Article

Joint Prior Learning for Visual Sensor Network Noisy Image Super-Resolution

Bo Yue [1,†], Shuang Wang [1,*,†], Xuefeng Liang [2,*,†], Licheng Jiao [1] and Caijin Xu [1]

[1] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an 710071, China; yuebo312@live.com (B.Y.); jlcxidian@163.com (L.J.); xcj0918xd.@163.com (C.X.)

[2] Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

* Correspondence: shwang.xd@gmail.com (S.W.); xliang@i.kyoto-u.ac.jp (X.L.); Tel.: +86-29-8820-2661 (S.W.); +81-75-753-5908 (X.L.)

† These authors contributed equally to this work.

Abstract: The visual sensor network (VSN), a new type of wireless sensor network composed of low-cost wireless camera nodes, is being applied for numerous complex visual analyses in wild environments, such as visual surveillance, object recognition, *etc*. However, the captured images/videos are often low resolution with noise. Such visual data cannot be directly delivered to the advanced visual analysis. In this paper, we propose a joint-prior image super-resolution (JPISR) method using expectation maximization (EM) algorithm to improve VSN image quality. Unlike conventional methods that only focus on upscaling images, JPISR alternatively solves upscaling mapping and denoising in the *E-step* and *M-step*. To meet the requirement of the *M-step*, we introduce a novel non-local group-sparsity image filtering method to learn the explicit prior and induce the geometric duality between images to learn the implicit prior. The *EM* algorithm inherently combines the explicit prior and implicit prior by joint learning. Moreover, JPISR does not rely on large external datasets for training, which is much more practical in a VSN. Extensive experiments show that JPISR outperforms five state-of-the-art methods in terms of both PSNR, SSIM and visual perception.

Keywords: visual sensor network; image super-resolution; image denoising; prior learning; EM algorithm

1. Introduction

A large or small visual sensor network (VSN) depending on spatially-distributed smart cameras is for sensing, communicating and fusing images of a scene from varied viewpoints. It has been applied to a variety of public applications, including security and area surveillance, tracking, environmental monitoring, *etc.* A VSN is generally equipped with low-cost camera nodes, which are simply stuck on walls or poles in a wild environment, such as dark lighting and dirt on the lens. Due to the working condition and limitations of the low-cost imaging acquisition device, most of the obtained frames are low resolution (LR) images with certain noise. Even a small amount of noise, which is inevitable in low-light conditions, reduces the visibility of details that could contain vital information.

In a VSN, one desires to obtain high resolution (HR) images with the least noise. However, constructing the HR image x from the observed LR image y with noise is a typically ill-posed. Mathematically, it can be modeled as:

$$y = AHx + n \tag{1}$$

$AHx + n$ is the mapping function of the image sensor from a real scene to the LR image. More specifically, A denotes the downsampling process; H is a blurring operation; and n represents the additive white Gaussian noise with variance δ^2. Solving this problem is indeed a non-trivial challenge due to the fewer rows than columns in the matrix AH.

To address this ill-posed issue, the prior information about the desired HR image may convert this problem into being well-posed. In recent years, this information has generally been divided into two categories: the explicit prior and the implicit prior. The first model reflects our basic understanding or assumption of the distribution or energy function of the target reconstructed image. It formulates the problem within the Bayesian framework by maximizing the probability of the HR image from the LR image. It can be generally modeled as:

$$x = \arg \min_{x} \|y - AHx\|_2^2 + \lambda R(x) \tag{2}$$

where $\|y - AHx\|_2^2$ is the likelihood function describing the probability relation between the LR image y and the original HR image x, $R(x)$ is the prior knowledge of the HR image showing how likely a reconstructed image x is. Therefore, the optimization strategy [1–3] and the selection of prior information [4–10] have been the core issues of explicit prior learning.

As opposed to explicit prior learning, implicit prior learning is somehow non-selective. The HR image prior can be implicitly defined, rather than explicitly given by the specific statistical distribution. It is intimately related to the learning-based super-resolution, which aims to learn the co-occurrence of local structures between the LR and HR images from the external/internal training database. Recently, two important foundations have been proposed, namely the coupled overcomplete dictionary learning [11–14] and the deep networks [15,16], for image super-resolution. They either learn an LR-HR overcomplete dictionary pair or network parameters from training datasets that contain a million co-occurrences of LR-HR image patches. Additionally, then, the dictionary pair or network parameter is the implicit prior to estimate the nonlinear mapping between the LR and HR images. The general model is:

$$x = F(y; w)$$
$$\text{s.t. } w = \arg \min_{w} \|x_{tr} - F(y_{tr}; w)\|_p^2 + \lambda R(w) \tag{3}$$

where x_{tr}, y_{tr} are the training data, F represents the mapping function, w is the dictionary pair or network parameters and $R(w)$ is the penalty on the parameters.

One can see that neither an explicit prior nor an implicit prior can completely super-resolve the noisy LR images in a VSN; because both of them concentrate on the upscaling mapping from the LR image to the HR image, but consider the denoising less. Moreover, they require large datasets for model training. This is almost impossible and unnecessary in a VSN. In this paper, we propose a joint-prior image super-resolution (JPISR) method using the expectation maximization (EM) algorithm for image quality improvement in a VSN. This EM algorithm alternatively solves upscaling mapping and denoising in the *E-step* and *M-step*. Specifically, we introduce a maximum a posteriori (MAP)-based HR image estimation method for the *M-step*, where the explicit prior serves as the likelihood estimation, and the implicit prior is regarded as the Bayesian prior estimation; in which we introduce a novel adaptive non-local group-sparsity image filtering method for likelihood estimation to adequately mine the explicit prior. Meanwhile, we induce the geometric duality [17,18] between the LR and HR images/patches into the implicit prior learning about every individual clean image patch, which can further enhance the super-resolution performance. Then, one can predict the mean and covariance of the target image patch from its respective LR image patch by a *Gaussian process* [19,20]. JPISR inherently integrates the above two priors' learning into one framework. Thanks to the non-local group-sparsity

and the geometric duality, this joint learning does not require external training data. Finally, we obtain the posterior of the HR image patch by the *Bayesian minimum mean-square error (BMMSE) estimator*.

The contributions of this paper can be summarized as follows:

- We propose a JPISR method based on the expectation maximization (EM) algorithm for image quality improvement in a visual sensor network, which can effectively reconstruct finer details and simultaneously suppress noises for smart sensing.
- In the *M-step* procedure, the novel adaptive non-local group-sparsity explicit prior serves as the likelihood estimation, and the geometric duality implicit prior is regard as the Bayesian prior estimation. They are effectively integrated into one framework by maximum a posteriori (MAP). Since there is no need for the external training data, this joint prior learning is very suitable for a VSN.
- When a pattern with a high frequency signal is simple, but rarely repeated in the image, we introduce a rotation invariance into non-local self-exemplars to increase the number of repeated image patches for explicit prior learning.

The rest of this paper is organized as follows. We briefly review the related visual sensor network and the state-of-the-art image super-resolution techniques in Section 2. Then, in Section 3, we explain the proposed EM algorithm for joint-prior learning image super-resolution in detail. Section 4 introduces the adaptive non-local group-sparsity image filtering method and the implicit prior learning by the *Gaussian process*. In Section 5, experimental results and comparisons with state-of-the-art methods are provided to show the effectiveness of the proposed method. Finally, we conclude the paper in Section 6.

2. Related Works

The visual sensor network has recently emerged as a new type of sensor-based intelligent system, which processes the captured image/video data locally and collaborates with other cameras over a network. Taking into account the high production cost of cameras, the VSN is often equipped with low-cost cameras. For the applications in wild environments, the LR image with noise seriously restricts the efficiency of the VSN, due to the targets being covered by a few pixels in size [6,21,22]. Thus, image super-resolution methods are desired to enhance the image quality.

