International Journal of Computer Vision, vol. 97, no. 2, pp. 123-147, April 2012

Energy-based Geometric Multi-Model Fitting

Hossam Isack · Yuri Boykov

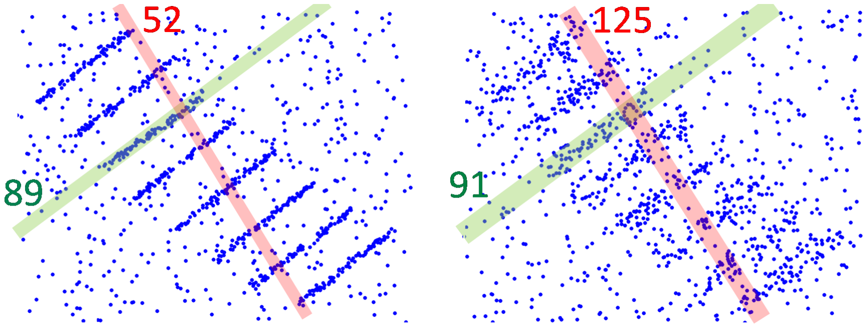
Received: June 10, 2010 / Accepted: June 7, 2011/ Published online: July 12, 2011

Abstract Geometric model fitting is a typical chicken-&-egg problem: data points should be clustered based on geometric proximity to models whose unknown parameters must be estimated at the same time. Most existing methods, including generalizations of RANSAC, greedily search for models with most inliers (within a threshold) ignoring overall classification of points. We formulate geometric multi-model fitting as an optimal labeling problem with a global energy function balancing geometric errors and *regularity* of inlier clusters. Regularization based on spatial coherence (on some near-neighbor graph) and/or label costs is NP hard. Standard combinatorial algorithms with guaranteed approximation bounds (e.g. α-expansion) can minimize such regularization energies over a finite set of labels, but they are not directly applicable to a continuum of labels, e.g. $\mathcal{R}^2$ in line fitting. Our proposed approach (PEARL) combines model sampling from data points as in RANSAC with iterative re-estimation of inliers and models parameters based on a global regularization functional. This technique efficiently explores the continuum of labels in the context of energy minimization. In practice, PEARL converges to a good quality local minimum of the energy automatically selecting a small number of models that best explain the whole data set. Our tests demonstrate that our energy-based approach significantly improves the current state of the art in geometric model fitting currently dominated by various greedy generalizations of RANSAC.

Keywords geometric models · α-expansion · graph cuts · sampling

H. Isack
Computer Science Department, University of Western Ontario
E-mail: habdelka@csd.uwo.ca

Y. Boykov
Computer Science Department, University of Western Ontario
E-mail: yuri@csd.uwo.ca



(a) low noise (b) high noise

Fig. 1 Blue dots are data points supporting 8 lines. In multi-model cases, detecting models by maximizing the number of inliers may work for low levels of noise (a). Higher noise levels require larger thresholds to detect inliers (b), but then, some random model (red) may have far more inliers than the true model (green). The integers show the number of model inliers for selected thresholds. Simplistic greedy selection of models with the largest number of inliers would fail in (b). This example illustrates a general problem (see Figs.7-9) for many multi-model fitting approaches greedily selecting one model at a time while ignoring the overall solution. It also motivates our global energy approach optimizing the quality of the whole solution.

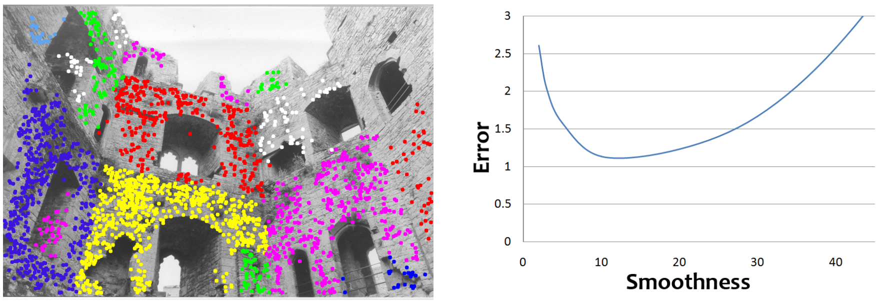
1 Introduction

We study a general case of geometric multi-model fitting problem where data is a mixture of outliers with points supporting unspecified number of models of some known type[1]. The majority of existing algorithms treat inlier classification and parameter estimation as isolated subproblems. Typically, each model is selected greedily by maximizing inliers within some fixed threshold. Popularized by RANSAC [14], this approach works well when data supports a single model, but we argue that it is fundamentally flawed in multi-model cases, see Fig.1.

RANSAC [14] is a well-known robust method for dealing with large number of outliers when data supports only one model. The main idea is to generate a number of model proposals by randomly sampling data points and then select one model with the largest set of inliers (a.k.a. consensus set) with respect to some fixed threshold. Many publications [30,34,40] proposed various generalizations of RANSAC for multi-model fitting. For example, [30,34] run RANSAC sequentially. Each iteration of these methods selects one randomly sampled model maximizing either the number of inliers or some similar threshold-based measure. Thereby identified model's inliers are removed from the set of data points before the next iteration looks for the next model. Other methods rely on different forms of greedy clustering, e.g. J-linkage [27], explicitly or implicitly maximizing the number of inliers within given threshold. One can also apply Hough transform to formulate multi-model fitting as a clustering problem in the space of model parameters and use *mean-shift* [9] to identify the modes in this Hough space. It is easy to see that this approach also greedily maximizes the number of inliers.

We argue that, in general, greedy selection of one model while ignoring the overall solution is a flawed approach to multi-model cases. Figure 1 shows a simple example illustrating the typical problem in the context of greedy inlier maximization: if the level of noise is increased, some random model can have a larger number of inliers than the true models. This

[1] For simplicity, we assume (parametric) models of same type. This is not essential.



(a) fitting homographies (stereo)

(b) estimation accuracy vs. λ

Fig. 2 Motivating spatial regularization in geometric model fitting. In many vision problems combining geometric errors and spatial coherence terms in energy (3) can be justified *generatively* because clusters of inliers are generated by regular objects (a). More over, spatial regularization can also be justified *discriminatively*. Plot (b) shows an average deviation from the ground truth for optimal models obtained in 100 randomly generated line-fitting tests as in Fig.4d. Each point in this plot corresponds to some fixed smoothness parameter λ in (3). Clearly, spatial coherence term significantly reduces estimation errors for some $\lambda > 0$.

also explains our results in Section 3 (Figs.7-9) demonstrating that many existing greedy methods work only on examples with low levels of noise and clutter.

1.1 Towards energy optimization

This paper argues that geometric multi-model fitting is better formulated as an optimization problem with a global energy functional describing the quality of the overall solution. An energy function sets some specific “goodness” criterion for different solutions and the sought optima can be seen as “objectively” the best solution with respect to this criterion. There are many problems in computer vision (e.g. segmentation, optical flows, stereo) routinely solved as optimization problems. Yet, we know only relatively few examples in vision [31,30,2,20] where some specific geometric multi-model fitting problems were approached using an energy-based formulation. Perhaps, limitations of these methods (see Sec.1.2) explain why many researchers in the community still prefer greedy heuristics. Our goal is a general energy-based framework for geometric multi-model fitting problems with efficient algorithms and wide applicability in computer vision.

There are several limitations for using standard energy-based methods for mixture models, such as EM or K-means, in geometric multi-model fitting problems in vision. In general, these methods may not be robust to outliers and noise. They are only guaranteed to find a local minimum and are known for sensitivity to initialization, e.g. see [30] and a detailed discussion in [13] (Sec.3.2). Models should be represented as probability distributions in EM, which is not always straightforward in geometric problems in vision. The standard versions of EM and K-means do not address spatial regularity explicitly. There are extensions of EM regularizing the number of models, e.g. using Dirichlet prior [3]. In the context of Gaussian mixture models (GMM), [13] combine Dirichlet sparsity prior with a large number of initial proposals, which is shown to better avoid local minima. In practice, the algorithm in [13] changes the energy functional when removing each redundant weak model. To achieve sufficiently strong model pruning effect, one should also use improper negative values of Dirichlet distribution parameter, see [13] and [11] (Fig.12). Both K-means and EM are more common in problems with a fixed number of models. For example, EM framework in MLESAC [29] is fixed to 2 models (inliers/outliers), and the method in [15] estimates a known number of motions in cases with relatively low noise[2]. Soft assignment of inliers is an advantage of EM algorithm in solving general mixture problems (e.g. GMM) where models can spatially overlap, but this may not be useful in geometric problems in vision where models typically have distinct spacial support, see Fig.2a. Standard K-means is also known to have a bias towards equally dividing the points among the models, see [11] (Fig.9).

