IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 October 2022, accepted 7 November 2022, date of publication 16 November 2022, date of current version 22 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3222525

METHODS

Luminance Compensation MEMC for Video Frame Interpolation

ZIXUAN XU[1,2], WENJING YING[2,3], HAO HE[2], QINGMENG ZHU[2], JIAN LIANG[2], AND HAIHUI WANG[1]

[1]School of Computer Science and Engineering (School of Artificial Intelligence), Wuhan Institute of Technology, Wuhan 430205, China
[2]Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
[3]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Corresponding author: Hao He (hehao21@iscas.ac.cn)

This work was supported in part by the National Natural Science Foundation (NSF) of China under Grant 62101552, and in part by the Key Deployment Program of the Chinese Academy of Sciences under Grant ZDRW-XH-2021-3-03.

ABSTRACT Video frame interpolation is an important technology in digital video processing, which has great impact on users' viewing experience. In particular, in medical or industrial application scenarios, the accuracy of the frame interpolation algorithm may also influence the diagnosis results. In addition, for videos based on ionizing radiation (e.g., X-rays), each frame exposure could cause damage to human tissues by ionizing radiation. Therefore, if a frame interpolation algorithm is introduced to display the same number of frames, it only needs to sample half of the frames and halve the exposure radiation doses, which is statistically promising to reduce human cancer rate caused by ionizing radiations (e.g., medical examinations). However, since there are errors in frame interpolation caused by luminance leap, existing works are not applicable in such scenarios. To solve this problem, this paper proposes a video interpolation algorithm based on luminance compensation MEMC (LC-MEMC). Firstly, a luminance compensation method based on the electromagnetic irradiation attenuation in human tissue is introduced to improve the performance of motion estimation and motion compensation (MEMC) and reduce matching errors caused by luminance leap. Secondly, LC-MEMC proposes an improved block matching approach, including i) a new search method from basic points to local points and ii) a block matching criterion that simplifies the calculation process. LC-MEMC improves the accuracy and processing speed of video interpolation from three perspectives: adding luminance compensation, improving the search strategy and optimizing the matching degree calculation method for each search position. We evaluated LC-MEMC on collected medical videos and achieved higher accuracy, faster processing speed, and significantly better viewing experience comparing with existing methods.

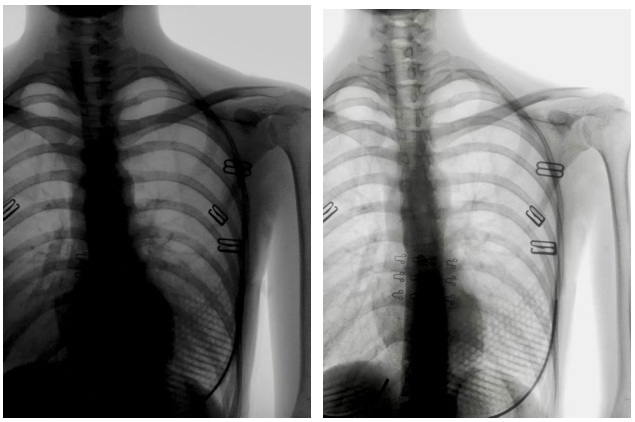
INDEX TERMS Video interpolation, luminance compensation, medical video, X-ray, MEMC.

I. INTRODUCTION

Advanced display equipment (e.g., projectors, flat panel TVs, etc.) usually support frame frequency doubling, which inserts intermediate frames between consecutive frames, to improve the users' viewing experience. Although some deep learning based approaches claim better performance on certain datasets, the shortage of training data and high cost of computation have limited the practical applications of such algorithms. As a matter of fact, the dominant video frame interpolation methods are still based on motion estimation and motion compensation (MEMC) technique which relies on none extra training data. However, in spite of success in common scenarios, MEMC are not applicable in some special and important areas, such as medical or industrial diagnosis applications due to the luminance leap in the videos, as shown in Figure 1.

Specifically, the acquisition of biomedical or industrial video involves X-ray exposure which changes based on the target composition and thickness, leading to high error rate in MEMC based algorithms. On the other hand, such ionizing radiation exposure can be hazardous to human health

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.



(a) Low luminance (b) High luminance

FIGURE 1. The luminance leap between frames in a video.

(which is accumulated over time and doses). A single full-body CT (electron computed tomography by X-ray) examination in a 45-year-old adult would result in an estimated lifetime attributable cancer mortality risk of around 0.08% [1]. It has been estimated that up to 2% of all cancers in the United States may be attributable to the radiation from CT scans [2]. In order to reduce the amount of radiation received by the patient, the simplest approach is to reduce the sampling rate and decrease the exposure time, which can result in the loss of some valuable information as well as larger layer spacing [3]. However, the reduction in sampling rate (e.g., a VARIAN X-ray imaging plate typically supports sampling rates of 30 fps or 15 fps) can cause flickering and incontinuity of medical video, hence frame rates below 24 fps puts a significant burden on the physician's diagnosis. Therefore, the trade-off between video quality and exposure time is important, and video frame interpolation seems promising to solve this if the luminance leap problem can be solved, as shown in Figure 2.

In traditional MEMC algorithms, motion estimation estimates the displacement of the object and get the motion vector, and motion compensation is used to adjust the displacement due to motion in the previous frame based on the obtained motion vector to get the predicted frame of the current frame as accurate as possible. The accuracy and reliability of the motion vectors are important indicators of the algorithm and the effectiveness of the video processing, while the selection of the prediction points and search methods determine the motion vectors. The existing methods include block matching, optical flow, pixel recursive algorithm, Bayesian algorithm, etc. The pixel blocks are usually used as a reference for motion estimation, and the sub-block with the highest similarity to the block to be matched in the current frame will be chose in the search region of the previous frame. To improve the speed of motion estimation, several search methods [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], block matching criteria, starting point prediction methods, and early termination strategies [30], [31], [32] have been proposed. For the accuracy of motion estimation, optimization algorithms such as hierarchical search [39], [40] and variable block size search [41] have been proposed. These methods achieve good frame interpolation performance on continuous motion and stable luminance videos, such as film and television, but still have high error rates for medical images. Most medical images are acquired by X-ray irradiation, such as Computerized Radiography (CR), Digital Radiography (DR), Electron computed tomography (CT), etc. X-rays have a strong penetration ability to ''observe'' soft tissue. X-rays will interact with different substances in the body, so that part of the energy is absorbed by different tissue of the body, while the remaining is received by the detector at the other end through the body. Therefore, due to the difference in tissues thickness of human body, there will be drastic luminance leap with the irradiation tube power changes during the irradiation imaging (in Figure 2).

However, although a delicate intermediate frame can be obtained via a traditional MEMC algorithm, when the luminance leaps as irradiation dose changes, it will seriously influence the prediction and matching accuracy. Therefore, the traditional MEMC-based video frame interpolation algorithm has several shortcomings in processing ionizing radiation videos: (1) The video sequences have nonlinear properties and luminance leaps that lead to non-applicability in MEMC; (2) The matching algorithm has high error rates when luminance leap occurs; (3) The search process has high computational consumption.

To solve the above problems, this paper proposes a video frame interpolation algorithm based on luminance compensation MEMC (LC-MEMC), to obtain smooth video with high utilization value without flicker. Firstly, a luminance compensation for illumination attenuation based on target thickness in motion estimation and motion compensation (MEMC) is proposed to solve the image discontinuity problem caused by luminance leap. Secondly, in order to obtain more accurate motion vectors, a block matching motion estimation method is also proposed, which introduces a fast search method from basic points to local points, and the block matching criterion selects the improved normalized cross-correlation (NCC) algorithm. The motion vector is obtained by motion estimation of the luminance-compensated frame, which then generates interpolation frames in motion compensation process. LC-MEMC aims to improve the accuracy and speed of video interpolation frames from three perspectives: adding luminance compensation, changing the search strategy, and optimizing the matching calculation at each search step.

The major contributions can be summarized as follows:

(1) To our best knowledge, it is the first attempt to introduce luminance compensation for frame interpolation based on illumination attenuation of target thickness, and design a luminance compensation MEMC (LC-MEMC). LC-MEMC is applicable to the medical area and has been put into the clinical application to reduce the doctors' visual fatigue, as well as halve the ionizing radiation exposure doses to the patients during medical fluoroscopy examinations.

