

Research Article

Spatiotemporal Super-Resolution Reconstruction Based on Robust Optical Flow and Zernike Moment for Video Sequences

Meiyu Liang, Junping Du, and Honggang Liu

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Junping Du; junpingdu@126.com

Received 14 March 2013; Accepted 18 June 2013

Academic Editor: Qingsong Xu

Copyright © 2013 Meiyu Liang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the spatiotemporal resolution of the video sequences, a novel spatiotemporal super-resolution reconstruction model (STSR) based on robust optical flow and Zernike moment is proposed in this paper, which integrates the spatial resolution reconstruction and temporal resolution reconstruction into a unified framework. The model does not rely on accurate estimation of subpixel motion and is robust to noise and rotation. Moreover, it can effectively overcome the problems of hole and block artifacts. First we propose an efficient robust optical flow motion estimation model based on motion details preserving, then we introduce the biweighted fusion strategy to implement the spatiotemporal motion compensation. Next, combining the self-adaptive region correlation judgment strategy, we construct a fast fuzzy registration scheme based on Zernike moment for better STSR with higher efficiency, and then the final video sequences with high spatiotemporal resolution can be obtained by fusion of the complementary and redundant information with nonlocal self-similarity between the adjacent video frames. Experimental results demonstrate that the proposed method outperforms the existing methods in terms of both subjective visual and objective quantitative evaluations.

1. Introduction

The resolution quality of the video sequences, which are collected by multisource vision sensors, plays an important role in the accurate moving targets recognition and tracking of the intelligent monitoring and control system. However, some factors such as light, inaccurate focusing, optical or motion blur, subsampling, and noise disturbance can negatively affect the visual quality of the video sequences. In this situation, the spatiotemporal super-resolution (SR) reconstruction technology [1] can provide an excellent solution, which can be described as in Figure 1. By making full use of the complementary and redundant information with similar but not exactly the same details in the different spatiotemporal scales between the adjacent video frames, video sequences with high resolution (HR) via fusion of several low resolution (LR) video frames can be produced, thus it would have great research significance and application potential for the intelligent monitoring and control.

In recent years, the spatiotemporal super-resolution reconstruction technology has become the focus of much

research [2–5]. Now lots of researchers make use of the complementary and redundancy features between multi-frame images and make endeavors to research the SR reconstruction via fusion of multiframe images, which aims to improve the spatial resolution of each image or video [6–9]. However, the traditional methods [10, 11] usually rely on accurate estimation of subpixel motion, which only constrain their applicability to video sequences with relatively simple motions such as global translation. Thus in recent years, some scholars have proposed a novel fuzzy registration scheme for probabilistic estimation of motion based on similarity match and introduced it into the super-resolution methods [12] for further improving the spatial resolution of image or video, which can effectively avoid the accurate estimation of subpixel motion. Using such a scheme, Protter et al. [13] proposed a nonlocal mean (NLM) based SR framework by extending the NLM filter [14, 15] concept successfully applied in the denoising field into the SR field. Su et al. proposed a spatially adaptive block-based SR model [16].

However, there are also some limitations which exist in the new developed fuzzy registration scheme. If different

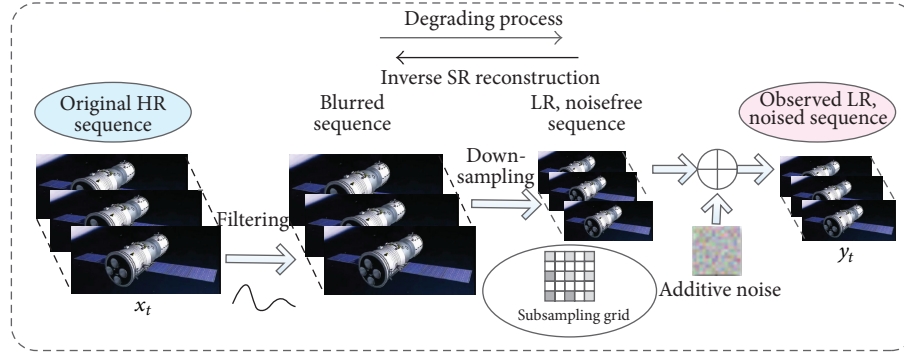


FIGURE 1: Description of the super-resolution reconstruction problem.

angle rotations exist in the image or video sequence, the correlation between corresponding pixels becomes weak, and then it will be difficult to use LR images or video frames effectively in the process of SR reconstruction. Moreover, if the LR images or video frames are noised, the reconstruction quality will be affected seriously. Thus considering good properties of rotation, translation, and scale-invariance of Zernike moment (ZM) [17, 18], we propose a fast fuzzy registration scheme based on ZM by using the self-adaptive region correlation judgment strategy, which could make an efficient similarity measure between region features in the spatiotemporal nonlocal domain for weight calculation. Based on that, we construct a novel spatiotemporal SR reconstruction model based on robust optical flow and ZM, which makes full use of the nonlocal self-similarity and redundant information in the different spatiotemporal scales between the adjacent video frames and produces video sequences with high resolution via fusion of several LR video frames. Meanwhile, the new model integrates spatial SR and temporal SR into a unified framework, which can improve both the spatial resolution and the temporal resolution and make the video sequences more clear and fluent. Different from the traditional SR reconstruction methods, the proposed method does not rely on accurate estimation of subpixel motion and can be adaptive to many kinds of complex motion patterns. The traditional motion vector based frame interpolation technology [19, 20] can also improve the temporal resolution, but because of the inevitable influence from the motion estimation error, the visual block or hole artifacts usually exist in the interpolation frame. Our proposed method can effectively overcome these artifacts while improving the spatial and temporal resolution.

The contributions of this paper are as follows. (1) We propose a novel spatiotemporal SR reconstruction model for video sequences based on robust optical flow and ZM. (2) We propose a robust optical flow motion estimation and compensation model based on motion details preserving. (3) By introducing the self-adaptive region correlation judgment strategy, we construct a fast fuzzy registration scheme based on ZM for better STSR with higher efficiency. (4) An efficient iterative curvature-based interpolation (ICBI) scheme is introduced to obtain the initial HR estimate of each LR video frame.

The remainder of the paper is organized as follows. Section 2 presents our proposed model architecture. Section 3 describes the algorithm implementation of our model. Section 4 gives the experimental results and analysis. Conclusions are presented in Section 5.

2. The Model Architecture

Given a low resolution video sequence $\{y_m[i, j, t]\}_{t=1}^T$, which is degraded by blurring B , downsampling D , noise disturbance ε , and frame missing from the high resolution video sequence $\{x_k[i, j, t]\}_{t=1}^T$, where $m \in M$, $k \in N$, $M \leq N$, N denotes the frame number in $\{x_k[i, j, t]\}_{t=1}^T$ and M denotes the frame number in $\{y_m[i, j, t]\}_{t=1}^T$. Our proposed model aims to reconstruct the initial high resolution video sequence $\{x_k[i, j, t]\}_{t=1}^T$, $k \in N$, from degraded low resolution video sequence $\{y_m[i, j, t]\}_{t=1}^T$, $m \in M$, by using the spatiotemporal reconstruction technology. Thus in this paper, a novel spatiotemporal SR reconstruction model (STSR) based on robust optical flow and Zernike moment is proposed, which integrates the spatial SR and temporal SR into a unified framework. The architecture of the proposed model is shown in Figure 2. It mainly includes the following three processes for the spatiotemporal SR reconstruction modeling.

First, by making the motion analysis in the spatiotemporal domain of the video sequence $\{y_m[i, j, t]\}_{t=1}^T$, $m \in M$, we propose a robust multilayered optical flow motion estimation method based on motion details preserving to obtain the motion vector (u, v) .

Then, according to the obtained motion vector (u, v) , we introduce an efficient biweighted fusion strategy to implement the spatiotemporal motion compensation, aiming at obtaining the compensated video sequence $\{y'_p[i, j, t]\}_{t=1}^T$, $p \in N$.

Finally, the fast fuzzy registration scheme based on ZM is proposed to implement the efficient spatiotemporal SR reconstruction and optimization for the sequence $\{y'_p[i, j, t]\}_{t=1}^T$, $p \in N$, via fusion of the nonlocal complementary and redundant information between the adjacent video frames, which can produce high quality video sequence $\{x_k[i, j, t]\}_{t=1}^T$, $k \in N$, with high spatiotemporal resolution.

