

NON-LOCAL SMOOTHNESS CONSTRAINTS FOR DISPARITY ESTIMATION IN A VARIATIONAL FRAMEWORK

Raffaele Gaetano, Giovanni Chierchia and Béatrice Pesquet-Popescu

TSI Department, TELECOM-ParisTech
46 rue Barrault, F-75634 Paris Cedex 13, FRANCE
{chierchi, gaetano, pesquet}@telecom-paristech.fr

ABSTRACT

The Non-Local Total Variation (NLTV) has been recently formalized to define new functionals for signal and image analysis, that strictly fit into the widely used variational framework but overcome the *locality* limitation of the classical TV.

This work lies in the context of disparity estimation in a variational framework, where Total Variation represents a common tool to impose a smooth behavior to the desired solution. Here, with reference to a recently proposed disparity estimation technique, several new smoothness constraints based on a NLTV formulation are presented, to prove the effectiveness of the non-local approach in encompassing structural prior knowledge in the problem. Results on several stereo pairs from the Middlebury database are very encouraging, and highlight the importance of a more accurate formulation of the smoothness constraint in the disparity estimation problem.

Index Terms— disparity estimation, total variation, variational estimation, non-local means, set-theoretic estimation

1. INTRODUCTION

Among the recent advances in imaging and video technology, the activities related to the acquisition and playback of 3D scenes are certainly occupying an important role. A direct consequence is the growing interest of research for the estimation of depth information from multi-view sources, an useful tool for a considerable range of applications: from 3D video coding and rendering for 3DTV or FTV (free point-of-view television) to the enhancement of machine vision systems.

When the source is a pair of rectified views of a scene, the depth information can be easily related to the *disparity* among the two views, which represents the difference between the projections of the 3D scene on each viewpoint in terms of displacement between the homologous points.

Research on disparity estimation techniques has been going on for years so far [1]. Earlier methods were mainly based on a *local* approach, seeking for correspondences among points of the two views only within a certain spatial window. These methods, despite their relatively reduced complexity, are likely to fail when no salient feature (textures, contours,

keypoints, etc.) emerge in the local windows to drive the matching process.

The methods based on a *global optimization approach* have been introduced in more recent times to address these issues. The key point of these techniques relies on the definition of an energy functional to minimize, to which the whole disparity field globally contributes. Along with discrete optimization techniques, e.g. based on graph cuts [2], a class of methods that proved particularly competitive is that of *variational* techniques, that apply a gradient descent approach to minimize the energy functional, like in [3, 4]. A technique of particular interest has been presented in [5], in which a variational formulation of the disparity estimation problem is proposed in the framework of convex optimization.

Enforcing a smooth behavior to the disparity map is a common practice, since the problem is ill-posed. To this aim, modeling discontinuities by means of the Total Variation (TV) is a widely used method in many imaging problems [6, 7, 3, 4]. However, the performance limitations of the TV minimization approach have been often highlighted for several applications, including image denoising [7] and optical flow estimation [8]. These limitations spring up in the form of over-smoothing across image contours and/or structural deformations, and are mainly due to the “locality” of the gradient operator involved in the computation of the TV, that only allows for the evaluation of low-level discontinuities, preventing the use of any structural (e.g. geometrical or textural) prior.

To overcome this limitation, a generalized form of TV has been recently formalized in [9], namely the Non-Local Total Variation (NLTV), which through the definition of *non-local derivatives* enables the weighted interaction of each pixel with any other in the image domain. Consequently, it makes possible to process not only color/intensity differences and other low-level image features, but also structural and textural features which can be extracted at higher scales. In [7] the efficiency of this approach is shown for image denoising, where the use of redundancies on a larger scale is proposed by means of a new definition of neighborhood: a pixel j belongs to the neighborhood of i if “a window around j looks like a window around i ”. A subject closer to the one of this paper is

discussed in [8], where a NLTV regularization is imposed on the desired optical flow by replacing classical derivatives by weighted pixel differences on large areas.

