# Scrolling Subtitle Processing of Frame Rate Up-Conversion Based on Global Text Motion Estimation

Guirui Li$^{(\boxtimes)}$, Xiaoyun Zhang, and Zhiyong Gao

Institute of Image Communication and Network Engineering,
Shanghai Jiao Tong University, Shanghai 200240, China
{lgrsjtu,xiaoyun.zhang,zhiyong.gao}@sjtu.edu.cn

**Abstract.** Motion compensated frame rate up-conversion (MC-FRUC) is a popular practical video processing technique to increase frame rate, which mainly composes of motion estimation (ME) and motion compensated frame interpolation (MCFI). However, for scrolling text embedded in videos, such as subtitle or scrolling TV news, ME usually fails to produce true trajectory when the scrolling motion of subtitle is irrelevance with background's movements. In this paper, we propose an effective scrolling subtitle post-processing technique to reduce artifacts such as text broken and motion blur around subtitle regions. True global text motion estimation is put forward to predict the motion vector of subtitle after detecting the text area in the intermediate frame. Finally, a fusion of subtitle and background method is proposed to obtain a correct intermediate frame. Experimental results show that the proposed algorithm can significantly reduce the broken artifact with low complexity.

**Keywords:** Frame rate up-conversion (FRUC) · Scrolling subtitle · Global text motion estimation · Motion compensation

## 1 Introduction

With the development of high definition (HD) video service and the progress in display technologies, high frame rate becomes a key element for high quality displaying. However, videos may be encoded at a low frame rate because of the limited transmission bandwidth, and are needed to be converted to a higher frame rate before displaying. Frame rate up-conversion (FRUC) is widely used for increasing frame rate by inserting new frames into the original moving sequence.

There are some simple approaches to achieve FRUC, such as frame repetition (FR) and frame linear interpolation (FA), but they do not utilize the motion of objects and cause artifacts like motion blur. Motion compensated frame rate up conversion (MC-FRUC) exploits the motion information between the previous and next reference frames to improve the accuracy of motion trajectories and achieve a higher video quality. A typical MC-FRUC algorithm generally consists of two stages: Motion estimation (ME) and Motion interpolation (MCFI). In the

ME process, the motion vectors (MV) are determined by using the unilateral or bilateral block matching (BM) methods. Then MCFI uses the estimated MVs to construct a new intermediate frame. Nowadays, many algorithms based on MC-FRUC [1–5] have been proposed. For example, Haan et al. [1] addressed to improve the accuracy of MV, blocking artifact has been reduced in [2,3], and halo effect at occlusion areas has been studied in [4,5]. Besides the problems mentioned above, scrolling subtitle is also a critical issue for high subjective quality.

In practical applications, subtitle is the text added to the videos or TV programs in post-production to display the content of the voice messages in caption format. These text information aims at helping the viewers understand the messages delivered by the video scenarios. For example, foreign language can be displayed in subtitle form after being transmitted. However, from the perspective of FRUC, there is little motion correlation between the subtitle and the original video content because the subtitle is added by post-production. So the text blocks's MVs obtained by ME based on spatial and temporal correlation are likely to be incorrect. And the intermediate frames will be interpolated with subtitle broken or motion blur.

Several subtitle processing methods for FRUC have been proposed to improve visual quality in text areas. In [6,7], stationary subtitle correction algorithms are proposed. In [6], the motion vectors of the text blocks are changed to be zero. Although it fixes motion vectors of the blocks with stationary text, severe artifacts may occur in moving background. A pre-post processing algorithm with stationary subtitle detection and in-painting is proposed in [7], which erases the stationary subtitle to do ME and MCFI. Then the intermediate frame is overlaid with subtitle at last. A horizontal scrolling text processing method proposed in [8] utilized the text detection method but it neglected the complex moving background.

