

RESEARCH

Open Access

Confidence-based adaptive frame rate up-conversion

Kyung-Yeon Min and Dong-Gyu Sim*

Abstract

In this article, we propose a new frame rate up-conversion (FRU) method for temporal quality enhancement. The proposed FRU algorithm employs gradual and adaptive motion estimation based on confidence priority for selecting more accurate motion vectors (MVs). In order to estimate accurate MVs, we adaptively alternate a search range and the order of blocks to be searched depending on the confidence level in hierarchical motion estimation. The precedence of the proposed algorithm is conducted based on the confidence level that is decided by the complexity of pixel values in a block. In addition, we perform bi-directional motion compensation and spatial linear interpolation to fill occlusion regions. In our experiments, we found that the proposed algorithm is about 2 dB better than several conventional methods. Furthermore, block artifacts and blur artifacts are significantly diminished by the proposed algorithm.

Keywords: Video coding, Frame rate up-conversion, Hierarchical motion estimation, Adaptive search range, Bi-directional motion estimation

1. Introduction

As the technology of display devices is developed, various high-performance display devices have become available to users in the market. While several current video coding and transmission specifications have been defined for relatively old devices having HD spatial resolution and a temporal rate of 30 fps, spatial and temporal resolutions of many commercial display devices have become higher [1]. Furthermore, they have the functionality to enhance spatial and temporal resolution during post-processing. Frame rate up-conversion (FRU) methods are suited for the post-processing of display devices for high perceptual quality, as FRU algorithms can enhance temporal quality without increasing the bitrates [1-14].

Up-conversion methods in a temporal axis can be used to produce an inter-frame in between two consecutive frames at a decoder side without any direct information for generation of the inter frame from an encoder. Many FRU algorithms have been developed, and they can be classified into two categories [3,6-22]. One approach employs frame repetition or linear interpolation regardless of object motion. Although this approach provides acceptable video

quality in absence of fast motion with minimum complexity, we can easily see motion jerkiness and blurring from moving objects at an interpolated frame. For the second category, motion compensation (MC) techniques are used to reduce these artifacts based on a block matching algorithm (BMA) based on redundancy between consecutive frames. This approach can improve the perceptual quality of videos by reducing annoying discontinuity between reconstructed frames and interpolated frames, as interpolated frames are generated by exploring the assumption of linear motion between consecutive frames.

However, it is not easy to estimate the real motion of objects and background, because motion estimation is regarded as an underdetermined problem. The block having the least sum of absolute difference (SAD) should be selected, even if there is no proper matching point for a target block. Therefore, several conventional algorithms employed weighted median filters, weighted zero MVs, and bi-directional motion estimation to refine or select true motion vectors (MVs). As it is hard to achieve satisfactory performance with only one constraint, two or more constraints were used in several conventional algorithms [3-9]. In order to remove blocking artifacts which are frequently founded at block boundaries in an interpolated frame, an overlapped block motion estimation and compensation (OBME/MC) method can be used.

* Correspondence: dgsim@kw.ac.kr
Department of Computer Engineering, Kwangwoon University, 447-1,
Wolgye-dong, Nowon-gu, Seoul, Korea

There are several existing algorithms to estimate true MV field [23]. True MVs would be helpful in generating interpolated frames. However, these algorithms require relatively a large number of computations. In order to utilize the redundancy of consecutive frames, we designed a new FRU algorithm based on BMA. As it is hard to find true MVs with only pixel values, the proposed method takes into account not only pixel values, but also motion relations between a target block and adjacent blocks in motion estimation. Note that the adjacent MVs are highly correlated, thus, accurate MVs in motion estimation could propagate positive effects to neighboring vectors, with the result that the accuracy of MVs in the neighboring blocks can be improved. In order to employ more accurate neighboring MVs in estimating an MV in a target block, the order and search range of motion estimation are determined based on the confidence levels of blocks. In addition, bi-directional MC and spatial linear interpolation are adopted in order to fill occlusion regions without blocking artifacts, because MVs of some objects cannot be estimated with linear motion estimations and/or some objects exist in only one-side frame. Note that BMA could suffer from blocking artifacts. In the proposed algorithm, the artifacts are alleviated by employing OBME/MC.

The remainder of this article is organized as follows. Section 2 presents previous works in brief. In Section 3, the proposed method is presented. Experimental results are given in Section 4. Finally, the conclusions are given in Section 5.

2. Conventional FRU algorithms

The many conventional FRU algorithms can be classified into two categories: non-motion compensated and motion compensated FRU methods. Repetition-based methods and linear interpolation methods have been proposed, and they can be classified as non-motion-compensated approaches [3-6]. The repetition-based FRU method is quite simple and can be fast worked. However, jerky artifacts are frequently observed, because the repetition method does not take into account moving objects. The linear interpolation method can reduce jerky artifacts, compared with the repetition methods [2]. Linear interpolation algorithms are not so complicated and moderately fast. Relatively high PSNR is expected at many target frames for slow-moving video. However, we can easily observe blur artifacts in the interpolated frames of videos with fast-moving objects.

Because the non-motion compensated methods do not make use of motion information, they cannot recover accurate interpolated frames for videos having high motion activity. To recover higher-frequency components in temporal domain, motion compensated interpolation algorithms have been proposed and they usually employ a BMA to estimate MVs for moving objects [4,7-9,17-20]. In order to improve the accuracy of MVs, conventional algorithms were proposed

to determine MVs by employing weighted median filter, weighted zero MVs, and/or bi-directional motion estimation algorithms [4,7,9]. On the other hand, the weighted median filter was employed to adjust estimated MVs [4,7]. This weighted median filter is performed based on the hypothesis that the majority of estimated neighboring MVs are likely to be correct. In addition, a weighted zero MV method was proposed to keep consistency of MVs in background regions by giving favors into zero MVs [6,7]. Moglia et al. [16] proposed a bi-directional ME which improves the accuracy of motion estimation with backward and forward frames. This FRU algorithm performs forward motion estimation to estimate an initial interpolated frame from the (t) frame to the ($t - 1$) frame. Then, the initial interpolation frame is divided into 4×4 , 8×8 , or 16×16 blocks. The bi-directional ME is performed, as the second stage motion estimation, to find the best matching blocks in the backward and forward frames from a block in the initial interpolated frame [9]. However, the block-based MC suffers from blocky artifacts. The OBME/MC method was proposed to reduce blocking artifacts. The OBME method performs motion estimation with partially overlapped blocks and the compensated frame is reconstructed by weighted summation of overlapped corresponding blocks [7-9,14]. Although the OBMC has high computational complexity, blocky artifacts can significantly be reduced.

