Performance of Wavelet-based Multiresolution Motion Estimation for Inbetweening in Old Animated Films

Dwi Ratna Sulistyaningrum, Mochamad Hariadi & Mauridhi Hery Purnomo

Department of Electrical Engineering, Institut Teknologi Sepuluh November,
Campus ITS, Sukolilo, Surabaya, 60111, Indonesia
Email: dwratna@gmail.com

Abstract. This paper investigates the performance of wavelet-based multiresolution motion estimation (MRME) for inbetweening in old animated films using three different MRME schemes. The three schemes are: coarse-to-fine with a wavelet-based MRME, one of Zhang's MRMEs, and an MRME in the spatial domain. In order to make a performance comparison of these MRME schemes, two video sequences were used for a simulation. The experimental results show that the coarse-to-fine method performed better than Zhang's MRME and the MRME in the spatial domain. The evaluation results on block size 9x9 indicate that the coarse-to-fine method had an average peak signal-to-noise ratio (PSNR) of 23.48 dB for the first sequence and 29.84 for the second sequence.

Keywords: *coarse-to-fine; inbetweening; multiresolution motion estimation; wavelet.*

1   Introduction

In order for a video to have the high visual and audio quality that modern audiences desire, the restoration of old animated films is required since they have some kind of damage, such as intensity flicker, noise and blotches [1], [2]. Old animated films were created by hand drawing. An animator makes a number of key frames and an assistant will then build the intermediate frames. As a result, the motion of the object appears jerky, which interferes with human vision. Inbetweening aims to smooth the object motion.

Inbetweening is a process of building intermediate frames between two images in order to show a smooth movement between the first and the second image. The intermediate frame is called 'in-between'. Inbetweening is also a key process in all other types of animation, including computer animation.

Many researchers have done research on inbetweening, but it remains an interesting study subject since the optimal solution for many problems has not been obtained yet. De Juan and Bodenheim [3] use gradient-based motion estimation on traditional animations. This method can process color animation videos, but it requires high computing power and in practice it cannot handle

displacements greater than 5 pixels [1]. Seah and Lu [4] use a modified hierarchical feature-based matching method for motion estimation to build in-between line drawings from a pair of input line drawings. This method uses features such as intensity, edge, angle, displacement orientation, and smoothness magnitude.

Most of the inbetweening research done is for 2-dimensional animation that is not old and in which the systems requires a lot of human intervention. The process of inbetweening in this study was imposed only on gray-level animated films that had some damage.

Inbetweening performance depends on the results of motion estimation. Motion estimation is a process to determine frame displacement in a sequence of pictures. There are several motion estimation algorithms, such as the block matching algorithm (BMA), gradient algorithm and phase correlation algorithm. The BMA is the most popular as it is easy to implement in software and hardware. According to research [5], [6], the block matching algorithm with a multiresolution approach can produce smoother motion vectors. In video coding and compression, wavelet-based motion estimation [7]-[9] has received much attention due to its superior performance compared to conventional motion estimation in the spatial domain. Zhang [8] proposes a matching block with a multiresolution approach on video coding based on wavelets. His method uses the wavelet transform to decompose a video frame into a set of sub-frames with different resolutions.

This paper investigates the performance of wavelet-based multiresolution motion estimation for inbetweening in old animated films based on three algorithms and three types of block sizes for matching.

2   Wavelet-Based Multiresolution Motion Estimation (MRME)

2.1   Concept of MRME

Multiresolution motion estimation (MRME) is a special case of the hierarchical block motion estimation (HBME) approach, using variable block sizes. It can reduce computational complexity because MRME estimates motion vectors hierarchically.

MRME in the wavelet domain is similar to MRME in the spatial domain. In the spatial domain, a video frame is decomposed into a number of frames with a different resolution using the Gaussian pyramid or the Laplacian pyramid. In the wavelet domain, a video frame is decomposed into a number of frames with a different resolution as well as with different spatial orientations.

In a wavelet-based MRME technique [7],[8], the motion estimation process is carried out in the wavelet domain. It starts with estimating the motion vectors at the lowest resolution level, where most of the image energy resides. The motion vectors thus obtained are used as a prediction for higher resolutions and as an addition to a smoothing factor. Motion vector at each block can be obtained by a block-matching algorithm.

