MULTIPLE OBJECTIVE FRAME RATE UP CONVERSION

Tak-Song Chong, Oscar C. Au, Wing-San Chau, Tai-Wai Chan

Dept. of Electrical and Electronic Engineering
Hong Kong University of Science and Technology
E-mail: eects@ust.hk, eeau@ust.hk, eecws@ust.hk, eedavid@ust.hk

ABSTRACT

In this paper, we propose a multiple objective frame rate up conversion algorithm (MOFRUC), which utilizes two different models. The first model is a constant velocity model that assumes the objects position is a linear function of time. The second model exploits the spatial correlation between neighboring blocks, and assumes the pixel intensity is highly correlated in a small local area. In this model, the perceptual quality of interpolated frame is also taken into account and the blocking artifact is minimized. Our proposed MOFRUC estimates the motion trajectory by the first model and interpolate the frame along the motion trajectory. At the same time, the algorithm refines the motion trajectory by maximizing a spatial correlation measurement defined in the second model and interpolates the frame with minimum blocking artifact. Simulation results show that our proposed MOFRUC outperforms other existing algorithms and produces high quality interpolated frame.

1. INTRODUCTION

Frame rate up conversion (FRUC) is the conversion between different video formats with different frame rates. In particular, it addresses the problem of converting video signal from a lower frame rate to a high frame rate. For example, it can convert a motion picture film operating at 24 frame/sec to higher frame rate for HDTV. In addition to display format conversion, FRUC is also employed in low bit-rate video coding such as video conferencing in which, the video is encoded at low frame rate and the decoder uses FRUC to increase the temporal resolution for better visual quality.

Traditional frame rate up conversion algorithms such as frame repetition and linear interpolation introduce annoying visual artifacts. Frame repetition causes jerky object motions while linear interpolation blurs the object in motion area. One way to improve the visual quality of the interpolated frame is to take the motion trajectory into account and interpolates the frame along the motion trajectory. In [1-3], motion estimation (ME) is performed to estimate the motion trajectory; however, it causes overlapped pixels and holes problem. While in [4], a similar method known as bi-direction motion estimation is employed to estimate the motion trajectory.

Motion trajectory estimation is different from motion estimation used in MPEG-1/2/4. The motion trajectory estimation aims to track object position in every single frame while the motion estimation aims to minimize the entropy of the residual signal of the displaced frame difference in MPEG-1/2/4. As a result, the estimated motion vector in motion estimation may not represent the true object motion. An inaccurate estimation of motion trajectory can cause annoying visual artifact in FRUC, especially blocking artifact. To solve this problem, several refinements in motion vector have been proposed. In [5], the motion vector is pre-processed according to an autoregressive model. In [6], the reliability of motion vector is analyzed and the most reliable motion vector is chosen for interpolating the new frame. In [3], the object motion in a block is modeled by multiple motion trajectories to improve the estimation of motion vector and to minimize blocking artifacts.

In this paper, we propose a multiple objective frame rate up conversion algorithm that employs bi-directional motion estimation to estimate the motion vector and exploits the spatial correlation between neighbouring blocks to improve the accuracy of estimation. The bi-directional motion estimation assumes a block moves with constant velocity, and uses a block dissimilarity measurement similar to the displaced frame difference in motion estimation process to estimate the motion trajectory. In addition, the spatial correlation of the blocks is considered to improve the accuracy in estimating the motion trajectory. The pixel intensity of the candidate blocks is analyzed and the candidate block that shows highest spatial correlation is chosen to interpolate the new frame.

This paper is organized as follows. The proposed algorithm is presented in Section 2. Section 3 shows the simulation results. Finally, concluding remarks are given in Section 4.

0-7803-9332-5/05/$20.00 ©2005 IEEE

2. MULTIPLE OBJECTIVE FRAME RATE UP CONVERSION ALGORITHM (MOFRUC)

Two different models are adopted in our proposed multiple objective frame rate up conversion (MOFRUC) algorithm. The first one is a constant velocity model, which assumes an object moves in constant velocity and its position can be modeled as a linear function of time. The motion trajectory is first estimated by bi-directional motion estimation, and the position of object in the to-be-interpolated frame is predicted by the motion trajectory in linear manner. The second one is a spatial correlation model, which exploits the spatial correlation to interpolate new frames and assumes the pixel intensity is highly correlated in a small local area. In the second model, a side matching distortion measure is introduced to measure the spatial correlation and act as a perceptual quality metric for measuring blocking artifact. Our proposed MOFRUC employs both models and finds the block that can best satisfy the constant velocity model and the spatial correlation model. To employ both models at the same time, the frame rate up conversion problem is formulated as a multiple objective optimization problem. The first objective is to maximize the fitness of the constant velocity model given a block and the second objective is to maximize the spatial correlation of interpolated block while minimizing the blocking artifact.

2.1. Constant Velocity Model

Constant velocity model assumes an object move in constant velocity within a small time interval. Suppose a object in frame t-1 moves to a new position in frame t+1 (figure 1) with a relative displacement of (+2u,+2v). If we assume constant velocity, the relative displacement of the object in frame t with respect to frame t-1 and frame t+1 is (-u,-v) and (+u,+v) respectively. With the estimated motion trajectory between frame t-1 and frame t+1, the position of the object at time t can be estimated.

