

Geometric Multi-Model Fitting with a Convex Relaxation Algorithm

Paul Amayo, Pedro Piniés, Lina M. Paz and Paul Newman
Oxford Robotics Institute
University of Oxford, UK

{pamayo, ppinies, linapaz, pneman}@robots.ox.ac.uk

Abstract

We propose a novel method for fitting multiple geometric models to multi-structural data via convex relaxation. Unlike greedy methods - which maximise the number of inliers - our approach efficiently searches for a soft assignment of points to geometric models by minimising the energy of the overall assignment. The inherently parallel nature of our approach, as compared to the sequential approach found in state-of-the-art energy minimisation techniques, allows for the elegant treatment of a scaling factor that occurs as the number of features in the data increases. This results in an energy minimisation that, per iteration, is as much as two orders of magnitude faster on comparable architectures thus bringing real-time, robust performance to a wider set of geometric multi-model fitting problems.

We demonstrate the versatility of our approach on two canonical problems in estimating structure from images: plane extraction from RGB-D images and homography estimation from pairs of images. Our approach seamlessly adapts to the different metrics brought forth in these distinct problems. In both cases, we report results on publicly available data-sets that in most instances outperform the state-of-the-art while simultaneously presenting run-times that are as much as an order of magnitude faster.

1. Introduction

Many interesting applications in computer vision such as homography estimation, plane detection, and motion estimation demand the ability to fit geometric models onto noisy data. This is a non-trivial task in that the scene typically consists of multiple geometric structures. Moreover, the observed data is likely to be contaminated with noise from different sources including measurement sensor noise and outliers. These are the main factors leading to biased solutions. Therefore, many multi-model fitting algorithms are driven fundamentally by their capacity to robustly deal with the complexity of the data and unknown distributions of data errors and outliers [15]. Extracted geometric models are often also drivers of other algorithms such as camera

calibration, robotic navigation, and 3D reconstructions creating an increased desire for algorithms that are not only robust to contamination but exhibit quick and repeatable run-times.

The most well-known solution to the geometric multi-model fitting problem found in the literature is RANSAC [5], a greedy approach consisting of two steps: in the first step, a set of model proposals is sampled from the model parameter space in a hypothesis-verification fashion. A refinement step is applied over the best-selected model, supported by the maximum number of inliers. To deal with multiple models, an extended solution like [19] suggests applying RANSAC sequentially over the remaining data points. Running alongside this are clustering methods that greedily maximise the inliers of geometric models [10]. Presenting approaches such as J-linkage [18] and its expansion to the continuous space T-linkage [8] that analyse and merge clusters in the preference space. However, as these greedy approaches do not implicitly consider the overall classification of the data to geometric models, their performance deteriorates in situations with high noise or outlier clutter.

Global energy-based approaches [7, 22, 4] have gained popularity by presenting a more general optimisation framework that jointly fit all models present in multi-structural data. In practice, global energy methods aim to find an optimal fitting solution by accounting for the model error in a data fidelity term for a given metric. In addition, the set of solutions can be constrained by encouraging spatial smoothness in a regularisation term. A more complete formulation also considers the number of models as a parameter to optimise for [7]. These approaches have been shown to outperform greedy methods such as RANSAC at the expense of intractable computational complexity with respect to the number of models fitted.

Minimisation of this energy can be performed via combinatorial algorithms, such as α -expansion [2], which operate in a discrete domain. These approaches have been largely applied to the multi-label image segmentation problem formalised as a graph cut Markov Random Field (MRF) based approach but when applied to geometric multi-model fitting have exhibited better performance than greedy methods [7]. However, attempts to parallelise this approach have so far

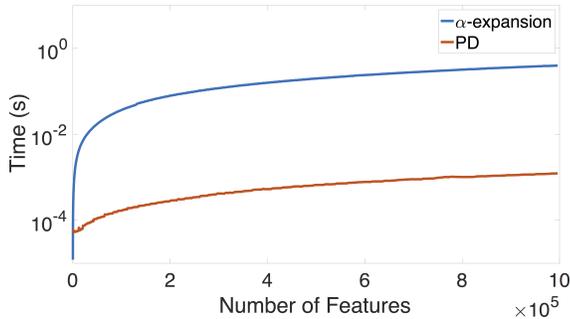


Figure 1: Comparison of the run-time per iteration of α -expansion and that of the primal-dual optimisation (PD) in an experiment with four models. The times are averaged over 100 runs on a machine with a 12 core Intel i7-5930K 3.50GHz CPU and a Nvidia GeForce GTX Titan Black 6048MB GPU. On this architecture, as the number of features increases so does the difference between the run times per iteration of the two methods. With the PD running at speeds two orders of magnitude (100x) faster than the α -expansion.

proven to be unsuccessful [14], leading to a linear increase of run-times with the number of features in the data and the number of geometric models to be fitted. Limiting their use with high-resolution data in applications with real-time demands.

In contrast, we present a general-purpose solution that exploits continuous multi-label optimisation, a framework that has been shown to be superior in terms of parallelisation and run-time performance [14]. Analogous to [7], the problem of geometric multi-model fitting is converted to a multi-labelling problem where each model is represented through a label function over the data points. The optimal solution to the relaxed problem is obtained efficiently through a first order primal-dual (PD) optimisation [3]. As this supports per-element evaluation it facilitates a parallel implementation on General Purpose Graphical Processing Unit (GPGPU) hardware.