As previously mentioned, the explicit prior and the implicit prior could make the image super-resolution problem well-posed. The explicit prior learning can be interpreted from the *Bayesian* perspective. Babacan *et al.* [1] proposed a variational Bayesian method to estimate the distributions of all unknowns. Zhao *et al.* [3] proposed a new fast super-resolution approach, which is successfully embedded into the alternating direction method of multipliers (ADMM) framework. Sun *et al.* [7] exploited the gradient profile priors for local image structures. Dong *et al.* [8] took advantage of the non-local similarity, sparse representation and autoregressive (AR) models in an image. Zhang *et al.* [9] modeled a natural image prior by a high-order Markov random field (MRF). Yang *et al.* [10] proposed a method that exploits self-similarities and group structural information of image patches. Li *et al.* [23] combined sparse representation and non-local similarity for image SR. Sajjad *et al.* [24] used an over-redundant dictionary based on effective image representations for image SR. The main differences among these methods are the optimization strategies and selections of prior information.

On the other hand, the implicit prior is learned from the mapping relationship from the LR image patch space to the HR image patch space. Considering that the image patches of different resolutions are of different linear spaces and each image patch can be represented by a vector, the relation between the LR image patch and HR image patch can be regarded as the relation between two vector spaces associated with the two image patches' styles. Under this assumption, a series of learning (mapping)-based approaches has been proposed to model the relationship between the LR image patch and the HR image patch. As the seminal work, the Markov random field was proposed by Freeman *et al.* [25] to model the relationship between the HR image patch and the LR one and

between the HR image patch and its neighboring patches. Then, the HR image is inferred by the likelihood maximum. The most well-known method is coupled dictionary learning based on sparse representation, which was proposed by Yang *et al.* [11,12]. As its core, the assumption is that the sparse representation of the HR image patch is the same as that of the LR one. Thus, the HR dictionary and the LR dictionary are united into a single dictionary, which is trained by classical dictionary learning methods. Considering that it is impossible to learn a universal model that creates such as a complex mapping relationship, Zhang *et al.* [14] proposed multiple linear mapping functions for super-resolution reconstruction. Recently, deep learning-based methods [15,16] were introduced to learn the relationship and show distinct advantage over existing state-of-the-art methods.

3. EM Scheme of Image Super-Resolution in a VSN

3.1. Problem Formulation

The low-cost VSN usually produces LR images/videos with noise. Image super-resolution aims to transform the LR image to the HR image, which is modeled in Equation (1). However, most super-resolution methods assume that input LR images have no noise. This is far from the reality in a VSN. Moreover, directly super-resolving the noisy LR image has little practicability. To address this issue, we introduce a missing datum, the so-called hidden variable (image) z, by dividing the image super-resolution degradation procedure into two problems:

$$\begin{cases} z = x + \alpha n_1 \\ y = Sz + n_2 \end{cases} \tag{4}$$

where we lump the downsampling operator and blurring operator into a single measurement matrix $S = AH$, $n_1 \sim \mathcal{N}(0, I)$ and $n_2 \sim \mathcal{N}(0, \delta^2 I - \alpha^2 SS^T)$ are independent Gaussian noises, such that $n = \alpha S n_1 + n_2 \sim \mathcal{N}(0, \delta^2 I)$, and α ($\alpha < \delta$) is a positive parameter controlling the distribution of noise.

Clearly, if z is given, we could obtain the HR image x by solving the equation $z = x + \alpha n_1$, which is a pure denoising problem, where αn_1 represents zero-mean noise with covariance $\alpha^2 I$. This key observation lets us be able to treat z as a missing datum and alternately estimate x and z in the *E-step* and *M-step* for simultaneous image upscaling and denoising.

3.2. EM Scheme

3.2.1. *E-Step*: The Likelihood of the Super-Resolution Model Approach Procedure

The *E-step* finds the conditional expectation value of the complete-data log-likelihood $\log p(y, z|x)$ with respect to the unknown z, the observed data y and the current parameter (the estimated HR image) for $\hat{x}^t$. The so-called *Q*-function is defined as follows:

$$Q(x, \hat{x}^t) = \mathbb{E}\left[\log p(y, z|x) | y, \hat{x}^t\right] \tag{5}$$

Equation (5) can be further reformulated as follows:

$$Q(x, \hat{x}^t) = -\frac{1}{2\alpha^2} \|x - \hat{z}^t\|_2^2 + k \tag{6}$$

where $\hat{z}^t$ is the *t*-th hidden image estimate z and k is a constant. The proof of Equation (6) is given in [26]. The *t*-th hidden image estimate $\hat{z}^t$ can be derived as follows.

Let $\hat{x}^t$ be the *t*-th HR image estimation. The estimation of the *t*-th hidden image estimate $\hat{z}^t$ is:

$$\hat{z}^t = \hat{x}^t + \frac{\alpha^2}{\delta^2} S^T (y - S\hat{x}^t) \tag{7}$$

3.2.2. *M-Step*: Image Denoising Procedure

The *M-step* is to maximize the expectation (*Q*-function) in the *E-step* by updating the estimated HR image x according to:

$$\hat{x}^{t+1} = \arg\max_{x} \left\{ Q(x, \hat{x}^t) - q(x) \right\} \tag{8}$$

where $q(x)$ is a penalty function to x. When the log prior of x is used as the regularization, the optimization of Equation (8) becomes the MAP estimation,

$$\hat{x}^{t+1} = \arg\min_{x} \left\{ \|x - \hat{z}^t\|_2^2 + \gamma \log \text{pr}(x) \right\} \tag{9}$$

where $\text{pr}(x)$ is the prior of x, and γ is the trade-off parameter. Thus, the *M-step* is indeed an image denoising procedure, which combines the reconstruction constraint and prior knowledge to further improve the quality of the estimated HR image.

Thus, the entire EM algorithm for solving noisy image super-resolution can be summarized as:

- Likelihood of the super-resolution model approach *(E-step)*:

$$\hat{z}^t = \hat{x}^t + \frac{\alpha^2}{\delta^2} S^T (y - S\hat{x}^t)$$

- HR image estimation from a hidden image *(M-step)*:

$$\hat{x}^{t+1} = \arg\min_{x} \left\{ \|x - \hat{z}^t\|_2^2 + \gamma \log \text{pr}(x) \right\}$$

One can see the *E-step* can be easily achieved because only a simple linear transformation combination is applied. It can also be considered as a likelihood approach or a gradient descent method. Thus, the major difficulty in our image super-resolution method is the *M-step*. That is how to estimate the HR image x from the hidden image z.

4. HR Image Estimation via Maximum A Posterior

As previously mentioned, *M-step* is an image denoising procedure. We then borrow the basic idea of the adaptive non-local group-sparsity methods [27,28], which perform effectively in image denoising. In our work, the non-local group-sparsity explicit prior serves as the likelihood estimation. It formulates the *M-step* as an optimal filter design problem and determines the spectral coefficients of the filter by considering a local Bayesian prior. However, this local Bayesian prior requires the statistical distribution of the image patch from a similar targeted database, which is not practical in a VSN. To alleviate this problem, we consider the geometric duality existing between the LR and HR image patches as the implicit prior and jointly learn it to predict the mean and covariance of the target image patch from its respective LR image patch, rather than external databases using the *Gaussian process* [19]. Thus, the explicit prior serving as the likelihood estimation and the implicit prior regarded as the prior estimation in *M-step* can be jointly learned by the MAP procedure.

4.1. Non-Local Group-Sparsity Explicit Prior Learning

HR image $\hat{x}^{t+1}$ estimation based on non-local group sparsity can be treated as a linear image patch denoising filter design procedure, which is described as below.

