Dense and Swift Mapping with Monocular Vision

Pedro Piniés, Lina Maria Paz, and Paul Newman

Abstract The estimation of dense depth maps has become a fundamental module in the pipeline of many visual-based navigation and planning systems. The motivation of our work is to achieve fast and accurate in-situ infrastructure modelling from a monocular camera mounted on an autonomous car. Our technical contribution is in the application of a Lagrangian Multipliers based formulation to minimise an energy that combines a non-convex data term with *adaptive pixel-wise regularisation* to yield the final local reconstruction. We advocate the use of constrained optimisation for this task. We shall show it is swift, accurate and simple to implement. Specifically we propose an *Augmented Lagrangian (AL)* method that markedly reduces the number of iterations required for convergence with more than 50% of reduction in all cases compared to the state-of-the-art approach. As a result, part of this significant saving is invested in improving the accuracy of the depth map. We introduce a novel per pixel *inverse depth uncertainty estimation* that allows us to apply adaptive regularisation on the initial depth map: high informative inverse depth pixels require less regularisation, however its impact on more uncertain regions can be propagated providing significant improvement on textureless regions. To illustrate the benefits of our approach, we ran our experiments on three synthetic datasets with perfect ground truth for textureless scenes. An exhaustive analysis shows that AL can speed up the convergence up to 90% achieving less than 4cm of error. In addition, we demonstrate the application of the proposed approach on a challenging urban outdoor dataset exhibiting a very diverse and heterogeneous structure.

1 Introduction

The creation of dense workspace models from cameras alone has long been a focus of robotics research. The mapping task is sometimes seen in a limited light as simply a pre-

Pedro Piniés · Lina María Paz · Paul Newman
Mobile Robotics Group, Dept. Engineering Science, University of Oxford,
17 Parks Rd., Oxford, OX1 3PJ, United Kingdom
e-mail: \{ppinies,linapaz,pnewman\}@robots.ox.ac.uk

cursor or at best dual for localisation. When maps were simply sparse collections of points [1] this narrow perspective was reasonable. But robots that can, through their own motion, produce *dense* reconstructions offer a new vista for autonomous and semi-autonomous plant inspection. But to do so the reconstruction process must be rapid allowing in-situ formation of the dense scene structure. This paper is about precisely that competency - creating dense depth maps *rapidly*.

Recent work has made clear the potential of variational methods in producing dense volumetric reconstructions of small workspaces under controlled lighting conditions [13, 10, 6]. In [13] the authors address the problem as a depth map estimation from a set of keyframes with corresponding camera poses obtained from a PTAM system. An energy function is optimised based on a data term that measures the photoconsistency over a set of small-baseline images, as well as total variation (TV) based regularisation term. This preserves sharp depth discontinuities due to occlusion boundaries, while simultaneously enforcing smoothness of homogeneous surfaces. The problem is stated as the minimisation of an energy functional comprising both terms by using an alternation scheme with a good initial seed. A similar approach is adopted in [10]. In this case, the solution relies on a primal-dual formulation successfully applied in solving variational convex functions that arise in many image processing problems [4]. Despite the non-convex nature of the energy functional for the depth map estimation, the authors provide theoretical insights to decouple the terms leading to a two-stage optimisation. Their solution is based on the application of the well known Quadratic Penalty (QP) method firstly introduced in [15] in the context of optical flow estimation with a similar energy formulation. In contrast to [13], an efficient cumulative discrete cost volume is considered to compute one of the terms allowing a robust initialization of the depth map before the optimisation. While [13] avoids an exhaustive point-wise search to find a minimum solution, [10] provides strategies that accelerate the search while achieving good accuracy. A different approach was introduced in [6]. Instead of optimising depth maps, a different energy functional over a 3D volume is formulated using a primal-dual algorithm for the minimisation. The authors use an implicit truncated signed distance function (TSDF) representation to compute the globally optimal fusion using a (TV regularised) convex energy. Then the surface is extracted by finding the zero level set of the accumulated TSDF. As input, the minimisation receives initial depth map estimates that are not required to be highly-accurate.

Despite these energy minimisation approaches reach soft real-time performance, their application to active tasks such as planning and obstacle avoidance is critical. For instance, in [2] the authors follow a DTAM based approach to estimate dense depth maps for live collision avoidance of a MAV. Their analysis shows that online generation of each depth map requires usually 900 primal dual iterations to converge with an estimated final error of 10cm, requiring a significant time of 500ms for this task. More recently the works of [5, 1, 8] introduce the use of the Augmented Lagrangian (AL) in the field of video restoration and general image inverse problems. As first paper contribution, we demonstrate the efficacy of the Augmented Lagrangian method [3] for dense depth map creation using monocular cameras which at the time of writing was the first time this had been done. Our experiments show that AL method dramatically reduces the number of iterations (more than 50%) required for the decoupling approach adopted in [10].

A second contribution lies in our consideration of how to progress from an initial guess to a final solution. In particular we need to reinforce pixels in the seed solution containing

[1] as in early SLAM formulations

plausible depth estimates and propagate its effect over those pixels with less accurate depth estimates. We advocate that large texture-less areas of the RGB images produce noisy and often grossly misleading meaningless regions in the initial depth map that greatly impede successful optimisation. We propose a inverse depth uncertainty estimation to calculate per-pixel adaptive confidences that aid the trade-off between the data fidelity term and the regularisation term. This provides a novel approach that affords us a principled way to only seed the optimisation with pixels from regions which should yield reliable depth estimates. Furthermore, we offer an illustrative study of the effect of three different photo-consistency measures. Our motivation is to understand the degree to which each affects final solution accuracy because each determines an initial seed solution for the optimisation.

