ABSTRACT

In this paper, we propose a new approach for estimating depth maps of stereo images which are prone to various types of noise. This method, based on a parallel proximal algorithm, gives a great flexibility in the choice of the constrained criterion to be minimized, thus allowing us to take into account different types of noise distributions. Our main objective is to present an iterative estimation method based on recent convex optimization algorithms and proximal tools. Results for several error measures demonstrate the effectiveness and robustness of the proposed method for disparity map estimation even in the presence of perturbations.

Index Terms—Proximity operator, total variation, tight frame, proximal algorithms, convex optimization, $\ell_p$-norm, Kullback-Leibler divergence.

1. INTRODUCTION

Given any two images of the same scene acquired by stereoscopic cameras, it can be seen that a degree of similarity exists between the two views. Indeed, the pixel at position $x$ in the left image $I_L$ corresponds to a pixel at position $x-u$ in the right image $I_R$: the disparity of those pixels is denoted by $u$. The matching problem amounts to searching for the disparity field $u$ which minimizes an error measure. Consequently, we can express the stereo matching problem as:

$$\min_u J(u),$$

where,

$$J(u) = \sum_{(x,y) \in D} \phi(I_L(x,y) - I_R(x-u(x,y),y))$$

and $\phi$ is assumed to belong to $\Gamma_0(\mathbb{R})$ which is the class of a proper lower-semi continuous convex function from $\mathbb{R}$ to $]-\infty, +\infty[$. $D \subset \mathbb{Z}^2$ is the considered finite image domain.

The recovery of the depth estimation has been increasingly used in a variety of fields such as movies, web networks, 3D reconstruction and computer games. During the last decades, various methods have been introduced in disparity estimation, especially in computer vision [1]. Stereo matching algorithms are generally classified into two categories: feature based and area based ones. Feature based methods match the feature elements between two images [2], while area based methods perform matching between pixels. Currently, the associated problems are often solved using discrete optimization techniques based on a global approach. Algorithms such as dynamic programming [3], graph cuts [4] and variational methods [5] were proposed. Note that variational-based disparity estimation methods demonstrated excellent performance compared with the state-of-the-art.

In [5], a convex energy function approximation of $J$ is derived and minimized subject to various convex constraints (modeled by convex sets $(C_1)_{1 \leq i \leq m}$) arising from prior knowledge and observed data. The optimization problem (1) to be solved then becomes

$$\min_{u \in L_i} J(u),$$

where $(L_i)_{1 \leq i \leq m}$ are linear operators. However, the employed algorithm requires not only the strict convexity of the criterion to be considered but a quadratic form of it.

This paper describes a new DDE (Dense Disparity Estimation) approach that generalizes the convex optimization approximation presented in [5]. The proposed method is based on some proximity operator ($\text{prox}$) properties and then, it is no longer limited to strictly convex quadratic criteria. Therefore, explicit expressions of the $\text{prox}$ of some norms such as $\ell_1$-norm and more generally $\ell_p$-norm with $p > 1$ or the Kullback-Leibler divergence allow us to apply such proximal algorithms to our disparity map estimation problem. By allowing a rich choice of distance functions, the proposed technique is well-suited for dealing with different types of noise corrupting the observed data, such as Poisson noise,
salt and pepper noise or Gaussian noise.

The remainder of the paper is organized as follows: Section 2 describes the basis tools employed in our approach. Section 3 lays out the problem statement. Results on the Middlebury dataset1 are provided in Section 4. Finally, some conclusions are drawn in Section 5.

2. BACKGROUND

We will now present some tools which are useful in the solution of our minimization problem. In what follows, we will focus on the definition of the proximity operators, total variation and tight frames.