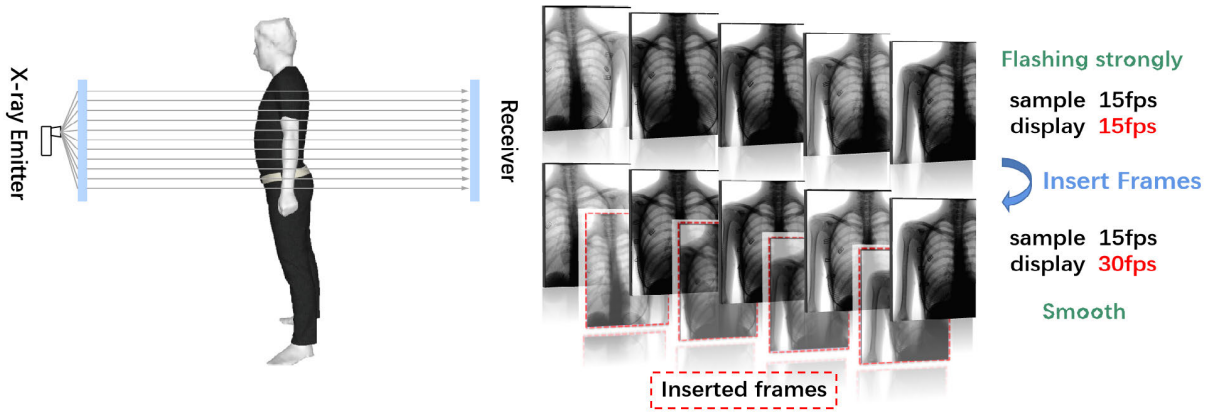


FIGURE 2. X-rays will appear intensity decay after passing through the human body, resulting in medical imaging affected by the thickness of human tissue, when the exposure dose or irradiation site changes will occur luminance leap. When filming with 15FPS and displaying at the same frame rate, there is a relatively strong flicker in the picture. After interpolating the frame to the video and displaying at 30FPS double frequency, the picture is continuous and smooth with better visual effect.

(2) This paper proposes a block matching method applicable to medical videos with less computation and more accurate matches. It reduces the complexity of the matching computation at each position and improves accuracy with faster searching speed.

(3) Evaluations on medical videos were carried out with other video frame interpolation algorithms. The experimental results show it not only improves automatic metrics (i.e., PSNR and SSIM close to the full search algorithm and NCP close to the rhombic method), but also has better visual performance.

II. RELATED WORKS

Video interpolation aims to increase video frame rate, which is often applied to tasks such as high frame rate video generation, slow-motion effect generation [4], view synthesis [5], video enhancement [6], video compression [7], etc. It has important practical significance for industries such as film, television, medical and criminal investigation, etc.

With the progress of video interpolation technology, different kinds of approaches have been proposed. The existing methods can mainly be classified into three categories: simple interpolation methods, methods based on motion estimation and motion compensation (MEMC), and Deep-learning-based methods. In this section, we briefly introduce these three kinds of methods.

A. SIMPLE INTERPOLATION METHOD

it is the simplest method in video interpolation, which directly inserts a repeated frame or a fused frame as the intermediate frame. Although this method is easy to operate, it does not take into account the motion information and luminance changes of the objects in the screen, so that the inserted frames are not highly continuous frames. For the original video with poor motion continuity and lag or flicker due to brightness changes, inserting repeated frames and averaging the fused frames before and after the two frames cannot achieve good results and will result in blurring, jittering and other poor viewing experience.

B. MOTION ESTIMATION (ME) AND MOTION COMPENSATION (MC) METHODS

this type of methods improves the insertion accuracy compared with the simple interpolation method. ME analyzes the current frame and the reference frame, and then performs MC according to the obtained motion vector to generate a motion continuous intermediate frame. The methods of motion vector acquisition can be classified into Block Matching, Bayesian Algorithm, Optical Flow, Pixel Recursive Algorithm, etc.

Block Matching Algorithm (BMA) is the most commonly used MEMC method, which uses a pixel block as a reference for motion estimation, finding the sub-block with the highest similarity to the block to be matched of the current frame in the search area of the previous frame, so that the motion of the object in each area can be well characterized by a parametric model. Therefore, the accuracy and efficiency of the block matching method depend on the choice of the block matching criterion and the search method. The commonly used block matching criteria are Sum of Absolute Difference (SAD), Absolute Mean Error Function (MAD), Minimum Mean Square Error Function (MSE) and Normalized Correlation Function (NCC), among which MSE has the highest accuracy but complex operation; MAD is slightly less accurate but easy to implement; SAD not only has a matching value that is equivalent to MSE, but also a greatly reduced computational effort; NCC is highly accurate and robust but slightly more computationally intensive. The choice of search methods has a great impact on the speed and accuracy of motion estimation, so researchers have worked on different fast search algorithms based on full search (FS) [8]. For example, Three-step Search (TSS) [9], New Three-step Search (NTSS) [10], Two Minimal Three-step Search [11], Four-step Search (FSS) [12], Simple Efficient Search (SES) [13], Diamond Search (DS) [14], [15], Diamond Cross Search (DCS) [16], Hexagonal Diamond Search (HDS) [17], Star-shaped Diamond Search (SD) [18], Adaptive Rood Pattern Search (ARPS), and many other methods [19], [20], [21].

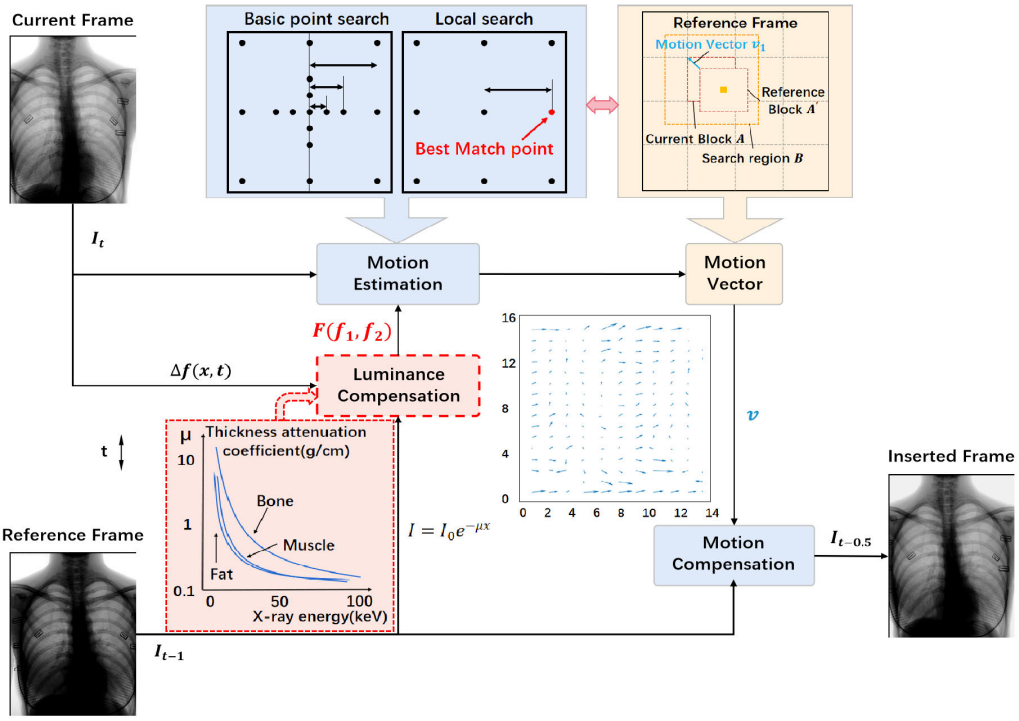


FIGURE 3. In the motion estimation and motion compensation framework, the Luminance compensation for illumination attenuation at target thickness is added to compensate the luminance difference between the current frame and the reference frame.

Block matching algorithms have been widely used as basic and accurate methods in regular videos. The disadvantages of such methods are that the accuracy of motion estimation is limited by the block size and it is not suitable for dealing with discontinuous values in the motion field. Therefore when the video has luminance instability based on ionizing radiation imaging, matching errors due to luminance leaps may occur regardless of the block matching criterion and search method chosen if compensation for luminance is not considered.

