Article

MAP-MRF-Based Super-Resolution Reconstruction Approach for Coded Aperture Compressive Temporal Imaging

Tinghua Zhang [1,2,*] and Kun Gao [1]

[1] School of Optics and Electronics, Beijing Institute of Technology, Beijing 100081, China; gaokun@bit.edu.cn
[2] Department of Optical and Electronic Equipment, Space Engineering University, Beijing 101416, China
* Correspondence: zth-gd@163.com; Tel.: +86-10-186-129-80800

Abstract: Coded Aperture Compressive Temporal Imaging (CACTI) can afford low-cost temporal super-resolution (SR), but limits are imposed by noise and compression ratio on reconstruction quality. To utilize inter-frame redundant information from multiple observations and sparsity in multi-transform domains, a robust reconstruction approach based on maximum a posteriori probability and Markov random field (MAP-MRF) model for CACTI is proposed. The proposed approach adopts a weighted 3D neighbor system (WNS) and the coordinate descent method to perform joint estimation of model parameters, to achieve the robust super-resolution reconstruction. The proposed multi-reconstruction algorithm considers both total variation (TV) and $\ell_{2,1}$ norm in wavelet domain to address the minimization problem for compressive sensing, and solves it using an accelerated generalized alternating projection algorithm. The weighting coefficient for different regularizations and frames is resolved by the motion characteristics of pixels. The proposed approach can provide high visual quality in the foreground and background of a scene simultaneously and enhance the fidelity of the reconstruction results. Simulation results have verified the efficacy of our new optimization framework and the proposed reconstruction approach.

Keywords: CACTI; TV; super-resolution; MAP-MRF; Belief Propagation (BP); WNS

1. Introduction

The information content within the entire 2D spatial and temporal extent of a dynamic scene cannot be captured and stored using conventional sampling strategies. Modern cameras have sufficient bandwidth to capture megapixel-scale resolution at limited frame rates. However, limited by the size of pixels and the reading circuit of the sensor, the imaging system may suffer motion aliasing caused by lower camera frame frequency or motion blur caused by longer exposure time, when tracking high-speed motion objectives. Compressive sensing (CS) strategies can take advantage of this knowledge by multiplexing high dimensional data stream onto detectors with fewer dimensions, both in terms of frame rate and spatial resolution. To deal with this problem, Coded Aperture Compressive Temporal Imaging (CACTI) [1] was proposed to improve the ability to catch fast-moving targets and serves as an effective measure to record and analyze the key events. It has become an important research direction in Compressive Sensing.

CACTI can afford low-cost temporal super-resolution based on the principle of optical field modulation and compressive sensing. The coded aperture is designed to bestow the Restricted Isometry Property (RIP) upon the observation model, to recover the images with high temporal, spectral, spatial or depth resolutions from observed images with limited resolutions; for example, the high-resolution imaging techniques such as focal plane coding, random phase modulation, and coded exposure imaging [2–4].

To meet the demand on temporal super-resolution imaging, Levin et al. [5] and Reskar et al. [6] applied the "flutter shutter" method to control the shutter within the exposure time with a binary pseudo-random sequence, thus turning the shutter into a broad-band filter. It was expected to overcome the motion aliasing and blur for sampling with low frame frequency, and retain more high-frequency information, to improve the imaging quality and frame frequency. However, the restoration in coded exposure imaging depends on point spread function (PSF) estimation and motion compensation.

Llull et al. proposed a coded aperture compressive temporal imager (CACTI) [1] that employs temporal mechanical translation of a passive coding element in an intermediate image plane to achieve low power space-time compressive measurement and to demonstrate CACTI's compression and decompression fidelity with 148-frame video reconstructions from single coded images captured with an experimental prototype [7]. It helps to achieve the high frame frequency for image sequences without the need to modify the performance standards of the imaging devices and integral time, nor of highly precise and fast exposure control, in comparison to hybrid cameras, camera arrays, and Digital Light Processing (DLP) and so on. Since it is more difficult to improve hardware performance, the CACTI has undoubtedly become more promising.

However, the reconstructing algorithm considerably affects the CACTI performance, and is closely related to the observing matrix, objective sparsity and compressing ratio. Xuejun Liao et al. [6] proposed the application of generalized alternating projection (GAP) in video compressive sensing, which adopts the weighted-$\ell_{2,1}$ norm to replace ℓ_1 norm for solving the convex optimization. The reconstruction error of GAP monotonically converges to zero, under a set of group-restricted isometry property (group-RIP) conditions. Thus, GAP is applicable for compressive sensing of natural images and videos. GAP features quick convergence, high calculating efficiency, and anytime convergence property. Most applications, particularly in real systems, utilize GAP to solve the compressive sensing problem in the transformation domain. However, the groups and weights are related to the anticipated importance of wavelet/discrete cosine transform (DCT) coefficients.

Roman Koller also verified the effectiveness of GAP and suggested the Constrained Least Squares (CLS) algorithm [8] using the constraint term of norm ℓ_2. In comparison with GAP and Orthogonal Matching Pursuit (OMP) [9], CLS can better balance reconstruction quality and speed. However, GAP has no obvious difference to CLS in terms of calculating efficiency, though it has much better peak signal to noise ratio (PSNR) performance. For OMP with similar PSNR evaluation, GAP is more efficient in calculation. Therefore, GAP still excels.

Yang J. et al. [10,11] proposed a Gaussian Mixture Model (GMM)-based inversion method that can reconstruct video frames with greater fidelity than other methods for certain synthetic and real data through the on-line or off-line dictionary learning. An online-updated GMM algorithm was robust in different scenes and motions, especially in simple movements. For fast and complicated movements, the offline-learned GMM algorithm with related training videos can yield promising results. In their comparison, however, the GAP algorithm had not been optimized for certain scenes and adopted only the default values for the weighting coefficient. Hence, it is necessary to choose the algorithm in accordance with the natural scenes. Additionally, the GMM algorithm is faster than GAP in parallel calculation, but slower in serial calculation.

The GAP algorithm is an undeniable success for CACTI. The GAP with transform regularization and GAP with total variation (TV) regularization algorithms are the state-of-the-art algorithms used in CACTI. The basic principles of these different algorithms are shown in Figure 1. The basic principle of GAP with transform regularization uses wavelet transform regularization in the spatial domain and DCT transform regularization in the temporal domain simultaneously [7], which is calculated only for the data from one observation, as shown in Figure 1b. However, the GAP algorithm cannot easily find a good transformation to be imposed on different data, and the selection of groups and weights affects the results significantly [12]. Moreover, this transformation will consume more computational time.

Therefore, Xin Yuan proposed the TV regularization method to optimize GAP [13]. By replacing the constraint term of weighting-$\ell_{2,1}$ norm with the TV regularization term, this method can help

realize the approximate performance of GAP, while simplifying it. However, the TV regularization was only verified against one kind of video image in simulated comparison, and only evaluated with fixed compression ratio. It has yet to be further proved with extensive experiments on compressive sensing of video with different sparsity and motion. Furthermore, the TV regularization is only imposed on the 2D image, which may result in information loss or over-smoothness of reconstructed images, as shown in Figure 1a.

While avoiding the image distortion, CACTI can considerably increase the temporal resolution. Nevertheless, the system requires the high-resolution mask or the spatial light modulator as an active coding element, which in turn reduces the luminous flux to the system, and thus renders the imaging results more sensitive to noise. Present methods conduct the video compression mainly in idealized conditions without sufficient consideration of the effect of sparsity and noises of natural scenes on the reconstruction algorithms.

This paper proposes a robust reconstruction approach to undermine the noise effect, and to stably adapt to a high compressive ratio. To utilize inter-frame redundant information and sparsity in multi-transform domains, a maximum a posteriori probability and Markov random field (MAP-MRF) model for CACTI is proposed. The proposed approach adopts the coordinate descent method to perform joint estimation on model parameters including spatiotemporal super-resolution images, motion vector flow, and temporal super-resolution image sequences.

The basic flow of the proposed approach is shown in Figure 1c, which is calculated for the data from multiple observations comparing the GAP algorithm. For simplicity, the basic flow is illustrated only with two observation data. Some close-ups marked by red or blue rectangles in each frame are enlarged and presented below the corresponding frames. The close-ups marked by red rectangles are still background and close-ups marked by blue rectangles are motion foreground. Compared with the existing GAP algorithm, the proposed approach can provide high visual quality in the foreground and background of a scene simultaneously and enhance the fidelity of the reconstruction results.
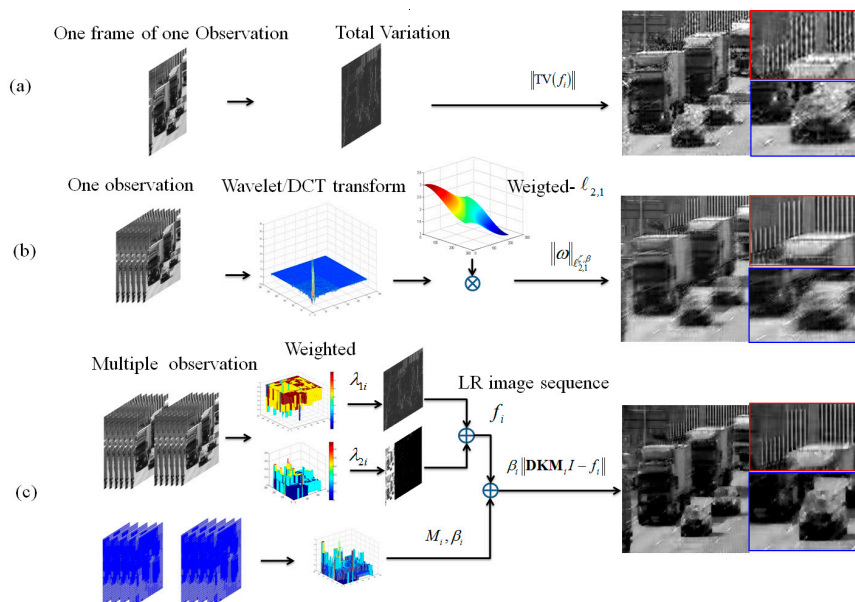

Figure 1. Basic flow of maximum a posteriori probability and Markov random field (MAP-MRF) based reconstruction approach. The close-ups marked by red rectangles represent the still background and the close-ups marked by blue rectangles represent the motion foreground. Proposed MAP-MRF-based algorithm uses more observation than other algorithms to utilize redundant inter-frame information. (**a**) the basic flow of generalized alternating projection (GAP) with total variation (TV) regularization; (**b**) the basic flow of GAP with wavelet and discrete cosine transform (DCT); (**c**) the basic flow of MAP-MRF-based approach.

The proposed reconstruction algorithm adopts multi-regularization constraints to reconstruct the video. It considers regularization with both total variation (TV) and norm in the wavelet domain into a minimization problem for compressive sensing, and solves it using a modified accelerated generalized alternating projection algorithm. The weighting coefficient for different regularization is resolved by the motion characteristics of pixels. Different from the existing GAP with transform regularization, the proposed multi-regularization algorithm does not need select the groups and group weights in the transform domain.

Moreover, most present algorithms only utilize the intra-frame information but not the redundant inter-frame information. Hence, this paper suggests the adoption of the sequence super-resolution (SR) algorithm to further improve the quality of the image sequence [14,15]. The reconstructed image sequence is estimated by the belief propagation algorithm based on the proposed Bayesian model for CACTI.

Table 1 summarizes frequently used notations in this paper. The paper has been organized into the following sections. Section 2 introduces the fundamental principles, forward model, and Bayesian model for CACTI image super-reconstruction; Section 3 introduces the proposal of a multi-regularization reconstruction algorithm and MAP-MRF-based reconstruction approach utilizing inter-frame redundant information; Section 4 describes the implementation of the proposed SR approach; and Section 5 involves the comparison and discussion of the proposed approach.

Table 1. Notations.

Symbols	Descriptions
$\mathbf{y}$	Observation of each low-speed coded snapshot1111
$\mathbf{x}$	Space-time data cube of the scene corresponding to $\mathbf{y}$
$\mathbf{\Phi}$	Observation matrix for CACTI
$\mathbf{n}$	Noises of the CACTI system
I	High-resolution image corresponding to one frame of $\mathbf{x}$
$\mathbf{D}$	Matrices correspond to down-sampling
$\mathbf{K}$	Matrices correspond to filtering with blur kernel K
$\mathbf{M}_i$	Warping matrix corresponding to motion vector
$Z_I(\eta), Z_\omega(\lambda), Z_K(\xi), Z_f(\zeta)$	Normalization constants respectively, depending on $\eta, \xi, \lambda, \zeta$ only
N_f	Compressive ratio in the temporal domain
$\mathbf{F}$	Matrices correspond to the Hadamard product
$\mathbf{w}$	Scale-invariant feature transform (SIFT) Flow velocity field of image sequence, $\mathbf{w} = [u, v]$
$\mathbf{H}$	Transformation matrix corresponding to Total Variation
η	Weight coefficients for Alternating Direction Method of Multipliers
β, λ	Weight coefficients for Accelerated Linearized Bregman algorithm
η_1, η_2	Weight coefficients for total variation and motion consistency SIFT Flow
α, d	Parameter for distance transform function used in SIFT Flow
λ_1, λ_2	coefficients for multiple regularizations

2. Coded Aperture Compressive Temporal Imaging (CACTI)

CACTI utilizes mechanical translation of a passive coding element to compress high-speed temporal information into low-frame rate video sequences via a process known as Code Division Multiple Access (CDMA). To explain the measurement process mathematically, as shown in Figure 1, the scene can be represented by a space-time data cube within one integration period T of the sensor. The high-speed spatiotemporal object voxels are patterned by a transmission function that shifts in time. Furthermore, distinct local coding structures are applied to each temporal channel prior to integration of the channels as limited-frame rate images on the detector. A high-speed estimation of scene may be reconstructed from each low-speed coded snapshot.