2.1. Proximity operator

The proximity operator of a convex function is a natural extension of the projection operator $P_C$ onto a nonempty closed convex set $C \subset \mathbb{R}^N$ [6, 7, 8]. The projection $P_{C_y}$ of a point $y \in \mathbb{R}^N$ onto $C$ is the solution to the problem:

$$\min_{x \in \mathbb{R}^N} \ i_C(x) + \frac{1}{2} \|x - y\|^2 \tag{4}$$

where $i_C \in \Gamma_0(\mathbb{R}^N)$ is the indicator function of $C$.

In a seminal paper [9], Moreau proposed the following extension of the notion of projection by replacing the function $i_C$ by an arbitrary function $f \in \Gamma_0(\mathbb{R}^N)$. Then, the problem can be rewritten as

$$\min_{x \in \mathbb{R}^N} f(x) + \frac{1}{2} \|x - y\|^2 \tag{5}$$

This problem admits a unique solution which is the proximity operator $\prox_f y$ of $f$ at $y$.

Proximity operators have very attractive properties [7] that make them particularly well-suited to design iterative minimization algorithms. In Table 1, we summarize some of them that will be needed to obtain closed form expressions of the required proximity operators.

2.2. Total variation

Let $H$ and $V$ be two spatial convolution operators corresponding respectively to the computation of horizontal and vertical discrete gradients. Then, a discrete version of the total variation (TV) for which proximity tools can be applied is the following [10, 11]:

$$TV(u) = \sum_{(x,y) \in L} \sqrt{|H u(x,y)|^2 + |V u(x,y)|^2} \tag{6}$$

It should be noticed that, in our simulations, we apply periodic convolutions, which means that the operators $H$ and $V$ are diagonalized by a Fast Fourier Transform.

<table>
<thead>
<tr>
<th>$f(x)$</th>
<th>$\prox_f x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi(x-z), z \in \mathbb{R}^N$</td>
<td>$z + \prox_\phi(x-z)$</td>
</tr>
<tr>
<td>$\phi(x/\rho), \rho \in \mathbb{R} \setminus {0}$</td>
<td>$\rho \prox_\phi(x/\rho)$</td>
</tr>
<tr>
<td>$\phi(Lx)$, semi-orthogonal, $L \in \mathbb{R}^{M \times N}, LL^T = \nu I, \nu &gt; 0$</td>
<td>$x + \nu^{-1}L^T(\prox_\phi(Lx) - Lx)$</td>
</tr>
</tbody>
</table>

Table 1. Some proximity operator properties ($\mathbb{I}$ denotes the identity matrix) [7].

2.3. Tight frame

Frame representations [12] and more precisely tight frame representations have been widely used during the last decade. Such transforms can be described by an analysis frame operator $F$ and a synthesis frame operator $F^T$. $F$ is said to be a tight frame when $F^T F = \mathbb{I}$, where $\nu > 0$.

A simple example of a tight frame is the union of $m$ orthonormal Haar bases over $1$ resolution level.

2.4. PPXA+ Algorithm

Many algorithms have been formulated in the literature to solve optimization problems like (1). In our case, we use the Parallel Proximal Algorithm (PPXA+) [14] which is a flexible tool. PPXA+ (algorithm 1) allows us to minimize a convex criterion $J$ on some closed convex constraint sets $(C_i)_{1 \leq i \leq m}$.

It consists of computing, in parallel, the projections onto the different convex sets $(C_i)_{1 \leq i \leq m}$ and the proximity operator of the criterion $J$.

Suppose that the following assumptions hold.

- $\sum_{i=1}^{m} L_i^T L_i + \mathbb{I}$ is an invertible matrix.

- $\exists \bar{u} \in \mathbb{R}^K$ such that $(\forall i \in \{1, \ldots, m\}) L_i \bar{u} \in \text{ri}(C_i)$ and $\bar{u} \in \text{ri}(\text{dom } J)$.

- $(\lambda_n)_{n \in \mathbb{N}}$ is a sequence of relaxation parameters such that $(\forall n \in \mathbb{N}) \lambda \leq \lambda_{n+1} \leq \lambda_n < 2$, where $\lambda \in [0, 2]$.