Bayesian method is an improved BMA; it is based on probabilistic statistical knowledge for data classification, which uses the motion vector of adjacent blocks to select the best search pattern adaptively [30] and develops an early termination strategy to reduce the number of invalid searches. Shen et al [31] proposed a Bayesian-decision-rule-based decision algorithm for coding unit size, and [32] proposed a new transformation unit decision algorithm based on it, which gives a specified early termination strategy based on the residual coefficients and block correlation. This method improves the accuracy and reduces the computational effort compared with BMA, but it still has limitations in dealing with luminance discontinuity values in motion fields and is only applicable to conventional videos such as movies and animations.

From the study of image pixel intensities, the optical flow method, which uses the time-domain variation and correlation of pixel intensities in an image to determine pixel locations, is also an advanced and effective method. Horn and Schunck [22] proposed to associate image grayscale values with a two-dimensional velocity field to allow efficient computation of the optical flow field; Lucas and Kanade [23] have since introduced constraints on the solution of the optical flow method by assuming that the optical flow field in space vector motion remains constant within the space, and the optical flow is calculated using the weighted least squares method. The actual motion scenes have motion blur, non-rigid motion, light reflection, and occluded regions causing poor robustness in computation, so methods to improve the optical flow objective function such as EpicFlow [24], EPPM [25], SPM-BP [26], and FullFlow [27] have also emerged successively, but they still do not overcome the limitation of algorithm complexity. Besides, the biggest limitation of the optical flow method is that it is only applicable to videos with small motion speed and distance and constant luminance, because the method is sensitive to light and the light change is very easy to affect the recognition effect. When the method is used for video interpolation of medical videos, the leap in luminance will cause the optical flow field to fail to reflect the motion of the target.

The Pixel Recursive Algorithm (PRA) uses the idea of recursion to perform pixel iterative operations around pixels in the gradient direction for the pixel data changes caused by object displacement, so that the successive operations converge to a motion vector. Pixel recursive search was introduced to motion estimation algorithms by Haan et al [28], which enabled motion vector computation to achieve sub-pixel precision; Tashlinskii [29] et al. proposed the stochastic gradient method to improve processing efficiency. However,

since each pixel of PRA is involved in the operation and the pixel spacing is small, such methods still have the limitations of high computational complexity and poor displacement tracking capability in practical application scenarios such as medical and industrial.

In general, the MEMC-based video frame interpolation algorithm still has research value, but it lacks the consideration of luminance variation and has limited application. The matching accuracy of this type of method still needs to be improved in real scenes, medical scenes and other changing complex situations, and the corresponding search speed also needs to be accelerated.

C. DEEP-LEARNING-BASED VIDEO INTERPOLATION METHODS

such methods, with Flow-Based Methods [4], [6], [33], [34], [43] and Kernel-Based Methods [35], [36], [37], [38] as the mainstream, have been actively developed. Among them, the representative methods with good performance are QVI [43] and DAIN [35], but the performance of both types of methods is limited by the underlying estimator, which may generate noise when complex occlusions are present in the video, resulting in noticeable artifacts. Later, Kalluri [42] et al. proposed FLAVR (CVPR2021) by using 3D CNN to reason about motion trajectories and attributes, with better interpolation accuracy and speed than previous work. However, deep learning-based methods rely on a large amount of data, which is more difficult to collect in most practical application scenarios, so this type of method also has a large limitation of use. Particularly, limited by their attributes such as uninterpretability and large data dependence, deep-learning methods and other methods still show the development trend of mutual promotion and complementary advantages in detail-sensitive video interpolation tasks such as medical and military.

In summary, all three types of existing methods have been well developed, especially the MEMC method is one of the most widely studied and has been applied tovarious commodity display devices, but there are still some limitations in the application in special fields such as medicine. Most of their implementations are based on the premise of luminance stability, ignoring the impact of luminance leaps on accuracy, thus failing to obtain smooth intermediate frames to form good visual effects, and even more failing to realize the practical use value in the medical field.

III. OUR APPROACH

In this section, we will first introduce the luminance compensation based on the illumination attenuation of the target thickness; then we will introduce the block matching algorithm we use. The overall illustration of the proposed LC-MEMC algorithm is shown in Figure 3.

A. LUMINANCE COMPENSATION FOR ILLUMINATION ATTENUATION AT TARGET THICKNESS

Biomedical video, for example, is primarily used to look inside the body and thus analyze lesions in bones and soft tissue, and therefore requires the aid of penetrating X-rays. The imaging principle of X-rays can be summarized as follows: electromagnetic wave starts from one end, passes through the body and is received by the detector at the other end, resulting in a two-dimensional image.

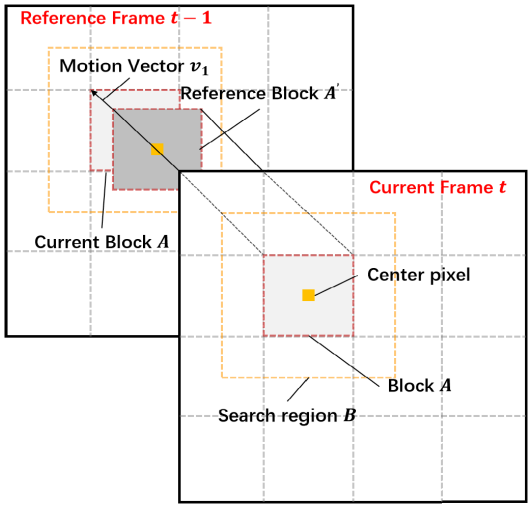
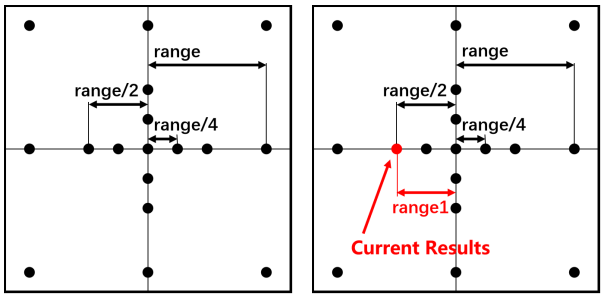


FIGURE 4. Motion estimation using block matching methods.

The brightness compensation of video interpolation is mainly aimed at the interpolation error caused by the brightness jump between adjacent frames in the imaging process. When X-rays pass through the target, they will be scattered and absorbed to form intensity attenuation. The degree of attenuation in the human body is mainly determined by atomic number and density and thickness of the tissue and organs. Therefore, when the imaging equipment irradiates different parts of the human body, there will be large brightness differences between adjacent frames of the video. Based on the fact above, this paper mainly compensates the brightness attenuation caused by different atomic numbers and density and thicknesses when inserting video frames. The attenuation amount is expressed by attenuation probability (determined by atomic number) and attenuation coefficient (determined by density and thickness).

1) ATTENUATION PROBABILITY

The attenuation probability is used to indicate the possibility of intensity attenuation and luminance leap occurring during the imaging process. After continuous X-rays pass through the human body, the low-energy part where the photoelectric effect occurs forms an intensity attenuation, and the high-energy part where the Compton effect occurs forms an intensity attenuation. The probability of occurrence of photoelectric effect is expressed by P_1, $P_1 \propto Z^3$; the probability of occurrence of Compton effect is expressed by P_2, $P_2 \propto Z^3$, where Z represents atomic number. Because the composition of different organs in the human body is different, the atomic number of different organs is also different, and thus the probability of forming attenuation of two effects is also different. In the human body, the attenuation of various soft tissue to



(a) Basic points (b) The current search results

FIGURE 5. Basic points search pattern.

X-rays is low, while the attenuation of bones and teeth to X-rays is high.

2) ATTENUATION VALUE

The attenuation coefficient indicates the degree of attenuation, and likewise indicates the degree of luminance leap. The fractional value of intensity attenuation that occurs when X-rays pass through a unit thickness of a material layer is expressed by the linear attenuation coefficient μ. In this paper, we approximated the intensity of X-rays after passing through the human body with $I = I_0 e^{-\mu x}$, where I_0 denotes the incident intensity and x denotes the thickness of the material layer.