2.1. Forward Model for CACTI

The forward model for CACTI is shown in Figure 2b. The space-time data cube of the scene is herein supposed to be $x' f_0(x', y', t)$ at the coded aperture. The data cube after the coded aperture is:

$$f_1(x', y', t) = T(x', y', t) f_0(x', y', t) \tag{1}$$

where $T(x', y', t)$ is the transmission function that represents the coded aperture shifting in time. The coding element and the detector pixel are of the same size Δ. Hence:

$$T(x', y', t) = \sum_{i,j,t} t_{i,j,t} \text{rect}\left(\frac{x}{\Delta} - i, \frac{y}{\Delta} - j\right) \tag{2}$$

In the equation, $t_{i,j,t}$ stands for the binary value at element (i, j) of the coded aperture at time t. After further propagation through the relay lens, the data cube at the detector plane is:

$$\begin{aligned} f_2(x, y, t) &= \iint f_1(x', , y', t) h(x' - x, y' - y) dx' dy' \\ &= \iint T(x', y', t) f_0(x', y', t) h(x' - x, y' - y) dx' dy' \end{aligned} \tag{3}$$

where $h(x' - x, y' - y)$ represents the shift-invariant optical impulse response of the relay lens.

The detector conducts integration of the changing scenes within the exposure time. The pixels of the detector array may induce the spatial discrete function as:

$$p(m, n; x, y) = \text{rect}\left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n\right) \tag{4}$$

As a result, the measurement at the pixel (m, n) of detector array is:

$$g_{m,n} = \int_T \iint f_2(x', y', t) p(m, n; x, y) dx' dy' dt \approx \sum_{i,j,k} t_{i,j,k} f_{i,j,k} \Omega_{i,j,m,n} \tag{5}$$

In the result, the function $\Omega_{i,j,m,n}$ represents the spatial position of light, and does not vary with the time; $t_{i,j,k}$ is the corresponding code model for the sub-image frame k, and $f_{i,j,k}$ is the corresponding scene. The maximal value of k should be the largest frame amount in the compression for one time period of observation. This relates to the compressive ratio of the system. In consideration of noises, the linear matrix can be written as:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n} \tag{6}$$

where $\mathbf{n}$ stands for the imaging noises of the system, and $\mathbf{\Phi}$ is the time-variant transfer function as:

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_{N_f} \end{bmatrix} \tag{7}$$

$$\mathbf{\Phi}_k = diag\begin{bmatrix} T_{1,1,k} & T_{2,1,k} & \cdots & T_{\sqrt{N},\sqrt{N},k} \end{bmatrix}, k = 1, \cdots, N_f \tag{8}$$

where N_f is the number of video frames reconstructed from one low-speed coded snapshot and N_f also represents the compressive ratio in the temporal domain.
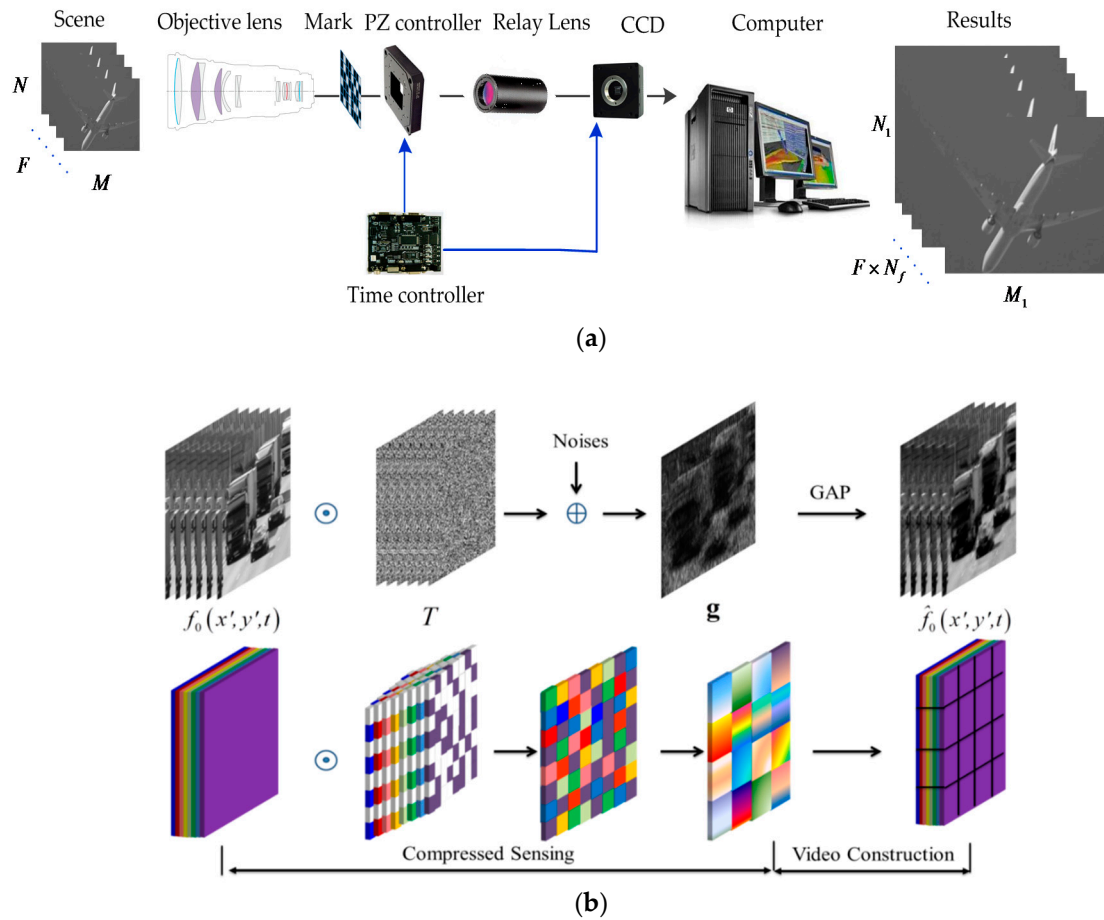
Figure 2. The illustration of the coding mechanisms within the coded aperture compressive temporal imaging (CACTI) system. (**a**) The system structure of CACTI; (**b**) the measuring principle of CACTI.

In the forward model of CACTI, matrix $\mathbf{\Phi}$ serves as the observation matrix in compressive sensing. The choice of the observation matrix is important to the reconstructed image quality irrespective of the selected reconstruction algorithm. To ensure the pseudo-randomness of the aperture code on all the moments of moving, the 50% random binary matrix was adopted to maintain the resolution of the coded aperture mask as the same as that of the sensor [1,16]. Reddy et al. [17,18] suggested the Gaussian mask with threshold, because the observation was most alike to the dense Gaussian matrix at that moment. Liu et al. [19] proposed the normalized Gaussian matrix to equalize the luminous flux for each pixel. This method could spare the decoding for the static area in the scene, but not meet the conditions for random coding. Roman Koller et al. [8] provided the Shifted-Normalized coding method, which combined the advantages of Llull's and Liu's methods [16]. For the same time periods of sampling, the luminous flux would remain the same for all pixels; the technique could be realized by the mask or liquid crystal light valve. In this paper, we used the Gaussian matrix for simplicity.

2.2. Bayesian Model for CACTI Super-Resolution (SR) Reconstruction

The Bayesian MAP [20] is adopted to carry out the joint estimation on high resolution (HR) image I, spatiotemporal transformation function M_i and super temporal resolution image sequence f_i as:

$$\{I^*, \{M_i\}^*, \{f_i\}^*\} = \underset{I, \{M_i\}, \{f_i\}}{\arg\max} \, p(I, \{M_i\}, \{f_i\} \,|\, \{y_i\}) \tag{9}$$

where the posterior is the product of prior and likelihood:

$$p(I, \{M_i\}, \{f_i\} | \{y_i\}) \propto p(I) \prod_i p(M_i) \prod_i p(f_i) p(y_0 | I, f_0) \prod_{i \neq 0} p(y_i | I, M_i, f_i) \tag{10}$$

Corresponding to motion vector matrix M_i, the motion vector is assumed to be separated into the horizontal component u_i and the vertical component v_i. With the gradient sparsity of motion vector, super temporal resolution image sequence, and the HR image as the image prior, the distribution of probability is:

$$p(I) = \frac{1}{Z_I(\eta)} \exp\{-\eta \|\nabla I\|\} \tag{11}$$

$$p(f_i) = \frac{1}{Z_K(\xi)} \exp\{-\xi(\|\nabla f_i\|)\} \tag{12}$$

$$p(M_i) = \frac{1}{Z_\omega(\lambda)} \exp\{-\lambda(\|\nabla u_i\| + \|\nabla v_i\|)\} \tag{13}$$

where ∇ is the gradient operator; $Z_I(\eta), Z_\omega(\lambda), Z_K(\xi)$ are the normalization constants respectively, depending on η, ξ, λ only.

With the known $p(\{f_i\})$ and $p(M_i)$ of the current frame, the prior probability distribution of the observed image can be presented as:

$$p(y_i | I, f_i, T_i) = \frac{1}{Z_f(\zeta)} \exp\{-\zeta \|y - \mathbf{DKM}_i \mathbf{F} I\|\} \tag{14}$$

where matrices $\mathbf{D}$ and $\mathbf{K}$ correspond to down-sampling and filtering with blur kernel K; $\mathbf{M}_i$ is the warping matrix corresponding to motion vector; matrices $\mathbf{F}$ correspond to the Hadamard (element-wise) product between the coding matrix and the frame in CACTI; $Z_f(\zeta)$ are the normalization constants respectively depending on ζ only.

After the probability distributions are obtained from both prior and likelihood, the method of coordinate descent is applicable for the solution.

The CACTI performance is directly related to the sparsity of objective scene, the velocity and manner of relative movement between the passive coding element and the pixel, the size of the coded element, as well as the velocity of the relative movement between the objective and the sensor. However, if these conditions are definite, the imaging results will be determined by the performance and robustness of the reconstruction algorithms.

3. Reconstruction Approach Based on Maximum a Posteriori Probability and Markov Random Field (MAP-MRF)

Based on the Bayesian model for CACTI, the model parameters are estimated by the coordinate descent method. The model parameters include a super temporal resolution image sequence, motion vector flow and high-resolution images. First, a robust MAP-MRF-based reconstruction approach is proposed, which renders the noises into total variation (TV) minimization problem for compressive sensing and solves it by accelerated linearized Bregman algorithm. This algorithm can perform preliminary estimation of the reconstruction image sequence, as the initial low resolution (LR) sequence shown in Figure 3. The estimation of motion vector for the weighted 3D neighbor system is estimated by scale-invariant feature transform (SIFT) Flow as shown in Figure 3. The proposed reconstruction algorithm utilizes the multi-regularization and the motion vector among different frames to reconstruct the frame. Furthermore, a super-resolution reconstruction approach using the Belief Propagation method is proposed to enhance the clarity and fidelity of the reconstruction results.