Many motion estimation algorithms have been proposed, however, one tool cannot produce satisfactory outcomes. Recent work tries to combine multiple tools to improve the accuracy of FRU. A combination of non-motion-based and motion-based methods was proposed [14]. This combined method generates multiple layers based on Gaussian and Laplacian pyramids. The target frame at the finer resolution is reconstructed with the up-sampled frame as well as the previous and next frames at the same resolution. The hierarchical algorithm is stable in estimating MVs by using not only the previous and next frames, but also interpolated frames in a coarse resolution. In general, MVs at the coarse resolution increase reliability and stability in further motion estimation because dominant MVs are computed at the coarse resolution. However, this hierarchical FRU method requires high computational complexity and it could still have blocky artifacts. Fujiwara and Taguchi [4] proposed a combination of weighted zero MV and weighted median filter. The block size is adaptively selected depending on the MV distribution of adjacent blocks. In addition, multi-frame FRU algorithm was proposed to reduce motion ambiguity instead of using two consecutive frames [11]. On the other hand, pixel-based auto-regressive model was presented and it would be useful in generating accurate interpolated frames; however, it requires relatively a large number of computations [12]. For real-time applications, a low complexity frame rate up-conversion (FRUC) was proposed by alleviate illumination change effects [13].

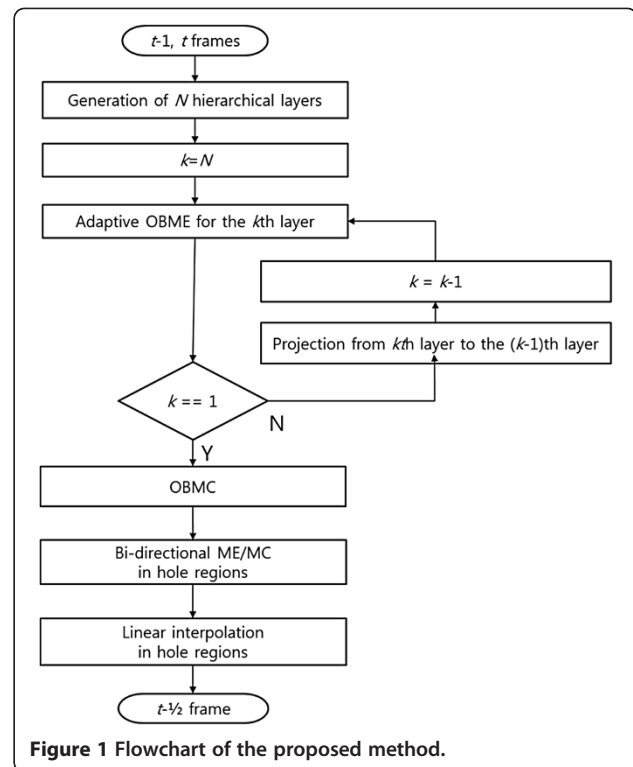
3. The proposed confidence-based adaptive FRU

Consecutive frames of a video have lots of similarity. Conventional FRU algorithms have been developed by utilizing temporal redundancy, as they perform motion estimation with pixel intensity within a fixed search range. However, it is hard to estimate true MVs, because there are many pixels at which estimated MVs are not unique. In order to decide on MVs, many conventional algorithms were proposed by finding a block having a smallest SAD for a target block. However, motion estimation in homogenous regions suffers from the aperture problem which does not have a unique solution even with several constraints. Furthermore, blocks in which wrong MVs are estimated can improperly influence motion estimation in neighboring blocks.

For the proposed algorithm, motion estimation is conducted based on highest confidence first. Unique motion can be determined in edge and corner regions unlike homogenous regions. In addition, we estimate MVs at blocks having complex texture without the aperture problem. Therefore, we can assume that the MV matched with a block having higher variance is more reliable than those matched with other blocks. Based on high confidence first, the reliability of the high confidence blocks can propagate into neighboring blocks. With this propagation, accuracy of motion estimation can dramatically be improved even in homogenous regions.

The proposed algorithm performs a hierarchical OBME according to a confidence evaluation with adaptive search ranges and adaptive ordering in motion estimation. Because edge and corner regions have a large variance in general, the confidence value is estimated by the variance of a block in the proposed algorithm. Therefore, the proposed algorithm can increase the probability that an estimated MV is more accurate, as the adaptive search range and block ordering are employed. In addition, as a hierarchical motion estimation method is utilized in the proposed method and, thus, we can prevent the solution from being stuck in local maxima and reduce the computational complexity.

Figure 1 is the flowchart of the proposed method. At first, two consecutive frames are used as input images for the proposed system. N hierarchical layers ($k = 1, \dots, N$) are generated and OBME is performed in each layer. Motion estimation is performed in a descending order of block variance, and the search range of each block is adaptively determined, depending on the MV in adjacent interpolated blocks at the interpolated frame. The search range is set not to overlap regions covered with neighboring reliable MVs. As a result, it can increase the accuracy of MVs at not only complex texture, but also homogenous regions. In addition, the motion estimation is performed from the coarse resolution to the fine resolution in order to improve performance of hierarchical motion estimation. The MV field is projected for the finer resolution using up-sampling filtering. Note that we



make use of the OBME to reduce blocky artifacts and stabilize the estimated MVs. OBMC is also employed to estimate the final MVs at the finest resolution layer of which size is the same as the original. While the conventional algorithm makes use of OBMC in all the layers, the proposed algorithm employs the OBMC in the last layer with interpolation of MV fields to reduce computational complexity.

In some regions, correspondence cannot be accomplished between the previous and next frames in cases that some regions in the two frames are occluded or occluding. Furthermore, MVs cannot be estimated for rotational movement because the proposed algorithm works with the assumption of translational rigid body motion. We can see several regions that are not compensated forward frame with OBMC and they are called hole regions. In the proposed algorithm, we perform a forward motion estimation to reconstruct a hole region with the neighboring area excluding the target hole pixels in matching. Then, the backward motion estimation is conducted for the hole regions again. The matched block having a smaller distance between them for each hole is selected. When the minimum distance value is larger than a threshold, a spatial linear interpolation is conducted in order to fill the occlusion regions to reduce blocking artifacts.

3.1 Generation of hierarchical layers

Hierarchical motion estimation can reduce computational complexity and avoid being stuck in local maxima.

The proposed method employs hierarchical motion estimation to enhance stability of motion estimation. For hierarchical motion estimation, MVs should be projected from a coarse layer (layer k) to the next enhanced layer (layer $k - 1$), as shown in Figure 2.

To reduce aliasing artifacts based on sub-sampling, we employ a simple low-pass filter prior to sub-sampling, when hierarchical layers are generated. With the down-scaling factor of 2, the down-sampled frame at the k th layer (L_k) can be computed by

$$L_k(j, i) = \frac{1}{4}(L_{k-1}(2j, 2i) + L_{k-1}(2j, 2i + 1) + L_{k-1}(2j + 1, 2i) + L_{k-1}(2j + 1, 2i + 1)) \quad (1)$$

where (j, i) is the index indicating the position of a pixel. Note that $j(i)$ ranges from 0 to $J - 1$ ($I - 1$) for the column (row) index.