3) COMPENSATION BASED ON LUMINANCE ATTENUATION

We define $F(f_1, f_2)$ to denote the luminance compensation function and $f_i(x, t)$ to denote the luminance attenuation function, with t denoting the irradiation site. For different irradiation sites and layer thicknesses, the luminance attenuation relative to the original X-ray should be $f_i(x, t) = I_0 - I$. In the video interpolation algorithm, we compensate the luminance of the previous frame compared to the reference frame, either positively or negatively, thus avoiding the effect of luminance leap on the motion vector calculation, denoted as $F(f_1, f_2) = ((f_1 - f_2))/2$.

B. BLOCK MATCHING VIDEO INTERPOLATION METHOD WITH LUMINANCE COMPENSATION

Based on the idea of motion estimation and motion compensation, LC-MEMC uses an improved block matching method incorporating luminance compensation to find the best matching position by calculating the similarity between image blocks. As shown in Figure 4, an image block of certain size (Block A) is set in the current frame, and the corresponding image block in the reference frame is Block A'. According to the search method and block matching criteria proposed in this paper, the image block with the highest similarity to block A is searched within a certain region (Search region B) near block A', and the displacement between the matched image blocks is analyzed to obtain the motion vector v_1 (the motion vector $v = v_1/2$ used in motion compensation).

Algorithm 1 Proposed Search of LC-MEMC

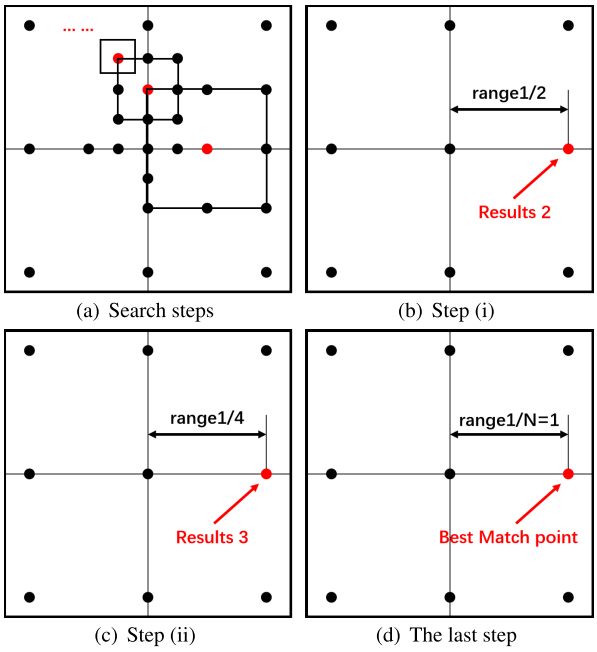
```
Function Found (int * x, int * y, int * max, intx1, inty1)
if Matchingcriterion(x1, y1) > max then
    *max ← Matchingcriterion(x1, y1);
    *x ← x1;
    *y ← y1;
end if
Procedure Search ()
int x, y, x2, y2, max;
//Basic points search
max ← Matchingcriterion(a, b);
Found (x, y, max, a + range, b + range);
. . . //Calculation for 16 surrounding basic points except the origin(a,b)
range1 ← abs(x);
//Local area search
while range1/2! = 1 do
    range1 ← range1/2;
    max ← Matchingcriterion(x2+range1, y2+range1);
    x ← x2 + range1;
    y ← y2 + range1;
    Found (x, y, max, x2 + range1, y2);
    . . . //Calculate for 7 surrounding points except the current search origin
end while
return (x,y);
```

1) SEARCH METHOD

The search method used by LC-MEMC is divided into two steps, as shown in Algorithm 1.

(1) Basic points search: as shown in Figure 5(a), the basic points search looks for 17 points consisting the origin, 8 points of the length of *range* from the origin, 4 points of *range*/2 from the origin and 4 points of *range*/4 from the origin, where the step *range* size is determined by the average speed of the human body moving in the image, which is set to 16 in this paper. In this step, block matching calculation and comparison are performed at 17 points (the block matching calculation guidelines are explained in detail in the next section). The point with the largest matching function value is the search result of the current step (*CurrentResult*). The search ends when the point with the largest value of the matching function is the center point, which is the final result (*BestMatchPoint*); otherwise the position of the point (*CurrentResult*) and the step between that position and the origin corresponding to *range*1 (as in Figure 5(b)) are recorded, and then the Local area search is entered.

(2) Local area search: the search steps of local area search are shown in Figure 6(a). Specifically, as in Figure 6(b), the search result (*CurrentResult*) of the basic point search is used as the search origin in Step(i), and the step is set to *range*1/2, searching the surrounding 8 points and finding



(a) Search steps (b) Step (i) (c) Step (ii) (d) The last step

FIGURE 6. Local area search.

the point with the largest value of the matching function as the new best matching point (*Result*2). As in Figure 6(c), in Step(ii), search again with the new best matching point (*Result*2) as the origin, with the step set to *range*1/4, and find the point (*Result*3) with the largest value of the matching function again. Repeat the above steps, as in Figure 6(d), when the search step is 1 for the last step to find the point with the largest value of the matching function and the end. The result (*BestMatchPoint*) is the peak point of the matching function in the region, and the coordinates of this point are the coordinates of the motion offset vector v_1.

The algorithm needs to search 33 points in the best case and 49 points in the worst case to find the peak point of the search area. Compared with the full search algorithm (FS)—which has the highest accuracy, it greatly reduces the number of matching operations and increases the speed a lot; compared with the diamond search algorithm (DS)—which has fewer search points, this algorithm slightly increases the number of search points and improves the accuracy of the search in the case of luminance change without significantly reducing the search speed.

2) BLOCK MATCHING CRITERION

An improved NCC block matching algorithm is also designed. The NCC algorithm has high accuracy and adaptability, is not affected by the linear transformation of the image gray value, and has a certain robustness for images with luminance leap. However, it is computationally complex, so we use the difference sum operation to simplify the algorithm to improve the matching efficiency.

The NCC matching algorithm determines the degree of matching by calculating the correlation value between the template image and the search image, and the position with the largest correlation value is the best position for the current template search. The size of the search image S is assumed to be $M \times M$ and the size of the template T is assumed to be $N \times N$ (where M and N represent image pixels, $M > N$). Template T translates the search on image S. The covered subgraph is denoted as $S^{i,j}$, and (i, j) is the coordinate of the top left vertex of the subgraph, then the normalized mutual correlation matching is defined as

$$NCC(i,j) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} S^{i,j}(m,n)T(m,n)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}\left[S^{i,j}(m,n) - \bar{S}^{i,j}\right]^2}} \times \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} S^{i,j}(m,n)T(m,n)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}[T(m,n) - \bar{T}]^2}} \quad (1)$$

The product of two arrays $f(x)$ and $g(x)$ of the same size $N(x = 1, 2, \ldots, K)$ is equivalently expressed as the product of one difference and the other progressive summation using the method of difference summation operation,

$$\sum_{x=1}^{K} f(x)g(x) = \sum_{x=1}^{K} F(x)G(x) \quad (2)$$

where

$$F(x) = f(x) - f(x+1) \quad (3)$$
$$G(x) = G(x-1) - g(x+1) \quad (4)$$
$$G(0) = 0 \quad (5)$$
$$f(K+1) = 0 \quad (6)$$

When this operation is used in Equation (1), all points within the template T are saved as an one-dimensional array $f(x)$, and the points corresponding to the template within the subgraph $S^{i,j}$ are saved as $g(x)$, with $F(x)$ denoting the difference array to $f(x)$ (as shown in Equation (3) and $G(x)$ denoting the progressive summation array to $g(x)$ (as shown in Equation (4), so that it is possible to transform the template T and the subgraph $S^{i,j}$ into a product operation on $F(x)$ and $G(x)$ according to Equation (2). In an image, the grayscale values of adjacent pixels do not differ much, and the differentiated array largely consists of 0s, 1s, and -1s, which can be ignored by multiplication operations, reducing the number of multiplication operations to a large extent. Moreover, in the actual template matching process, the template is fixed, so the differentiation on the template only needs to be performed once, which greatly reduces the time consumption.

IV. EXPERIMENTS

A. EXPERIMENTAL CONFIGURATION

1) EXPERIMENTAL ENVIRONMENT

the experiments in this paper were programmed with MATLAB R2018b. The hardware environment was Intel(R) Core(TM) CPU of 2.50GHz, 8G running memory. The software environment was Windows 10 operating system. The experiments used Toshiba X-ray tube, VARIAN DR flat panel detector,and the radiation frequency were 15 fps.