In order to motivate our general approach, we first demonstrate some energy-based interpretation of the basic RANSAC algorithm [14]. This interpretation is limited to a simple case when data supports only one model (e.g. one line). The main goal of RANSAC is to find parameters L of the model with the largest number of inliers within some threshold T. This can be represented as minimization of energy

$$E(L) = \sum_p ||p - L||$$

where

$$||p - L|| = \begin{cases} 0 \text{ if } \ dist(p, L) < T \\ 1 \text{ otherwise} \end{cases}$$

and $dist(p, L)$ is Euclidean distance between data point p and the nearest point on model L. In this paper $||p - L||$ will generally denote an arbitrary error measure for point p and geometric model L. RANSAC's energy $E(L)$ counts inliers for L using 0-1 measure $||p-L||$ above. Note that the standard RANSAC algorithm is a heuristic for maximizing the number of inliers, but in some cases it is possible to find the global optimum [24,39]. Standard RANSAC also includes an additional step refining model parameters L by minimizing the sum of squared errors for inliers. Thus, a more principled optimization-based formulation of RANSAC leads to MSAC energy [29] using truncated Euclidean errors

$$||p - L|| = \begin{cases} dist^2(p, L) \text{ if } \ dist^2(p, L) < T \\ \quad T \qquad \text{otherwise.} \end{cases}$$

Note that RANSAC or MSAC optimize $E(L)$ only over model parameters L and inliers are identified *implicitly* from threshold T in the corresponding error measures $||p - L||$.

Now assume that data supports multiple models. If the number of models is known (say K) it could be possible to formulate geometric model-fitting as optimization of energy $E(L_1, L_2, ...L_K)$ over K model parameters. As in the earlier example with a single model, this approach needs some implicit assignment of inliers to models. In multi-model case, however, this could be non-trivial. As shown in Fig.3b, simple thresholding may assign a point to several models. Interestingly, the EM framework for mixture models [3,29,15] corresponds to energy $E(L_1, L_2, ...L_K)$. EM uses implicit "soft" classification of inliers computed in an intermediate optimization step. Even though the standard version of EM algorithm needs the number of models to be known, there are many generalizations of EM that could be worth studying in the context of geometric applications in vision. However, we prefer to focus on a fairly different energy formulation based on explicit "hard" classification of inliers. As shown in Fig.2a, in many problems in computer vision geometric models have non-overlapping spatial support, which better corresponds to hard assignment of inliers.

We formulate geometric multi-model fitting as an optimal labeling problem. Consider the general case when the data supports some unknown number of models. In principle, in

[2] From personal communications with A. Gruber.

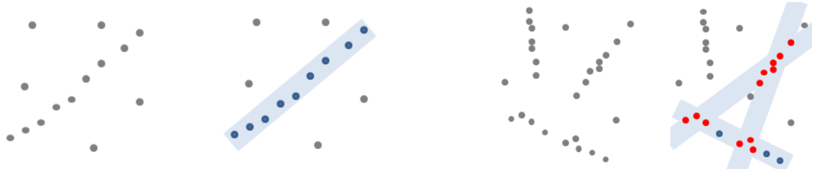


Fig. 3 Examples of inlier classification from thresholding. If data is known to support one model (a) then thresholding identifies inliers (blue) for any model without ambiguity. In case of 3 models (b), simple thresholding may not disambiguate inliers (red) between the models.

this case each data point p can have a separate model L_p. Model fitting could be formulated as minimization of energy $E(\mathbf{L})$ over labeling $\mathbf{L} = \{L_p | p \in P\}$ of points in data set P. Since labeling $\mathbf{L}$ explicitly assigns models to data points, inliers support $\{p | L_p = L\}$ for any specific model L does not have to be implicitly deduced from some threshold.

If the goal is only to minimize the fitting errors, as in our one-model example, then

$$E(\mathbf{L}) = \sum_p ||p - L_p||. \tag{1}$$

where $||p - L||$ could be an arbitrary error measure. Obviously, this functional would not work well as the globally optimal solution will independently fit some model L_p to each point p. This corresponds to overfitting: every point is assigned some perfectly fit model and there are no outliers. It is clear that model fitting errors (1) must be combined with some energy term regularizing the labeling. One special "outlier" label could be added as well.

One form of regularization for (1) could be to fix the number of allowed distinct models/labels. Then, energy (1) corresponds to the standard K-means algorithm. This approach, however, does not work if the exact number of models is not known a priori. Recently, Li [20] proposed a soft form of regularization for the number of models by combining geometric errors with the label count penalty

$$E(\mathbf{L}) = \sum_p ||p - L_p|| + \beta \cdot |\mathcal{L}_{\mathbf{L}}| \tag{2}$$

where $\mathcal{L}_{\mathbf{L}}$ is the set of distinct models (labels) assigned to points by labeling $\mathbf{L}$. Ten year earlier Torr [30] suggested even more general form of such regularization where each distinct model (label) gets a penalty defined by the model's complexity instead of some fixed constant β. This approach allows to fit models of different types. In general, geometric model-fitting using energies like (2) is a very interesting idea, but specific algorithms for minimizing such energies proposed in [30] and [20] are fairly limited, see Sec.1.2. We also argue that spatial regularity of inliers is required in many typical vision problems, see Fig.2a.

This paper proposes two specific general forms of regularization in the context of geometric model fitting. In particular, we consider spatial regularization (3)

$$E(\mathbf{L}) = \sum_p ||p - L_p|| + \lambda \cdot \sum_{(p,q)\in\mathcal{N}} w_{pq} \cdot \delta(L_p \neq L_q),$$

where $\mathcal{N}$ is some neighborhood (e.g. edges on some near-neighbor graph), and a more general combination of spatial smoothness with label counts (5)

$$E(\mathbf{L}) = \sum_p ||p - L_p|| + \lambda \cdot \sum_{(p,q)\in\mathcal{N}} w_{pq} \cdot \delta(L_p \neq L_q) + \beta \cdot |\mathcal{L}_{\mathbf{L}}|.$$

While spatial regularization is very common in vision in general, it is not common in geometric model fitting. In part, this could be explained by the fact that spatial coherence is hard to justify generatively in applications where data points are i.i.d. samples. But in computer vision, see Figure 2a, one can defend generative models of spatial regularity. Figure 2b also suggests that spatial regularization may work discriminatively even for i.i.d. data[3].

Summary of contributions This work demonstrates significance of efficient combinatorial optimization methods for a wide range of geometric applications in computer vision. Surprisingly, such methods are overlooked in geometric model fitting, even though they are very common in segmentation, dense stereo, optical flows, and other problems. We believe that we contribute a new approach and significant algorithmic ideas specific to general geometric multi-model fitting. We see our two main contributions as follows:

- We propose a general practical energy-based framework for geometric model fitting particularly suitable for a wide range of applications in vision. To the best of our knowledge, energies (3) and (5) were not used for general geometric model fitting problems in the past. We demonstrate conceptual advantages and significant practical improvements over the state-of-the-art methods on a large number of generic applications in vision (line/plane fitting, homography estimation, rigid motion detection). In particular, we argue against typical greedy heuristics currently dominating in geometric model fitting and hope that our work would encourage a wider use of energy optimization methods well known in other areas of computer vision.
- Energies (3) and (5) can be addressed by existing powerful combinatorial optimization techniques with guaranteed optimality bounds (e.g. α-expansion [5]) only in cases of finite set of labels. This limits their use for geometric model fitting where the space of model parameters is a continuum, e.g. $\mathcal{R}^2$ in line fitting. We propose a practical method (PEARL) for efficiently exploring the continuum of labels (model parameters) in the context of energy-based geometric model fitting. PEARL may also be juxtaposed with the contribution in [13] where a multitude of initial proposals combined with a sparsity prior are shown to reduce sensitivity of EM to local minima. Likewise, we show that combining a large number of initial random proposals with combinatorial algorithms produces robust solutions for our discrete model fitting functionals.

Our general approach alleviates dependence of many previous geometric model fitting methods on thresholding. The proposed methods for optimizing energies (3) and (5) work quite differently from greedy selection of models by the largest number of inliers. Our approach is robust to high levels of noise and clutter. It automatically computes on optimal set of labels/models with a good fit to data points.

In order to apply standard discrete optimization algorithms to energies (3) and (5), we generate a large number of proposed labels (models) by sampling data points as in RANSAC. The goal of this step is to prune the search space (the continuum) of model parameters. As in RANSAC, the number of sampled models should be sufficiently large to guarantee with some level of confidence that at least one sample was generated from inliers for each true model. Such finite pool of labels is likely to contain good model candidates. However, in contrast to RANSAC-style methods we rely on optimization of a global energy functional to select some small subset of models from this large (but finite) pool of

[3] One can not expect spatial regularization to work well for i.i.d. data, in general. Line fitting examples in Section 2 are a special case where it does work for i.i.d. points. We use these line fitting examples only to illustrate the basic operations of our algorithm. The primary target of our model fitting approach are applications in vision (Sec.3) where spatial coherence is well-founded.

proposals. Exploration of the continuum of labels (model parameters) is further significantly improved by iterating inlier segmentation for a finite set of labels and re-estimation of these labels (model parameters) from their inliers. Both steps minimize the same energy $E(\mathbf{L})$ and correspond to a coordinate descent converging to a local minimum of the energy. Such iterative refinement of model parameters and inlier classification allows to generate better solutions from a smaller set of initial samples even in single model fitting, see Fig.6.

