MATLAB Implementation of New True-Motion Estimation technique and its Application to Motion-Compensated Temporal Frame rate up conversion

Madhukumar R K[1] Sarika Tale[2]
[1]M.Tech
[1]Department of Digital Electronics and Communication Systems
[1]VTU Centre for Post Graduation Studies, Bengaluru

Abstract— In multimedia communication video plays an importance, because it gives the very perfect flow of frames or image for our visual effect with good realistic view experience. So in this paper gives a new low complexity true motion estimation technique is introduced for video processing application, like motion-compensated temporal frame rate up conversion (MCTFRUC) or motion compensated temporal frame interpolation (MCTFI).In basically main role of motion estimation is to produce the motion vectors for reduce the temporal redundancy. Where as in this paper explains true motion estimation which is mainly used to track the motioned object as closely as possible by imposing the explicit and/or implicit smoothness constraints on block matching algorithm. To get the good interpolated frames the dense motion vector field is obtained for both forward and backward motion vectors is applied by adding both. Then the performance parameter is calculated and is produced, at last the good quality of video frame is produced with smoothly.

Key words: frame rate up conversion, frame interpolation, motion estimation, true motion, clustering, video processing, and motion compensated temporal frame interpolation (MCTFI), structural similarity(SSIM),peak signal to noise ratio(PSNR).

I. INTRODUCTION

Motion estimation has plays a very important role in video coding and several video processing application like such as denoising, de interlacing and frame rate up conversion/ frame interpolation. Motion estimation mainly employed to produce the temporal correlation between the video frames either to reduce the Temporal redundancy for video coding or to enhance the visual video quality for video processing.

In video processing some of the arguments are there like video processing may potentially uses the existing motion blocks from the decoder via motion vector post processing to keep complexity low however this may not usually be a feasible option this infeasible could be due to either difficulty of using motion vector or lack of available motion vector. For example, to increase the frame rate MV post processing may be preferred for resource-limited handheld devices, whereas ME algorithm can be employed for high visual video quality demanding applications, such as broadcasting, television and multimedia players[1],[2]. Besides, human visual system (HVS) tolerates artifacts better in small displays compared with large displays due to the increased angular resolution; hence, handheld devices can tolerate lower complexity inferior methods compared with large displays. It is not easy to accurately estimate motion since ME is an ill-posed problem [3,4]. Furthermore, video coding, they are usually termed as true-motion estimation. True motion estimation technique used to emphasize the fact that their objective is to track the projected object motion rather than to reduce the temporal redundancy. Projected object motion usually results in a coherent motion vector field (MVF) except on the motion boundaries; however, regular ME algorithms may not necessarily produce a coherent MVF since their objective is to minimize the prediction error or number of bits required to code the prediction error. Hence, regular ME algorithms are susceptible to give wrong motion trajectories. To obtain a coherent MVF, TME technique further impose smoothness constraints on MVs by using their spatiotemporally neighboring blocks

So to utilize the motion blocks in video processing in decoder side I used the technique called as clustering of motion blocks which gives the implicit smoothness on motion blocks and also by using this clustered motion blocks it is helpful to produce the dense motion field then after the use to produce the good temporal frame rate up conversion. To know the quality of interpolated frames need to find out the PSNR and SSIM for interpolated frames and original frames. And then lastly the good video frames is obtained and produced.to distinguish ME schemes used in video processing applications from regular ME techniques used in

II. METHODOLOGY OF THE PROJECT

A. BLOCK MATCHING ALGORITHM:

The purpose of a block matching algorithm is to find a matching block from a frame *i* in some other frame *j*, which may appear before or after *i*. This can be used to discover temporal redundancy in the video sequence, increasing the effectiveness of inter frame video compression and television standards conversion. Block matching algorithms make use of an evaluation metric to determine whether a given block in frame *j* matches the search block in frame *i*.