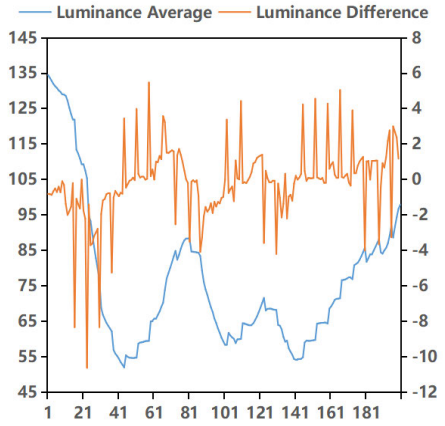


FIGURE 7. Luminance leap statistics.

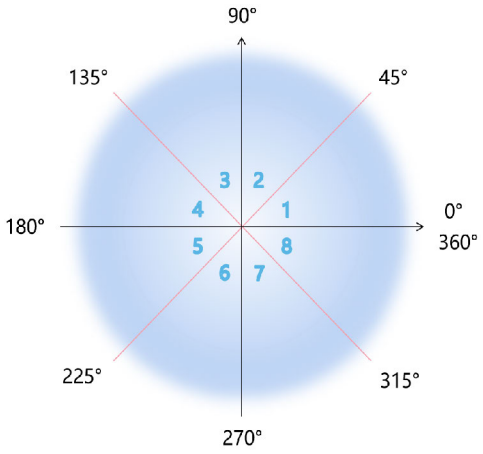


FIGURE 8. Area division diagram.

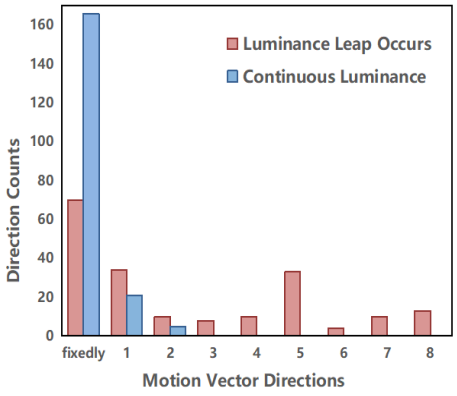


FIGURE 9. Distribution of motion vectors.

2) DATA

we collected a large number of human clinical X-ray videos for research. Due to privacy related reasons, the data cannot be made public at present, and we reported the results of the experiments in this paper.

B. EXPERIMENTS FOR ISSUE ANALYSIS

In order to verify the role of luminance compensation on medical video interpolation, we first counted the luminance variation of medical video. Then we analysed the effect of luminance variation on motion vectors, and obtain the effect on video quality, that is, video interpolation effects, so as to prove the necessity of luminance compensation.

1) LUMINANCE LEAPING PHENOMENON IN MEDICAL VIDEO

We divided the video sequence into 200 consecutive frames, and computed the average luminance of each frame and the luminance variation between adjacent frames. The statistical results in Figure 7 show that due to the special nature of the imaging device, the transmitter will automatically adjust the transmitting power according to the thickness of the human body, thus causing a luminance jump on the display, for example, the luminance difference between frame 16 and frame 17 is -8.3245, between frame 23 and frame 24 is -10.6134, etc. Such a luminance difference will greatly affect the visual effect of the video.

2) MOTION VECTOR DIRECTION STATISTICS

We analyzed the effect of luminance change on motion vectors and inferred the effect on video interpolation by dividing the direction of motion vectors into 8 regions in the plane according to angles (Figure 8) 0°∼45°, 45°∼90°, ⋯, 215°∼360° respectively. We selected adjacent frames in the video, recorded the motion vectors with values less than 0.5 between two frames as *fixedly*, and counted the distribution of other motion vectors in the 8 regions. The results in Figure 9 show that in the case of normal luminance change, most of the motion vectors are concentrated in the same region and a few are distributed in other regions; however, when there is a luminance leap, the motion vectors are almost evenly distributed in all regions. It can be seen that when there is a leap in luminance, some blocks are incorrectly matched, and the distribution of motion vectors is scattered, which affects the generation of motion estimation and transition frames.

TABLE 1. The impact of luminance compensation on video quality.

Methods/Criteria	PSNR	SSIM
With Luminance Compensation	**34.8033**	**0.9224**
Without Luminance Compensation	32.3409	0.9001

3) THE NECESSITY OF LUMINANCE COMPENSATION

We used the same search method and matching function to conduct experiments with and without adding luminance compensation, and counted the PSNR, SSIM and their average values for 200 consecutive frames. As the results shown in Table 1 and Figure 10, the mean values of PSNR and SSIM with luminance compensation are higher than without luminance compensation, and this is also the case in the statistics of most frames. In addition, we analyze the luminance change against Figure 7 and get that when there is a luminance leap in a certain frame of the video (e.g., the 17th and 24th frames), the video frame interpolation

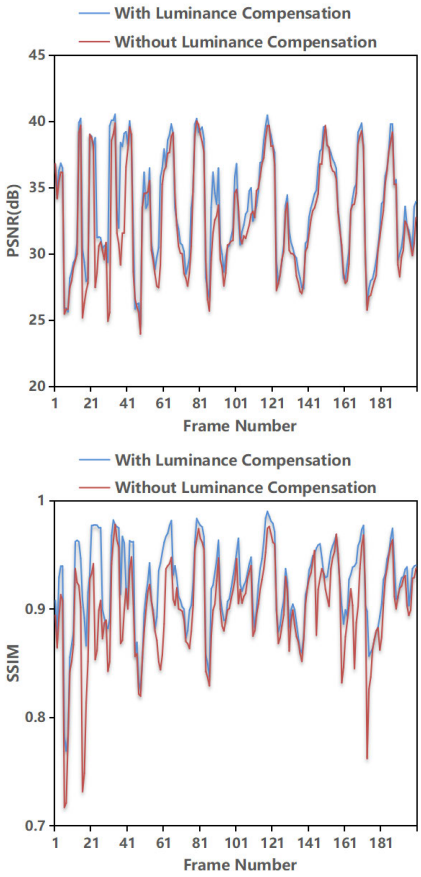


FIGURE 10. The impact of luminance compensation on video quality (PSNR and SSIM).

method without luminance compensation has lower PSNR and SSIM. After adding luminance compensation, $PSNR_{17}$ increased from 25.1787 to 29.3788, $SSIM_{17}$ increased from 0.7317 to 0.9082 while $PSNR_{24}$ increased from 27.4693 to 38.7607, $SSIM_{24}$ increased from 0.8536 to 0.9777, and the interpolation quality of the nearby frames also improved. This shows that adding luminance compensation can reduce the impact of luminance leap when doing video interpolation, thereby improving the accuracy of interpolation and video quality.

C. COMPARISON EXPERIMENTS

In order to reflect the advancedness of LC-MEMC compared with other methods, besides reflecting the necessity of luminance compensation in the previous section, we also make comparative experiments with traditional MEMC methods from two perspectives of search method and matching function, respectively. In addition to that, we also compare with advanced deep learning methods. Finally, several visual effect examples are used to further demonstrate the differences of several methods distinctly.

1) COMPARISON OF DIFFERENT SEARCH METHODS

We conducted comparative studies between the proposed method in this paper (LC-MEMC) and the following methods

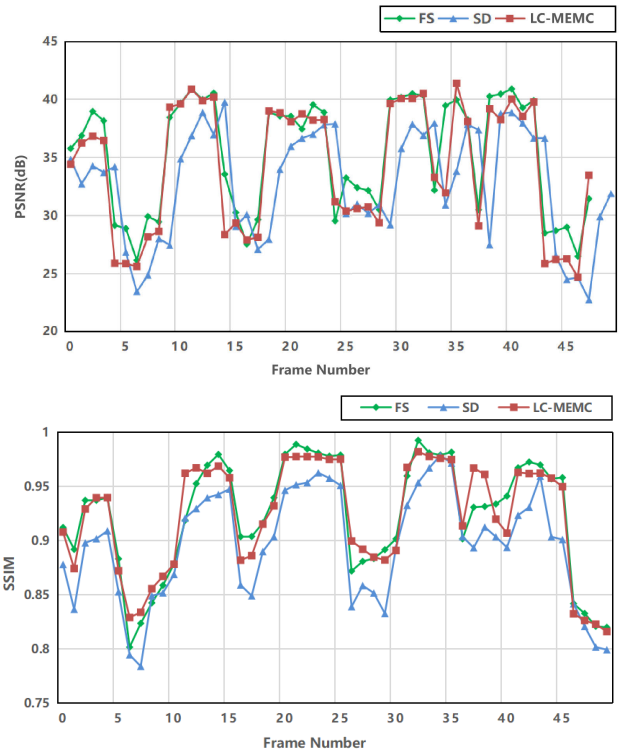


FIGURE 11. The video quality (expressed by PSNR and SSIM) of some consecutive frames of LC-MEMC, SD and FS reference results.