3.2 Overlapped block motion estimation based on adaptive search ranges with confidence levels

Interpolated frames can be generated with block-based motion estimation algorithms; however, BMAs generally suffer from blocking artifacts. Because the proposed method is also one of the block-based algorithms, we employed OBME/MC to reduce blocking artifacts. Thus, the OBMC can increase computational complexity of the proposed algorithm.

In the proposed algorithm, the search range is adaptively adjusted at all the hierarchical layers according to the confidence level, with proper projection of the estimated MVs between adjacent layers. The confidence level of a block is defined by a block variance. Thus, the confidence level for

each block is computed by $C(m, n, t) = \sum_{j=0}^M \sum_{i=0}^N |f(t, j, i) -$

$\bar{f}(t, j, i)|^2$, where $f(t, j, i)$ represents the intensity on the (j, i) th position at the time t and $\bar{f}(t, j, i)$ is the mean value of the $M \times N$ window. The confidence-based hierarchical OBME is sequentially conducted in descending order of the confidence levels. For conventional FRU algorithms, motion

estimation is performed with a fixed search range, which suffers from wrong overlapping in the (t) frame. Wrong estimated MVs could also affect adjacent blocks even though accurate MVs are estimated at the adjacent block. Therefore, the search range of the proposed algorithm is adaptively set by reliable MVs of neighboring blocks. That can prevent incorrect overlapping among blocks. The restriction can be interpreted as a smoothness constraint on MVs for homogenous regions.

Figure 3 shows several examples of how to determine adaptive search ranges at the $(t - 1)$ frame in motion estimation. Note that matched blocks are denoted with gray and they have been matched in previous matching steps, because variance of the matched block is larger than those of neighboring blocks. Then, we can restrict a search range denoted by a solid box for the neighboring target block and the search range is smaller than the original maximum rectangle, as shown in Figure 3a. If there are not any matched neighboring blocks, the search range is set to the maximum one, as the first example in Figure 3b. If there are matched blocks, the search range can reduce based on matching consistency. Restriction of a search range can prohibit overlapping of two adjacent blocks in an interpolated frame. However, the proposed algorithm employs OBME/MC and the overlapping for OBMC is accounted for the reduction of artificial blocking. Note that the corresponding block in the restricted search range (solid line) of the $(t - 1)$ frame is selected by finding a block having the smallest SAD value between pixels in the target block and searched block by

$$SAD(m, n, MV_j, MV_i, t) = \sum_{j=0}^M \sum_{i=0}^N |f(t, j + Mm, i + Nn) - f(t - 1, j + Mm + MV_j, i + Nn + MV_i)| \quad (2)$$

where f means pixel value, (t, j, i) means temporal index and position indices, and $M \times N$ means the block size. (m, n) and (M_j, M_i) represent the target block indices and MV for the (m, n) th target block. Note that each searched block should be in the search area that is determined in

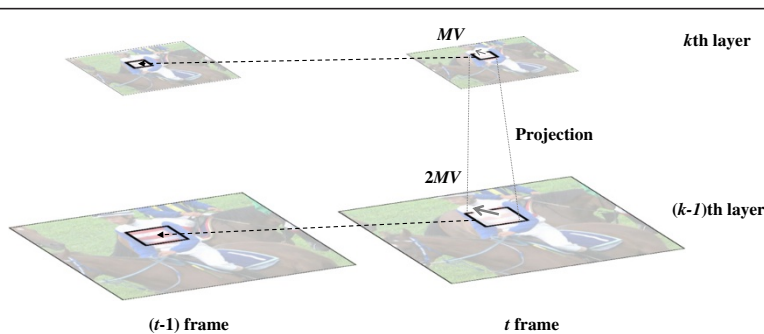
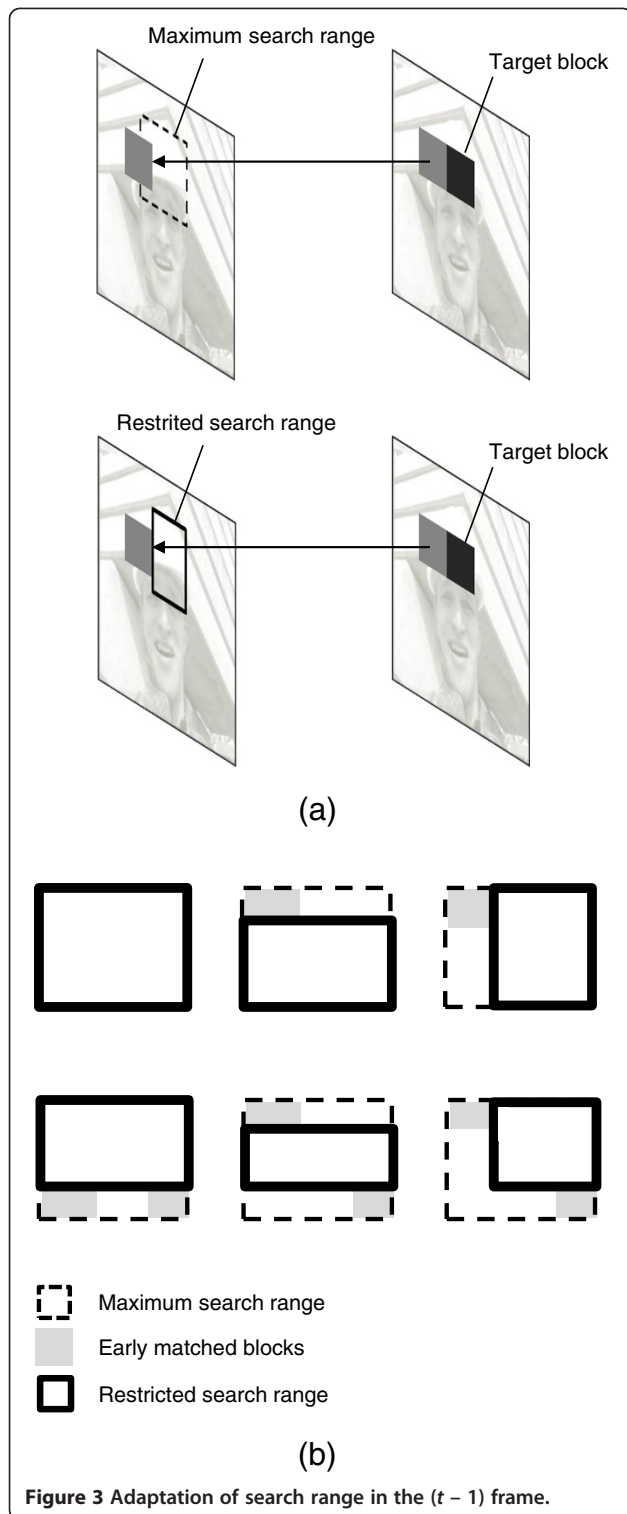


Figure 2 Projection of MV in between consecutive layers.



the previous step, as shown in Figure 3. In this study, we employed the full search algorithm in the search area.

