain, September 2005.

[5] J. Yang, D. Schonfeld, and M. Mohamed, "Robust Video Stabilization Based on Particle Filter Tracking of Projected Camera Motion", *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 19, no.7, july 2009, pp. 945-954.

[6] D. Pang, H. Chen, and S. Halawa, "Efficient Video Stabilization with Dual-tree Complex wavelet transform," *EE368 Project Report, Spring*, 2010.

[7] C. N. de Juan and B. Bodenheimer, "Re-using Traditional Animation: Methods for Semi-automatic Segmentation and Inbetweening", *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, Aire-la-Ville, Switzerland, Switzerland, 2006, pp. 223–232.

[8] M. Hagiwara, M. Abe and M. Kawamata, "Estimation Method of Frame Displacement for Old Films using Phase-only Correlation", *Journal of Signal Processing*, 8, 5, 2004, pp.42-429.

[9] J. Kim, C. Yang, D. Lim and Y. Ho, "Multiresolution Motion Compensation in the Wavelet Domain for Scalable Video Coding", *Visual Communication and Image Processing*, Vol. 4671, 2002, pp.1103-114.

[10] S. Usama, B. Šimák, "A Low-complexity Wavelet Based Algorithm for Inter-frame Image Prediction", *Acta Polytechnica* Vol. 42 No. 2, 2002.

[11] Y.Q. Zhang, and S. Zafar, "Motion Compensated Wavelet Transform Coding for Color Video Compression", *IEEE Trans., Circuits and Systems for Video Tech.*, Vol. 2, Sept. 1992, pp. 285-296.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12] B. Sriram, and S. Thiyagarajan, "Hybrid Transformation Technique for Image Compression", *Journal of Theoretical and Applied Information Technology*, Vol.41 No.2, July 2012.

[13] O. Buisson, S. Boukir, and B. Besserer, "Motion compensated film restoration", *Machine Vision and Applications*, vol. 13, no.4, Feb. 2003, pp. 201–212.

[14] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: a Paradigm Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of the ACM*, vol. 24(6), 1981, pp. 381–395.

[15] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum. "Full-frame Video Stabilization with Motion Inpainting", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 7, July 2006, pp. 1150-1163,.

[16] H. Chang, S. Lai, and K. Lu, "A Robust Real-time Video Stabilization Algorithm", *Journal of Visual Communication & Image Representation*, vol. 17, 2006, pp. 659–673.

[17] L. Marcenaro, G. Vernazza, and C. S. Regazzoni, "Image Stabilization Algorithms for Video-surveillance Applications", *Proceedings 2001 International Conference on Image Processing*, vol. 1, 2001, pp. 349–352.