Given a noisy image patch $q \in \mathbb{R}^d$ from the hidden image $\hat{z}^t$, estimate a linear transform operator (a filter) $\mathcal{A} \in \mathbb{R}^{d \times d}$ to make the estimation $p - \mathcal{A}q$ have the *minimum mean squared error (MMSE)*:

$$\mathcal{A} = \arg\min_{\mathcal{A}} \mathbb{E} \left[\|\mathcal{A}q - p\|_2^2 \right] \tag{10}$$

where p is the ground truth image patch. In general, we assume that $\mathcal{A}$ is symmetric and square. Thus, we can gain a better understanding of this linear transform by performing the *singular value decomposition* $\mathcal{A} = U\Lambda U^T$, where the dictionary $U \in \mathbb{R}^{d \times d}$ is an orthonormal matrix satisfying $U^T U = I$, and $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_d\} \in \mathbb{R}^{d \times d}$ is the diagonal matrix containing the spectral coefficients. Therefore, the optimization problem Equation (10) is rewritten by:

$$(U, \Lambda) = \arg\min_{U, \Lambda} \mathbb{E}\left[\left\|U\Lambda U^T q - p\right\|_2^2\right] \tag{11}$$

If the dictionary U is known, the MAP estimator of Λ has a closed-form solution, especially when the dictionary U is square and unitary. $U^T q$ is an operation that transforms the image patch from the space domain to the frequency domain. The symbol Λ is an element-wise shrinkage operator of the transform spectrum, where the true signal component is reserved and the noise is suppressed by the element-wise operation. Thus, the true information can be successfully separated from the noise by shrinkage. The denoised image patch is then obtained by multiplication by U, which transforms from the frequency domain back to the space domain. Repeat this process until all of the image patches have been denoised.

Thus, it is natural to determine the basis matrix U and to compute the MAP estimator of Λ. However, constructing dictionary U of image patch q has two issues: what input should we use, and how do we train the dictionary? We give the solution below.

Similar to the idea of the dictionary learning strategy in [29], the input training samples are obtained directly from the similar patches $\{p_j\}_{j=1}^k$, and this is modeled in the unit of the group. In other words, we search the similar patches within the noisy image. The similar patch is selected if the Euclidean distance between the patch q and the patch p_j is less than a threshold value. Each group $\{p_j\}_{j=1}^k$ is represented by the form of a matrix, denoted by P, which is composed of patches similar to the patch q.

With this first issue solved, we concentrate on the dictionary U learning from group P. Luo *et al.* [28] pointed out that a good dictionary U should satisfy the following two properties: first, the projected vector $\{U^T p_j\}_{j=1}^k$ should be similar in both magnitude and location, which is based on the observation that similar patches have a similar decomposition [27]; second, each projected vector $U^T p_j$ should be sparse. The more non-Gaussian image patches we have, the easier it is to distinguish from Gaussian noise because the noise is not sparse. Hence, this idea is more effective for denoising.

In order to satisfy the above-mentioned two criteria, we propose the group sparsity that is represented by a joint sparsity pattern; in which we introduce a so-called $\ell_{1,2}$ mixed norm matrix $\|A\|_{1,2} = (\sum_i \|a_i\|_1^2)^{1/2}$ where a_i is the i-th column of matrix A. Thus, the $\ell_{1,2}$ norm of the matrix $U^T P$ is minimized as follows:

$$\underset{U}{\text{minimize}} \quad \left\|U^T P\right\|_{1,2} \quad s.t. \quad U^T U = I \tag{12}$$

where the objective function is to minimize the joint sparsity error, and the constraint is to impose the orthogonality of U. This problem seems to be complex, but actually is identical to *principal component analysis (PCA)* [28,29]. We then compute:

$$[U, S] = \text{eig}\left(PP^T\right) \tag{13}$$

where U is the eigenvector matrix and S denotes the corresponding eigenvalue matrix.

When the dictionary of U has been learned, we will compute the optimal Λ by the *Bayesian minimum mean-square error (BMMSE)*. Most image denoising techniques learn "universal" image priors from a variety of scenes to guide the denoising for all kinds of images. Then, the Λ is obtained by

a simple hard thresholding whose value is empirically gained. Obviously, such Λ is less accurate. Luo [28] considered an image patch prior $f(\boldsymbol{p})$ learned from the similar targeted database, which makes Λ much more specific and accurate. However, having a targeted database is not practical in a VSN. We then consider that the geometric duality exists between the LR and HR image patches as the implicit prior and jointly learn it to predict the mean and covariance of the target image patch from its respective LR image patch, rather than the external database using the *Gaussian process* [19].

To have $f(\boldsymbol{p})$, we assume that the mean μ and covariance Σ of $f(\boldsymbol{p})$ are known. Then, the optimal Λ is derived by the following lemma:

Let $f(\boldsymbol{q}|\boldsymbol{p}) = \mathcal{N}\left(\boldsymbol{p}, \alpha^2 I\right)$, and let $f(\boldsymbol{p}) = \mathcal{N}\left(\mu, \Sigma\right)$ for any vector μ and matrix Σ; then, the optimal Λ that minimizes Equation (11) is:

$$\Lambda = \left(\mathrm{diag}\left(\boldsymbol{G} + \alpha^2 I\right)\right)^{-1}\mathrm{diag}\left(\boldsymbol{G}\right) \tag{14}$$

where $\boldsymbol{G} \overset{\mathrm{def}}{=} \boldsymbol{U}^T \mu \mu^T \boldsymbol{U} + \boldsymbol{U}^T \Sigma \boldsymbol{U}$.

4.2. Geometric Duality Implicit Prior Learning

In practice, $f(\boldsymbol{p})$ is unknown in Equation (14), so we cannot estimate the *a posteriori* probability of Λ. Instead, we could approximate the distribution of $f(\boldsymbol{p})$ from the geometric duality prior between the LR and HR image patches, rather than from a set of training example images, through the *Gaussian process*.

The obvious advantage of the *Gaussian process* over regression problems is that we can obtain the predictive distribution of the test output sample, such as the mean and covariance, rather than giving a hard assignment. We can directly estimate the output by learning a predictive function $g(x) : \mathcal{X} \to \mathcal{Y}$ from training data. Note that, here, x and y are defined at the local patch level, where x are the patches from the HR image and y are the corresponding patches from the LR interpolated image. We use a non-parametric model, which assumes a *Gaussian process* prior $y = g(x) \sim \mathcal{GP}(m(x), k(x_i, x_j))$ with $m(x) = 0$. The joint distribution of the training outputs and the test outputs is:

$$p(y|x) \sim \mathcal{N}\left(0, \begin{bmatrix} K(x^{\mathrm{tr}}, x^{\mathrm{tr}}) & K(x^{\mathrm{tr}}, x^{\mathrm{te}}) \\ K(x^{\mathrm{te}}, x^{\mathrm{tr}}) & K(x^{\mathrm{te}}, x^{\mathrm{te}}) \end{bmatrix}\right) \tag{15}$$

where $y = [y^{\mathrm{tr}}, y^{\mathrm{te}}]$, $y^{\mathrm{tr}} = [y_1^{\mathrm{train}}, \ldots, y_N^{\mathrm{train}}]$ are N training output samples and $y^{\mathrm{te}} = [y_1^{\mathrm{test}}, \ldots, y_M^{\mathrm{test}}]$ are M test output samples. $x = [x^{\mathrm{tr}}, x^{\mathrm{te}}]$ are the corresponding input samples. $K(x^{\mathrm{tr}}, x^{\mathrm{te}})$ denotes the $N \times M$ matrix of the covariances evaluated at all pairs of training input and test input points and similarly for the other entries $K(x^{\mathrm{tr}}, x^{\mathrm{tr}})$, $K(x^{\mathrm{te}}, x^{\mathrm{te}})$ and $K(x^{\mathrm{te}}, x^{\mathrm{tr}})$. For an input sample x^{test}, the posterior over the output sample y^{test} has a simple Gaussian form: $p(y^{\mathrm{test}}|x^{\mathrm{tr}}, y^{\mathrm{tr}}, x^{\mathrm{test}}) \sim \mathcal{N}(\mu_y, \Sigma_y)$, where:

$$\begin{aligned} \mu_y &= K(x^{\mathrm{te}}, x^{\mathrm{tr}})(K(x^{\mathrm{tr}}, x^{\mathrm{tr}}))^{-1} y^{\mathrm{tr}} \\ \Sigma_y &= K(x^{\mathrm{te}}, x^{\mathrm{te}}) - K(x^{\mathrm{te}}, x^{\mathrm{tr}})(K(x^{\mathrm{tr}}, x^{\mathrm{tr}}))^{-1} K(x^{\mathrm{tr}}, x^{\mathrm{te}}) \end{aligned} \tag{16}$$

