Neurocomputing 181 (2016) 76–85

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Motion-compensated frame interpolation with weighted motion estimation and hierarchical vector refinement

Dan Guo *, Zhihong Lu

School of Computer and Information, Hefei University of Technology, Hefei, China

ARTICLE INFO

Article history:
Received 6 January 2015
Received in revised form
6 June 2015
Accepted 18 June 2015
Available online 14 December 2015

Keywords:
Weighted motion estimation
Motion vector clustering
Hierarchical motion vector refinement
Motion-compensated frame interpolation

ABSTRACT

Large amounts of data have been gathered by techniques such as mobile devices, remote sensing and cameras. With the rapid development of high-definition digital TV and multimedia information systems, motion-compensated frame interpolation (MCFI) has become a widely used tactic for frame rate up-conversion (FRUC) to improve the visual property of video. In this paper, we propose a novel MCFI algorithm based on weighted motion estimation (Weighted ME), motion vector (MV) segmentation, and hierarchical vector refinement. We first present an MV clustering based on the Weighted ME to obtain more accurate MVs and partition a frame into moving areas and background. To capture scene changes, we then apply a hierarchical vector refinement scheme to the moving areas that consists of three steps: pre-screening, reclassification, and smoothing. In this scheme, we use an overlapped block ME (OBME) method that uses multi-candidate pre-screening to discover unreliable MVs and protect the moving area edge structures from being damaged. We then employ the bidirectional prediction difference (BPD) to identify outliers using strong correlation with adjacent blocks. Meanwhile, an adaptive vector median filter (AVMF) is adopted to refine the block size, which can effectively smooth blocking artifacts and ghost effects. Experiments show the algorithm achieves a high image quality both subjectively and objectively. In particular, it has a good adaptability in video sequences with fast motion and complex backgrounds.

1. Introduction

Ininformation technology, people often try to find specialized content using on-hand big data, such as in web mining, multimedia retrieval [1,2], video content analysis [3–11] or intelligent transportation research [12–14]. Large amounts of data have been gathered by techniques such as mobile devices, remote sensing and cameras. With the rapid development of high-definition digital TV and multimedia information systems in the past few decades, massive numbers of videos are now available that have been gathered by mobile devices and cameras and need improved display quality. How to improve the frame rate of videos to achieve high visual quality has aroused widespread attention. In recent years, frame rate up-conversation (FRUC) as a post-processing tool to convert between two display formats with different frame rates has become a research focus. Various FRUC algorithms have been developed, such as frame repetition and temporal frame averaging [15–18]. These methods can provide an acceptable image quality, but they often lead to annoying jerky and blurry moving object artifacts for fast motion or complex scenes. Motion-compensated frame interpolation (MCFI) is another widely used tactic. MCFI methods predict the intermediate frame using either linear or nonlinear interpolators to adapt to scene changes; hence they can achieve a better image quality than FRUC algorithms. However, the main problem with MCFI is its failure to represent true motion. By considering only the motion distribution on object boundaries, it is very challenging to produce high quality, artifact free interpolated frames.

To address this problem, many algorithms that are performed at the decoder have been proposed. Motion estimation (ME) and motion compensation (MC) are the two key steps of MCFI. Most MCFI methods use the block-matching algorithm (BMA) in ME, which is called block-based ME. This method splits the current frame into several $N \times N$ sized blocks and assumes that all pixels in a block have the same motion vector (MV), hence it has low complexity and is easy to implement with respect to pixel-wise ME [19]. However, the MVs generated by block-based ME at the encoder are usually obtained by minimizing residual energy rather than finding the true motion, so directly using the received MVs at the decoder often introduces annoying artifacts such as ghost effects and blocking artifacts. Many ME algorithms have consequently been proposed to obtain a more accurate MV field, such as overlapped block ME [20], bilateral ME [21–23], recursive search

* Correspondence to: School of Computer and Information, Hefei University of Technology, 193 Tunxi Road, Hefei, Anhui 230009, China.
E-mail addresses: guodan@hfut.edu.cn (D. Guo), luzhihong5842@gmail.com (Z. Lu).

http://dx.doi.org/10.1016/j.neucom.2015.06.102
0925-2312/

[24], and three-dimensional recursive search [25]. Although these methods sometimes can enhance the accuracy of ME, they cannot always reflect the true motion trajectory.

In addition, some exhaustive motion re-estimation methods have proposed to obtain true MVs, which may prevent their use on mobile devices because of their higher complexity. Many MV processing methods have been proposed to reduce the computational complexity in the field of real-time video applications. Dane et al. [26] presented a smoothing method to reduce blocking artifacts, but the method introduces more ghost effects. In addition, filters have been widely applied to obtain a smoother MV field, such as the vector median filter [27], adaptive vector median filter [28], and trilateral filter [29]. These methods have the advantage of removing outliers, but they may amend the MVs excessively and even lead to abnormal MVs.

