A Video Super Resolution Framework Using SCoBeP

Nafise Barzigar, Aminmohammad Roozgard, Pramode Verma, and Samuel Cheng

***Abstract*—Super resolution as an exciting application in image processing was studied widely in the literature. This paper presents new approaches to video super resolution, based on sparse coding and belief propagation. First, find candidate match pixels on multiple frames using sparse coding and belief propagation. Second, incorporate information from these candidate pixels with weights computed using the Nonlocal-Means (NLM) method in the first approach or using SCoBeP method in the second approach. The effectiveness of the proposed methods is demonstrated for both synthetic and real video sequences in the experiment section. In addition, the experimental results show that our models are naturally robust in handling super resolution on video sequences affected by scene motions and/or small camera motions.**

***Index Terms*—Super Resolution, Sparse Coding, Belief Propagation, Nonlocal-Means Filter**

I. Introduction

SUPER RESOLUTION tries to combine several low resolution (LR) images from a scene and produces one higher resolution image with better optical resolution. This is an inverse problem that is commonly tackled by integrating denoising, deblurring, and upsampling.

Fig. 1 illustrates this inverse process and presents how the LR sequence may be modeled using an original higher resolution frame. During imaging, the blurring effect can be modeled by the optical point spread function (PSF). The scene may then be warped due to camera or object motion. Moreover, the motion effect might not be the same for all frames in the sequence. A fixed decimation operator is typically used to model the effect of sampling by the image sensor. The operator is characterized by the resolution ratio between the original higher resolution frame and the LR sequence. The noise, which in most applications assumed to be white i.i.d. Gaussian, is added to the LR frames. The outcome of the super resolution reconstruction problem depends on the involved operators and noise characteristics of the above mentioned model.

A wide variety of super resolution methods have been studied in the last two decades [1]–[10]. Huang and Tsai were the first to address the multiframe super resolution problem using a frequency domain approach that works for band limited and noise-free images [11]. Later, it was extended by others, such as Kim *et al.* who proposed a super resolution method on noisy and blurred images [12]. Pleg and Irani [1] also

The authors are with the Department of Electrical and Computer Engineering, the University of Oklahoma, Tulsa, OK, 74135.
E-mail: barzigar@ou.edu

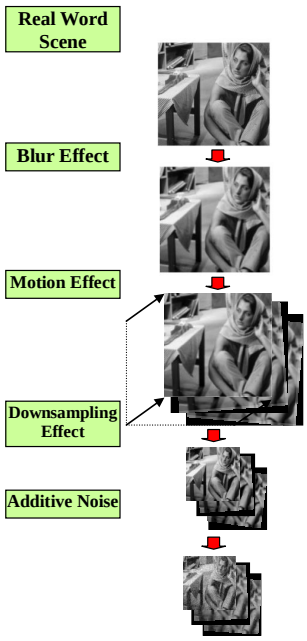


Fig. 1: A general model of multi-frame super resolution.

suggested a different approach for the super resolution problem based on the iterative backprojection (IBP) method adopted from computer aided tomography (CAT). Recently, an iterative multiframe super resolution method was presented in [9] that relied on extending the steerable kernel method in space-time. However, the approach assumes the input frames only contain smooth textures. Also, it has difficulties to estimate a pixel-wise motion in regions with the larger motion [13].

In dealing with camera position variation, a few attempts have been made through a global motion model [4], [5], [11], [14]. Bonchev and Alexiev suggested a method of super resolution that used the information from several LR frames by controlling the camera position in frequency domain when taking frames [5]. Also, in [14] a maximum a posteriori (MAP) was adopted to provide coarse estimates of rotation and translation between images. The authors claimed that such estimation step provided enough accuracy to effectively remove the effect of the rotational and coarse (super-pixel) translational motion between the images. Although that algorithm incorporates smoothness priors as a constraint to reconstruct the HR images, using these smoothness priors might not lead to smooth results [15]. A number of super resolution approaches using Total Variation (TV) regularization terms have been explored in the last decade, e.g., the approach by

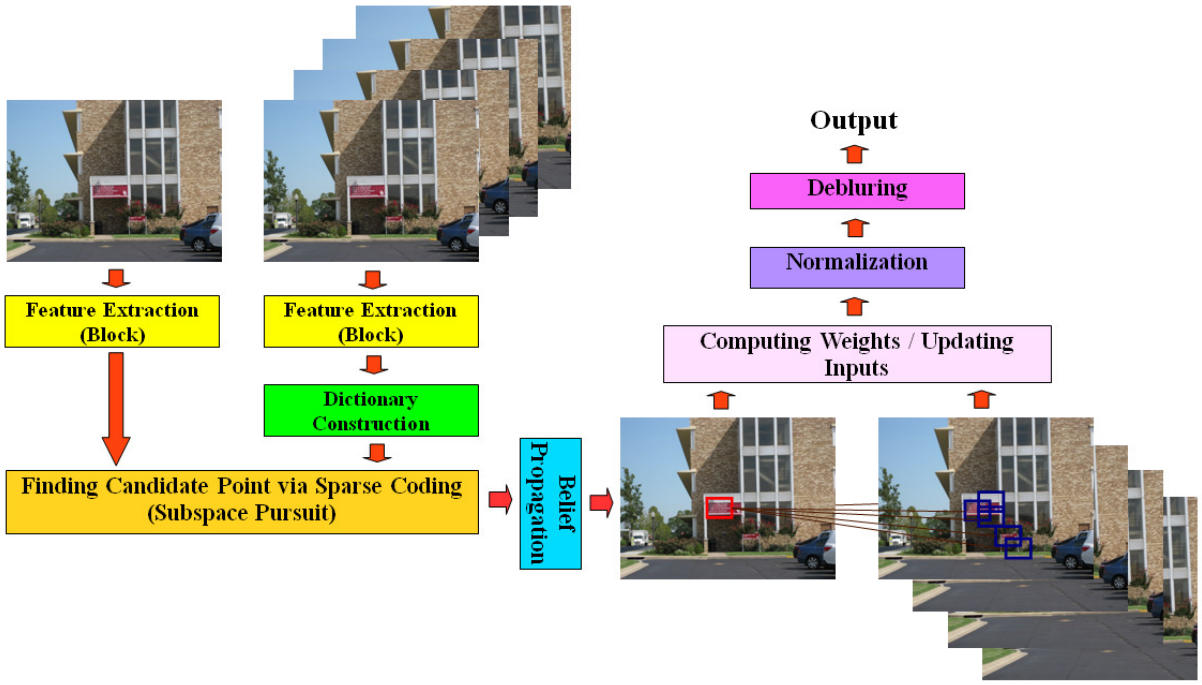


Fig. 2: Block diagram representation of our models.

Mitzel *et al.* in [7], where their method is not restricted to any particular motion model and they do not assume that the motion is known. In another work, Farsiu *et al.* suggested a multiframe super resolution method by applying constraints on the L_1 norm of both the bilateral TV regularization term and the data fusion term to produce a sharp, high resolution (HR) image [4]. The researchers also registered the LR images with respect to a reference frame before starting the super resolution process. Liu and Sun in [13] proposed a Bayesian framework for adaptive video super resolution that deals with video super resolution by also simultaneously estimating underlying large motion. Moreover, they jointly estimated the flow field and the noise level in a coarse-to-fine manner on a Gaussian image pyramid using the HR image and the blur kernel.

In this work, we focus on video frames that suffer from non-homogeneous noise, atmosphere or camera blur, motion and down-sampling effects. Also, as real videos can be taken from both fixed or movable cameras, we also consider frames affected by scene motions and/or small camera motions. Note that as we will see in the coming sections, the approach we introduced here works well for such scenario. Our method is complementary to approaches such as [13], which considers the superresolution of sequences with larger camera motions but little or no scene motions.

We propose to solve the super resolution problem using a novel framework taking advantage of two recently developed techniques: SCoBeP [16] and Nonlocal-Means (NLM) [17]. Our approaches are based on the concept of Sparse Coding and Belief Propagation (SCoBeP) which is earlier introduced in [16] for 2-D signals (images). It turns out that the technique is well-suited for super resolution of video (a 3-D signal) as we explore in this paper.