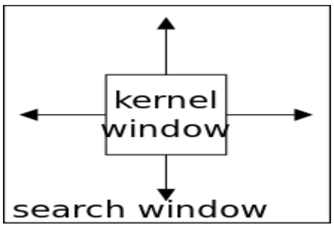

Fig. 1: Block matching concept

Block matching algorithm is employed for more efficient and hardware friendly implementation. Some researchers employed some other method also like gradient-based, object based and transform domain approach. BMA is most widespread algorithm used in video coding and video processing applications partly due to its straightforward, simpler more efficient and hardware friendly implementation. BMAs target to minimize an objective function which is also called as block distortion measure(BDM) is in form

$$E_D(d)= \sum_{n\in B} \rho(I_{k1}[n+d],I_{k2}[n]) \qquad (1)$$

Where I_{k1} and I_{k2} are images at times k1 and I_{k2} d is the displacement of n or MV at n, B is the set of all pixels in I_{k2} ρ is the matching criterion. Usually $\rho(x,y) = |x-y|^p$ is used by setting $p = 1$ for implementation simplicity, which is usually referred to as sum of absolute differences or errors (SAD or SAE). To achieve smoother MVFs, several different methods have been attempted in the literature. Imposing smoothness constraint can be done either explicitly or implicitly, or both. Approaches impose implicit smoothness constraint through the use of multi-resolution or predictive search for BMA. Enforcing explicit smoothness constraint is generally achieved by one of the following three ways: 1) adding a penalty term to (1), 2) using a different matching criterion, $\rho(\cdot,\cdot)$, and 3) post-processing the MVF obtained from (1).

B. MOTION ESTIMATION:

This is the very important one part of my project to find out the moving object motion in the video sequences. In case of motion estimation the first of all we need to collect the all frames in the video. After collecting all the frames it is necessary to convert the all frames which is in the coloured format into the gray scale images.

First the reference frames is to be taken from first frames of video sequences then the motion is estimated by using all other frames by using the following process is takesVideo is collection of the frames so first I need to collect the all frames from video After collecting the all frames .then going to convert the coloured image into the gray scale image The first frames is taken as the reference frame and all other frames are considered as the anchor frames. In next convert the both reference and anchor frames into the 4*4 block size Then the converted 4*4 block is now need to perform the Euclidian norm which gives the difference value which is matched the block or not.The correlated output which is the value of the block after the Euclidian norm operation. The correlated value is more than that of threshold valued then that block has some motion .then that block is need to show as whitened and this called motion block then the remaining block is produced as like only. Totally the above process is going to repeat for all blocks and all other frames Then after the find out the motion block collect the frames and produce the video which is gives the where the motion is happening in the sequence wise.

C. .IMPOSSING SMOOTH.NESS.:

Using a modified matching function is usually motivated by intuition, whereas adding a penalty term can be theoretically justified by using a Bayesian MAP estimator, which imposes certain prior distribution on the model parameters. The a posteriori probability distribution of the motion field.

$$P (Dk = d_k|I_k = I_k;I_{k-1}) \qquad (2)$$

Equation (2) is used to obtain the MAP estimate by rewriting its Bayes equivalent as

$$P (I_k = I_k|D_k = d_k;I_{k-1}). P (Dk = d_k; I_{k-1}) \qquad (3)$$

Where D_k is a vector random field, d_k is one of its realization I_k is a scalar random field, and I_k is one of its realization. Then the MAP estimate of d_k is computed as follows:

$$\widehat{dk} = \arg\min_d(P (I_k = I_k|D_k = d_k;I_{k-1}). P (Dk = d_k; I_{k-1}) . P (Dk = d_k; I_{k-1}) \qquad (4)$$

where the first term is related to the observation model measuring how well d_k models the change, and the second term serves as a motion model explaining the prior information contribution of the random field D_k , such as its smoothness.