1.2 Related work

Our work proposes, justifies, and validates a wide class of regularization energies and a powerful iterative optimization technique as a general framework for geometric multi-model fitting particularly suitable in vision. Other geometric model fitting works have used separate elements of our approach such as RANSAC-style random sampling [30,20] or EM-style iterations [2], but none have combined them in a single optimization framework. We also use a more general form of regularization (5) than any earlier geometric fitting methods. Our experiments show that our general energy-based approach works better than many state-of-the-art algorithms in this area. In other settings (segmentation, stereo) some elements of our framework have been used in various application-specific ways [38,2,25,37].

Probably the earliest efforts to formulate an energy-based framework for geometric model fitting in vision is due to [31]. They optimize a likelihood function over binary indicator variables associated with a multitude of proposed models under the uniqueness constraint: each data point could be an inlier for at most one model. The corresponding integer programming problem is solved with a generic branch-and-bound solver.

Another early paper by Torr [30] carefully justifies a version of model-fitting energy like (2) from an *information criterion*. The specific optimization technique used in [30] was EM initialized by a few models selected via *sequential* RANSAC[4]. As pointed out in [30], the solution generated by EM strongly depends on the quality of the initial models. Figures 1 and 9 show that greedy procedures like sequential RANSAC may be non-robust.

Variants of sequential RANSAC are also commonly used as a preprocessing step for dense MRF-based segmentation of image pixels with geometric labels. For example, Wills et al. [35] address the problem of finding large constant motions in optical flow imagery as follows. First, they use an extension of sequential RANSAC to detect a few rigid motions from pairs of matched discrete features. Then, they assign image pixels to these motion layers based on color consistency and Potts regularization using α-expansion algorithm [5].

Some related MRF-based formulations used convergent iterative re-estimation of geometric models. Birchfield & Tomasi [2] estimate affine geometric models in a way specific to dense narrow base-line stereo. They combine photoconsistency of pixels with spatial regularization on a grid. Unlike [35], their initial geometric models are estimated from a disparity map generated by another stereo algorithm. The most noticeable overlap of our approach and [2] is iterative use of α-expansion and model re-estimation steps. After [2] and [25] such EM-style optimization became fairly common for different problems in vision. In contrast to [2], however, our framework is suitable for a significantly more general set of geometric problems. For example, instead of photoconsistency we optimize geometric errors and combine them with more general forms of regularization, e.g. (5). Our method is more concerned with fitting to sparse data. Finally, we do not need to run other algorithms to initialize. Our experiments in Sec.3.1 show noticeable improvement of accuracy on examples from [2].

[4] The actual term was introduced in [34] a few years later.

Zabih & Kolmogorov [37] also use iterative optimization as in [2,25] specifically in the context of image segmentation. They use standard spatial regularization like in (3) to cluster image pixels into spatially coherent segments with automatically estimated color models. The color models are initialized in an application specific way. In contrast, we work with a very different problem of geometric model fitting studying more general label cost functional (5). In fact, our recent work [11] with additional coauthors shows that energy (5) may significantly improve image segmentation results. Minimum description length (MDL) interpretation of (5) is well known in segmentation literature for some time [19,38].

Schindler and Suter [26] proposed a related optimization method for geometric model fitting based on an approximation of label cost functional (2) without spatial regularity term. Like in our work, the goal in [26] is to detect geometric models using some global energy optimization instead of greedy heuristics like sequential RANSAC. They formulate a quadratic pseudo-boolean optimization (QPBO) problem over indicator variables defined for models proposed from data. These binary variables are similar to those in [31]. However, instead of enforcing the exact *uniqueness constraint* [31] (see above), the energy formulation in [26] makes an assumption that each data point is an inlier for no more than 2 proposed models. This strong assumption requires a pre-processing *data analysis* step that prunes the set of initially sampled models leaving only a relatively small set of good candidates. The actual optimization over binary indicator variables for such candidate models is performed using standard *Taboo-search* algorithm. Iterative re-estimation of model parameters seems impossible is this framework because assignment of models to data points is done implicitly.

The paper by Li [20] is probably the most closely related prior work. Similar to [30, 26], it formulates general geometric model fitting functional (2) and studies it in the context of rigid motion estimation, which we also consider as one of the applications in Sec.3.3. Instead of the greedy approach of [30], [20] uses LP relaxation of (2). This could be slow. To speed up the method, [20] uses several application specific heuristics to significantly prune the set of proposed models. More importantly, [20] does not guarantee any optimality of the discrete solution obtained after rounding and the quality of such optimization could be an issue. These problems do not allow [20] to use EM-style iterative optimization that, in our experience, can significantly improve model fitting results. A better optimization of energy (2) with some optimality guarantees is discussed in [11].

In this paper, we argue that (5) is generally a better energy for geometric model fitting problems in vision. We found that per-label regularization term proposed in [30,20] is a practically useful addition to standard spatial regularization (3). Fig.10 shows one illustrative example where penalty for using each distinct label encourages the merging of isolated clusters supporting nearly identical models. Similarly, the results on real motion detection sequences in Figure 21 fail to merge spatially isolated background clusters into one motion if label counts are not a part of the energy. To optimize the third term in energy (5) one can use a simple and fast merging step in combination with standard α-expansion optimizing the first two terms in the energy. This merging heuristic is discussed at the end of Sec.2. Alternatively, [11] provides an extension of α-expansion algorithm that automatically accounts for the third term in (5) incorporating it into each expansion step as a high-order clique. The technical details of such extension is a subject of [11]. The main focus of this paper is to demonstrate that a general algorithmically solid optimization approach to geometric model fitting with either (3) or (5) is a significantly better alternative to greedy generalizations of RANSAC-style thresholding currently dominant in geometric problems in vision.

The structure of our paper Section 2 presents our general method for fitting multiple models to sparse data points. For simplicity, most of the details are explained in the context of energy

(3). Energy (5) is introduced in the end of the section. Section 3 provides evaluation on real data in narrow-base stereo, wide-base stereo/reconstruction, and rigid motion estimation.

2 Our approach (PEARL)

This section described our algorithm for geometric model fitting in detail. For simplicity, the main ideas are illustrated in the context of synthetic multi-line fitting examples. Section 3 validates our approach for estimating affine transformations, homographies, and rigid motion models in the context of computer vision.

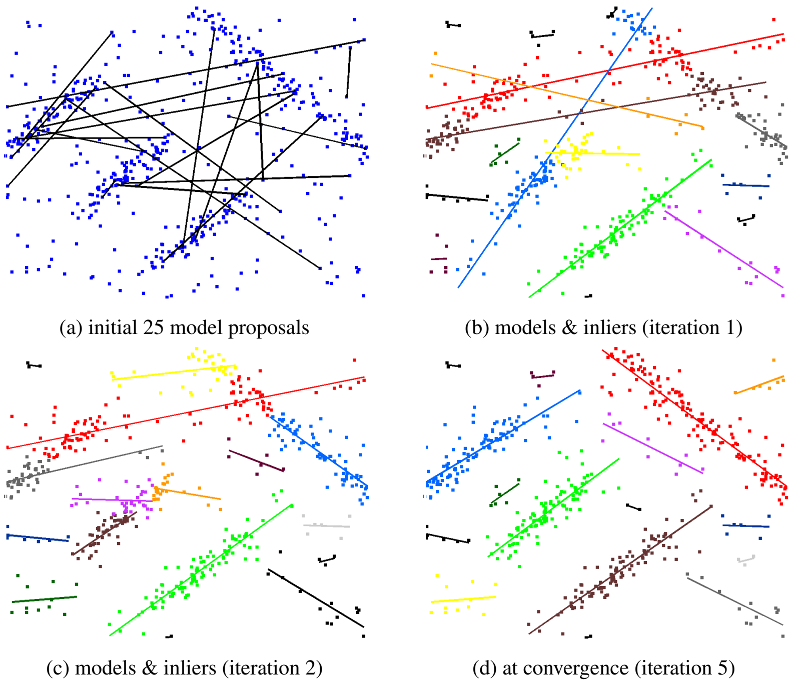
We use regularization labeling framework to assign models to data points. Regularization energy can combine geometric fit errors with spatial smoothness term (3) and a label count penalty (5). As long as the number of labels is finite (≤ 10000 or so), such approach can be handled by graph-based optimization methods, e.g. α-expansion [5,11]. In our case the labels are models described by n real-valued parameters. Therefore, we should find a practical way to restrict the huge search space of labels $\mathcal{R}^n$. The first step is to *propose* a finite set of plausible models (labels). In the next step each label is *expanded* to estimate its spatial support, or to classify inliers. Once the inliers are fixed, the models (labels) can be *re-estimated* by minimizing the geometric errors - the first term in our energy functionals. As both *expand* and *re-estimate* are guaranteed to decrease the energy, one can converge to a local minimum by iterating over these two steps. One can also iterate *propose* steps either by further sampling the data points, or by generating some new proposed labels (models) from the currently supported models, e.g. by merging them. Below we provide more details about our model-fitting algorithm in the context of energy (3). Most of them also apply to energy (5) introduced in the end of the section.