The SR reconstruction accuracy of image sequence is highly dependent on the estimation accuracy of the inter-frame motions between the LR observations. To address error propagation, we adopt a coarse-to-fine scheme that significantly improves the performance. The basic idea is to roughly

estimate the HR image, then gradually update motion estimation and reconstruct the HR image from coarse to fine using the multi-regularization algorithm and BP algorithm. Considering the computational complexity, we only estimate the two-scale resolution image sequence and update the motion estimation once time.
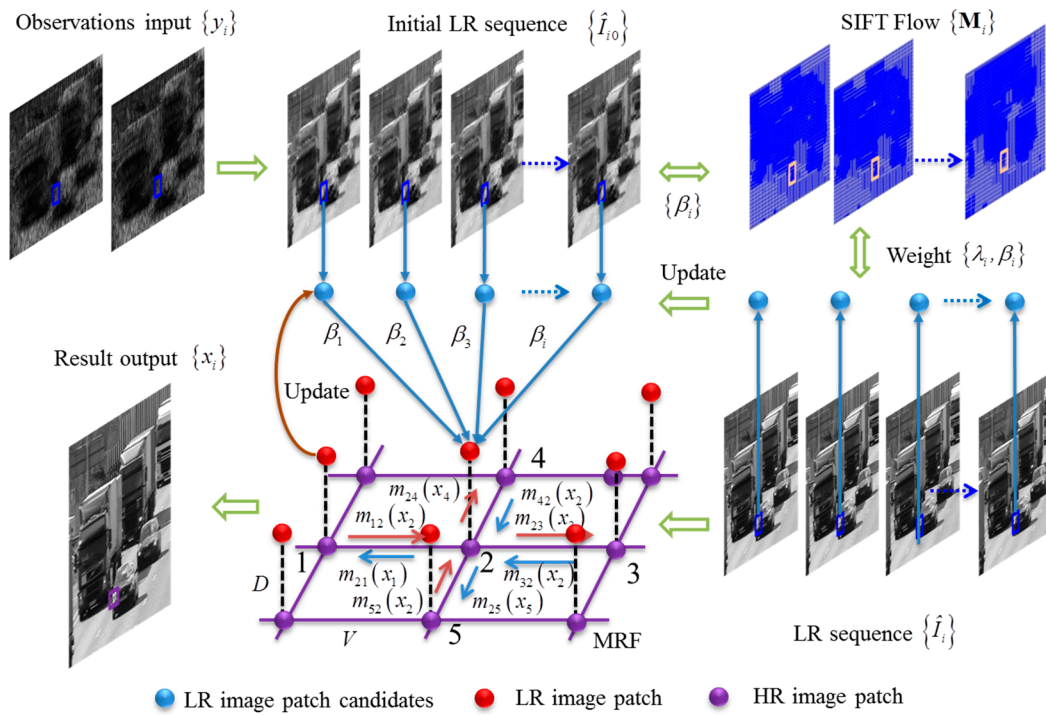

Figure 3. The framework of the MAP-MRF-based approach. The initial estimation of video sequence which is observed by multiple snapshot, and the low resolution (LR) image patch are candidates for MRF model. Then motion vectors field estimation by SIFT Flow, and which are used to calculate weighted coefficients of the 3D weight NS and multi-regularization. Then LR sequence are reconstructed by proposed multi-regularization algorithm, the LR image patch are candidates for MRF model. Finally, the high resolution (HR) image is constructed by Belief Propagation algorithm.

3.1. Preliminary Estimation of Reconstruction Image Sequence

In accordance with the forward model mentioned above, the N_f frames of video can be reconstructed from one observation of coded images y. The reconstructive algorithm requires Equation (6) to reach the results. For most algorithms, data fidelity needs to be enforced by using the quadratic term and the regularization term.

The total variation-based algorithms have demonstrated high performance on various CS problems, which can improve the fidelity and robustness of solvers [21,22]. The problem in Equation (6) can be represented as:

$$\min_{x}\|\mathrm{TV}(x)\| \leq C, \quad \text{subject} \quad \text{to} \quad \Phi x = y \tag{15}$$

where C is the radius of $\ell_1 - \text{ball}$ based on TV of the signal, and $\mathrm{TV}(x)\sum_{i,j} \sqrt{(\nabla_x x)_{i,j}^2 + (\nabla_y x)_{i,j}^2}$.

The alternating projection algorithm is adopted to solve the above equation:

$$\left(x^{(t)}, \theta^{(t)}\right) = \operatorname*{argmin}_{x,\theta} \frac{1}{2}\|x - \theta\|_2 + \lambda\|\mathrm{TV}(\theta)\|, \quad \text{subject} \quad \text{to} \quad \Phi x = y \tag{16}$$

where λ is the regulation coefficient and can be set as 0.07–0.15; t is the number of iteration times. Equation (16) can be solved by alternating the iteration algorithm with updated θ and x. For a given θ, the update of x is simplified to:

$$x^{(t+1)} := \underset{x}{\arg\min} \frac{1}{2} \|y - \mathbf{\Phi}x\|_2^2 + \frac{\eta}{2} \|x - \theta^{(t)}\|_2^2 \tag{17}$$

It can be converted into the Euclidean Projection problem directly or resolved with Alternating Direction Method of Multipliers (ADMM) algorithm [23]. The two methods are equivalent to each other as proven in literature [13]. For simplicity, the parameter can be set to values as 1.

$$x^{(t+1)} = \theta^{(t)} + \mathbf{\Phi}^T \left(\mathbf{\Phi}\mathbf{\Phi}^T \right)^{-1} \left(y - \mathbf{\Phi}\theta^{(t)} \right) \tag{18}$$

Given x, the sub-problem about updating θ can be regarded and solved as the denoising or deblurring problem [24]. Taking the noises into account, the sub-problem can be rewritten as:

$$\theta^{(t+1)} := \underset{x}{\arg\min} \|x^{(t+1)} - \theta\|_2^2 + \beta \|b - \text{TV}(\theta) - t\|_2^2 + \lambda \|\text{TV}(\theta)\| \tag{19}$$

The problem can be solved with the Accelerated Linearized Bregman algorithm, which accelerates the classical gradient descent method in that it reduces the iteration complexity significantly without increasing the per-iteration computational effort [25]. β is the weight and can be set as 0.15 for most applications.

This algorithm can be applied in three easy steps:

$$\theta^{(t+1)} := \underset{\theta}{\arg\min} \|x^{(t+1)} - \theta\|_2^2 + \beta \|b - \text{TV}(\theta) - t\|_2^2 \tag{20}$$

$$b^{(t+1)} := \underset{b}{\arg\min} \beta \|b - \text{TV}(\theta) - t\|_2^2 + \lambda \|\text{TV}(\theta)\| \tag{21}$$

$$t^{t+1} := t^t + \left(\theta^{t+1} - b^{t+1} \right) \tag{22}$$

Equation (22) is explicit and convenient for evaluation. Equation (20) is a differentiable optimization problem, which can be solved by Gauss-Seidel algorithm [26]. Equation (21) can be solved efficiently by using shrinkage as:

$$b^{t+1} := \max\left(\left| \nabla\theta + t^{t+1} \right| - \eta, 0 \right) \frac{\nabla\theta + t^{t+1}}{|\nabla\theta + t^{t+1}|} \tag{23}$$

Therefore, the preliminary estimation of reconfiguration results can be solved by Equations (18) and (20)–(23).

3.2. Estimation of Motion Vector by Scale-Invariant Feature Transform (SIFT) Flow algorithm

Based on the preliminary estimation of reconfiguration results, we can estimate the motion vectors between different frames. The typical optical-flow method may be applicable for motion estimation and compensation, but is faced with challenges from large-scale movement, sheltering judgment and processing, non-rigid movement, changing illumination, small object, blurred motion, noise of camera system, and so on [27]. To overcome the weaknesses in the optical-flow method, Ce Liu proposed the SIFT Flow velocity field theory [28] to estimate the dense correspondences across complex scenes such as illumination changes. This method can help to match the images involving different objects under varied visual angles, altered spatial positions, or scales.

Based on the Bayesian model and Equation (13), if the SIFT Flow is represented as $\mathbf{w}(p) = [u(p), v(p)]^T$ for point p, the SIFT Flow $\mathbf{w}(p)$ for point p should be expressed as

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\arg\min}\sum_p \left(\|I_1(p) - I_2(p + \mathbf{w}(p))\|_2^2, t \right) + \eta_1 \sum_p (|u| + |v|) \\ + \eta_2 \sum_{(p,q)\in\varepsilon} ((\alpha|u(p) - u(q)|, d) + (\alpha|v(p) - v(q)|, d)) \tag{24}$$

The first constraint term refers to the constancy of SIFT descriptors; the second to sparsity of motion vectors; the third to the motion consistency. Since the truncated ℓ_1 norms are adopted in this objective function, the sequential belief propagation of distance transform function may further ease the complexity and achieve better convergence. $\eta_1, \eta_2, \alpha, d$ are set to values as described in the original literature [28].

3.3. Improved Reconstruction Algorithm Using Multiple Regularization

To utilize inter-frame redundant information and sparsity in multiple transform domains, the proposed reconstruction algorithm adopts multi-regularization constraints to reconstruct the video. Reconstructing the images with multiple regularizations is equivalent to the following constraint problem:

$$\min_{\omega} \lambda_1 \|\omega\|_{\ell_{2,1}^\xi} + \min_x \lambda_2 \|\text{TV}(x)\|, \quad subject \quad to \quad \mathbf{\Phi}x = y \tag{25}$$

where y is the observed value, $\mathbf{\Phi}$ is the observation matrix, x is the signal to be estimated, ω is the coefficient corresponding to the transformation domain, and $\|\omega\|_{\ell_{2,1}^\xi}$ represents the weighting group of norm of ω, defined as $\|\omega\|_{\ell_{2,1}^\xi} = \sum_{k=1}^K \|\omega_{\zeta_k}\|_2$, and $\text{TV}(x) = \sum_{i,j} \sqrt{(\nabla_x x)_{i,j}^2 + (\nabla_y x)_{i,j}^2}$. The wavelet transform uses 3-level Haar wavelet and the temporal transform uses DCT.

The alternating projection algorithm is adopted to solve the above equation:

$$\left(x^{(t)}, \theta^{(t)} \right) = \underset{x,\theta}{\arg\min} \frac{1}{2}\|x - \theta\|_2 + \lambda_1 \|\omega(\theta)\|_{\ell_{2,1}^\xi} + \lambda_2\|\text{TV}(x)\|, \quad subject \quad to \quad \mathbf{\Phi}x = y \tag{26}$$

where λ_1, λ_2 is the regulation coefficient, and t is the number of iteration times. λ_1, λ_2 can be solved by:

$$\lambda_1 = \alpha^{|u|+|v|}, \lambda_2 = 1 - \alpha^{|u|+|v|} \tag{27}$$

where the motion vector is assumed to be separated into the horizontal component u and the vertical component v, α should be set to value as $0 \le \alpha \le 1$, and $\alpha = 0.7$ in this paper.

Equation (25) can be solved by alternating iteration algorithms with updated θ and x. For a given θ, the update of x is simplified to:

$$x^{(t+1)} := \underset{x}{\arg\min} \frac{1}{2}\|y - \mathbf{\Phi}x\|_2^2 + \frac{1}{2}\|x - \theta^{(t)}\|_2^2 + \lambda_2\|\text{TV}(x)\| \tag{28}$$

Given x, the update of θ is represented as the sub-problems:

$$\theta^{(t)} := \underset{\theta}{\arg\min} \frac{\eta}{2}\|x^{(t)} - \theta\|_2^2 + \lambda_1 \|\omega(\theta)\|_{\ell_{2,1}^\xi} \tag{29}$$

The accelerated GAP algorithm can speed up the convergence of its reconstruction error using adaptive adjustment. Equation (28) is solved by:

$$x^{(t+1)} = \theta^{(t)} + \mathbf{\Phi}^T \left(\mathbf{\Phi}\mathbf{\Phi}^T \right)^{-1} \left(y^{(t)} - \mathbf{\Phi}\theta^{(t)} \right) + \lambda_2\mathbf{H} \tag{30}$$

where $\mathbf{H}$ is the Transformation matrix of Total Variation, and:

$$y^{(t)} = y^{(t-1)} + \left(y - \mathbf{\Phi}\theta^{(t-1)}\right), \forall t \geq 1 \tag{31}$$

Equation (29) is further developed into the following sub-problems:

$$\theta_{\Gamma_k}^{(t)} = x_{\Gamma_k}^{(t)} \max\left\{1 - \frac{\lambda^{(t)}}{\left\|x_{\Gamma_k}^{(t)}\right\|_2^2}, 0\right\} \tag{32}$$

It has been proved that both grouping and weights play an important role for enhancing the reconstruction accuracy in GAP with weighted $\ell_{2,1}$ norm, which demonstrates the importance of exploiting the inherent structural sparsity of natural images. The groups seem to have yielded greater improvements than the weights, probably because the former capture the structural sparsity of solvers in a more direct way. To simplify the implementation of the multi-regularization algorithm, we only use the group when comparing with GAP using the weighted $\ell_{2,1}$ norm.

3.4. Spatiotemporal Super-Resolution Reconstruction of Image Sequences by Belief Propagation

Given the current motion vector and preliminary estimation of image sequence, the estimation can be conducted on HR images with the following equation as:

$$\hat{I} = \underset{I}{\arg\min} \sum_p \left(\|\mathbf{DK}I - f_0\| + \mu(\|\text{TV}(I)\|)\right) + \sum_{(p,q)\in NS} \left(\beta_i \|\mathbf{DKM}_i I - f_i\|\right) \tag{33}$$

where the first term is the data term incorporating the data distortion and TV total variation regularization, and the second is the penalty term. $\mathbf{D}$ is the down-sampling matrix, $\mathbf{K}$ is the blur kernel matrix, $\text{TV}(I)$ represents the total variation operator of I, $\mathbf{M}_i$ the matrix of flow $\mathbf{w}$ which can be calculated by optical flow or SIFT Flow methods, f_0 the image on current frame, f_i the preliminary estimation of image, and NS is the weighted neighborhood system where Point p exists. μ is the weight parameter which are set to values as $\mu = 0.7$.

However, there should be a way to reflect the contribution of different frames to the current super-resolution reconstruction, to demonstrate the spatiotemporal relationship in the process. Accordingly, the weighted neighborhood system is proposed herein based on the motion vector.