In section 2 we briefly review the approach presented in [10] to build an initial depth map from monocular frames. Dealing with non-convex data terms requires careful attention, thus section 3 is devoted to explain the so often used Quadratic Penalty method and the proposed Augmented Lagrangian method. How to estimate per-pixel depth uncertainties for adaptive regularisation is introduced in section 4. An evaluation of the precision and convergence of the complete approach on monocular synthetic datasets with perfect Ground Truth is described in 5. Also, we demonstrate the application of the proposed approach on challenging urban outdoor dataset exhibiting a very diverse and heterogeneous structure. Finally, we draw our conclusions in section 6.

2 Building an Initial Seed

As in [10], to obtain an initial depth map for our optimisation algorithm, we build a cost volume $\mathbf{C}_r$ that accumulates, for a uniformly sampled set of inverse depths ξ_j, $j = 1 : d$, the photo-consistency error of overlapping images. The reason for using an inverse depth representation being that a uniform discretisation of ξ produces a uniform sampling of epipolar lines in an image.

Figure 1 shows a 2D top view of the process used to initialise each voxel of the cost volume. Given a pixel $u_i \in \mathbf{u}$ in a reference image I_r and an inverse depth ξ_j the corresponding pixel in a neighbouring image $I_k \in \mathbf{I}(r)$, where $\mathbf{I}(r)$ is the set of images that overlap with I_r, is given by the warp

$$\mathbf{w}_k(u_i, \xi_j) = \pi(T_{kr}\pi^{-1}(u_i, \xi_j)) \tag{1}$$

where $\pi(\mathbf{x})$ describes a perspective projection of a 3D point $\mathbf{x}$, $\pi^{-1}(u_i, \xi_j)$ is the back-projection of a pixel u_i with inverse depth ξ_j and $T_{kr} \in SE(3)$ is the relative transformation between cameras corresponding to images I_k and I_r.

We have studied the effect of different photo-consistency measures in the accuracy of depth map estimates. In particular, we have tested, for different window sizes W, the Sum of Squared Differences (ρ^{SSD}), the Sum of Absolute Differences (ρ^{SAD}) and the Normalised Cross Correlation (ρ^{NCC}) which are described in table 1.

The average photometric error $\mathbf{C}_r(u_i, \xi_j)$ for all images $I_k \in \mathbf{I}(r)$ and for each inverse depth ξ_j is given by:

$$\mathbf{C}_r(u_i, \xi_j) = \frac{1}{|\mathbf{I}(r)|} \sum_{k \in \mathbf{I}(r)} \rho^*_{ij}(I_k, u_i, \xi_j) \tag{2}$$

Table 1 Similarity metrics

Metric	Definition	Equation
Sum of Square Distances	ρ_{ij}^{SSD}	$\sum_{i\in W}\|I_r(u_i) - I_k(\mathbf{w}_k(u_i,\xi_j))\|_2$
Sum of Absolute Distances	ρ_{ij}^{SAD}	$\sum_{i\in W}\|I_r(u_i) - I_k(\mathbf{w}_k(u_i,\xi_i))\|_1$
Normalized Cross Correlation	ρ_{ij}^{NCC}	$\frac{\sum_{i\in W} I_r(u_i) I_k(\mathbf{w}_k(u_i,\xi_i))}{\sqrt{\sum_{i\in W} I_r^2(u_i)\sum_{i\in W} I_k^2(\mathbf{w}_k(u_i,\xi_i))}}$

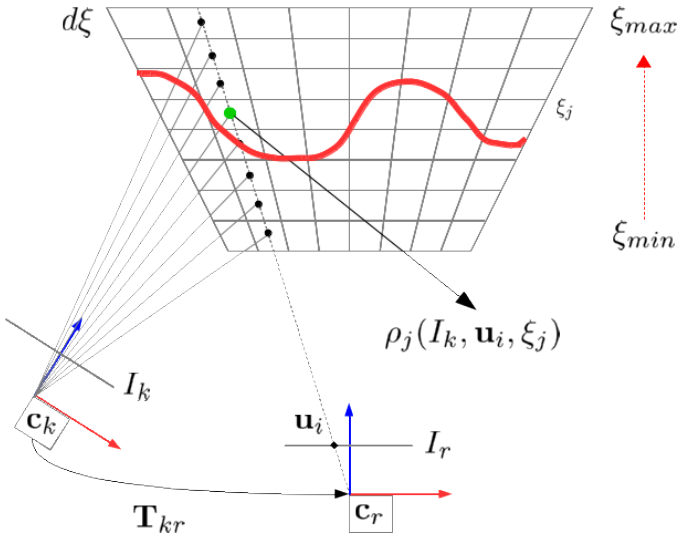


Fig. 1 This example illustrates the process of building the "data fidelity" term for our energy minimisation problem. A discretised cost volume is built to accumulate the photo-consistency error: for each pixel u_i in a reference image frame I_r, we back-project the pixel along a discrete set of inverse depth distances ξ_j in the interval $[\xi_{max}\ \xi_{min}]$ obtaining the 3D pose for the centre of each voxel in the cube. Then each voxel centre gets projected into the current image frame (I_k, c_k) and we compare the corresponding intensities according to a predefined similarity metric ρ_{ij}^*. The results of these comparisons are stored in the corresponding cells. This process is repeated for all overlapping image frames $I_k \in \mathbf{I}(r)$. The final average minimum cost is calculated according to Eq.(2). This calculation is achieved by inducing an exhaustive search in the cost volume. The corresponding per-pixel initial inverse depthmap is associated with the voxel at the minimum cost rendering a surface as illustrated with the red curve.

where $|\mathbf{I}(r)|$ is the number of images that overlap with I_r and ρ_{ij}^* represents the chosen similarity metric.