Full Search (FS), Three-step Search (TSS), New Three-step Search (NTSS), Four-step Search (FSS), Diamond Search (DS), Diamond Cross Search (DCS), Star-shaped Diamond Search (SD) on medical images (all based on luminance compensation and the same block matching criteria). We used the video processed by FS as the reference results, the mean peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as video quality measures, and the average Number of Calculated Points (NCP) required per macroblock and the average time to process a video as evaluation criteria for method complexity. As shown by the results in the conventional MEMC methods section of Table 2, compared with the FS reference results, SD has the lowest NCP and processing time (only 20.81, 0.5613), but its PSNR and SSIM are at a lower level, so the SD algorithm processes fastest but with lower accuracy; similarly, other fast search algorithms (TSS, NTSS, FSS, DS, and DCS) have fewer NCP and Time but PSNR and SSIM are also much lower than the reference results, which means these algorithms have the fastest processing speed but lower accuracy in medical images. In contrast, our LC-MEMC has the PSNR and SSIM closest to the FS reference results, possessing high accuracy, and the NCP and Time are not significantly increased compared to the fast search algorithm, which has improved the processing speed to a great extent compared to FS reference results.

Therefore, LC-MEMC is a method that combines the advantages of FS (reference results) and fast search algorithm, and is able to achieve fast and high accuracy in the processing of medical images.

TABLE 2. Comparison of our method with other methods.

	Methods/Criteria	PSNR	SSIM	NCP	Time(s)
Conventional MEMC Methods	FS(reference results) [8]	35.3101	0.9231	201.43	3.7452
	TSS [9]	32.0782	0.8919	26	0.5929
	NTSS [10]	32.7514	0.8987	28	0.5933
	FSS [12]	32.2891	0.8993	24.93	0.5902
	DS [14] [15]	31.7681	0.8861	21.13	0.5867
	DCS [16]	32.0284	0.8876	21	0.5835
	SD [18]	30.5743	0.8789	**20.83**	**0.5613**
	LC-MEMC	**34.8033**	**0.9224**	34.1	0.5947
Deep Learning Methods	QVI [43]	31.8746	0.8742	/	0.7044
	DAIN [35]	30.9753	0.8700	/	0.6911
	FLAVR [42]	33.9443	0.9047	/	0.5934

TABLE 3. Comparison of different block matching functions.

Methods/Criteria	NCP	Time(s)	PSNR
$\mathrm{LC-MEMC}_{MSE}$	34.1	0.3012	34.4727
$\mathrm{LC-MEMC}_{MAD}$	34.1	0.2653	34.1675
$\mathrm{LC-MEMC}_{SAD}$	34.1	0.2064	34.4812
$\mathrm{LC-MEMC}_{NCC}$	34.1	0.2472	34.6605
LC-MEMC	**34.1**	**0.2035**	**34.8033**

Since LC-MEMC has the highest accuracy and SD has the fastest search speed, we counted the video quality (expressed by PSNR and SSIM) and processing speed (expressed by NCP and Time) of LC-MEMC, SD and FS reference results for some consecutive frames to show the difference between them. The results in Figure 11 indicate that for almost all frames, the video quality of our method is comparable to that of the FS reference results, even surpassing the reference results around frames 9-13 and 36, while the DS method is slightly inferior. In Figure 12, we counted the NCP and processing time of several methods for 10 consecutive frames. We can see that, LC-MEMC is close to the most efficient SD algorithm with respect to the reference results. The NCP of LC-MEMC is usually only about 1/6 of that of FS reference results, and the NCP of SD is 1/10 of that of FS, with only about 6% difference between the two; in terms of processing time, LC-MEMC is also close to SD, only 1/10 of that of reference results. In summary, LC-MEMC achieves a much higher processing speed while maintaining a high accuracy rate.

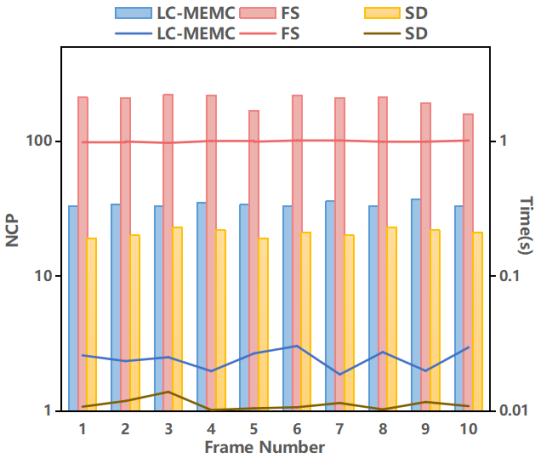


FIGURE 12. NCP and processing time of the methods in this paper (LC-MEMC), SD and FS reference results.

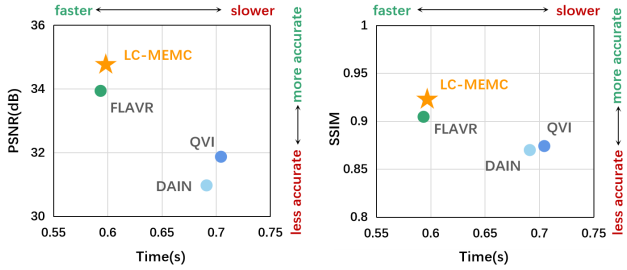


FIGURE 13. Comparison of accuracy and processing time of LC-MEMC, QVI, DAIN and FLAVR.

2) COMPARISON OF DIFFERENT BLOCK MATCHING FUNCTIONS

In our method, an improved NCC matching criterion is used in order to reduce the computational complexity of the algorithm. Experiments compare our method with different classical matching functions (using luminance compensation and our proposed search method). The methods include $\mathrm{LC-MEMC}_{MSE}$, $\mathrm{LC-MEMC}_{MAD}$, $\mathrm{LC-MEMC}_{SAD}$, $\mathrm{LC-MEMC}_{NCC}$ and LC-MEMC. We calculated the NCP, computing time and PSNR for 50 consecutive frames of video. The results in Table 3 below show that replacing the matching function does not affect NCP; MSE is considerably accurate but the processing speed is significantly slower; MAD slightly improved the processing speed but the accuracy is the lowest; SAD and NCC have better performance with medium processing speed and accuracy; in contrast, our improved NCC has the fastest processing speed and the highest accuracy.

3) COMPARISON WITH ADVANCED METHODS OF DEEP LEARNING

We compare the approach in this paper with representative state-of-the-art methods for deep learning on our data by counting their average PSNR, SSIM and the average time to process a video. Methods include QVI [43] for Flow-Based Methods, DAIN [35] for Kernel-Based Methods, and the advanced FLAVR [42]. It should be mentioned that the experiments in this section were run on a Quadro RTX 5000 GPU. The results in the deep learning methods section of Table 2 and Figure 13 indicate that the QVI

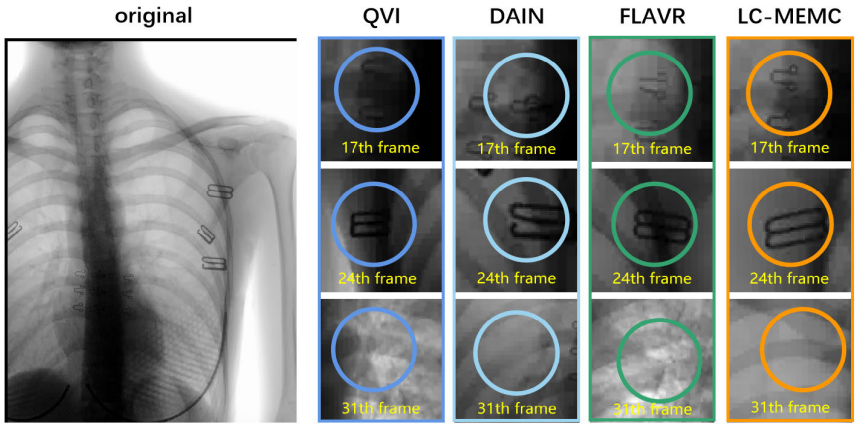


FIGURE 14. The visual effects of the interpolated frames generated by the deep learning method and LC-MEMC when there is a large leap in luminance.