2.1 *Propose* initial labels $\mathcal{L}_0$

First, our method uses random sampling of data points to *propose* an initial finite set of models $\mathcal{L}_0 \subset \mathcal{R}^n$, where n is the number of parameters describing each model ($n = 2$ for lines and $n = 6$ for affine models). The idea of generating models by sampling the data points is borrowed from RANSAC [14]. The required number of initial models $|\mathcal{L}_0|$ is one of the parameters of RANSAC-based methods. It depends on the number of data points, the number of outliers, the number of estimated models, the minimum number of points requited to estimate each model, and on desired level of confidence. The exact analysis for the case of estimating a single model is given in [14]. Its adaptation to multi-model case is in [34,40]. The number of initial models $|\mathcal{L}_0|$ for PEARL is analyzed in [18] (see one example in Sec.3.3.1). Our experiments suggest that in practice PEARL often needs far fewer samples than the theoretical estimate due to converging iterations that significantly improve probability of an accurate model reconstruction from rough initial guesses.

2.2 Energy formulation

Once initial finite set of proposed models $\mathcal{L}_0 \subset \mathcal{R}^n$ is known (see Fig.4(a)), we can *expand* the models to estimate their spatial support. We use MRF-based regularization framework and α-expansion optimization [5] to assign models to data points. The set of current models in $\mathcal{L}_0$ is interpreted as a set of current labels. Assume that P is a set of data points and that



(a) initial 25 model proposals

(b) models & inliers (iteration 1)

(c) models & inliers (iteration 2)

(d) at convergence (iteration 5)

Fig. 4 Illustration of PEARL's iterations. (a) proposals generated by random sampling, (b-d) re-estimation of models and their inliers after several iterations of *expand* and *re-estimate* steps for energy (3) or (5). Note that the algorithm may converge to good models even from a small set of rough guesses. In this example we did not use an *outlier label*, see Sec.2.3, and an optimal set of lines in (d) "explains" all data points.

$L_p \in \mathcal{R}^n$ is a label (model) assigned to a given data point $p \in P$. Then, PEARL method estimates models and their spatial support (inliers) by optimizing the following energy of labeling $\mathbf{L} = \{L_p | p \in P\}$

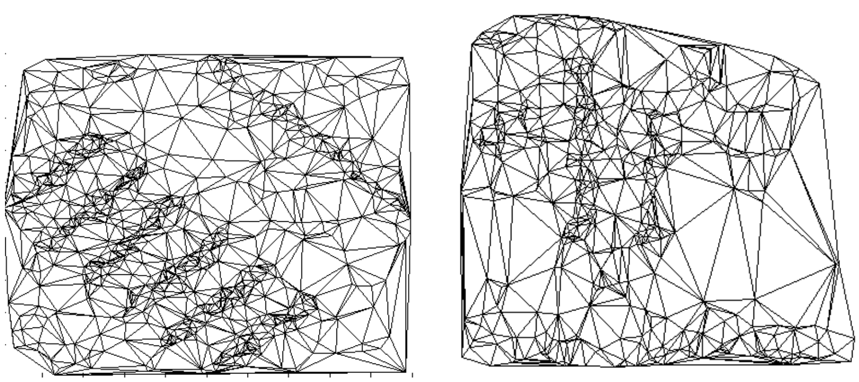
$$E(\mathbf{L}) = \sum_p ||p - L_p|| + \lambda \cdot \sum_{(p,q)\in\mathcal{N}} w_{pq} \cdot \delta(L_p \neq L_q). \tag{3}$$

The first term $||p - L||$ in (3) measures geometric error between point p and model L. For example, the line fitting examples in this section use "perpendicular distance" between 2D point $p = (x, y)$ and line $L = (a, b)$[5]

$$||p - L|| = \left(\frac{|y - ax - b|}{\sqrt{a^2 + 1}}\right)^2$$

which is the distance from p to the nearest point on line L. Robust (truncated) measures are also possible. Term $||p - L||$ corresponds to the log-likelihood $\ln \Pr(p|L)$ when energy (3) is interpreted as an MRF-based posterior energy. Thus, the use of quadratic distance for $||p - L||$ is equivalent to assuming Gaussian distribution for errors. Clearly, optimal labeling L for (3) depends on specific choice of geometric measure $||p - L||$.

[5] E.g. points (x, y) on line $L = (a, b)$ satisfy $y = ax + b$. Note that this basic representation of lines does not cover vertical lines. Alternatively, one can use 2 polar parameters, or 3 homogeneous parameters.



(a) “Lines” data, Fig.8 a

(b) “Clorox” data, Fig.13 b

Fig. 5 We use Delaunay triangulations of data points. K-nearest points or other techniques can be used as well, particularly for higher dimensional data. We did not observe much difference in practical performance.

The second term of energy (3) is a smoothness prior. It assumes some specific neighborhood system $\mathcal{N}$ for the data points. For example, the neighborhood system could be based on a triangulation of points, see Fig.5. In this paper we use the Potts model, e.g. [5], where $\delta(\cdot)$ is 1 if the specified condition inside parenthesis holds, and 0 otherwise. Weights w_{pq} set discontinuity penalties for each pair of “neighboring” data points. For example, the synthetic line fitting examples in this section use weights w_{pq} inversely proportional to the distance between points p and q because closer points are a priori more likely to fit the same model

$$w_{pq} = \exp - \frac{||p-q||^2}{\varsigma^2}.$$

In all of our synthetic line experiments ς was constant and it was chosen heuristically to be 5. Examples in Section 3 use constant weights $w_{pq} = 1$. Besides Potts (piece-wise constant) prior, one can also consider piece-wise smooth priors. Such priors would allow small variation of model parameters between data points.

2.3 Outliers

In the context of multi-model fitting the term *outlier* may become somewhat philosophical. For example, Fig.4(d) shows an optimal solution with respect to energy (3) where a small set of lines explains all data points. In this case the word “outlier” is subject to an interpretation. In particular, one could use any specific “outlier criterion” to classify weak models, e.g. those with sufficiently small number of inliers.

In this paper we mostly use a different approach common in MRF-based literature. We introduce a special “outlier model” and the corresponding label $\emptyset$ which is always present in the pool of labels when minimizing energy (3) or (5). Any point p assigned this label is considered an *outlier* in this paper[6]. In contrast to real geometric models, label $\emptyset$ has constant fidelity measure $||p - \emptyset|| = \gamma$ for all points $p \in P$. Intuitively, the outlier label corresponds to a “uniform” model. Typical weak models incur regularization penalties (smoothness and

[6] Some alternative ways to define outliers in our general energy-based framework are discussed in [18].

label costs) while explaining only a few points. Thus, outlier label is often optimal for points that otherwise would be assigned some weak model. For example, compare the optimization results in Fig.4(d) obtained without the explicit outlier label and the results in Figs.7-9(f) where $\emptyset$ was added into the pool of labels when minimizing energy (3).

2.4 *Expand* and *re-estimate* labels

Energy (3) can be minimized using α-expansion algorithm [5] for labels $\alpha \in \mathcal{L}_0$. In this case, it is possible to interpret α-expansions as a competition among model-labels for spatial support; models with the best-fit to data points find the largest number of spatially coherent "inliers", while most of the "erroneous" models get no inliers.

Once inliers are computed, model labels in $\mathcal{L}_0 \subset \mathcal{R}^n$ with non-empty set of inliers can be re-estimated as follows. Note that the first term of energy (3) can be represented as

$$\sum_p ||p - L_p|| \quad = \quad \sum_{L \in \mathcal{L}_0} \; \sum_{p \in P(L)} ||p - L||$$

where $P(L) = \{p \in P | L_p = L\}$ denotes a set of inliers for label L. Clearly, we can minimize this expression by re-estimating parameters of each model $L \in \mathcal{L}_0$

$$\hat{L} = arg \min_l \sum_{p \in P(L)} ||p - l||. \tag{4}$$

We replace each label L with non-empty support $P(L)$ by label $\hat{L} \in \mathcal{R}^n$ which has a better fit to points in $P(L)$. Finally, after discarding all labels with no inliers, we obtain a new set of labels $\mathcal{L}_1$. Note that this operation does not affect the second (smoothness) term in (3) unless two labels L, L' become equal after re-estimation $\hat{L} = \hat{L}'$ (in this case, the smoothness energy also decreases). Clearly, the described operation of changing the set of labels

$$\mathcal{L}_0 \to \mathcal{L}_1$$

can only decrease the energy (3).

There are many known methods for optimizing the sum of geometric errors $||p - L||$ in (4). Optimization method may depend on specific choice of measure $||p - L||$. For example, the minimum sum of squares of orthogonal errors in our lines-fitting examples could be obtained using a standard closed formula. A large number of other examples of geometric or algebraic error measures $||p - L||$ and different numerical methods for optimizing them are widely discussed in the computer vision literature, e.g. see [17]. Our approach can incorporate many of these error functions $||p - L||$.

Figure 4(b) visualizes clusters of inliers and re-estimated models $\mathcal{L}_1$ obtained in two separate steps described above: *expand* (inlier classification) and *reestimate* (model parameters). In some cases it could be useful to iterate the *propose* step as well. For example, new labels can be generated by merging or splitting clusters of inliers. One interesting example of "merging" is described in the context of example in Figure 10 in the end of this section.