The image patches in Figure 3 represent these image patches from different frames. The space vector $\mathbf{v}_0^k$, which represents the motion vector between the k frame and its adjacent frame $k + N$ (N is the number of frames for SR reconstruction, and N can be larger than N_f in order to utilize the information between multi-observations y_i), can be obtained by the fitting of the motion vectors among different frames, such as $\mathbf{v}_1^k$, $\mathbf{v}_2^k$, $\mathbf{v}_3^k$ and $\mathbf{v}_4^k$:

$$\mathbf{v}_0^k = \mathbf{v}_1^k + \mathbf{v}_2^k + \mathbf{v}_3^k + \mathbf{v}_4^k \tag{34}$$

The inter-frame motion vectors of different frames can be projected onto the motion vectors between the current frame and the frame N, to mark the spatiotemporal proximity between the pixels. In this way, the weight coefficient can be settled for spatiotemporal SR reconstruction:

$$w_{ij} = \frac{\|\mathbf{v}_0^k \mathbf{v}_j^k\|}{\|\mathbf{v}_0^k\| \|\mathbf{v}_j^k\|} \tag{35}$$

Therefore, β_i is the weight factor of WNS and can be expressed as:

$$\beta_i = \omega_{fk}(p,q) = 1 - \xi \exp\left(-\gamma \left|\frac{\|\overrightarrow{v}_0^{\,k}\overrightarrow{v}_f^{\,k}\|}{\|\overrightarrow{v}_0^{\,k}\|\|\overrightarrow{v}_f^{\,k}\|}\right|\right) \tag{36}$$

where ξ is the normalization coefficient ensuring $\beta_i \subset [0,1]$; γ is the attenuation coefficient, which can be set as 1 for most application.

MAP-MRF can be adopted for the estimation on multi-frame reconstructed HR images and results in the energy minimization problem being suitable for inference algorithms based on graph cuts or belief propagation [29].

As an iterative method for the probabilistic inference of the probability graph model, the BP (Belief Propagation) algorithm is one of the key algorithms in solving MRF problems. In the BP algorithm, all the information is passed on in a parallel way, as shown in Figure 3. The nodes exchange information between one another to update the entire status of MRF. The algorithm works as the robust iterative approximate computation of MRF, which can reach the optimal solution through convergence.

The standard BP [30] form of Equation (33) is represented as:

$$E(f) = \sum_{p \in} D_p(f_p) + \sum_{(p,q) \in NS} V(f_p, f_q) \tag{37}$$

where, as shown in Figure 3, $D_p(f_p)$ is cost of assigning label f_p to pixel p, which is referred to the data cost as the term $\sum_p (\|\mathbf{DK}I - f_0\| + \mu(\|\mathrm{TV}(I)\|))$; $V(f_p, f_q)$ is cost of assigning label f_p and f_q to two neighboring pixels, and which is normally referred to as the discontinuity cost as the term $\sum_{(p,q) \in NS} (\beta_i \|\mathbf{DKM}_i I - f_i\|)$.

To initialize the belief networks, m_{pq}^t is defined as the message that node p sends to a neighboring node q at iteration t, and can be expressed as:

$$m_{pq}^t(f_q) = \min_{f_p}\left(V(f_p - f_q) + D_p(f_p) + \sum_{s \in NS(p)q} m_{sp}^{t-1}(f_p)\right) \tag{38}$$

The initial value of m_{pq}^0 is 0. $m_{pq}^t[0\ldots k]$ is the vector with the scale k. After T times of iteration, the belief vector b_q is calculated for each node:

$$b_q(f_q) = D_q(f_q) + \sum_{p \in NS(q)} m_{pq}^T(f_q) \tag{39}$$

Following this, there is the MRF stable solution $f_q^* = \min\{b_q(f_q)\}$ at each node.

The standard implementation of BP algorithm needs $O(nk^2T)$, where n is the number of pixels in the image, k is the number of possible labels for each pixel and T is the number of iterations. Besides current methods to speed up calculation, the possible label statuses, i.e., message length k, can be limited in order to reduce the calculating burden. In SR reconstruction, the initial value of HR image is obtained by the linear interpolation of current frame. The gray scale ranges within the changing gray-scale values in WNS. In this way, the computation will be considerably simplified.

4. Implementation of Improved Reconstruction Approach Based on MAP-MRF

The implementation of the proposed reconstruction approach described in Section 3 is shown in Figure 4. The key processes are the joint estimation of MAP-MRF model parameters using coordinate descent method, and the image reconstruction using multiple regularizations.
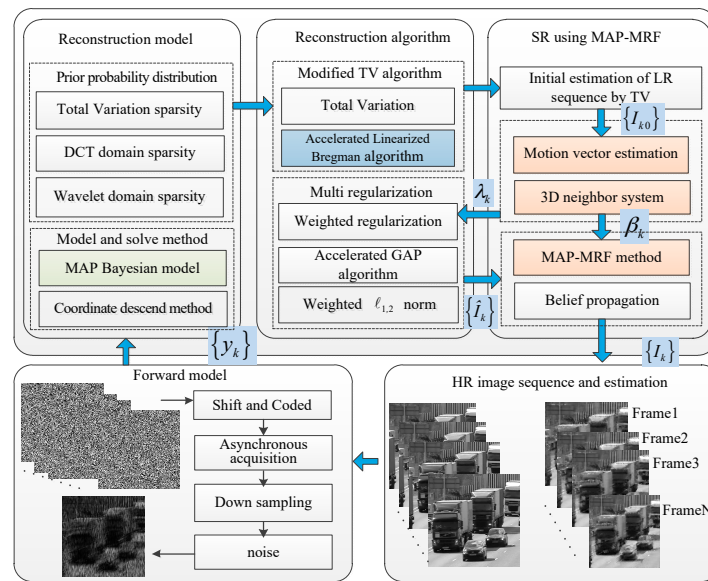
Figure 4. Framework of the proposed MAP-MRF-based reconstruction approach for CACTI.

In accordance with the aforementioned analysis, the procedures of MAP-MRF-based reconstruction may be summarized as follows Algorithm 1.

Algorithm 1. The procedures of MAP-MRF-based reconstruction.

Input: Observation image y, coding matrix Φ up-sampling factor s, numbers of frames for compressive imaging N_f, and the number of frames for SR reconstruction N.

Step 1: Estimate the preliminary reconfiguration results of multi-frames from the one observation y by proposed TV algorithm and Equations (18)–(23);

Step 2: Initialize times of iteration k, the initial estimation of HR image $I^{(0)}$ (bilinear interpolation method);

Step 3: judge the ending of iteration by $\left| I^{(k-1)} - I^{(k)} \right| < \varepsilon$ or $k \geq \mathrm{I}_{max}$, I_{max} the maximal times of iteration:

① 　Estimate motion vector $\omega_i^{(k)}$ by SIFT Flow algorithm as Equation (24);

② 　Estimate the frames $f_i^{(k)}$ by multi-regularization algorithm and Equations (26)–(32);

② 　Initialization: $m = 1$, $I^{(k,0)} = I^{(k-1)}$;

④ 　Loop until $\left| I^{(k,m)} - I^{(k,m-1)} \right| < \varepsilon$

　　　a. 　compute weight β_i by Equation (36);

　　　b. 　iterative estimation with BP algorithm by Equations (37)–(39);

　　　c. 　$m = m + 1$, returning to iteration;

⑤ 　$k = k + 1$

Output: $I = I^{(k)}$

---

5. Simulation and Discussion

We demonstrate the proposed reconstruction approach on a variety of scenes and compare the results with those of current leading reconstruction algorithms. We mainly analyze the performance of joint estimation of MAP-MRF model parameters and the reconstruction using multi-regularization. In Sections 5.1 and 5.2, we mainly verify the robustness of the proposed reconstruction algorithm with multi-regularization under different compression ratios and noise conditions. In Section 5.3, we verify the performance of SR reconstruction approach based on the MAP-MRF for CACTI image

reconstruction. The implementation and the ground truth have been adopted from the works of Xin Yuan et al. and http://www.disp.duke.edu/projects/CACTI/.

The performances of major existing algorithms, such as the GAP and GAP-acc + TV algorithm, are compared. The GAP algorithm has demonstrated excellent performance on compressive sensing problems, which utilizes generalized alternating projection to solve the compressive sensing problem in the transformation domain, that is, the wavelet or DCT domain. GAP + b signifies the GAP with wavelet/DCT transformation and block weight. GAP + TV algorithm utilizes generalized alternating projections to solve the compressive sensing problem with total variation, and GAP-acc + TV algorithm denotes the accelerated GAP + TV algorithm. "MAP-MRF-based SR" indicates the joint estimation of image sequences constructed by MAP-MRF-based approach, and "proposed algorithm" denotes the reconstruction algorithm with multi-regularization only in this section.

5.1. Reconstruction Performances under Different Temporal Compression Ratios

Based on the imaging principle of CACTI, the compression ratio, coded matrix, and motion feature affect the results significantly. The coded matrix and motion features are restricted by the hardware system and observation scene. The compression ratio is relevant to the stepping precision of the motion controller of the active coded element and the pixel size of the detector, which is the main specification of CACTI.

The simulation results of the proposed multi-regularization algorithm compared with other algorithms are shown in Figures 5 and 6. There are two videos synthesized for reconstruction. These synthetic videos differ in their features. The traffic video is dominated by a moving foreground, but the eye-blink video has more of a fixed background. The traffic video contains multiple moving objects, while the eyeblink video only contains one moving object.
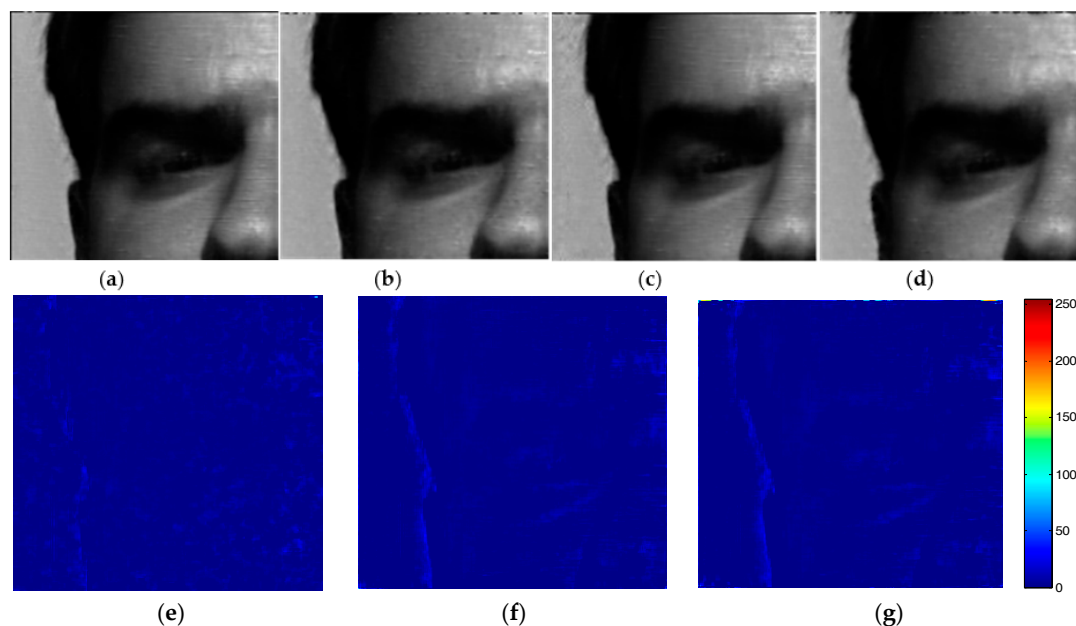