The 2D discrete wavelet transform (DWT2) of image $f(x, y)$ with M-level can be expressed as a sequence of sub-images [8] as follows:

$$\left\{ S_{2^K} \left[W_{2^1}^j\right]_{j=V,H,D}, \cdots, \left[W_{2^K}^j\right]_{j=V,H,D} \right\} \tag{1}$$

where $\left\{ S_{2^K}, K=0,1,2,\dots,M \right\}$ shows a set of approximations of $f(x, y)$ with a resolution of $\left\{1, 2^{-1}, 2^{-2}, \cdots, 2^{-M}\right\}$. S_0 is the original image, while S_2 is the approximation of $f(x, y)$ at the next resolution 2^{-1}. $W_{2^M}^j$ is detail image in resolution 2^{-M} at location j, where V, H and D indicate vertical, horizontal, and diagonal, respectively.

Zhang and Zafar [8] have proposed four MRME schemes. These schemes are classified based on their different approach of reference motion vector estimation in the lowest resolution and motion vector refinement in high and mid-frequency sub-bands. Wavelet-based coarse-to-fine schemes [10] and (S_8 + *refine*) Zhang's scheme will be the subject of further investigation in this study.

2.2   Wavelet-Based Coarse-to-Fine Scheme

In a coarse-to-fine scheme [10], the motion vector is estimated only at the low-pass band. Motion vectors at a given resolution level can be predicted from the motion vectors at a lower resolution, by multiplying them by two. A smoothness factor is added after the motion estimation process is done around that prediction. Let $V_k(x, y)$ represent the motion vectors centered at (x, y) for the sub-image at level k. Then the estimation of $V_k(x, y)$ is given by

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

$$\text{for } k = 0, 1, 2, \dots, M$$

where M is the level of decomposition and δ_k is a refinement factor.

In this study, the original frame was decomposed into two levels using the 2D discrete wavelet transform (DWT2). The lowest resolution only had 1/4 pixels of the original image, but it contained a large percentage of the total energy. In each low-pass band, block matching estimation was performed within twenty four neighboring pixels. Figure 1 shows the structure of the wavelet-based coarse-to-fine algorithm.
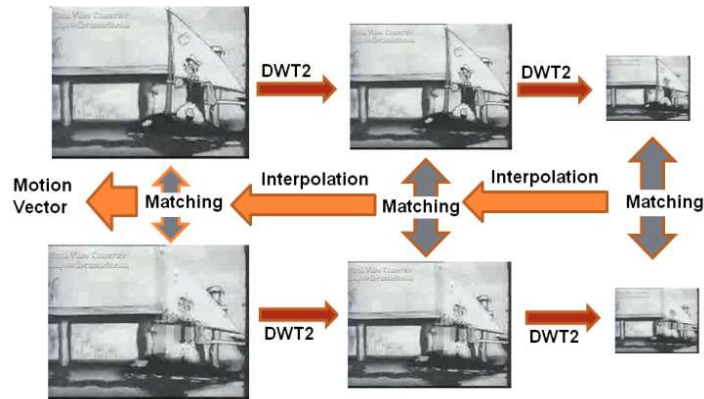

Figure 1   The structure of the wavelet based coarse-to-fine algorithm.

2.3   ($S_8 + $ *refine*) Zhang's Scheme

Zhang [8] has proposed several techniques for motion estimation, one of which, the scheme-III ($S_8 + $ *refine*) technique, provides superior performance over the others. A total of seven sub-bands were obtained with a two-level wavelet analysis. The motion relationships among the resolution levels are shown in Figure 2.