As a summary of our approaches, we first build an over-complete dictionary out of all block features of LR frames as shown in Fig. 2. Different from [18], we are not generating HR/LR patch pairs from the frames by exploiting self-similarities. For each pixel of the initial estimate of the HR frame, we then select a set of candidate pixels out of the constructed dictionary using sparse coding [19]. The match score of each candidate pixel will be evaluated taking both local and neighboring information into account using belief propagation [20]. The best matches will be selected as the candidates with the highest scores. An occluded pixel or any pixel not covered by the LR frames is likely to be identified since the match scores in this case will be significantly smaller than a typical maximum score when a match pixel actually exists. Finally, in our first proposed method, the NLM approach exploits similarity in patches around candidate pixels to average out the noise among similar patches [6] and in our second proposed method, a pixel is reconstruct from multiple candidate pixels with the weights extracted directly from the output of SCoBeP.

In the experiment section, we also illustrate that the proposed methods can perform well on real LR videos (besides "phantom" LR videos generated artificially) and can reconstruct image edges with high fidelity. Although the NLM filtering has shown great potentials for image denoising and superresolution [17], it is only effective when a reference patch can be identified to accurately represent the targeted patch. Moreover, NLM approaches generally have very high computational complexity. We will show in this paper reference patches can be effectively found by SCoBeP. Further, we will also show that the NLM step may be skipped completely (as demonstrated in the second method) with only a small performance penalty but a significant (about three times) speed-up.

TABLE I: Summary of Notation

Notation	Description
$\mathcal{X}$	the reconstructed HR frame
$\mathcal{Y}_t$	the LR frame at t
y_t	the interpolated LR frame at t
T	the number of available LR frames
$\mathcal{Z}$	the blurred version of the reconstructed HR frame
$\mathcal{N}(q,l)$	a neighborhood around pixel $[q,l]$
R_{ql}	the matrix that extracts a patch centered around pixel $[q,l]$
r	the resolution (magnification) ratio
n	the number of candidate pixels
X_{ql}	the vectorized patch at pixel $[q,l]$ in the reference frame
D_t	the dictionary constructed from the t^{th} LR frame $\mathcal{Y}_t$
α_{qlt}	the sparse representation vector of a patch centered around pixel $[q,l]$ in t^{th} LR frame
Y_{qlt}	the vectorized patch at pixel $[q,l]$ in the t^{th} LR frame $\mathcal{Y}_t$
$\mathcal{W}[i,j,q,l,b,t]$	the weight mapping from a pixel in the reference frame to b^{th} candidate pixel in the t^{th} interpolated LR frame y_t
cp	an $n \times 2$ matrix storing the locations of the candidate pixels
ρ_{qlt}	the prior probability of pixel $[q,l]$ in the reference frame mapping to the b^{th} candidate pixel in t^{th} interpolated LR frame

The rest of this paper is structured as follows. We give a brief summary of related work and review the background of super resolution and NLM filter in the next section. In Section III, we introduce our proposed methods: SCoBeP-SR and SCoBeP-NLM. Implementation issues are also presented in detail in this section. Section IV presents the experimental results and compares our results with that of the existing super resolution methods. Finally, future work is outlined and concluding remarks are made in Section V.

II. Related Work And Background

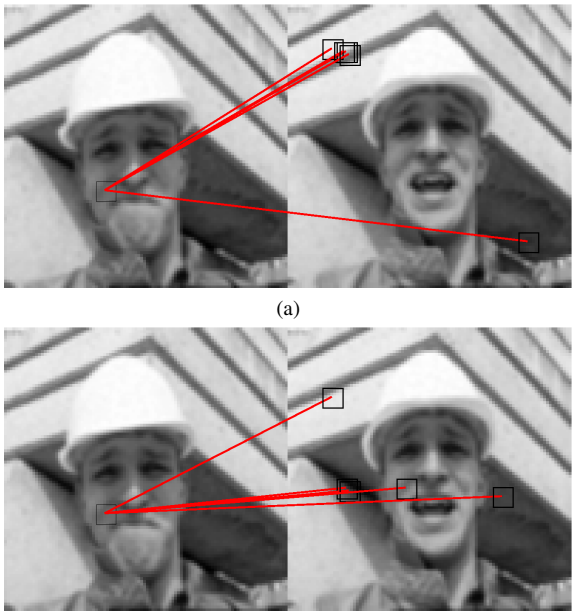
Sparse representation [21] and self-similar-based techniques [3], [10] have been used in super resolution in recent years. In this section, we will review how some of the recent works use these techniques in recovering the downsampled signals and computing the similarity of image patches.

In [21], [22], a large set (of the order of a hundred thousand) of patches randomly sampled from natural images to train an LR and a HR dictionaries. The main idea consists of seeking in the database for a sparse representation of each patch of the LR input, followed by using this representation to generate the HR output. Yang *et al.* [18] proposed a super-resolution method that exploits self-similarities and group structural constraints of image patches using only one single input frame. In this algorithm, the patch self-similarity within the image is exploited and the group sparsity then will be introduced for better regularization in the reconstruction process. Another recent example based on an enhanced sparse representation in transform domain is block-matching 3-D filter (BM3D) [23], which uses a block matching technique to find a set of similar 2D blocks. Danielyan *et al.* have extended (BM3D) in [8] for image and video super resolution. They produce a sparse representation of the true signal in the transform domain to exploit the similarity among the blocks. In contrast to the sparse representation approaches discussed above where they use information from only one corresponding pixel per LR frame to reconstruct a target pixel, our first approach incorporates the NLM method to take advantage information from multiple matched pixels for the reconstruction.

We now turn to a discussion of certain works associated with self-similar-based technique. Plenty of works have emerged lately based on self-similarity for natural image and video processing. The self-similarity property shows that the image content desires to repeat itself within some neighborhoods. Non local self-similarity has been effectively applied to many aspects of image processing [3], [24], [25]. Following this insight, Buades *et al.* used this approach in image denoising, which is known as the NLM method [17]. The NLM method was used also in image restoration explicitly exploits self-similarities in natural images [3], [17]. Liu and Freeman in [26] proposed a video denoising approach to use an approximate k-nearest neighbor (AKNN) algorithm to approximately but rapidly seek the most similar patches for a given video. As pointed out in [27], [26] takes into account only similar blocks for a given video and thus could be classified as a "closest structure" method, while one can call the original NLM method [17] a "closest space" method in the sense that it uses only closest blocks in a small window. Moreover, Marial *et al.* in [25] extended the NLM method in denoising and demosaicking using the idea that similar patches have similar sparsity patterns. Also, in [10], Zhang *et al.* proposed a non-local kernel regression method for image and video super resolution, which exploits both non-local self-similarity and local structural regularity in a single model. Distinct from the local kernel regression, the NLM method estimates the value of a pixel from all possible patches collected from a search area, and breaks the locality constraint in the restoration algorithms. Protter *et al.* [6] generalized this denoising method to perform multiframe super resolution reconstruction with no explicit motion estimation. In that work, computing the similarity of video frame patches resulted in probabilistic estimates of motion.

Prior works have been limited to block matching in restricted neighborhoods. These neighborhoods determine the candidate matches of target pixels and thus have a significant impact on SR performance. However, they have always been assigned with limited sizes and regular shapes (e.g., as rectangular blocks) in prior works and hence often do not include the best match patches. Due to this poor block matching, the prior techniques could suffer from block artifacts in some test cases [6], [28].

In contrast, the advantage of SCoBeP registration is that the chosen candidates will have better "diversity" when compared with the AKNN or even the *exact* K-nearest neighobor (KNN) approach (see Fig. 3). This originates from the induced orthogonality of the patches when sparsity is imposed in the solution. In the KNN case, when a smooth patch is incorrectly matched to a patch, the next best "matches" are likely around the neighborhood of the wrong patch and it ends up incorrect matching for all patches. The better "diversity" actually affords SCoBeP a larger search window compared with other registration without sacrificing the robustness of the approach [16]. In this paper, we take advantage of SCoBeP to select from each LR frame a set of candidate pixels which are likely to be most similar to the target pixel. As a result, for each pixel



(a) (b)

Fig. 3: Candidate points obtained by KNN and sparse coding. The images in (a) shows that KNN tends to result in candidate points with poor diversity. And thus it can easily miss including the true corresponding point as one of its candidate points. In contrast, the images in (b) show that the candidate points of sparse coding tend to diversify and thus is more likely to include the true corresponding point.

and per LR frame, we have an irregular neighborhood that can include any pixel in the frame. This significantly improves the block matching performance that directly links to the overall SR performance.