MC is another important step of MCFI. Approaches to interpolate the intermediate frames to reduce the artifacts have also been proposed. Overlapped block motion compensation (OBMC) [30] was widely applied to reduce blocking artifacts by using adjacent blocks. When OBMC is applied to all blocks uniformly, the quality of the interpolated frame may be degraded because of over-smoothed edges. Cao et al. [18] used adaptive-weighted motion compensated interpolation to construct the intermediate frame, but it is not able to handle fast motion. Moreover, some approaches use spatial transform techniques to avoid blocking artifacts, such as control grid interpolation [31–33]. However, these techniques assume that the whole frame sequence has a consistently globally smooth motion field model, which usually does not happen in practice. These techniques also require much higher computational complexity. The work in [34] reconstructed intermediate frames using an affine transformation based on SIFT features. However, it does not work well on frames with repetitive texture. Therefore, these methods can also cause distortion or other artifacts, and how to achieve a higher image quality using MCFI remains a challenging problem.

In this work, we focus on the detection of moving objects based on MV processing. We correct unreliable MVs to solve the occlusion issues so that occluded areas or deformed objects can be discovered effectively. Fig. 1 shows the overall process of our algorithm. First, to improve the accuracy of the MV field, we propose a Weighted ME that sums different measurement criteria for unidirectional ME. Second, MV clustering is used to partition a frame into moving areas and background. The MVs in the moving areas are usually unreliable and likely to produce visual artifacts. Next, to make these MVs reliable, we present a hierarchical MV processing method to amend them. The method is a gradual refinement process that includes three steps: pre-screening, reclassification, and smoothing. We develop the overlapped block ME (OBME) method using multi-candidate pre-screening to recalculate MVs. This can also protect the moving areas' edge structure from being damaged by the block segmentation. After recalculation, some unreliable MVs become reliable. We employ the bidirectional prediction difference (BPD) to identify these, and outliers with low BPD can be detected by strong correlation with adjacent blocks. This approach is also used in the smoothing step based on an adaptive vector median filter (AVMF) to correct the remaining unreliable MVs. Finally, the bidirectional frame interpolation is used to compensate for the motion. Because we gradually reduce the numbers of unreliable MVs instead of dealing with all the MVs at once, our method has low complexity.

The main contributions of this paper are two-fold: 1) a Weighted ME is proposed to obtain the true MV, and 2) we propose a low-complexity and hierarchical optimization MCFI method based on MV segmentation and refinement processing. As a result, the interpolated results can remove most of the ghost artifacts and obtain clearer object contours. Experiments show the algorithm not only achieves a high image quality both subjectively and objectively, but also better adapts to video sequences with fast motion and complex background than other methods.

The remainder of the paper is organized as follows. A discussion of ME methods is given in Section 2. We then briefly describe MV clustering in Section 3. Section 4 details the MV hierarchical refinement scheme, including MV pre-screening, MV reclassification, and MV smoothing. Section 5 introduces bidirectional MC. Experimental results are presented in Section 6. Section 7 summarizes the paper.

2. ME

Conventional ME methods use the BMA, which usually minimizes the residual energy rather than finds the true motion; hence the received MVs cannot always reflect the motion trajectory of objects. To enhance the accuracy of the received MVs, in this section we analyze the disadvantages of two conventional ME methods and propose a weighted sum as a measurement criterion for unidirectional ME.

2.1. Bilateral ME

The bilateral ME method [21–23] was recently proposed. It splits the to-be-interpolated frame into several N × N sized blocks and determines the MV of a block by assuming that the position shift in the previous frame and following frame are symmetric. Let f_{t-1} and f_{t+1} denote the previous frame and the following frame, respectively, and let f_t represent the interpolated frame. For each block B in f_t, suppose v is the position shift forward to f_{t+1}, and that of f_{t-1} is the opposite. Thus, we can determine the MV by minimizing the sum of bidirectional absolute difference (SBAD), which is simply defined as:

$$\mathrm{SBAD}(v) = \sum_{i,j \in B} |f_{t-1}(i-v_x, j-v_y) - f_{t+1}(i+v_x, j+v_y)|, \quad (1)$$

where $v=(v_x, v_y)$ and $f(i,j)$ are the MV and pixel of location (i,j), respectively.

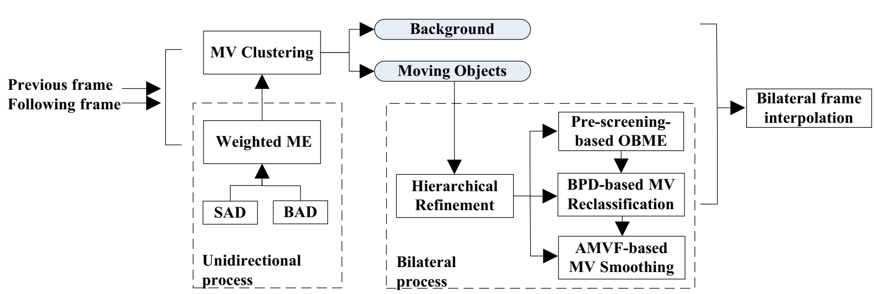

Fig. 1. Overview of the proposed algorithm.