To solve (4) it is assumed that the displaced frame difference (DFD) is a zero-mean Gaussian distribution; hence, the first term can be written as a product of zero-mean Gaussians. In addition, it is assumed that D_k is a MRF; so, the second term is a Gibbs distribution specified by cliques and a potential function. Using these assumptions and two-element cliques for MRF, (3) can be recast as

$$\widehat{dk} = \arg\min_d(\sum_n ||(I_k[n]-,I_{k-1}[n])||^2 + \lambda \sum_{l\in Nn} ||(d[n]-d[l])||^2 \qquad (5)$$

where Nn is a set of neighbors of n (e.g, 4- or 8-neighbors), and ||•|| denotes Euclidian norm. λ is equal to $2\sigma^2$, where σ^2 is the variance of the DFD; λ controls the amount of smoothness. The first term measure's how well each MV candidate d matches by comparing the intensity variation, and the second term penalizes the deviation from its neighbors. For minimization of this function different approaches are proposed, such as simulated annealing, iterated conditional modes, highest confidence first; however, they are very complex, and incompatible for real-time applications. In addition, even though this minimization gives rise to smooth MVF, it is inconvenient since the MVF is smooth even at the object boundaries.

D. D.CLUSTERING OF MOTION BLOCKS:

In this step the grouping of motion blocks which is similar in nature, which is also gives the perfect motioned object.

To cluster the MVs blocks in the spatiotemporal neighborhood, it is desired to have a clustering algorithm that would adaptively give different number of clusters based on the maximum cluster distance specified at input, along with the additional constraint that this will take a fixed number of iterations.

The algorithm takes the sample set X, maximum number of clusters K, and the maximum cluster distance D as input; after execution it outputs the cluster means m and cluster labels c. Since the sample set X is formed by zero MV and MVs of the spatiotemporal neighbors, zero MV is used in the initialization. Mean of the first cluster is set to zero and others to undefined. Similarly, label of the first sample is set to zero and others to undefined. Initially, number of formed clusters is set to one; depending on the displacement of MVs, the number will adaptively increase until K. Elements in the set, starting from second till last, are tested for the closest cluster and then the maximum distance to the elements of this cluster are calculated; if they are in the D proximity of a cluster, then they become a member of this cluster[7], otherwise they start a new cluster. Finally, cluster mean of the assigned cluster is updated and continued with the next sample in the set. clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).This gives the motion block grouping and also it specifies the true motion of blocks with their indices. Which is also obtain the true motion estimation.consider the all the motioned blocks which is comes after the motion estimation let takes the first block and then find out the Euclidian norm to next block. If the blocks matches means take there index and store with the 1st block repeat the above step until the last block in that frame then after repeating we store the indices min value along with the matched blocks then take the next block and continue above process for un clustered blocks And also there matched block indices .Whole process repeat for all other blocks which are not matched before the steps above. This gives the clustered motion blocks for the one frame .At last repeat entire steps to all other frames and stores the matched data and indices of different motion blocks.

E. MOTION COMPENSATION TEMPORAL FRAME RATE UPCONVERSION:

This is the method of producing the extra frames in between the existing frames for the purpose of better video quality by increasing the frames rate depends on the motion.

To enable better quality multiple interpolation frames between two existing frames UME is employed instead of BME; to better handle occlusion regions effectively for better quality both forward and backward TME are used. Occluded areas exist in one of the two successive frames; depending on these areas being covering or uncovering corresponding former or latter frame has the necessary information, and the interpolated frame should be constructed accordingly. In addition, to prevent blocking artifacts in the interpolated frame, dense motion field at the interpolation instant will be obtained and used for temporal interpolation.

1) Obtaining Dense Motion Field:

Applying backward and forward TME between frames Ik and I_{k+1} will result in MVFs $\mathbf{d}^{b}_{k+1}$ and $\mathbf{d}^{f}_{k}$, respectively, where superscript denotes the direction. Although temporally shifting them to obtain the MVFs at k+α is more straightforward compared with projecting them, it gives inaccurate MVs especially for large values. Hence, these MVFs will be projected to k + α for better interpolated image quality. Although the use of UME results in a contiguous MVF at its associated anchor frame, the projected MVF at interpolation instant may not be contiguous. In fact, it will have overlap and hole regions due to differing MV values of neighboring blocks as shown in fig. 2

2) Interpolation:

Interpolation means that adding extra things in between the existing one. In the video processing the interpolation means the adding the extra frames in between the existing frames, for the purpose of smoothening the motion object and to get the good video quality.