Figure 1 shows an illustration of the *Gaussian process* for image patches' prior learning. In our setting, we use t-th estimated HR image $\hat{x}^t$ in the *M-step* as a training output, the corresponding interpolated image (obtained by interpolating the image $\hat{y}^t$ using the bicubic function, where $\hat{y}^t = S\hat{x}^t$ as the training input and the LR interpolated image (obtained by interpolating the LR image $\boldsymbol{y}$) as the testing input. Each 7×7 patch from the training input and the corresponding patch from the training output form a predictor-target training pair. In order to be more specific according to the local prior, the training is carried out in a 30×30 overlapped region separately from which the training pairs come.
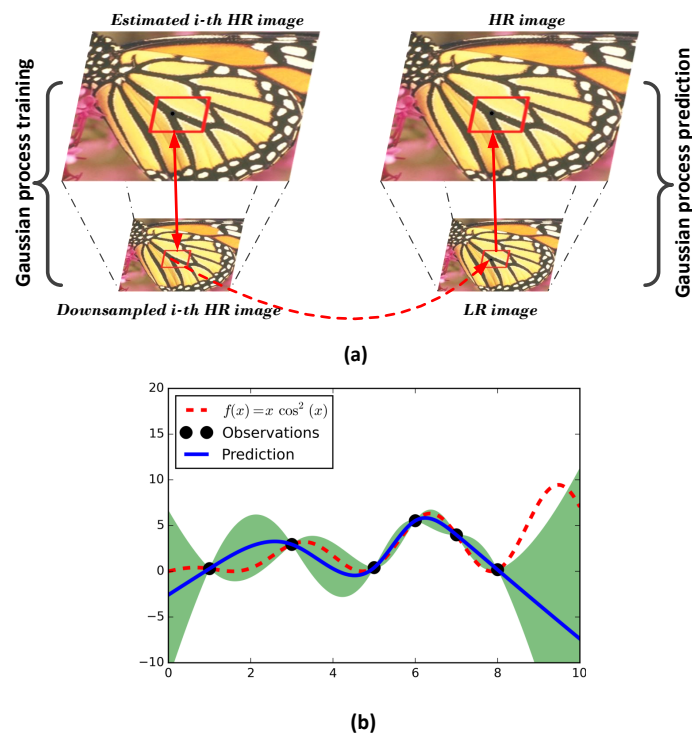
Figure 1. Illustration of the *Gaussian process* for image patch prior learning. (**a**) The overview of image patch prior learning. The left half is the training process, and the right half is the prediction process; (**b**) Example of one-dimensional data. The black points correspond to the observed training data points. The red dotted curve represents the true function. The blue solid curve is the prediction mean, and the green area is the variance.

4.3. Improving Similar Patches Match by Introducing Rotation Invariance

As mentioned in Subsection 4.1, the dictionary U is computed from the reference patches. Their group matrix P is composed of non-local patches similar to the patch q, where the patch selection/matching is performed by measuring the Euclidean distance similarity between the patch q and each of the patches from its image. The non-local similarity matching is based on the fractal nature of images, which suggests that patches of a natural image recur within the same image.

However, one problem is that the pattern may rarely repeat in the image. Thus, the non-local characteristics will not be sufficiently expressive to cover all of the patches. One expands the internal patch search space by allowing geometric variations achieved by affine transformations [30]. We propose a rotated non-local self-exemplars strategy for similar image patches' matching and improve the performance.

Our rotated non-local self-exemplars strategy finds that the most similar k images patches come from the rotated poses, rather than the original pose in the image. It alleviates the drawback induced by Euclidean distance similarity matching, which is computed by summating the square-error between the targeted pixels and corresponding pixels. By rotating the image, the similar image patches can be successfully matched by the Euclidean distance. In the illustration in Figure 2, we can see that the leftmost is the input noisy image; the rotated images are second from left; the image patches' matching is accomplished in the third from the left; and the noisy image patch is showed in the rightmost. An example of the rotated non-local self-exemplars strategy is given in Figure 3.
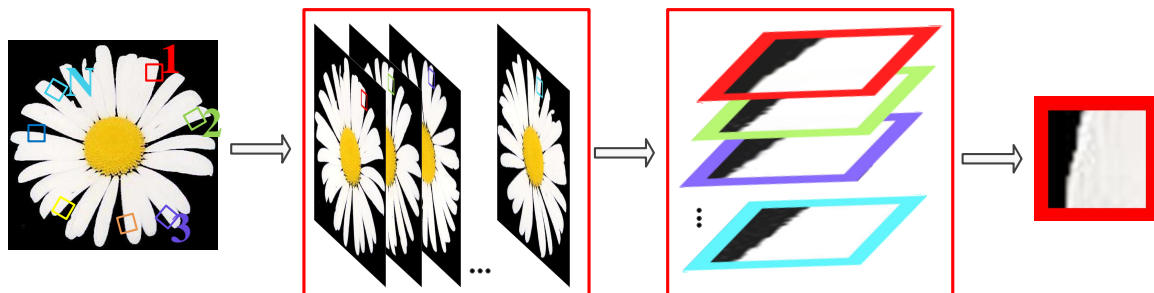
Figure 2. The rotated non-local self-exemplars strategy for the image filtering method.
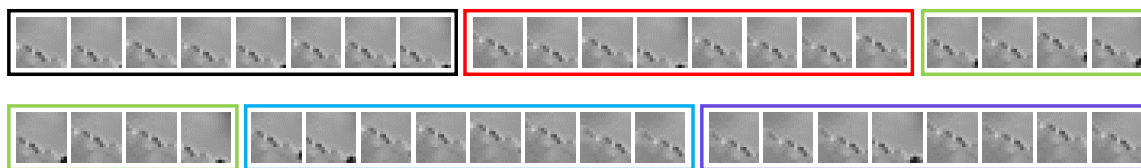
Figure 3. The rotated non-local self-exemplars strategy for the image patches' matching method. The patches in black box are from non-rotated image (rotation angle $\theta = 0°$). The patches in the red box are from the image with rotated angle $\theta = 45°$. The patches in the green box are from the image with rotated angle $\theta = 90°$. The patches in the blue box are from the image with rotated angle $\theta = 135°$. The patches in the purple box are from the image with rotated angle $\theta = 180°$.

5. Experiment Results and Discussion

To demonstrate the superior performance of our proposed super-resolution method, we first compare it to bicubic interpolation and five representative methods in the image super-resolution field on noiseless images. These are Yang's method based on the sparse representation prior [11], *Gaussian process* regression for super-resolution (GPR) [20], Zhang's method based on Markov random field prior learning [9], Dong's method based on adaptive sparse domain selection and adaptive regularization (ASDS) [8] and adjusted anchored neighbor regression (ANR) [31]. Then, a comparison is done on the synthesized noisy images to show the robustness of our algorithm to noise.

5.1. Experimental Configuration

In order to evaluate the image super-resolution results with objective measures, the LR images (training or test images) are generated from the original HR image by a 7×7 Gaussian blurring operator with a standard deviation of 1.6 and then downsampled by a factor of three, which is similar to the actual VSN cameras' imaging. Considering that the HR images of the VSN are hard to acquire, we use the basic open images in experiments. As most VSN applications use grey scale images/videos, we apply our algorithm to the illuminance channel only. For the other two color layers (Cb, Cr), we enlarge them using bicubic interpolation. For the noisy LR images in Section 5.3, the Gaussian noise is added to the LR images generated. As the parameter α in the EM algorithm impacts the effectiveness, we empirically set $\alpha = 0.8\delta + 1$. In the *M-step* HR image estimation procedure, the size of the image patch is set to 7×7, with one overlap pixel between adjacent patches. In the similar patches matching procedure, the number k of similar image patches is set to 40. The rotated angle of the non-local self-exemplars strategy is set to $\theta = 0°, 45°, 90°, 135°, 180°$, respectively. We evaluate the results of various methods both visually and qualitatively in *peak signal to noise ratio* (PSNR) and *structural similarity index measurement* (SSIM). Note that since we only work on the illuminance channel, the reported PSNR and SSIM are carried out only on the illuminance channel. Additionally, we evaluate the super-resolution capability of the different algorithms using twenty benchmark test images used in [31].