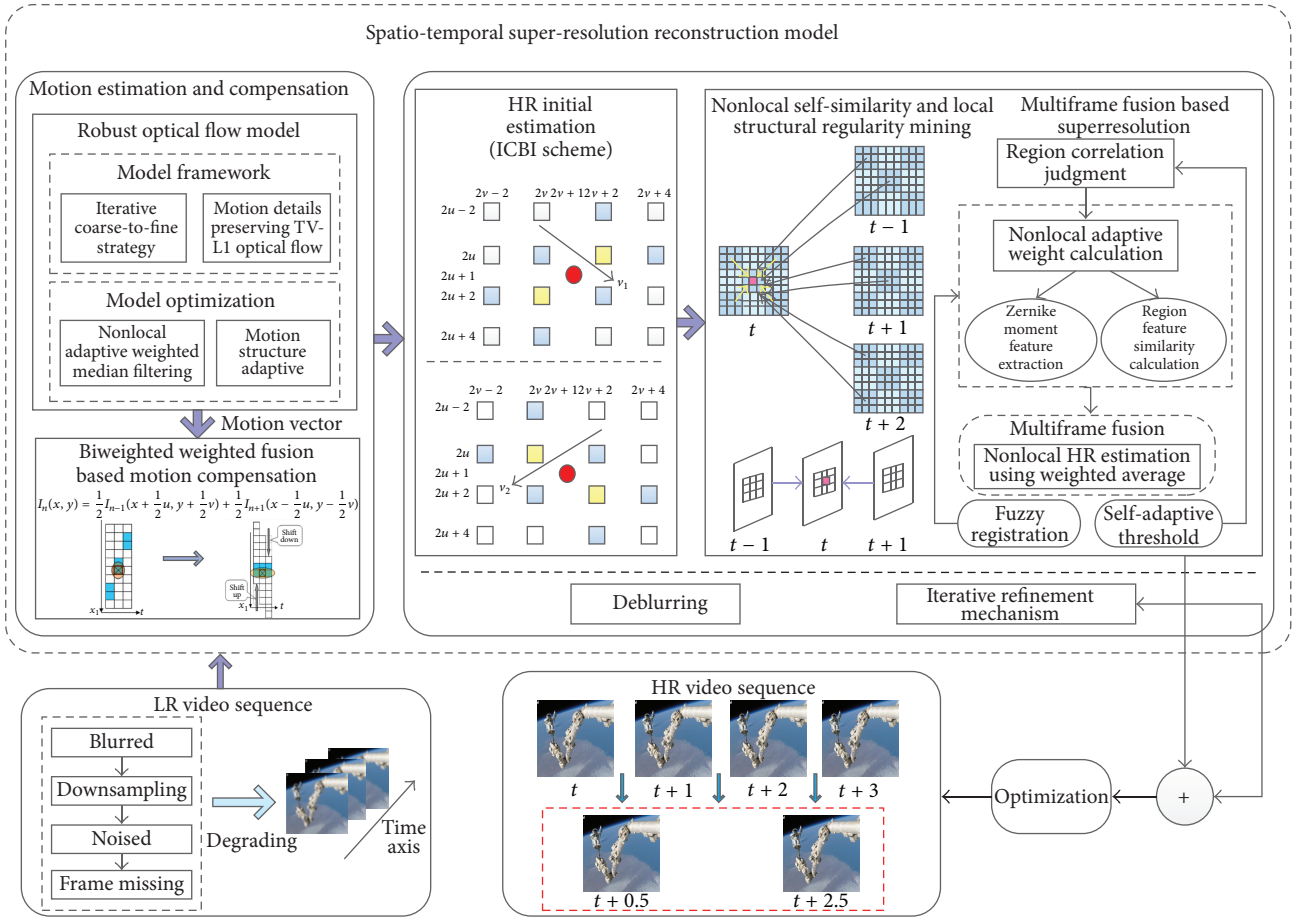


FIGURE 2: Architecture of the proposed model.

This process mainly includes three operations: initial HR estimation based on ICBI scheme, iterative multiframe fusion, and deblurring.

3. The Algorithm Implementation of the Model

3.1. Robust Optical Flow Motion Estimation Model Based on Motion Details Preserving. Owing to the brightness constancy and motion smoothness constraints, the traditional optical flow motion estimation methods are not usually robust to light noise, and lack the strong ability of motion details preserving [21]. In order to solve this problem, we propose a novel robust optical flow motion estimation model. In our model, the iterative multiresolution coarse-to-fine strategy and the Total Variation (TV) idea are applied in the model framework, which could effectively avoid falling into the local optimum and further improve the time efficiency as well.

Traditional optical flow calculation methods based on TV usually have higher computational complexity, and also when larger displacements exist in the video sequence, the error of motion estimation becomes larger. In order to overcome this problem, the iterative multi-resolution layered mechanism

based on Gaussian pyramid is introduced in our model to calculate optical flow from coarse to fine in the multi-resolution scales. The final motion vectors are obtained by adding the offset got in the higher solution to the motion vector got in the lower solution. The new mechanism cannot only effectively improve time efficiency, but can also obtain more reliable motion vectors.

Given two adjacent frames I_1 and I_2 in the video sequence, in the traditional TV based optical flow model, the process to obtain the motion vectors by optical flow motion estimation is equal to solving the minimum speed vector (u, v) of the objective energy function shown as follows:

$$\begin{aligned} E(u, v) &= E_d(u, v) + \lambda E_r(u, v) \\ &= \mu (I_2(x + u, y + v) - I_1(x, y)) + \lambda \times \eta (\nabla u, \nabla v), \end{aligned} \quad (1)$$

where $E_d(u, v)$ denotes the data item and $E_r(u, v)$ denotes the regularization item, which are calculated, respectively, as follows:

$$\begin{aligned} E_d(u, v) &= \sum_{x, y} |I_2(x + u_{x, y}, y + v_{x, y}) - I_1(x, y)|, \\ E_r(u, v) &= \sum_{x, y} (|\nabla u_{x, y}| + |\nabla v_{x, y}|). \end{aligned} \quad (2)$$

In the above calculation method, the minimum process for data term $E_d(u, v)$ is based on the brightness constancy constraint, so it usually makes the traditional model greatly influenced by some factors such as light, shadows, or occlusion. And for the regularization item $E_r(u, v)$, it is based on the motion smoothness constraint, but motion is often discontinuous near the contours or edges of the video frame, thus it usually makes the motion estimation perform worse for the motion discontinuity points.

Thus, on the basis of the traditional methods, we make improvements and optimization for the optical flow model, in order to further improve the model robustness and motion estimation accuracy. The specific details are shown as follows.

First, in order to enhance the model robustness to the factors such as light noise, we make improvement for the data item $E_d(u, v)$ in the optical flow model and construct a new data item with the combination constraints of brightness constancy and gradient constancy, which is calculated as follows:

$$\begin{aligned} E_d(u, v) &= \sum_{x,y} 0.5 \left| I_2(x + u_{x,y}, y + v_{x,y}) - I_1(x, y) \right| \\ &\quad + 0.5\zeta \left| \nabla I_2(x + u_{x,y}, y + v_{x,y}) - \nabla I_1(x, y) \right| \\ &= \sum_{x,y} 0.5 \left| I_x u_{x,y} + I_y v_{x,y} + I_t \right| \\ &\quad + 0.5\zeta \left| \nabla I_2(x + u_{x,y}, y + v_{x,y}) - \nabla I_1(x, y) \right|, \end{aligned} \quad (3)$$

where the parameter ζ is a weight adjustment factor between the two constraints. $I(x, y)$ denotes luminance value for the pixel (x, y) at time point t . I_x , I_y , and I_t denote partial derivatives on x , y , and t for $I(x, y)$, and (u, v) denotes motion vector obtained by optical flow estimation.

Second, in order to effectively protect the motion discontinuity and edge details, inspired by the ideas of [22, 23], a motion structure adaptive strategy is introduced in the regularization item of our optical flow model. The new improved regularization item is defined as follows:

$$E_r(u, v) = \sum_{x,y} \omega(x, y) \left(|\nabla u_{x,y}| + |\nabla v_{x,y}| \right), \quad (4)$$

where $|\nabla u_{x,y}| + |\nabla v_{x,y}|$ is the traditional TV regularization operation. $\omega(x, y)$ is the adaptive weight to protect motion details, which is calculated as follows:

$$\omega(x, y) = \exp \left(-|\nabla I_1(x, y)|^k \right). \quad (5)$$

Results from large quantities of experiments show that when the parameter k is set to 0.8, the motion estimation obtains best performance.