Figure 5. Reconstructed frames 14 for the eye blink dataset by multi-regularization algorithms for comparison: (**a**) ground truth; (**b**) result constructed by GAP-acc + TV algorithm; (**c**) result constructed by GAP + b algorithm; (**d**) result constructed by proposed multi-regularization algorithm; (**e**) the residual images of (**b**); (**f**) the residual images of (**c**); (**f**) the residual images of (**d**).
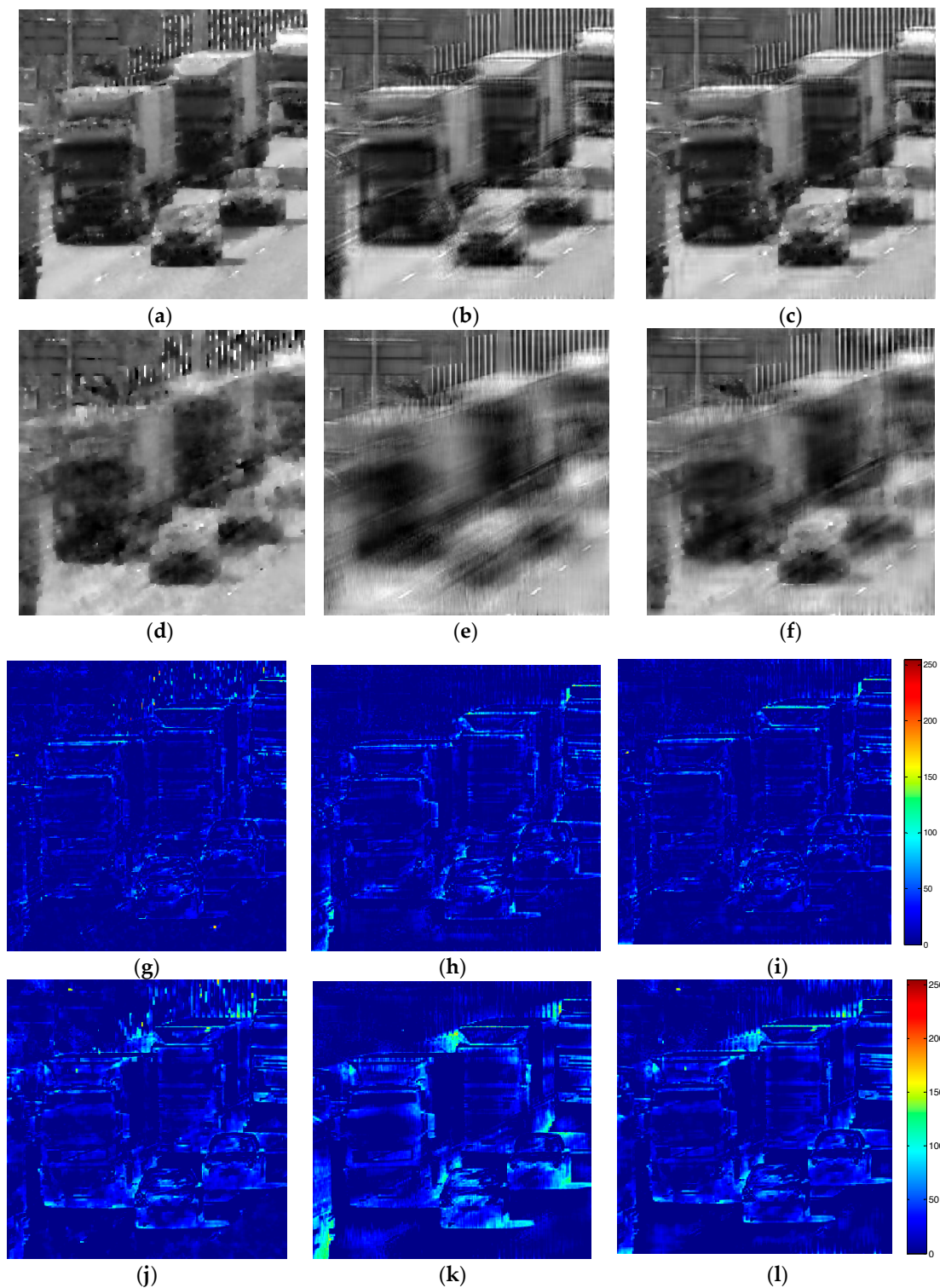
Figure 6. Reconstructed frame 8 for the traffic dataset by multi-regularization algorithms for comparison: (**a**) frame reconstructed by GAP-acc + TV algorithm with $N_f = 8$; (**b**) frame reconstructed by GAP + b algorithm with $N_f = 8$; (**c**) frame reconstructed by proposed algorithm with $N_f = 8$; (**d**) frame reconstructed by GAP-acc + TV algorithm with $N_f = 24$; (**e**) frame reconstructed by GAP + b algorithm with $N_f = 24$; (**f**) frame reconstructed by proposed algorithm with $N_f = 24$; (**g**) the residual images of (**a**); (**h**) the residual images of (**b**); (**i**) the residual images of (**c**); (**j**) the residual images of (**d**); (**k**) the residual images of (**e**); (**l**) the residual images of (**f**).

High-speed video of an eye blink, from closing to opening, is reconstructed from a single coded snapshot for $N_F = 14$. It is noticeable that the eye is the only part of the scene that moves. There is little aesthetic difference between these reconstructions.

Reconstructed results using multi-regularization algorithm and residual for the traffic dataset are shown in Figure 6. Frames reconstructed with GAP + b look better in the background but exhibit ghosting effects in the foreground due to ambiguities in the solution. The GAP-acc + TV can generally reconstruct the frames with sharper scenes than GAP + b but exhibits information loss and over-smoothness of reconstructed results. Compared with other algorithms, the proposed multi-regularization algorithm can provide richer texture information and sharper structure both in the background and foreground. When $N_f = 24$, the reconstructed result drops significantly, but the proposed multi-regularization algorithm can still identify the target in the scene. Visual quality evaluations on the reconstructed frames are consistent with the peak-to-peak signal-to-noise ratio (PSNR) results and residual images shown in Figure 7.

Figure 7 plots the PSNR curves and reconstruction error curves against the frame for comparison of the algorithms with the traffic dataset under the temporal compression ratio from 4 to 48 frames with a single coded snapshot. The performance of all algorithms drops significantly when the compression ratio is larger than 24. As shown in Figure 6d–f, the reconstructed frame cannot be useful to target recognition. Therefore, the proposed multi-regularization algorithm becomes the better choice in terms of balancing reconstruction quality and robustness.

The performance of GAP + b drops significantly when the compression ratio grows. As Figure 8 demonstrates, the proposed algorithm performs significantly better than GAP + b and GAP-acc + TV in terms of PSNR when the compression ratio is less than 24. It is robust enough to be applied under available compression ratios for video CS.
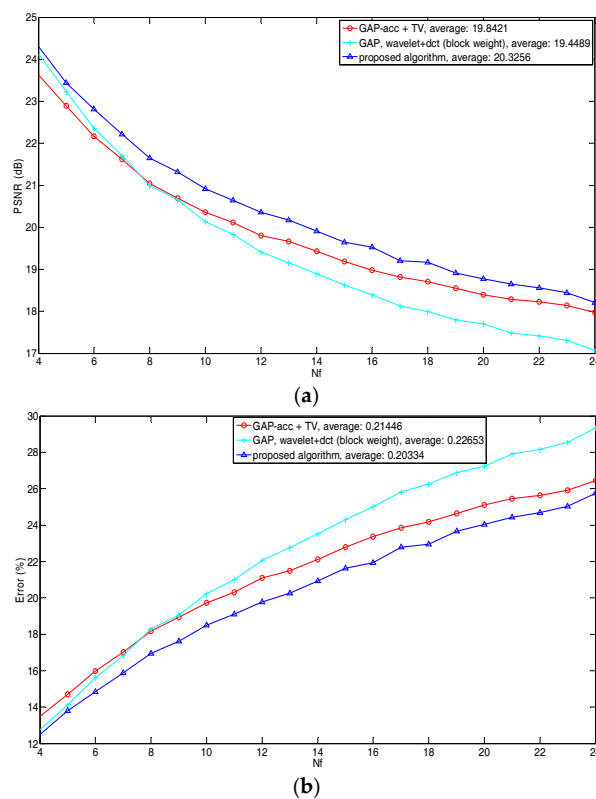

(a)

(b)

Figure 7. Comparison of reconstruction results of traffic datasets with compression ratio range from 4 to 24: (**a**) peak-to-peak signal-to-noise ratio (PSNR) for reconstructed frame of Traffic dataset under different reconstructed ratio; (**b**) reconstruction error for reconstructed frame of Traffic dataset under different reconstructed ratio.
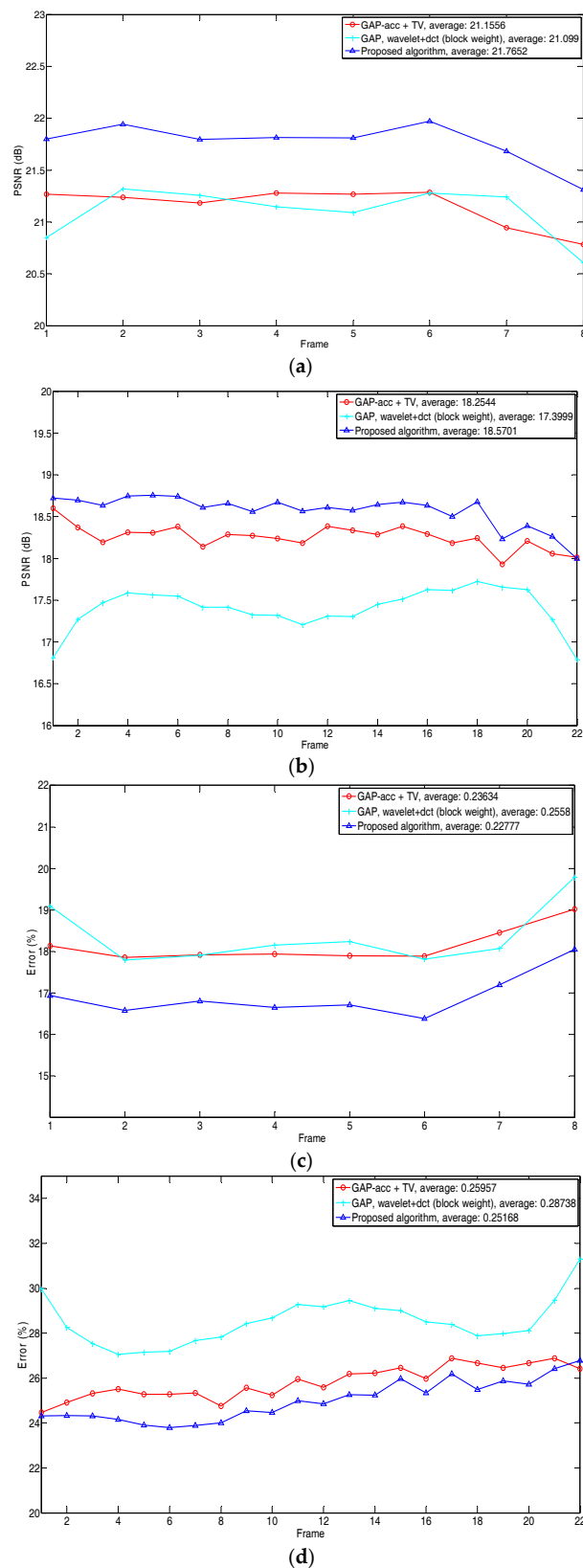
(**a**)

(**b**)

(**c**)

(**d**)

Figure 8. Comparison of reconstruction results of different datasets: (**a**) PSNR for each reconstructed frames of Traffic dataset (N_f = 8); (**b**) PSNR for each reconstructed frames of Traffic blink dataset (N_f = 22); (**c**) reconstruction error for each reconstructed frames of Traffic dataset (N_f = 8); (**d**) reconstruction error for each reconstructed frames of Traffic blink dataset (N_f = 22).

5.2. Reconstruction Performances under Varied Noise Conditions

In accordance with the CACTI Imaging Model, the imaging performance is significantly affected by noise. Therefore, the comparison is made under different noise conditions in this section. Figure 9 shows the performances of different algorithms with the white Gaussian noise variances 0.01 and 0.05.