The motion vector in a sub-band at a given resolution level, which is predicted from a motion vector at the lowest resolution level, is multiplied by an appropriate power of two. A refinement factor is added to this prediction after carrying out a motion search around this motion prediction. The equation of the motion vector prediction can be written as

$$V_k^j(x, y) = 2\, V_{k-1}^j(x, y) + \delta_k^j(x, y) \tag{3}$$

$$\text{for } k = 1, 2, \ldots, M \quad \text{and } j = V, H, D$$

where V_k^j is the motion vector at sub-band level k with orientation j and δ_k^j is the refinement factor determined by the result of the motion estimation around the predicted value $2\ V_{k-1}^j$.
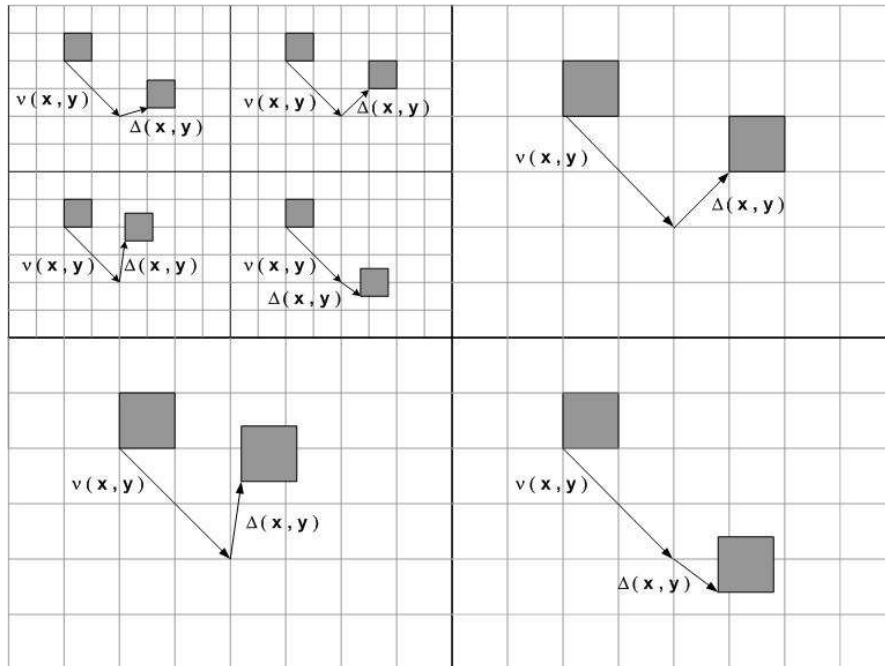

Figure 2 Motion relationship in a two-level variable-block-size MRME scheme.

3   Wavelet-based Inbetweening in Old Animated Films

Figure 3 shows a block diagram of the inbetweening system for old animated films proposed by Sulistyaningrum [10]. It consists of several processing units. First, correcting intensity flicker; second, estimating motion vectors; third, smoothing motion vectors resulted from step 2; fourth, generating intermediate frames between two frames.

Old animated films produced between 1930-1940 were the object of this experiment. Such old films usually have some kind of damage, such as intensity flicker, blotches, noise and line jitter. Intensity flicker is the unnatural temporal fluctuation in perceived image intensity that does not originate from the original frame. In this work, the handled damage type was only intensity flicker. A histogram matching method [11] was applied to correct flicker. This method is based on histogram equalization by applying a uniform histogram.
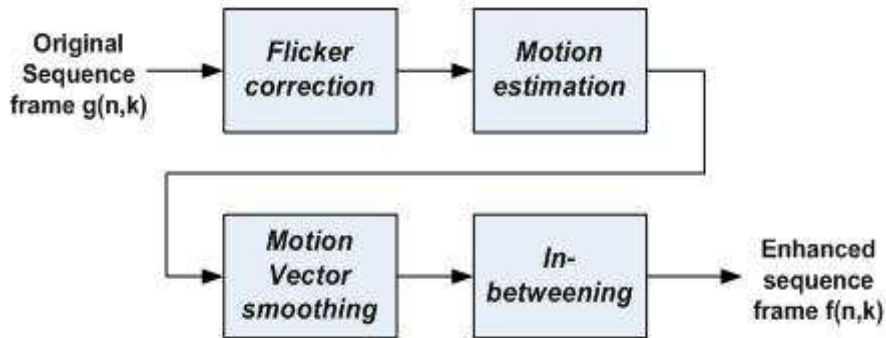
Figure 3 Block diagram of the inbetweening system for old animated films.