2.5 Algorithm and its properties

Both *expansion* (inlier classification) and *reestimation* steps decrease energy (3). Thus, we can iterate over these steps until convergence, see Fig.4 (b-d). We can stop the iterations when a new round of α-expansion does not change inliers. As soon as the spatial support of the current models (labels) stops changing, re-estimation of the models (4) can not improve geometric error term. It is clear that this iterative algorithm converges to a local minimum. PEARL algorithm (Propose Expand and Re-estimate Labels) is summarized here:

1) **Propose**:
— at initialization, set i=0, randomly sample data to get $\mathcal{L}_0$, may add label $\emptyset$ (Sec.2.3)
—* (optional for $i > 0$) sample more or merge/split current models in $\mathcal{L}_i$
2) **Expand**:
— run α-*expansion* [5] for energy (3) or (5)[7] and for $\alpha \in \mathcal{L}_i$
— if the energy does not decrease, stop
3) **Re-estimate Labels**:
— solve (4) and obtain a new set of labels $\mathcal{L}_{i+1}$
— set i = (i+1), go to step 2 (or optional to *).

Figure 6 gives idea on how accuracy of estimated models depends on the number of initial randomly sampled proposals $|\mathcal{L}_0|$. For simplicity, we generated synthetic data supporting only one line. This also allows to juxtapose PEARL with standard RANSAC. Both methods used the same initial set $\mathcal{L}_0$ of randomly sampled model proposals. RANSAC basically selected the best model in $\mathcal{L}_0$ with the largest number of inliers w.r.t. some fixed threshold. It was easy to tune energy (3) so that an optimal data labeling is binary with the following two labels: one line label and outlier label $\emptyset$. Depending on initialization, PEARL's iterative *expand* and *re-estimate* steps (illustrated in Fig.4) would converge to different local minima. Normally, larger set of initial random proposals $\mathcal{L}_0$ leads to better solutions. Particularly for small $|\mathcal{L}_0|$ PEARL may output 2 line models, in which case we reported the error for the model with the largest support. In general, Fig.6 shows that by minimizing energy (3) PEARL can output a line significantly better than the best line in $\mathcal{L}_0$. In contrast, RANSAC strongly relies on a larger number of initial samples to find accurate estimates. Some generalizations of RANSAC explicitly improve the initial proposals, e.g. by re-sampling the inliers of the strongest initial models [8]. Similarly to RANSAC, their generalization to multi-model problems could be problematic.

Figures 7-9 compare PEARL to existing geometric multi-model fitting methods [40, 27] and mean-shift [9] which were discussed in the introduction. The synthetic multi-line examples were generated using different levels of Gaussian noise and different number of uniformly distributed outliers. The previous methods were tuned to get the best results for each specific level of noise and clutter. To demonstrate robustness of PEARL, in each test we tuned only one parameter σ in the geometric error measure $||p - L|| = -\ln G_\sigma(p - L)$ where $G_\sigma(\cdot)$ is a Gaussian distribution function and $p - L$ is the distance from point p to line L. Note that PEARL obtains very similar results when parameter σ is automatically estimated as described in the context of example in Figure 11.

For PEARL and multi-RANSAC the data points were uniformly sampled while for J-linkage and mean-shift we used distance-based sampling which helps J-linkage and mean-shift algorithms. Sampling closer points increases the probability that the sampled model is closer to one of the peaks in the Hough transform.

[7] In case of energy (5) one can use an extra *merging* operation or an extension of α-*expansion* [11].

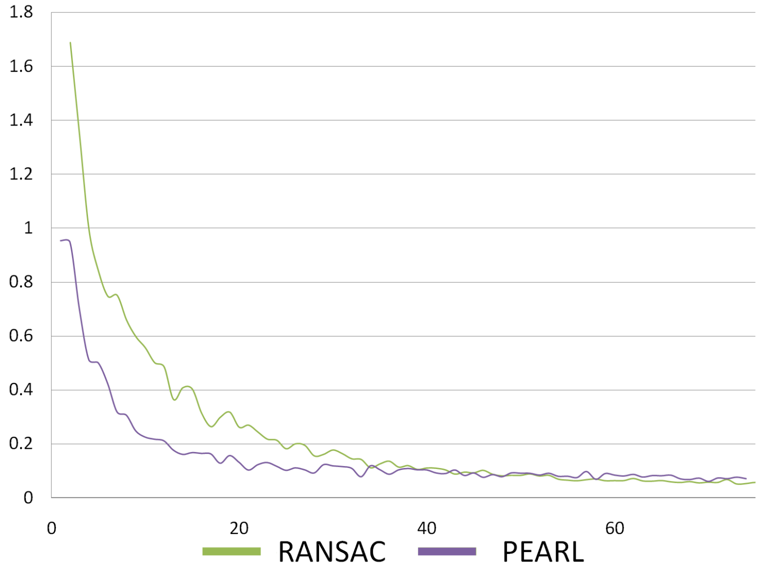
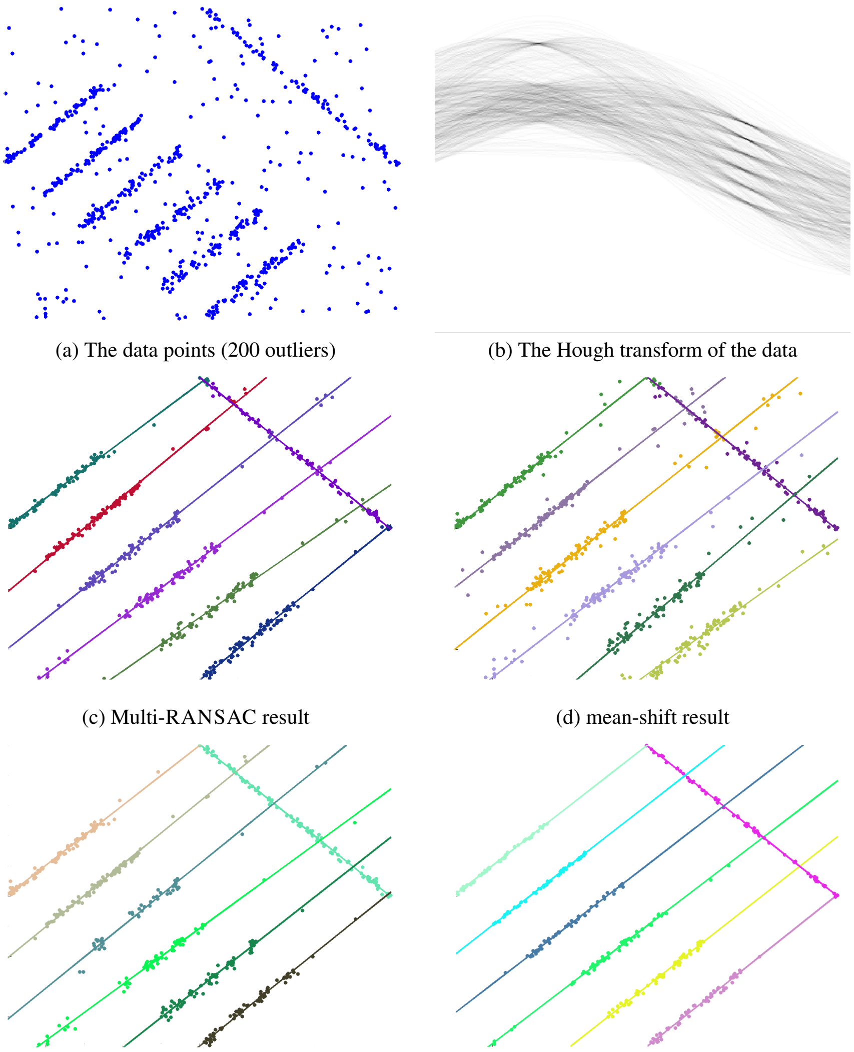


Fig. 6 PEARL vs RANSAC for synthetic data sets with 80% uniformly distributed outliers and 20% noisy inliers supporting one line model. In each test both methods used the same initial set $\mathcal{L}_0$ of randomly sampled model proposals. X-axis is the number of initial models $|\mathcal{L}_0|$. Y-axis shows estimation errors w.r.t. parameters of the ground truth line model. The errors are averaged over 400 randomly generated tests. RANSAC basically selects the best model from the initial set of proposals $\mathcal{L}_0$. By iteratively minimizing global energy (3) PEARL can converge to a much better model than those in the initial set $\mathcal{L}_0$. In contrast, standard RANSAC strongly relies on a larger number of initial samples to find accurate estimates.

Mean-shift and J-linkage have no constraints on the number of models they generate. Figures 7-9 show their strongest 7 models. Multi-RANSAC had to be given the exact number of models. Compared to mean-shift and J-linkage, PEARL finds a very small number of models giving with the optimal fit to the data. But, in addition to correctly identified true models, it can “hallucinate” a few models among outliers (as in Fig.4d). Such weak models can be automatically filtered out by setting a very conservative limit on the minimum number of inliers.

The results for standard methods in Figures 7-9 are consistent with our earlier observations in Figure 1: common greedy heuristics selecting one model with a large score (e.g. number of inliers) independently from the overall solution could be problematic in multi-model problems. As illustrated in Fig.1, random models may have higher scores (more inliers) than the true ones. This explains why many standard methods work relatively well only for the low noise example in Figure 7.