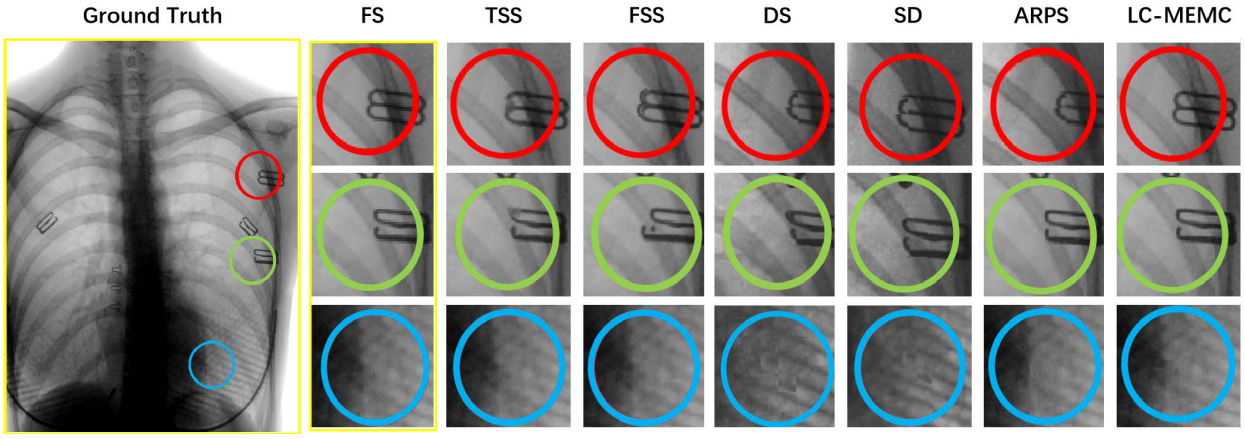


FIGURE 15. Comparison of our proposed method LC-MEMC with FS reference results and five existing methods.The pictures are from the medical image dataset we collected.

and DAIN methods have a slower processing speed, and the PSNR below 32 dB and SSIM below 0.9, so it has a lower utilization value. FLAVR has improved its processing speed and accuracy to a large extent, with an accuracy of 33.9443 PSNR and 0.9047 SSIM. LC-MEMC improves PSNR by 0.859 and SSIM by 0.0177 compared to the advanced method FLAVR, and the runtime is comparable to the advanced method FLAVR (0.5947s vs. 0.5934s). So, in comparison, LC-MEMC has the highest utilization value in medical imaging.

In addition, we focus on the visual effects of the interpolated frames generated by the deep learning method and LC-MEMC when there is a large leap in luminance. Take frames 17, 24, and 31 as example, the borders around the bones and organs are clear and the lines around the metal clips are smooth in the original image, but the interpolated frames generated by the three deep learning methods have different degrees of errors. As shown in Figure 14, the three rows indicate the interpolated frames generated by the different methods, with frames 17, 24 and 31 as the reference frames, respectively. In the interpolated frames generated by QVI, missing and overlapping metal clips and large blurs around bones and organs; With DAIN, the two metal clips appear offset and overlapped respectively, while the bones appear truncated; FLAVR similarly left two metal clips with varying degrees of loss and cheapness, and small truncations of the bone. These errors have a significant impact on the doctor's diagnosis in practical medical applications, while there are no obvious errors in any of the three interpolated frames generated by the method LC-MEMC in this paper. Therefore, it is shown that the deep learning method also cannot achieve good interpolation of frames for medical videos with luminance leaps.

D. VISUAL EFFECT COMPARISON AND CASE STUDY

It is well known that existing quantitative evaluations can only measure the performance of algorithms to a certain extent and are not a substitute for human visual perception, and that the continuity of videos cannot be measured by numerical standards. We find that the visual qualitative difference and exploitable value of our method compared with other methods is much larger than the numerical difference. Therefore,

we again compared the visual effects of LC-MEMC with other conventional MEMC methods on the acquired medical images. Figure 15 shows the interpolated frames generated by different methods. Different methods have different matching results in detail positions such as metal clip and around organs. We mark three details with circles in red, green and blue respectively. At the locations of the metal clip with clear boundary (red and green circles), the five rapid methods TSS, FSS, DS SD and ARPS have matching errors, in which TSS causes partial boundary blurring, FSS has partial boundary duplicate matching, DS and SD make part of the clip misaligned, ARPS also has a small range of misalignment and absence, and LC-MEMC has similar effect with FS (reference results) and is close to ground truth. At the location of the around organs, FS, TSS and FSS are blurred to some extent, TSS and FSS are blurred more severely, DS is partially misaligned, SD is partially wrong and has distorted lines, ARPS is also partially distorted and missing, and LC-MEMC results are similar to the reference results, with only a small degree of shadow diffusion, closest to ground truth. In terms of the number of markers in error, TSS and FSS have two serious errors and one minor error that can be ignored; DS and SD have three serious errors; ARPS has two serious errors and one minor error; and LC-MEMC has only one negligible minor error similar to the FS reference result.

To sum up, for ionizing radiation videos, when the irradiation dose changes or the local motion speed is large, the matching errors of traditional MEMC methods will greatly increase, resulting in visual discontinuity. In contrast, LC-MEMC shows the best ability to process details in inserted frames, which obtains delicate and smooth transition frames and reaches higher processing speed.

V. CONCLUSION

In this paper, we propose a video interpolation algorithm based on luminance compensation MEMC (LC-MEMC). For ionizing radiation videos where luminance leaps exist and thus are difficult to process with existing image processing methods, we propose a luminance compensation of electromagnetic wave irradiation attenuation based on target thickness, which, together with the simplified block matching method proposed in this paper for searching accurate computation, maximizes processing speed while improving accuracy. To the best of our knowledge, this paper is the first to propose luminance compensation to assist video interpolation and put it into clinical application in the medical field. We have experimentally verified that the medical video processed by this method is smooth and fluent without flickering, achieving high utilization value of reducing the visual fatigue of doctors and halving the radiation to patients.

REFERENCES

[1] D. J. Brenner and C. D. Elliston, "Estimated radiation risks potentially associated with full-body CT screening," *Radiology*, vol. 232, no. 3, pp. 735–738, Sep. 2004.

[2] D. J. Brenner and E. J. Hall, "Computed tomography—An increasing source of radiation exposure," *New England J. Med.*, vol. 357, no. 22, pp. 2277–2284, 2007.

[3] G. Valizadeh, F. B. Mofrad, and A. Shalbaf, "Impacts of spherical harmonics shape descriptors on the inter-slice interpolation of MR images," in *Proc. 26th Nat. 4th Int. Iranian Conf. Biomed. Eng. (ICBME)*, Nov. 2019, pp. 26–30.

[4] H. Jiang, D. Sun, V. Jampani, M.-H. Yang, E. Learned-Miller, and J. Kautz, "Super SloMo: High quality estimation of multiple intermediate frames for video interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9000–9008.

[5] X. Qing, Li Shuai, H. Aimin, and Z. Qinping, "Deep learning for digital geometry processing and analysis: A review," *Comput. Res. Develop.*, vol. 56, no. 1, p. 155, 2019.

[6] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman, "Video enhancement with task-oriented flow," *Int. J. Comput. Vis.*, vol. 127, no. 8, pp. 1106–1125, 2019.

[7] G. Lu, X. Zhang, L. Chen, and Z. Gao, "Novel integration of frame rate up conversion and HEVC coding based on rate-distortion optimization," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 678–691, Feb. 2018.

