SUPER RESOLUTION METHOD ADAPTED TO SPATIAL CONTRAST

Tien Ho-Phuoc, Antoine Dupret, Laurent Alacoque

CEA/LETI-MINATEC, Grenoble, France

ABSTRACT

In this paper we present an effective method for super resolution (SR). The proposed method is inspired by the Non-Local Means (NLM) algorithm but shows a lot of improvements over it. Firstly, we show the sensitivity to the standard deviation parameter of the NLM algorithm in detail regions that has low contrast. This analysis leads to an improvement of the NLM algorithm by selecting the best neighbors for each pixel in the SR image to be reconstructed. Secondly, this novel selection of neighbors is combined with a segmentation step in order to build a SR framework that can be spatially adapted to contrast. The proposed method allows reconstruction of SR images that conserves low-contrast detail while ensuring noise canceling.

Index Terms — super resolution, non-local mean, spatial contrast

1. INTRODUCTION

For many wavelengths, image sensors feature large pixels that limit their resolution. This is especially the case when one considers infrared or terahertz imaging. In order to get higher resolution images, super resolution (SR) technique may be preferred to interpolation. Although the SR problem was first carried out in the frequency domain [1], most of SR algorithms have been implemented in the spatial domain, which provides more flexibility concerning motion estimation or regularization [2-6]. The SR problem is often formulated as a degradation model that presents different steps to generate low resolution (LR) images from an original SR one, including motion (warp matrices), blurring effect, down-sampling, and noise [7].

Generally, motion estimation plays the most important role in super resolution: this step computes the shift, represented by motion vectors, between sequential LR images to be registered in order to form the SR image. Yet inaccurate motion vectors can much degrade the super-resolved image, which can have a quality lower than the LR images. Originally used for denoising [8], the Non-Local Means (NLM) algorithm has recently been considered as a candidate for super resolution without motion estimation.

Since the NLM method reconstructs a SR pixel value as a weighted sum of LR neighboring pixel values, it works as a low-pass filter. In the NLM SR algorithm, it is highly important to choose weights providing a good compromise between two contradictory objectives: denoising and detail conservation.

In the present paper, we first analyze the influence of the standard deviation of the Gaussian kernel on the weights of the neighboring pixels in high contrast zones and low contrast zones. Then we propose a modification of the NLM SR algorithm in order to better satisfy at the same time the two contradictory objectives: denoising and detail conservation. We introduce a new parameter to control the number of neighboring pixels and combine it with the Gaussian kernel to match the two objectives. This number of neighbors is chosen differently according to the uniformity of a region. Furthermore, a compact segmentation method is also proposed to detect uniform regions from non-uniform ones.

2. NON-LOCAL MEANS ALGORITHM REVISITED

Protter’s work [9-11] is the first one applying the NLM principle to the SR problem. The NLM SR problem is then considered under the view of an optimization problem. Thus, the pixel value \( x_i \) (of pixel \( i \)) in the SR image is reconstructed as follows:

\[
x_i = \frac{\sum_{j \in \Omega(i)} w_{ij} y_j}{\sum_{j \in \Omega(i)} w_{ij}}
\]

(1)

The indices \( i, j, (x, y) \) represent the positions (pixel values, respectively) of the corresponding pixels. \( y_j \) is the pixel value of pixel \( j \) in the LR neighborhood of pixel \( i \), \( \Omega(i) \) set of LR neighbors of pixel \( i \). Thus:

\[
\# \{ \Omega(i) \} = K(2R_S + 1)^2
\]

(2)

with \( K \) the number of LR images and \( R_S \) the radius of the search area. \( w_{ij} \) represents the similarity between pixel \( i \) in the SR image and pixel \( j \) in a LR image. These weights are computed by a Gaussian kernel as a function of distance \( d_{ij} \):

\[
w_{ij} = \exp \left( -\frac{d_{ij}^2}{2\sigma_w^2} \right)
\]

(3)

where \( d_{ij} \) is defined as the distance – or dissimilarity – between two patches around the two pixels \( i \) and \( j \). To compare two patches of different resolutions (the patch around \( i \) is SR while the patch around \( j \) is LR), a down-
sampling step is applied to the patch around \( i \) to bring it to the same resolution as the patch around \( j \) (up-sampling is also possible; however a down-sampling is preferred as it reduces computational complexity without significant impact on the distance computation). Although the distance can be computed by several methods, we used the normalized Euclidean distance because of its simplicity and efficiency.