In Figure 1 a comparison of the run-time per iteration of α -expansion¹ and a CUDA implementation of the PD optimisation on dedicated hardware is shown. It can be seen that there is a clear reduction in run-time per iteration, of up to two orders of magnitude, through the PD optimisation as the number of features increases.

We summarise the main contributions in this paper as follows:

- We propose a novel global energy-based approach to fit and segment multi-structural data via a CONvex Relaxation ALgorithm (CORAL). Unlike greedy meth-

ods, which maximise the number of inliers, this approach efficiently searches for a soft assignment of points to models by minimising the energy of the overall classification.

- We provide a designed energy functional that encompasses spatial regularisation on a continuous label function while simultaneously minimising the number of labels. This without making any assumptions on the underlying data structure.
- We demonstrate the adaptability of the approach to two important vision applications including multi homography estimation and plane detection from RGB-D images. These applications show the versatility of our approach to different metrics with different norms used in both instances. Further, they show that when applied to both simulated and publicly available datasets with real imagery, the proposed approach outperforms other global energy based methods for geometric multi-model fitting in terms of both speed and accuracy.

This paper is organised as follows. In Section 2 the proposed framework is described using a continuous energy formulation. Section 2.1 describes the two-stage convex relaxation algorithm. The practical advantages of the geometric model fitting framework are demonstrated in Section 3. Finally, advantages of the proposed method are discussed and conclusions drawn in Section 4.

2. Energy Minimisation for Multi-Model Fitting

In this work geometric multi-model fitting is framed as an optimisation problem where the quality of the solution is linked to an energy functional. The first term of the energy considers the geometric error of all data points to their corresponding models/labels. The energy also encompasses prior knowledge about locality by ensuring that data points that are spatially close have a higher likelihood of belonging to the same model. Finally, a different sort of regularisation that promotes compactness is imposed by favouring solutions that explain the data using as few models as possible. Equation 1 represents these three ideas in a general formulation.

$$\underbrace{\sum_{l=1}^L \int_{\Omega} \rho_l(\mathbf{u}, \phi_l(\mathbf{u})) d\Omega}_{\text{Data Energy}} + \lambda \underbrace{\sum_{l=1}^L \int_{\Omega} \omega_{\mathcal{N}} R(\nabla_{\mathcal{N}} \phi_l(\mathbf{u})) d\Omega}_{\text{Smoothness Energy}} + \underbrace{\beta L}_{\text{Model Count Energy}} \quad (1)$$

The first term, ρ_l , in Equation 1 is a data fidelity term, defined over the data points $\mathbf{u} \in \Omega$, where $\Omega \subset \mathbb{R}^m$ repre-

¹Code obtained from [20]

sents a continuous domain. The data term captures the geometric cost of a data point supporting a particular model. The assignment of data points to their respective models is encapsulated through the indicator function $\phi_l(\cdot)$:

$$\phi_l(\mathbf{u}) = \begin{cases} 1 & \text{if } \mathbf{u} \in L_l \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where the uniqueness in the label assignment is achieved by adding the constraint $\sum_{l=1}^L \phi_l(\mathbf{u}) = 1$.

In practice, some data points might not be explained by a geometric model. For this case, a special label, \emptyset , representing the outlier model is added. In this way, a constant cost, γ , can be assigned to these points. The model cost for this model is given by $\rho_\emptyset(\mathbf{u}, \phi_\emptyset(\mathbf{u})) = \gamma$.

The smoothness energy term in Equation 1 takes into account locality by promoting a homogeneous assignment of labels to neighbouring points. The $\nabla_{\mathcal{N}}$ operator calculates the gradient of the indicator function over the neighbourhood, \mathcal{N} , of a point. The function R is designed to penalise points that belong to the same neighbourhood but do not share the same model. This penalty can be evaluated using different norms, $|\cdot|_p$. For instance, the standard 4-connected lattice implemented in graph cut algorithms [2] can be obtained by implementing $\nabla_{\mathcal{N}}$ using forward differences combined with a Manhattan norm $|\cdot|_{1,1}$ ². In the continuous domain this is equivalent to penalising the anisotropic total variation over the gradient of the indicator function [11].

The parameter λ controls the trade-off between the smoothness/locality cost and the model cost. The weights, $\omega_{\mathcal{N}}$, allow for finer control of the dependencies between neighbouring data points through an appropriate metric e.g. distance.

Finally, the model count energy term in Equation 1 penalises the number of models L by adding a constant cost β per model.

2.1. Convex Relaxation Algorithm

The constraint in Equation 2 makes the problem combinatorial and NP-hard so it can only be approximately solved. A fast known relaxation approach [23] that transforms the original problem into a convex one is thus used. While this relaxation is not the tightest, it produces good results in practice. The relaxation is based on allowing $\phi_l(\mathbf{u})$ to take values in the interval $\phi_l(\mathbf{u}) \in [0, 1]$ instead of the binary set $\{0, 1\}$. As a result, for a fixed number of models L , Equation 1 with the linear equality constraint in $\phi_l(\mathbf{u})$ becomes a convex optimisation problem.