Once the cost volume is computed, an inverse depth map $\xi_r(\mathbf{u})$ over the whole set of pixels $\mathbf{u}$ can be recovered by searching for the minimum cost for each pixel:

$$\xi_r(\mathbf{u}) = \arg\min_{\xi_j} \mathbf{C}_r(\mathbf{u}, \xi_j) \tag{3}$$

Since $\xi_r(\mathbf{u})$ is usually noisy, it will be used as initial seed for the optimization algorithm explained in the next section. Without loss of generality and to improve readability, we will drop the subindex r and will refer only to ξ and $\mathbf{C}$ in the remaining of the paper.

3 Dealing with non-convex data terms

In this section we show how we can improve the initial crude depth map using search over a regular partitioning which replaces the so called "winner-takes-all approach" described in Eq.(3). The searched solution $\xi(\mathbf{u})^*$ minimises the energy functional:

$$\min_{\xi} E(\xi) = \int_{\Omega} w(\mathbf{u})||\nabla\xi(\mathbf{u})||_{\varepsilon} + \lambda \mathbf{C}(\mathbf{u}, \xi(\mathbf{u}))d\mathbf{u} \tag{4}$$

where $\Omega \in \mathscr{R}^2$ is the depth map domain, $w(\mathbf{u})$ is a per pixel weight based on I_r gradient that reduces the effect of regularization across image edges, $||\nabla\xi(\mathbf{u})||_{\varepsilon}$ is the Huber norm and λ is a parameter used to define the trade-off between the regulariser and the data term. After discretising the domain Ω, a depth map is redefined as the set $\xi = [\ldots, \xi_{ij}, \ldots]$. Therefore, we can express the previous equation as:

$$\min_{\xi} E_R(\xi) + \lambda E_D(\xi) \tag{5}$$

where $E_R(\xi)$ is the regularisation term and $E_D(\xi)$ is the data term that corresponds with the information stored in the cost volume. In order to solve Eq.(5), we will make use of the iterative Primal Dual optimisation algorithm presented in [4]. This algorithm requires both the regulariser and the data term to be convex, however $E_D(\xi)$ is not a convex function. One solution to this problem is to decouple both terms and solve instead the following equivalent constrained optimisation

$$\begin{aligned} \min_{\xi,\eta} \quad & E_R(\xi) + \lambda E_D(\eta) \\ \text{s.t.} \quad & \xi = \eta \end{aligned} \tag{6}$$

The advantage of the decoupling approach is that it allows us to independently solve for the regulariser term using convex optimisation methods and for the data term using a simple exhaustive search in the cube. Obviously, both problems are in fact coupled by the constraint. In the following subsections we will discuss the main possible solutions of the previous constraint optimisation problem: The Quadratic Penalty (QP) and the Augmented Lagrangian (AL) whose numerical implementation is illustrated in **Algorithm 1**. The interested reader can find a more detailed discussion of these and more general techniques for constraint minimisation in [3].

3.1 Quadratic Coupling Penalty

We briefly describe the algorithm proposed in [10] in order to obtain an improved $\xi(\mathbf{u})^*$ depth map solution from the initial seed in eq. 3. This approach is based on eliminating the constraints through the use of a coupling penalty function. Popularly a simple quadratic penalty function suffices. Using this approach, Eq.(6) is minimised by sequentially solving an unconstrained minimisation problem of the form

Algorithm 1 ξ = EnergyMinimisation($\eta, \theta, \varepsilon, \alpha$)

{Initialization of variables:}
$\tau, \sigma > 0, \gamma \in [0,\ 1], \theta \in [0,\ 1]$
{For each pixel ij}
$\xi_{ij}^0 = \eta_{ij}, \mathbf{p}_{ij}^0 = 0$
$\bar{\xi}_{ij} = \xi_{ij}^0$
while $t \leq N$ **do**
  {Update Dual}
  $\mathbf{p}_{ij}^{t+1} = \frac{\mathbf{p}_{ij}^t + \sigma w_{ij} \nabla \bar{\xi}_{ij}^t}{1+\sigma\varepsilon}$
  $\mathbf{p}_{ij}^{t+1} = \mathbf{p}_{ij}^{t+1} / \max(1, |\mathbf{p}_{ij}^{t+1}|)$
  {Update Primal}
  $\xi_{ij}^{t+1} = (\xi_{ij}^t + \tau w_{ij} \nabla \cdot \mathbf{p}_{ij}^{t+1} + \frac{\tau}{\theta^t} \eta_{ij}^t - \tau \alpha_{ij}^t) / (1 + \frac{\tau}{\theta})$
  {Relaxation}
  $\bar{\xi}_{ij}^{t+1} = \xi_{ij}^{t+1} + \gamma(\xi_{ij}^{t+1} - \xi_{ij}^t)$
  η_{ij}^{t+1} = SubpixelSearch($\xi^{t+1}, \theta, \mathbf{C}, \lambda, \alpha^t$)
  $\alpha_{ij}^{t+1} = \alpha_{ij}^t + \frac{1}{\theta}(\xi_{ij}^{t+1} - \eta_{ij}^{t+1})$
end while

Algorithm 2 η = SubpixelSearch($\xi, \theta, \mathbf{C}, \lambda, \alpha$)