The contributions of the proposed approach can be summarized as follows: (i) Identifying the best similar matches based on the local and the geometric characteristics using factor graph, which efficiently trade off both characteristics optimally. (ii) Incorporating SCoBeP to efficiently sift through pixels from all LR images and thus resulting much more accurate non-local candidates for subsequent estimation. (iii) Exploiting both SCoBeP output and the NLM technique in calculating weights to facilitate tradeoff between computational complexity and performance.

A. Classic Super Resolution

Super resolution reconstruction attempts to estimate one high quality result $\mathcal{X}$ out of several lower resolution and potentially noisy images $\{\mathcal{Y}_t\}_{t=1}^{T}$. A popular way to model LR images $\{\mathcal{Y}_t\}_{t=1}^{T}$ from a pseudo HR image $\mathcal{X}$ is through a sequence of operations including geometrical wrapping F_t, linear space-invariant blurring H, spatial decimation D_t, and zero-mean white Gaussian noise ε_t. The model can be summarized with the following equation:

$$\mathcal{Y}_t = DHF_t\mathcal{X} + \varepsilon_t, \quad t = 1, 2, 3, ..., T, \tag{1}$$

where T is the number of available LR frames. Note that we assume H and D are identical for all frames in the sequence.

The recovery of $\mathcal{X}$ from $\{\mathcal{Y}_t\}_{t=1}^{T}$ using the above mentioned model requires us to solve an inverse problem. The maximum *a posteriori* probability estimate of $\mathcal{X}$ can be obtained by minimizing the following objective function with respect to $\mathcal{X}$:

$$\epsilon^2_{MAP}(\mathcal{X}) = \frac{1}{2}\sum_{t=1}^{T} \|DHF_t\mathcal{X} - \mathcal{Y}_t\|_2^2 + \lambda \cdot TV(\mathcal{X}), \tag{2}$$

where the first summation term ensures that the projections of the estimate $\mathcal{X}$ looks similar to the LR images and the second term, $\lambda \cdot TV(\mathcal{X})$, acts as a prior and helps to remove artifacts from the final solution and improves the rate of convergence [29].

Since H and F_t are space-invariant operators in (2), they can be considered as block circulant matrices (assuming a cyclic boundary treatment) that they commute [4], [30]. This allows one to solve (2) in the following two steps [3], [4], [6], [30]. First, minimize the following penalty function with respect to $\mathcal{Z}$:

$$\epsilon^2_{ML}(\mathcal{Z}) = \frac{1}{2}\sum_{t=1}^{T} \|DF_t\mathcal{Z} - \mathcal{Y}_t\|_2^2, \tag{3}$$

where $\mathcal{Z}$ can be interpreted as a blurred version of the HR frame $\mathcal{X}$ and thus should be approximately equal to $H\mathcal{X}$. This step estimates the blurry high-resolution image $\mathcal{Z}$ from the collection of the low resolution images $\mathcal{Y}$. For a more general case with multiple input patches, we will modify ϵ^2_{ML} in (3) to (15) as shown in Section II-C.

Then, impose the constraint of the closeness of $\mathcal{Z}$ and $H\mathcal{X}$ and incorporate back the regularization term to obtain the following objective function:

$$\epsilon^2_{MAP}(\mathcal{X}) = \|H\mathcal{X} - \mathcal{Z}\|_2^2 + \lambda \cdot TV(\mathcal{X}), \tag{4}$$

where $\mathcal{X}$ can be obtained through minimizing (16) in Section III-C. Since H is usually singular, this stage is an underdetermined problem and needs regularization (see [31], [32] for more detail).

In summary, one can break the minimization problem in (2) in two steps:

1) compute a blurred version of HR $\mathcal{Z}$ by minimizing (3).
2) estimate the deblurred frame $\mathcal{X}$ from the found blurred HR $\mathcal{Z}$ in step 1.

As the second step only involves the classic deblurring problem, many potential techniques can be applied here. In our proposed approaches, we adopt the *Adaptive Kernel Total Variation* (AKTV) regularized locally-adaptive kernel regression in a variational approach developed by Takeda *et al.* [33], which can simultaneously interpolate and deblur in one integrated step. However, one can generally incorporate any deblurring techniques into the proposed method.

B. Background of NLM Filter

The whole entity of a self-similar object is exactly like or similar to a part of itself. As a consequence, parts of it can show the same statistical properties at many scales. Based

on this presumption, non-local self-similarity techniques have been widely used in areas such as image denoising [17], [34], texture synthesis [24], and super resolution [3], [6], [10]. For example, the NLM filter, which is based on the assumption that image content is likely to repeat itself within its neighborhood, is applied successfully to image denoising. Its key idea is that one can denoise a pixel $[i,j]$ by performing weighted average around its neighborhood [17]. More precisely, denote $\mathcal{Y}[i,j]$ as the intensity of pixel $[i,j]$, then the intensity of denoised pixel $[q,l]$, $\mathcal{X}[q,l]$, can be written as

$$\mathcal{X}[q,l] = \frac{\sum_{(i,j)\in\mathcal{N}(q,l)} \mathcal{W}[i,j,q,l]\mathcal{Y}[i,j]}{\sum_{(i,j)\in\mathcal{N}(q,l)} \mathcal{W}[i,j,q,l]}, \tag{5}$$

where $\mathcal{N}(q,l)$ denotes a neighborhood around pixel $[q,l]$, and $\mathcal{W}[i,j,q,l]$ is a weight that is decreased with the distances between pixels $[i,j]$ and $[q,l]$, and increased with the similarity of the patches centering at the two pixels. The formula in (5) describes the NLM filter where denoising each pixel is done by averaging all pixels in its neighborhood. However, this averaging is not performed blindly and instead each pixel in the relevant neighborhood is assigned a weight which corresponds to the probability that the pixel $\mathcal{Y}[i,j]$ and the pixel $\mathcal{X}[q,l]$, prior to the additive noise degradation, had the same value.

NLM filter computes the weight based on both radiometric proximity and geometric proximity between the pixels. The radiometric part is estimated by computing the Euclidean distance between two image patches centered around these two included pixels. Let us consider $R_{q,l}$ as the matrix that extracts a patch with fixed and predefined size of $g \times g$ pixels at its position $[q,l]$ in the image. Hence, $R_{q,l}\mathcal{Y}$ is equivalent to the $g\times g$ matrix representing the extracted patch of $\mathcal{Y}$ at position of $[q,l]$. As NLM estimation is a zero-order regression, only the zero-order basis is used for estimation. Therefore, the NLM weights look like

$$\begin{aligned}\mathcal{W}[i,j,q,l] =& exp\left\{-\frac{\|R_{q,l}\mathcal{Y} - R_{i,j}\mathcal{Y}\|_2^2}{2\sigma^2}\right\} \\ &\times f(\sqrt{(q-i)^2+(l-j)^2}),\end{aligned} \tag{6}$$

where σ manages the effects of radiometric differences between two patches and when the intensities of the two patches are far away, the weight becomes very small and thus can be ignored. Whereas the function f is in charge of the geometric distance, and it may have many forms such as a Gaussian, a box function, or a constant [3], [6]. Since there are various other ways to choose the weights in (5), in this paper we will restrict our choice to SCoBeP [16] and NLM as described in Sections III-B and III-C.

C. NLM for Super Resolution

Since self-similarities exist in most natural images, one can also use the NLM algorithm to take advantage the non-local similarity property of natural images in the superresolution problem.

In essence, one may extract a target patch information from multiple patches instead of one patch per each LR frame. This allows us to modify ϵ^2_{ML} in (3) instead to [3], [6]

$$\begin{aligned}\epsilon^2_{ML}(\mathcal{Z}) &= \frac{1}{2}\sum_{t=1}^{T}\sum_{[q,l]\in\mathcal{I}}\sum_{[ri,rj]\in\mathcal{N}(q,l)} \mathcal{W}[i,j,q,l,t] \\ &\quad\times \left\|DR^H_{q,l}\mathcal{Z} - R^L_{i,j}\mathcal{Y}_t\right\|_2^2,\end{aligned} \tag{7}$$

where $\mathcal{I}$ is the set of pixel coordinates of the entire frame $\mathcal{X}$, $\mathcal{N}(q,l)$ is a neighborhood of the pixel $[q,l]$, and $\mathcal{W}[i,j,q,l,t]$ can be interpreted as a weight that the pixel $[q,l]$ in the reference frame should be mapped to the pixel $[i,j]$ in the t^{th} LR frame $\mathcal{Y}_t$. $R^H_{q,l}$ and $R^L_{i,j}$ are defined as the HR and LR patch extraction operators respectively, where the size of the extracted patches are related to the resolution ratio r as follows. Let the size of patches extracted by $R^L_{i,j}$ and $R^H_{q,l}$ be $g \times g$ and $k \times k$, respectively. We have $k = r(g-1)+1$. Note that k is not set precisely as rg to avoid the need of extrapolation. The detail in computing $\mathcal{W}$ will be deferred to Section III-B and III-C.