The solution adopted to solve the energy in Equation 1 depends on the selection of the functionals ρ_l and R . In this work the outlier-robust L_1 norm is opted for, as it is non-smooth its non-differentiable intrinsics exclude the use of

²In the 2D case, $|\nabla_{\mathcal{N}}\phi_l(\mathbf{u})|_{1,1} = |\nabla_x\phi_l(\mathbf{u})| + |\nabla_y\phi_l(\mathbf{u})|$

Algorithm 1: Primal-Dual Optimisation

```

Initialisation;
 $\tau, \alpha > 0, \quad \theta \in [0, 1];$ 
 $\phi^0 = \bar{\phi}^0 = \Psi^0 = 0;$ 
while  $k < N$  do
    Dual Step;
     $\Psi^{k+1} = \pi_\Psi(\Psi^k + \tau \nabla \bar{\phi}^k);$ 
    Primal Step;
     $\phi^{k+1} = \pi_\phi(\phi^k - \alpha(\rho_l(\mathbf{u}, \phi^k) + \lambda \nabla^T \Psi^{k+1}));$ 
    Relaxation step;
     $\bar{\phi}^{k+1} = \phi^{k+1} + \theta(\phi^{k+1} - \phi^k)$ 
end

```

standard optimisation techniques. However, recent achievements on continuous optimisation [3] show that non-smooth priors used in similar relaxed convex problems can be transformed into saddle point problems from which existing efficient convex optimisation techniques can be applied. In this transformation, the indicator function over the smoothing term is represented by its dual function as shown in Equation 3.

$$\int_{\Omega} |\nabla \phi_l(\mathbf{u})|_p d\Omega = \max_{\Psi_l(\mathbf{u})} \int_{\Omega} \nabla \phi_l(\mathbf{u}) \cdot \Psi_l(\mathbf{u}) d\Omega \quad (3)$$

$$s.t. |\Psi_l(\mathbf{u})|_{p^*} \leq 1 \quad (4)$$

where $\Psi_l(\mathbf{u}) : \Omega \rightarrow \mathbb{R}^2$ is known as the dual function of $\phi_l(\mathbf{u})$ and both $|\cdot|_p$ and $|\cdot|_{p^*}$ are dual norms. The optimal values of $\phi(\cdot)$ and $\Psi(\cdot)$ are obtained from the first-order primal-dual optimisation shown in Algorithm 1. As this optimisation performs per-point evaluations, it supports a parallel GPGPU implementation.

The two-stage approach for minimising the global energy given in Equation 1 is summarised as follows.

1. Data and Smoothness Energy optimisation:

Algorithm 1 shows the steps of the primal-dual optimisation that minimises the data and smoothness energy terms. Where τ, α are the step size parameters of the algorithm, with θ controlling the relaxation. The values of these are determined using the diagonal preconditioning scheme [17].

To fulfil the constraint in Equation 4, the gradient ascent step of the dual variable gets projected onto the feasible set with the projection, $\pi_\Psi(\cdot)$, defined by:

$$\pi_\Psi(\Psi) = \frac{\Psi}{\max(1, |\Psi|_{p^*})} \quad (5)$$

Similarly, the function $\pi_\phi(\cdot)$ projects the gradient descent step of the primal variable onto the simplex $\phi_l(\mathbf{u}) \in [0, 1] \mid \sum_{l=1}^L \phi_l(\mathbf{u}) = 1$ as described in [12].

Algorithm 2: CORAL Energy Minimisation

```

{Initialisation};
Propose  $s$  models;
while not converged do
  Primal-Dual Optimisation;
  Merge Models;
  Re-estimate models;
end

```

2. Model Count Optimisation:

Although it is not explicitly formulated into the primal-dual optimisation, the minimisation of the smoothness energy reduces the model count energy by enforcing a single model for spatially connected regions. In this way redundant models are reduced as they have no support in the data leading to a more compact solution.

This does not hold for non-spatially connected regions. To account for this an extra explicit step is performed that merges separated models with similar parameters. In the presence of noise, merging two models results in an increase in the data energy. If this increase is however less than β , the global energy still decreases, ensuring an optimal compact solution.

Analogous to the PEARL algorithm [7], the proposed optimisation cannot be directly applied to continuous data as the number of possible labels for a model with p parameters is \mathcal{R}^p . To reduce the search space, stochastic sampling for a finite number of models, s , should be performed. Moreover, to re-estimate models after the primal-dual optimisation is completed, a one-to-one correspondence between indicator functions and models is needed. To this end, the continuous labels are thresholded by selecting the maximum value of the indicator function at each data point. The data energy can be further reduced by re-estimating models based on the assignment given by the thresholded solution. This provides a new initialisation point to the inner primal-dual optimisation. This is continued until the global energy converges, hence the CORAL algorithm shown in Algorithm 2.

