A NONSmooth GroPh-BASED APPROACH TO LIGHT FIELD SUPER-RESOLUTION

Mattia Rossi  Mireille El Gheche  Pascal Frossard

École Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4)

ABSTRACT
In this article we propose a new super-resolution algorithm tailored for light field cameras, which suffer by design from a limited spatial resolution. To do so, we cast the light field super-resolution problem into an optimization problem, where the particular structure of the light field data is captured by a nonsmooth graph-based regularizer, and all the light field views are super-resolved jointly. In our experiments, we show that the proposed method compares favorably to the state-of-the-art light field super-resolution algorithms in terms of PSNR and visual quality. In particular, the nonsmooth graph-based regularizer leads to sharper images while preserving fine details.

Index Terms— light field, super-resolution, graph

1. INTRODUCTION
A light field camera behaves as a compact camera array, providing multiple simultaneous images of a 3D scene from slightly different points of view on a regular grid [1]. The captured data is referred to as the light field [2] and is potentially suitable for a wide range of applications, such as depth estimation, view synthesis, and 3D modeling. However, the light field views exhibit a significantly lower resolution than images from traditional cameras, and this represents a limit for many light field applications. It is therefore crucial to improve the resolution of the light field views through super-resolution techniques.

It is important to note that the light field data is characterized by a particular structure that needs to be preserved when augmenting the resolution of the single light field views. As a consequence, off-the-shelf single-frame super-resolution algorithms [3–5], which are completely blind to the light field structure, and traditional multi-frame super-resolution algorithms [6–8], whose global warping model is too general to capture the complexity of the light field structure, are not ideal candidates for light field super-resolution. Motivated by this observation, in this article we present a new super-resolution algorithm which explicitly takes the light field structure into account by modeling the inter view correlation with a graph.

A few super-resolution algorithms developed explicitly for light field data have already been proposed in the literature, but they exhibit one or more drawbacks. For example, in their light field super-resolution framework [9], Wanner and Goldluecke propose to first compute a disparity map with sub-pixel precision at each low resolution view of the light field, and then to use the computed maps to super-resolve sequentially each light field view within a multi-frame alike super-resolution method. However, disparity estimation is a very challenging task at low spatial resolution and the disparity errors translate into significant artifacts in the super-resolved light field views.

In a different framework, Mitra and Veeraraghavan propose a light field super-resolution algorithm based on a learning procedure [10]. The low resolution light field is decomposed into multiple light fields with small spatial resolution, light field patches, and each one is assigned a unique disparity value. A Gaussian Mixture Model prior for light field patches is learnt offline for each disparity value and then employed, within a MAP estimator, to super-resolve each light field patch. However, first, the reconstruction quality depends on the selected light field training data, which is not readily available yet; second, the constant disparity assumption within each patch leads to severe artifacts at depth discontinuities in the super-resolved light field views.

Building on our previous work [11, 12], we propose a new light field super-resolution algorithm that targets the drawbacks of the methods in [9, 10]. We develop a super-resolution algorithm that augments the resolution of all the views together, while relying only on a very rough estimate of the disparity at each view, and without any offline learning procedure. In particular, light field super-resolution is cast into a global optimization problem, whose objective function comprises three terms. The first one enforces data fidelity, by constraining each high resolution view to be consistent with its low resolution counterpart. The second one is a warping term, which gathers for each view the complementary information encoded in the others. The third one is a novel graph-based prior, which regularizes the high resolution views by enforcing the geometric structure of the light field. Differently from our previous work [12], where a quadratic graph-based regularizer is employed, the new graph-based regularizer is nonsmooth. This represents a substantial difference, as quadratic regularizers are known to induce a low-pass filtered solution, while the adopted nonsmooth regularizer does
not; it better preserves textures, edges, and fine structures in
general. We solve the new optimization problem using the
Forward-Backward Primal-Dual method [13], and we show
that the new algorithm leads to significantly better reconstruc-
tion results, especially in terms of visual quality.