Obtained dense motion fields at the interpolation instant are used to generate the interpolated frame. Existence of occlusion areas between neighboring frames complicates the interpolation process since they may exist in either one of them. Occlusion areas are either covering or uncovering depending on the former or latter frame has the necessary information. An uncovering area at frame k + 1 does not have its correspondence at frame k; hence, the backward MVs in this area may not be accurate. To accurately find the motion of this occlusion area frame k + 2 has to be used as it would have the corresponding area. Similarly, for an covering area at frame k, frame k − 1 has to be utilized to find the corresponding MV of this covering area. Therefore, to effectively handle occlusion regions between two frames, one additional frame in each direction is required, resulting in total four frames. This, however, increases the complexity and storage requirements of the solution. To enable a low complexity solution, we employ only two neighboring frames and gracefully obtain the interpolated images. To handle the occlusion areas using two frames, some methods employ occlusion detection mechanisms and use their detection results to interpolate the occlusion area by selecting the pixels from the corresponding neighboring frame depending on the area is covering or uncovering. However, such hard switching methods at pixel or block level may produce strong local temporal artifacts that are very annoying during the playback. It's usually preferred to have a global degeneration than a strong local distortion .As a result, occlusion areas are dealt with implicitly using both forward and backward MVFs. In the proposed method hard switching is avoided by employing soft decision for occlusion areas. Both forward and backward MVFs are used to generate the interpolated frame by mixing the forward and backward interpolated frames. take the second frame and find out the motion blocks by using motion estimation method and which is formed clusters of the motion blocks also. take the third frame and also find out the motion blocks by using the motion estimation method and which is formed clusters of the motion blocks also. With there corresponding indices and motion blocks. creating an empty matrix to insert to create new frame. inserting the plane frame along with the blocks which is have not motion. find out the mean with averaging both the value of the previous and current motion blocks then insert it into the corresponding location. that frame is store in between the current and previous frame. this gives interpolation of frame. then repeat above process for taking all other frames. Finally we get the interpolated video sequences.

This results in increasing the frames in video for producing the better quality video by adopting the true motion estimation. Then lastly the performance parameter is obtained. The distortion introduced into image/video can be classified into three categories in regard to the human visual perception as sub-, near-, and supra-threshold distortions [5], [6]. Sub threshold distortion is below just-noticeable difference (JND) and usually unperceivable by the HVS; near-threshold distortion is slightly above JND; and, supra-threshold distortion generally appears in a structured form, such as blockiness, ghosting, and blurring. While PSNR has good performance on measuring near-threshold distortion, SSIM has good performance on measuring both near- and supra-threshold artifacts [6]. Since supra-threshold dominates the human visual Perception. The Structural Similarity (SSIM) index is a method for measuring the similarity between two images. The SSIM index can be viewed as a quality measure of one of the images being compared provided the other image is regarded as of perfect quality this will be compared with the interpolated frames and original frames.

III. BLOCK DIAGRAM OF THE PROJECT

This section explains the basic block diagram for my project First take the video and which is contains the number of frames. In the those frames is there any moving object is there then we have to find out the motion estimation. after finding out motion estimation we have to make smoothening on motion blocks by explicitly and implicitly then the clustered output of motion blocks which is gives the true motion. then using these motion blocks the extra frames are introduced in between the existing one by motion compensated temporal frame interpolation at last the performance parameter is to be calculated to know the quality of videos .