As we discussed in the introduction, coherence between inliers is often a good assumption particularly for problems in vision. Figure 10 shows one typical example where this assumption could be violated. Clearly, one of the intersecting lines cannot be assigned spatially connected group of inliers. More over, optimal solution for energy (3) can not merge two models with very similar parameters if their inliers are spatially separated (Fig.10b). However, a simple postprocessing after each *expansion* step can merge separated groups of inliers with similar models (Fig.10c) if the “average” optimal model increases the sum of geometric errors by no more than some predefined threshold β. In fact, this merging op-



(a) The data points (200 outliers)

(b) The Hough transform of the data

(c) Multi-RANSAC result

(d) mean-shift result

(e) J-Linkage result

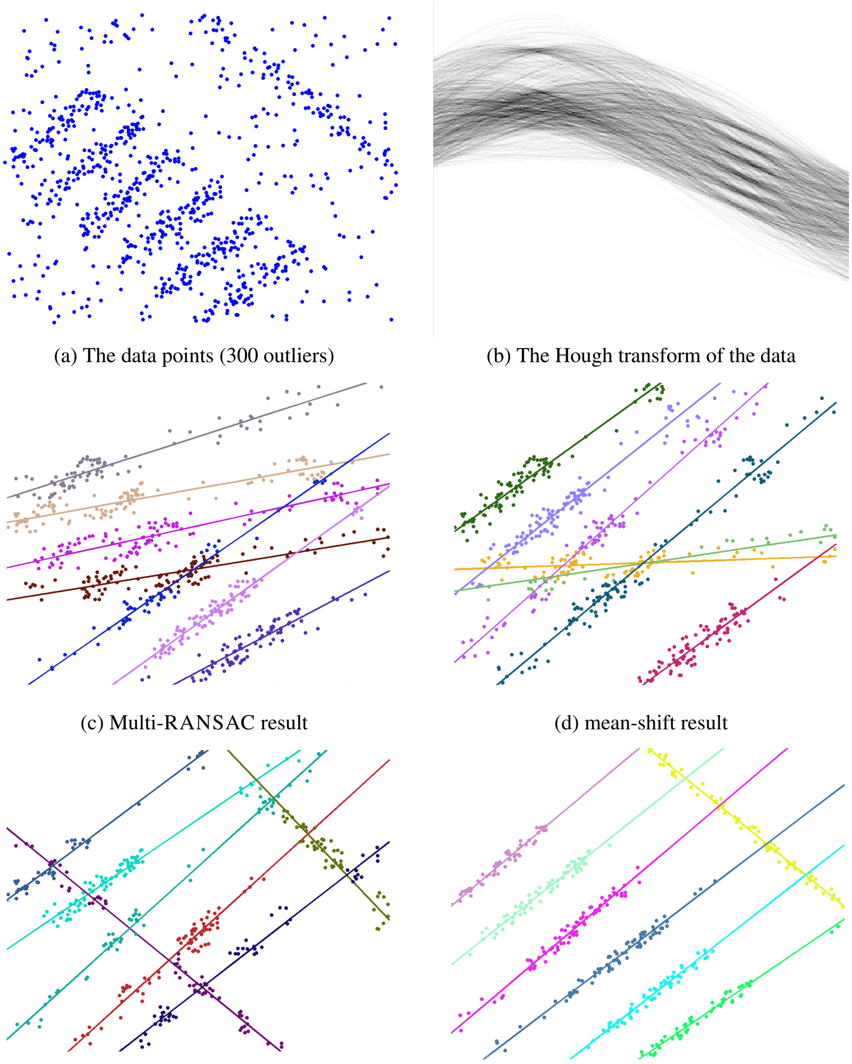
(f) PEARL result

Fig. 7 Comparing the results for fitting lines to noisy data points. The data points were perturbed with a low level of Gaussian noise ($\sigma = 0.01$) and 200 random outliers were added. Outliers represent 25% of the data.

eration can be seen as an optimization step if the energy function gets an additional term penalizing the number of models/labels with non empty support $|\mathcal{L}_{\mathbf{L}}|$

$$E(\mathbf{L}) = \sum_{p} ||p - L_p|| + \lambda \cdot \sum_{(p,q)\in\mathcal{N}} w_{pq} \cdot \delta(L_p \neq L_q) + \beta \cdot |\mathcal{L}_{\mathbf{L}}|. \qquad (5)$$

Similar merging operations were also used in [38] for a continuous version of this label cost energy. Instead of the proposed merging heuristic, energy (5) can be also minimized using an extension of α-expansion algorithm [11]. Note that ideas in [30] allow to generalize the label



(a) The data points (300 outliers)

(b) The Hough transform of the data

(c) Multi-RANSAC result

(d) mean-shift result

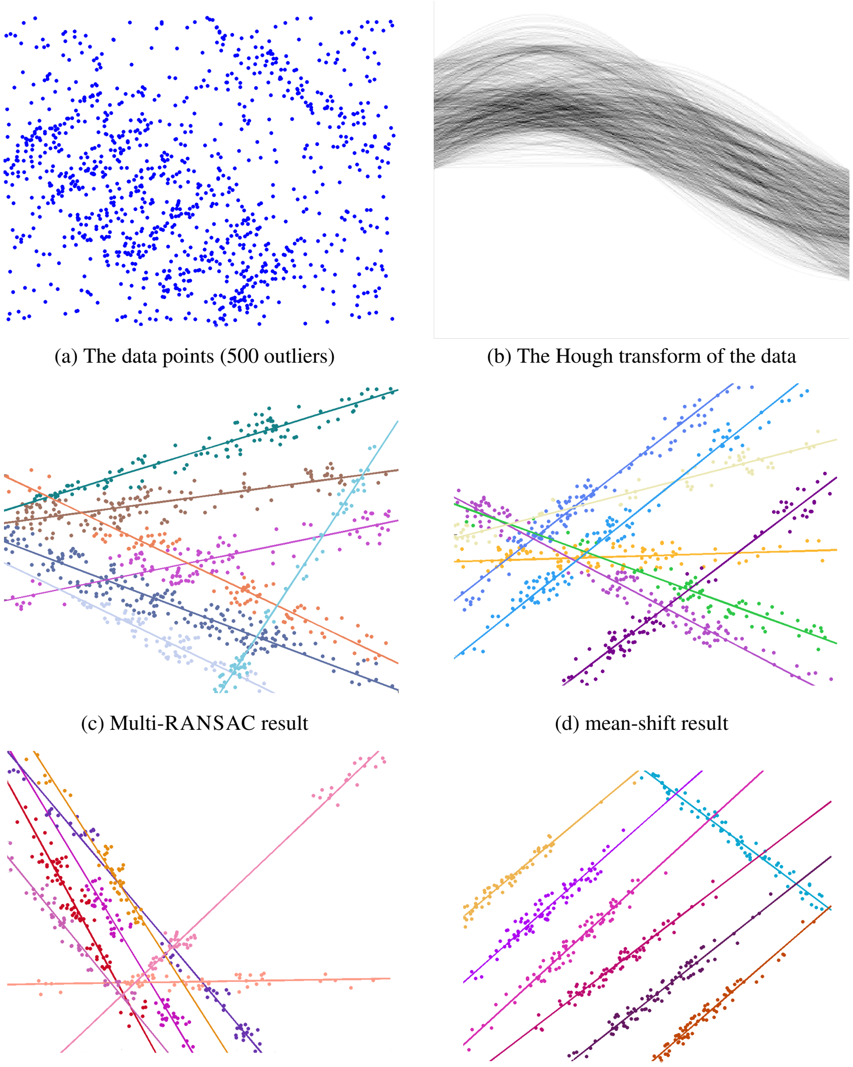
(e) J-Linkage result

(f) PEARL result

Fig. 8 Comparing the results for fitting lines to noisy data points. The data points were perturbed with medium Gaussian noise ($\sigma = 0.02$) and included 300 random outliers. Outliers represent 35% of the data.

costs in energy (5) in order to work with models of different complexities. This extension of our optimization framework is straightforward and left as a simple exercise for the reader.