Finally, a heuristic nonlocal median filtering item E_{WNL} is introduced in our model, which can make the optimization for optical flow motion vector on each level using the adaptive weighted median filtering. This process can be formulated as follows:

$$E_{WNL} = \sum_{x,y} \sum_{(i,j) \in N_{x,y}} \omega_{x,y,i,j} \left(|\hat{u}_{x,y} - \hat{u}_{i,j}| + |\hat{v}_{x,y} - \hat{v}_{i,j}| \right), \quad (6)$$

where $N_{x,y}$ is the set of neighbors of pixel (x, y) in a large nonlocal region. $\hat{u}_{x,y}$ and $\hat{v}_{x,y}$ denote the flow field estimate. $\omega_{x,y,i,j}$ is the adaptive weight factor for the pixel (i, j) . Inspired by the idea in [24], the weight $\omega_{x,y,i,j}$ is determined according to the spatial distance, the color-value distance, and the occlusion state, which is calculated as follows:

$$\omega_{x,y,i,j} = \frac{1}{25} \times \exp \left\{ -\frac{|i - i'|^2 + |j - j'|^2}{2\sigma_1^2} - \frac{|I(i, j) - I(i', j')|^2}{4\sigma_2^2} + \frac{o(i', j')}{o(i, j)} \right\}, \quad (7)$$

where $I(i, j)$ and $I(i', j')$ denote color vectors in the Lab space, $o(i, j)$ and $o(i', j')$ denote occlusion variables, and $\sigma_1 = 7$, and $\sigma_2 = 7$.

Owing to the pixels in the occlusion regions between the adjacent video frames lack correspondence, the motion vectors estimated from these regions are usually not accurate. Thus we need to make a further optimization for the estimated motion vectors by occlusion perception refinement. Comprehensively considering the flow difference and the pixel projection difference, we apply the following method to detect the occlusion region and solve the occlusion variable $o(i, j)$.

Consider

$$\begin{aligned} o(x, y) &= N(d(x, y), \sigma_d) \times N(e(x, y), \sigma_e), \\ d(x, y) &= \begin{cases} \text{div}(x, y), & \text{div}(x, y) < 0 \\ 0, & \text{others} \end{cases} \\ &= \begin{cases} \frac{\partial}{\partial x} u_{x,y} + \frac{\partial}{\partial y} v_{x,y}, & \text{div}(x, y) < 0 \\ 0, & \text{others} \end{cases} \\ e(x, y) &= I(x, y) - I(x + u_{x,y}, y + v_{x,y}), \end{aligned} \quad (8)$$

where $N(\cdot)$ follows nonnormal Gaussian prior assumption with zero mean. $d(x, y)$ denotes the flow difference; $e(x, y)$ denotes the pixel projection difference. In our experiment, $\sigma_d = 0.3$ and $\sigma_e = 20$.

Based on the above three points of optimization, we construct the objective energy function of our new optical flow model shown in (9). The final optical flow motion vector (u, v) with high accuracy is obtained by minimizing

$$E(u, v) = E_d(u, v) + \alpha E_r(u, v) + \beta E_{WNL}, \quad (9)$$

where parameters α and β are weight adjustment factor between $E_d(u, v)$, $E_r(u, v)$, and E_{WNL} .

In the optimization process using the heuristic nonlocal median filtering item E_{WNL} , we design a strategy to further improve the time efficiency for our optical flow model. Given the estimated optical flow, first we detect the motion boundaries using a Canny edge detector and then dilate these edges with a 5×5 mask to obtain flow boundary regions. In

these regions, we apply the adaptive weight in (7) in a 15×15 nonlocal window. But in the nonboundary regions, we apply the equal weight in a 5×5 window for the median calculation.

3.2. Spatiotemporal Motion Compensation Using the Biweighted Fusion Strategy. After the motion vector (u, v) is obtained using the optical flow model introduced in Section 3.1, the spatiotemporal motion compensation scheme is introduced to predict the intermediate missed frames in the video sequence, which aims to obtain their initial estimation.

The traditional motion compensation strategies based on single directional motion vectors usually produce some visual artifacts such as block effect, frame distortion, and motion blurring, which significantly influence the video visual quality. To some extent, some complex strategies can obtain better performance, but too much time cost. In order to obtain better compensation effects, and meanwhile improve the algorithm time efficiency, we introduce an efficient biweighted fusion strategy to make the spatiotemporal motion compensation. The predicted frame pixel energy value can be determined according to the following calculation method:

$$I_n(x, y) = \lambda_1 \times I_{n-1}(x + 0.5 \times u, y + 0.5 \times v) + \lambda_2 \times I_{n+1}(x - 0.5 \times u, y - 0.5 \times v). \quad (10)$$

Comprehensively, considering the algorithm time complexity and compensation accuracy, the parameters λ_1 and λ_2 are both set to be 0.5 in our experiment.

Based on Sections 3.1 and 3.2, the general framework of our proposed optical flow motion estimation and spatiotemporal motion compensation is shown in Figure 3.

3.3. Enhanced Spatiotemporal Super-Resolution Reconstruction Based on Robust Optical Flow and Zernike Moment. Through the experimental analysis, we can see that some artifacts of hole usually exist in the spatiotemporal predicted frames, which is mainly caused by the inevitable optical flow motion estimation errors. Thus in order to overcome this problem, through making full use of the nonlocal self-similarity and redundant information in the spatiotemporal domain between the adjacent video frames, we propose a fast fuzzy registration scheme based on ZM and then apply multiframe information fusion strategy to construct an efficient enhanced spatiotemporal super-resolution reconstruction model to reconstruct and optimize the predicted frames, which aims to obtain predicted frames with more pleasing visual quality and further improve the temporal resolution of the video sequence. Moreover, this new model can implement the spatial resolution reconstruction for the initial LR video sequence and finally obtain a video sequence with high spatiotemporal resolution.

Different from the traditional SR reconstruction approaches, the new scheme is not dependent on the accurate estimation of subpixel motion, which implements SR reconstruction by mining the nonlocal self-similarities between several adjacent video frames. Because of good

rotation, translation, and scale-invariance properties of ZM, the proposed model can perform better when some complex motion patterns exist in the video sequence such as arbitrary angles of rotations. What is more, ZM is not sensitive to noise, thus our model also has well noise robustness. The implementation of our proposed STSR model mainly includes the following three steps.

Step 1. An efficient iterative curvature-based interpolation scheme is introduced to obtain the initial HR estimation $\{Y_p[i, j, t]\}_{t=1}^T$, $p \in N$, of the LR video sequence $\{y'_p[i, j, t]\}_{t=1}^T$, $p \in N$, after motion compensation.

Step 2. On the basis of the initial HR estimation $\{Y_p[i, j, t]\}_{t=1}^T$, $p \in N$, each video frame is super-resolved using the multiframe information fusion strategy based on the proposed fast fuzzy registration scheme.

Step 3. The deblurring operation and iterative refinement mechanism are applied to optimize the super-resolved video sequence, which aims to further improve the reconstruction quality. And finally the high quality video sequence $\{x_k[i, j, t]\}_{t=1}^T$, $k \in N$, with high spatiotemporal solution is obtained.

In the first step, we introduce an efficient iterative curvature-based interpolation scheme (ICBI) [25] to provide better HR initial estimation for the second step, which would significantly influence the weight calculation in the first iteration of the next fusion process. Compared with traditional interpolation schemes, ICBI scheme used in our method, based on the continuity of the second-order derivatives and energy curvature, not only is simple and extremely effective in removing blurring or jaggy artifacts, but also has real-time performance. In this scheme, a rough estimate of the energy of each interpolated pixel $I(2u + 1, 2v + 1)$ is calculated as follows:

$$\begin{aligned} & I(2u + 1, 2v + 1) \\ &= \begin{cases} \frac{I(2u, 2v) + I(2u + 2, 2v + 2)}{2}, & v_1(2u + 1, 2v + 1) < v_2(2u + 1, 2v + 1), \\ \frac{I(2u + 2, 2v) + I(2u, 2v + 2)}{2}, & v_1(2u + 1, 2v + 1) \geq v_2(2u + 1, 2v + 1), \end{cases} \\ & v_1(2u + 1, 2v + 1) \\ &= I(2u - 2, 2v + 2) + I(2u, 2v) + I(2u + 2, 2v - 2) \\ &\quad - 3I(2u, 2v + 2) - 3I(2u + 2, 2v) + I(2u, 2v + 4) \\ &\quad + I(2u + 2, 2v + 2) + I(2u + 4, 2v), \\ & v_2(2u + 1, 2v + 1) \\ &= I(2u, 2v - 2) + I(2u + 2, 2v) + I(2u + 4, 2v + 2) \\ &\quad - 3I(2u, 2v) - 3I(2u + 2, 2v + 2) \end{aligned}$$