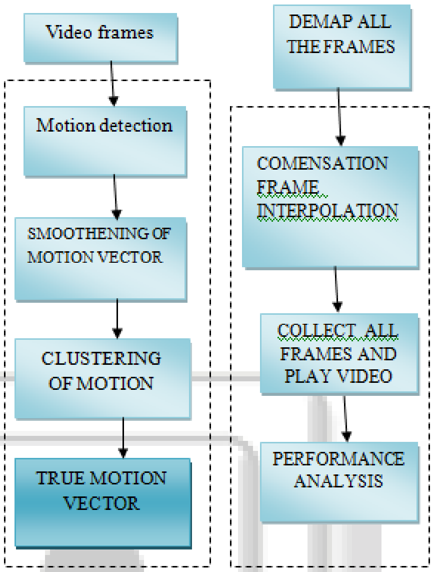

Fig.2 block diagram of true motion estimation and compensated temporal frame interpolation

i *video* :

Video is the continuous moving frames ,collect all the frames from video. For the calculations of the motion detection

A. *motion detection:*

Motion detection is the process of detecting a change in position of an object relative to its surroundings or the change in the surroundings relative to an object. with the help of the eight neighbors search method and produces the motion vectors .

B. *Smoothening of motion vector:*

In statistics and image processing, to smooth a data set is to create an approxi-mating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures/rapid phenomena. In smoothing, the data points of a signal are modified so individual points (presumably because of noise) are reduced, and points that are lower than the adjacent points are increased leading to a smoother signal. Smoothing may be used in two important ways that can aid in data analysis the aim of smoothing is to give a general idea of relatively slow changes of value with little attention paid to the close matching of data values, while curve fitting concentrates on achieving as close a match as possible.

C. *Clustering of motion vector:*

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).To cluster the MVs in the spatiotemporal neighborhood, it is desired to have a clustering algorithm that would adaptively give different number of clusters based on the maximum cluster distance specified at input, along with the additional constraint that this will take a fixed number of iterations.

D. *True motion vector:*

After clustering of motion vector the true motion is obtained the motion vector field which is mainly tracked object motion with the back ground is same as image and moving object is in white blocks. After the encoding of the data with the predicted value and the other information in decoder side the original video frames are produced by assembling the all blocks.

E. *Demap all the frames:*

After encoding the predicted value of true motion vector and other blocks are get together for producing the original video frames. in decoder side.

F. *Interpolation:*

Interpolation is a method of constructing new data points within the range of a discrete set of known data points, Obtained dense motion fields at the interpolation instant are used to generate the interpolated frame. This means the additional extra frames are produces in between the existing frames to produce the better video quality of frames.

G. *performance parameter:*

1) *PSNR:*

PSNR is most commonly used to measure the quality of reconstruction of lossy comp-ression codecs. The signal in this case is the original data, and the noise is the error introduced by compression. PSNR is an approximation to human perception of reconstruction quality. Although a higher PSNR generally indicates that the reconstruction is of higher quality. PSNR is most easily defined via the mean squared error (*MSE*).

2) *SSIM:*

To define the image is fully matched with other image means the metric index SSIM should be nearer to 1. And it is not similar means it should be zero. The Structural Similarity (SSIM) index is a method for measuring the similarity between two images. The SSIM index can be viewed as a quality measure of one of the images being compared, provided the other image is regarded as of perfect quality This will be compared with the interpolated frames and original frames.

IV. RESULTS

This section gives the implemented results, that is how to find out true motion and its result gives the tracked object motion. And the reconstructed video in decoder side then it gives the interpolated video sequences and frames then lastly the performance metrics like the PSNR, SSIM, and MSE and before interpolation how many number of frames is there and after the interpolation the number of frame present.

The fig.3 shows that the detected motion object between frame 1 and frame 30. It gives the detected motion is in whitened blocks and the reamining areas are kept as same like both the frames.The detected motined object is of blocked form with the clusterd of these motion blocks gives the true motion vectors

Fig.3: frame no.1

Fig. 4: frame no.30

Fig. 5: true motion detecion

The oveall project is implemented in mat lab simulink so I get the overall out put in the graphical user interface(GUI) window. This winow shows in fig. 4