{Accelerated search:}
$r = \sqrt{2\theta\lambda(C_{ij}^{max} - C_{ij}^{min})}$
{Exhaustive search for $\eta_{ij} \in [\xi_{ij} - r, \xi_{ij} + r]$}
$\eta_{ij}^{aux} = \arg\min_{\eta_{ij}} \frac{1}{2\theta}(\xi_{ij} - \eta_{ij})^2 + \lambda C_{ij}(\eta_{i,j}) + \alpha_{i,j}(\xi_{i,j} - \eta_{i,j})$
{Subpixel refinement:}
$\nabla E^{aux} = \lambda \nabla C_{ij}(\eta_{ij}^{aux}) + \frac{\eta_{ij}^{aux} - xi_{ij}}{\theta} - \alpha_{ij}$
$\nabla^2 E^{aux} = \lambda \nabla^2 C_{ij}(\eta_{ij}^{aux}) + \frac{1}{\theta}$
$\eta_{ij} = \eta_{ij}^{aux} - \nabla E^{aux} / \nabla^2 E^{aux}$

$$\min_{\xi,\eta} E_R(\xi) + \frac{1}{2\theta}\|\xi - \eta\|_2^2 + \lambda E_D(\eta) \tag{7}$$

where $E(\xi, \eta) \to E(\xi)$ as $\theta \to 0$. In general, the main disadvantages of this approach, reported in [3], are its slow convergence and ill-conditioning for small values of θ. Nevertheless, for the depth map estimation problem, this algorithm has shown an admirable performance in practice. Note that Lagrange multipliers play no direct role in this method. The new energy functional in eq. 7 allows us to split the minimisation into two different problems that are alternatively solved until convergence:

- First, for a fixed η solve:

$$\min_{\xi} E_R(\xi) + \frac{1}{2\theta}\|\xi - \eta\|_2^2 \tag{8}$$

  which corresponds to the well known TV-ROF convex denoising problem that can be solved using a primal-dual algorithm [4]. In this case η represents a noisy image whereas ξ is the searched denoised result.
- Second, for a fixed ξ solve:

$$\min_{\eta} \frac{1}{2\theta}\|\xi-\eta\|_2^2+\lambda E_D(\eta) \tag{9}$$

this optimisation is performed by a point-wise exhaustive search followed by an accelerated subpixel refinement for each η in the cost volume as it is explained in [10]. We show a general implementation of the primal dual solution along with the sub-pixel refining steps in **Algorithms 1** and **2**. Lines 6-16 illustrate the main iterative per-pixel primal dual algorithm. Line 9: ascend gradient step to update the dual variable $\mathbf{p}$. Line 11: descend gradient step to update the primal variable ξ. The parameters τ and σ are calculated via preconditioning [11].

3.2 Lagrange Multipliers

We must now briefly mention the role of Langrange Multipliers as a precursor to our use of the "Augmented Lagrangian" in the next section. The original constrained optimisation Eq.(6) can be transformed to an unconstrained minimisation problem by introducing the Lagrangian function

$$E_R(\xi)+\alpha^T(\xi-\eta)+\lambda E_D(\eta) \tag{10}$$

where α is a Lagrange multiplier associated with the original constraint. In this approach the Lagrange multiplier is treated on an equal basis with the variables ξ,η, which means that in order to solve the unconstrained problem we have to iterate as well for α. Although there exist different methods to iteratively update ξ,η, α and solve the Lagrangian equation, we are going to concentrate on the Augmented Lagrangian method explained in the next subsection.

3.3 Augmented Lagrangian

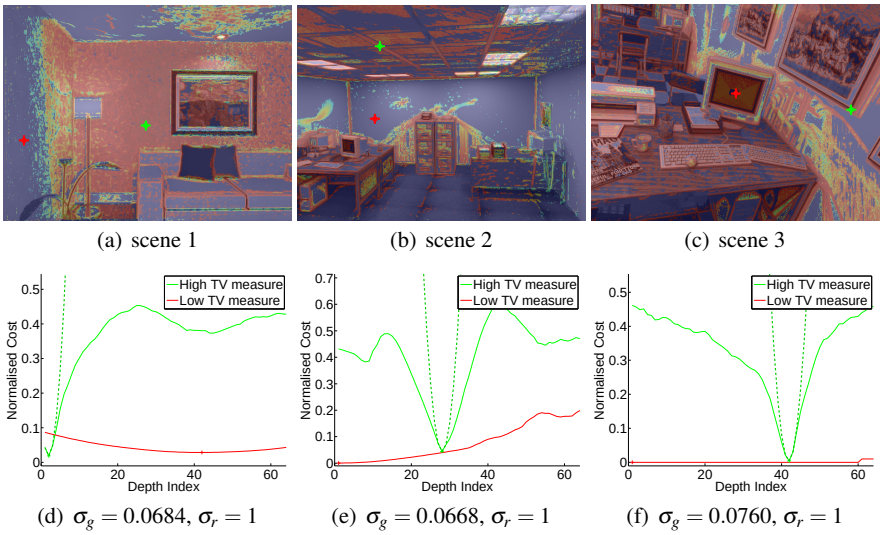
The Augmented Lagrangian belongs to a class of methods called *methods of multipliers* in which the penalty regularization is combined with the Lagrange Multipliers method. The resultant objective function, called the *Augmented Lagrangian*, is sequentially minimized to obtain a solution to the original constrained problem. In our case the augmented Lagrangian is given by

$$E_R(\xi)+\alpha^T(\xi-\eta)+\frac{1}{2\theta}\|\xi-\eta\|_2^2+\lambda E_D(\eta) \tag{11}$$

The main advantages of this method over the previous ones are: First, convergence can be attained even when θ does not decrease to zero improving the stability of the algorithm. Second, there exists a simple update of the Lagrange Mulltiplier α that tends to make it converge faster to its proper value than pure Lagrange Multipliers approaches [3].