The function determining the weights used for the weighted sum (Eq.1) plays an important role in the NLM super resolution method. Usually, a Gaussian kernel is selected for weights computation as in Eq. 3. The weighted sum constitutes a spatially variant low-pass filter – the parameter \( \sigma_w \) (Eq. 3) determines its cut-off frequency. Thus, a larger \( \sigma_w \) corresponds to a lower cut-off frequency. Generally, a large \( \sigma_w \) leads to better denoising; at the risk of over-smoothing.

Generally, a large \( \sigma_w \) leads to better denoising; at the risk of over-smoothing.

For NLM super resolution of noiseless or low-noise images, a small enough \( \sigma_w \) can lead to a small number of neighboring pixels significantly contributing to the weighted sum in both low-contrast and high contrast zones. Thus, it ensures detail conservation as well as denoising in the SR image. However, in the presence of higher noise, we must increase \( \sigma_w \) for the denoising purpose. In order to evaluate the number of significantly contributing neighbors for a SR pixel, let us designate \( N_{90} \) the number of the most similar neighboring pixels that contain of 90% of the sum of all weights. A high value of \( N_{90} \) means that the non-zero weights are distributed over a large number of neighbors. Figure 1.d shows \( N_{90} \) as a function of std. As predicted, when std increases, \( N_{90} \) also increases but more rapidly in the low contrast case than that in the high contrast case.

3. PROPOSED METHOD

From the above analysis we propose improvements to the NLM SR algorithm. Firstly, based on uniformity an image is segmented into two categories: uniform regions and non-uniform regions (containing details). Secondly, according to the uniformity of each region different numbers of neighbors can be determined in order to match the two contradictory objectives – denoising and detail conservation. Our complete method is described below.

3.1. Segmentation

This step computes the uniformity map – consisting of uniform and non-uniform regions – of the LR image to be super-resolved. The segmentation of uniform regions from non-uniform ones can be carried out by any existing segmentation algorithms in the literature. Here we adopted a simple algorithm based on the Harris detector; its implementation is based on image blocks in order to reduce computational complexity. We first apply a Gaussian low-pass filter (of 5×5 pixels) to the image and then compute the horizontal (\( I_H \)) and vertical (\( I_V \)) derivatives of its result. For each block \( k \) (of 8×8 pixels) of the LR image in question, the covariance matrix based on its derivatives is generated:

\[
A_k = \begin{bmatrix}
\langle I_H^2 \rangle_k & \langle I_H I_V \rangle_k \\
\langle I_H I_V \rangle_k & \langle I_V^2 \rangle_k
\end{bmatrix}
\]

(4)

where \( \langle \rangle_k \) computes the mean value in block \( k \).

We use the absolute Harris value of this block to determine whether it is uniform:

\[
h_k = |\det(A_k) - \alpha [trace(A_k)]^2|
\]

(5)

The main idea is that in uniform regions, both two eigenvalues of the matrix \( A_k \) are very small or negligible. Therefore, the Harris value is very close to zero. By contrast, if the block comprises lines or corners, which
correspond to one or two non-negligible eigenvalues respectively, the Harris value will be significantly different from zero. Thus, a block is considered uniform if its absolute Harris value is smaller than a threshold. The rest regions are considered non-uniform:

$$U(p_k) = \begin{cases} 0 & \text{if } h_k \geq \gamma \max_k \{h_k\} \\ 1 & \text{otherwise} \end{cases}$$