In this work, the aim is to convey the principles of NLTV smoothing into the discontinuity model originally proposed in [5]. The smoothness constraint is here reformulated in the NLTV framework, and several alternative definitions of the weighting function will be introduced and tested.

2. REFERENCE VARIATIONAL FRAMEWORK

The estimation of the disparity field for a rectified stereo pair, namely the *stereo matching* problem, consists in seeking a correspondence between pixels of the two images that represent the projection of the same point of the three-dimensional scene. The associated displacements between homologous pixels constitute the target disparity field u . In the variational framework, this problem is classically formulated as the minimization of an energy $J(u)$, jointly taking into account a data likelihood term $L(u)$ and a regularization term that controls the smoothness of the final solution. $L(u)$, positive definite, is directly proportional to the difference between one image of the stereo pair and the other compensated by means of the disparity field u . The Taylor expansion of the functional J around an initial estimate u_0 of the disparity [3, 8] enables the use of variational techniques for minimization.

In this work, we refer to both the variational formulation and optimization method introduced in [5], in which the authors rewrite the problem in the convex optimization framework. More precisely, founding on the principles of set theoretic estimation [10], a solution is sought that provides the minimum cost in terms of data likelihood, subject to an ideally arbitrary number of constraints, each expressed in the form of a closed convex set and whose intersection determines the *feasibility set* of the problem. In particular, two constraint sets have been used based on straightforward properties of disparity field, namely a *limited disparity range set* $R = \{u : u(p) \in [u_{min}, u_{max}], \forall p\}$ and a TV based *smoothness set* defined as

$$S = \{u : \text{TV}(u) = \sum_p \sqrt{(\nabla_x u(p))^2 + (\nabla_y u(p))^2} < \tau\}, \quad (1)$$

$\nabla u = (\nabla_x u, \nabla_y u)$ being the gradient of the disparity u . The minimization problem can be eventually written as

$$\underset{u}{\text{minimize}} \quad L(u) + \iota_R(u) + \iota_S(\nabla u), \quad (2)$$

where, for any closed convex set C , ι_C is the *characteristic function* of C ¹.

An efficient convex optimization tool based on proximal splitting, namely the *Parallel ProXimal Algorithm* (PPXA+), has been recently introduced [11] to minimize a sum of convex and lower-semicontinuous functions. In the most

¹In the convex analysis literature, the characteristic function of a convex set C , namely $\iota_C(x)$, has value 0 if $x \in C$, $+\infty$ otherwise

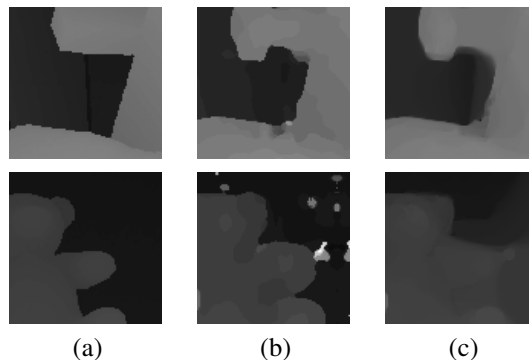


Fig. 1. Details of disparity maps for the *Teddy* stereo pair: (a) ground truth, (b) initial estimate of the disparity (MAE = 0.86), (c) disparity map provided by PPXA+ with TV constraint (MAE = 0.75).

general form, each function of the sum may take as argument a different linearly transformed version of the minimizing variable. Since the characteristic functions are convex and lower-semicontinuous, it is easy to notice that the problem in Eq. 2 can be tackled using PPXA+ if the convexity and lower-semicontinuity condition is verified also for the likelihood term $L(u)$. This is the case in this work, where we rely on a L^1 -norm based likelihood function.

3. NON-LOCAL TOTAL VARIATION CONSTRAINTS

A major concern about the smoothness constraint used in [5] is that the isotropic Total Variation constraint in Eq. (1) takes into account all discontinuities in the disparity map in the same way. This means that the smoothing action induced by the PPXA+ algorithm on the current estimate at each iteration makes no discrimination between undesired discontinuities, due to errors in the initial estimate, and admissible ones, occurring in correspondence of edges among objects at different depths in the scene. As a main effect, the final solution is generally not immune by over-smoothing through “real” disparity edges or keeping undesired block effects resulting from the initial estimation.