As in the Quadratic Penalty section, Eq.(11) is minimized by alternatively solving the following sub-problems until convergence

- First, for a fixed η solve:



(a) scene 1 (b) scene 2 (c) scene 3

(d) $\sigma_g = 0.0684$, $\sigma_r = 1$ (e) $\sigma_g = 0.0668$, $\sigma_r = 1$ (f) $\sigma_g = 0.0760$, $\sigma_r = 1$

Fig. 2 Adaptive selection of λ. To weight the contribution of each pixel in the data term, we estimate the uncertainty of the depth represented as a Gaussian distribution on the cost along the inverse depth range, with mean centred in the depth for which the cost is minimum. First row shows the pixel-wise uncertainty overlapping the reference image for three synthetic datasets. Green crosses represent examples of highly informative pixels u_g, while red crosses determine pixels u_r with more uncertainty. Second row shows variability of the cost along 64-discrete inverse depth index values for the two examples. The fitted Gaussian is illustrated for the green case.

$$\min_{\xi} E_R(\xi) + \alpha^T(\xi - \eta) + \frac{1}{2\theta}\|\xi - \eta\|_2^2 \tag{12}$$

using a primal-dual algorithm [4] since the previous optimization is convex in ξ

- Second, for a fixed ξ solve:

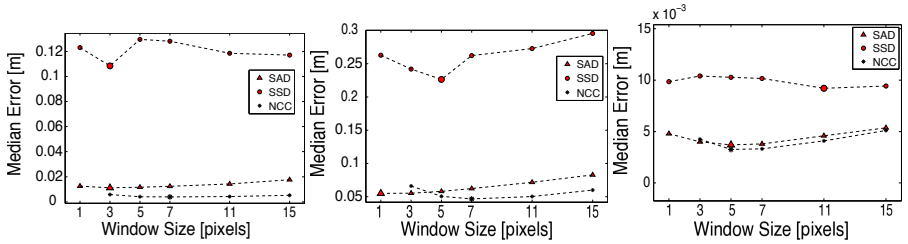
$$\min_{\eta} \alpha^T(\xi - \eta) + \frac{1}{2\theta}\|\xi - \eta\|_2^2 + \lambda E_D(\eta) \tag{13}$$

using a point-wise exhaustive search for each η in the cube.

- Third, update α

$$\alpha = \alpha + \frac{1}{\theta}(\xi - \eta) \tag{14}$$

In contrast to the Quadratic Penalty method, we have introduced the new variable α. Although it implies a change in the numerical implementation, the iterations required for convergence are substantially reduced as we will show in section 5. In particular, it affects the update of the primal variable ξ (line 11 in algorithm 1) as well as the accelerated search and smoothing step for sup-pixel accuracy (algorithm 2, lines 4 and 6). For better readability, we have highlight in red color the differences between the QP and AL numerical implementations. Notice that the changes between both algorithms are minimal.



(a) scene 1 (b) scene 2 (c) scene 3

Fig. 3 Median error obtained after optimisation on three different synthetic scenes. For each similarity metric (SAD, SSD, NCC), the plots show the optimal window size to achieve the minimum error. In general, NCC yields more accurate results on all datasets (see the scale of y-axis).

4 Adaptive Regularisation

In this paper we also exploit the concept of the uncertainty on the inverse depth to reinforce regularisation on non-informative depth map regions. Regularisation plays an important role in achieving highly accurate depth-maps in small scenes. However, depending on the quality of the metric used as well as the initial depth seed, the effect of the regularisation does not necessarily provide a positive impact on the final solution. The lack of texture in some regions of the scene (blank walls, texture-less surfaces, ...) generates in fact a non-informative cost along the volume. Figure 2 bottom shows, for two different pixels u_g and u_r in the reference image, the corresponding set of cost values store in the cube along the inverse depth interval $[\xi_{max}\ \xi_{min}]$. Notice that the 1D cost functions present a low or high variability depending on whether the pixel belongs to a texture-less region u_r (flat walls, floor, roof, ...) or to an informative one u_g. For each pixel $u_i \in \mathbf{u}$ in the reference image, we can estimate the inverse depth uncertainty using the following second order approximation,

$$C(u_i,\xi) \sim C(u_i,\xi^*) + (\xi-\xi^*)\nabla C(u_i,\xi)|_{\xi=\xi^*} + \frac{1}{2}\nabla^2 C(u_i,\xi)|_{\xi=\xi^*}(\xi-\xi^*)^2 \quad (15)$$

where $C(u_i,\xi^*)$ represents the minimum cost along the sampled distances. Figure 2 bottom, shows the quadratic approximation of the cost function at a particular pixel. Note that the quadratic is naturally centred at the sampled depth ξ^* at which the cost is minimum. To associate uncertainties with per-pixel inverse depth estimates we look at the curvature of the correlation surface, i.e., how strong the minimum in the cost volume is at the winning inverse depth [14]. Under the assumption of small noise, photometrically calibrated images, and densely sampled inverse depth, the uncertainty is approximated by a normal distribution synthesised as follows

$$\xi(u_i) \sim \mathcal{N}(\xi^*(u_i),\Sigma_\xi) \quad (16)$$

where the variance is locally estimated by the hessian $\Sigma_\xi \propto 1/\nabla^2 C(u_i,\xi)|_{\xi=\xi^*}$ in the inverse depth point where the cost is minimum. Eq.(15) allows us to calculate a per-pixel adaptive trade off $\lambda(\mathbf{u})$ between the data fidelity term and the regularisation depending on the quality of the information in the initial depth map.