(6)

where \(p_k\) represents the positions of all pixels in block \(k\). \(U(p)=1\) indicates that a pixel \(p\) belongs to an uniform region and reciprocally for \(U(p)=0\). The parameters \(\alpha\) and \(\gamma\) are empirically optimized: \(\alpha=0.06\) and \(\gamma=0.0001\). From our experiment, we observed that only small uniform region is extracted in tested natural images using this threshold value \((\gamma)\). For texture images, almost no uniform region is extracted.

The binary image \(U\) is then interpolated by using “nearest values” to have the same size of the SR image.

3.2. Selection of limited number of neighbors

We proposed to use a discontinuous weighting function as follows by introducing a parameter controlling the number of neighbors:

$$w_{ij} = \begin{cases} \exp \left( - \frac{d_{ij}^2}{2\sigma_w^2} \right) & \text{if } S(d_{ij}) \leq N_{Th} \\ 0 & \text{otherwise} \end{cases}$$

(7)

with \(S(d_{ij}) = \# \{ l \in \Omega(i) \mid d_{ij} > d_{il} \} + 1\).

Hence, \(S(d_{ij})\) can be considered as the ascending order of \(d_{ij}\). \(N_{Th}\) is the limited number of neighbors accounted for the weighted sum. The proposed function preserves the differences of neighboring pixels in contributing to the SR pixel value. On the other hand, a limited number of neighbors take into account non-linearity to reduce the effect of the long-tailed distribution of the Gaussian kernel, which causes the over-smoothing artifact.

It is important to note that in high contrast regions, distances are widely distributed and there are only a limited number of neighbors having small distances that can significantly contribute to the weighted sum. Hence, using a limited number of neighbors \(N_{Th}\) has almost no impact on the SR image. If we choose \(N_{Th}\) equal to the total number of neighbors, the weighting function in Eq.\(7\) becomes the one in the NLM algorithm.

Hence, in uniform regions, we use all neighbors to reconstruct a SR pixel value (like the NLM algorithm) while in non-uniform regions, a small number \(N_{Th}\) is used. A question is how to choose \(N_{Th}\)? In presence of noise \(N_{Th}\) must be chosen in such a way that it can satisfy the compromise between noise cancelling and conservation of detail. We pose \(N_{Th} = n_{Th} \times K\) where \(K\) is the number of LR images. From our experiments, \(n_{Th}\) should be less than 10; and \(N_{Th}\) is less than 10% of the total number of neighboring pixels.

It is worth noting that the introducing of an extra parameter \(N_{Th}\) can make the SR image less sensitive to the standard deviation. Hence, in this case it is easier to choose this latter parameter for a good denoising.

3.3. Combination of the two regions

In order to create the final SR image, we combine super resolution effectuated in the two types of region. In other words, in uniform regions a SR pixel \(x_i\) is reconstructed using the weights \(w_{ij}\) computed according to Eq.\(3\). In non-uniform regions, \(w_{ij}\) obtained from Eq.\(7\) are used to compute \(x_i\). For a SR pixel in the transition zones that are around the frontiers between these two types of region, we effectuate a linear combination of two SR pixels obtained from the two above methods at this position. The uniformity values in the transition zones are created by convoluting the image \(U\) with a Gaussian filter having size of \(7\times7\) pixels. Hence, in the transition zones, a SR pixel value \(x_i\) is reconstructed as follows:

$$x_i = (1-U(i))x_{i,1} + U(i)x_{i,2}$$

(8)

with \(x_{i,1}, x_{i,2}\) is the SR pixel value reconstructed by using \(w_{ij}\) according to Eq.\(7\) and Eq.\(3\), respectively.