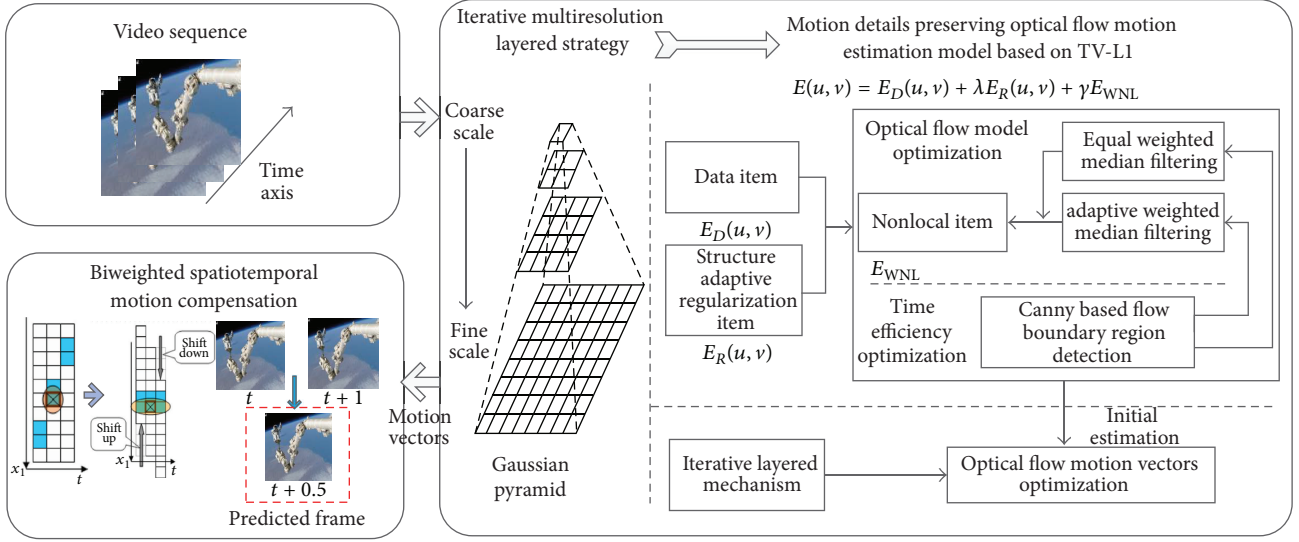


FIGURE 3: Framework of optical flow motion estimation and spatiotemporal motion compensation.

$$\begin{aligned}
 &+ I(2u - 2, 2v) + I(2u, 2v + 2) \\
 &+ I(2u + 2, 2v + 4),
 \end{aligned} \tag{11}$$

where $v_1(2u + 1, 2v + 1)$ and $v_2(2u + 1, 2v + 1)$ are local approximations of the second-order derivatives along the two diagonal directions using the eight neighboring pixels.

However, the pixel energy obtained above is only a rough estimate and we need continuous iterative refinement in some way. Following Giachetti and Asuni [25], the rough estimate of energy for each pixel is modified according to (12), and then we obtain an initial HR estimate of each LR video frame with higher quality.

$$\begin{aligned}
 E(2x + 1, 2y + 1) &= \alpha E_c(2u + 1, 2v + 1) \\
 &+ \beta E_e(2u + 1, 2v + 1) \\
 &+ \gamma E_i(2u + 1, 2v + 1),
 \end{aligned} \tag{12}$$

where α , β , and γ are adjustment factors that control the proportions of the curvature continuity energy E_c , the curvature enhancement energy E_e , and the isolevel curve smoothing energy E_i . We calculate these three energies using (5), (9), and (10) in [25].

Once the initial HR estimate of each LR video frame is obtained, the next step (Step 2) is the core of the algorithm, which is established on the basis of the fast fuzzy registration scheme based on ZM. We make weight calculation by mining the nonlocal self-similarity patterns between the frame to be super-resolved and each LR video frame, and then the HR estimation of the frame to be super-resolved could be obtained by weighted average. The weight $\omega_{\text{zer}}[k, l, i, j, t]$ of each pixel in the nonlocal neighboring region is calculated based on the similarity with ZM. On the basis of [17], to improve the time efficiency, we introduce a region correlation judgment strategy controlled by a self-adaptive threshold,

which yields a spatiotemporal adaptive model. Moreover, it is beneficial for mining the most similar patterns to calculate the weight based on similarity, and thus SR quality can also be improved to some extent. To describe our improved method, we first provide the following definitions.

Definition 1. A video frame F is divided into many regions of equal size, and each region is divided into 5×5 patches. The total number of pixels in each region is Num, and pixels energies are denoted by $p_1, p_2, \dots, p_{\text{Num}}$, respectively. We define $\bar{E}(x, y)$ as the average energy for the region centered on pixel (x, y) , calculated as

$$\bar{E}(x, y) = \sum_{i=1}^{\text{Num}} \frac{p_i}{\text{Num}}. \tag{13}$$

Definition 2. Given two regions centered on pixels (k, l) and (i, j) , denoted by $R(k, l)$ and $R(i, j)$, respectively, the corresponding feature vectors extracted from these two regions are $\text{ZM}(k, l)$ and $\text{ZM}^T(i, j)$. The feature similarity between these two regions is defined as

$$\begin{aligned}
 &\text{RFS}(R(k, l), R(i, j)) \\
 &= \exp \left\{ -\frac{\sum \|\text{ZM}(k, l) - \text{ZM}^T(i, j)\|_2^2}{\varepsilon^2} \right\},
 \end{aligned} \tag{14}$$

where $\text{ZM}(k, l)$ and $\text{ZM}^T(i, j)$ represent two ZM feature vectors for pixel (k, l) and pixel (i, j) in the nonlocal neighboring search region, calculated as

$$\begin{aligned}
 \text{ZM}(k, l) &= (\text{ZM}_{00}, \text{ZM}_{11}, \text{ZM}_{20}, \text{ZM}_{22}, \text{ZM}_{31}, \text{ZM}_{33}), \\
 \text{ZM}^T(i, j) &= (\text{ZM}_{00}^T, \text{ZM}_{11}^T, \text{ZM}_{20}^T, \text{ZM}_{22}^T, \text{ZM}_{31}^T, \text{ZM}_{33}^T),
 \end{aligned} \tag{15}$$

where the ZM with order n and repetition m ($0 \leq n \leq \infty, |m| \leq n, |n-m| = \text{even}$) of a video frame is defined as

$$\text{ZM}_{nm} = \frac{n+1}{\pi} \sum_x \sum_y V_{nm}^*(x, y) f(x, y),$$

$$\begin{aligned} V_{nm}(x, y) = & \sum_{k=|m|, |n-k|=\text{even}} (-1)^{(n-k)/2} \left[\frac{(n+k)}{2} \right]! \\ & \times \left(\left[\frac{(n-k)}{2} \right]! \left[\frac{(k-m)}{2} \right]! \left[\frac{(k+m)}{2} \right]! \right)^{-1} \\ & \times \rho^k e^{im\theta}, \end{aligned} \quad (16)$$

where $\rho = \sqrt{x^2 + y^2}$, $\theta = \tan^{-1}(y/x)$, $V_{nm}^*(x, y)$, represents the complex conjugate of $V_{nm}(x, y)$.

In our SR reconstruction model, the region correlation judgment is first applied to divide the regions centered on all

pixels (i, j) in the search region for pixel (k, l) into related and unrelated regions; only related regions are used to calculate the weight, which can further improve the time efficiency. For the region correlation judgment, a self-adaptive threshold δ_{adap} is introduced. If two regions are related, δ_{adap} is defined as

$$|\bar{E}(k, l) - \bar{E}(i, j)| < \delta_{\text{adap}}. \quad (17)$$

The self-adaptive threshold is adaptively determined by the average energy $\bar{E}(k, l)$ for the region centered on pixel (k, l) . This leads to more accurate judgment of region correlation. The self-adaptive threshold is calculated as

$$\delta_{\text{adap}} = \lambda \bar{E}(k, l), \quad (18)$$

where λ is an adjustment factor that controls δ_{adap} . Experiments confirmed that better SR quality is obtained when λ is set to 0.08.