2. THE LIGHT FIELD

We first introduce the framework studied in this paper. We
consider the light field $U$ as the output of an $M \times M$ array
of cameras, each one equipped with an $N \times N$ pixel sensor.
We denote with $U_k \in \mathbb{R}^{N \times N}$ the $k$-th view in the light field
in column major order, with $k \in \{1, 2, ..., M^2\}$. Also, we
denote with $U_k(x, y) \in \mathbb{R}$ the intensity of the pixel at $(x, y)$
within the same view, with $x, y \in \{1, 2, ..., N\}$.

The particular light field structure is summarized in Fig-
ure 1. Under the Lambertian assumption, the projections of
a pixel $U_k(x, y)$ on the eight neighboring views fall on a vir-
tual square (in red in Figure 1) whose size depends on the
disparity value $d_{x,y}$ associated to $U_k(x, y)$, with $d_{x,y}$ not nec-
essarily integer. This is referred to as the Square Constraint
in [12]. In particular, due to the registration of the light field
views, the projection of the pixel $U_k(x, y)$ falls along a row
or a column in the horizontally and vertically adjacent views,
respectively, while it falls along a diagonal trajectory in the
four diagonally adjacent views. Later on, we will refer to this
configuration as the light field structure.

3. PROBLEM FORMULATION

The light field super-resolution problem deals with the re-
cov ery of the high resolution views $U_k \in \mathbb{R}^{N \times N}$ of the
light field $U$ from the corresponding low resolution views
$V_k \in \mathbb{R}^{(N/\alpha) \times (N/\alpha)}$ of the light field $V$, with $\alpha \in \mathbb{Z}^+$
as the super-resolution factor. We cast the super-resolution
problem into the minimization of the following objective function:

$$u^* \in \arg\min_u \ f_1(u) + \lambda_2 f_2(u) + \lambda_y g(u) \quad (1)$$

where $\lambda_2$ and $\lambda_y$ are positive constants balancing the three
terms. Hereafter, $u_k$ and $u$ denote the vectorized versions
of the view $U_k$ and the whole light field $U$, respectively.

The first term in Eq. (1) enforces the consistency between the
high and low resolution versions of the same view, and it is
typically referred to as the data fidelity term:

$$f_1(u) = \sum_k \|SB u_k - v_k\|_2^2. \quad (2)$$

where $B \in \mathbb{R}^{N^2 \times N^2}$ and $S \in \mathbb{R}^{(N/\alpha)^2 \times N^2}$ denote a blurring
and a sampling matrix, respectively.

The various low resolution views in the light field cap-
ture the scene from slightly different perspectives, therefore,
details dropped by digital sensor sampling at one view may
survive in another one. In order to gather at one view the com-
plementary information from the others, we can enforce that
each high resolution view $u_k$ can generate, by warping, the
other low resolution views $v_{k'}$ in the light field, with $k' \neq k$.
This is the purpose of the second term in Eq. (1):

$$f_2(u) = \sum_k \sum_{k' \in N_k} \|SBF^k_{k'} u_k - v_{k'}\|_2^2 \quad (3)$$

where the matrix $F_{k'}^{k'} \in \mathbb{R}^{N^2 \times N^2}$ is such that $F_{k'}^{k'} u_k \simeq u_{k'}$
and it is typically referred to as a warping matrix, while $N_k$
denotes the set of the eight views surrounding $u_k$. We do
not rely on a precise estimate of the warping matrices, as
this would require us to carry out an expensive disparity esti-
mation with sub-pixel precision at each light field view. We
rather adopt the coarse Square Constraint estimation method
in [12], which relies only on a rough disparity estimation step,
as the noise introduced by the imprecise warping matrices is
handled by the third term in Eq. (1), our regularizer.