In this paper, a low complexity post-processing algorithm is proposed to deal with the scrolling subtitle problem with complex motion background. Firstly, the scrolling text location in the intermediate frame will be detected. After that, true global text motion estimation will be proposed to predict subtitle's motion trajectory. Finally, the final intermediate frames are obtained by combining the initial interpolated background with the complete subtitle.

This paper is organized as follows. In Sect. 2, the motion estimation in the subtitle area is investigated and analyzed. And the proposed algorithm for FRUC with subtitle processing is presented in Sect. 3. Then Sect. 4 shows the experimental results and assesses the performance of the proposed algorithm using subjective and objective evaluations. Finally, the conclusion is drawn in Sect. 5.

## 2 Analysis of the Motion Estimation

Motion estimation (ME) is an extremely important process to approximate the true motions between consecutive frames. Bilateral or unilateral block matching methods are always conducted in ME stage because of the low computational complexity.

In our ME process, motion vectors for intermediate frames are predicted through bilateral motion estimation (BME) [9] which utilizes the spatial and temporal information. Specifically, the motion vector is obtained by comparing the matching blocks in previous and next reference frames. And it is important to note that a block's MV is initially estimated by searching for neighboring blocks' MV as its candidates. We compute the sum of bilateral absolute differences (SBAD) to measure the reliability of the motion vectors. The candidate that is closer to the true MV usually have smaller SBAD in probability.

Let $B_{i,j}$ denote a block in the intermediate frame and $m$ is a pixel in that block, $f_n$ and $f_{n+1}$ are the two reference frames. For the candidate motion vector $v$, SBAD is calculated as:

$$SBAD(B_{i,j}, v) = \sum_{m \in B_{i,j}} |f_n(m - v) - f_{n+1}(m + v)| \qquad (1)$$

BME can provide acceptable motion vectors when the video sequence contain simple motions. However, it does not trace reliable motion trajectories in the scrolling subtitle areas where blocks with text or without text both exist. Generally, subtitle's scrolling orientation is little correlated with background's moving direction as the subtitle is added by post-production. The candidate MV set for a text block usually contains background's MV. And the BME process usually results in incorrect MVs which do not track the true scrolling trajectory for subtitle.
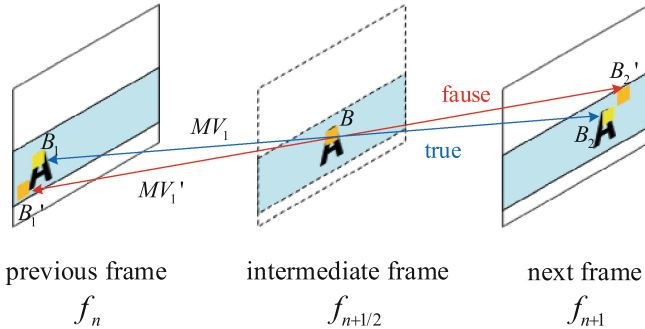


**Fig. 1.** Bilateral motion estimation for a subtitle block

As illustrated in Fig. 1, $B$ is a text block in intermediate frame, $MV_1'$ and $MV_1$ are the candidate MVs, $B_1'(B_1)$ and $B_2'(B_2)$ are the corresponding matching blocks in previous frame and next frame respectively. $MV_1'$ will be determined as the optimal MV for block $B$ when the SBAD value of $MV_1'$ is the smallest. In that case, the true $MV_1$ will not be chosen. So it is difficult to track the true scrolling trajectory of subtitle even if the correct MV is included in the candidate set. Consequently, the subtitle will be interpolated broken. For
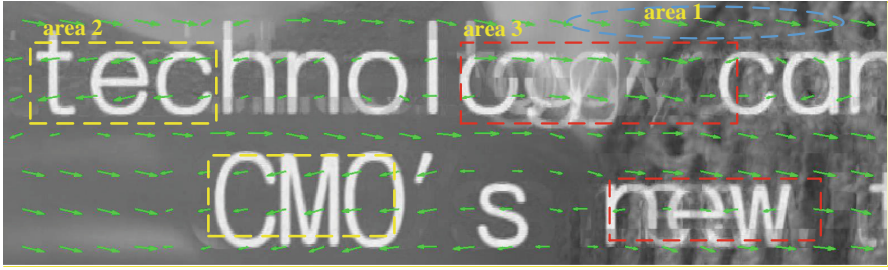
**Fig. 2.** Subtitle interpolation with traditional FRUC algorithm

example, a portion of the letter "A" was interpolated with background's pixels and resulting in text broken.