4. RESULTS

Now we apply the proposed method for the super resolution of real sequences. It is important to note that this method relates to the fusion step in the SR framework often used in the literature [2-3, 9-10]. Indeed, after the fusion step we obtain a blurred SR image that needs to be deblurred. Yet, the deblurring step can sharpen but also alter or even deteriorate a fused SR image. In the present paper, a simple method is used, i.e. Matlab blind deconvolution with a few iterations to help preserving the fused image intrinsic quality and, hence, to allow evaluating the impact of the proposed method on the fusion step.

The chosen test sequences are the popular Foreman and Suzie sequences. We first tested our method with the Foreman sequence. The process to generate a LR image sequence is as follows. Each frame of the original sequence is blurred by a PSF (Point Spread Function) that is an average function of \(3\times3\) pixels. A resulting frame is then down-sampled by factor 3 and is then added to Gaussian noise of std 5 in order to generate the LR image sequence (\(K=30\) frames). All 30 frames of the LR sequence are used to reconstruct a SR image: we obtained finally a sequence of SR images.

The parameters used for the proposed method are following: standard deviation of the Gaussian kernel is 5.5, patch size \(9\times9\) pixels, search area \(13\times13\) pixels, \(N_{Th}=5\times K\). The NLM method used the same parameters except that all neighbors were used for the reconstruction of a SR pixel value. As stated above, the deblurring step was carried out...
The difference between NLM and the proposed method is not very important. Although the improvement in small details perception is clearly visible in presented zoomed images, mathematical criteria fail to reflect this improvement correctly. That can be explained by the fact that pixel-to-pixel difference in low contrast regions is small, which leads to similar PSNR in NLM and our method, but this difference cannot reflect the difference in human quality perception.

To our understanding, in the NLM method, the denoising objective is achieved at the expense of detail conservation. By contrast, in the proposed method, using a limited number of neighbors, combined with segmentation, ensures the denoising objective and, at the same time, well conserves details in the SR image.

5. CONCLUSIONS

In this paper we presented a new effective NLM-based method for super resolution. The proposed method consists in limiting the number of neighbors taken into account in the non-uniform regions. The advantage of this method is that it satisfies both contradictory objectives of the super resolution problem: denoising and detail conservation. These objectives are very hard to be obtained with the NLM method. The experiment showed that the proposed method performs well with highly noisy image sequences. Besides, we also presented a compact segmentation method based on the Harris detector. Uniformity can also be generalized to multi-values in place of binary values, which in turn leads the number of neighbors to adapt gradually to uniformity.

The proposed method can also be considered as one using non-linear combination of neighbors by thresholding the number of neighbors. The non-linearity property of the fusion step may be efficient in conserving more details in the SR image. This problem is worth being further studied in the future.

Inspired from the NLM method, our proposed super resolution method allows omitting explicit motion estimation and, hence, can be applied to image sequences with general motion. Recently, studies on bilateral filtering and kernel regression have enriched the super resolution methods without explicit motion estimation [13-14]. Also not taking motion into account, the sparse representation approach has showed promising results for super resolution by using a single low-resolution input image combined with learning method [15]. Such classes of SR methods are expected to be able to solve the SR problem more generally and efficiently.

6. REFERENCES


<table>
<thead>
<tr>
<th>TABLE 1 – Quantitative assessment of the proposed algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Foreman PSNR (dB)</td>
</tr>
<tr>
<td>Foreman SSIM</td>
</tr>
<tr>
<td>Suzie PSNR (dB)</td>
</tr>
<tr>
<td>Suzie SSIM</td>
</tr>
</tbody>
</table>

While PSNR is a classic evaluation criterion that is rather mathematical, SSIM [12] constitutes a more perceptual evaluation since it is based on the structure similarity between two images.

For each criterion, a reconstructed SR image is compared to the corresponding ground truth; the criterion value is then averaged over all frames of the sequence. In both criteria, the results are improved from Lanczos to NLM and then again from NLM to the proposed method, although the difference between NLM and the proposed method is not very important.