To better appreciate such effects, let us observe the details in Fig.1: the initial disparity (second column) is evidently affected by errors w.r.t. the ground truth (first column), but presents globally sharp contours, while the result obtained with PPXA+ algorithm using the TV constraint (PPXA+/TV from now on) of Eq. 1, although globally closer to the ground truth, suffers from over-smoothing of “real” contours. To mitigate this effect, one possibility is to limit the intensity of the smoothing action, both enlarging the constraint set, by augmenting τ in Eq. (1), or reducing the contribution of the smoothing action in the weighted averaging step of PPXA+.

Given these observations, in this work several new discontinuity measures have been tested, that rely on the principles

of Non-Local Total Variation introduced in Section 1. In the context of this work, these characteristics are mainly used to adapt the smoothing action to the different structural contexts in which discontinuities happen in the disparity map.

Formally, the Non-Local Total Variation operator used in this work has the following form:

$$\text{NLTV}(u) = \sum_{p \in \mathcal{D}} \sum_{q \in \mathcal{N}_p} \omega_{pq} |u(p) - u(q)|, \quad (3)$$

with $\mathcal{N}_p \subseteq \mathcal{D}$ being a suitably chosen neighborhood of pixel at position p and ω_{pq} the weight controlling the interaction of pixel p with the pixel q belonging to its neighborhood. Note also that the L^2 -norm of the original TV formulation has been replaced by the anisotropic L^1 -norm in Eq. (3): this choice is justified by the better preservation of corners provided by this norm in the TV framework [9]. Using the NLTV as discontinuity measure changes the problem formulation of Eq. 2 into:

$$\underset{u}{\text{minimize}} \quad L(u) + \iota_R(u) + \iota_{S'}(F u), \quad (4)$$

where $S' = \{u : \text{NLTV}(u) < \tau\}$, and $F = (F_1, \dots, F_{|\mathcal{N}_p|})^\top$ is a concatenation of linear operators such that:

$$(F_i u)(p) = u(p) - u(q_i), \quad \text{with } q_i \in \mathcal{N}_p. \quad (5)$$

In the presented non-local framework, the definition of a discontinuity measure is complete once a choice of the weights ω_{pq} is given for each neighborhood \mathcal{N}_p . In the following, several options for this choice will be proposed and discussed, and a comparison among the different solutions will be provided in Sect. 4. All the weighting functions are introduced for the case of grayscale images, to comply with the reference PPXA+ based technique.

3.1. Constant Weights

A simple solution consists in equally weighting all the points in every neighborhood. Considering a fixed size neighborhood \mathcal{N}_p , this choice amounts to fixing

$$\omega_{pq} = \frac{1}{|\mathcal{N}_p| - 1} \quad \forall p, q. \quad (6)$$

Note that such choice transforms the general formulation of Eq. (3) into a sort of multi-directional formulation of the classical anisotropic TV, with gradients in the various directions expressed as absolute differences between the central pixel and each pixel in the neighborhood. Improvements may be expected due to the finer description of directional discontinuities.

3.2. Patch-based Weights

A second more farsseeing choice makes use of an approach inspired by the Non-Local Means denoising algorithm described in [7]: given two locations p and $q \in \mathcal{N}_p$, the weight

ω_{pq} will be dependent on the similarity between the $K \times K$ patches B_p, B_q built around the pixels p and q on the texture image (the image of the stereo pair onto which the disparity is mapped). More precisely:

$$\omega_{pq} = \exp\left(\frac{-G_\sigma * \|B_p - B_q\|^2}{h^2}\right), \quad (7)$$

where $\|B_p - B_q\|^2$ is the L^2 -norm difference between the two patches in vector form, G_σ is a Gaussian kernel with standard deviation σ and $h > 0$ is a filtering parameter.