## 5.2. Comparison with Six Super-Resolution Algorithms

In this subsection, we evaluate the performance of six super-resolution methods (including the bicubic interpolation method) in comparison with the proposed algorithms on twenty benchmark test images used in [31], *i.e.*, *Water lily* (256 × 256), *Butterfly1* (256 × 256), *Starfish* (256 × 256), *Bike* (256 × 256), *Butterfly2* (256 × 256), *Leaves* (256 × 256) and *Roof* (256 × 256). To visualize the performance difference, we magnify the region (red box) in each test image. Figures 4–6 show the visual comparisons. Then, we give the partially enlarged visual views in Figure 7. Figure 8 gives the numerical results. Figure 4a shows the result of the bicubic interpolation algorithm, which includes both visually displeasing blurred textural details and serious jagged artifacts along the edges. Figure 4b shows the result obtained by Yang's method. Although this method produces sharper edges than bicubic interpolation, there is no further detail added. Figure 4c generates a relative high-quality HR image with many fine details. There are still some unpleasing artifacts along major edges and an over-smoothed region. Figure 4d illustrates the result of the GPR method. This method can produce an HR image of high quality and with rich artifacts. However, the result has both jagged edge artifacts and the preservation of annoying textural details, especially around the edge of the lotus leaf. Figure 4e shows the result obtained by Zhang's method. It produces over-smoothed results and eliminates much of the image details. It fails to reconstruct fine image edges. Figure 4f illustrates the result of the ASDS method. It performs well in synthesizing many fine details, but there are some noticeable blurring details along dominant edges. Our proposed method (Figure 4g) can produce a visually comparable result to the ASDS method. The enhanced result is more faithful to the original HR image in terms of finer details and sharper edges by exploiting repetitive patterns to suppress the unexpected artifacts that most example learning-based approaches will produce. Figures 5–7 validate the above description. To sum up, our method reduces the annoying artifacts and leads to a more faithful super-resolution reconstruction. To further validate our proposed super-resolution algorithm, we present the PSNR and SSIM comparisons of the methods in Figure 8. In the comparative super-resolution methods, Zhang's method has higher values for PSNR and SSIM. However, its visual effect is the worst. We can see that our proposed method is consistently better than the compared methods, not only for the pleasing visual results, but also the better PSNR and SSIM.
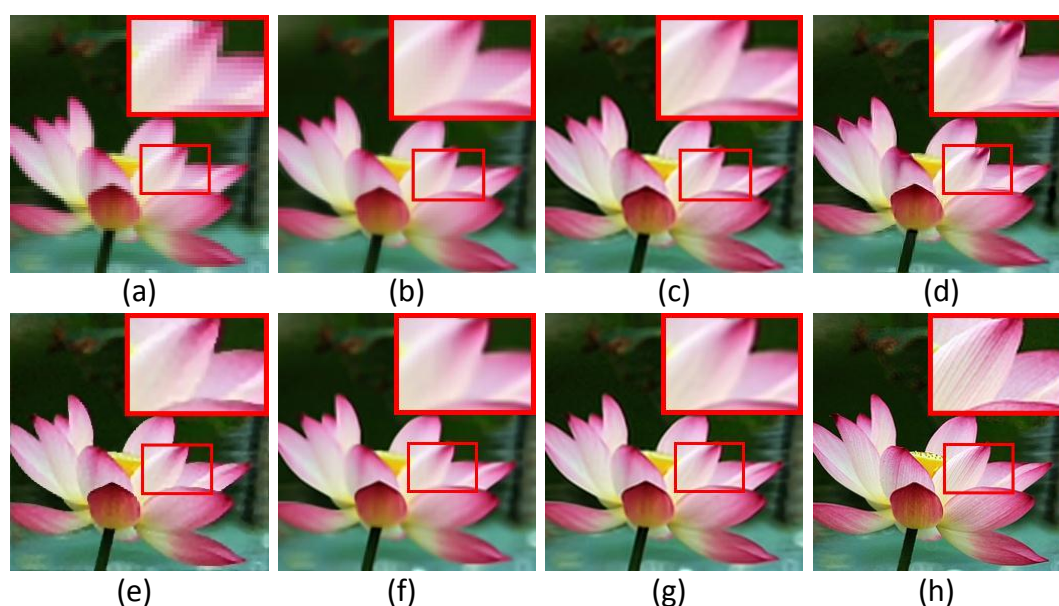
Figure 4. Comparisons with various image super-resolution methods on the image of *Water lily*. (**a**) Bicubic interpolation; (**b**) Yang's method [11]; (**c**) anchored neighbor regression (ANR) [31]; (**d**) *Gaussian process* regression (GPR) [20]; (**e**) Zhang's method [9]; (**f**) adaptive sparse domain selection (ASDS) [8]; (**g**) joint-prior image super-resolution (JPISR) method; (**h**) ground-truth.
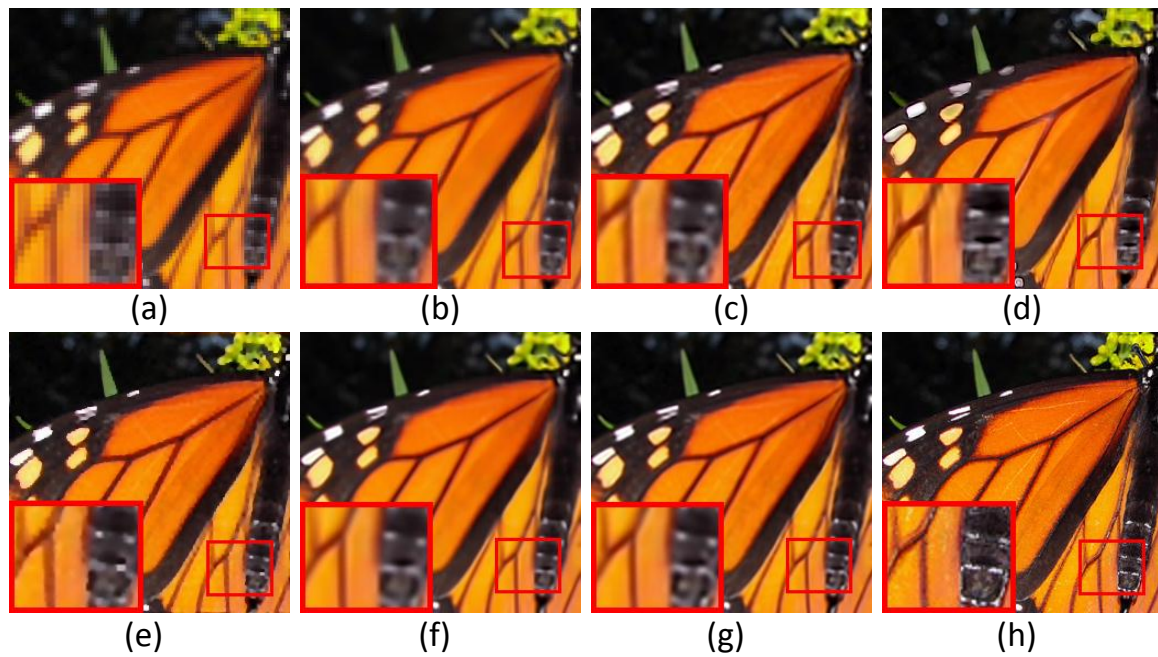
Figure 5. Comparisons with various image super-resolution methods on the image of *Butterfly1*. (**a**) Bicubic interpolation; (**b**) Yang's method [11]; (**c**) ANR [31]; (**d**) GPR [20]; (**e**) Zhang's method [9]; (**f**) ASDS [8]; (**g**) JPISR method; (**h**) ground-truth.
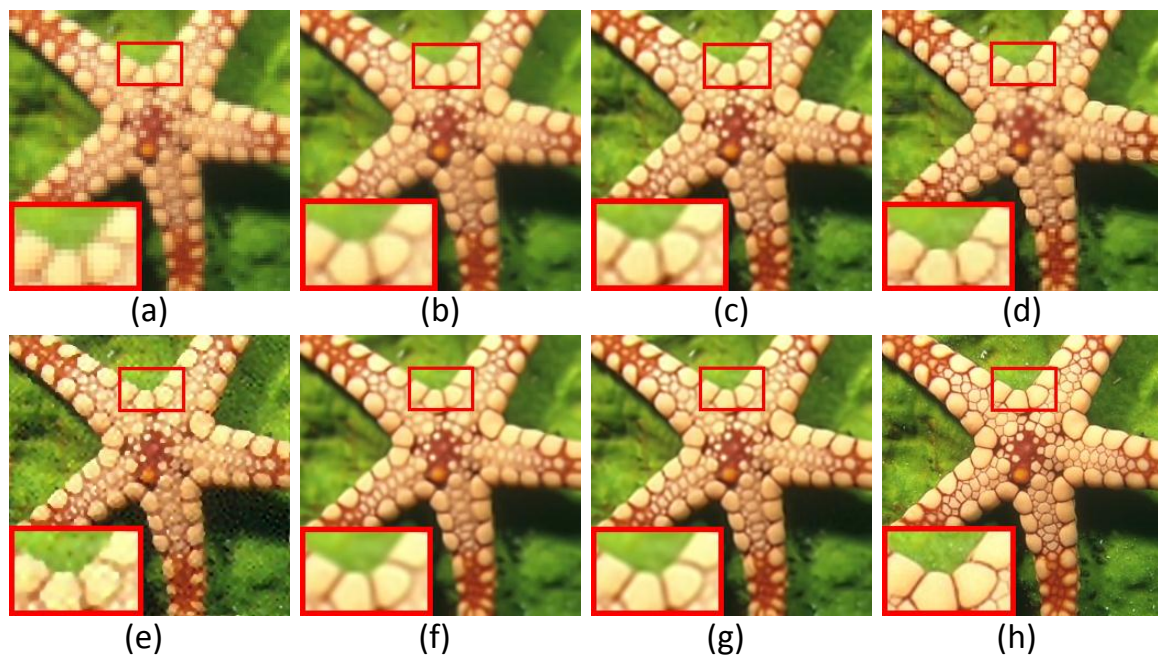