Figure 11 demonstrates another interesting feature of our optimization approach. Unlike many previous multi-model fitting methods [34, 40, 34] using fixed thresholds, PEARL can identify multiple models with different levels of noise. For example, this can be achieved as follows. Assuming that geometric errors for inliers correspond to Gaussian noise, one can set geometric error penalty $||p - L_p||$ according to the negative logarithm of the normal



(a) The data points (500 outliers)

(b) The Hough transform of the data

(c) Multi-RANSAC result

(d) mean-shift result

(e) J-Linkage result

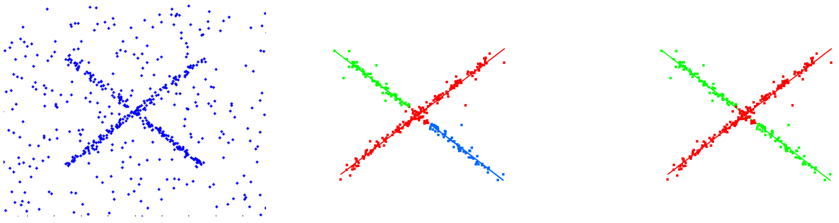
(f) PEARL result

Fig. 9 Comparing the results for fitting lines to noisy data points. The data points were perturbed with high Gaussian noise ($\sigma = 0.025$) and 500 random outliers were added. Outliers represent 45% of the data.

distribution function

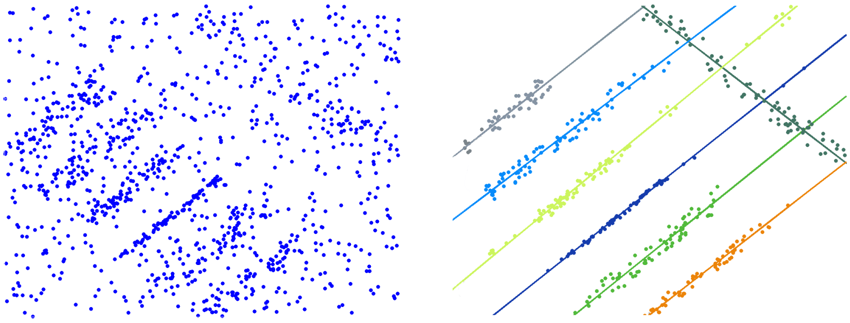
$$||p - L_p|| \;=\; -\ln G_\sigma(p - L_p)$$

where $p - L_p$ is the distance from p to the assigned model L_p. Here one assumes some known σ parameter corresponding to the distribution's variance. If models come with unknown different levels of noise, one can estimate extended labels $\tilde{L}_p = \{L_p, \sigma_p\}$ combining geometric model parameters L_p with the corresponding unknown noise-level σ_p. In this case



(a) data (300 outliers) (b) minimum of energy (3) (c) *merging*, energy (5)

Fig. 10 Intersecting lines example. Optimization of energy (3) may leave spatially isolated groups of inliers assigned to 2 models even if their parameters are infinitesimally close (b). Per-label costs in energy (5) solves this problem (c) but requires either an additional merging operation or an extension of α-expansion [11]. This example may also suggest EM-style soft assignments of labels to points near the intersection. However, in vision (Sec.3) we get occlusions (not intersections), which better motivate our "hard" assignments of labels.



(a) models with different noise levels (b) PEARL result based on (6)

Fig. 11 Fitting lines with different noise levels. The inliers in (a) were generated with different levels of Gaussian noise. 40% of the data are outliers. In (b) PEARL estimated labels combining geometric model parameters with unknown noise variances using error measure (6). Previous multi-model fitting methods use fixed thresholds to identify inliers, which would not work in this case.

one can use error measure

$$||p - \tilde{L}_p|| \;=\; -\ln G_{\sigma_p}(p - L_p). \tag{6}$$

Figure 11 shows that this approach can correctly estimate both geometric and statistical noise-level parameters for each model.

3 Experimental results

This section validates our model fitting technique (PEARL) on multi-view reconstruction data sets supporting multiple models. Our experiments used affine models (Sec.3.1) and homographies (Sec.3.2). Data points were obtained by matching SIFT[21] features on rectified image pairs in narrow-based stereo, and on uncalibrated wide-base pairs.

3.1 Estimating multiple affine models

In this section we apply PEARL to estimate affine transformation in the context of rectified narrow-base stereo. We use SIFT [21] features as points of interest, since they are scale and rotation invariant. They are also partially invariant to illumination and 3D camera view point changes. Matches between pairs of points in two images are found using exhaustive search along the corresponding scan line. In principle, it is possible to replace exhaustive search with "smarter" methods as in [1,23].

We will use the notation (x_l, y_l), (x_r, y_r) to describe the coordinates of the image feature on the left image p_l and right image p_r.The symbol p denotes a pair of matching points (x_l, y_l, x_r, y_r). Restricting the search for matching pairs to the corresponding epipolar lines[8] corresponds to imposing an additional constraint $|y_l - y_r| < \epsilon$ for some small threshold ϵ.

A planar homography has only three degrees of freedom for rectified images. In this case the epipole $\mathbf{e} = [1\ 0\ 0]^T$ is at infinity and the fundamental matrix can be written as

$$F = [\mathbf{e}]_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

where $[\mathrm{e}]_x$ describes the skew-symmetric matrix of the vector e. Following [12], a planar homography satisfies the following constraint $H^T F + F^T H = 0$. Then, it can be shown that any planar homography H for a rectified stereo pair is an affine transformation

$$A = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

with 3 degrees of freedom corresponding to parameters a, b, and c. This transformation can be uniquely identified from three matching pairs. We first generate a finite set of initial model proposals A_0 by randomly sampling three pairs of matching points and by computing parameters a, b, c for the corresponding models.

One simplistic way to measure geometric error between matching pair $p = (x_l, y_l, x_r, y_r)$ and model A is a non-symmetric (left) *transfer error* from $p_l = (x_l, y_l)$ to $p_r = (x_r, y_r)$

$$||p - A|| = |A \cdot p_l - p_r|^2 = \Delta_x^2 \tag{7}$$

where

$$\Delta_x = (ax_l + by_l + c - x_r)$$

is a *horizontal shift* between $A \cdot p_l$ and p_r along the epipolar line, see Fig.12(a). This basic approach assumes that the *vertical shift* between $A \cdot p_l$ and p_r

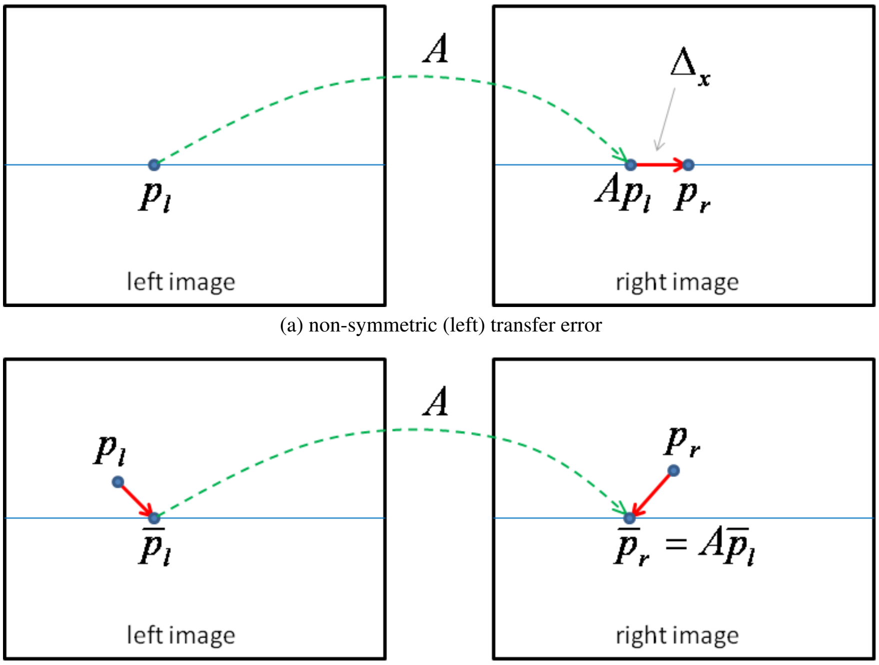
$$\Delta_y = (y_l - y_r)$$

is zero. This could be justified because our matched pairs $p = \{p_l, p_r\}$ are points on the same scan lines ($|y_l - y_r| < \epsilon$) and affine transformation A respects such (epipolar) lines.

Alternatively, one can use the *reprojection error* [17] illustrated in Fig.12(b). This approach treats p_l and p_r as noisy observations of some unknown "true" points $\bar{p} = \{\bar{p}_l, \bar{p}_r\}$ estimated by minimizing the observation noise, as follows

$$||p - A|| = \min_{\bar{p}} |\bar{p}_l - p_l|^2 + |\bar{p}_r - p_r|^2 \quad s.t. \quad \bar{p}_r = A \cdot \bar{p}_l.$$

[8] These are scan lines for rectified stereo images.



(a) non-symmetric (left) transfer error

(b) reprojection error

Fig. 12 Geometric fitting errors $||p - A||$ for rectified stereo. Assuming matched points $p = \{p_l, p_r\}$ are on the corresponding epipolar (scan) lines $\Delta_y = y_l - y_r = 0$, the left *transfer error* is a horizontal shift Δ_x between p_r and Ap_l shown in (a). The standard *reprojection error* (b) is obtained by treating points p_l and p_r as noisy observations of some hidden "true" pair of matched points $\bar{p} = \{\bar{p}_l, \bar{p}_r\}$ such that $\bar{p}_r = A \cdot \bar{p}_l$ and which minimize the observation noise $\min_{\bar{p}} |\bar{p}_l - p_l|^2 + |\bar{p}_r - p_r|^2$.

For $\bar{p} = \{p_l, Ap_l\}$ the objective function above equals $|A \cdot p_l - p_r|^2$ and, therefore, optimization over all $\bar{p}$ should give an error smaller than (7). That is, our transfer error (7) can over-estimate the observation noise, as defined by the constrained optimization problem above. The difference could be particularly significant if plane A is near-horizontal. Note that the optimal "true" pair $\bar{p}$ corresponding to the minimum observation error is typically[9] located on a scan line different from those containing data points p_l and p_r.