An interpolation result for subtitle areas with the estimated MVs is presented in Fig. 2. In this video, the background is moving right and the subtitle is scrolling horizontally from right to left. So the scrolling motion of subtitle is weakly correlated with the movement of background. The broken text is interpolated with bilateral ME and MCFI which did not consider the problem of scrolling subtitle. The MVs in "area 1" denote the background's motion. Text in "area 2" is complete with correct MVs and is broken in "area 3" with inaccurate motions. It is pretty clear from the vector plots in "area 3" that the motion vectors estimated for these text blocks are misled by background's MVs in "area 1". In summary, motion vectors estimated with traditional BME methods could not correctly approximate the true motions of the text.

## 3   The Proposed Algorithm

The proposed FRUC algorithm is dedicated to improve the performance of the motion estimation and the interpolation in the subtitle areas. Figure 3 shows the proposed FRUC system. Firstly, using the information in previous and next reference frame, the bilateral motion estimation method is performed to predict the motion vectors for the intermediate frame. In this process, the MVs for text blocks may be wrongly estimated. So the true global text motion estimation is proposed to predict the motion trajectory of the text after detecting the subtitle in the intermediate frame. Finally, the proposed motion compensation is executed by a two-stage interpolation in the subtitle area. The subtitle and background are interpolated respectively to minimize the interference with each other. And the final intermediate frames are the background interpolation results overlaid with the complete subtitle.

Compared with the traditional FRUC algorithms, the proposed algorithm employs the global text motion estimation and two-pass compensation techniques. The global text motion vector extracted for the moving subtitle improves the reliability of the BME method. And thus leads to better subtitle interpolation.
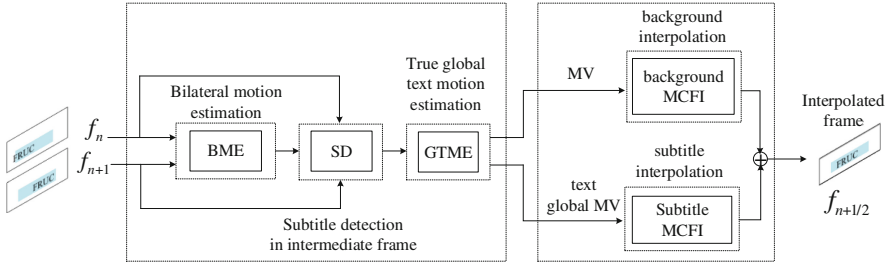
**Fig. 3.** The proposed FRUC system

### 3.1 Subtitle area detection

For videos with scrolling subtitle, we could not locate the text blocks directly within the intermediate frame since this frame does not exist originally. So the traditional text detection techniques, such as texture-based and edge-based techniques, could not be applied to identify text blocks. In our algorithm, motion irrelevance and subtitle's features are combined to locate subtitle area in the intermediate frame.

Text blocks and background blocks are likely spatial neighborhood. So the candidate MV set initialized for a text block usually contains the background's MV. But the motion of the text is irrelevance with background's movement. Besides, text blocks probably contain more complex texture than background blocks. Thus larger SBAD values are produced in these areas. And it is easy to mark some text blocks whose spatial candidate MVs and the final determined MV have larger SBAD values than a threshold. Therefore, we can obtain the initial marking area which contains the incorrect text blocks in intermediate frames.