As for the first step, one can show that the optimum $\mathcal{Z}$ can be computed as [6]

$$\mathcal{Z}[q,l] = \frac{\sum_{t=1}^{T}\sum_{[ri,rj]\in\mathcal{N}(q,l)} \mathcal{W}[i,j,q,l,t]\mathcal{Y}_t[i,j]}{\sum_{t=1}^{T}\sum_{[ri,rj]\in\mathcal{N}(q,l)} \mathcal{W}[i,j,q,l,t]}. \tag{8}$$

III. Proposed Method

The key to apply NLM to super resolution efficiently depends on how we can identify the appropriate neighboring set ($\mathcal{N}(q,l)$) for each pixel and also how we can choose the appropriate weighting function. In particular, the neighborhood $\mathcal{N}(q,l)$ has significant effect on the performance of the NLM filter. The neighborhood should be sufficiently large to take advantage "non-local" benefit of the algorithm. However, this also significantly increases the complexity of the algorithm.

Ideally, we would like the neighborhood set $\mathcal{N}(q,l)$ to cover the entire frame. That is, to allow each pixel to take into account information from any pixel of every LR frame and let the weight variable $\mathcal{W}[i,j,q,l,t]$ to take care of the significance of the contribution. This, of course, will lead to unrealistic computational load if we blindly look into every pixel of every LR frame. What we need is an intelligent preprocessing step to identify pixels that are likely to provide useful information to the target pixel no matter where the formers locate. The described problem above is closely related to image registration, and we want to look for multiple matches from each reference frame (i.e., an LR frame in this case).

While many registration methods can be used, we chose to use SCoBeP [16] for the aforementioned purpose as SCoBeP naturally identifies multiple matched pixels and returns the corresponding match scores as needed in this application. In summary, for each pixel in the initial estimate of the HR frame, we use SCoBeP to select from each LR frame a set of n candidate pixels which are likely to be most similar to the target pixel. The similarity between a target pixel and a candidate pixel, which will be characterized by a weight, will

be used by the SCoBeP method as to be described in Section III-B or the NLM filter as to be described in Section III-C.

For the rest of this section, we will review the SCoBeP registration technique in the context of superresolution and provide the implementation details of our proposed super resolution methods, SCoBeP-SR and SCoBeP-NLM, which are based on sparse coding, belief propagation and NLM. We divide the super resolution process into two steps as shown in Sections III-A and III-B or III-C.

A. Use SCoBeP [16] to compute the locations and prior probabilities of candidate pixels

The proposed method described here is inspired by our recent work, SCoBeP [16]. First, we extract the features from all interpolated LR frames $\{y_t\}_{t=1}^{T}$ and the reference frame $\mathcal{X}$. To extract the features, we consider a patch of size $(2h+1)^2$ containing neighboring pixels around each pixel on the reference and LR frames, where h is a positive integer. For each pixel $[p,q]$ in the reference frame $\mathcal{X}$, we vectorized the patch centered around the pixel $[p,q]$ to a feature vector $X_{ql} \in \mathbb{R}^{S\times 1}$, where $S=(2h+1)^2$.

In this paper, we focus ourselves on only using block features even though the proposed approach can generally be applied to other features (such as SIFT-features). Thus, each feature considered here is essentially a vectorized block centered around a pixel in a frame.

Second, to match the extracted features of the reference frame to the corresponding extracted features of the t^{th} interpolated LR frame y_t, we create a dictionary which contains all feature vectors of y_t. More precisely, a dictionary $D_t \in \mathbb{R}^{S\times MN}$ (M and N are the height and width of the interpolated LR frame y_t minus h from each side) is constructed with all possible vector $Y_{qlt} \in \mathbb{R}^{S\times 1}$ as D_t's column vectors, where Y_{qlt} is created in the same manner as X_{ql} but from the t^{th} interpolated LR frame y_t instead. Thus, we can write D_t as

$$D_t = [Y_{1,1,t}Y_{1,2,t}\cdots Y_{1,N,t}Y_{2,1,t}\cdots Y_{M,N,t}]. \quad (9)$$

We then normalize dictionary D_t to guarantee the norm of each feature vector to be 1.

Third, to identify the candidate Y_{qlt} that looks most similar to the input X_{ql} in the reference frame, we apply sparse coding to each extracted features of the reference frame. Sparse coding will reconstruct a reference patch at pixel $[q,l]$ as a linear combination of LR patches. Denote α_{qlt} as the sparse vector where each element corresponds to a coefficient in this combination. Note that α_{qlt} should be sparse, i.e., it should be 0 for most coefficients.

Mathematically, we try to solve the following sparse coding problem of finding the most sparse coefficient vector α_{qlt} such that

$$X_{ql} = D_t\alpha_{qlt}. \quad (10)$$

The sparse vector α_{qlt} is the representation of X_{ql}, which has few number of non-zeros coefficients. Thus, α_{qlt} describes how to construct X_{ql} as a linear combination of a few columns (also referred to as atoms) in D_t. The locations of the nonzero coefficients in α_{qlt} specifically point out which Y_{ql} in the dictionary D_t is used to build X_{ql} and the values of the non-zero coefficients in α_{qlt} show what "portions" thereof are used for its construction. As shown in Fig. 4, one expected that most of the coefficients in α_{qlt} obtained by sparse coding are zero, and the bases of those non-zero coefficients correspond to the highlighted gray columns in D_t. Thus, X_{ql} can be written as a sparse linear combination of those highlighted gray columns.

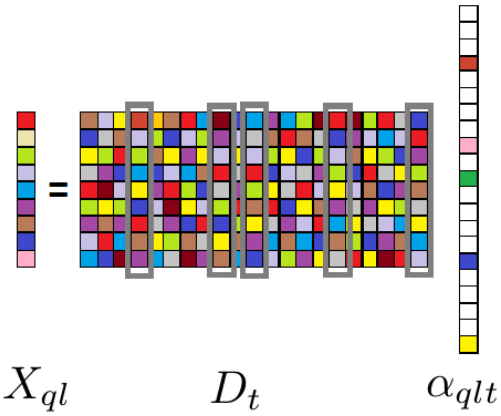


Fig. 4: Sparse representation of a feature vector X_{ql} with a dictionary D_t: α_{qlt} as a sparse vector constructs the feature vector X_{ql} using a few columns (highlighted in gray) of dictionary D_t.

To solve (10), besides linear programming, many other suboptimal techniques have been proposed including orthogonal matching pursuit [35], Subspace Pursuit (SP) [36] and gradient projection [37]. In this work, we employed Subspace Pursuit (SP) [36]. After finding the sparse representation vector α_{qlt}, to select the n candidate pixels, we simply pick those corresponding to n largest absolute value of coefficients in α_{qlt}. We denote cp_{qlt} as an $n\times 2$ matrix storing the locations of these candidate pixels and ρ_{qlt} as the length-n vector storing the corresponding values of α_{qlt}. We will take the normalized $|\rho_{qlt}|$ as a prior probability of matching the reference patch at $[q,l]$ to a patch of the interpolated LR frame y_t taking only local characteristics into account but ignoring geometric characteristics of the matches.

Finally, to incorporate geometric characteristics, we model the problem by a factor graph and apply belief propagation to update probabilities ρ_{qlt} (for more details, see [16]).