3. Experiments and Applications

In this section the proposed algorithm is employed to address the extraction of planar regions in images for two different problems: the first corresponds to the known multi-homography estimation from two views. The second is plane detection from a single RGB-D image. The motivation is driven by the fact that detection of geometric structure from images is of widespread importance to applications such as camera calibration, camera motion estimation, and surface reconstruction. We argue that the existing approaches could be prohibitive for online, real-time function-

ality. The implicit assumption in both cases is that scenes consist mostly of man-made objects commonly found in urban outdoor and indoor environments (e.g. buildings, walls, screens, desks, etc.). The image setup choice (i.e. colour images versus range data) allows us to deal with different situations. On the one hand, indoor environments usually expose texture-less surfaces. In addition, matching algorithms will be affected by low-light conditions thus motivating the use of an RGB-D sensor. On the other hand, outdoor scenes are more abundant in texture. However, the use of a depth sensor is practically limited by its maximum range. A two-view homography estimation approach for plane detection with feature correspondences would result in a better choice. Under these assumptions, the versatility of the proposed formulation is shown in the next sections with no special distinction on the distribution of data over the space.

For the run-time experimentation presented in the following sections, a machine equipped with a 12-core Intel i7-5930K 3.50GHz CPU and a GeForce GTX TITAN Black 6048MB GPU was utilised.

3.1. Two-View Multi-Homography Estimation

Our first application considers two views of a static scene with multiple planes. Given a sparse set of n pixel correspondences in homogeneous coordinates between the two views, $\mathbf{u}_i = (\mathbf{u}_i^1, \mathbf{u}_i^2) \in \mathbb{R}^2$, $i = 1 \dots n$, the homography $\mathbf{H}_{21} \in \mathbb{R}^{3 \times 3}$ establishes the mapping of pixels from the first view to the second view through an observed plane. This operation is denoted by $\mathbf{u}_i^2 = \mathbf{H}_{21} \mathbf{u}_i^1$. We aim to find the classification of pixel matches w.r.t several homographies while simultaneously rejecting outliers. A simplified version of the energy in Equation 1 for this problem is proposed in Equation 6,

$$\sum_{l=1}^L \left(\frac{1}{2} \sum_{i=1}^n (\|D(\mathbf{u}_i, \mathbf{H}_{12})\|_{\Sigma_{12}} + \|D(\mathbf{u}_i, \mathbf{H}_{21})\|_{\Sigma_{21}}) \phi_l(\mathbf{u}) + \lambda \sum_{i=1}^n \omega_{\mathcal{N}} |\nabla_{\mathcal{N}} \phi_l(\mathbf{u})|_{1,1} \right) + \beta L \tag{6}$$

The data term is chosen to account for the symmetric transfer error of the re-projection operation³. The $\nabla_{\mathcal{N}}$ operator takes into account the variation over the label function on a specified neighbourhood, \mathcal{N} . In this work, a four-connectivity neighbourhood is used although it can be easily extended to any connectivity pattern. $\omega_{\mathcal{N}}$ penalises pixels that are far away in terms of Euclidean distance. Model initialisation is carried out by applying the Direct Linear Transform (DLT) algorithm [6].

³Here, D is the Mahalanobis distance $\|D(\mathbf{u}_i, \mathbf{H}_{ab})\|_{\Sigma_{ab}} = (\mathbf{u}_i^a - \mathbf{u}_i^{a'})^T \Sigma_{ab}^{-1} (\mathbf{u}_i^a - \mathbf{u}_i^{a'})$. Σ_{ab} represents the propagated covariance matrix through the mapping induced by the corresponding homography.

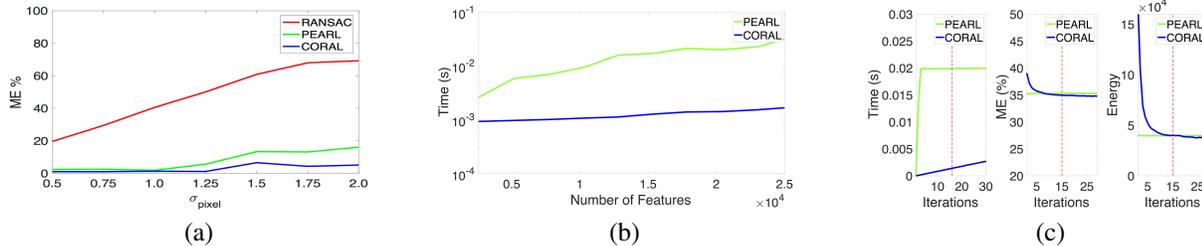


Figure 2: Results from the simulation experiment on two-view homography segmentation. In (a) the performance of RANSAC, PEARL, and CORAL are compared in an experiment with increasing sensor noise, σ_{pixel} , in the absence of outliers. The ME was then evaluated over ten different two-view pairs with CORAL outperforming both PEARL and RANSAC. In (b) with the sensor noise σ_{pixel} set to one the run-times of the two energy minimisation algorithms were compared as the number of features in the simulation was increased, CORAL yielding a run-time that is up to an order of magnitude faster. (c) shows the evolution of the run-times, ME and energy of the two energy minimisation algorithms when the number of features is twenty thousand. It can be seen that PEARL converges within fewer iterations. However, this still results in a slower run-time at the point of energy equivalence shown by the red dotted line. CORAL ultimately converges to a faster run-time (10x) as well as a lower ME and energy.

The accuracy and run-time performance of CORAL was analysed firstly through a controlled simulated experiment before extending the analysis to the extensively validated Adelaide benchmark dataset [21].