As anticipated, the third term in Eq. (1) is a regularizer. It models
each pixel in the high resolution light field as a vertex of a
graph, where edges connect one pixel to its projections
on the other eight surrounding views, as shown in Figure 1.
Formally, we define the regularizer as follows [14]:

\[ g(u) = \sum_i \sqrt{\sum_{j \sim i} w(i, j) (u(i) - u(j))^2} \]  

(4)

where \( w(i, j) > 0 \) is an edge weight capturing the similarity between the pixels \( u(i) \) and \( u(j) \), with \( j \sim i \) the set of pixels \( u(j) \) directly connected to \( u(i) \). The regularizer in Eq. (4) enforces the directly connected pixels in the graph, i.e., those which are the projection of the same 3D point in the scene, to share similar values, thus promoting the light field structure. The square root in the nonsmooth regularizer of Eq. (4) is chosen in order not to over penalize those pixels whose intensity differ significantly from that of its directly connected neighbors in the graph. This prevents the low-pass tendency of the quadratic regularizer in our work [12].

In practice, as shown in Figure 1, the projection of a pixel \( U_k(x, y) \) in a neighboring view \( U_{k'} \) does not lie at integer spatial coordinates. Therefore, in our graph we aim at connecting the pixel \( U_k(x, y) \) to those pixels of \( U_{k'} \) that enclose its projection, i.e., the green pixels in Figure 1. To detect these pixels, we define a search window centered at the pixel \( U_k(x, y) \), and we compute the following weight between the pixel \( U_k(x, y) = u(i) \) and each pixel \( U_{k'}(x', y') = u(j) \) in the considered window:

\[ w(i, j) = \exp \left( -\frac{\|P_k(x, y) - P_{k'}(x', y')\|_F^2}{\sigma^2} \right) \]

(5)

where \( P_k(x, y) \) is a square patch centered at \( U_k(x, y) \), \( \| \cdot \|_F \) is the Frobenius norm, and \( \sigma \) is a tunable constant. The procedure is repeated for each one of the eight surrounding views. The shape of the search window varies from view to view: a 1D search window is sufficient for the horizontally and vertically adjacent views, while a 2D window is preferred for the diagonally adjacent ones, as shown in Figure 1. Finally, among the weights computed for the adjacent view \( U_{k'} \), we keep only the \( n \) highest ones, with \( n \) equal to 2 and 4 for the views equipped with a 1D and 2D windows, respectively, as represented in Figure 1.

4. SUPER-RESOLUTION ALGORITHM

We now have all the ingredients to solve our problem in Eq. (1), whose objective function is the sum of a smooth function \( f = f_1 + \lambda_2 f_2 \) and a nonsmooth one \( \lambda_g g \). In particular, the nonsmooth function \( g \) can be expressed as the \( \ell_{1,2} \)-norm composed with a discrete difference operator:

\[ g(u) = \|Tu\|_{1,2} \]

(6)

where

\[
Tu = \left[ \sqrt{w(1, j)}(u(1) - u(j)) \right]_{j=1}^{j=1}, \ldots, \left[ \sqrt{w((NM)^2, j)}(u((NM)^2) - u(j)) \right]_{j=(NM)^2}^{j=(NM)^2},
\]

and the \( i \)-th row in the above formula is a vector in \( \mathbb{R}^{(j-i)} \), with \( | \cdot | \) the set cardinality. As a consequence, the solution of the problem in Eq. (1) requires an algorithm to deal with convex problems involving nonsmooth functions and linear operators. For this reason, we resort here to the Primal-Dual Proximal Methods [13,15–19]. In the convex setting, the key tool of these methods is the proximity operator [20] of a lower semicontinuous convex function \( \varphi : \mathbb{R}^K \to (-\infty, +\infty] \), defined as \( \text{prox}_{\varphi}(z) = \arg\min_x \varphi(x) + (1/2)\|x - z\|^2 \), \( \forall z \in \mathbb{R}^K \). Proximal methods provide a unifying framework that allows one to address a wide class of convex optimization problems involving nonsmooth penalizations and hard constraints.