We assume the operations such as warping and blurring in the maximum *a posteriori* probability equation (2) are known. However, this is not true in practice. In particular, while the blurring operation can be approximated to be more or less constant over the entire scene, the warping operation could vary from pixels to pixels. One way to handle the unknown warping problem is to adopt the Bayesian formulation and integrate all possible warping operations, this is of course too expensive to compute. Fortunately, the SCoBeP step has already provided us some candidate match locations. So (8) becomes

$$\mathcal{Z}[q,l] = \frac{\sum_{t=1}^{T}\sum_{b=1}^{n}\sum_{[i,j]|[ri,rj]\in\mathcal{N}(cp_{qlt}^{(b)})}\mathcal{W}[i,j,q,l,b,t]\mathcal{Y}_t[i,j]}{\sum_{t=1}^{T}\sum_{b=1}^{n}\sum_{[i,j]|[ri,rj]\in\mathcal{N}(cp_{qlt}^{(b)})}\mathcal{W}[i,j,q,l,b,t]}, \quad (11)$$

Algorithm 1 Super Resolution framework using SCoBeP - estimate version of HR frame $\mathcal{X}$

Inputs : LR and noisy frames $\{\mathcal{Y}_t\}_{t=1}^{T}$, resolution ratio r, weight patch R, frame number t_c and the maximum number of iterations

Initialize :

- Set $\mathcal{Z}$ as the bicubic interpolated frame of $\mathcal{Y}_{t_c}$

Iterate : while the maximum number of iterations is not reached

Use SCoBeP to find candidate pixels : For each increased resolution frame y_t

- Extract dense feature
- Construct dictionary D_t
- Find the initial estimates of candidate pixel probabilities ρ_{qlt} and candidate pixel locations cp_{qlt}
- Apply belief propagation to refine ρ_{qlt} and cp_{qlt}

Find the blurred HR frame : For each pixel location $[q,l]$ on the HR frame $\mathcal{Z}$ and for $b \in \{1,2,\cdots,n\}$, for each pixel location $[i,j]$ such that $[ri,rj] \in \mathcal{N}(cp_{qlt}^{(b)})$

- Compute weights:

  1) SCoBeP-SR weights:
  $\mathcal{W}[i,j,q,l,b,t] = I(cp_{qlt}^{b} = [i,j])\rho_{qlt}^{(b)}$

  OR

  2) SCoBeP-NLM weights:
  $$\mathcal{W}[i,j,q,l,b,t] = \rho_{qlt}^{(b)} exp\left\{-\frac{\left\|R_{q,l}\mathcal{Z} - R_{cp_{qlt}^{(b)}} y_t\right\|_2^2}{2\sigma^2}\right\} \times f(\sqrt{(q-ri)^2+(l-rj)^2+\xi(t-t_c)^2})$$

- Compute $\mathcal{Z}$, the blurred version of reconstructed HR frame:
$$\mathcal{Z}[q,l] = \frac{\sum_{t=1}^{T}\sum_{b=1}^{n}\sum_{[i,j]|[ri,rj]\in\mathcal{N}(cp_{qlt}^{(b)})} \mathcal{W}[i,j,q,l,b,t]\mathcal{Y}_t[i,j]}{\sum_{t=1}^{T}\sum_{b=1}^{n}\sum_{[i,j]|[ri,rj]\in\mathcal{N}(cp_{qlt}^{(b)})} \mathcal{W}[i,j,q,l,b,t]}$$

End of Iteration

Perform deblurring : $\mathcal{X}$ =TV$deblur(\mathcal{Z})$

Output : a HR frame $\mathcal{X}$

where $\mathcal{W}[i,j,q,l,b,t]$ can be interpreted as the weight mapping from a pixel in the reference frame to b^{th} candidate pixel in the t^{th} interpolated LR frame y_t. Since there are various other ways to choose the weights in (11), in this paper we will restrict our choice to SCoBeP-SR weights and SCoBeP-NLM weights as described in the next subsections.

B. Calculate Weights for SCoBeP-SR

As SCoBeP has naturally identified pixels that are most likely to be relevant to a target pixel and also output the corresponding "weight" of the relevant pixels. Thus, we have introduced and implemented a new SCoBeP based SR algorithm, SCoBeP-SR, where "mixing" weights and candidates are extracted from the SCoBeP step only.

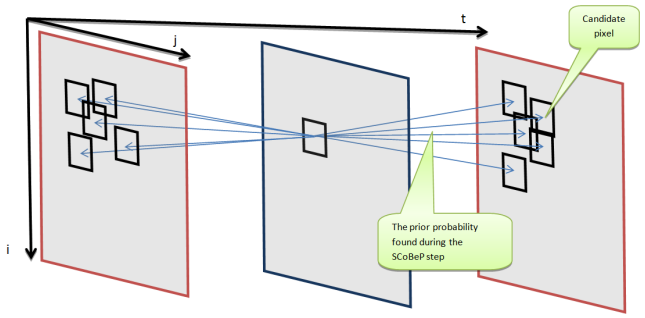


Fig. 5: Candidate pixel and weight computation in SCoBeP. For the patch in the middle frame, SCoBeP weights the found candidate pixels along the space-time.

The method for calculating $\mathcal{W}[i,j,q,l,b,t]$ for SCoBeP-SR is based on the materials that have been developed in section III-A. As some candidate locations and the corresponding belief are available from the SCoBeP output. We will simply assign the weights as zero except the candidate locations and the weights precisely as the beliefs output from SCoBeP. More precisely, we define

$$\mathcal{W}[i,j,q,l,b,t] = I(cp_{qlt}^{b} = [i,j])\rho_{qlt}^{(b)}, \tag{12}$$

where $I(cp_{qlt}^{b} = [i,j])$ is an indicator function which is equal to 1 if $cp_{qlt}^{b} = [i,j]$ and 0 otherwise. To maintain the original formulation, the neighborhood function of a patch $\mathcal{P}$ will just equal to the patch itself. That is, $\mathcal{N}(\mathcal{P}) = \mathcal{P}$.

C. Calculate Weights for SCoBeP-NLM

In this subsection, the method for estimating $\mathcal{W}[i,j,q,l,b,t]$, based on the materials that have been developed in Sections II-B and III-A, is proposed as follows:

$$\mathcal{W}[i,j,q,l,b,t] = \rho_{qlt}^{(b)} exp\left\{-\frac{\left\|R_{q,l}\mathcal{Z} - R_{cp_{qlt}^{(b)}} y_t\right\|_2^2}{2\sigma^2}\right\} \times f(\sqrt{(q-ri)^2+(l-rj)^2+\xi(t-t_c)^2}), \tag{13}$$

where t_c is the frame number of the output frame (see Fig. 5), and ξ is a scaling factor taking into account the difference in scale along the temporal and spatial dimensions. Note that we denote here $\rho_{qlt}^{(b)}$ as the b^{th} element of ρ_{qlt} (the probability of pixel $[q,l]$ in the reference frame mapping to the b^{th} candidate pixel in t^{th} interpolated LR frame), and $cp_{qlt}^{(b)}$ as the b^{th} row of cp_{qlt} (the b^{th} candidate location described by cp_{qlt}). Hence, $R_{cp_{qlt}^{(b)}}$ in (13) extracts a patch at the position $cp_{qlt}^{(b)}$ from frame y_t. To follow the notation easily, we summarized them in Table I.

Note that computing weights involves the knowledge of the unknown frame $\mathcal{Z}$. For first iteration, the weights are computed by using an estimated version of $\mathcal{Z}$, which is a scaled-up

frame generated by a conventional image interpolation algorithm such as bicubic, bilinear, or the lanczos method [38]–[40]. For the remaining iterations, the weights are computed using the estimated $\mathcal{Z}$ obtained in the previous iteration. The main procedure for our proposed methods are summarized in Algorithm 1, and also graphically depicted in Fig. 2. Note that in Algorithm 1, one can either pick SCoBeP-SR weights or SCoBeP-NLM weights for computing the weights.