Based on the above ideas, we construct the following enhanced weight calculation formula based on similarity calculation with ZM.

$$\begin{aligned} \omega_{\text{ezer}}[k, l, i, j, t] &= \frac{1}{C(k, l)} \times \text{RFS}(R(k, l), R(i, j)) \\ &= \begin{cases} \frac{1}{C(k, l)} \times \exp \left\{ -\frac{\sum \| \text{ZM}(k, l) - \text{ZM}^T(i, j) \|_2^2}{\varepsilon^2} \right\}, & |\bar{E}(k, l) - \bar{E}(i, j)| < \delta_{\text{adap}}, \\ 0, & \text{otherwise,} \end{cases} \quad (19) \\ C(k, l) &= \sum_{(i, j) \in N_{\text{nonloc}}(k, l)} \exp \left\{ -\frac{\sum \| \text{ZM}(k, l) - \text{ZM}^T(i, j) \|_2^2}{\varepsilon^2} \right\}, \end{aligned}$$

where (k, l) denotes the pixel to be super-resolved and (i, j) denotes the pixel in the nonlocal neighboring region $N_{\text{nonloc}}(k, l)$ of (k, l) . The parameter ε controls the decay rate of the exponential function and the weight. $C(k, l)$ is a normalization constant.

It is worth noting that the higher the ZM order is, the more sensitive it is to the noise. Thus, in our experiments we only calculated the first third-order moments, including $\text{ZM}_{00}, \text{ZM}_{11}, \text{ZM}_{20}, \text{ZM}_{22}, \text{ZM}_{31}$, and ZM_{33} .

When the weight $\omega_{\text{ezer}}[k, l, i, j, t]$ is determined, the HR estimation of each pixel in the video frame to be super-resolved can be obtained by the weighted average for the pixels in its nonlocal neighboring region of each LR video frame. Suppose that $Z = HX$; the objective function of the blurring HR estimate Z can be obtained by the following energy function:

$$\begin{aligned} \hat{Z}_{\text{SR}}^1[x] &= \sum_{(k, l) \in \Psi} \sum_{t \in [1, \dots, T]} \sum_{(i, j) \in N_{\text{nonloc}}(k, l)} \omega_{\text{ezer}}[k, l, i, j, t] y_t[i, j] \end{aligned}$$

$$\times \left(\sum_{(k, l) \in \Psi} \sum_{t \in [1, \dots, T]} \sum_{(i, j) \in N_{\text{nonloc}}(k, l)} \omega_{\text{ezer}}[k, l, i, j, t] \right)^{-1}, \quad (20)$$

where Ψ represents the video frame to be super-resolved.

Finally in the third step, we introduce a stronger adaptive Kernel regression (AKR) deblurring mechanism [26] to deal with the blur, which is applied in the results of the second step of multiframe fusion. The desired HR video frame X can be obtained by minimizing the following objective function:

$$\hat{Z}_{\text{SR}}^2[x] = \|Z - HX\|_2^2 + \gamma \text{AKR}(X), \quad (21)$$

where γ denotes the weighted parameter of the AKR deblurring process.

To further improve the SR reconstruction quality, the result needs to be iteratively refined for further optimization. The result after each iteration could provide basis for more accurate weight calculation in the next iteration.

TABLE 1: Evaluation for the optical flow motion estimation performance (without noise).

Video sequence	Performance evaluation	HS	BA	Classic-C	Classic-NL	Proposed
Rubber Whale	EPE	0.118	0.097	0.091	0.076	0.075
	AAE	3.798	3.156	2.947	2.401	2.342
	Time cost (s)	109.95	378.90	395.46	162.18	160.20
Grove2	EPE	0.204	0.181	0.176	0.169	0.164
	AAE	2.853	2.664	2.415	2.155	2.128
	Time cost (s)	173.64	194.63	193.51	133.04	128.83

TABLE 2: Evaluation for the optical flow motion estimation performance (with noise).

Video sequence	Performance evaluation	HS	BA	Classic-C	Classic-NL	Proposed
Rubber Whale	EPE	0.263	0.224	0.202	0.196	0.192
	AAE	8.814	7.530	6.682	6.553	6.537
	Time cost (s)	116.56	382.91	415.92	221.40	206.70
Grove2	EPE	0.236	0.182	0.164	0.119	0.112
	AAE	3.252	2.609	2.372	1.653	1.650
	Time cost (s)	126.57	415.29	496.72	435.95	386.56

4. Experimental Results and Analysis

4.1. Experimental Data Set and Evaluation Indices. In order to validate the effectiveness of our proposed model and algorithm, two groups of experiments were designed. In the first group of experiments, we made evaluation for our proposed optical flow motion estimation model in terms of the estimation accuracy and time efficiency and made comparison and analysis with some existing methods. And the second group of experiments was designed to evaluate our proposed spatiotemporal SR reconstruction model in terms of subjective visual evaluation and three objective quantitative indices, the peak signal-to-noise ratio (PSNR), mean structural similarity index (MSSIM), and root-mean-square error (RMSE). In the experiments, we used the spatial video sequence taken from YOUKU website (<http://www.youku.com/>) and the standard video sequences taken from <http://trace.eas.asu.edu/yuv/index.html> website.

The SR methods were assessed by subjective visual evaluation and three quantitative indices, the peak signal-to-noise ratio (PSNR), mean structural similarity index (MSSIM), and root-mean-square error (RMSE), calculated as follows:

PSNR

$$= 10 \log_{10} \frac{255^2}{(1/(M \times N)) \sum_{i=1}^M \sum_{j=1}^N (R(i, j) - F(i, j))^2} \text{ dB},$$

$$\text{SSIM}(x, y) = \frac{(2\eta_x\eta_y + e_1)(2\sigma_{xy} + e_2)}{(\eta_x^2 + \eta_y^2 + e_1)(\sigma_x^2 + \sigma_y^2 + e_2)},$$

$$\text{MSSIM}(x(i), \hat{x}(i)) = \frac{1}{L} \sum_{i=1}^L \text{SSIM}(x(i), \hat{x}(i)),$$

$$\text{RMSE} = \sqrt{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (R(i, j) - F(i, j))^2}, \quad (22)$$

where M and N denote the length and width of the video frame. R and F denote the reconstructed frame and the original frame, respectively. η_x and η_y are the mean. σ_x and σ_y are the standard deviation for the original and reconstructed frame, respectively. σ_{xy} is the covariance for the original and reconstructed frames. e_1 and e_2 are constants. L is the number of frame blocks. For greater PSNR, the reconstructed frame is closer to the original. The closer MSSIM ($0 \leq \text{MSSIM} \leq 1$) is to 1, the greater the similarity is between the original and reconstructed frame structures. For smaller RMSE, the reconstructed frame is closer to the original.

4.2. Experimental Results. In the first group of experiments, our improved optical flow motion estimation model was assessed by two quantitative indices, average end-point error (EPE) and average angular error (AAE). And also the existing four methods (HS [27], BA [27], Classic-C [28], and Classic-NL [24]) were introduced for comparison with our approach. In this experiment, we chose two video sequences of Rubber Whale and Grove2 from the standard optical flow Middlebury database [29] to evaluate the performance of our optical flow model. The parameters ζ , α , and β in our optical flow model are set to 1, 10^2 , and 1, respectively. The EPE and AAE values and the time efficiency for the two sequences in the different optical flow methods under the conditions without noise and with noise are shown in Tables 1 and 2, respectively. The results from Table 1 show that compared with some existing methods, the overall performance of the proposed method is optimal, which obtains lower average EPE and AAE values and smaller time cost. From results shown in Table 2, we can see that the proposed model has good noise robustness, which can also perform better under the conditions with noise disturbance.

Figure 4 gives the optical flow maps and motion vector graphs for the spatial sequences which were obtained by the proposed optical flow model. As can be seen from the optical flow maps and flow vectors shown in Figure 4, the proposed

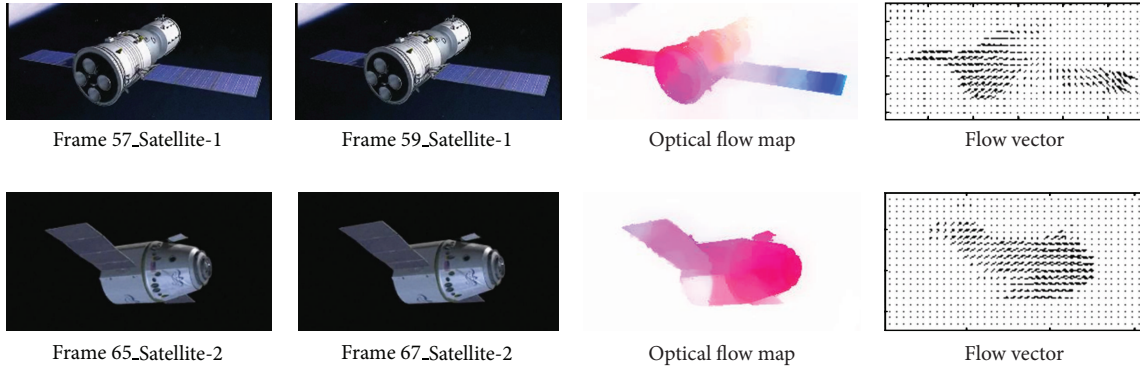


FIGURE 4: Optical flow map and motion vector graph of the spatial sequences.