3.1.1 Simulation Experiments

To characterise the performance and robustness of CORAL with respect to existing algorithms, experiments were first run in a controlled simulation environment. The simulation environment consists of three planes, placed mutually orthogonally to each other. This configuration resembles, for instance, the end of a corridor or the corner of a building such that two walls and the ground are simultaneously observed. Uniform sampling of the planes creates the point features observable from two frames by a camera with associated noise, σ_{pixel} . Samples of the simulation environment are shown in Figure 3.

CORAL was first tested under different values of σ_{pixel} in the absence of any outliers. This was compared to a implementation of sequential RANSAC. In addition, a comparison to PEARL using the open-source implementation available in [20] was made with the Misclassification Error (ME) given in Equation 7 used as the metric to quantify the accuracy.

$$ME = \frac{\#\text{Misclassified points}}{\#\text{points}} \quad (7)$$

Results from the three approaches are shown in Figure 2. It is clear that the global energy approaches outperform RANSAC as the sensor noise increases, in fact CORAL reports better results than PEARL for the largest noise value. The higher ME reported by RANSAC in the presence of noise can be explained by analysing Figure 3, this shows the triangulated noisy points in the simulation environment for $\sigma_{pixel} = [0.5, 1.5, 1.5]$ together with the

ground-truth planes from which they were originally sampled from. Colour coding of membership to a particular model is used to show the results of the multi-homography extraction for the three stated approaches.

For $\sigma_{pixel} = 1.5$, RANSAC selects models that, though geometrically valid, are not consistent with the ground truth planes. See for example the combination of green and red points on the two vertical planes. The energy approaches are more robust to this kind of noise as the smoothness prior ensures convergence to a better solution which is consistent with the ground truth. This is shown in the last column of Figure 3.

By varying the number of point features sampled from the mutually orthogonal planes, an analysis of the run-time performance of CORAL was undertaken. With results showing CORAL’s time performance to be as much as an order of magnitude faster than PEARL as shown in Figure 2(b). While digging deeper into the simulation result when the number of features was set to twenty thousand, in Figure 2(c) shows that CORAL not only outputs a faster and more accurate ME but also converges to a slightly lower energy.

3.1.2 Benchmark Experiments

With the performance and robustness of the proposed approach verified in simulation, the proposed approach was bench-marked against the state-of-the-art on real imagery. The Adelaide dataset [21] is used in this evaluation. It consists of image pairs with extracted keypoints and manually labelled ground-truth. The performance on this dataset of different multiple model fitting approaches is available in [10]. From these results, a comparison of CORAL to T-linkage [8], J-linkage [18], RPA [9], SA-RCM [16], Grdy-RansaCov and ILP-RansaCov [10] was carried out. These results are shown in Table 1.

	J-Lnkg	T-Lnkg	RPA	SA-RCM	Grdy-RansaCov	ILP-RansaCov	Multi-H	CORAL
mean	25.50	24.66	17.20	28.30	26.85	12.91	4.40	4.2117
median	24.48	24.53	17.78	29.40	28.77	12.34	2.41	3.48

Table 1: Misclassification error for two view plane segmentation in the Adelaide dataset. Part of the results in this table are reported in [10].

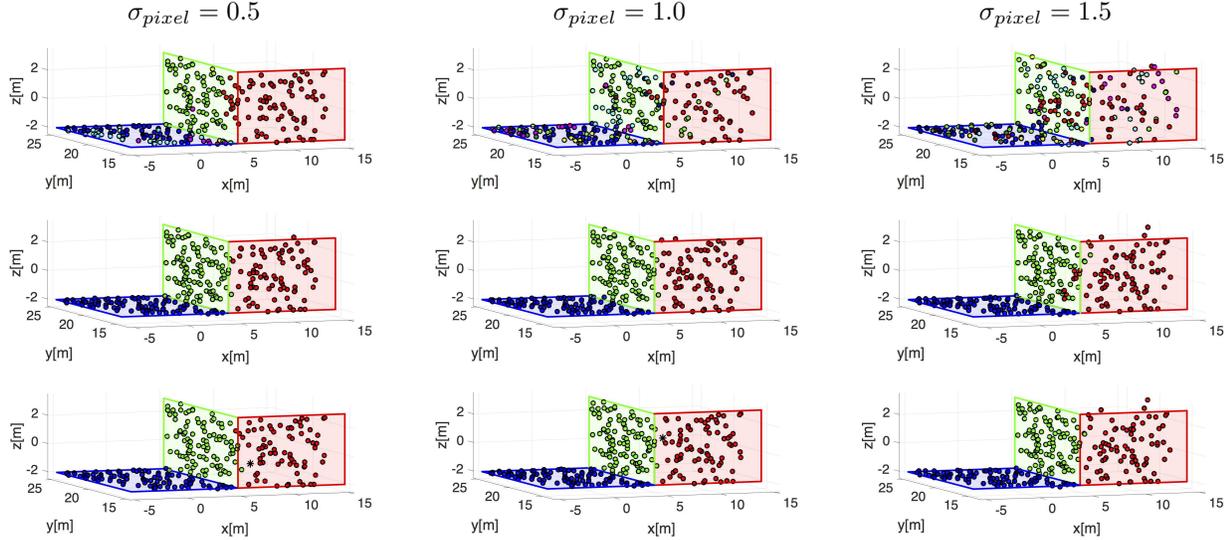


Figure 3: The triangulated positions of noisy pixels together with the ground truth planes (red, green, and blue patches) are shown under different values of σ_{pixel} . The results of the multi-homography detection are displayed by colour coding the point classification to their assigned models. RANSAC, PEARL and CORAL results are shown in the first, second and third row respectively.