Two algorithms of wavelet-based MRMEs were investigated for their performance. In addition, the original MRME algorithm was also investigated. To begin the process of motion vector estimation, the original image frames were decomposed into two layers to build the pyramids. The Laplacian Pyramid approach was applied to the original MRME, while DWT2 was applied to the wavelet-based MRME. In the wavelet-based coarse-to-fine algorithm, motion vectors were estimated only at the low-pass band. In the Zhang algorithm [8], motion vectors were estimated at all levels of each sub-band. A detailed explanation of both methods has been given in Section 2.3.

The motion-vector smoothing algorithm consists of two steps: detecting and smoothing the outlier. An average motion vector surrounding block is used for motion vector smoothing in [12]. The average motion vector is defined as follows:

$$v_m = \tfrac{1}{9} \sum_{i=1}^{9} v_i$$

$$D_c = abs(v_m - v_1) \tag{4}$$

$$D_n = \tfrac{1}{8} \sum_{i=2}^{9} abs(v_m - v_i)$$

Where v_m denotes the mean value of v_1 and all neighbors. After computing D_c and D_n, then v_1 is outlier, if $D_c > D_n$.

Vector median filtering (VMF) is applied for smoothing outlier in [13]. VMF is used to smooth the motion vector field by using the median of motion vector

input adjacent to the vector. If $d_i = \sum_{j=1}^{n} \left\| v_j - v_i \right\|$, the vector median is defined as the one that has the smallest distance from all vectors in V, i.e. $med(V) = v_k \in V$ so that $d_k = \min\{d_i\}$.

Inbetweening can be done by interpolation. Interpolation is a key process for each type of animation, including computer animation. The object interpolation approach was applied to inbetweening in this study. Therefore, object segmentation had to be done first before inbetweening can take place. The process of object segmentation uses a threshold method while the interpolation uses cubic interpolation.

4   Experiment and Results

Our simulation used two video sequences to investigate the wavelet-based MRME. It used a Popeye sequence with a dimension of 240 x 320 pixels and a Felix the Cat sequence with a dimension of 240 x 320 pixels. The mean absolute difference (MAD) was used for the block-matching algorithm of the wavelet-based MRME. Assuming that a motion block has size N x M, the MAD is defined as:

$$MAD(dx, dy) = \frac{1}{NM} \sum_{m=1}^{M} \sum_{n=1}^{N} \left| I_k(m,n) - I_{k-1}(m+dx, n+dy) \right| \qquad (5)$$

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

$$(MV_x, MV_y) = \min_{(dx, dy) \in R^2} MAD(dx.dy) \qquad (6)$$

This experiment used 101 consecutive frames for each test sequence. 50 even frames were removed and 50 new even frames were generated from 51 odd frames using the proposed and existing algorithms. The selected frames had a static background and a dynamic foreground. The PSNR was calculated from the constructed even frames with respect to the original even frames, as follows:

$$MSE = \frac{1}{NM} \sum_{x=1}^{N} \sum_{y=1}^{M} (f_{org}(x,y) - f_{cons}(x,y))^2$$

$$PSNR_i = 10 \times \log_{10} \left(\frac{255^2}{MSE} \right)$$

where *MSE* is the mean square error, *N* is the vertical size of frame, *M* is the horizontal size of the frame, $f_{org}(x, y)$ is the pixel value of the original even frame at position (x, y) and $f_{cons}(x, y)$ is the pixel value of the constructed even frame at position (x, y). The average PSNR, denoted by PSNR$_{avg}$, is given as

$$PSNR_{avg} = \frac{1}{K}\sum_{i=1}^{K} PSNR_i \qquad (7)$$

where *PSNR$_i$* is the measured *PSNR* for frame-*i*, and *K* is the total number of constructed even frames.

This experiment was carried out in order to investigate the performance of a wavelet-based multiresolution motion estimation for inbetweening in old animated films for three algorithms and three types of block size. The comparison results are shown in Figure 4, Figure 5 and Figure 6. The figures indicate that the coarse-to-fine scheme provided an improvement of the PSNR for all test sequences over the spatial MRME and Zhang's MRME. The performances of the coarse-to-fine method and the spatial MRME method provided similar PSNR values for a constant block size of 9x9 and variable block sizes of 5, 9, 17.