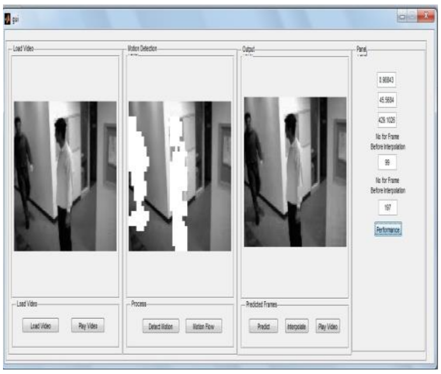

Fig. 6: GUI window of overall project

To discuss the results I am considering a video as vipemen.avi format which is consists of total frames 100. And it is of 160*120 format. Out of 100 frames first frame is the reference frame. With the help of this frame am finding out the true motion for all remaining frame by tracking of motion as closely as possible. After true motion estimation it has clusterd motion blocks ,by using these motion motion blocks am going to demap all the frames and lastly produce the original video of 100 frames. After demap final step is to produce the interpolated frames between the existing frames. Basically 99 frames are considering as original frames after inserting the interpolated frame totally 197 frames are formed shows fig.5 .And then playing all the frames in GUI. It shows slowly moving frames with smooth video.

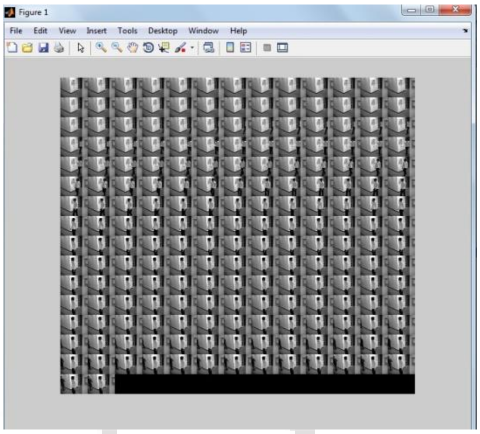

Fig.7: Total number of the frames after interpolation

After interpolation the performance metrics is calculated in between the extracted interpolated frame and the original frames. The MSE ,PSNR and SSIM is produced by using the formula for corresponding frames.

Totally 99 frames am comparing and producing the performance metrics. Out of all the maximum is like 45.56 dB of PSNR ,and 0.988 of SSIM it shows in GUI of window.so the SSIM is very nearest to 1 I can conclude that the interpolated frames are good similarity to the original frames.

The PSNR and SSIM for different video frames are given in fig. 6 and 7 respectively.And the finally SSIM map for all different frames is shown in the fig. 8 which shows the how much map interpolated frames with the original frames. Which is gives the how the frames matches to the interpolated frames with the original frames .so SSIM is very close to 1 it conclude that the frames which are produced are very good. And also produces the best videos quality of frames.