We solve the problem in Eq. (1) using the Forward-Backward Primal-Dual FBPD method [13], due to its straightforward implementation. In the FBPD method, detailed in Algorithm 1, each iteration evaluates the gradient \( \nabla f \) of the function \( f = f_1 + \lambda_2 f_2 \) and the proximity operator of the \( \ell_{1,2} \)-norm. The gradient \( \nabla f \) is defined as follows:

\[
\nabla f \ u = 2 \sum_k (SB)^T (SBu_k - v_k) + 2 \sum_{k, k' \in N_k} \lambda_2 (SBF_{k}^k)^T (SBF_{k'}^k u_{k'} - v_k) .
\]

The \( \ell_{1,2} \)-norm proximity operator is listed in [21, Eq. (16)].

Algorithm 1 FBPD [13]

| Initialization | \( \text{Choose} \ u_0^{[0]} \in \mathbb{R}^{(NM)^2} \) and set \( z^{[0]} = Tu_0^{[0]} \), \( \text{set} \ \tau > 0 \) and \( \omega > 0 \) such that \( \tau (\beta/2 + \omega \lambda_g \|T\|^2) < 1 \).
| For \( l = 0, 1, \ldots \) | \( \hat{u}^{[l]} = \nabla f(u^{[l]}) + T^T z^{[l]} \)
| \( u^{[l+1]} = u^{[l]} - \tau \hat{u}^{[l]} \)
| \( \hat{z}^{[l]} = T (2u^{[l+1]} - u^{[l]}) \)
| \( z^{[l+1]} = (z^{[l]} + \omega \hat{z}^{[l]}) - \omega \text{prox}_{\lambda_g \|T\|_{1,2}} \left( \frac{z^{[l]} + \omega \hat{z}^{[l]}}{\omega} \right) \)

5. EXPERIMENTS

We test the proposed super-resolution algorithm with graph-based nonsmooth prior, GB-NS hereafter, on the HCI light field dataset [22], and we compare it to two state-of-the-art light field super-resolution algorithms: [10] and our previous algorithm GB-SQ [12]. The latter relies on a graph model of the light field as well, but binds it to a quadratic regularizer. We also provide the results of a simple bilinear interpolation of the single views as a baseline. Similarly to [10], we crop each light field to a \( 5 \times 5 \) array of views, i.e., \( M = 5 \).

The matrices \( B \) and \( S \) implement a \( \alpha \times \alpha \) box filter and a regular sampler, respectively. For a fair comparison between
the methods GB-SQ and GB-NS, which rely on the same construction procedures for the warping matrices and the graph, we select the same parameter values: we set the size of the patch $P$ to $7 \times 7$ pixels and $\sigma = 0.7229$. Finally, we empirically set $\lambda_2 = 0.15$ and $\lambda_g = 0.0055$ in the problem in Eq. (1) and solve it twice: at the first round, the warping matrices and the graph are built on a bilinearly interpolated version of the light field views, then, the obtained high resolution light field is used to build the warping matrices and the graph for the second round. Finally, for our experiments on the algorithm in [10] we use the code provided by the authors.