Table 1 shows the average PSNR for all test sequences. The test resulted in three MRME schemes and three various block sizes on all sequences. The variable block size was varied (5x5, 9x9, and 17x17) from top to bottom levels. The constant block size was varied (5x5 and 9x9). The table indicates that the coarse-to-fine MRME provided the best PSNR for all block sizes, especially in block size 9x9. In the coarse-to-fine scheme, the motion vectors are estimated only at the low-pass band so that the noise in the high-pass band is not processed. This causes the coarse-to-fine MRME to be more resistant to noise than Zhang's MRME scheme.

The coarse-to-fine MRME and the spatial MRME gave a similar PSNR for the 9x9 block size and the variable block sizes because the wavelet transformation used in the experiments was the Haar wavelet, which generates sub-images similar to those of the Laplacian pyramid. The block size affects the accuracy of the ME. The larger the block size, the more accurate it is; however, computation time is also longer.
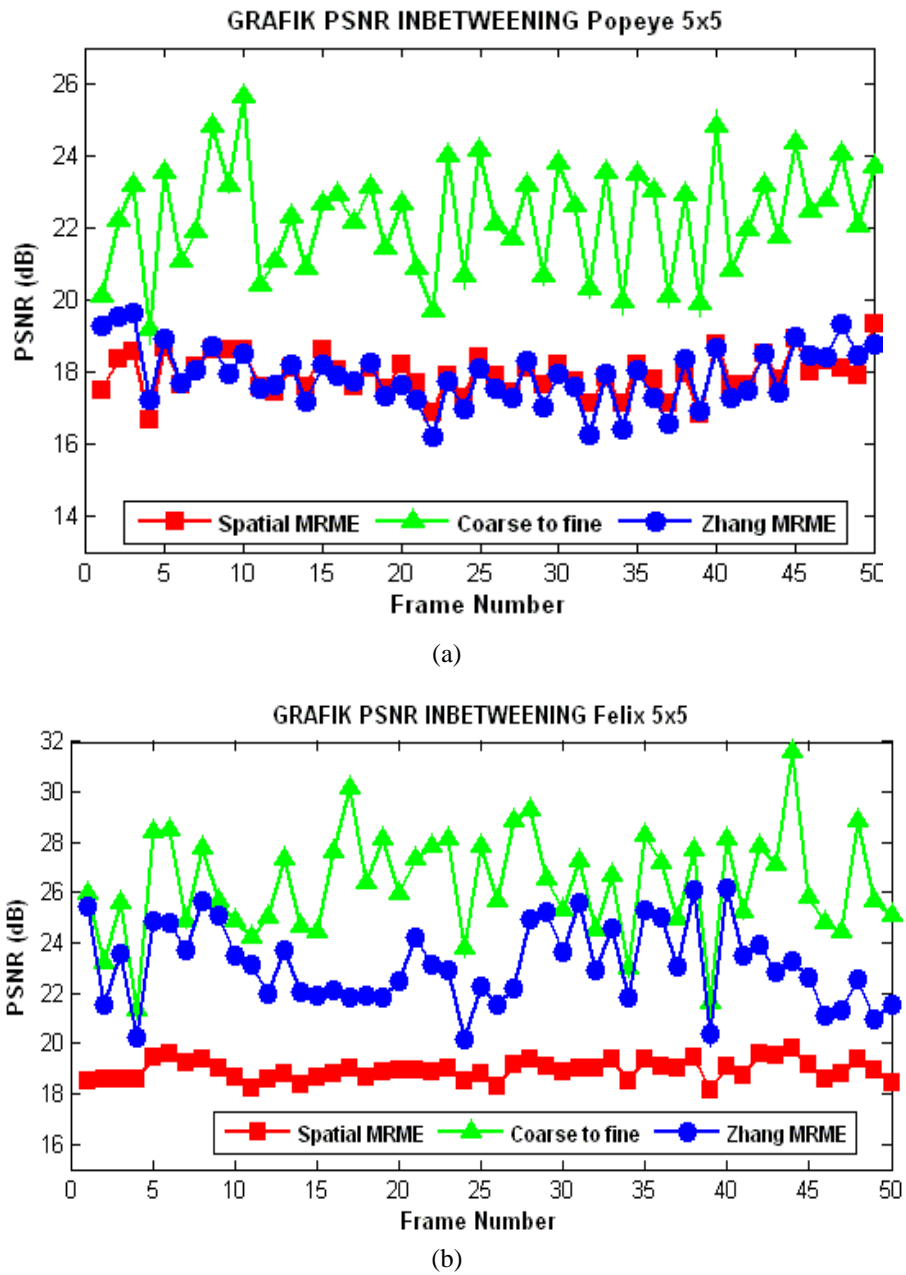
(a)