As shown in Figure 9, the performances of all reconstructive algorithms decrease significantly in term of visual quality, although the PSNR still maintains a high value. The PSNR of reconstructed results for the traffic dataset with noises is shown in Figure 10. The average PSNR value of GAP + b method is decreases nearly 2 dB, compared with the results without noise. The average PSNR value of GAP + b method decreases nearly 2 dB, compared with the results without noise. Note that reconstructed frames by GAP + b look better but still with more blurring. Note that reconstructed frames by GAP + b looks better but still with more blurring. Because of its TV regulation reconstruction, the GAP-acc + TV algorithm can suppress the noise and thus results in smoother local details, but with the obvious missing texture. Compared with other algorithms, the proposed approach can suppress noise and motion blur and provide richer texture information, which is verified by the PSNR curve in Figure 10.
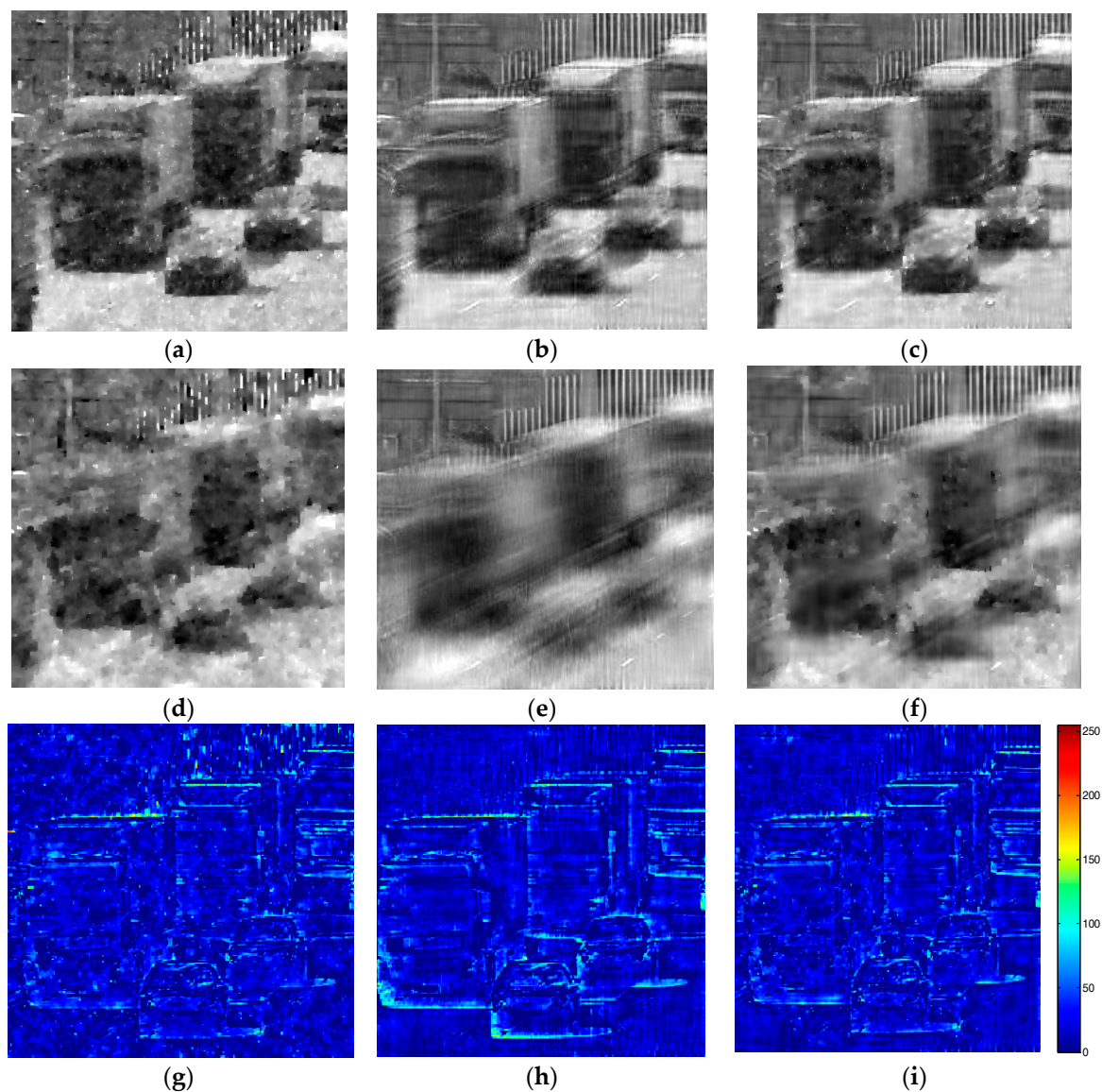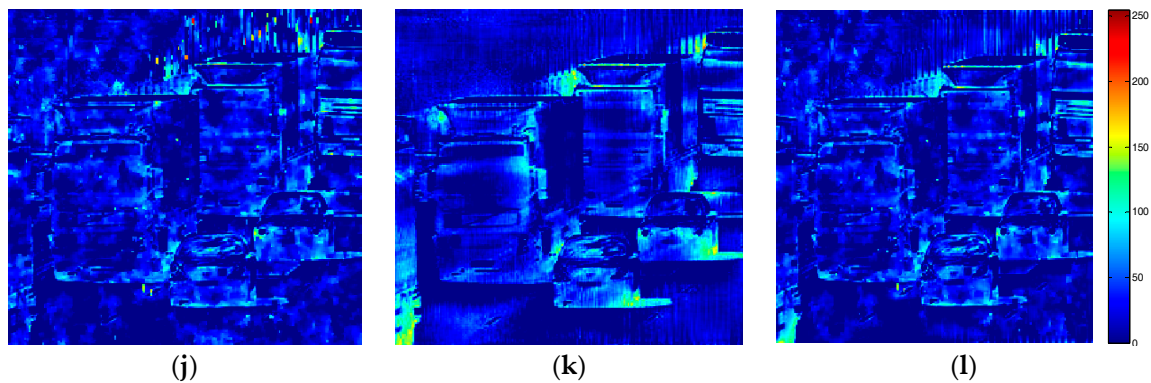

Figure 9. *Cont.*

(j)　　　　　　　　　　　**(k)**　　　　　　　　　　　**(l)**

Figure 9. Comparison of reconstruction results under noise conditions (variances = 0.01): (**a**) frame reconstructed by GAP-acc + TV algorithm (variances = 0.01, $N_f = 8$); (**b**) frame reconstructed by GAP + b algorithm (variances = 0.01, $N_f = 8$); (**c**) frame reconstructed by proposed algorithm (variances = 0.01, $N_f = 8$); (**d**) frame reconstructed by GAP-acc + TV algorithm (variances = 0.05, $N_f = 24$); (**e**) frame reconstructed by GAP + b algorithm(variances = 0.05, $N_f = 24$); (**f**) frame reconstructed by proposed algorithm (variances = 0.05, $N_f = 24$); (**g**) the residual images of (**a**); (**h**) the residual images of (**b**); (**i**) the residual images of (**c**); (**j**) the residual images of (**d**); (**k**) the residual images of (**e**); (**l**) the residual images of (**f**).

(a)

(b)
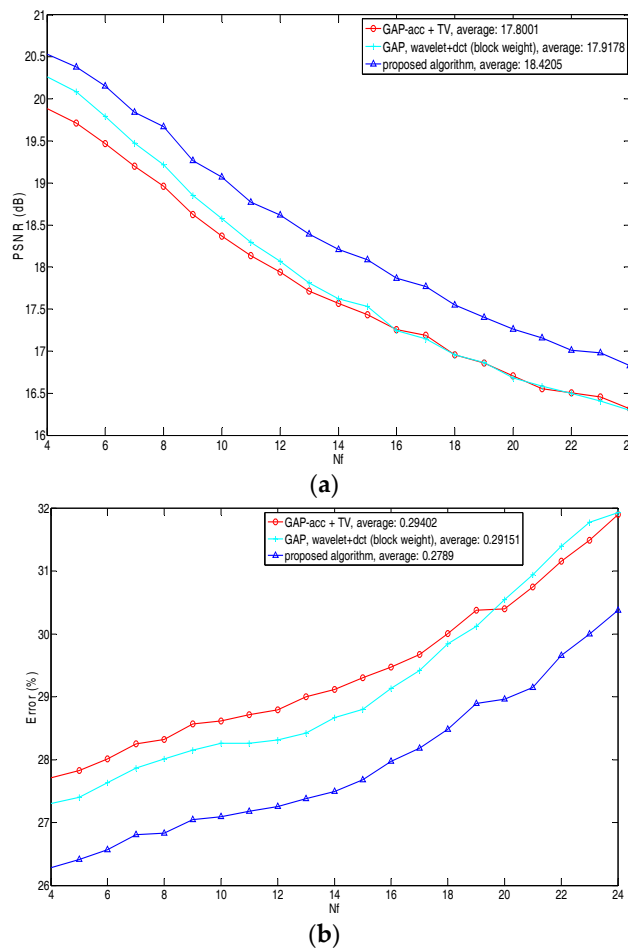
Figure 10. Performance comparison of CACTI reconstruction algorithms under noise conditions: (**a**) PSNR for reconstructed frame of Traffic dataset under different reconstructed ratio; (**b**) reconstruction error for reconstructed frame of Traffic dataset under different reconstructed ratio.

The performance of GAP-b drops significantly with the noises. From Figure 11, the proposed algorithm performs significantly better than GAP + b and GAP-acc + TV in terms of PSNR when the compression ratio is less than 24 under noise condition. It is robust enough to be applied for real world video CS in term of both noise and high compression ratio.
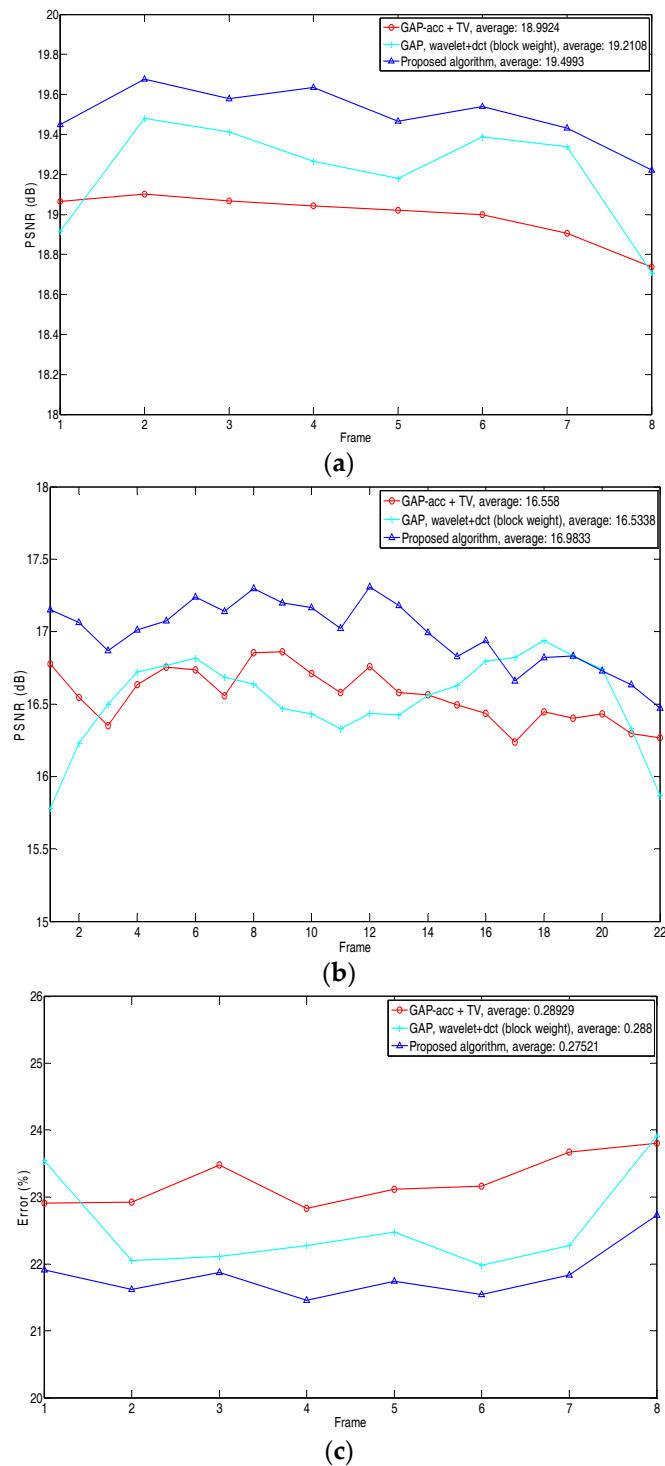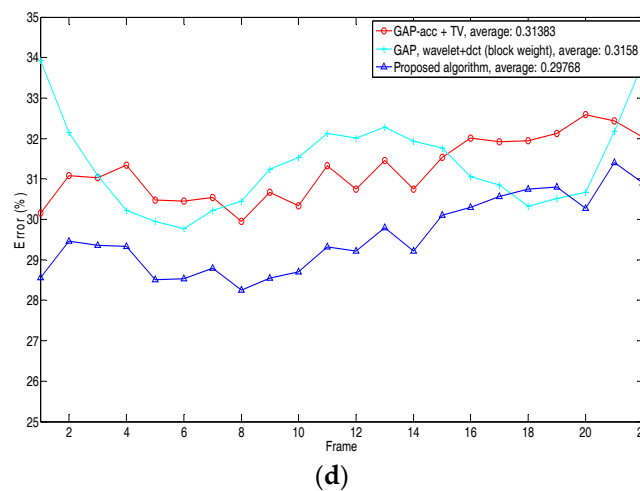

(a)

(b)

(c)

Figure 11. *Cont.*

(d)

Figure 11. Performance comparison of CACTI reconstructive algorithms under noise conditions (variances = 0.05): (**a**) PSNR for reconstructed frames of Traffic with $N_f = 8$; (**b**) PSNR for reconstructed frames of Traffic with $N_f = 22$; (**c**) reconstruction error for reconstructed frames of Traffic with $N_f = 8$; (**d**) reconstruction error for reconstructed frames of Traffic with $N_f = 22$.

5.3. Performance of Spatiotemporal Super-Resolution Reconstruction

According to the comparison and analysis in Section 5.2, for the given noise conditions and compression ratios, the current reconstruction algorithms cannot significantly improve the quality of reconstructed images. Therefore, more consideration is drawn on the inter-frame non-redundant information of the reconstructed image sequences. In this section, we conduct simulations to verify the effectiveness of the MAP-MRF-based SR approach for CACTI. We compare performances of the GAP + b algorithm, the Gap-acc + TV algorithm, the proposed multiple regularization reconstruction algorithm, and the 3D Steering Kernel Regression algorithm which is classical spatiotemporal super resolution algorithms.