As a summary of our approaches, we were able to write and minimize our cost function which has two terms:

$$\epsilon^2_{MAP}(\mathcal{X}) = \frac{1}{2}\sum_{t=1}^{T}\sum_{[q,l]\in\mathcal{I}}\sum_{b=1}^{n}\sum_{[i,j]|[ri,rj]\in\mathcal{N}(cp^{(b)}_{qlt})}\mathcal{W}[i,j,q,l,b,t] \times \left\|DR^H_{q,l}H\mathcal{X} - R^L_{i,j}\mathcal{Y}_t\right\|_2^2 + \lambda\cdot TV(\mathcal{X}). \quad (14)$$

As described in Section II-A, we followed [3], [4], [6] and decomposed (14) into two steps:

1) compute a blurred version of HR $\mathcal{Z}$ which $\mathcal{Z} = H\mathcal{X}$ by minimizing

$$\epsilon^2_{ML}(\mathcal{Z}) = \frac{1}{2}\sum_{t=1}^{T}\sum_{[q,l]\in\mathcal{I}}\sum_{b=1}^{n}\sum_{[i,j]|[ri,rj]\in\mathcal{N}(cp^{(b)}_{qlt})}\mathcal{W}[i,j,q,l,b,t] \times \left\|DR^H_{q,l}\mathcal{Z} - R^L_{i,j}\mathcal{Y}_t\right\|_2^2, \quad (15)$$

2) estimate the deblurred frame $\mathcal{X}$ from the found blurred HR $\mathcal{Z}$ in step 1:

$$\epsilon^2_{MAP}(\mathcal{X}) = \|H\mathcal{X} - \mathcal{Z}\|_2^2 + \lambda\cdot TV(\mathcal{X}), \quad (16)$$

We introduced the first step in Sections III-A, III-B, and III-C, and as the second step is the conventional deblurring problem, many works can be applied here, which we simply adopted (AKTV) regularized locally-adaptive kernel regression in a variational approach developed by Takeda *et al.* [33].

We denote cp_{qlv} as an $n\times 2$ matrix storing the locations of these candidate pixels and ρ_{qlv} as the length-n vector storing the corresponding values of α_{qlv}. Each coefficient in ρ_{qlv} serves as a prior probability of matching the reference patch at $[q,l]$ to a LR patch of $\mathcal{Y}_v$ taking only local characteristic into accounts but ignoring geometric characteristics of the matches. Finally, to incorporate geometric characteristics, we model the problem by a factor graph and apply belief propagation to update probabilities ρ_{qlv} (for more details, see [16]).

IV. Experimental Results

In this section, we consider in two separated subsections with two different sets of experiments[1]. We utilize four test sequences (Miss America, Foreman, Suzie and Stefan) in Section IV-A to compare the performance of our proposed methods with the state-of-the-art methods [6], [9], [41]. In that section we will first generate synthetic LR sequences and next apply super resolution methods to the degraded sequences. We will then compare the results to the ground truth (the original sequences). Also, in Section IV-B, we will illustrate additional examples that will assess our super resolution methods for real video sequences. Comparison will be made against the multi-image super resolution method proposed by Farsui *et al.* [4], 3-D ISKR method [9], super resolution Using TV prior method [41] and a single image up-sampling using the Lanczos algorithm [40], which were implemented using the software provided by their authors.

A. Evaluation On Synthetic Sequences

In this section, to evaluate our performance, we present some super resolution examples using existing sequences such as Miss America, Foreman, Suzie, and Stefan. The sequences in this section contain object motions only in the scene and no camera movement. All tests in this section were processed in the following manner: All 30 frames were involved in the reconstruction of each frame. The similar block size used for computing weight (R) was 13×13 and was not changed for various tests. The low patch extraction operator $R^L_{i,j}$ extracts only one pixel, therefore the $R^H_{q,l}$ extracts a patch of size 3×3 pixels. Also, the search area (the size of neighborhood $\mathcal{N}$) is 31×31 pixels. We set the parameter $\sigma = 2.2$ and the maximum number of iterations equal to 2 for all sequences.

To generate the LR frames, first, we degrade the test sequences by blurring the videos with a 3×3 uniform point spread function (PSF) and downsampling them by a resolution ratio of $3:1$ in both horizontal and vertical directions. Then the white Gaussian noise with standard deviation of $\sigma_{noise} = 2$ is added to each frame. Two of the selected LR sequences, Foreman and Suzie for frame numbers $8, 13, 23$, and frame numbers $3, 23$ are shown in Figs. 6(a), and 8(a), respectively. Then, we upscale the degraded videos using the Lanczos interpolation [40], the GNL-Means method [6], and our proposed methods. Figs. 6(b)–(e), and 8(b)–(e), respectively, show the results.

The graphs[2] in Fig. 7 show the frame by frame PSNR values of Miss America, Foreman and Suzie. Our proposed methods beats the GNL-Means method in all frames by a significant margin for all sequences. The average PSNR values for our proposed methods and the compared methods are shown in the caption of Fig. 7.

The PSNR results of 8^{th} and 13^{th} frames of the Miss America sequence are summarized in Table II, showing that the proposed methods again constantly outperform the current state-of-the-art methods. Note that the results from 3-D ISKR method is cited directly from [9]. In Fig. 9, we show the PSNR result and a clear visual comparison on the Suzie sequence. As shown in Fig. 9, although the GNL-Means method [6] acts well at regular-structured areas, it suffers from block artifacts[3] due to poor block matching. In contrast, our proposed methods performs remarkably well for both regular and detail structures and is free of these artifacts. In Fig. 10, we further show

[1]The image frames of the result sequences using SCoBeP-NLM and SCoBeP-SR are available at http://students.ou.edu/B/Nafise.Barzigar-1/software/SCoBeP-NLM.html.

[2]The PSNR results of 3D-ISKR [9] are not listed as they are not available in their original paper.

[3]Please note that we have adopted the terminology "block artifact" from [10]. The terminology is different from the artifacts typically found in low bitrate compressed image by old JPEG.

(a) (b) (c) (d) (e)

Fig. 6: Results for the 8^{th}, 13^{th} and 23^{th} frame from the "Foreman" sequence. From Left column to Right column: LR frame; GNL-Means [6]; Lanczos interpolation [40]; result of the proposed SCoBeP-NLM; result of the proposed SCoBeP-SR. Also, the PSNR values for all the frames are shown in Fig. 7(b).

the results of Foreman sequence compared with the GNL-Means method [6], 3-D ISKR method [9] and NLKR [10]. The super resolution results on Miss America sequence in frames 8 and 13 and Stefan sequence are also given in Figs. 11 and 12, respectively for visual comparison. The proposed methods outperform the other methods by notable improvement.

Moreover, we examined how our methods perform under various noise levels. We added white Gaussian noise with standard deviation σ_n (varying from 0 to 2) to the LR sequences, where the sequence with $\sigma_n = 0$ was degraded by the downsampling process only. Table III shows that both SCoBeP-NLM and SCoBeP-SR are able to produce fine details when the noise level is increasing. For a clear comparison on varying noise level, we show the results of a noise added Foreman sequence in Fig. 13, where we compared our algorithms with the state-of-the-art 3-D ISKR [9].

TABLE III: Noise Addition: PSNR for 1^{st} frame of Foreman Sequence

	Lanczos [40]	**3-D ISKR [9]**	**SCoBeP-NLM**	**SCoBeP-SR**
$\sigma_n = 0.00$	28.51	28.94	29.88	29.76
$\sigma_n = 1.20$	28.44	28.93	29.86	29.74
$\sigma_n = 1.60$	28.36	28.87	29.83	29.73
$\sigma_n = 2.00$	28.25	28.86	29.75	29.68

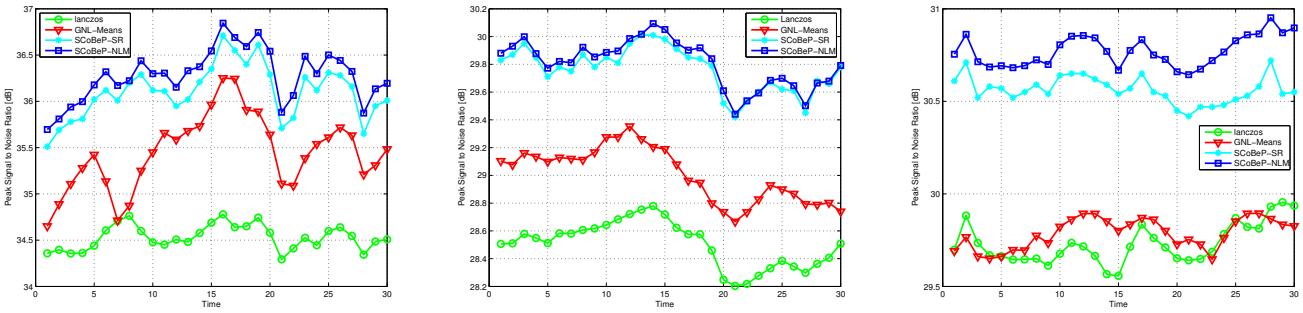
B. Evaluation on Real video Sequences

In this section, we turn to some real sequences, where we apply our proposed methods directly to the captured sequences without altering the frames. Note that there are no published methods that have tested on real sequences. As no standard sequence is available, we have captured a sequence for testing and with camera motion intentionally introduced. We choose the multi-image super resolution method proposed by Farsui *et al.* [4], 3-D ISKR method [9], super resolution Using TV prior method [41] and a single image up-sampling using the Lanczos algorithm [40] for comparison because their source codes are available publicly. Since no ground truth is available for a real sequence, we cannot evaluate the resulting HR frames with objective measure such as PSNR. However, the perceptual quality illustrate the robustness of our proposed methods on real videos.