(b)

Figure 4 PSNRs of the spatial MRME, the coarse-to-fine MRME and Zhang's MRME with block size 5x5 for test sequences: (a) Popeye, (b) Felix.

(a)

(b)

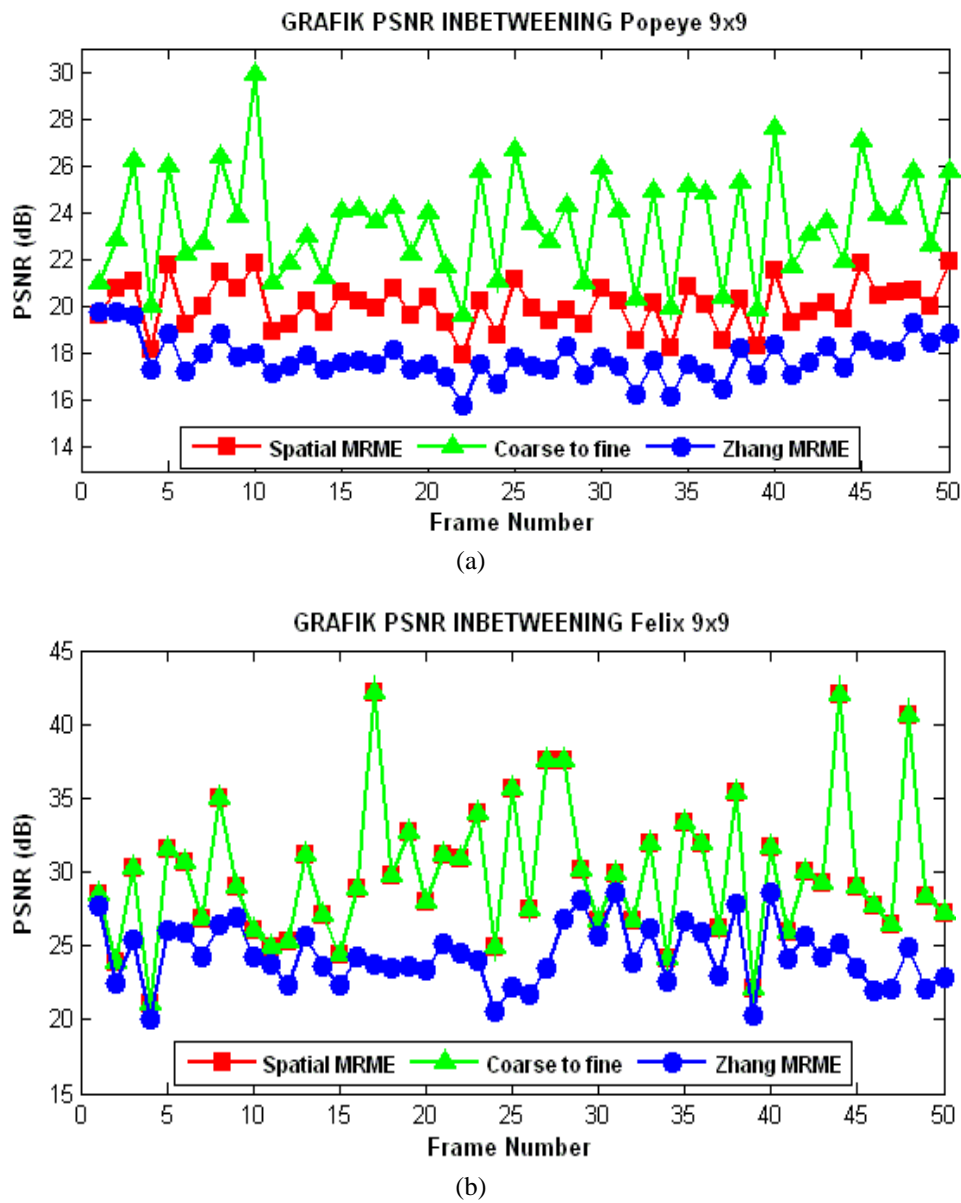**Figure 5** PSNRs of the spatial MRME, the coarse-to-fine MRME and Zhang's