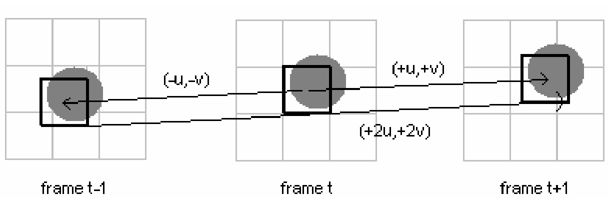

Fig. 1 The motion trajectory over three consecutive frames

Let frame t be the to-be-interpolated frame which is inserted between frame t-1 and frame t+1 and the pixel intensity at time t be f(x,y,t). For each N-by-M block B in frame t, our goal is to find two blocks, one in frame t-1 and one in frame t+1 that pass through B. The dissimilarity measurement is defined as the pixel intensity difference, which is showed as follows:

$$C(u,v) = \sum_{(x,y)\in S(B)} | f(x+u, y+v, t+1) - f(x-u, y-v, t-1) | \quad (1)$$

where S(B) is the spatial location of block B.

A small value of C suggests the two blocks in frame t-1 and frame t+1 respectively are similar and most likely belongs to the same object. Therefore, there is a large chance that the motion trajectory passes through B.

Hence the reconstructed block B is interpolated by simple averaging along the motion trajectory:

$$f(x,y,t) = \frac{f(x+u^*, y+v^*, t+1) + f(x-u^*, y-v^*, t-1)}{2}, (x,y) \in S(B) \quad (2)$$

where

$$(u^*, v^*) = \arg \min_{(u,v)} C(u,v)$$

2.2. Spatial Correlation Model

The spatial correlation model exploits the spatial correlation between neighboring blocks and tries to find the block that has highest correlation. This model is widely used in error concealment algorithm to recovery the lost motion vector during transmission of compressed bitstream in unreliable channel [7]. In this model, the side matching distortion is introduced for measuring spatial correlation.

Let (x,y) be the upper left coordinate of current block B, and $f(x,y,t), \hat{f}(x,y,t)$ be the pixel intensity at location (x,y) at time t of the neighbouring block and the candidate block respectively. The side matching distortion D is defined as the absolute difference between pixel values across the macroblock boundary (above, below, left and right), as follows (figure 2):

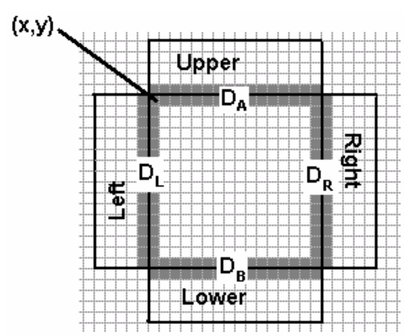

Fig .2 The current block and the neighbouring block.

$$D_A = \sum_{x=0}^{N-1} | \hat{f}(x,y) - f(x,y-1) |$$

$$D_B = \sum_{x=0}^{N-1} | \hat{f}(x,y+M-1) - f(x,y+M) |$$

$$D_L = \sum_{y=0}^{M-1} | \hat{f}(x,y) - f(x-1,y) |$$

$$D_R = \sum_{y=0}^{M-1} | \hat{f}(x+N-1,y) - f(x+N,y) |$$

$$D = D_A + D_B + D_L + D_R \quad (3)$$

A small value of D indicates the boundary pixels of the candidate block match well with the boundary pixels of the neighboring blocks. By calculating the side matching distortion of all the candidate blocks, we can choose the best one that has lowest side matching distortion. It is worthy to note that the side matching distortion D is also a measurement for blocking artifact. A large value of D indicates a strong blocking artifact while a small value of D indicates a smooth transition across block boundary. As a result, the side matching distortion is employed in our frame rate up-conversion algorithm to improve the perceptual quality of the interpolated frames.

2.3. Multiple Objective Minimization

The frame rate up conversion is formulated as a multiple objective minimization problem, which minimized the block dissimilarity measure and the side distortion. The algorithm is described as below:

Step1: Initialization.

The to-be-interpolated frame t is first interpolated by constant velocity model only. For each N-by-M block in frame t, we find the motion vector that minimizes the dissimilarity measurement *C(u,v)* (Eqn. 1), and interpolate the block by Eqn. 2. We denote the interpolated frame t as $f^0(x,y,t)$.

Step 2: Iteration.