Then, the sequence $(u_n)_{n \in \mathbb{N}}$ generated by Algorithm 1 converges to a solution to Problem (3), provided that such a solution exists.

1http://cat.middlebury.edu/stereo/newdata.html

2dom $J$ is the domain of $J$ and the relative interior of a set $C$ is denoted by ri$C$. 

2
Based on (2) and (7), we deduce that:

\[
\text{to perform a Taylor expansion of the nonlinear term}
\]

where

\[
\text{then:}
\]

\[
\text{vexity of the function}
\]

\[
\text{φ}
\]

The function defined in (2) is nonconvex despite the convexity of the function φ. To alleviate this problem, we propose to perform a Taylor expansion of the nonlinear term \(I_R(x - u(x, y), y)\) around an initial estimate \(\bar{u}(x, y)\). We have then:

\[
I_R(x - u(x, y), y) \simeq I_R(x - \bar{u}(x, y), y) - (u(x, y) - \bar{u}(x, y)) I_R^T(x - \bar{u}(x, y), y),
\]

(7)

where \(I_R^T\) is the horizontal gradient of the right image. Based on (2) and (7), we deduce that:

\[
J(u) = \sum_{(x,y) \in D} \phi(T(x, y) u(x, y) - r(x, y))
\]

(8)

where

\[
T(x, y) = I_R^T(x - \bar{u}(x, y), y),
\]

(9)

\[
r(x, y) = I_R(x - \bar{u}(x, y), y) + \bar{u}(x, y) T(x, y) - I_L(x, y).
\]

(10)

Since the occluded image areas (denoted by \(O\)) yield very large disparity errors, they will be discarded in the expression of the similarity criterion. This is performed by rewriting J as:

\[
J(u) = \sum_{(x,y) \in D\setminus O} \phi(T(x, y) u(x, y) - r(x, y))
\]

(11)

The proposed PPXA+ algorithm provides an efficient solution to solve this problem. Our main contribution here lies in the various choices that can be made for the error measure. As already mentioned, the algorithm is based on iterating computations of the proximity operator of J and projections onto convex sets related to constraints. We will first give the explicit forms of the proximity operators of interest in this work which correspond to error measures J taken as \(\ell_p\)-norms or Kullback-Leibler divergence.

3.2. Error measures

3.2.1. \(\ell_p\)-norm

Based on Table 1, it can be deduced that

\[
\tilde{u}(s) = \text{prox}_{\gamma^{-1}|T(s)-\gamma^{-1}r(s)|^2} u(s)
\]

\[
= \frac{1}{T(s)} \left[ \text{prox}_{\gamma^{-1}|T(s)|^2} (T(s)u(s) - r(s)) + r(s) \right]
\]

where \(s = (x, y)\). Then the proximity operator of \(\gamma^{-1}|T(s)|^2\phi\) with \(\phi = | \cdot |^p\) and \(\xi(s) = T(s)u(s) - r(s)\), is [15]:

\[
\text{prox}_{\gamma^{-1}|T(s)|^2\phi}(\xi(s)) =
\]

\[
\begin{cases}
\text{sign(}\xi(s)) \max(\xi(s) - \gamma^{-1}|T(s)|^2, 0), & \text{if } p = 1;
\xi(s) \gamma^{-1}|T(s)|^2 + 1, & \text{if } p = 2;
\text{sign(}\xi(s)) \sqrt{1 + \frac{1}{2}} \sqrt{\frac{1}{\gamma^{-1}|T(s)|^2} + 1}, & \text{if } p = 3;
\left( \frac{\mu + \xi(s)}{8\gamma^{-1}|T(s)|^2} \right)^\frac{1}{p} - \left( \frac{\mu - \xi(s)}{8\gamma^{-1}|T(s)|^2} \right)^\frac{1}{p}, & \text{if } p = 4.
\end{cases}
\]

(12)

3.2.2. Kullback-Leibler divergence

One can note that (7) can be rewritten as

\[
I_R(x - u(s), y) = r(s) + I_L(s) - u(s)T(s) = \tilde{r}(s) - u(s)T(s) = \xi(s).
\]