MRME with block size 9x9 for test sequences: (a) Popeye, (b) Felix.
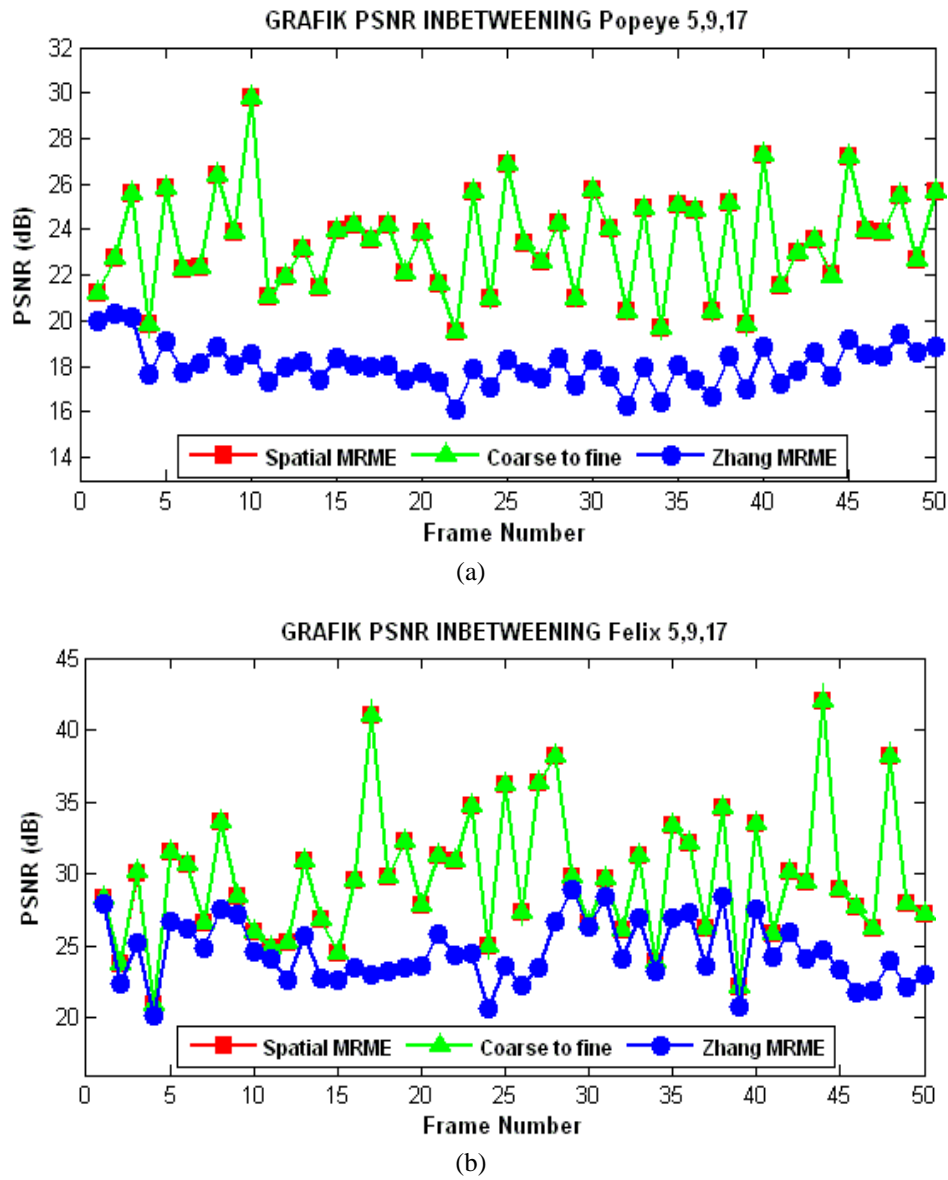
(a)

(b)

Figure 6 PSNRs of the spatial MRME, the coarse-to-fine MRME and Zhang's MRME with variable block sizes 5, 9, 17 for test sequences: (a) Popeye, (b) Felix.