Table 2 Convergence Analysis for AL and QP

Scene 1, range = [1.655 3.445] [m]							
	Median Error [m]		Energy		$\|\xi-\eta\|_2$		% iter saved
Metric	AL	QP	AL	QP	AL	QP	
SAD 3	0.0111	0.0107	3283.77	3233.44	0.0452	0.0500	57 %
SSD 3	0.1084	0.1459	1288.70	1342.70	0.0466	0.0130	74 %
NCC 7	0.0038	0.0032	26081.11	27804.42	0.0480	0.0497	63 %
Scene 2, range = [1.102 6.186] [m]							
	Median Error [m]		Energy		$\|\xi-\eta\|_2$		% iter saved
Metric	AL	QP	AL	QP	AL	QP	
SAD 1	0.0549	0.0551	8278.24	8317.93	0.0426	0.0456	66 %
SSD 5	0.2264	0.2821	8141.76	8383.51	0.0406	0.0236	72 %
NCC 7	0.0467	0.0488	49449.81	49356.62	0.0462	0.0496	55 %
Scene 3, range= [0.773 5.953] [m]							
	Median Error [m]		Energy		$\|\xi-\eta\|_2$		% iter saved
Metric	AL	QP	AL	QP	AL	QP	
SAD 5	0.0037	0.0043	11410.13	11456.60	0.0460	0.0190	84 %
SSD 11	0.0092	0.0089	2594.75	2577.19	0.0482	0.0098	90 %
NCC 5	0.0032	0.0032	75876.31	75601.43	0.0423	0.0433	67 %

Analysis of Errors, Energy convergence and constraint fulfill at the final solution for both the Augmented Lagrange (AL) and the Quadratic Penalty (QP) methods using different similarity measures to obtain the initial seed.

$$\lambda(u_i) \propto \frac{1}{\Sigma_\xi} \tag{17}$$

Figure 2 top, shows the output image that results after the calculation of the per pixel variance for three synthetic indoor datasets. Notice that the reference image is overlapped for better interpretation.

5 Results

5.1 Evaluation on Indoor Synthetic Datasets

We have conducted our experiments on three synthetic indoor scenes that provide high precision depth maps from images taken at 30Hz [7, 9][2][3]. Our chosen scenes consider both close and far objects from the camera and partial occlusions. We first evaluate the influence of the similarity metric used to obtain the initial solution. Recall that the metrics under evaluation are the SSD, the SAD and the NCC. After executing the AL optimisation algorithm for each metric, we calculate the median error of the depth-map solution with respect to the ground truth. In order to compare the accuracy of AL and QP algorithms we will calculate:

$$cost(\mathbf{u}) = median(\|\xi(\mathbf{u})_{GT} - \xi^*(\mathbf{u})\|_1) \tag{18}$$

Figure 3 shows for all scenes the median errors obtained for window sizes ranging in the interval $W = [1 \dots 15]$. This preliminary analysis shows that, for the correct window size,