Figure 6. Comparisons with various image super-resolution methods on the image of *Starfish*. (**a**) Bicubic interpolation; (**b**) Yang's method [11]; (**c**) ANR [31]; (**d**) GPR [20]; (**e**) Zhang's method [9]; (**f**) ASDS [8]; (**g**) JPISR method; (**h**) ground-truth.
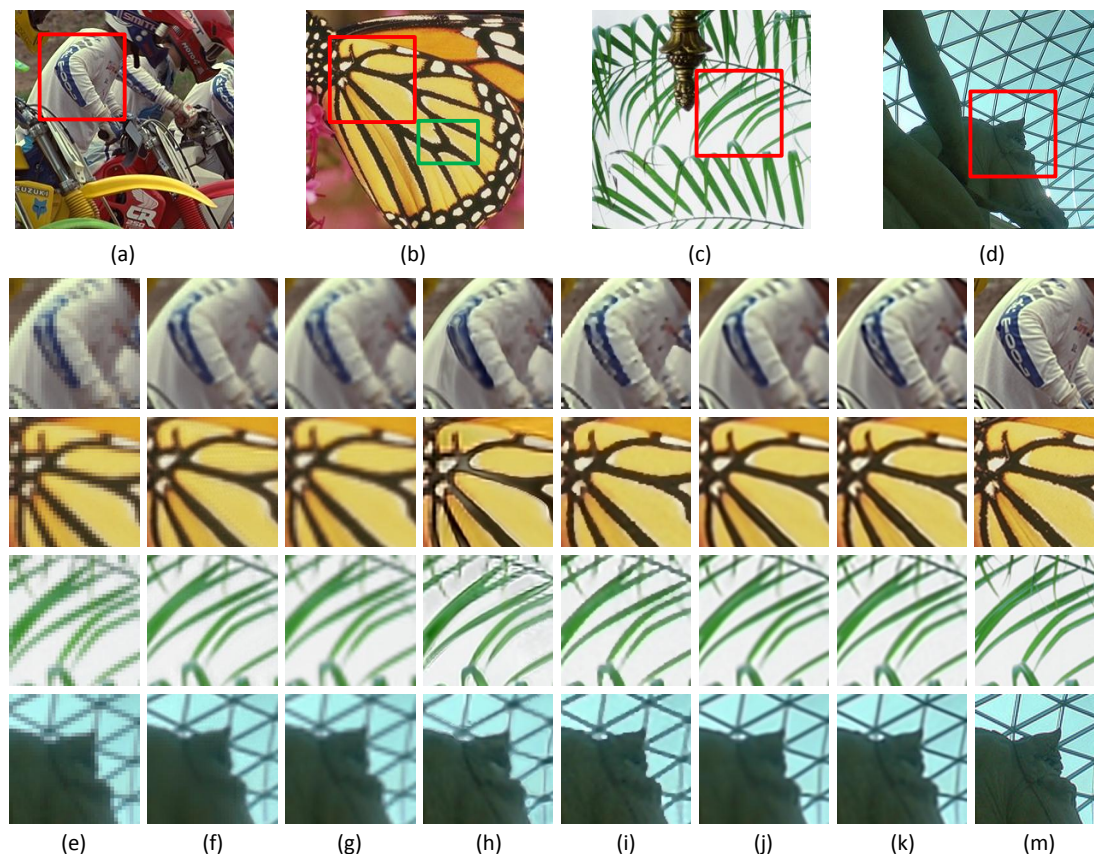
Figure 7. The magnified super-resolution views of the region (red box), where each row represents the same LR image region and each column represents the same method. (**a**) *Bike*; (**b**) *Butterfly2*; (**c**) *Leaves*; (**d**) *Roof*; (**e**) bicubic interpolation; (**f**) Yang's method [11]; (**g**) ANR [31]; (**h**) GPR [20]; (**i**) Zhang's method [9]; (**j**) ASDS [8]; (**k**) JPISR method; (**m**) ground-truth.
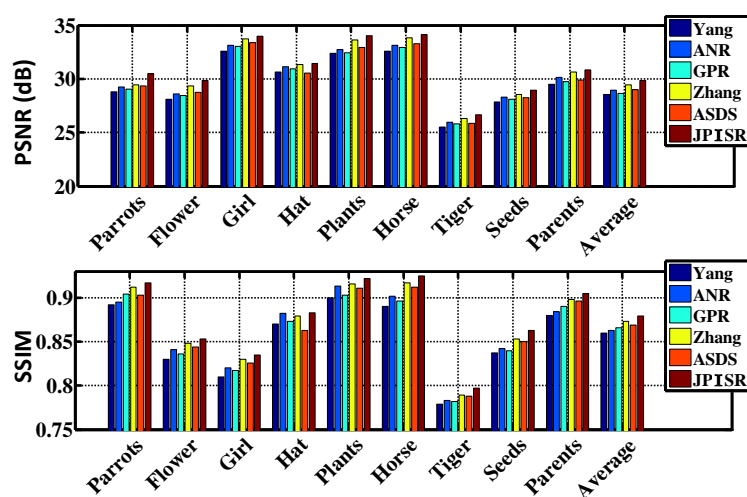
Figure 8. PSNR and SSIM of Yang's method [11], ANR [31], GPR [20], Zhang's method [9], ASDS [8] and the JPISR method.