For detailed comparison, one arbitrarily selected frame from each of the reconstructed SR sequences is shown in Figure 12. It should be noted that our result is an improvement of the time and spatial resolution simultaneously, and the results were scaled to facilitate the comparison to others. The estimation results of different algorithms are used as the initial input of the LR sequence, and the output results reconstructed by MAP-MRF-based approach can reflect the error propagation. As shown in Figure 12i,j, both GAP + b algorithm and GAP-acc + b algorithms cause the error propagation, especially at the edge of the motion target. In the proposed MAP-MRF approach, the multi-regularization algorithm and coarse-to-fine scheme can address error propagation and significantly improve the performance.
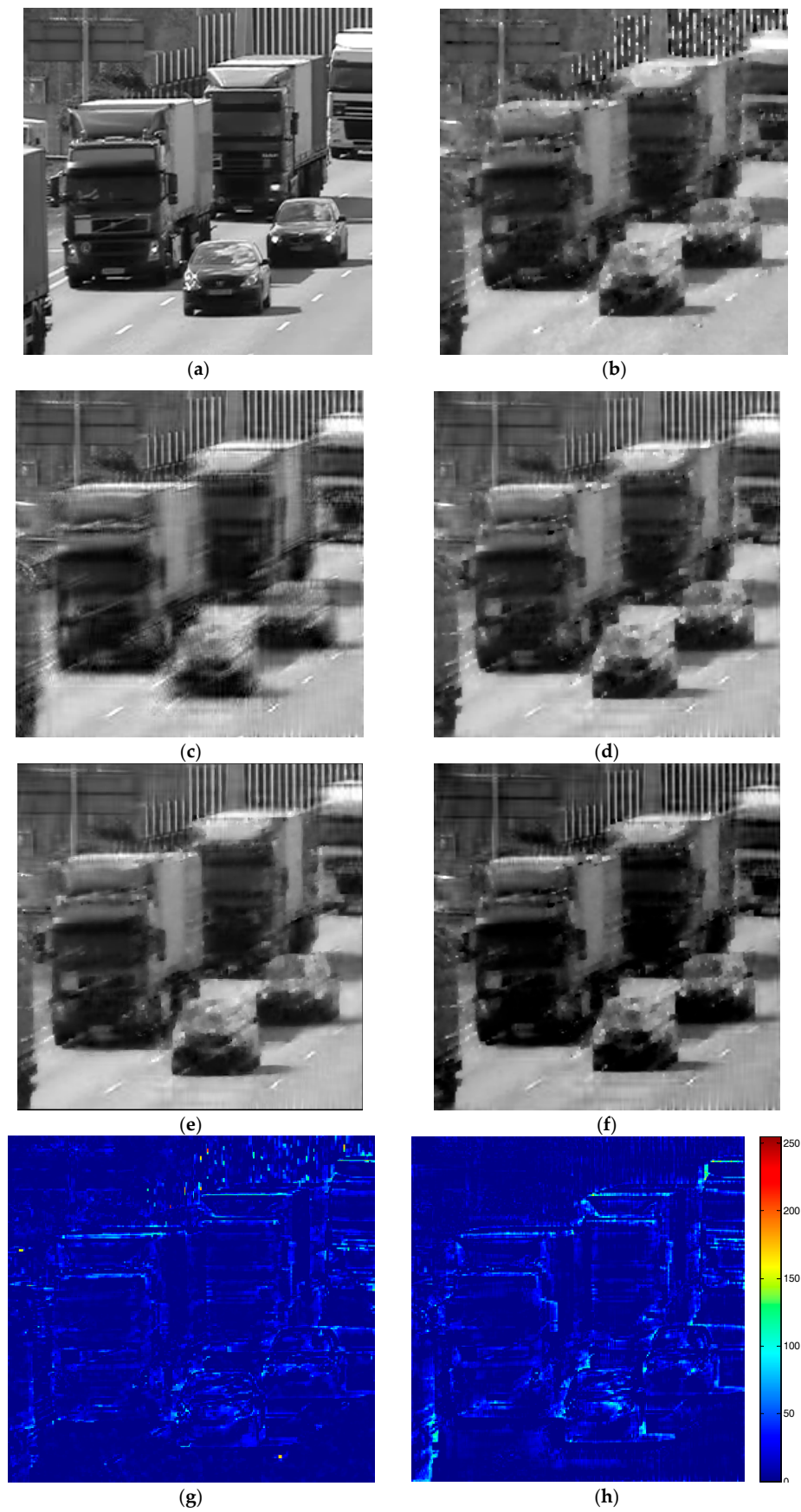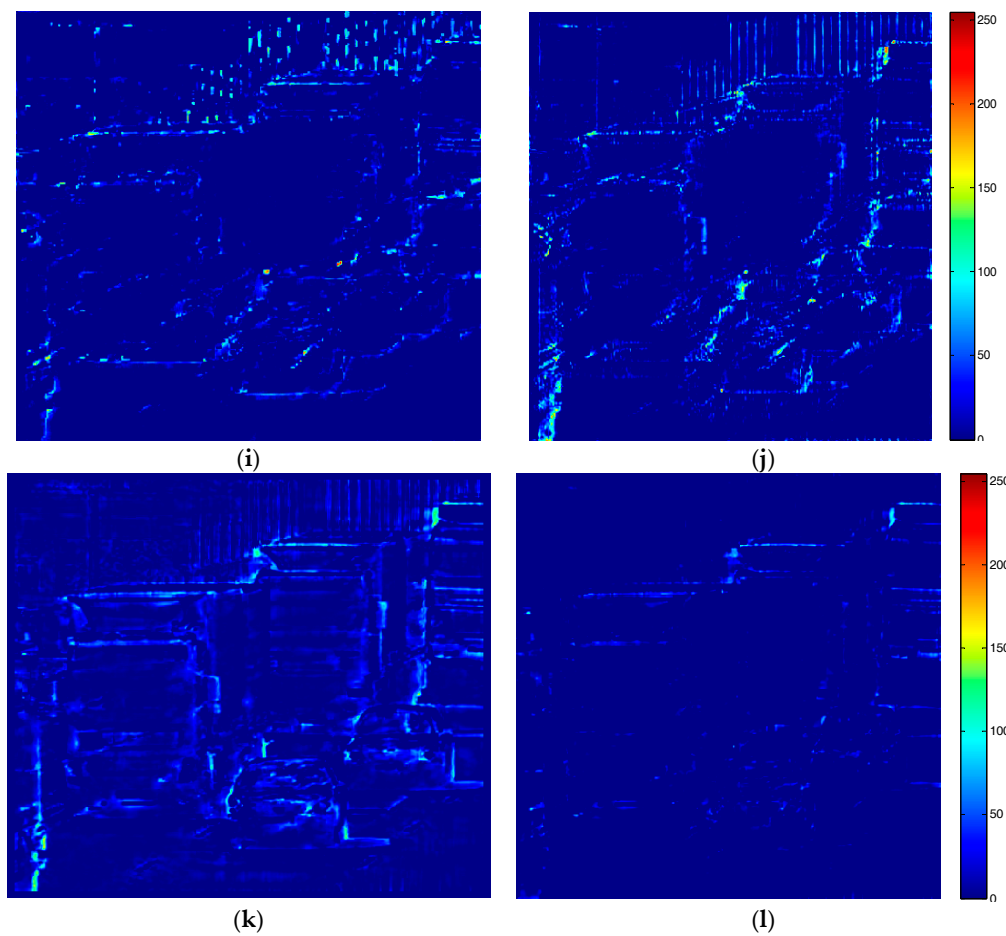
(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 12. *Cont.*

Figure 12. SR results of CACTI reconstruction results of traffic dataset with $N_f = 8$: (**a**) ground truth; (**b**) reconstructed frame 1 by GAP-acc + TV; (**c**) reconstructed frame 1 by GAP + b; (**d**) reconstructed frame 1 by proposed multi-regularization algorithm; (**e**) reconstructed frame 1 by spatial temporal algorithm super reconstruction based on 3D Steering Kernel Regression; (**f**) reconstructed frame by MAP-MRF-based spatial temporal super reconstruction approach; (**g**) the residual images of (**b**); (**h**) the residual images of (**c**); (**i**) the residual of image reconstructed by MAP-MRF-based approach with the initial LR sequence reconstructed GAP-acc + TV; (**j**) the residual of image reconstructed by MAP-MRF-based approach with the initial LR sequence reconstructed GAP + b; (**k**) the residual images of (**e**); (**l**) the residual images of (**f**).

Some close-ups of the reconstructed results are enlarged and presented below the corresponding frames. As proven in Figure 13, the approach proposed herein causes less loss of the edge information according to the profile comparison, and noticeably improves the local spatial resolution.

In accordance with the edge information, the MAP-MRF-based SR approach can heighten the spatiotemporal resolution, and meanwhile preserve the texture information and produce smoother edges with sharper transitions and less motion blurring, compared with other algorithms. This is mainly because the WNS and multi-regularization in our method effectively exploit the spatiotemporal coherences. In summary, through the spatiotemporal SR, the visual quality is improved significantly, especially in terms of fidelity in the foreground.
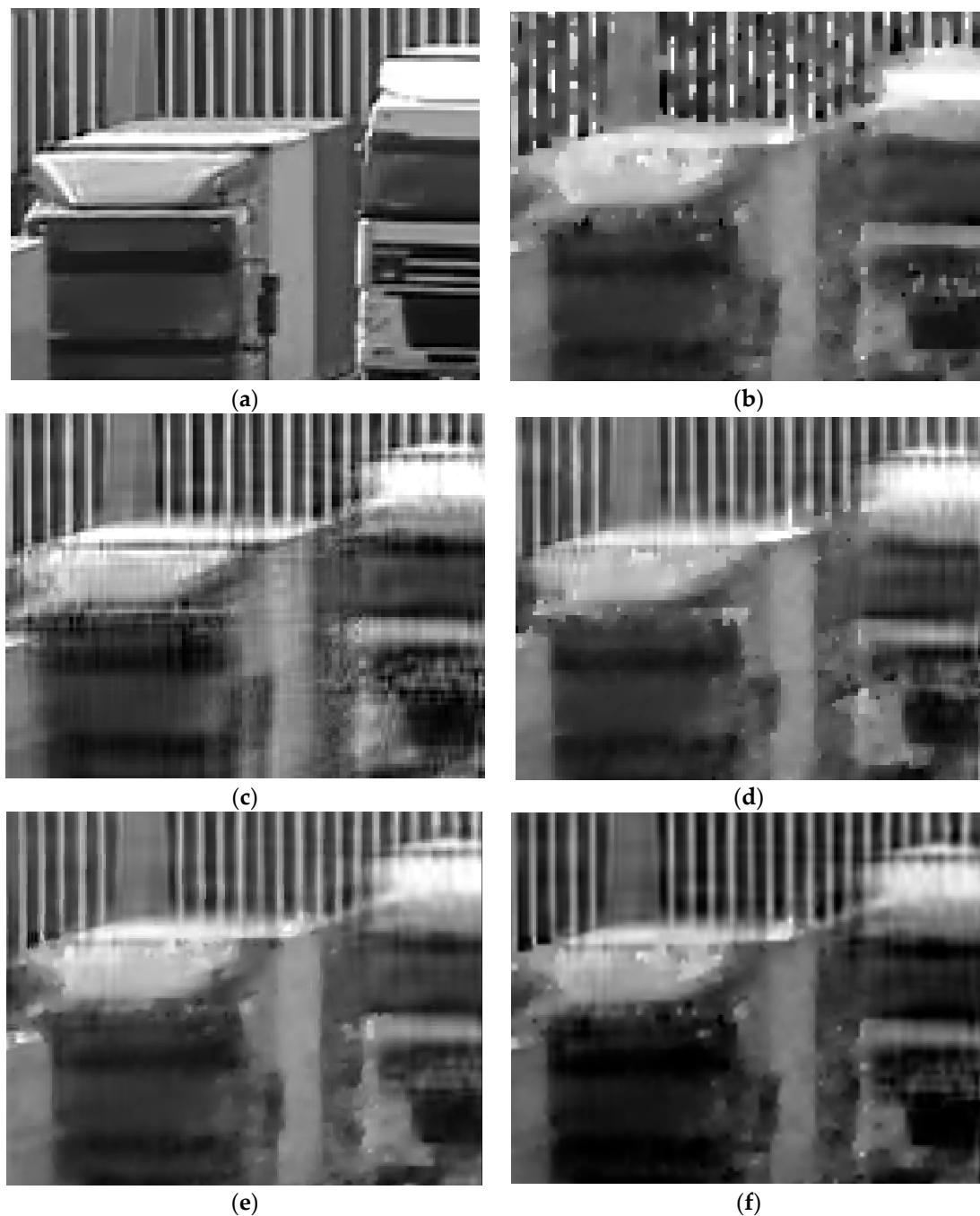
Figure 13. Close ups of the reconstruction results of traffic dataset: (**a**) ground truth; (**b**) close up of reconstructed frame by GAP-acc + TV; (**c**) close up of reconstructed frame by GAP + b; (**d**) close up of reconstructed frame by proposed multi-regularization algorithm; (**e**) close up of reconstructed frame by 3D Steering Kernel Regression; (**f**) close up of reconstructed frame by MAP-MRF-based reconstruction approach.