Fig. 14 shows the superresolution results for a real Navajo Sculpture video sequence (70×80 pixels, 30 frames). One can see some "blocking" artifacts in the original sequence due

TABLE II: PSNR for 8^{th} and 13^{th} frames of Miss America Sequence

Miss America Sequence	**Nearest Neighborhood**	**Lanczos [40]**	**GNL-Means [6]**	**3-D ISKR [9]**	**SCoBeP-NLM**	**SCoBeP-SR**
8^{th} frame	32.97	34.76	34.49	35.53	36.28	36.20
13^{th} frame	32.74	34.48	35.33	35.15	36.33	36.02



(a) Miss America (b) Foreman (c) Suzie

Fig. 7: PSNR values of each super resolved frame by Lanczos [40], GNL-Means [6], and the proposed method for (a) the results of Miss America shown in Fig. 11, (b) the results of Foreman shown in Fig. 6, and (c) the results of Suzie shown in Fig. 8. The average PSNR values for all frames for the Miss America example are $34.12[dB]$ (Lanczos), $35.09[dB]$ (GNL-Means [6]), $35.73[DB]$ (SCoBeP-SR) and $35.94[dB]$ (SCoBeP-NLM) and the average PSNR values for the Foreman example are $28.51[dB]$ (Lanczos), $29.01[dB]$ (GNL-Means [6]), $29.71[DB]$ (SCoBeP-SR) and $29.80[dB]$ (SCoBeP-NLM), and also the average PSNR values for the Suzie example are $29.73[dB]$ (Lanczos), $29.79[dB]$ (GNL-Means [6]), $30.56[DB]$ (SCoBeP-SR) and $30.77[dB]$ (SCoBeP-NLM), respectively.

(a) (b) (c) (d) (e)

Fig. 8: Results for the 3^{th} and 23^{th} frame from the "Suzie" sequence. From Left column to Right column: LR frame; GNL-Means [6]; Lanczos interpolation [40]; result of the proposed SCoBeP-NLM; result of the proposed SCoBeP-SR. Also, the PSNR values for all the frames are shown in Fig. 7(c).

to its low resolution as shown in Fig. 14(a). In Fig. 14(m), we illustrate the ability of our proposed methods in removing these artifacts and resulting in a clear output. We also show the superresolution results by the Lanczos interpolation [40], Farsui *et al.* method [4], 3-D ISKR method [9], and the TV prior method [41] with three time magnification per each dimension (i.e., an output resolution of 210×240 pixels) in Fig. 14(c)–(i), respectively. As shown in Fig. 14, the Farsui *et al.* method [4] and the TV prior method [41] introduce severe block artifacts (near the *mouth* and the *eyes* in Figs. 14(e) and 14(i) respectively), and the 3-D ISKR method [9] does not preserve the line texture well and generates the ghost image as shown in Fig. 14(g). In contrast, our proposed methods do not suffer from these artifacts.

The computational complexity of SCoBeP-SR can be determined by considering the following two steps: 1) computing the locations and prior probabilities of the candidate pixels, 2) calculating weights via the NLM and SCoBeP or only SCoBeP. The complexity associated with the computing the location and weights of the candidate pixels takes 70% of the overall complexity in SCoBeP-NLM. Since we replace NLM weight with SCoBeP weight in Algorithm 1 and we found the probabilities for the candidate matches in the previous step it can significantly reduce computation complexity and storage requirement. Just to put things into perspective, note that the current implementation requires approximately 700 s per frame for the Algorithm 1 using SCoBeP-NLM weights, and 230 s for Algorithm 1 using SCoBeP-SR weights in the most demanding case like "Navajo" sequence with high-resolution frame size of 250×220, with the current pure Matlab implementation on a Pentium 3 GHz (11-GB RAM) machine. In comparison, ISKR takes approximately 5784 s per frame.

V. Conclusion

In conclusion, we have proposed two novel and efficient super resolution methods based on SCoBeP [16] and Nonlocal-Means (NLM) techniques, which finds corresponding patches

Fig. 9: Video super resolution for Suzie sequence: (frame 28 and 18 with the resolution ratio 3, PSNR in brackets). From Left column to Right column: Ground truth; LR frame; GNL-Means [6] [PSNR: 29.87 - 29.86]; Lanczos interpolation [40] [PSNR: 29.41 - 29.27]; SCoBeP-NLM [PSNR: 30.95 - 30.75]; SCoBeP-SR [PSNR: 30.55 - 30.71].

Fig. 10: Video super resolution for Foreman sequence: From Left column to Right column: Ground truth; GNL-Means [6]; 3-D ISKR [9]; NLKR [10]; result of the proposed SCoBeP-NLM.

using sparse coding and demonstrates competitive results in both the synthetic and the real sequences. Our techniques perform super resolution by first running sparse coding over an overcomplete dictionary constructed from the LR frames to gather possible match candidates. Belief propagation is then applied to eliminate bad candidates and to select optimum matches. Finally, in the SCoBeP-NLM, the NLM approach exploits similarity in patches around candidate pixels to average out the noise among similar patches. While the algorithm performs favorably comparing with other recent approaches as illustrated in the experimental results, the algorithm is quite complex and we realized that the source of most computation is originated from the NLM component. As SCoBeP has naturally identified pixels that are most likely to be relevant to a target pixel and also output the corresponding "weight" of the relevant pixels. This suggested us that NLM is probably not essential in our SCoBeP based SR algorithm. Thus, we have also implemented a SCoBeP based SR algorithm, SCoBeP-SR, where "mixing" weights and candidates are extracted from the SCoBeP step only.

We conducted experiments on both the synthetic and the real video sequences, where our approaches work well for both types of sequences demonstrating the effectiveness and robustness of our approaches. Furthermore, unlike many existing super resolution approaches targeting to LR frames that have been pre-registered manually [2] or have assumed a stationary camera [3], [10], the proposed method can handle a sequence captured with a moving camera and do not require

Fig. 11: Video super resolution for Miss America sequence; frame 8 (top) and frame 13 (bottom): From left to right: Ground truth; Lanczos interpolation [40]; GNL-Means [6]; 3-D ISKR [9]; super resolution Using TV prior [41]; SCoBeP-NLM; SCoBeP-SR.

Fig. 12: Video super resolution for Stefan sequence: From top to bottom column: LR frame; 3-D ISKR [9]; result of the proposed SCoBeP-NLM.

Fig. 13: Video super resolution for Foreman sequence with noise: we added synthetic additive white Gaussian noise (AWGN) to the input LR sequence, with the noise level $\sigma_n = 1.20$ (left) and $\sigma_n = 2.00$ (right). From top to bottom column: Noisy LR; 3-D ISKR [9]; SCoBeP-NLM; SCoBeP-SR.

preprocessing of the sequence.

As SCoBeP provides decent results in images with both significantly and slightly varying viewpoints [16], [42], hence, it will be useful to a wide range of applications such as de-interlacing, surveillance application and medical image super resolution. As for future work, we plan to extend our

Fig. 14: Multi-frame super resolution for real frames: "Navajo" sequence. (a) LR frame; (b) Lanczos interpolation; (c) Farsui *et al.* [4] method; (d) 3-D ISKR [9]; (e) super resolution Using TV prior [41]; (f) SCoBeP-NLM; (g) SCoBeP-SR.

approaches to these areas.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and the anonymous reviewers for their constructive comments and suggestions. We would also like to thank Mrs. Renee Wagenblatt and Mrs. Sepideh Darbandi for editing the manuscript.

REFERENCES

[1] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.

[2] J.C.W. Chan, J. Ma, P. Kempeneers, F. Canters, J. Vandenborre, and D. Paelinckx, "An evaluation of ecotope classification using superresolution images derived from chris/proba data," *IGARSS, July*, pp. 6–11, 2008.