(13)

Then, the criterion to be minimized takes a form slightly different from (11):

\[
J(u) = \sum_{(x,y) \in D\setminus O} \Phi(I_L(s), \xi(s))
\]

(14)

where \(\Phi(I_L(s), \xi(s))\) is the Kullback-Leibler divergence function given by:
\( \Phi(I_L(s), \zeta(s)) = \begin{cases} I_L(s) \ln \left( \frac{|I_L(s)|}{\zeta(s)} \right) + \zeta(s) - I_L(s), & \text{if } (I_L(s), \zeta(s)) \in [0, +\infty]^2 \\ \zeta(s), & \text{if } I_L(s) = 0, \zeta(s) \geq 0 \\ +\infty, & \text{otherwise.} \end{cases} \) 

(15)

The properties in Table 1 allow us to re-express the proximity operator \( \tilde{u}(s) \) at \( u(s) \) of the function \( \gamma^{-1} \Phi(I_L(s), \tilde{r}(s) - T(s) \cdot \cdot) \) as:

\[
(\forall s \in D \setminus O) \quad \tilde{u}(s) = \frac{1}{T(s)} \left( \tilde{r}(s) - \text{prox}_{\gamma^{-1}|T(s)|^2 \Phi(I_L(s), \cdot)}(\tilde{\zeta}(s)) \right)
\]

where \( \tilde{\zeta}(s) = \tilde{r}(s) - \tilde{u}(s)T(s) \). If \( I_L(s) \in [0, +\infty] \), then

\[
\text{prox}_{\gamma^{-1}|T(s)|^2 \Phi(I_L(s), \cdot)}(\tilde{\zeta}(s)) = \text{prox}_{\gamma^{-1}|T(s)|^2 \Phi(I_L(s), \cdot)}(\tilde{r}(s) - I_L(s) \ln(\cdot) + \cdot)(\tilde{\zeta}(s))
\]

\[
= \tilde{\zeta}(s) - \gamma^{-1}|T(s)|^2
\]

\[
+ \frac{\sqrt{|\tilde{\zeta}(s) - \gamma^{-1}|T(s)|^2|^2 + 4\gamma^{-1}|T(s)|^2I_L(s)|}}{2}.
\]

Furthermore, it can be shown that the above expression remains valid if \( I_L(s) = 0 \). Let us now turn our attention to the convex constraints that can be introduced.

### 3.3. Convex constraints

The motivation behind introducing these constraints is to incorporate additional prior knowledge about the solution.

First, we will constrain the range values of the disparity \( u \) by setting a first constraint:

\[
S_1 = \{ u | u_{\text{min}} \leq u \leq u_{\text{max}} \}.
\]

(16)

Then, we can impose an upper bound on the total variation measure. Indeed, this constraint allows us to recover piecewise and homogeneous areas with sharp edges:

\[
S_2 = \{ u | TV(u) \leq \tau_u \}.
\]

(17)

Finally, one can introduce some prior information on the wavelet coefficients of the disparity. Based on Section 2.3, we can represent the associated constraint in terms of Besov space \( B_{1,1}^{s} \) semi-norm as:

\[
S_3 = \{ u | \sum_{j \geq 1, k \in \mathbb{Z}^2, \omega \in \{H, V\}} |c_{j,k,o}^B(u)| \leq \kappa \}.
\]

(18)

One can link these constraints \( S_1^{1} \leq i \leq S_3 \) to some convex sets \( (C_1)_{1 \leq i \leq 3} \) by noticing that \( S_1 = C_1, S_2 = L_2^{-1}(C_2) \) where \( L_2 = \begin{bmatrix} H & V \end{bmatrix} \) and \( S_3 = L_3^{-1}(C_3) \) with \( L_3 = F \). The projection onto \( C_1 \) is straightforward. The projections onto the convex sets \( C_2 \) and \( C_3 \) were implemented using the iterative projection algorithm proposed in [16].