In addition, scrolling features of subtitle can be adopted to obtain a more accurate location. First of all, it is assumed that the luminance value of subtitle is unified and visually recognizable from the background. Moreover, the scrolling subtitle generally appears in consecutive frames with the location area remaining unchanged. If a block is marked in the intermediate frame in a block row $X$, we will check the blocks of the same row in previous and next reference frames. The row $X$ will be identified as the subtitle area when it contains text blocks in both two reference frames. And this block row $X$ is detected as subtitle area in the intermediate frame.

### 3.2 True global text motion estimation

As shown in Fig. 2, the pixels of the text area may be well interpolated or badly interpolated by traditional BME algorithm. Although not all the motion vectors for text area are correct, there still exist true motion vectors among them. So it is feasible to extract the true MVs for text blocks in the text area.

Moreover, subtitle generally appears in consecutive frames with scrolling. It should also be noted that the scrolling motion of subtitle is global. So a global motion vector could be estimated for all of the text blocks. And there is only one global motion which can represent the true scrolling movement. We refer to this global motion as GTMV (global text motion vector) in our algorithm.

ME results for text blocks in Fig. 2 are exploited to conduct a statistical analysis shown in Fig. 4. In the horizontal scrolling subtitle applications, the vertical component of MVs for text blocks approximate zero. So we performed the statistics with horizontal component of MVs.
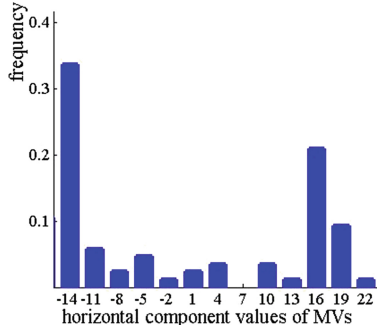


**Fig. 4.** Frequency statistics of text blocks MVs

In Fig. 4, the x-coordinate describes the horizontal component of MVs while the y-coordinate is the corresponding frequency for each MV. After the statistics, the top $n$ MVs with highest frequency are obtained. And these MVs can constitute a candidate set for GTMV, denoted as $GMV_i (i = 1, 2, ...n)$. For a text block in the intermediate frame, its matched blocks in previous and next frames should also be text blocks. Besides, the SBAD can be calculated to measure the reliability of the motion vectors. If MV is closer to the true motion, the SBAD will have smaller value. Therefore, for each candidate $GMV_i$, the reliability $R(i)$ is calculated to measure which candidate is the global motion of text. And the GTMV is determined by minimizing $R(i)$.

$$R(i) = \frac{SBAD(i)}{C(i) + \varepsilon} \ (i = 1, 2, ...n) \tag{2}$$

$$C(i) = \frac{M_{i,p} + M_{i,n}}{2 * N * N} \ (i = 1, 2, ...n) \tag{3}$$

where $i = 1, 2, ...n$ represents that there are $n$ candidates for GTMV. $N * N$ is the total text pixel number in a block. $M_{i,p}$ and $M_{i,n}$ are the text pixel number in the block of previous frame and next frame respectively with candidate $GMV_i$. $C(i)$ is computed to illustrate the confidence of whether the matching block is text or not. $SBAD(i)$ denotes the SBAD value for the candidate $GMV_i$. And $\varepsilon$ is a smoothing parameter to avoid singular value of $C(i)$.

### 3.3   MCFI in Subtitle Area

In subtitle area, there is no guarantee whether a block-based FRUC algorithm could perfectly divides the text and background pixels into different blocks. For a text block in which the non-text pixels account for a large proportion, the movement of the non-text part is inconsistent with the motion of scrolling text. Hence, the text block may be interpolated partly broken because the GTMV is accepted as its optimal MV. Therefore the incomplete separation decreases the visual quality of interpolated frames.

As we mentioned above, it is difficult to construct complete subtitle and background in a single-pass MCFI with block-based ME and MC. In our proposed system, we conduct a two-pass compensation for the subtitle pixels and the background respectively.