We decide each block corresponded with any block at the $(t - 1)$ frame if the associated SAD is smaller than a threshold. If so, the block is compensated for the

interpolated frame. Otherwise, it is left not interpolated as a hole region. Due to occlusion, we can see several hole regions in the interpolated frame. The hole regions are compensated by bi-directional ME/MC or spatial interpolation.

3. 3 Overlapped block MC

The proposed method makes use of OBMC, in order to reduce the block artifacts, as mentioned before. We generate the interpolated frame with the average pixel values of matched blocks which are decided by motion estimation between previous and next frames. We perform the OBMC only on the last layer and the MC is defined by

$$f\left(t + \frac{1}{2}, j + \frac{MV_j}{2}, i + \frac{MV_i}{2}\right) = \frac{f(t - 1, j + MV_j, i + MV_i) + f(t, j, i)}{2} \quad (3)$$

for each block. The block is overlapped with neighboring blocks in one pixel, as shown in Figure 4. Depending on the location, one to four overlaps occur. As shown in Figure 4, four prediction values are used for the four corner points. For boundary pixels, two prediction values are used. For the center region, there is no overlap, as shown in the figure. For the regions, the interpolated values are computed by averaging all the overlapped pixel values.

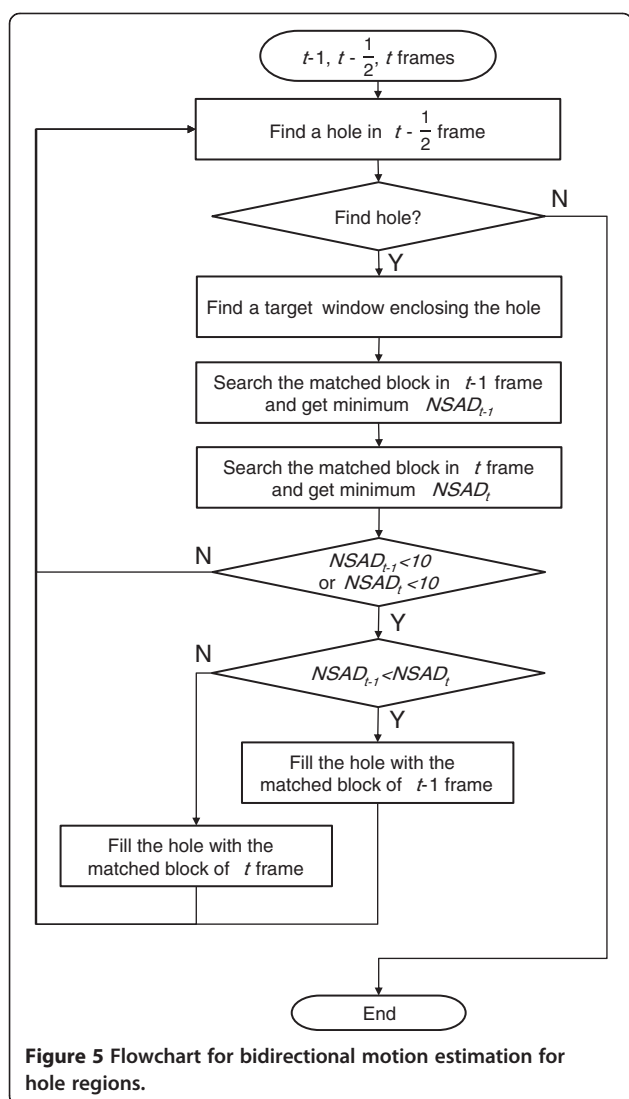
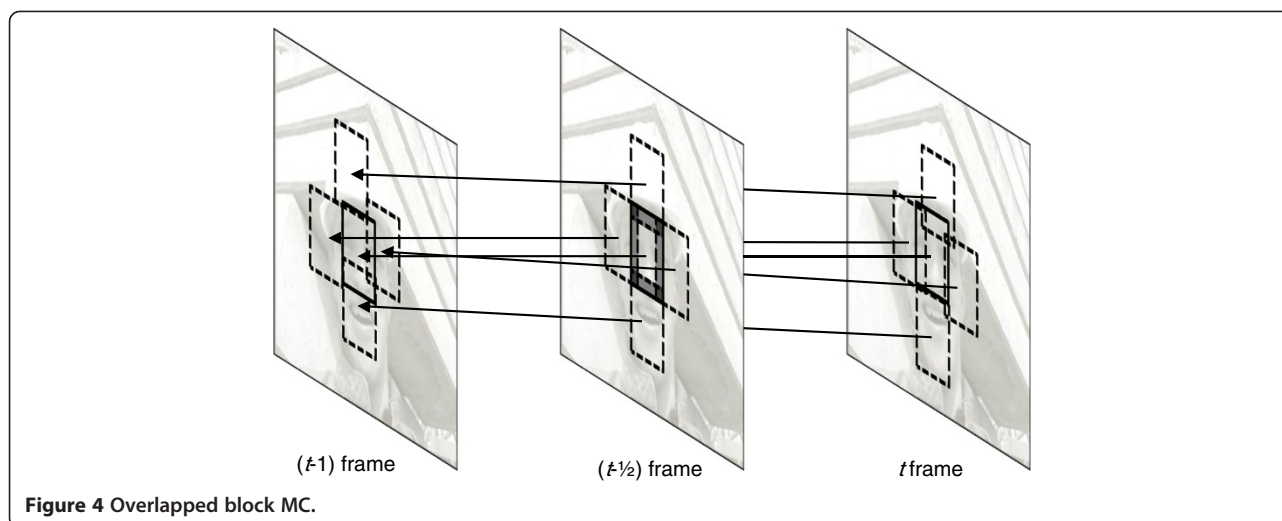
3. 4 Bidirectional motion estimation in hole regions

When an object is shown in only one of the consecutive frames, SAD could be bigger than the threshold, because the correct correspondence cannot be accomplished. As a result, the interpolated frame could have several occlusion regions.

Figure 5 shows the block diagram of the proposed bidirectional motion estimation for hole regions. The hole regions should be filled with forward or backward frames. In the proposed algorithm, a matched area is detected with neighboring pixels of a hole region, as shown in Figure 5. The hole is filled with the region enclosed by the match area. This algorithm can work if an object appears in a forward or backward frame.