From Table 1 it can be seen that the best mean performance is achieved using our algorithm. The better median performance reported by Multi-H [1] can be explained by its specialised initialisation for homographies using affine transformations that have been shown to be superior to the DLT. The proposed approach would also benefit from this specialised initialisation; however, for consistent comparison with the state-of-the-art an initialisation based on the DLT for model generation is retained.

A subset of the images with the detected homographies are shown in Figure 4, with the membership to different models colour-coded. From this it can be seen that the proposed formulation is able to accurately deal with the varied range of models present in a scene.

3.2. Plane detection with RGB-D images

The second case of structure detection involves the extraction of planes from a single RGB-D image. In this scenario, the regularity of the pixel grid is leveraged by working with an inverse depth representation on a per-pixel basis. It can be shown that if two pixels, \mathbf{u} and \mathbf{u}^* , belong to the same planar surface in 3D, their inverse depths, $\xi(\mathbf{u})$ and

$\xi(\mathbf{u}^*)$, satisfy the following equation:

$$\xi(\mathbf{u}) - \xi(\mathbf{u}^*) = \langle \mathbf{w}, \mathbf{u} - \mathbf{u}^* \rangle \quad (8)$$

where $\langle \cdot, \cdot \rangle$ represents the inner product between two vectors. $\mathbf{w} = (w_u, w_v)$ codifies the projection of the 3D plane normals into the image plane. A proof of the validity of this expression is given in the supplementary material.

In this application, the energy to be minimised is given by:

$$\sum_{l=1}^L \left(\int_{\Omega} (\|\xi(\mathbf{u}) - \xi(\mathbf{u}^*) - \langle \mathbf{w}, \mathbf{u} - \mathbf{u}^* \rangle\|_{\sigma_{\xi}}) \phi_i(\mathbf{u}) d\Omega + \lambda \int_{\Omega} \omega_{\mathcal{N}_r}^{\alpha} |\nabla_{\mathcal{N}} \phi(\mathbf{u})|_{1,2} d\Omega \right) + \beta L \quad (9)$$

where $\omega_{\mathcal{N}}^{\alpha} = e^{-\|\nabla_{\mathcal{N}} I(\mathbf{u})\|_{\alpha}}$ and I represents the intensity of the corresponding RGB image. These weights serve as a measure of edginess that can be controlled with the parameter α thus aiding to preserve sharp discontinuities between objects. Note that the Mahalanobis distance on the inverse depth was used as the data term. The remaining elements in this equation are similar to the general-purpose energy proposed in Equation 1.

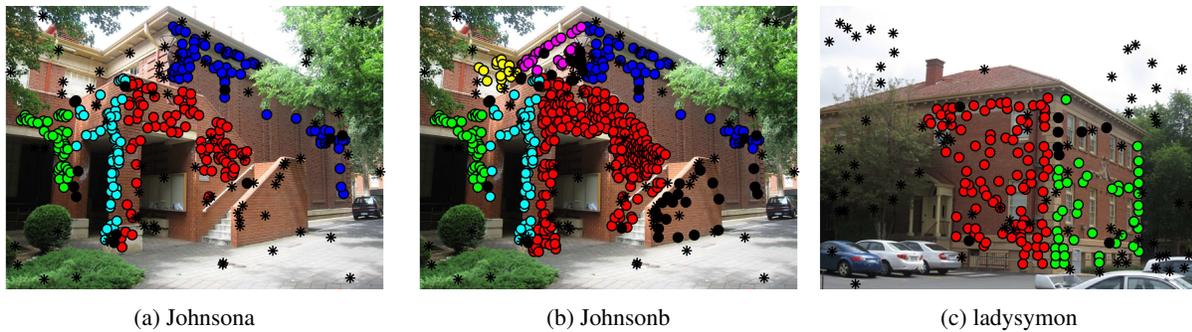


Figure 4: Sample of images from the Adelaide dataset, with the membership to the different homographies given by CORAL indicated by the different colours. Crosses are used to signify points that belong to the uniform outlier label, \emptyset , while points that are mislabelled are shown in black through all figures.