The proposed global text motion estimation has improved the accuracy of the motion vectors for text blocks. So complete subtitle pixels should be interpolated with GTMV. We utilize the bilateral motion compensation to generate intermediate frame which is determined as follows:

$$f_{n+1/2}(m) = \frac{1}{2}[f_n(m - v) + f_{n+1}(m + v)] \tag{4}$$

where $f_n$ and $f_{n+1}$ are the previous and next reference frames, $f_{n+1/2}$ is the intermediate frame, $v$ represent the MV used to interpolate new frames.

Using the previous and next reference frames, the motion compensation for subtitle pixels can be conducted subsequently with the true global text motion vectors. And the integrity of subtitle can be well guaranteed.

The incomplete separation between the text and the background may lead to broken background even if text pixels have been well interpolated with GTMV. So a special motion compensation for the background around subtitle is particularly essential. Because the background is covered by text pixels, it should be reconstructed with reference frames whose subtitle is removed. The holes left by removed text can be replaced with the spatial information provided by pixels around subtitle. In order to avoid a blurry in-painting result, we utilize the directional interpolation method proposed in [10]. The information on local edges and textures will be extracted within blocks and then used for filling the text pixels.

After subtitle removing and in-painting, the subtitle pixels' damage on background can be effectively eliminated. Then the background can be interpolated with the regular MCFI as shown in (4).

The final intermediate frames are the background interpolation results overlaid with the complete subtitle. Therefore, the proposed interpolation method has two advantages which come from its complete background and the appropriate subtitle attributes.

## 4   Experiment

In order to demonstrate the performance of our method, we use the HD($1920 * 1080$) test sequences with scrolling subtitle and moving background.

The experiment results are examined in comparison with the traditional FRUC method which does not consider the scrolling subtitle problem.

One of the compared results with and without scrolling subtitle processing algorithm is depicted in Fig. 5. The sequence "CMO" contains scrolling subtitle and moving background. The two motions are inconsistent with each other.
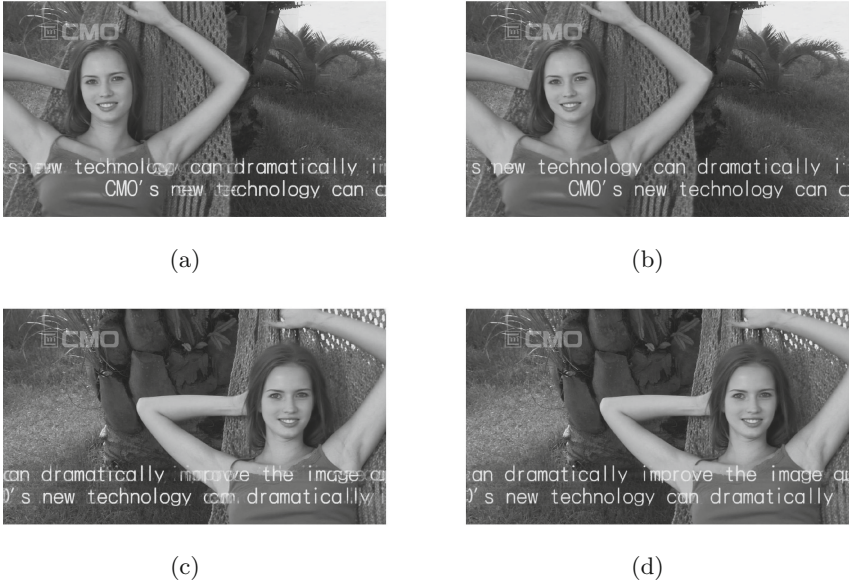


(a)                                        (b)

(c)                                        (d)

**Fig. 5.** FRUC results with and without subtitle processing. The interpolated frames with traditional FRUC method are shown in left column and improved ones using the proposed algorithm are shown in the right. (a) (PSNR = 27.94) and (b) (PSNR = 28.75) are reconstructed frames of $28^{th}$. (c) (PSNR = 27.65) and (d) (PSNR = 28.93) are reconstructed frames of $58^{th}$.