Figure 6 shows an example in which the bi-directional motion estimation is performed at occlusion regions. We assume that white region (a) is an occlusion region and the other region in the block is an interpolated region in

$$NSAD_{t-1} = \frac{1}{\sum_{j,i} M(j,i)} \sum_j \sum_i M(j,i) \left(f\left(t - \frac{1}{2}, j, i\right) - f(t - 1, j + MV_j, i + MV_i) \right)^2 \quad (4)$$

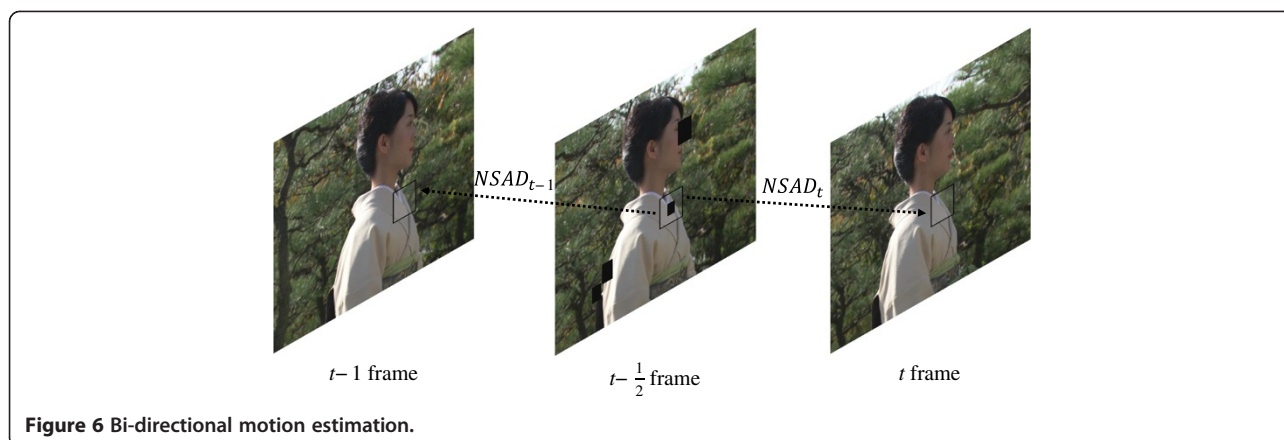


the previous stage. Then, we make use of a block enclosing a to find minimum normalized cost between $NSAD_{t-1}$ and $NSAD_t$. The cost $NSAD_{t-1}$ can be denoted by where $M(i,j)$ is a mask that represent hole pixel or not. $M(i,j)$ is set to 0 when the pixel is in a hole in the interpolated frame $f(t-1/2,j,i)$. Otherwise, it is set to 1.

Only interpolated pixel values are used in computing the NSAD. The corresponding block in the $(t-1)$ frame is identified by finding minimum $NSAD_{t-1}$ with only an interpolated square block enclosing the interpolated region in the $(t-1/2)$ frame. Then, we also find the corresponding block in the (t) frame by computing $NSAD_t$ with the interpolated region in the $(t-1/2)$ frame. Either block is selected by finding the smaller NSAD value between the previous or next frames. If the minimum normalized SAD is less than a threshold, we compensate the hole pixels with the corresponding block.

3.5 Spatial linear interpolation

While consecutive frames in a video have lots of temporal similarity, it is hard to find temporal similarity in part of an image for several reasons such as occlusion and homogenous background. This causes many small holes and blocking artifacts to appear because of the compensation of motion estimation at such regions in the interpolated frame. In order to compensate the occlusion regions without blocking artifacts, we employ a spatial linear interpolation using only spatial correlation. Figure 7 shows an example of spatial interpolation being conducted at the interpolated frame. We assume that the shaded block is the hole pixels and the white regions are interpolated pixels. A pixel a in the occlusion region is compensated with four interpolated pixels from four directions. In computing the



interpolation value, we make use of four weighted pixel values and it can be defined by

$$f\left(t + \frac{1}{2}, j, i\right) = \frac{(D_{left}^{-1}f(t + \frac{1}{2}, j, i - D_{left}) + D_{right}^{-1}f(t + \frac{1}{2}, j, i + D_{right}) + D_{up}^{-1}f(t + \frac{1}{2}, j - D_{up}, i) + D_{down}^{-1}f(t + \frac{1}{2}, j + D_{down}, i))}{D_{left}^{-1} + D_{right}^{-1} + D_{up}^{-1} + D_{down}^{-1}} \quad (5)$$

where D represents the distance between the target pixel and the nearest neighboring interpolated pixel. Spatial linear interpolation makes blur artifacts. However, occlusion regions are very small and frames change fast. Therefore, people cannot exactly see the blur artifacts clearly.

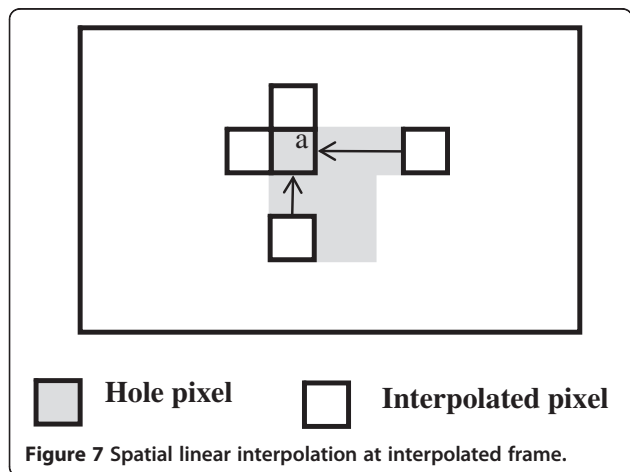
4. Experimental results and discussion

For performance evaluation, we used six test sequences: ‘Akko’, ‘Ballroom’, ‘Exit’, ‘Flamenco2’, ‘Race1’, and ‘Rena’. The sequences include slow-, middle-, and fast-translational/rotational motions and camera panning. ‘Race1’ is an outdoor sequence and the others are indoor sequences. The color format and dimension of all the sequences are 4:0:0 YUV and 640×480 pixels, respectively. In addition,

performance of the proposed and conventional methods [4-6] was evaluated for not only original videos, but also

decoded videos from H.264/AVC bitstreams. The bitstreams were coded with JM17.2. The frame rate for ‘Akko’, ‘Flamenco2’, ‘Race1’, and ‘Rena’ sequences is 15 fps and that for ‘Ballroom’ and ‘Exit’ is 12.5 fps. These sequences were used for video compression standardization. In our experiment, we used 50 frames for each sequence. We used four QP values (22, 27, 32, and 37). Two layers are used in the proposed hierarchical motion estimation based on the trade-off between complexity and motion estimation accuracy. The block size is set to 16×16 in all the layers and 7×7 in the bi-directional ME. In addition, the search range is set to ± 32 in the coarse layer and ± 16 in the fine layer. Hierarchical OBME/MC is conducted when a SAD value between two matched blocks is smaller than the threshold (6000). Otherwise, the block is set to a hole. For the hole, bi-directional interpolation is used when the normalized SAD is smaller than 10. Otherwise, the spatial linear interpolation is used to fill hole regions. All the thresholds are empirically determined. Computational loads are evaluated with the Intel dual core system.