The calculation of regulation coefficient and weighted coefficient for the proposed approach depends on the estimation of motion vectors. The results of motion vectors and weight parameters using SIFT Flow algorithm are shown in Figure 14.
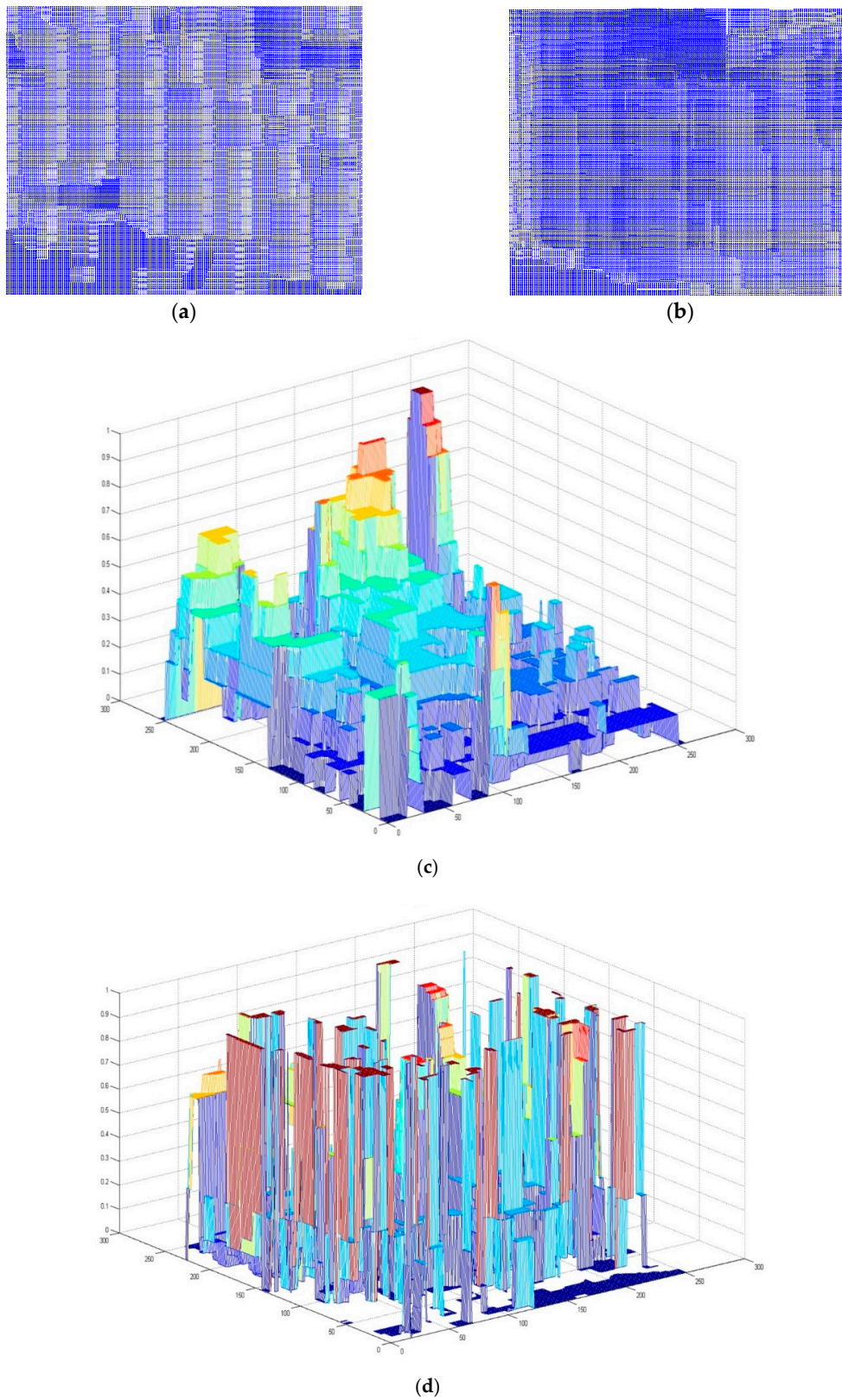
(**a**)

(**b**)

(**c**)

(**d**)

Figure 14. The motion vector between different frames: (**a**) the motion vector between frame 1 and frame 2; (**b**) the motion vector between frame 1 and frame N_f; (**c**) the regulation coefficient of multi-regularization algorithm; (**d**) the coefficient for the proposed MAP-MRF-based reconstruction approach.

With the known real-world HR images, the reconstruction performance is evaluated by PSNR; for unknown ground truth images, the effect is indicated by blind image quality indices [31]. We adopt a no-reference quality evaluation metric, that is, the blind image quality index (BIQI) [32], which will generate a quality score based on the video frame distortion statistics. The score is within 0–100, and the smaller the score is, the better the quality will be.

After the assessment, the SR reconstruction sequence by 3D Steering Kernel Regression gets a score of 53.9386, while the proposed MAP-MRF-based reconstruction approach is 44.3304. The proposed MAP-MRF-based approach enjoys a better indication of performance in a blind image quality index. In accordance with both the objective indexes and subjective evaluation, the proposed MAP-MRF-based approach is generally better than the other ones. Therefore, this simulation verifies the efficacy of the framework and the performance of the proposed reconstruction approach.

Suppose the image has h^2 pixels, the time and space complexity of this dual-layer BP is $O(h^4)$. The coarse-to-fine algorithm can speed up computation, whose complexity is $O(h^2 log h)$. The multi-regularization algorithm and reconstruction algorithm using Accelerated Linearized Bregman algorithm require $O(h log h)$ time in each iteration respectively. For these 256×256 images ($N_f = 8$) in one observation, the running time of GAP-acc + TV is 14.35 s, the running time of GAP + b is 38.80 s, the running time of the proposed algorithm with multiple regularizations is 28.53 s, and the whole running time of the proposed MAP-MRF-based approach using BP method is 657.65 s without any optimization on a computer with quad-core 3.4 GHz Intel i7-4770 CPU and 8 GB memory. Further speedup can be achieved through GPU implementation of the BP algorithm since this algorithm can be parallelized. We leave this as future work.

6. Conclusions

CACTI can afford the low-cost temporal super-resolution, but there is image degradation in the process of reconstruction. The imaging results are mainly determined by the performance and robustness of the reconstructive algorithms.

The GAP with transform regularization algorithm and the GAP with the TV regularization algorithm are the state-of-the-art algorithms used in CACTI. However, the GAP algorithm cannot easily select the groups and group weights in the transform domain, and the selection of groups and weights affects the results significantly. The GAP + TV algorithm utilizes TV regularization only imposed on the 2D frame, which may result in information loss or over-smoothness of reconstructed results, although the TV algorithms are almost parameter-free.

A maximum a posteriori probability and Markov random field (MAP-MRF) super-resolution approach is proposed to enhance the clarity and fidelity of the reconstruction results. It is based on a weighted 3D neighbor system and adopts the coordinate descent method to perform joint estimation on model parameters, including motion vector flow and SR results.

The key processes of the proposed approach are the joint estimation of the MAP-MRF model using coordinate descent method, and the image reconstruction using multi-regularization. A multi-reconstruction algorithm is proposed in this paper to reconstruct high-speed videos from CACTI, and simulation results demonstrate its robustness and efficiency. The proposed approach renders noise into a total variation minimization problem for compressive sensing and utilizes the redundant information among multi-frames.

The proposed approach can provide richer texture information and sharper structure. Extensive simulation results verify the robustness and efficacy of the proposed approach. Future work may focus on incorporating this optimization framework into the hardware system.

References

1. Llull, P.; Liao, X.; Yuan, X.; Yang, J.; Kittle, D.; Carin, L.; Sapiro, G.; Brady, D.J. Coded aperture compressive temporal imaging. *Opt. Express* **2013**, *9*, 698–706. [CrossRef] [PubMed]

2. Agrawal, A.; Veeraraghavan, A.; Raskar, R. Reinterpretable imager: Towards variable post-capture space, angle and time resolution in photography. *Comput. Graph. Forum* **2010**, *29*, 763–772. [CrossRef]

3. Don, M.L.; Fu, C.; Arce, G.R. Compressive imaging via a rotating coded aperture. *Appl. Opt.* **2017**, *56*, B142–B153. [CrossRef] [PubMed]

4. Tsai, T.H.; Llull, P.; Yuan, X.; Brady, D.J.; Carin, L. Spectral-temporal compressive imaging. *Opt. Lett.* **2015**, *40*, 4054–4057. [CrossRef] [PubMed]

5. Levin, A.; Fergus, R.; Durand, F.; Freeman, B. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* **2007**, *26*, 70–78. [CrossRef]

6. Raskar, R.; Agrawal, A.; Tumblin, J. Coded exposure photography: Motion deblurring using fluttered shutter. *ACM Trans. Graph.* **2006**, *25*, 795–805. [CrossRef]

7. Liao, X.; Li, H.; Carin, L. Generalized alternating projection for weighted-$\ell_{2,1}$ minimization with applications to model-based compressive sensing. *SIAM J. Imaging Sci.* **2014**, *7*, 797–823. [CrossRef]

8. Koller, R.; Schmid, L.; Matsuda, N. High spatio-temporal resolution video with compressed sensing. *Opt. Express* **2015**, *23*, 15992–16007. [CrossRef] [PubMed]

9. Gupta, M.; Mitsunaga, T.; Hitomi, Y.; Gu, J.; Nayar, S.K. Video from a single coded exposure photograph using a learned over-complete dictionary. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 287–294. [CrossRef]

10. Yang, J.; Yuan, X.; Liao, X. Video compressive sensing using Gaussian mixture models. *IEEE Trans. Image Process.* **2014**, *23*, 4863–4878. [CrossRef] [PubMed]

11. Zhang, S.; Jiao, L.; Liu, F.; Wang, S. Global low-rank image restoration with Gaussian mixture model. *IEEE Trans. Cybern.* **2017**, *99*, 1–12. [CrossRef] [PubMed]

12. Yuan, X.; Llull, P.; Liao, X.; Yang, J.; Sapiro, G.; Brady, D.J.; Carin, L. Low-cost compressive sensing for color video and depth. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 3318–3325. [CrossRef]

13. Yuan, X. Generalized alternating projection based total variation minimization for compressive sensing. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 2539–2543. [CrossRef]

14. Nasrollahi, K.; Moeslund, T.B. Super-resolution: A comprehensive survey. *Mach. Vis. Appl.* **2014**, *25*, 1423–1468. [CrossRef]

15. Shechtman, E.; Caspi, Y.; Irani, M. Space-time super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 531–545. [CrossRef] [PubMed]

16. Llull, P.; Yuan, X.; Liao, X. Temporal compressive sensing for video. In *Compressed Sensing and Its Applications. Applied and Numerical Harmonic Analysis*; Boche, H., Calderbank, R., Kutyniok, G., Vybíral, J., Eds.; Springer International Publishing: Basel, Switzerland, 2015; pp. 14–18. ISBN 978-3-319-16041-2.

17. Veeraraghavan, A.; Reddy, D.; Raskar, R. Coded strobing photography: Compressive sensing of high speed periodic videos. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 671–686. [CrossRef] [PubMed]

18. Gao, L.; Li, J.; Li, C.; Wang, L.V. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature* **2014**, *516*, 74–77. [CrossRef] [PubMed]

19. Liu, D.; Gu, J.; Hitomi, Y. Efficient space-time sampling with pixel-wise coded exposure for high-speed imaging. *IEEE Trans. Softw. Eng.* **2014**, *36*, 248–260. [CrossRef]

20. Liu, C.; Sun, D. On Bayesian adaptive video super resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 346–360. [CrossRef] [PubMed]

21. Hosseini, M.S.; Plataniotis, K.N. High-accuracy total variation for compressed video sensing. *Mathematics* **2014**, *23*, 3869–3884. [CrossRef] [PubMed]

22. Li, C.; Yin, W.; Jiang, H.; Zhang, Y. An efficient augmented lagrangian method with applications to total variation minimization. *Comput. Optim. Appl.* **2017**, *56*, 507–530. [CrossRef]

23. Hong, M.; Luo, Z.Q. On the linear convergence of the alternating direction method of multipliers. *Math. Program.* **2012**, *162*, 1–35. [CrossRef]

24. Zhu, M.; Wright, S.J.; Chan, T.F. Duality-based algorithms for total-variation-regularized image restoration. *Comput. Optim. Appl.* **2010**, *47*, 377–400. [CrossRef]

25. Liu, H.; Liu, S.; Zhang, Z.; Sun, J.; Sun, J. Adaptive total variation-based spectral deconvolution with the split Bregman method. *Appl. Opt.* **2014**, *53*, 8240–8248. [CrossRef] [PubMed]

26. Beck, A.; Teboulle, M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **2009**, *18*, 2419–2434. [CrossRef] [PubMed]

27. Liang, M.; Du, J.; Li, X. Spatio-temporal super-resolution reconstruction based on robust optical flow and Zernike moment for video sequences. *Math. Probl. Eng.* **2013**, *745752*, 1–15. [CrossRef]

28. Liu, C.; Yuen, J.; Torralba, A. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *33*, 978–994. [CrossRef] [PubMed]

29. Mudenagudi, U.; Banerjee, S.; Kalra, P.K. Space-time super-resolution using graph-cut optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 995–1008. [CrossRef] [PubMed]

30. Yasuda, M.; Kataoka, S.; Tanaka, K. Statistical analysis of loopy belief propagation in random fields. *Phys. Rev. E* **2015**, *92*, 725–732. [CrossRef] [PubMed]

31. Moorthy, A.K.; Bovik, A.C. A two-step framework for constructing blind image quality indices. *IEEE Signal Process. Lett.* **2010**, *17*, 513–516. [CrossRef]

32. OG-IQA Software Release. Available online: http://live.ece.utexas.edu/research/quality/og-iqa_release.zip (accessed on 5 October 2015).