As shown in Fig. 5(a) and (c), the interpolated subtitle has a poor visual quality. Some of the texts, such as "technology" and "can", are broken since they are interpolated with background's pixels rather than the text pixels. The subtitle information has been damaged severely which affects the visual experience. In Fig. 5(b) and (d), the proposed algorithm achieves considerable improvement on the completeness of subtitle. Besides, it also demonstrates that the proposed algorithm provides a better quality objectively by an average 1 dB gain in PSNR.

Figure 6 presents the text processing results for some letters. Letters "improve" and "technology" are partly broken or fuzzy when using traditional FRUC. With the proposed subtitle processing method, letters are interpolated completely and clearly.
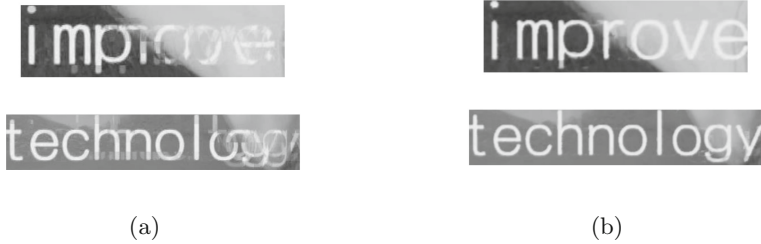
(a) (b)

**Fig. 6.** Letters correction results

## 5 Conclusion

In this paper, we proposed a practical method with low complexity to solve the broken artifacts in scrolling subtitle regions with moving background. By taking sufficient statistical analysis of the ME results and the discriminative features of subtitle, the quality of the interpolated frames is improved subjectively and objectively. Moreover, it is compatible to the current block-based FRUC architecture.

## References

1. de Haan, P.W.G., Biezen, H.H., Ojo, O.A.: True-motion estimation with 3-D recursive search block matching. IEEE Trans. Circ. Syst. Video Technol. **3**(5), 368–379 (1993)
2. Choi, B.D., Han, J.W., Kim, C.S.K., Ko, S.J.: Overlapped block motion compensation: an estimation-theoretic approach. IEEE Trans. Image Process. **3**(9), 693–699 (1994)
3. Choi, B.D., Han, J.W., Kim, C.S., Ko, S.J.: Motion compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation. IEEE Trans. Circ. Syst. Video Technol. **17**(4), 407–416 (2007)
4. Bellers, E.B., van Gurp, J.W., Janssen, J.G.W.M., Braspenning, J.R., Wittebrood, R.: Solving occlusion in frame-rate up-conversion. In: Proceedings of IEEE International Conference on Consumer Electronics, pp. 1–2 (2007)
5. Hong, W.: Low-complexity occlusion handling for motion-compensated frame rate up-conversion. In: Proceedings of IEEE International Conference on Consumer Electronics, pp. 1–2 (2009)
6. Kang, S.J., Yoo, D.G., Lee, S.K., Kim, Y.H.: Multiframe-based bilateral motion estimation with emphasis on stationary caption processing for frame rate up-conversion. IEEE Trans. Consum. Electron. **54**(4), 1830–1838 (2008)
7. Guo, Z.Y., Gao, L.C.: Stationary subtitle processing for real-time frame rate up-conversion. In: Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–4 (2013)

8. Gim, G.Y., Kim, Y.J., Ahn, T.G., Park, S.H.: Horizontal scrolling text processing for frame rate conversion. In: Proceedings of IEEE International Conference on Consumer Electronics-Berlin (2012)
9. Choi, B.T., Le, S.H., Ko, S.J.: New frame rate up-conversion using bi-directional motion estimation. IEEE Trans. Consum. Electron. **46**(3), 603–609 (2000)
10. Wang, A.D., Vincent, P.B., Klepko, R.: Motion-compensated frame rate conversionpart II: new algorithms for frame interpolation. IEEE Trans. Broadcast. **56**, 142–149 (2010)