Based on the assumption that a change in the disparity generally corresponds to a change into the textural properties of the involved objects, this choice allows to reduce the contribution to NLTV of a discontinuity in the disparity map that matches with an important change in the texture image. Correspondingly, the smoothing action across this ‘‘probably real’’ discontinuity is limited.

3.3. Similarity-Proximity based Weights

When computing weights through patch differences, as described in the previous section, the main issues concern the choice of the patch size K . Using bigger patch sizes, differences in the visual appearance may generally be better caught, but the similarity between two ‘‘close’’ patches whose central pixels lie across an edge may result enhanced, and the corresponding weight be too high. On the contrary, small patches become too sensitive to noise and micro-textural variations.

To avoid this trade-off, another possibility is to refer to the weighting strategy described in [12], which measures changes in the visual appearance by observing single pixel values and their Euclidean distance on the image support, namely:

$$\omega_{pq} = \exp\left[-\left(\frac{\Delta c_{pq}}{\gamma_c} + \frac{\|p - q\|}{\gamma_p}\right)\right], \quad (8)$$

where $\Delta c_{pq} = |I(p) - I(q)|$ is the absolute intensity difference between values of p and q , observed on the texture image I that matches with disparity, and $\gamma_c > 0, \gamma_p > 0$ are parameters depending respectively on the image values domain and the size of the neighborhood \mathcal{N}_p . This choice provides a solution similar to [8], with the main difference that the smoothness prior is expressed as a constraint set instead of a regularization term, to match the formulation of Eq. (2).

3.3.1. Texture-based Similarity

As a final variant proposed in this work, a solution is sought that might use a more accurate description of local visual differences without incurring all the drawbacks of the patch-based approach described in Sec. 7. To this aim, the weight expression of Eq. (8) is here enforced by a difference measure between the texture descriptors associated to each pixel

	<i>Venus</i>	<i>Teddy</i>	<i>Cones</i>
<i>GlobalGCP</i>	0.489	0.272	0.443
<i>HistoAggr</i>	0.788	0.278	0.434
PPXA+	0.631	0.330	0.645
<i>CurveletSupWgt</i>	0.831	0.299	0.832
<i>GC+occ</i>	1.166	0.303	0.585

Fig. 3. Comparison of the PPXA+/TSP method with other techniques on the Middlebury website. The results are expressed in mean average error (MAE). See <http://vision.middlebury.edu/stereo> for the full list of references.

location p , based on the occurrence of intensity values within a $T \times T$ patch around p , namely:

$$\omega_{pq} = \exp \left[- \left(\frac{\Delta c_{pq}}{\gamma_c} + \frac{\|p - q\|}{\gamma_p} + \frac{\Delta H_{pq}}{\gamma_h} \right) \right] \quad \text{with (9)}$$

$$\Delta H_{pq} = \sum_{i=1}^B |\mathbf{H}_p(i) - \mathbf{H}_q(i)|. \quad (10)$$

Here, the \mathbf{H}_p vector represents the histogram of intensity values within the patch around p . To ensure consistency to the texture descriptors among the various pixels, a fixed number B of equally spaced bins is determined for each histogram, always covering the whole intensity range.

This choice is motivated by the fact that, using such statistical descriptors, the sensibility w.r.t. noise and micro-textural variations occurring on surfaces at the same depth is limited. Apart from gaining in accuracy on such surfaces, this choice also allows for the use of (relatively) smaller patch sizes, thus implying a more accurate weight selection also close to the edges of the disparity map.

4. EXPERIMENTAL RESULTS

The effectiveness of the proposed modifications has been proved by testing the different versions of the algorithm on several grayscale images of the Middlebury database [1].

The initialization u_0 is provided by means of a simple dense block matching algorithm with Normalized Cross Correlation (NCC) similarity measure and 11×11 sized blocks. In all the experiments, the value of the upper bound τ is then chosen as a fraction $\alpha TV(u_0)$ of the TV computed on the initial estimate: for the classical TV case, α is set to 0.8 while for all NLTV cases $\alpha = 0.65$. This difference is due to the fact that smaller values of τ cause over-smoothing if the TV constraint is used, while using NLTV makes possible to increase the intensity of the smoothing action. Moreover, for all the experiments with NLTV, the neighborhood \mathcal{N}_p has size 5×5 , while the patches are 7×7 pixels when using both the weights of Eq. (7) and (10). Finally, $h = 35$ in Eq. (7), $\gamma_c = 5$, $\gamma_p = 2.5$ and $\gamma_h = 5$ in both Eq. (8) and (10). All these parameter values are determined on an heuristic basis,

provided that the algorithm does not show critical instabilities w.r.t. them. General PPXA+ parameters are set as in [5].