TABLE 3: Average PSNR, MSSIM, and RMSE values in the four SR reconstruction methods.

Video sequence	Evaluation index	SR reconstruction methods			
		POCS	NL-SR	ZM-SR	STSR (proposed)
Satellite-1	PSNR (/dB)	31.5435	32.2815	32.5435	34.2740
	MSSIM	0.7621	0.7747	0.8032	0.8530
	RMSE	0.0054	0.0056	0.0049	0.0032
Satellite-2	PSNR (/dB)	20.5592	27.5057	28.6850	32.1723
	MSSIM	0.8915	0.8979	0.9072	0.9483
	RMSE	0.0146	0.0079	0.0071	0.0051
Forman	PSNR (/dB)	34.5542	35.1881	35.3928	36.6380
	MSSIM	0.6533	0.8051	0.8171	0.8491
	RMSE	0.0054	0.0049	0.0046	0.0038
Suzie	PSNR (/dB)	41.1444	44.3360	44.7565	45.4872
	MSSIM	0.7031	0.7974	0.8006	0.8236
	RMSE	0.0077	0.0055	0.0054	0.0046

TABLE 4: The SR time efficiency in the two SR reconstruction methods.

Video sequence	ZM-SR	STSR (proposed)
Satellite-1 (color)	91.43 s	42.78 s
Satellite-2 (color)	57.90 s	38.40 s
Suzie (gray)	19.83 s	10.97 s
Forman (color)	71.16 s	40.01 s

method can accurately detect the motion area of spatial target and also preserve the motion edges well, thus it is effective to use it for the motion estimation of spatial sequences.

In the second group of experiments, in order to verify the performance of the proposed SR reconstruction model and algorithm, three experiments were designed for comparison of our method (STSR) with existing approaches, POCS, NL-SR [13], and ZM-SR [17]. In our method, to obtain better SR reconstruction results and also improve time efficiency, we applied STSR to each frame using its adjacent six frames in the video sequence. In our STSR method, the region size for the weight calculation is 3×3 , and the optimal search region size is set to 5×5 .

In Experiment 1, we used two spatial video sequences of Satellite-1 (20 frames/s, 592×256 /frame) and Satellite-2 (20 frames/s, 640×346 /frame) and two standard video

sequences of Forman (20 frames/s, 352×288 /frame) and Suzie (20 frames/s, 351×240 /frame), each of which was blurred using a 3×3 uniform mask, decimated by a factor of 1:2 (for each frame), and then contaminated by additive Gaussian noise with $\sigma = 2$. And meanwhile, the even frames were gotten rid of, and then the frame rate was compressed to 10 frames/s. Then by the two times spatial and temporal super-resolution reconstruction, we tried to reconstruct the HR video sequences from the LR video sequences.

Figure 5 gives PSNR, MSSIM, and RMSE values of the POCS, NL-SR, ZM-SR, and STSR methods for the four video sequences. The average PSNR, MSSIM, and RMSE values of the four methods are shown in Table 3. Table 4 shows the time cost of the multiframe fusion process in the ZM-SR and STSR methods. It lists the average SR time for each iteration of each video frame in the spatial sequences and the standard

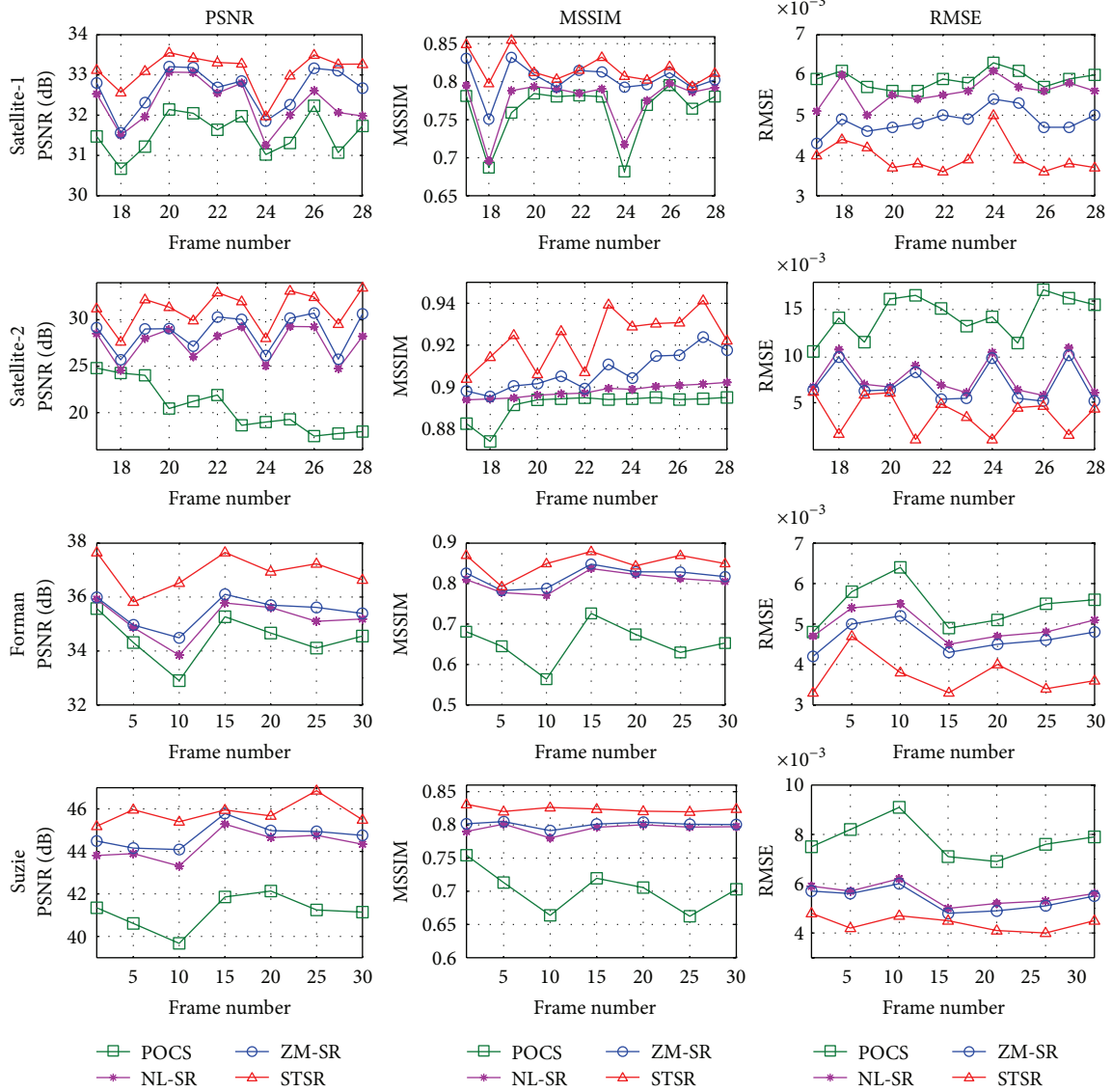


FIGURE 5: PSNR, MSSIM, and RMSE values in the four SR reconstruction methods.

Forman and Suzie sequences. Through a further analysis from the experimental results, we can see that compared with traditional methods, the proposed STSR method shows obvious advantages. On one hand, it yields better results with higher PSNR and MSSIM values and lower RMSE values. On the other hand, it has higher time efficiency, which only costs half of that for ZM-SR. The main reason is the region correlation judgment strategy controlled by a self-adaptive threshold. On one hand, it is beneficial for mining the most similar patterns to calculate the weight based on a similarity measure, and on the other hand, only the most related regions are used for the weight calculation rather than all the regions.