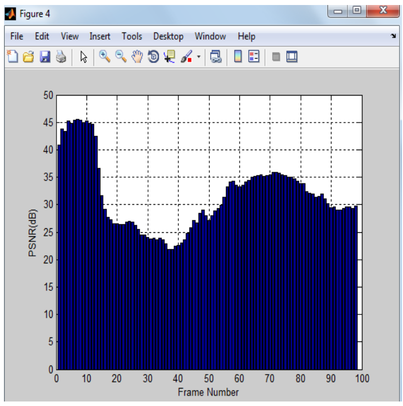

Fig.8: PSNR for all different frames

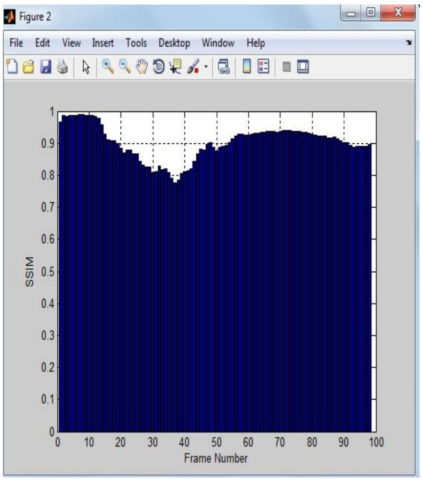

Fig.9: SSIM value for all different frames

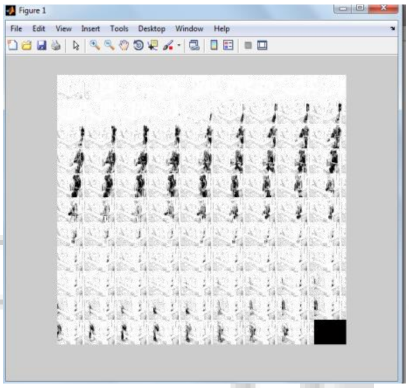

Fig. 10: SSIM map for all different frames

Totally I got results by using proposed method to find out the true motion estimation for motioned object tracking and later onwards compensated temporal frame interpolation frames is produce and the performance metric is also calculated and shown it in as graph. At last the video is also produced it is very smoothly plays.

V. CONCLUSION

A new true-motion estimation technique and its application to MCFRUC is presented Computational-complexity, regularity, and memory bandwidth are considered when designing the algorithm so that a low-complexity. The proposed TME algorithm imposes implicit and explicit smoothness constraints on BMA after finding out motion estimation the clustering of motion blocks is used which is gives the coherent motion vector field. Then by using the true motion estimation the motion compensated temporal frame is obtained .lastly the performance parameter is also calculated .The SSIM of both interpolated and original frame is getting near to one, this shows the interpolated frame are good after the production of video, then the good video sequences are getting and it play very smoothly.

REFERENCES

[1] A.-M. Huang and T. Nguyen, “Correlation-based motion vector processing with adaptive interpolation scheme for motion-compensated frame interpolation,” IEEE Trans. Image Process., vol. 18, no. 4, pp. 740–752, Apr. 2009.

[2] G. de Haan, P. W. A. C. Biezen, H. Huijgen, and O. A. Ojo, “True motion estimation with 3-D recursive search block matching,” IEEE Trans. Circuits Syst. Video Technol., vol. 3, no. 5, pp. 368–379, Oct. 1993.

[3] A. M. Tekalp, “Digital Video Processing”. Upper Saddle River”, NJ: Prentice-Hall, 1995.

[4] J. Wang, D. Wang, and W. Zhang, “Temporal compensated motion estimation with simple block-based prediction,” IEEE Trans. Broadcast., vol. 49, no. 3, pp. 241–248, Sep. 2003.

[5] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, “Modeling visual attention’s modulatory aftereffects on visual sensitivity and quality evaluation,” IEEE Trans. Image Process., vol. 14, no. 11, pp. 1928–1942, Nov. 2005

[6] K.-C. Yang, A.-M. Huang, T. Q. Nguyen, C. C. Guest, and P. K. Das, “A new objective quality metric for frame interpolation used in video compression,” IEEE Trans. Broadcast., vol. 54, no. 3, pp. 680–690, Sep. 2008.

[7] B.-T.Choi, S.-H. Lee, Y.-J. Park, and S.-J. Ko, “Frame rate upconversion using the wavelet transform,” in Proc. Int. Conf. Consum. Electron., Los Angeles, CA, Jun. 2000, pp. 172–173

[8] A.-M. Huang and T. Nguyen, “A multistage motion vector processing method for motion-compensated frame interpolation,” IEEE Trans. Image Process., vol. 17, no. 5, pp. 694–708, May 2008.

[9] S. Dikbas, T. Arici, and Y. Altunbasak, “Fast motion estimation with interpolation-free sub-sample accuracy,” IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 7, pp. 1047–1051, Jul. 2010.

[10] M. Biswas and T. Nguyen, “A novel motion estimation algorithm using phase plane correlation for frame rate conversion,” in Proc. Rec. 36th Asilomar Conf. Signals, Syst. Comput., vol. 1. Nov. 2002, pp. 492–496

[11] J. Wang, D. Wang, and W. Zhang, “Temporal compensated motion estimation with simple block-based prediction,” IEEE Trans. Broadcast., vol. 49, no. 3, pp. 241–248, Sep. 2003.