Results for four of the tested images are reported in Fig. 2, along with the achieved *Root Mean Square Error* (RMS), *Mean Average Error* (MAE) and *Bad Pixel* (BAD) quality measures used in the Middlebury dataset [1]. Results on other images, not reported for brevity, exhibit a similar performance improvement. To avoid unfair comparisons due to the fact that no color information is used here, a detailed evaluation will be provided only w.r.t. the reference technique. For sake of completeness, a comparison with some state-of-the-art techniques for some of the tested images is shown in Fig. 3, w.r.t. the MAE figure.

Remarkably, the overall quality of the result is improved with almost all the NLTV based proposed variants. In particular, all the techniques using a non uniform weighting strategy achieve a significant gain, while the only PPXA+/CTV solution seems unable to increase performances, resulting even worse than the reference technique in one case. Also at a visual inspection, the increased sharpness of the contours in the disparity maps is evident. However, as remarked in other works using NLTV regularization, the results are not immune from the so-called *staircasing effect*, exhibiting as the tendency to produce piecewise constant regions, mainly affecting surfaces with a relevant gradient nature like *Venus*: it is worth noticing that the BAD indicator is particularly sensible to this effect, while the MAE and RMS measures, more affected by over-smoothing errors along the contours, have a more regular progression and also enhance the performance assessment w.r.t. other state-of-the-art techniques.

As expected, the PPXA+/TSP solution is providing the best results in average, proving that the accurate use of visual appearance to “validate” discontinuities in the disparity field may be a rewarding choice. Only in one case, for the *Cones* stereo pair, the results assess slightly worse than using PPXA+/SP: this is much probably due to the significant presence of contours at higher frequencies, around which the (window-based) texture descriptors may loose reliability.

Clearly, the use of the NLTV based smoothness constraint generally implies an increase of the computational complexity, proportional to the size of the neighborhood. However, as highlighted in [5], one of the qualifying points of PPXA+ concerns its parallelization potential: in our case, the computation of multiple gradients simply extends the degree of task-parallelization, potentially reducing the impact of the newly introduced NLTV based constraint on computational time.

5. CONCLUSIONS

In this work, the advantages of Non-Local Total Variation for the definition of smoothness constraints in the context of disparity estimation have been inspected. With reference to a recently proposed variational technique, a NLTV reformulation of the discontinuity model is here presented, along with sev-


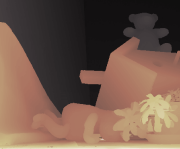
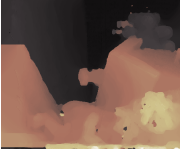
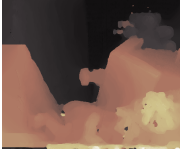
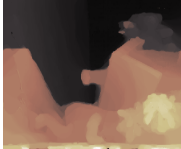
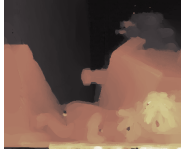
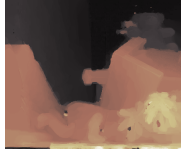


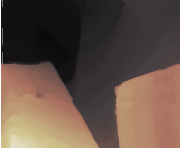
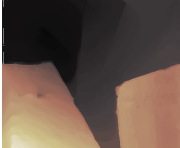

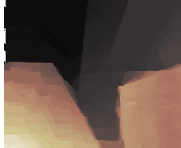
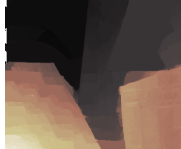