Figure 6 gives the visual effects of the spatial resolution reconstruction in the four different methods (POCS, NL-SR, ZM-SR, and STSR) for the four video sequences. From Figure 6 we can see that, because of relying on the accurate estimation of subpixel motion, POCS method is influenced

by the errors of motion estimation, which produces some ghosting phenomenon (see the red rectangular box), so it is not adaptive to the sequences with some complex motion patterns. Compared with POCS, NL-SR, ZM-SR, and STSR perform better in the SR reconstruction, because of that they do not rely on the accurate estimation of subpixel motion and also have a denoising effect to some extent. However, compared with NL-SR and ZM-SR which usually produce blur and jagged effects, such as some textures in the sequences (see the local textures marked in the red rectangular box), the proposed STSR method yields more pleasing visual effects that contain richer details and clearer edges and contours.

Figure 7 gives the visual effects of the temporal resolution reconstruction in the two different schemes (ICBI and STSR) for the two spatial video sequences, which aim to predict and reconstruct the missed or distorted video frames. From results shown in Figure 7, we can see that because of the

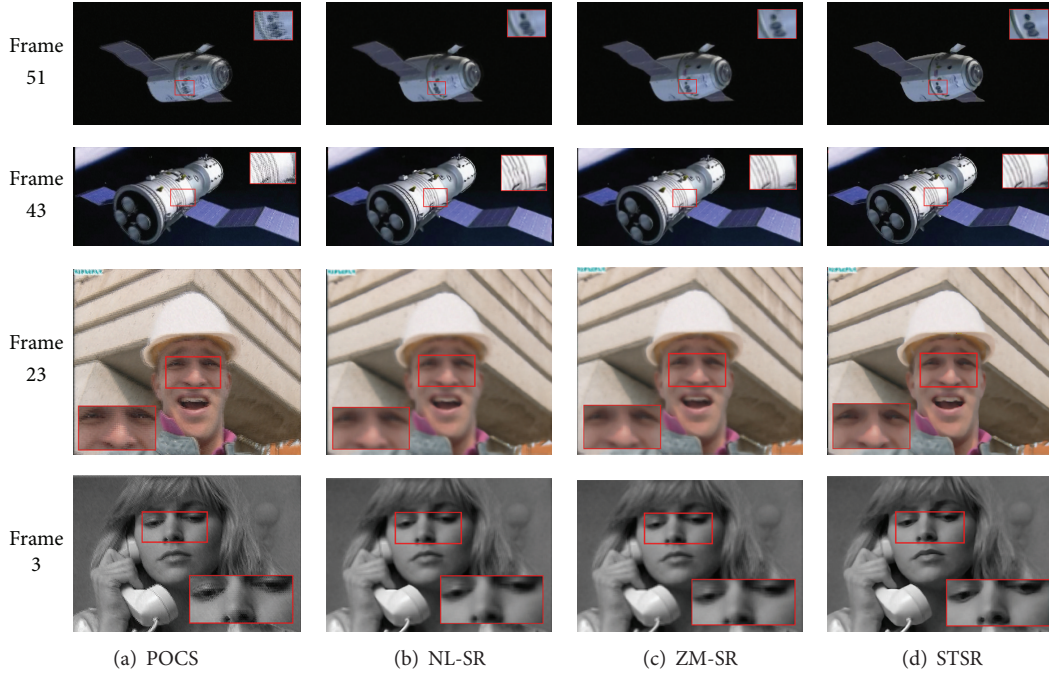


FIGURE 6: Visual effects for the spatial super-resolution reconstruction in the STSR model.

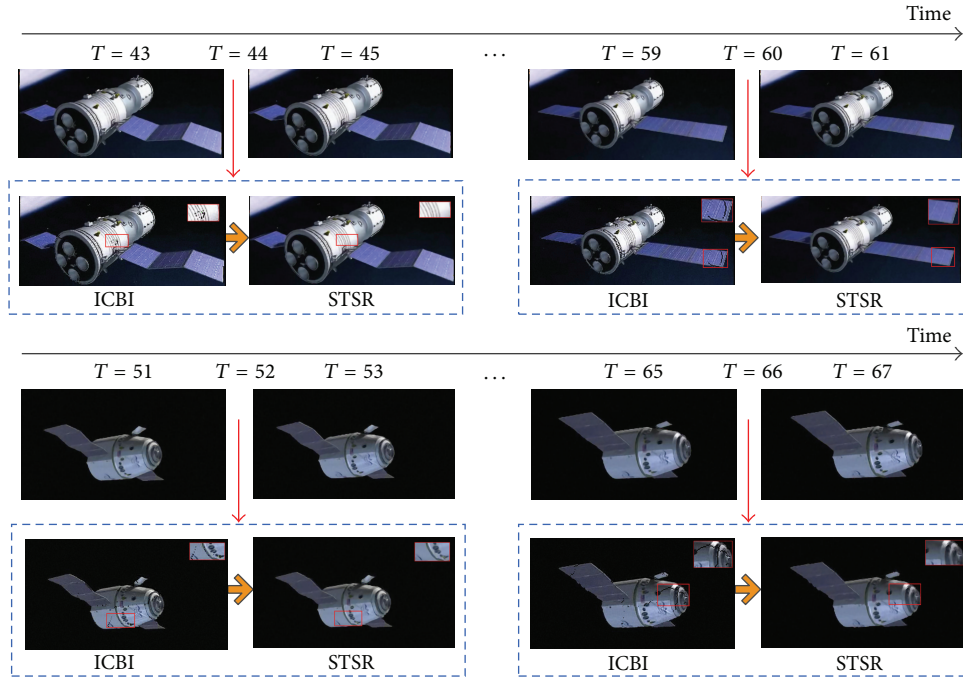


FIGURE 7: Visual effects for the temporal super-resolution reconstruction in the STSR model.

influence of noise and optical flow motion estimation errors, the traditional single frame based ICBI scheme usually produces some black hole effects in the reconstructed frames (see the local texture effects marked in the red rectangular box). But compared with ICBI, our method can effectively overcome this problem, and the reason mainly lies in the fact that we applied the multiframe information fusion strategy

for SR reconstruction by making full use of the nonlocal self-similarity and redundant information in the spatiotemporal domain between the adjacent video frames.

In Experiment 2, we used the two spatial sequences of Satellite-1 and Satellite-2 to test the effectiveness of the noise robustness and rotation invariance of the proposed method and made comparison with traditional POCS, NL-SR, and

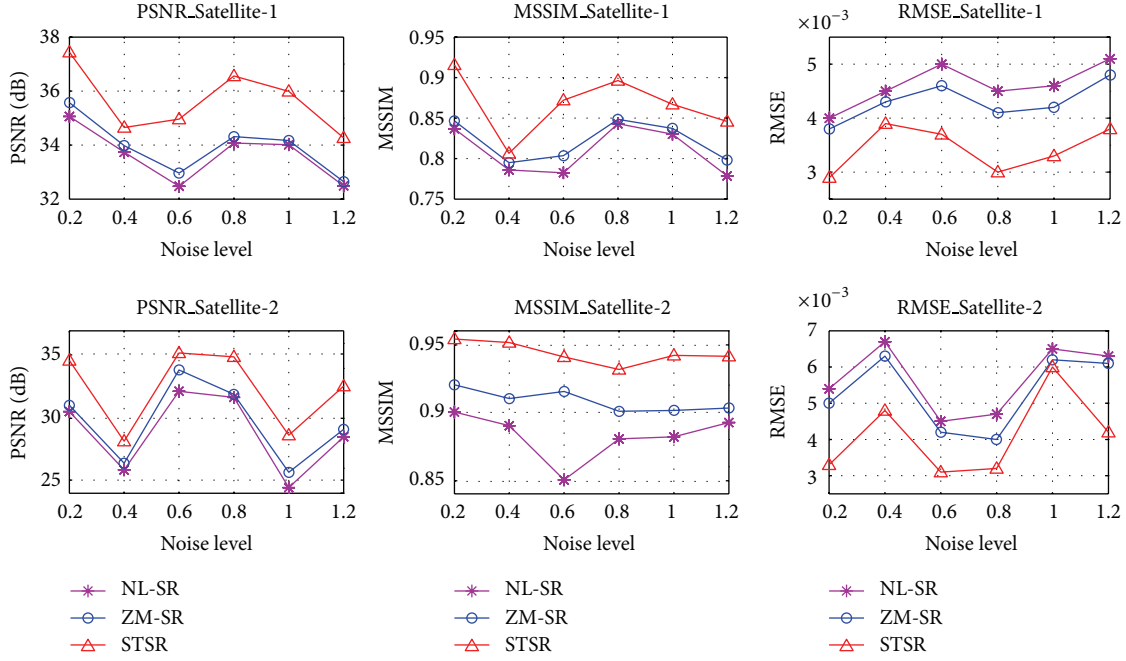


FIGURE 8: PSNR, MSSIM, and RMSE values of the spatial video sequences under different noise levels.