5.3. Comparison on Noisy Images

In practice, the LR image is often noise corrupted in the VSN due to the working condition, which makes the super-resolution more challenging. We added Gaussian white noise (with standard deviations of 5, 10, 15 and 20, respectively) to the LR images, and the reconstructed HR images are

shown in Figures 9–11. From Figures 9 and 10, we can see that Yang's method, ANR, Zhang's method and the ASDS method are sensitive to noise. The severe noise-caused artifacts can be found around the edges. The GPR method results in over-smoothed HR images. In contrast, the proposed method is more robust to noise. Figure 11 gives the more detailed visual results for comparison. Meanwhile, we give the numerical results under different noise levels in Figure 12b,c. It clearly shows that JPISR has a distant capability of suppressing noise, especially for sever noise.
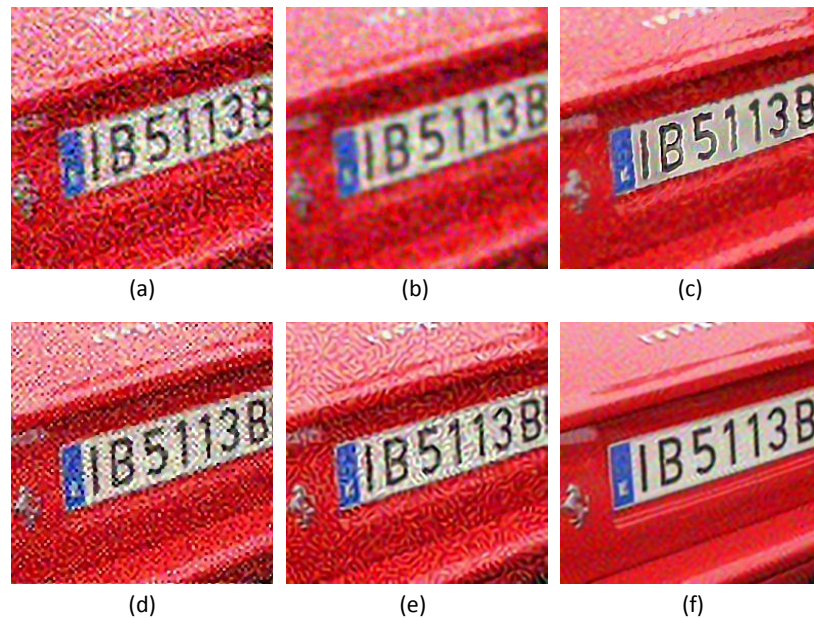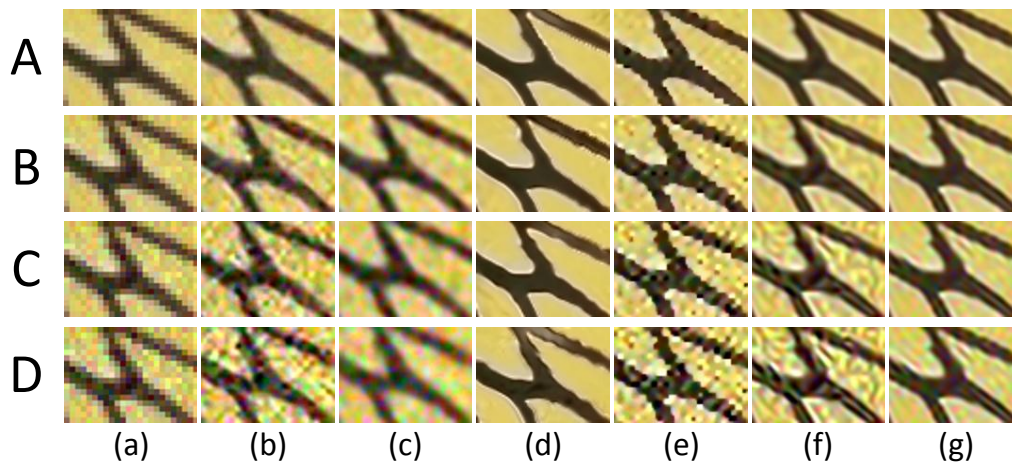

Figure 9. The super-resolution result on LR image *Car tail* with a noise deviation of 20. (**a**) Yang's method [11]; (**b**) ANR [31]; (**c**) GPR [20]; (**d**) Zhang's method [9]; (**e**) ASDS [8]; (**f**) JPISR method.
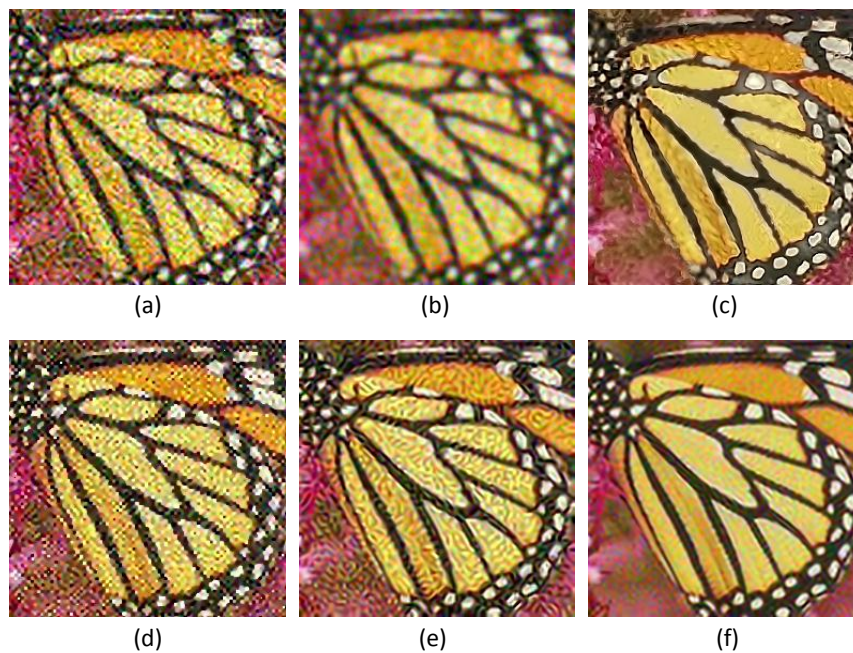
Figure 10. The super-resolution result on LR image *Butterfly 2* with a noise deviation of 20. (**a**) Yang's method [11]; (**b**) ANR [31]; (**c**) GPR [20]; (**d**) Zhang's method [9]; (**e**) ASDS [8]; (**f**) JPISR method.

Figure 11. The magnified super-resolution views of region (green box) in Figure 7, where each row represents the same noise variance and each column represents the same method. Rows A, B, C, D are the results with noise deviations of 5, 10, 15, 20, respectively. (**a**) Bicubic interpolation; (**b**) Yang's method [11]; (**c**) ANR [31]; (**d**) GPR [20]; (**e**) Zhang's method [9]; (**f**) ASDS [8]; (**g**) JPISR method.

5.4. Comparison of Data Size with PSNR

It is well known that most super-resolution methods require large external datasets for training. However, this is not practical in the VSN. in contrast, JPISR uses the non-local group-sparsity and the geometric duality for joint-prior learning, which does not rely on the external training data. Figure 12a gives the comparison of data sizes and PSNR values of five methods in the experiment. We can see that all other methods need enormous training data, except for JPISR and GPR. However, our method still achieves the highest PSNR value, even compared to Zhang's method.
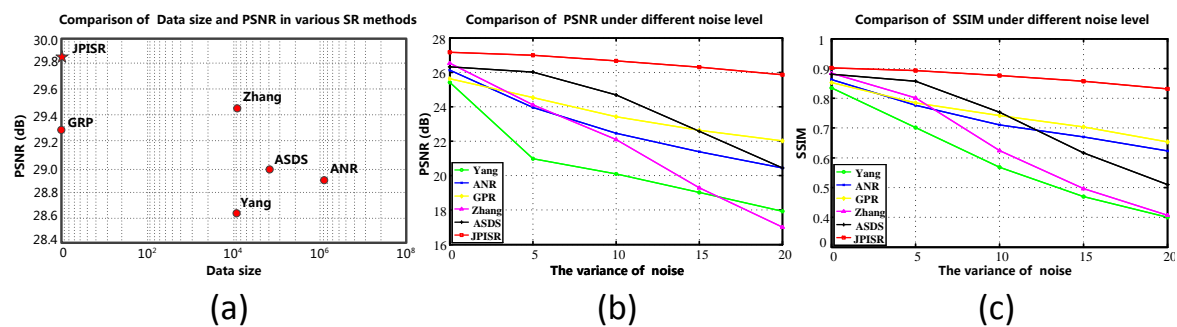
Figure 12. (**a**) Comparison with five super-resolution methods on PSNR values against data sizes; (**b**) comparison of PSNR under different noise levels; (**c**) comparison of SSIM under different noise levels.