[2] http://www.doc.ic.ac.uk/ ahanda/VaFRIC/index.html

[3] http://www.doc.ic.ac.uk/ ahanda/HighFrameRateTracking/downloads.html

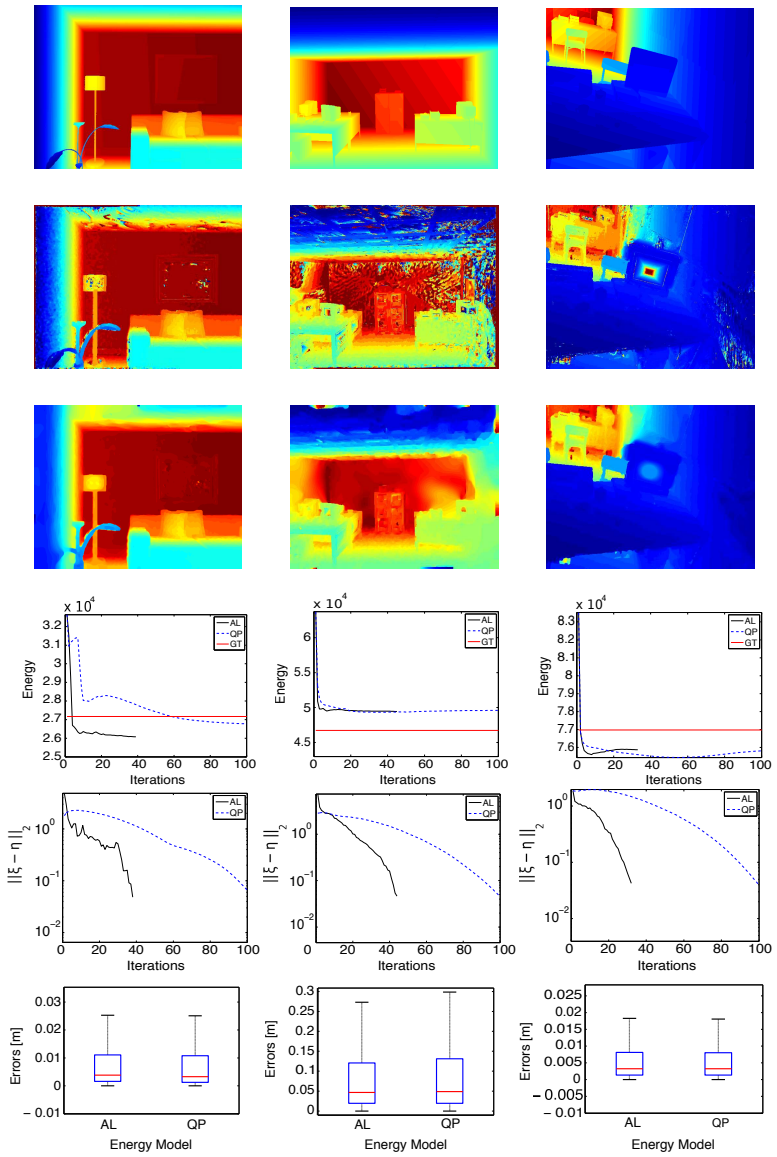


Fig. 4 Convergence and Accuracy Analysis for the proposed Augmented Lagrangian (AL) method in comparison to the common Quadratic Penalty (QP) approach. The experiments are shown for three synthetic scenes: left, scene 1; middle, scene 2; right, scene 3. First row, ground truth depth map. Second row, Initial seed obtained with a NCC-based cost volume at the optimal window size as reported in table 2. Third row, achieved depth-map solution. Fourth row, energy evolution over all iterations. Fifth row, evolution of the constraint $\|\xi - \eta\|_2$ per iteration. Notice that AL (black solid line) outperforms QP (blue light line) to converge at the final solution. Energy is evaluated at the ground truth (GT) which constant value is displayed with a red line. Sixth row, boxplots of the error distributions of the per pixel inverse depth map estimates. The tops and bottoms of each box are the 25th and 75th percentiles of the samples, respectively. The distances between the tops and bottoms are the inter-quartile ranges. The line in the middle of each box is the sample median. AL and QP achieve high accurate depth-maps with similar error distributions. However, AL achieves the final solution faster than QP.

the NCC measure achieves the best results. This can be a consequence of the NCC invariance to illumination changes. Since the NCC is usually costly to evaluate we can also see that the SAD even with a window size of 1 performs relatively well and can be used in case of computation constraints. "Median Error" column in table 2, shows for the AL and QP algorithms the lowest median errors obtained for all similarity measures at their optimal window size. Observe that NCC produces the best results and that the QP and AL algorithms produce similar accurate estimates.

We also studied the convergence properties of the AL and QP algorithms described in the paper. In order to obtain a fair comparison, we have applied the same stop criteria to both methods: First, the relative decrease in the energy minimization has to be below a given threshold (to assure we can not make much progress) and second, the equality constraint is considered to be fulfilled if $\|\xi - \eta\| \leq 5e-2$. Figure 4 shows the energy evolution for both algorithms using NCC with optimal window size for the initial seed. In two of the three synthetic scenes (Fig. 4 second,third column) both methods converge to similar final energy and constraint values. Notice that, in the limit, ξ and η must achieve the same values, thus the decouple energies for AL and QP should approximate very well the original energy in equation 4. However, the most important advantage of the AL method, which is one the contributions of this paper, over the QP method is its faster convergence requiring fewer iterations to achieve the same result. In figure 4, second row, we observe how the quadratic constraint decreases rapidly for AL and so the energy falls to its minimum value. Table 2 column nine, shows the gain percentage of AL with respect to the number of iterations required for QP. The proposed approach requires 50% less iterations till convergence for all cases. Fig. 4, sixth row, shows the histogram of the errors for AL and QP. Note that the accuracy of the solution is not traded for speed.

5.2 Dense reconstruction of outdoor scenes

Our goal is to show that the AL method in combination with adaptive regularisation improve the appearance of the point cloud capturing the diverse shapes present in outdoor environments. Our motivation is that while a sparse map provides a compact representation for autonomous navigation, higher level robot tasks can require denser maps to improve scene understanding. We have a forwards-facing camera mounted on a car travelling forwards and sensing distant objects with a low parallax. This leads us to rely on an improved regularisation method to reinforce depth on critical parts of the scene. In our case, a suitable assumption is to expect to find many affine surfaces in the environment, like roads, pathways, building facades or vehicle surfaces.

The input to our pipeline consists of only two consecutive image frames gathered by a camera at 25Hz. This choice enables us to estimate the depth of dynamic objects (particularly important in urban environments), which could be potentially disregarded by a long sequence integration. The sensor is mounted on a car that traverses a city environment. Figure 5, shows the reconstruction of three different scenes with heterogeneous geometry (walls, roads and vegetation). To track the camera, we employ our own scaled Visual Odometry system [12].

Figure 5 first row, shows the per pixel inverse depth uncertainty. As it is expected, road surfaces and distant regions exhibit low information. The use of the per-pixel adaptive