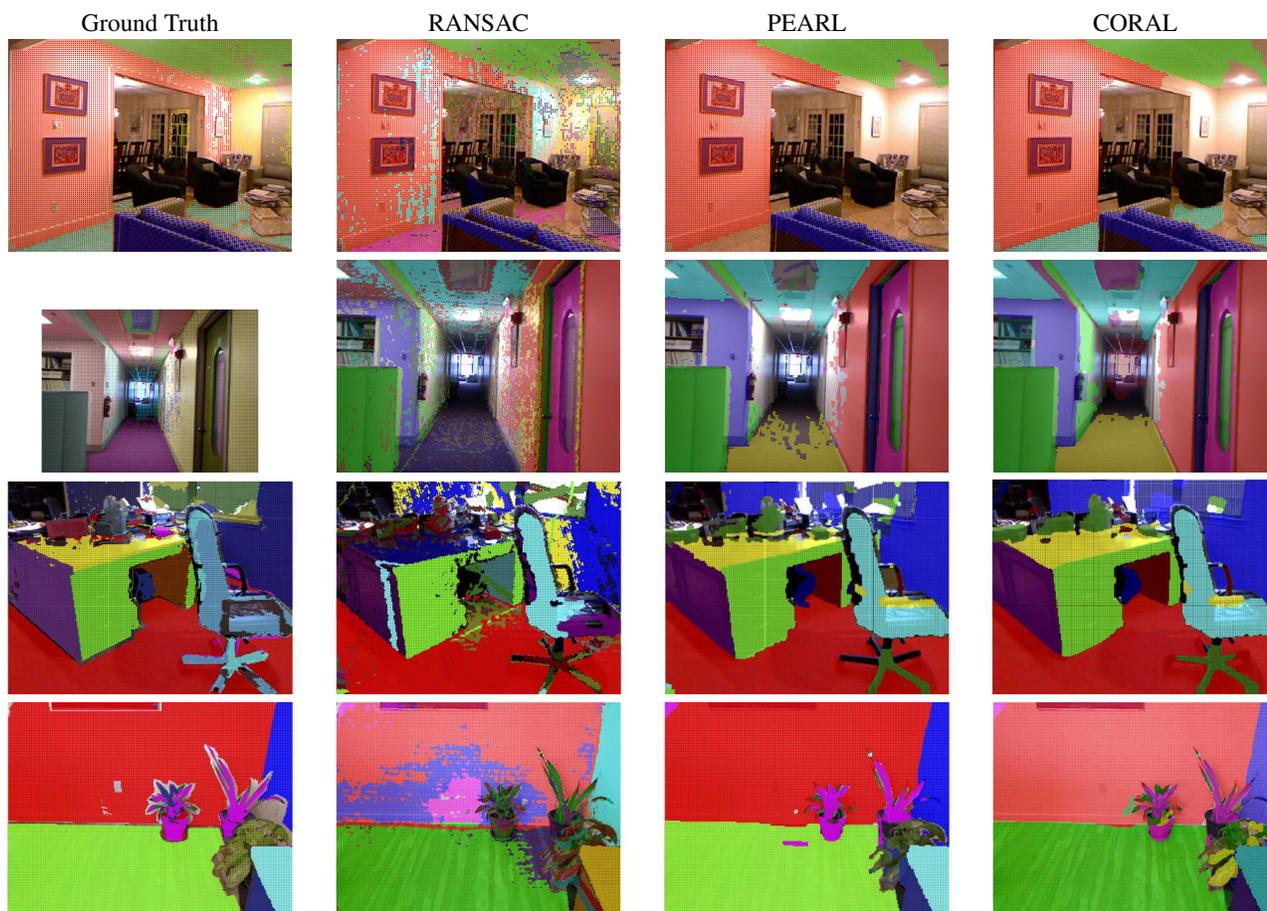


Figure 5: Plane segmentation on a sample of RGB-D images from the NYU-Depth dataset. with the Ground Truth, RANSAC, PEARL, and CORAL results presented for each image. Pixel membership to a plane model is shown by superposing on a pixel a colour-coded point that assigns it to a specific plane model.

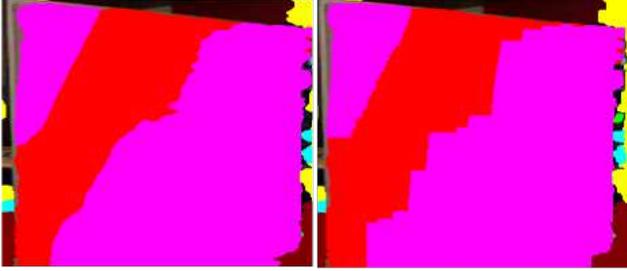


Figure 6: Intermediate optimisation result showcasing the difference between the $|\cdot|_{1,2}$ norm (left) and the Manhattan norm (right) used by PEARL on the smoothing optimisation. The Manhattan norm preserves edges aligned with the axes leading to a stair-case effect when dealing with dense data, whereas our flexible formulation allows for the selection of the more appropriate $|\cdot|_{1,2}$ norm alleviating this effect in the smoother result (left).

To evaluate the performance of CORAL for plane detection using RGB-D images, the NYU-depth dataset [13] was used. This contains 1449 tuples of RGB, depth, and labelled images for multiple instances of objects each of resolution 420 x 560. For the evaluation a subset of 232 images containing scenes where either the walls, ceilings, desks, floor or a combination of all of these surfaces are observed were selected. This ensures that the selection contains images with significant planar regions. To obtain a ground truth, the labels provided were employed and planes fitted to individual instances of objects in the scene with a minimum number of inliers. In addition, planes with similar planar parameters were merged to reduce redundancy.

Four samples of the dataset are shown in Figure 5. Each column represents the ground truth, RANSAC, PEARL and CORAL solutions with a colour-code representing the different labels. The results show that the energy based methods produce more consistent plane models aligned with the expected ground truth. CORAL solutions show greater quality compared to PEARL solutions, in particular considering the first two images. In such cases, CORAL is able to identify more plane models than PEARL.