Figure 8a–c shows original ‘Akko’ frames of which frame indexes are 14, 15, and 16, respectively. Note that the 14th and 16th frames are used as input frames for FRU. The objective of the FRU is to estimate an interpolation frame which is close to the original 15th frame. Figure 8d shows an intermediate interpolated frame with the proposed hierarchical OBME/MC for the input frames (Figure 8a, c). Figure 8e is the interpolated frame performing the hierarchical OBME/MC and bidirectional ME/MC. Figure 8f is the final interpolated frame by the proposed FRU. Owing to the confidence-



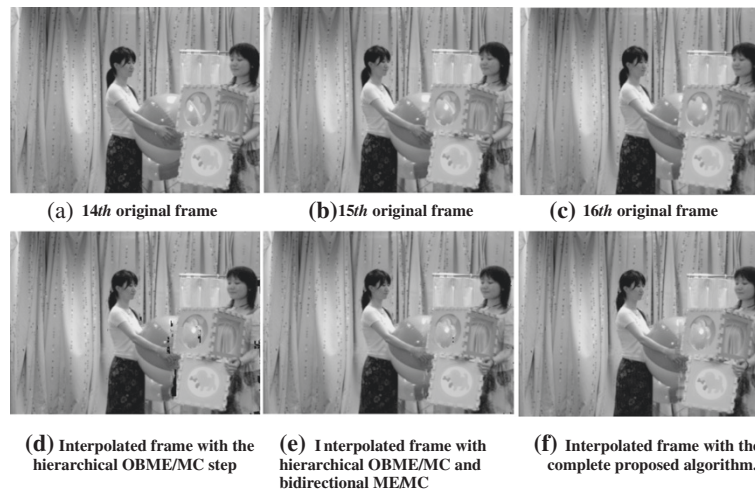


Figure 8 Interpolated frame at each step for 'Akko' sequence.

based iteration, the estimated MVs of blocks having complex textures propagate the adjacent blocks in the homogenous regions. Therefore, the interpolated frame by the proposed algorithm does not have visual artifacts, as shown in Figure 8d. However, regions near moving objects can be influenced by the MVs of the moving objects. Furthermore, the regions, in general, belong to covered or uncovered regions. In order to properly compensate the occluded regions, bi-directional ME and weighted spatial interpolation are employed in the proposed method. When the occluded background regions are homogenous, we hardly see visual artifacts even with inaccurate MVs from

that of a moving object. However, when the occluded regions have complicated textures, some visual artifacts can be found not only conventional algorithms, but also the proposed algorithm, as shown in Figure 8e, f.

Figure 9 shows the original and interpolated frames by the proposed and conventional algorithms for the 'Flamenco2' sequence. For comparative study, we selected three conventional algorithms based on hierarchical ME (HME) [4], diverse block sizes (DBS) [5], and multiple reference frames (MRF) [6]. We can see that the proposed algorithm is better around edge regions in terms of blur and blocking artifacts for arms, face, and dresses areas. The interpolated frames by

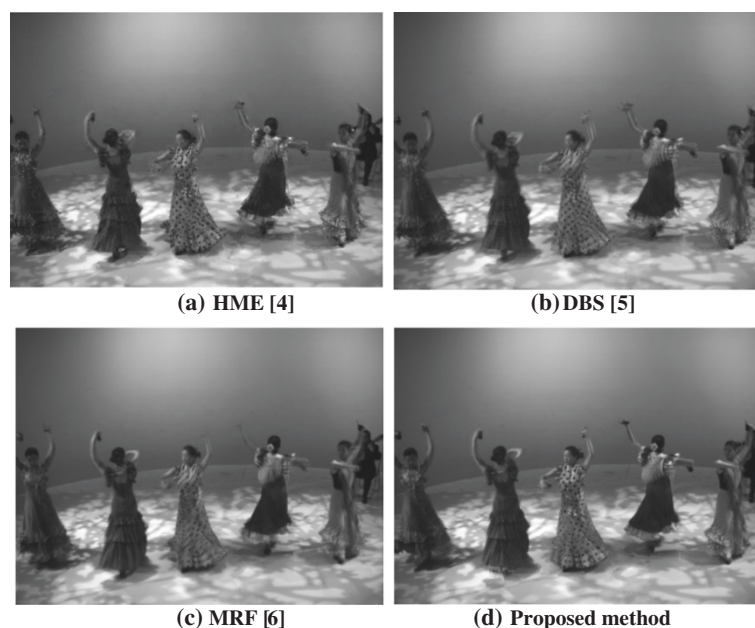


Figure 9 Interpolated frames of the conventional and proposed algorithms for 'Flamenco2' sequence.

Table 1 PSNR comparison of interpolation frames with the proposed and conventional algorithm

Test sequence	QP	HME [4]	DBS [5]	MRF [6]	Proposed
Akko	Original	24.70	31.29	32.10	34.12
	22	24.68	31.04	32.09	33.23
	27	24.66	30.72	31.91	33.23
	32	24.62	29.96	31.40	32.54
	37	24.52	28.85	30.59	31.43
	Average	24.62	30.14	31.50	32.61
Ballroom	Original	24.11	27.54	28.28	29.19
	22	24.08	27.47	27.95	28.86
	27	24.03	27.35	27.84	28.68
	32	23.93	26.94	27.59	28.35
	37	23.76	26.33	27.11	27.63
	Average	23.95	27.02	27.62	28.38
Exit	Original	32.63	35.97	37.32	37.12
	22	32.56	35.96	37.02	36.95
	27	32.45	35.79	36.67	36.66
	32	32.08	35.01	35.83	35.66
	37	31.45	33.66	34.50	34.28
	Average	32.14	35.11	36.01	35.89
Flamenco2	Original	27.32	28.70	30.97	31.41
	22	27.33	28.65	31.29	31.18
	27	27.33	28.53	31.24	31.03
	32	27.31	28.38	31.03	30.23
	37	27.25	27.93	30.60	30.23
	Average	27.31	28.37	31.04	30.67
Race1	Original	20.69	25.51	24.16	28.98
	22	20.64	25.34	24.11	28.37
	27	20.61	25.08	24.02	28.20
	32	20.54	24.63	23.84	27.84
	37	20.46	23.92	23.60	27.60
	Average	20.56	24.74	23.89	28.00
Rena	Original	31.82	33.80	35.14	35.67
	22	31.56	33.56	34.85	35.17
	27	31.45	33.39	34.58	34.89
	32	31.24	32.87	34.04	34.41
	37	30.91	32.09	33.26	33.45
	Average	31.29	32.98	34.18	34.48
Average		26.64	29.73	30.71	31.67

conventional algorithms have artifacts from incorrect interpolation (e.g., the woman's arm) and blur artifacts. However, the interpolated frame by the proposed method is without blur and blocking artifacts.