[8] K. Yi, Y.-H. Lee, and J.-H. Joo, "A fast video decoding technique by means of converting input video stream into forward-oriented format stream in little-endian systems," in *Proc. 7th Int. Conf. Multimedia, Comput. Graph. Broadcast. (MulGraB)*, Nov. 2015, pp. 15–18.

[9] D. Gangodkar, P. Kumar, and P. Kumar, "Real-time motion detection using block matching algorithms on multicore processors," *Int. J. Inf. Commun. Technol.*, vol. 3, no. 2, pp. 131–147, Feb. 2011.

[10] T. Koga, "Motion compensated interframe coding for video-conferencing," in *Proc. Nat. Telecommun. Conf.*, 1981, pp. G5.3.1–G5.3.5.

[11] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, no. 4, pp. 438–442, Aug. 1994.

[12] H. A. Basher, "Two minimum three step search algorithm for motion estimation of images from moving IR camera," in *Proc. IEEE Southeastcon*, Mar. 2011, pp. 384–389.

[13] L.-M. Po and W.-C. Ma, "A novel four-step search algorithm for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 313–317, Jun. 1996.

[14] J. Lu and M. L. Liou, "A simple and efficient search algorithm for block-matching motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 429–433, Apr. 1997.

[15] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 287–290, Feb. 2000.

[16] X. Wang, W. Wan, J. Zhang, and Y. Ma, "Research on the motion estimation with a novel octagon cross diamond search algorithm," in *Proc. Asia Pacific Conf. Postgraduate Res. Microelectron. Electron. (PrimeAsia)*, Sep. 2010, pp. 89–92.

[17] R. A. Manap, S. S. S. Ranjit, A. A. Basari, and B. H. Ahmad, "Performance analysis of hexagon-diamond search algorithm for motion estimation," in *Proc. 2nd Int. Conf. Comput. Eng. Technol.*, 2010, pp. V3-155–V3-159.

[18] D. Kerfa and A. Saidane, "An efficient algorithm for fast block matching motion estimation using an adaptive threshold scheme," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 24173–24184, Sep. 2020.

[19] B.-G. Kim, S.-K. Song, and P.-S. Mah, "Enhanced block motion estimation based on distortion-directional search patterns," *Pattern Recognit. Lett.*, vol. 27, no. 12, pp. 1325–1335, Sep. 2006.

[20] Z. Pan, R. Zhang, W. Ku, and Y. Wang, "Adaptive pattern selection strategy for diamond search algorithm in fast motion estimation," *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 2447–2464, Jan. 2019.

[21] A. P. S. Immanuel and J. Anitha, "An unvarying orthogonal search with small triangle pattern for video coding," in *Smart Intelligent Computing and Applications*. Singapore: Springer, 2019, pp. 43–52.

[22] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1980.

[23] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, Aug. 1981, vol. 81, no. 1, pp. 674–679.

[24] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-preserving interpolation of correspondences for optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1164–1172.

[25] L. Bao, Q. Yang, and H. Jin, "Fast edge-preserving PatchMatch for large displacement optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3534–3541.

[26] Y. Li, D. Min, M. S. Brown, M. N. Do, and J. Lu, "SPM-BP: Sped-up PatchMatch belief propagation for continuous MRFs," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4006–4714.
[27] Q. Chen and V. Koltun, "Full flow: Optical flow estimation by global optimization over regular grids," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4706–4714.
[28] G. de Haan and P. W. A. C. Biezen, "Sub-pixel motion estimation with 3-D recursive search block-matching," *Signal Process., Image Commun.*, vol. 6, no. 3, pp. 229–239, Jun. 1994.
[29] A. G. Tashlinskii, P. V. Smirnov, and M. G. Tsaryov, "Pixel-by-pixel estimation of scene motion in video," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vols. 24, pp. 61–65, May 2017.
[30] H. R. Tohidypour, M. T. Pourazad, and P. Nasiopoulos, "A low complexity mode decision approach for HEVC-based 3D video coding using a Bayesian method," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 895–899.
[31] X. Shen, L. Yu, and J. Chen, "Fast coding unit size selection for HEVC based on Bayesian decision rule," in *Proc. Picture Coding Symp.*, May 2012, pp. 453–456.
[32] L. Shen, Z. Zhang, X. Zhang, P. An, and Z. Liu, "Fast TU size decision algorithm for HEVC encoders using Bayesian theorem detection," *Signal Process., Image Commun.*, vol. 32, pp. 121–128, Mar. 2015.
[33] Z. Liu, R. A. Yeh, X. Tang, Y. Liu, and A. Agarwala, "Video frame synthesis using deep voxel flow," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4463–4471.
[34] T. Xue, B. Chen, and J. Wu, "Frame interpolation with multi-scale deep loss functions and generative adversarial networks," U.S. Patent 11 122 238, 2021.
[35] W. Bao, W.-S. Lai, C. Ma, X. Zhang, Z. Gao, and M.-H. Yang, "Depth-aware video frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3703–3712.
[36] W. Bao, W.-S. Lai, X. Zhang, Z. Gao, and M.-H. Yang, "MEMC-Net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 933–948, Mar. 2021.
[37] X. Cheng and Z. Chen, "Video frame interpolation via deformable separable convolution," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 10607–10614.
[38] H. Lee, T. Kim, T.-Y. Chung, D. Pak, Y. Ban, and S. Lee, "AdaCoF: Adaptive collaboration of flows for video frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5316–5325.
[39] K. M. Nam, J.-S. Kim, R.-H. Park, and Y. S. Shim, "A fast hierarchical motion vector estimation algorithm using mean pyramid," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 4, pp. 344–351, Aug. 1995.
[40] N. Al-Najdawi, M. Noor Al-Najdawi, and S. Tedmori, "Employing a novel cross-diamond search in a modified hierarchical search motion estimation algorithm for video compression," *Inf. Sci.*, vol. 268, pp. 425–435, Jun. 2014.
[41] G. Senbagavalli and R. Manjunath, "Motion estimation using variable size block matching with cross square search pattern," *Social Netw. Appl. Sci.*, vol. 2, no. 8, pp. 1–9, Aug. 2020.
[42] T. Kalluri, T. Kalluri, D. Pathak, M. Chandraker, and D. Tran, "FLAVR: Flow-agnostic video representations for fast frame interpolation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1–17.
[43] X. Xu, T. Kalluri, D. Pathak, M. Chandraker, and D. Tran, "Quadratic video interpolation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1106–1125.

ZIXUAN XU received the B.E. degree in digital media technology from the Wuhan Institute of Technology, in 2020, where she is currently pursuing the M.S. degree in software engineering with the School of Computer Science and Engineering (School of Artificial Intelligence). She is also participating in the joint training at the Institute of Software, Chinese Academy of Sciences. Her research interests include computer vision and digital image processing.

WENJING YING received the B.S. degree in computer science from the Worcester Polytechnic Institute, Worcester, MA, USA, in 2021. She is currently pursuing the M.S. degree in software engineering with the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. She worked as a Visiting Scholar and a Research Intern with the Institute of Software, Chinese Academy of Science, in 2021. Her research interest includes the area of scalable systems.

HAO HE received the Ph.D. degree from the University of Chinese Academy of Sciences, in 2019. From 2019 to 2021, he worked with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. He is currently an Assistant Professor at the Institute of Software, Chinese Academy of Sciences. His research interests include computer vision and natural language processing.

QINGMENG ZHU received the M.S. degree in computer application from the University of Chinese Academy of Sciences. He is currently pursuing the Ph.D. degree with the Institute of Software, Chinese Academy of Sciences. From 2014 to 2019, he worked with the Institute of Software, Chinese Academy of Sciences, as a Project Manager to develop several national projects. His research interests include the areas of time-space big data processing, cloud computing, and artificial intelligent.

JIAN LIANG received the B.S. degree in geomatics from the China University of Geosciences, Beijing, China, in 2009, and the M.S. degree in cartography and geographic information system from the Graduate University of Chinese Academy of Sciences, Beijing, in 2012. His research interests include the areas of big data processing and cloud computing.

HAIHUI WANG received the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Image Recognition & Artificial Intelligence, Huazhong University of Science and Technology. He is currently a Professor at the Wuhan Institute of Technology. His publications span several research areas, and his research interests include related to several topics, including digital image processing, computer vision, target detection, image understanding and recognition, and intelligent computing.