Additionally the Manhattan norm used to penalise smoothness in PEARL was compared to the $|\cdot|_{1,2}$ norm used in this experiment. Focusing on an intermediate optimisation result in Figure 6, it can be seen that the Manhattan norm preserves edges aligned with the axes leading to a stair-case effect when dealing with dense data, the change in the norm allows for a much smoother result as shown.

Table 2 condenses the ME results for all three methods. As observed, CORAL outperforms PEARL in three of the four instances and over the whole test set. To quantify performance over the test set, an optimal value of λ was trained using a quarter of the images in the subset.

	RANSAC	PEARL	CORAL
Image 1	22.96	16.24	13.95
Image 2	28.70	20.59	17.12
Image 3	36.60	26.30	25.30
Image 4	15.72	7.77	7.83
Test set	29.38	23.04	18.99

Table 2: Mislabelling error for RGB-D segmentation.

Run-times for the two energy optimisation algorithms were evaluated and shown in Table 3. CORAL runs on average around twenty times faster than PEARL on this task, even as the number of models increases.

	$L=4$	$L=8$	$L=16$
PEARL	0.163843	0.659035	1.42582
CORAL	0.007628	0.026909	0.071628

Table 3: Time taken in seconds for the energy minimisation for PEARL and CORAL evaluated and averaged over the images shown in Figure 5. CORAL implementation was on average about twenty times faster even as the number of models, L , varied.

4. Discussion and Conclusions

In this work CORAL, a global energy minimisation algorithm for geometric multi-model fitting, has been introduced. Our general solution uses a convex relaxation for model assignment before leveraging advanced optimisation techniques in the continuous domain. Our approach inherits all the benefits described in the available survey presented in [14]. The advantages of CORAL over discrete solutions analysed in previous sections are mainly driven by its potential for parallel implementation compared to state-of-the-art methods that require sequential evaluations of labels. CORAL intrinsically boosts run-time performance by simultaneously handling per-point evaluations over the labels as demonstrated in the previous sections. This makes our approach more suitable for geometric model extraction in applications with real-time performance constraints.

Our formulation also allows for flexibility on the norms applied, which we characterise through the structure detection problem from images in two different scenarios. All without degradation of performance, as results that outperform or are at least equivalent to the best performances achieved using state-of-the-art methods are reported.

In summary, CORAL incorporates powerful optimisation machinery into the solution of geometric multi-model fitting. Offering an algorithm that is simultaneously able to robustly extract accurate models in the presence of contamination and offer improved time performance guarantees over the state-of-the-art.

References

- [1] D. Barath, L. Hajder, and J. Matas. Multi-h: Efficient recovery of tangent planes in stereo images. *BMVC, 27th British Machine Vision Conference.*, 28:32, 2016. 6
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001. 1, 3
- [3] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. 2, 3
- [4] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *International journal of computer vision*, 96(1):1–27, 2012. 1
- [5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 1
- [6] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 4
- [7] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *International journal of computer vision*, 97(2):123–147, 2012. 1, 2, 4
- [8] L. Magri and A. Fusiello. T-linkage: A continuous relaxation of j-linkage for multi-model fitting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3954–3961, 2014. 1, 5
- [9] L. Magri and A. Fusiello. Robust multiple model fitting with preference analysis and low-rank approximation. In *BMVC*, pages 20–1, 2015. 5
- [10] L. Magri and A. Fusiello. Multiple model fitting as a set coverage problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3318–3326, 2016. 1, 5, 6
- [11] S. A. J. Marsden, L. S. S. Wiggins, L. Glass, R. Kohn, and S. Sastry. *Interdisciplinary Applied Mathematics*. Springer, 1993. 3
- [12] C. Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of R^n . *Journal of Optimization Theory and Applications*, 50(1):195–200, 1986. 3
- [13] P. K. Nathan Silberman, Derek Hoiem and R. Fergus. Indoor segmentation and support inference from rgb-d images. In *ECCV*, 2012. 8
- [14] C. Nieuwenhuis, E. Toeppe, and D. Cremers. A survey and comparison of discrete and continuous multi-label optimization approaches for the potts model. *Int. Journal of Computer Vision*, 104(3):223–240, 2013. 2, 8
- [15] T. T. Pham, T.-J. Chin, K. Schindler, and D. Suter. Interacting geometric priors for robust multimodel fitting. *IEEE Transactions on Image Processing*, 23(10):4601–4610, 2014. 1
- [16] T. T. Pham, T.-J. Chin, J. Yu, and D. Suter. The random cluster model for robust geometric fitting. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1658–1671, 2014. 5
- [17] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1762–1769. IEEE, 2011. 3
- [18] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *European conference on computer vision*, pages 537–547. Springer, 2008. 1, 5
- [19] P. H. Torr. Geometric motion segmentation and model selection. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 356(1740):1321–1340, 1998. 1
- [20] O. Veksler and A. Delong. Multi-label optimization. <http://vision.csd.uwo.ca/code>, 2016. [Online; accessed 19-November-2016]. 2, 5
- [21] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter. Dynamic and hierarchical multi-structure geometric model fitting. In *2011 International Conference on Computer Vision*, pages 1044–1051. IEEE, 2011. 5
- [22] J. Yu, T.-J. Chin, and D. Suter. A global optimization approach to robust multi-model fitting. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2041–2048. IEEE, 2011. 1
- [23] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *VMV*, pages 243–252, 2008. 3