Due to space constraints, we report only the results of our experiment for the super-resolution factor $\alpha = 2$. For each light field in the HCI dataset, Table 1 reports the average PSNR (dB) of the $M^2$ reconstructed views. The proposed method GB-NS achieves the higher PSNR on five out of twelve light fields, and in the light fields buddha and stillLife is only 0.02 dB away from the highest PSNR values, achieved by [10] and GB-SQ, respectively. Concerning the remaining five light fields couple, cube, maria, pyramide and statue, here the PSNR happens to be a misleading index, as in these light fields GB-NS achieves better visual results than its competitors [10] and GB-SQ. These five light fields share a particular structure: the foreground hosts an object that changes from light field to light field, while the background consists of the same panel with a fined detailed tree motif. At low resolution, the panel details are completely lost, and building a meaningful graph is really challenging, both for GB-NS and GB-SQ. In fact, no significant visual difference can be perceived between the panels reconstructed by GB-NS and GB-SQ. However, thanks to the smooth nature of its regularizer, GB-SQ reconstructs a smoother panel texture where the error is spread over the whole surface, and this biases the PSNR measure. On the other hand, in each one of the five considered light fields, the object in the foreground exhibits significantly sharper edges in the reconstruction by GB-NS than in those by GB-SQ and [10], where the same edges appear blurred and pixelated. An example is reported in Figure 5, which represents two details from the light fields pyramide (left) and statue (right), for $\alpha = 2$. From the top to the bottom, the original High Resolution light field and the reconstructions of GB-NS, GB-SQ and [10], respectively. GB-NS provides sharper edges than its competitors GB-SQ and [10] in both the light fields.

![Fig. 2. Super-resolved details from the bottom right-most views of the light fields pyramide (left) and statue (right), for $\alpha = 2$. From the top to the bottom, the original High Resolution light field and the reconstructions of GB-NS, GB-SQ and [10], respectively. GB-NS provides sharper edges than its competitors GB-SQ and [10] in both the light fields.](image)

### Table 1. Mean PSNR on the HCI dataset [22], for $\alpha = 2$.

<table>
<thead>
<tr>
<th>Light Field</th>
<th>Bilinear</th>
<th>[10]</th>
<th>GB-SQ</th>
<th>GB-NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>buddha</td>
<td>35.22</td>
<td><strong>39.12</strong></td>
<td>39.00</td>
<td>39.09</td>
</tr>
<tr>
<td>buddha2</td>
<td>30.97</td>
<td>33.63</td>
<td>34.41</td>
<td><strong>34.54</strong></td>
</tr>
<tr>
<td>couple</td>
<td>25.52</td>
<td>31.83</td>
<td><strong>33.51</strong></td>
<td>33.43</td>
</tr>
<tr>
<td>cube</td>
<td>26.06</td>
<td>30.99</td>
<td><strong>33.28</strong></td>
<td>33.11</td>
</tr>
<tr>
<td>horses</td>
<td>26.37</td>
<td>33.13</td>
<td>32.62</td>
<td><strong>33.59</strong></td>
</tr>
<tr>
<td>maria</td>
<td>32.84</td>
<td>37.03</td>
<td><strong>37.25</strong></td>
<td>37.02</td>
</tr>
<tr>
<td>medieval</td>
<td>30.07</td>
<td>33.34</td>
<td>33.45</td>
<td><strong>33.50</strong></td>
</tr>
<tr>
<td>mona</td>
<td>35.11</td>
<td>38.32</td>
<td>39.37</td>
<td><strong>40.05</strong></td>
</tr>
<tr>
<td>papillon</td>
<td>36.19</td>
<td>40.59</td>
<td>40.70</td>
<td><strong>41.56</strong></td>
</tr>
<tr>
<td>pyramide</td>
<td>26.49</td>
<td>33.35</td>
<td><strong>35.41</strong></td>
<td>35.09</td>
</tr>
<tr>
<td>statue</td>
<td>26.32</td>
<td>32.95</td>
<td><strong>35.61</strong></td>
<td>35.43</td>
</tr>
<tr>
<td>stillLife</td>
<td>25.28</td>
<td>28.84</td>
<td><strong>30.98</strong></td>
<td>30.96</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

We presented a new light field super-resolution algorithm that super-resolves all the light field views jointly, thanks to a graph-based regularizer that captures their correlation. In particular, we showed that coupling a graph model of the light field with a non smooth regularization permits to reconstruct high resolution views characterized by a higher visual quality, often supported by the value of the PSNR. The proposed algorithm could further benefit from a non quadratic penalty in the warping term, as this could better handle the outlier pixels due the roughly estimated warping matrices.
7. REFERENCES