Table 1 shows PSNRs of the interpolated images with the proposed and conventional algorithms [4-6]. The

PSNRs are evaluated with the original videos and degraded ones with multiple QP values. Note that 50 interpolated frames are used for each sequence. The target frame rate is twice of the input frame rate. PSNR values of the proposed algorithm are higher than those of conventional algorithms for not only original but also decoded videos. Because the conventional algorithms refine MVs in the raster scanning order, early wrong

Table 2 Quality comparison of interpolation frames with the proposed and conventional algorithms in terms of various metrics

Test sequence	Metric	HME [4]	DBS [5]	MRF [6]	Proposed
Akko	NQM	9.68	19.05	19.25	21.00
	IFC	2.19	5.26	5.40	5.58
	SSIM	0.76	0.94	0.93	0.95
	VIF	0.33	0.67	0.67	0.72
	UQI	0.55	0.86	0.84	0.86
Ballroom	NQM	8.82	14.70	15.40	15.84
	IFC	2.49	4.15	4.35	4.28
	SSIM	0.81	0.89	0.91	0.90
	VIF	0.36	0.56	0.58	0.58
	UQI	0.62	0.73	0.75	0.75
Exit	NQM	17.25	27.36	29.72	28.18
	IFC	1.67	2.97	3.20	3.16
	SSIM	0.85	0.92	0.94	0.93
	VIF	0.31	0.51	0.53	0.53
	UQI	0.47	0.56	0.60	0.60
Flamenco2	NQM	14.32	17.69	20.66	20.30
	IFC	1.30	2.20	2.76	2.63
	SSIM	0.84	0.89	0.93	0.93
	VIF	0.23	0.36	0.45	0.44
	UQI	0.57	0.69	0.74	0.74
Race1	NQM	7.28	15.50	13.45	18.52
	IFC	0.75	2.43	1.95	3.00
	SSIM	0.64	0.80	0.77	0.89
	VIF	0.11	0.34	0.26	0.44
	UQI	0.29	0.58	0.50	0.66
Rena	NQM	8.47	12.40	12.77	13.61
	IFC	2.45	4.57	5.24	5.05
	SSIM	0.92	0.96	0.97	0.97
	VIF	0.45	0.69	0.76	0.73
	UQI	0.58	0.76	0.80	0.79
Average	NQM	8.09	13.22	13.59	14.88
	IFC	1.53	3.10	3.28	3.42
	SSIM	0.66	0.74	0.75	0.77
	VIF	0.25	0.44	0.45	0.48
	UQI	0.43	0.60	0.60	0.63

Table 3 PNSs in terms of various input frame rates for degraded video

Test sequence	Input frame rate (fps)	Output frame rate (fps)	HME [4]	DBS [5]	MRF [6]	Proposed
Akko	15.0	30	24.68	31.04	32.84	33.23
	7.50		20.87	26.48	20.43	29.35
	3.750		18.55	21.09	20.79	24.15
	Average		21.37	26.20	24.69	28.91
Ballroom	12.5	25	24.08	27.47	28.31	28.86
	6.25		20.90	23.64	21.08	24.65
	3.125		19.04	20.74	20.24	20.48
	Average		21.34	23.95	23.21	24.66
Exit	12.5	25	32.56	35.96	37.37	36.95
	6.25		28.58	33.17	28.73	34.00
	3.125		26.90	29.84	29.12	30.51
	Average		29.35	32.99	31.74	33.82
Flamenco2	15.0	30	27.33	28.65	31.39	31.18
	7.50		24.37	26.22	24.67	27.46
	3.75		22.47	24.15	24.67	24.74
	Average		24.72	26.34	26.91	27.79
Race1	15.0	30	20.64	25.34	24.33	28.37
	7.50		18.43	21.37	19.34	24.04
	3.75		17.05	18.70	18.84	18.94
	Average		18.71	21.80	20.84	23.78
Rena	15.0	30	31.56	33.56	35.08	35.17
	7.50		28.05	30.88	28.13	31.85
	3.75		26.76	28.49	28.64	29.17
	Average		28.79	30.98	30.62	32.06
Average			24.05	27.04	26.33	28.51

estimation can improperly influence on consecutive blocks. However, as the proposed method estimates MVs with adaptive modification of the search range according to confidence levels, the estimated MVs can have a proper effect on consecutive blocks. Therefore, the proposed algorithm is better than the three conventional algorithms by around 5, 1.9, and 0.9 dB for the original input and decoded videos.

Table 4 Computational complexity comparison of the proposed and conventional algorithms for interpolating frame

Test sequence	HME [4] (%)	DBS [5] (%)	MRF [6] (%)
Akko	73	33	45
Ballroom	74	37	48
Exit	71	30	42
Flamenco2	73	32	44
Race	77	42	52
Rena	73	32	44

Table 2 shows the objective evaluations of the interpolated frames by the proposed and conventional methods for original videos in terms of various metrics. However, we know that the subjective quality is quite important in real applications. For more various quality evaluations, we added performance evaluation results with several metrics: noise quality measure (NQM) [24], information fidelity criterion (IFC) [25], structural similarity (SSIM) [26], visual information fidelity (VIF) [27], and universal quality index (UQI) [28]. Some of them are widely used for perceptual evaluation. As you know, PSNR and NQM are used for objective quality. For perceptual quality, IFC can give the quantity of mutual information between original and degraded sequences. SSIM is referred as an SSIM index and it is known to be highly correlated with the subjective quality. VIF can quantify the mutual information of distorted videos in terms of the human visual system. UQI estimates degradation as a combination of three factors that are loss of correlation, luminance distortion, and contrast distortion. Note that larger values with the metrics mean superior visual

quality. These objective metrics cannot give exact subjective assessments; however, they have been verified to be highly correlated with subjective quality. Table 2 shows the objective evaluations of the interpolated frames by the proposed and conventional methods. All the average values of NQM, IFC, SSIM, VIF, and UQI of the proposed are higher than those of conventional methods, as shown in the table. That implies the proposed method has quite good performance for not only objective quality but also perceptual quality.

Performances of the proposed and conventional methods were also evaluated for various conversion ratios for degraded frames by the H.264/AVC with QP of 22. The frame rates of the test sequences are 25 and 30 fps, and we evaluated PSNRs of the interpolated frames by reducing the input frame rates to 12.5, 6.25, and 3.125 fps for 25 fps output videos and 15, 7.5, and 3.75 fps for 30 fps output videos, respectively. As shown in Table 3, we found that the PSNRs for interpolated frame drop as the input rate decreases. With lower frame rates, larger MVs should be estimated and larger occlusion areas are dealt with. Thus, we can say that the PSNR drop could significantly reduce for practical applications with higher input frame rates. However, the proposed algorithm is better than the conventional algorithms even for the degraded videos having much lower input rates. For higher input rate, we can expect much higher PSNR outcomes.

Table 4 shows relative computational loads of the conventional algorithms with respect to that of the proposed algorithm to evaluate computational complexity. The relative complexity ratio is defined by

$$\frac{C_P - C_C}{C_P} \times 100 \quad (6)$$

where C_P is the computation time of the proposed method and C_C is the computation time of a conventional algorithm. The proposed algorithm requires more computational time than the conventional algorithms, because the proposed method sequentially performs motion estimation in descending order of block variances. To speed up the proposed algorithm, a frame can be divided into multiple slices. Then, each slice can be performed in parallel with some loss of estimation accuracy.