[3] M. Protter and M. Elad, "Super resolution with probabilistic motion estimation," *IEEE Trans. Image Process*, vol. 18, no. 8, pp. 1899–1904, 2009.

[4] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process*, vol. 13, no. 10, pp. 1327–1344, 2004.

[5] S. Bonchev and K. Alexiev, "Improving super-resolution image reconstruction by in-plane camera rotation," in *Information Fusion (FUSION), 13th Conference on*. IEEE, 2010, pp. 1–7.

[6] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the non-local-means to super-resolution reconstruction," in *IEEE Trans. Image Process*, 2009, p. 36.

[7] D. Mitzel, T. Pock, T. Schoenemann, and D. Cremers, "Video super resolution using duality based tv-l 1 optical flow," *Pattern Recognition*, pp. 432–441, 2009.

[8] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Image and video super-resolution via spatially adaptive block-matching filtering," in *Proceedings of International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*. Citeseer, 2008.

[9] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Trans. Image Process*, vol. 18, no. 9, pp. 1958–1975, 2009.

[10] H. Zhang, J. Yang, Y. Zhang, and T. Huang, "Non-local kernel regression for image and video restoration," *Computer Vision–ECCV 2010*, pp. 566–579, 2010.

[11] R.Y. Tsai and T.S. Huang, "Multiframe image restoration and registration," *Advances in computer vision and Image Processing*, vol. 1, no. 2, pp. 317–339, 1984.

[12] SP Kim, NK Bose, and HM Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 6, pp. 1013–1027, 1990.

[13] C. Liu and D. Sun, "A bayesian approach to adaptive video super resolution," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 209–216.

[14] G.K. Chantas, N.P. Galatsanos, and N.A. Woods, "Super-resolution based on fast registration and maximum a posteriori reconstruction," *IEEE Trans. Image Process*, vol. 16, no. 7, pp. 1821–1830, 2007.

[15] X. Li, K.M. Lam, G. Qiu, L. Shen, and S. Wang, "Example-based image super-resolution with class-specific predictors," *Journal of Visual Communication and Image Representation*, vol. 20, no. 5, pp. 312–322, 2009.

[16] N. Barzigar, A. Roozgard, S. Cheng, and P. Verma, "Scobep: Dense image registration using sparse coding and belief propagation," *Journal of Visual Communication and Image Representation*, 2012.

[17] A. Buades, B. Coll, J.M. Morel, et al., "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2006.

[18] Chih-Yuan Yang, Jia-Bin Huang, and Ming-Hsuan Yang, "Exploiting self-similarities for single frame super-resolution," in *Computer Vision–ACCV 2010*, pp. 497–510. Springer, 2011.

[19] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[20] F.R. Kschischang, B.J. Frey, and H.A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on information theory*, vol. 47, no. 2, pp. 498–519, 2001.

[21] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process*, vol. 19, no. 11, pp. 2861–2873, 2010.

[22] J. Wang, S. Zhu, and Y. Gong, "Resolution enhancement based on learning the sparse association of image patches," *Pattern Recognition Letters*, vol. 31, no. 1, pp. 1–10, 2010.
[23] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process*, vol. 16, no. 8, pp. 2080–2095, 2007.
[24] A.A. Efros and T.K. Leung, "Texture synthesis by non-parametric sampling," in *iccv*. Published by the IEEE Computer Society, 1999, p. 1033.
[25] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2272–2279.
[26] Ce Liu and William T Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in *Computer Vision–ECCV 2010*, pp. 706–719. Springer, 2010.
[27] Quoc Bao Do, Azeddine Beghdadi, and Marie Luong, "Combination of closest space and closest structure to ameliorate non-local means method," in *Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2011 IEEE Symposium on*. IEEE, 2011, pp. 134–141.
[28] S. Villena, M. Vega, R. Molina, and A.K. Katsaggelos, "Bayesian super-resolution image reconstruction using an l1 prior," in *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*. IEEE, 2009, pp. 152–157.
[29] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.
[30] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004.
[31] Ruimin Pan and Stanley J Reeves, "Efficient huber-markov edge-preserving image restoration," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3728–3735, 2006.
[32] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 460–489, 2005.
[33] H. Takeda, S. Farsiu, and P. Milanfar, "Deblurring using regularized locally adaptive kernel regression," *Image Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 550–563, 2008.
[34] A. Buades, B. Coll, and J.M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 60–65.
[35] A. Yang, A. Ganesh, Z. Zhou, S. Sastry, and Y. Ma, "Fast l1-minimization algorithms and an application in robust face recognition: a review," in *Proceedings of the International Conference on Image Processing*, 2010.
[36] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, 2009.
[37] R. Maleh, AC Gilbert, and MJ Strauss, "Sparse gradient image reconstruction done faster," in *IEEE International Conference on Image Processing, ICIP*, 2007, vol. 2, pp. 77–80.
[38] W. Burger and M.J. Burge, *Principles of digital image processing: core algorithms*, Springer-Verlag New York Inc, 2009.
[39] R. Keys, "Cubic convolution interpolation for digital image processing," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 6, pp. 1153–1160, 1981.
[40] G. Wolberg, Institute of Electrical, and Electronics Engineers. Computer Society, *Digital image warping*, vol. 3, IEEE computer society press California, 1990.
[41] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Variational bayesian super resolution," *Image Processing, IEEE Transactions on*, vol. 20, no. 4, pp. 984–999, 2011.
[42] A. Roozgard, N. Barzigar, S. Cheng, and P. Verma, "Medical Image registration using sparse coding and belief propagation," in *The 34th Annual IEEE International Conference of the Engineering in Medicine and Biology Society, San Diego.*, 2012.

Nafise Barzigar received the B.S. degree in software engineering from the Iran University of Science and Technology, Tehran, Iran, and the M.S. degree in electrical and computer engineering from the University of Oklahoma, Tulsa (OU), in 2006 and 2012, respectively. She is currently pursuing the Ph.D. degree in electrical and computer engineering at OU. Her research interests are in image and video processing (registration, denoising, interpolation, motion estimation, and super-resolution) and inverse problems.

Aminmohammad Roozgard received the B.S. degree in software engineering from the Iran University of Science and Technology, Tehran, Iran, and the M.S. degree in computer engineering from the Sharif University of Technology, Tehran, Iran, in 2006 and 2009, respectively. He is currently pursuing the PhD. degree in electrical and computer engineering from the University of Oklahoma, Tulsa (OU). Aminmohammad's research interests are in the area of image and video processing (registration, denoising, interpolation, motion estimation, and super-resolution), and genome privacy protection.

Pramode Verma is Director of the Telecommunications Engineering Program in the School of Electrical and Computer Engineering of the University of Oklahoma-Tulsa. He also holds the Williams Chair in Telecommunications Networking. Prior to joining the University of Oklahoma in 1999, Dr. Verma held a variety of professional and leadership positions in the telecommunications industry at AT&T Bell Laboratories and Lucent technologies. His last position with Lucent Technologies was as Managing Director, Business Development-Global Service Providers Business. Dr. Verma obtained his doctorate in Electrical Engineering from Concordia University in Montreal, Canada in 1970 and an MBA from the Wharton School of the University of Pennsylvania in 1984. He is the author/co-author of over 100 publications and several books in telecommunications, computer communications, and related fields. He is a past president of the International Council for Computer Communication, a Washington D.C.-based global organization; a senior member of the Institute of Electrical and Electronics Engineers, New York, and is registered as a Professional Engineer, Province of Ontario, Canada.

Samuel Cheng received the B.S. degree in Electrical and Electronic Engineering from the University of Hong Kong, and the M.Phil. degree in Physics and the M.S. degree in Electrical Engineering from Hong Kong University of Science and Technology and the University of Hawaii, Honolulu, respectively. He received the Ph.D. degree in Electrical Engineering from Texas A&M University in 2004. He worked in Microsoft Asia, China, and Panasonic Technologies Company, New Jersey, during the summers of 2000 and 2001. In 2004, he joined Advanced Digital Imaging Research, a research company based near Houston, Texas, as a Research Engineer. Since 2006, he joined the School of Electrical and Computer Engineering at the University of Oklahoma and is currently an associate professor. His research interests include information theory, image/signal processing, and pattern recognition.