Table 1 Average PSNR of test video sequences.

Sequences	Schema	Constant Block		Variable Block
		5x5	9x9	5x5,9x9,17x17
Blow Me Down Popeye (1933)	MRME spatial domain	17.9244	20.03717	23.4286
	Coarse-to-fine	22.2571	23.4805	23.4286
	Zhang's MRME	17.8945	17.7746	18.0369
Felix the Cat in Hollywood (1923)	MRME spatial domain	18.9243	29.8436	29.6829
	Coarse-to-fine	26.3320	29.8445	29.6842
	Zhang's MRME	23.1284	24.3541	24.4985

5   Conclusion

In this paper a comparative study was discussed of the performance of a wavelet-based MRME for inbetweening in old animated films. Our findings show that the coarse-to-fine method was one of the best methods for inbetweening in old animated films. The evaluation results on block size 9x9 indicate that the coarse-to-fine method had an average PSNR of 23.48 dB for the Popeye sequence and 29.84 dB for the Felix sequence.

References

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

[2]   Van Roosmalen, P.M.B., *Restoration of Archived Film and Video,* Ph.D. Thesis, TU Delft, Delft, 1999.

[3]   De Juan, C. & Bodenheimer, B., *Re-using Traditional Animation: Methods for Semi-Automatic Segmentation and Inbetweening*, Eurographics/ACM SIGGRAPH Symposium on Computer Animation, Boulic R., Pai D., (Eds.), 2006.

[4]   Seah, H.S. & Lu, J.*, Computer-Assisted Inbetweening of Line Drawings: Image Matching*, In 7th International Conference on Computer Aided Design and Computer Graphics, CAD & CG, Kunming, China, pp. 193-200, August 2001.

[5]   Kamil, M.U., Veterli, M. & LeGall, D., *A Multiresolution Approach to Motion Estimation and Interpolation with Application to Coding of Digital HDTV*, IEEE, 1990.

[6]   Kim, M.J., Lee, Y.G., Jong & Beom R.A., *A Fast Multiresolution Block Matching Algorithm for Multiple-Frame Motion Estimation*. IEICE Ttrans. Inf. & Syst., **E88–D**(12), 2005.

[7]   Usama, S. & Šimák, B., *A Low-complexity Wavelet Based Algorithm for Inter-frame Image Prediction*, Acta Polytechnica, **42**, 2002.

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

[9]   Kim, J.T., Yang, C.M., Lim, D.K. & Ho, Y.S., *Multiresolution Motion Compensation in the Wavelet Domain for Scalable Video Coding*, Proc. SPIE, Visual Communication and Image Processing 2002, **4671**, 2002.

[10] Sulistyaningrum, D.R., Hariadi, M. & Hery Purnomo, M., *Motion Estimation Using Coarse-to-fine Based on Wavelet Multiresolution for Inbetweening in Old Animation Films*, in Proceedings of International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, 2010.

[11] Naranjo, V. & Albiol, A., *Flicker Reduction in Old Films*, in Proceedings of IEEE International Conference on Image Processing (ICIP '00), **2**, pp. 657–659, Vancouver, Canada, September 2000.

[12] Kang, S.J., Cho, K.R. & Kim, Y.H., *Motion Compensated up-Conversion using Extended Bilateral Motion Estimation*, IEEE Trans. Consumer Elec., **53**(4), pp 1758-1767, Nov. 2007.

[13] Dane, G. & Nguyen, T., *Motion Vector Processing for Frame Rate Up Conversion*, in Proc. 2004 IEEE International Conferences on Acoustics, Speech, and Signal Processing, **3**, pp.309-312, May 2004.