TABLE 5: PSNR, MSSIM, and RMSE values for STSR model under different noise models.

Video sequence	Gaussian noise			Poisson noise			Mixed Poisson-Gaussian noise		
	PSNR	MSSIM	RMSE	PSNR	MSSIM	RMSE	PSNR	MSSIM	RMSE
Satellite-1	34.2740	0.8530	0.0032	36.6014	0.9093	0.0032	32.6448	0.7947	0.0042
Satellite-2	32.1723	0.9483	0.0051	33.8618	0.9989	0.0036	27.9808	0.8948	0.0072
Forman	36.6380	0.8491	0.0038	39.2632	0.9335	0.0027	37.4941	0.8729	0.0033
Suzie	45.4872	0.8236	0.0046	44.8339	0.8277	0.0052	44.9418	0.8244	0.0051

ZM-SR methods. The two sequences of Satellite-1 (Frames 55–60) and Satellite-2 (Frames 62–67) were blurred using a 3×3 uniform mask, decimated by a factor of 1:2 (for each frame) and contaminated by additive noise with mean 0 and standard deviation 0.2, 0.4, 0.6, 0.8, 1.0, and 1.2; then some frames were processed with a slight angle. Moreover, the even frames were gotten rid of. Then we made the two times spatial and temporal SR reconstruction for the two sequences. PSNR, MSSIM, and RMSE values for the POCS, NL-SR, ZM-SR, and STSR methods under different noise levels are shown in Figure 8, which demonstrate that, compared with traditional methods, STSR method with a filter parameter of $\varepsilon = 10$ shows better performance, with higher PSNR and MSSIM values and lower RMSE values for all noise levels tested. Results for the first two experiments demonstrate that the STSR method performs better regardless of whether rotations occur or not. Thus the performance of our method shows that it has higher rotation invariance effectiveness and also is not sensitive to noise.

Furthermore, we made Experiment 3 to test the performance of our proposed STSR model under some different noise models (Gaussian, Poisson, and mixed Poisson-Gaussian) for the sequences of Satellite-1, Satellite-2, Forman, and Suzie, and the experimental results are shown in Table 5.

PSNR, MSSIM, and RMSE results shown in Table 5 demonstrate that our proposed method yields better performance for Gaussian, Poisson, and mixed Poisson-Gaussian noise. Thus we can see that our method can be applied to some other noise models except the white Gaussian noise model and can also perform better.

5. Conclusions

A novel model and algorithm are proposed in this paper to implement the spatiotemporal super-resolution reconstruction for the video sequences. In our model, a motion details preserving based optical flow motion estimation model is first proposed to obtain the motion vectors, and then an efficient biweighted fusion strategy is introduced to implement the spatiotemporal motion compensation. Then combining the good properties of rotation, translation, and scale-invariance of ZM, we propose a fast fuzzy registration scheme based on ZM by using the self-adaptive region correlation judgment strategy, and then the final video sequences with high spatiotemporal resolution can be obtained by fusion of the complementary and redundant information with nonlocal self-similarity between the adjacent video frames. Moreover, the new model integrates spatial SR and temporal SR into

a unified framework, which can improve both the spatial resolution and the temporal resolution and make the video sequences more clear and fluent. Different from the traditional SR reconstruction methods, the proposed method does not rely on accurate estimation of subpixel motion and can be adaptive to many kinds of complex motion patterns. Experimental results demonstrate that the proposed method outperforms the existing methods in terms of both subjective visual and objective quantitative evaluations and has higher rotation invariance effectiveness and noise robustness.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206), National Natural Science Foundation of China (nos. 91024001 and 61070142), and Beijing Natural Science Foundation (no. 4111002).

References

- [1] O. Shahar, A. Faktor, and M. Irani, "Space-time super-resolution from a single video," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 3353–3360, June 2011.
- [2] Y. An, Y. Lu, and Z. Yan, "Spatial-temporal motion compensation based video super resolution," in *Proceedings of the 10th Asian Conference on Computer Vision*, vol. 6493 of *Lecture Notes in Computer Science*, no. 2, pp. 282–292, 2011.
- [3] U. Mudenagudi, S. Banerjee, and P. K. Kalra, "Space-time super-resolution using graph-cut optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 995–1008, 2011.
- [4] J. Chen, J. Nunez-Yanez, and A. Achim, "Video super-resolution using generalized Gaussian Markov random fields," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 63–66, 2012.
- [5] S. P. Belekos, N. P. Galatsanos, and A. K. Katsaggelos, "Maximum a posteriori video super-resolution using a new multi-channel image prior," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1451–1464, 2010.
- [6] Q. Yuan, L. Zhang, and H. Shen, "Multiframe super-resolution employing a spatially weighted total variation model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 3, pp. 379–392, 2012.
- [7] C. Liu and D. Sun, "A Bayesian approach to adaptive video super resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 209–216, June 2011.
- [8] Z. Xiong, X. Sun, and F. Wu, "Robust web image/video super-resolution," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2017–2028, 2010.
- [9] M. Shimano, T. Okabe, I. Sato, and Y. Sato, "Video temporal super-resolution based on self-similarity," in *Proceedings of the 10th Asian Conference on Computer Vision*, vol. 6492 of *Lecture Notes in Computer Science*, no. 1, pp. 93–106, 2011.
- [10] J. Tian, T. Hou, and M. Li, "Spatio-temporal adaptive super-resolution reconstruction of video based on POCS frame," *Application Research of Computers*, vol. 28, no. 7, pp. 2778–2781, 2011.
- [11] A. L. D. Martins, A. L. M. Levada, M. R. P. Homem, and N. D. A. Mascarenhas, "MAP-MRF super-resolution image reconstruction using maximum pseudo-likelihood parameter estimation," in *Proceedings of IEEE International Conference on Image Processing (ICIP '09)*, pp. 1165–1168, November 2009.
- [12] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 1958–1975, 2009.
- [13] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 36–51, 2009.
- [14] N. Dowson and O. Salvado, "Hashed nonlocal means for rapid image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 485–499, 2011.
- [15] H. Zheng, A. Bouzerdoun, and S. L. Phung, "Wavelet based nonlocal-means super-resolution for video sequences," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 2817–2820, September 2010.
- [16] H. Su, L. Tang, Y. Wu, D. Tretter, and J. Zhou, "Spatially adaptive block-based super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1031–1045, 2012.
- [17] X. Gao, Q. Wang, X. Li, D. Tao, and K. Zhang, "Zernike-moment-based image super resolution," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2738–2747, 2011.
- [18] N. Vretos, N. Nikolaidis, and I. Pitas, "3D facial expression recognition using Zernike moments on depth images," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 773–776, September 2011.
- [19] C. Xu, Y. Chen, Z. Gao, Y. Ye, and T. Shan, "Frame rate up-conversion with true motion estimation and adaptive motion vector refinement," in *Proceedings of the 4th International Congress on Image and Signal Processing (CISP '11)*, pp. 353–356, October 2011.
- [20] N. Jacobson and T. Q. Nguyen, "Scale-aware saliency for application to frame rate upconversion," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2198–2206, 2012.
- [21] C. Lei and Y.-H. Yang, "Optical flow estimation on coarse-to-fine region-trees using discrete optimization," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 1562–1569, October 2009.
- [22] A. Wedel, D. Cremers, T. Pock, and H. Bischof, "Structure- and motion-adaptive regularization for high accuracy optical flow," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 1663–1668, October 2009.
- [23] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1293–1300, June 2010.
- [24] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2432–2439, June 2010.
- [25] A. Giachetti and N. Asuni, "Corrections to "Real-time artifact free image upscaling,"" *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2361–2369, 2012.
- [26] H. Takeda, S. Farsiu, and P. Milanfar, "Deblurring using regularized locally adaptive Kernel regression," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 550–563, 2008.
- [27] D. Sun, S. Roth, J. Lewis, and M. J. Black, "Learning optical flow," in *Proceedings of the 10th European Conference on Computer Vision (ECCV '08)*, vol. 5304 of *Lecture Notes in Computer Science*, pp. 83–97, Marseille, France, 2008.

- [28] A. Wedel, T. Pock, and C. Zach, "An improved algorithm for TV- L^1 optical flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis*, vol. 5604 of *Lecture Notes in Computer Science*, pp. 23–45, 2009.
- [29] S. Baker, S. Roth, D. Scharstein, M. J. Black, J. P. Lewis, and R. Szeliski, "A database and evaluation methodology for optical flow," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV '07)*, pp. 1–8, October 2007.