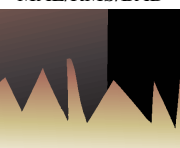
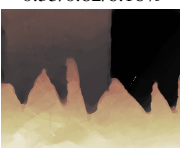
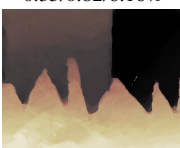
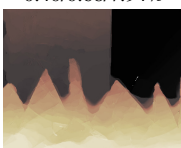

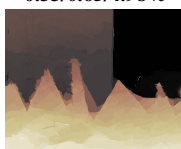
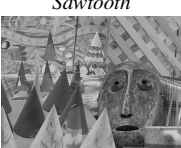




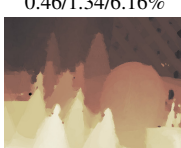
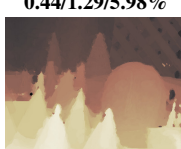
Left Image	Ground Truth	PPXA+TV	PPXA+CTV	PPXA+PB	PPXA+SP	PPXA+TSP
						
Teddy	MAE/RMS/BAD	0.75/2.03/16.81%	0.75/2.13/16.84%	0.68/1.85/15.92%	0.66/1.81/14.91%	0.63/1.77/14.25%
						
Venus	MAE/RMS/BAD	0.35/0.82/6.16%	0.35/0.82/6.16%	0.40/0.68/7.94%	0.34/0.66/5.61%	0.33/0.65/4.98%
						
Sawtooth	MAE/RMS/BAD	0.50/1.32/6.90%	0.50/1.32/6.90%	0.51/1.31/6.98%	0.46/1.34/6.16%	0.44/1.29/5.98%
						
Cones	MAE/RMS/BAD	0.69/2.13/12.91%	0.69/2.13/12.91%	0.64/2.02/10.81%	0.64/2.02/10.56%	0.64/2.03/10.90%

Fig. 2. Disparity estimation of some images of the Middlebury database. Columns 3 to 7: results using the reference technique with classical TV constraint (PPXA+TV), constant weighting (PPXA+CTV), patch-based weighting (PPXA+PB), similarity-proximity based (PPXA+SP), similarity/proximity based with texture descriptors (PPXA+TSP).

eral weighting strategies that allow to take advantage of different structural features of the source images. Experimental results proved this choice rewarding, and motivating further research to address newly opened issues (staircasing effect) and optimize the resulting technique for implementation on parallel architectures.

6. REFERENCES

- [1] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proc. IEEE Workshop Stereo and Multi-Baseline Vision (SMBV 2001)*, 2001.
- [2] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. Eighth IEEE Int. Conf. Computer Vision ICCV 2001*, 2001, vol. 2, pp. 508–515.
- [3] W. Miled, J.-C. Pesquet, and M. Parent, "A convex optimization approach for depth estimation under illumination variation," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 813–830, 2009.
- [4] R. Ben-Ari and N. Sochen, "Stereo matching with Mumford-Shah regularization and occlusion handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2071–84, Nov. 2010.
- [5] M. El Gheche, C. Chau, J.-C. Pesquet, J. Farah, and B. Pesquet-Popescu, "Disparity map estimation under convex constraints using proximal algorithms," in *Proc. of IEEE Workshop on Signal Processing Systems, SIPS 2011*, Beirut, Lebanon, 2011.
- [6] P. L. Combettes and J.-C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. on Im. Proc.*, vol. 13, no. 9, pp. 1213–1222, 2004.
- [7] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.
- [8] M. Werlberger, T. Pock, and H. Bischof, "Motion estimation with non-local total variation regularization," in *2010 IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*. June 2010, pp. 2464–2471, IEEE.
- [9] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling Simulation*, vol. 7, no. 3, pp. 1005, 2009.
- [10] P. L. Combettes, "The foundations of set theoretic estimation," in *Proceedings of the IEEE*, 1993, vol. 81.
- [11] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," to appear in *Pacific Journal of Optimization*, 2012. Available: http://www.optimization-online.org/DB_HTML/2010/11/2825.html
- [12] K. Yoon and I. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–6, Apr. 2006.