### 4. SIMULATION RESULTS

The objective of this section is to present numerical results obtained with the proposed method for a variety of synthetic stereo noisy images. In [17], we provided some comparisons with the state-of-the-art DDE method presented in [5]. We briefly report the results of the application of our approach to the standard Middlebury stereo dataset. The objective of this first experiment is to demonstrate the validity of the proposed method without any perturbation. Fig. 1 shows the stereo image pairs considered in this work, namely Corridor, Saw, Teddy and Cones, along with their ground truth images. Our method allows us to compute a smooth disparity map with accurate depth discontinuities as also shown by the preliminary results in [17]. We specify in each case the considered error measure and constraints (the choice was made according to the best performances). We evaluate our method using two error measures between the computed field \( u_c \) and the ground truth field \( u_{\text{ref}} \):

\[
\text{SNR} = 10 \log_{10} \left( \frac{\sum_{s \in D} |u_{\text{ref}}(s)|^2}{\sum_{s \in D} |u_c(s) - u_{\text{ref}}(s)|^2} \right),
\]

(19)

\[
\text{MAE} = \frac{1}{N} \sum_{s \in D} |u_c(s) - u_{\text{ref}}(s)|,
\]

(20)

where \( N \) is the number of pixels. In the presence of noise, we evaluate our method using different kinds of perturbation. To introduce a significant noise variation, we modified the left and the right images by using the \text{imnoise} function in Matlab. Fig. 2 shows the computed disparity maps from the noisy stereo pairs. As we can see, our method is less affected than DDE by noise changes and provides accurate depth maps.

We display in Fig. 2 results obtained with the proposed error measures (\( \ell_1 \)-norm for salt and paper noise, Kullback-Leibler divergence for Poisson noise and quadratic norm for Gaussian noise). Different constraints were also taken into account: we chose the one that leads to the best results either for DDE or PPXA+. As shown in the last column, our approach outperforms the DDE method [5], which proves the effectiveness of our method even in the presence of Gaussian noise. Furthermore, as we can see in the first and second columns, numerical values and visual comparisons on the obtained disparity maps confirm that our PPXA+ algorithm produces good estimates in the presence of other kinds of noise. Finally, Fig. 3 displays the MAE with respect to the Poisson noise intensity \( \alpha \) (a smaller \( \alpha \) corresponding to a higher noise). PPXA+ algorithm outperforms the DDE approach especially when the noise intensity is high.

### 5. CONCLUSIONS

An efficient proximal method that deals with disparity estimation problems has been proposed in this paper. Being very flexible, it allows us to minimize various criteria (not necessarily differentiable nor strictly convex), which can be useful
when perturbations are present during image acquisition. The effectiveness of the proposed approach was demonstrated on different stereo image pairs, while taking into account different kinds of perturbations. In the future, we plan to extend our approach to the case where illumination variations exist.

6. REFERENCES


\[ C_1 + C_3: \text{SNR}=15.19 \, \text{dB}, \text{MAE}=0.58 \]
\[ C_1 + C_2: \text{SNR}=11.24 \, \text{dB}, \text{MAE}=1.82 \]
\[ C_1 + C_2: \text{SNR}=16.96 \, \text{dB}, \text{MAE}=2.08 \]
\[ C_1 + C_3: \text{SNR}=15.68 \, \text{dB}, \text{MAE}=0.5 \]
\[ C_1 + C_2: \text{SNR}=12.19 \, \text{dB}, \text{MAE}=1.54 \]
\[ C_1 + C_2: \text{SNR}=17.67 \, \text{dB}, \text{MAE}=1.59 \]

**Fig. 2.** From left to right: Salt and paper noise, Poisson noise, Gaussian noise. From top to bottom: noisy left image, DDE results, PPXA+ results.

**Fig. 3.** MAE with respect to Poisson noise intensity \( \alpha \) applied to Cones pair.