Let k>0 be the iteration index. A new cost function is defined as followed:

$$E(u,v) = C(u,v) + \mu D'(u,v)$$

where $C(u,v)$ is the dissimilarity measurement (Eqn. 1) and μ is a constant to control the relative importance of each model and $D'(\vec{v})$ is a modified version of side matching distortion (Eqn. 3) defined as follows:

$$D'(u,v) = D'_A(u,v) + D'_B(u,v) + D'_L(u,v) + D'_R(u,v) - \max(D'_A(u,v), D'_B(u,v), D'_L(u,v), D'_R(u,v))$$

where

$$D'_A(u,v) = \sum_{x=0}^{N-1} \left| \frac{f(x+u,y+v,t+1) + f(x-u,y-v,t-1)}{2} - f^{k-1}(x,y-1) \right|$$

and $D'_B(u,v)$, $D'_L(u,v)$, $D'_R(u,v)$ are defined similarly

Hence the reconstructed block B in k[th] iteration is:

$$f^k(x,y,t) = \frac{f(x+u^*,y+v^*,t+1) + f(x-u^*,y-v^*,t-1)}{2}$$

where

$$(u^*,v^*) = \arg \min_{(u,v)} E(u,v)$$

Step 3: Termination

Repeat step 2 until k=N for some predefined N or the algorithm converges.

3. RESULTS

Five CIF sequences (30 fps), namely foreman, table tennis, flower, mobile and akiyo are tested and different frame rate up conversion are compared with the proposed MOFRUC. Each CIF sequence is first 2:1 sub-sampled in temporal domain into 15fps and then, the interpolated frame are compared with the original CIF sequences.

Table 1 shows the PSNR of the up converted video sequences of different FRUC algorithms, including simple averaging, bi-directional motion estimation, uni-directional motion estimation, motion compensated temporal interpolation (MCTI) in [1] and the proposed MOFRUC. In Table 1, the proposed MOFRUC achieves a PSNR gain ranging from 0.32dB to 7.95 dB over other existing algorithm with an average gain of 1.93dB. This shows that the proposed algorithm provides better objective quality than others in terms of PSNR.

Figure 3 shows a typical interpolated frame (frame 3) of the foreman sequence using various algorithms. In frame 3, foreman rotates his head. As rotational motion is poorly approximated by translational motion, most of the algorithms fail and the interpolated frame tends to have serious blocking artifacts due to incorrect motion trajectory estimation. In contrast, the proposed MOFRUC gives perceptually much better interpolated frames with little blocking artifacts. It is because the proposed MOFRUC employs the spatial correlation model which ensures the reconstructed block match well with the neighbouring block and produces interpolated frame with smooth transition across the block boundary.

3. CONCLUSION

In this paper, a multiple objective frame rate up conversion algorithm is proposed. It employs a constant velocity model and a spatial correlation model. By utilizing two different models through multiple objective optimization technique, the motion trajectory can be estimated much more accurately with minimum visual artifacts and thus producing high quality interpolated frame.

7. ACKNOWLEDGEMENT

This work has been supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China (project no. HKUST6203/02E), the Innovative Technology Commission (project no. ITS/122/03) and HKUST (project no. DAG04/05.EG34).

8. REFERENCES

[1] C. K. Wong and O. C. Au, "Fast motion compensated temporal interpolation for video", *Proc. SPIE: Visual Communications and Image Processing*, pp.1108-1118, May 1995.

[2] Chi-Wah Tang and O. C. Au, "Unidirectional motion compensated temporal interpolation", *Proc. Of IEEE Int. Sym. on Circuits and Systems*, Vol. 2, pp. 1444 – 1447, Jun 1997.

[3] Sung-Hee Lee, Ohjae Kwon and Rae-Hong Park, "Weighted-adaptive motion-compensated frame rate up-conversion", *IEEE Transactions on Consumer Electronics*, Vol. 49, No. 3, pp. 485-492, Aug. 2003.

[4] Tao Chen, "Adaptive temporal interpolation using bidirectional motion estimation and compensation", *Proc. Of IEEE Int. Conf. on Image Processing*, Vol. 2, pp. 313-316, Sep. 2002.

[5] G. Dane and T.Q Nguyen, "Motion vector processing for frame rate up conversion", *Proc. Of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 3, pp 309-312, May 2004.

[6] H. Sasai, S. Kondo and S. Kadono, "Frame-rate up-conversion using reliable analysis of transmitted motion", *Proc. Of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 257-260, May 2004.

[7] W. M. Lam, A.R. Reibman and B. Liu, "Recovery of lost or erroneously received motion vectors", *Proc. Of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 417-420, Apr 1993.

	Frame averaging	Bidirectional ME	Uni-directional ME	MCTI in [1]	Proposed MOFRUC
Foreman	29.22	31.79	29.43	30.62	**32.62**
Table	30.93	31.61	30.84	31.44	**32.17**
Flower	22.45	28.95	28.4	28.77	**30.4**
Mobile	24.74	25.99	26.28	27.05	**28.41**
Akiyo	45.13	46.52	46.3	46.46	**46.78**

Table 1: The PSNR of different frame rate up conversion algorithms.

Fig. 3a. Frame 3 of original "foreman".

Fig. 3c. Interpolated frame 3 of "foreman" using bidirectional ME.

Fig. 3e. Interpolated frame 3 of "foreman" using MCTI in [1].

Fig. 3b. Interpolated frame 3 of "foreman" using frame averaging.

Fig. 3d. Interpolated frame 3 of "foreman" using uni-directional ME.

Fig. 3f. Interpolated frame 3 of "foreman" using proposed MOFRUC.