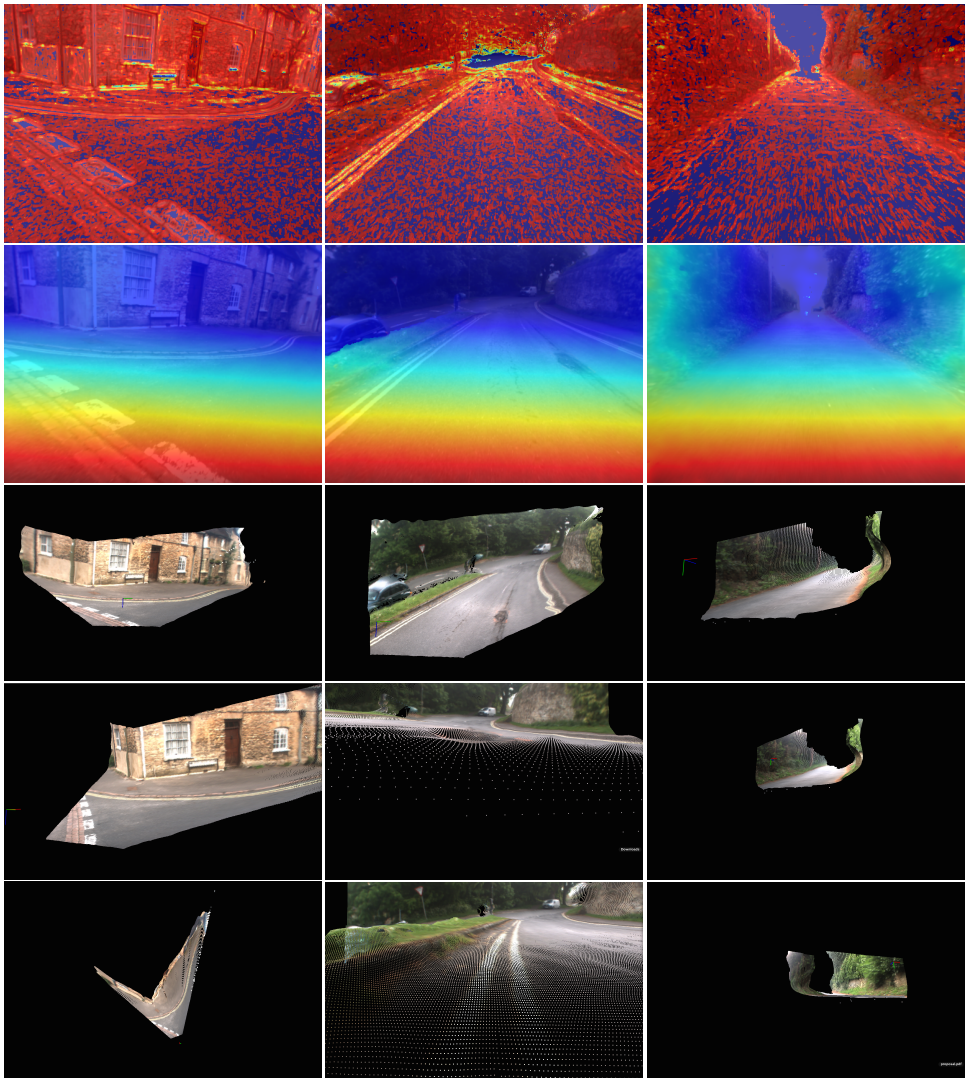


Fig. 5 3D reconstruction of outdoor scenes from monocular images. The use of Adaptive regularisation improve the appearance of the point cloud capturing the diverse shapes present in the environment. First row, pixel-wise depth uncertainty. Second row, Inverse depth map obtained after 30 primal dual iterations. Third-fifth rows, different camera views of the final 3D dense reconstruction.

regularisation allows us to recover most of the structure. A video showing more details of the execution of the algorithms is available at (`http://youtu.be/LrNv9QCKH1s`).

6 Conclusions

We have shown the efficacy of the Augmented Lagrangian method for depth map estimation using monocular cameras. As a result we can substantially reduce the number of iterations required for convergence, more than 50% of reduction in all cases, compared to state of the art algorithms based on Quadratic Penalty methods. We have also performed an exhaustive study of different photo-consistency measures SSD, SAD and NCC and different windows sizes in order to improve the accuracy of the initial depth map used as seed in the optimisation algorithm. As was expected, NCC provides the best results due to its intrinsic properties to cope with illumination changes. Finally, we introduce a novel per pixel inverse depth uncertainty estimation that allows us to apply adaptive regularisation on the initial depth map: high informative inverse depth pixels require less regularisation, however its impact on more uncertain regions can be propagated providing significant improvement on textureless regions.

References

1. Afonso, M., Dias, J., Figueiredo, M.A.T.: An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems. IEEE Trans. on Image Processing **20**(3), 681–695 (2011)
2. Alvarez, H., Paz, L., Sturm, J., Cremers, D.: Collision avoidance for quadrotors with a monocular camera. In: Experimental Robotics - The 14th International Symposium on Experimental Robotics, ISER 2014, Marrakesh, Morocco. Springer (2014)
3. Bertsekas, D.P.: Constrained optimization and lagrange multiplier methods. Computer Science and Applied Mathematics, Boston: Academic Press **1** (1982)
4. Chambolle, A., Pock, T.: A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)
5. Chan, S.H., Khoshabeh, R., Gibson, K.B., Gill, P.E., Nguyen, T.Q.: An augmented lagrangian method for total variation video restoration. IEEE Trans.s on Image Processing **20**(11), 3097–3111 (2011)
6. Graber, G., Pock, T., Bischof, H.: Online 3D reconstruction using Convex Optimization. In: 1st Workshop on Live Dense Reconstruction From Moving Cameras, Workshops ICCV 2011, Barcelona, Spain
7. Handa, A., Newcombe, R.A., Angeli, A., Davison, A.J.: Real-Time Camera Tracking: When is High Frame-Rate Best? In: ECCV 2012, Florence, Italy
8. Li, C., Yin, W., Jiang, H., Zhang, Y.: An efficient augmented Lagrangian method with applications to Total Variation minimization. Computational Optimization and Applications **56**(3), 507–530 (2013)
9. Nardi, L., Bodin, B., Zia, M.Z., Mawer, J., Nisbet, A., Kelly, P.H.J., Davison, A.J., Luján, M., O'Boyle, M.F.P., Riley, G., Topham, N., Furber, S.: Introducing SLAMBench, a performance and accuracy benchmarking methodology for SLAM. In: ICRA 2015
10. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: In: ICCV 2011, pp. 2320–2327. IEEE Computer Society, Washington, DC, USA
11. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: (ICCV) 2011, Barcelona, Spain., pp. 1762–1769
12. Smith, M., Baldwin, I., Churchill, W., Paul, R., Newman, P.: The new college vision and laser data set. The International Journal of Robotics Research **28**(5), 595 – 599 (2009)
13. Stühmer, J., Gumhold, S., Cremers, D.: Real-time dense geometry from a handheld camera. In: DAGM 2010, pp. 11–20. Darmstadt, Germany (2010)
14. Szeliski, R.: Computer Vision: Algorithms and Applications, 1st edn. Springer-Verlag New York, Inc., New York, NY, USA (2010)
15. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime TV-L1 optical flow. In: Pattern Recognition DAGM, pp. 214–223. Heidelberg, Germany (2007)