6. Conclusions

Due to its flexibility and low cost, the VSN has attracted more interest in the last few years and is expected to play a major role in the evolution of smart sensing, data collaborative processing and communication capabilities. Unfortunately, in many cases, the images captured by live cameras are often of low resolution with noise due to the environment or equipment limitations. To make the quality of the captured image more suitable for analysis in various surveillance applications, we proposed a novel framework of prior-adaptive image super-resolution based on the EM algorithm. It inherently combined the super-resolution characteristics (implicit prior) with the image filtering method (explicit prior) to upscale and denoise the LR images captured from the low-cost visual nodes in the VSN.

In addition, the proposed joint-prior learning does not rely on the external training data, which is versatile for hostile environments, such as video surveillance in the wild, traffic monitoring, *etc.*

Although our method shows potential for VSN image SR, two aspects need to be considered in our future research. First, our proposed method is relatively time consuming, which cannot meet the real-time requirement. Second, all SR methods can only be trained under a specific degrading process, which restricts practical use. In future work, we will do a parallel implementation on GPU acceleration. Meanwhile, we plan to estimate the camera's degradation parameters by the blind restoration strategy.

Author Contributions: Bo Yue proposed the original algorithm and wrote this paper; Shuang Wang and Xuefeng Liang revisited the paper and supervised a whole process; Licheng Jiao and Caijin Xu gave some valuable suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Babacan, S.D.; Molina, R.; Katsaggelos, A.K. Variational Bayesian super resolution. *IEEE Trans. Image Process.* **2011**, *20*, 984–999.
2. Unger, M.; Pock, T.; Werlberger, M.; Bischof, H. A convex approach for variational super-resolution. *Patt. Recogn.* **2010**, 313–322.
3. Zhao, N.; Wei, Q.; Basarab, A.; Kouame, D.; Tourneret, J.-Y. Fast Single Image Super-Resolution. Available online: http://arxiv.org/abs/1510.00143 (accessed on 15 February 2016).
4. Efrat, N.; Glasner, D.; Apartsin, A.; Nadler, B.; Levin, A. Accurate blur models *vs.* image priors in single image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Austria, 10–13 December 2013; pp. 2832–2839.
5. Kang, W.; Yu, S.; Ko, S.; Paik, J. Multisensor Super Resolution Using Directionally-Adaptive Regularization for UAV Images. *Sensors* **2015**, *15*, 12053–12079.
6. Sajjad, M.; Mehmood, I.; Baik, S. Sparse representations-based super-resolution of key-frames extracted from frames-sequences generated by a visual sensor network. *Sensors* **2014**, *14*, 3652–3674.
7. Sun, J.; Sun, J.; Xu, Z.; Shum, H.-Y. Image super-resolution using gradient profile prior. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, USA, 24–26 June 2008; pp. 1–8.
8. Dong, W.; Zhang, D.; Shi, G.; Wu, X. Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. *IEEE Trans. Image Process.* **2011**, *20*, 1838–1857.
9. Zhang, H.; Zhang, Y.; Li, H.; Huang, T.S. Generative Bayesian image super resolution with natural image prior. *IEEE Trans. Image Process.* **2012**, *21*, 4054–4067.
10. Yang, C.-Y.; Huang, J.-B.; Yang, M.-H. Exploiting Self-Similarities for Single Frame Super-Resolution. In *Computer Vision—ACCV*; Springer Berlin Heidelberg: Berlin, Germany, 2010; pp. 497–510.
11. Yang, J.; Wright, J.; Huang, T.S.; Yi, M. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 497–510.
12. Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; Huang, T. Coupled dictionary training for image super-resolution. *IEEE Trans. Image Process.* **2012**, *21*, 3467–3478.
13. Roman, Z.; Michael, E.; Matan, P. On single Image Scale-Up Using Sparse-Representations. In *Curves and Surfaces*; Springer Berlin Heidelberg: Berlin, Germany, 2012; pp. 711–730.
14. Zhang, K.; Tao, D.; Gao, X.; Li, X.; Xiong, Z. Learning Multiple Linear Mappings for Efficient Single Image Super-Resolution. *IEEE Trans. Image Process.* **2015**, *24*, 846–861.
15. Cui, Z.; Chang, H.; Shan, S.; Zhong, B.; Lin, X.; Chen, Z. Deep Network Cascade for Image Super-Resolution. In *Computer Vision—ECCV 2014*; Springer International Publishing: Cham, Switzerland, 2014; pp. 49–64.

16.	Dong, C.; Chen, C.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In *Computer Vision—ECCV 2014*; Springer International Publishing: Cham, Switzerland, 2014; pp. 184–199.

17.	Li, X.; Orchard, M.T. New edge-directed interpolation. *IEEE Trans. Image Process.* **2001**, *10*, 1521–1527.

18.	Chang, H.; Yeung, D.-Y.; Xiong, Y. Super-resolution through neighbor embedding. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Washington, WA, USA, 27 June–2 July 2004; doi:10.1109/CVPR.2004.1315043.

19.	Rasmussen, C.E. Regression. In *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; pp. 32–58.

20.	He, H.; Siu, W.-C. Single image super-resolution using Gaussian process regression. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 20–25 June 2011; pp. 449–456.

21.	Zhu, J.; Javed, O.; Liu, J.; Yu, Q.; Cheng, H.; Sawhney, H. Pedestrian Detection in Low-Resolution Imagery by Learning Multi-scale Intrinsic Motion Structures (MIMS). In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2015; pp. 3510–3517.

22.	Jiang, N.; Liu, W.; Su, H.; Wu, Y. Tracking low resolution objects by metric preservation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 20–25 June 2011; pp. 1329–1336.

23.	Li, J.; Wu, J.; Deng, H.; Liu, J. A Self-Learning Image Super-Resolution Method via Sparse Representation and Non-Local Similarity. *Neurocomputing* **2015**, doi:10.1016/j.neucom.2015.07.139.

24.	Sajjad, M.; Mehmood, I.; Baik, S.W. Image super-resolution using sparse coding over redundant dictionary based on effective image representations. *J. Vis. Commun. Image Represent.* **2015**, *26*, 50–65.

25.	Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-based super-resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65.

26.	Figueiredo, M.; Nowak, R.D. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.* **2003**, *12*, 906–916.

27.	Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Non-local sparse models for image restoration. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 2272–2279.

28.	Luo, E.; Chan, S.H.; Nguyen, T.Q. Adaptive Image Denoising by Targeted Databases. *IEEE Trans. Image Process.* **2015**, *24*, 2167–2181.

29.	Zhang, J.; Zhao, D.; Gao, W. Group-based sparse representation for image restoration. *IEEE Trans. Image Process.* **2014**, *23*, 3336–3351.

30.	Huang, J.-B.; Singh, A.; Ahuja, N. Single Image Super-Resolution from Transformed Self-Exemplars. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.

31.	Timofte, R.; Vincent, D.; GoolLuc, V. A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution. In *Computer Vision—ACCV 2014*; Springer International Publishing: Cham, Switzerland, 2014; pp. 111–126.