However, as the proposed method refines MVs based on reliability, more effective motion estimation can be performed. The search range and order of block processing depend on block variances and are adaptively adjusted. For the higher block variance, the more accurate motion estimation is performed. In addition, we use spatial linear interpolation in the occlusion region, so we can compensate occlusion regions without blocking

artifacts. This leads to improved perceptual quality and PSNR, especially at edge regions.

5. Conclusions

In this article, we proposed a new adaptive FRU method based on confidence levels. In addition, we employed the HME with the adaptively modified search range according to the confidence level. With the proposed algorithm, the accurate MVs influence neighboring vector fields and we can estimate consistent MV field to generate accurate interpolated frames. Experimental results show that the proposed algorithm is better than several conventional methods in not only objective but also perceptual quality. However, the proposed method requires high computational complexity due to the sequential confidence-based ME. Further work can focus on reduction of computational complexity of the FRU algorithms for real-time processing.

Competing interests

The authors declare that they have no competing interests.

Acknowledgement

This work was partly supported by the IT R&D program of MKE/KEIT [10039199, A Study on Core Technologies of Perceptual Quality based Scalable 3D Video Codecs], the grant from the Seoul R&BD Program (SS110004M0229111), and the Research Grant of Kwangwoon University in 2012.

Received: 4 January 2012 Accepted: 15 January 2013

Published: 6 February 2013

References

1. R Castagno, P Haavisto, G Ramponi, A method for motion adaptive frame rate up-conversion. *IEEE Trans. Circuits Syst. Video Technol.* **6**(5), 436–446 (1996)
2. H Sasai, S Kondo, S Kadono, Frame-rate up-conversion using reliable analysis of transmitted motion information, in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (, Osaka, Japan, vol. 5, May 2004), pp. 257–260
3. G-I Lee, B-W Jeon, R-H Park, S-H Lee, Hierarchical motion compensated frame rate up-conversion based on the Gaussian/Laplacian pyramid, in *IEEE International Conference on Consumer Electronics* (Los Angeles, CA, June 2003), pp. 350–351
4. S Fujiwara, A Taguchi, Motion-compensated frame rate up-conversion based on block matching algorithm with multi-size blocks, in *International Symposium on Intelligent Signal Processing and Communication Systems* (Tokyo, Japan, Dec. 2005), pp. 353–356
5. Y-K Chen, A Vetro, H Sun, SY Kung, Frame-rate up-conversion using transmitted true motion vectors, in *Proceedings of the 2nd IEEE International Workshop on Multimedia Signal Processing* (Redondo Beach, CA, Dec. 1998), pp. 622–627
6. SJ Kang, DG Yoo, SK Lee, Y Kim, Multiframe-based bilateral motion estimation with emphasis on stationary caption processing for frame rate up-conversion. *IEEE Trans. Consum. Electron.* **5**(4), 1830–1838 (2008)
7. Y-T Yang, Y-S Tung, J-L Wu, Quality enhancement of frame rate up-converted video by adaptive frame skip and reliable motion extraction. *IEEE Trans. Circuits Syst. Video Technol.* **17**(12), 1700–1713 (2007)
8. C Cafforio, F Rocca, S Tubaro, Motion compensated image interpolation. *IEEE Trans. Commun.* **35**(2), 215–222 (1990)
9. G-I Lee, B-W Jeon, R-H Park, S-H Lee, Frame rate up-conversion using the wavelet transform, in *International Conference on Consumer Electronics* (Los Angeles, CA, June 2000), pp. 172–173
10. T-H Tsai, H-Y Lin, High visual quality particle based frame rate up conversion with acceleration assisted motion trajectory calibration. *J. Disp. Technol.* **8**, 6 (2012)

11. D Wang, L Zhang, A Vincent, Motion-compensated frame rate up-conversion—Part I: fast multi-frame motion estimation. *IEEE Trans. Broadcast.* **56**, 2 (2010)
12. Y Zhang, D Zhao, S Ma, R Wang, W Gao, A motion-aligned auto-regressive model for frame rate up conversion. *IEEE Trans. Image Process.* **19**, 5 (2010)
13. C Wang, L Zhang, Y He, Y-P Tan, Frame rate up-conversion using trilateral filtering. *IEEE Trans. Circuits Syst. Video Technol.* **20**(6), 886–893 (2010)
14. B-D Choi, S-J Ko, J-W Han, C-S Kim, Frame rate up-conversion using perspective trans. *IEEE Trans. Consum. Electron.* **52**(3), 975–982 (2006)
15. J Zhai, K Yu, J Li, S Li, A low complexity motion compensated frame interpolation method, in *IEEE International Symposium on Circuits and Systems* (Kobe, Japan, May 2005), pp. 4927–4930
16. D Alfonso, D Bagni, D Moglia, Bi-directionally motion-compensated frame-rate up-conversion for H.264/AVC decoder, in *47th International Symposium EL/MAR* (Zadar, Croatia, June 2005), pp. 41–44
17. J Shi, J Malik, Normalized cuts and image segmentation. *IEEE Trans. Pattern. Anal. Mach. Intell.* **22**(8), 888–905 (2000)
18. S-H Lee, S Yang, Y-Y Jung, R-H Park, Adaptive motion-compensated interpolation for frame rate up-conversion, in *IEEE International Conference on Consumer Electronics* (Los Angeles, CA, June 2002), pp. 68–69
19. T Chen, Adaptive temporal interpolation using bidirectional motion estimation and compensation, in *IEEE International Conference of Image Processing* (Princeton, NJ, vol. 2, Sept. 2002), pp. 313–316
20. T Thaipanich, PH Wu, CC Kuo, Low complexity algorithm for robust video frame rate up-conversion technique. *IEEE Trans. Consum. Electron.* **55**(1), 220–228 (2009)
21. KY Min, SN Park, JH Nam, DG Sim, SH Kim, Distributed video coding based on adaptive block quantization using received motion vectors. *KICS J.* **35**(2), 172–181 (2010)
22. YL Lee, T Nguyen, High frame rate motion compensated frame interpolation in high-definition video processing, in *ICASSP 2010* (Dallas, TX, Mar. 2010), pp. 858–861
23. G Haan, P Biezen, H Huijgen, OA Ojo, True-motion estimation with 3-D recursive search block matching. *IEEE Trans. Circuits Syst. Video Technol.* **3**(5), 368–379 (1993)
24. N Damera-Venkata, T Kite, W Geisler, B Evans, A Bovik, Image quality assessment based on a degradation model. *IEEE Trans. Image Process.* **9**(4), 636–650 (2000)
25. H Sheikh, A Bovik, G de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **14**(12), 2117–2128 (2005)
26. Z Wang, A Bovik, H Sheikh, E Simoncelli, Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
27. H Sheikh, A Bovik, Image information and visual quality. *IEEE Trans. Image Process.* **15**(2), 430–444 (2006)
28. Z Wang, A Bovik, A universal image quality index. *IEEE Signal Process. Lett.* **9**, 81–84 (2002)

doi:10.1186/1687-6180-2013-13

Cite this article as: Min and Sim: Confidence-based adaptive frame rate up-conversion. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:13.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