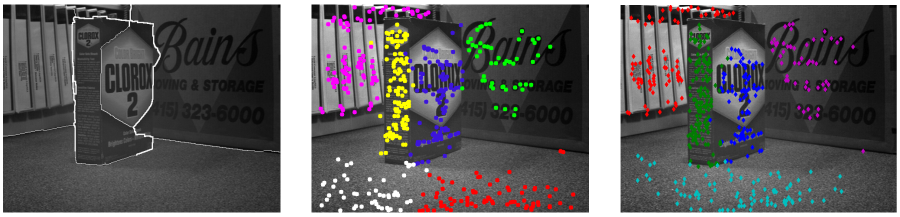
If p_l and p_r are treated as noisy observations, these points do not need to respect the epipolar geometry. Thus, we can drop the constraint $\Delta_y = 0$ when using the reprojection error. Following the definition in the previous paragraph, one can derive the following closed formula for the reprojection error specific to our affine transformations

$$||p - A|| = \frac{2\Delta_x^2 \; + \; (1 + a^2 + b^2)\Delta_y^2 \; - \; 2b\Delta_x\Delta_y}{2a^2 + b^2 + 2}. \tag{8}$$

Figures 13 (a-c) compare affine model fitting results regenerated by BT [2] and results generated by PEARL for two different geometric error measures $||p - A||$ in (7) and (8). We applied PEARL to energy

$$E(\mathbf{A}) = \sum_p ||p - A_p|| + \lambda \sum_{(p,q)\in\mathcal{N}} \delta(A_p \neq A_q) + \beta \cdot |\mathcal{A}|$$

[9] Except when A is an exactly vertical plane.



(a) BT [2]
segmentation of pixels

(b) PEARL (7)
classification of inliers

(c) PEARL (8)
classification of inliers

Fig. 13 Results for ”Clorox” stereo pair [2]. (a) Dense pixel segmentation by BT [2] uses photoconsistency. (b,c) Sparse inlier classification by PEARL using geometric fit measures (7,8).

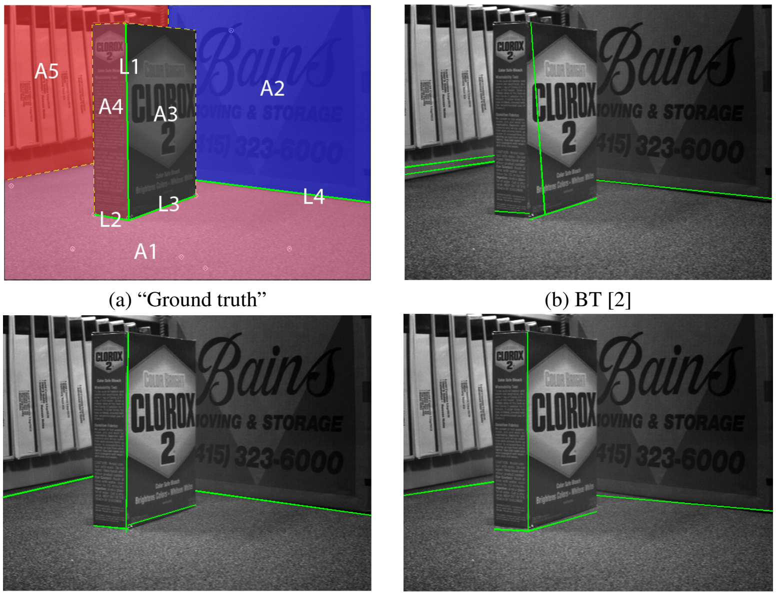
Line	BT [2]	PEARL (7)	PEARL (8)
L_1	34.55	8.80	4.37
L_2	16.83	4.82	7.5
L_3	5.56	13.27	9.53
L_4	5.99	4.46	8.47
Total	62.93	31.35	29.87

Table 1 Geometric errors for lines in Fig.14 where affine models intersect. Errors are computed as the sum of distances between the ground truth line segment corners and the computed lines. The errors for sequential RANSAC and J-linkage, see Fig.15, were significantly larger than those above.

where $\mathbf{A} = \{A_p | p \in P\}$ is an assignment of affine models to data points p extracted using SIFT and $|A|$ is the number of used affine models. Our neighborhood graph $\mathcal{N}$ is a triangulation of points p_r in the right image. BT [2] uses dense segmentation of pixels based on photoconsistency. This measure does not work well in texturelss regions and they have to rely on intensity edges (static cues) to detect the boundaries between regions supporting different models. In contrast, PEARL labels a sparse set of distinct features based on geometric errors and spatial regularization. Adding label cost term would allow our methods to connect spatially disconnected parts of the same model as in Figure 10.

Our results in Figures 13 (b,c) demonstrate that specific choice of geometric measure $||p - A||$ can significantly change the results. For example, minimizing the horizontal transfer error in equation (7) worked well for all vertical planes in the scene, but it split the ground plane into two, see Fig.13(b). The reprojection errors (8) worked significantly better than the transfer errors. This is particularly obvious for the ground plane. As mentioned earlier, the transfer error (7) significantly over-estimates the observation noise for near-horizontal planes. This is fairly analogous to the consequences of using “vertical” point-to-line distance $||p - L|| = |y - ax - b|^2$ for fitting near-vertical lines $L = (a, b)$ to 2D points $p = (x, y)$.

In order to provide some quantitative comparison between the affine models generated by BT [2] and PEARL, we used a “ground truth” image (Fig.14(a)) where we manually extracted the lines corresponding to intersecting planes. Assuming that two intersecting planes π_1 and π_2 are represented by the affine models A^{π_1} and A^{π_2}, the homogenous vector representing the line of intersection is defined as the first row of the matrix $(A^{\pi_1} - A^{\pi_2})$. Therefore, such lines can be computed from the models estimated by either BT or PEARL. Table 1 compares the accuracy of these lines with respect to our ground truth.



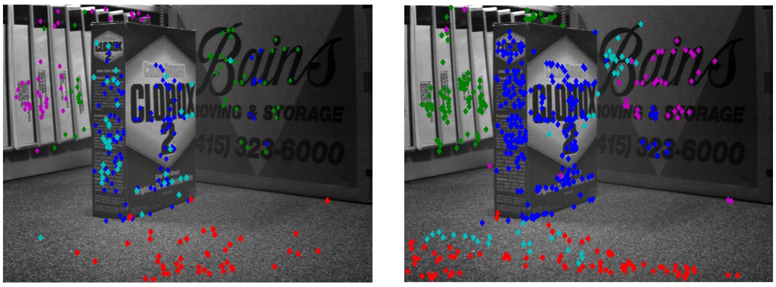
(a) “Ground truth”

(b) BT [2]

(c) PEARL (7)

(d) PEARL (8)

Fig. 14 Comparison of results by BT [2] and PEARL. Lines are computed for all pairs of intersecting planes (affine models).



(a) sequential RANSAC

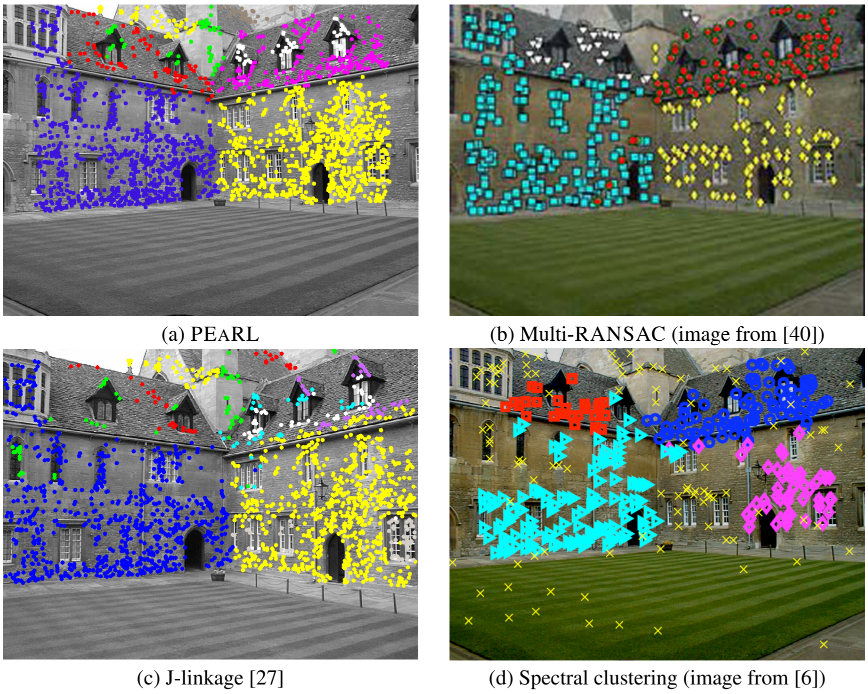
(b) J-linkage [27]

Fig. 15 Typical results for affine model fitting in Sec.3.1 via sequential RANSAC (a) and J-linkage [27]. We used 5000 uniform samples for RANSAC. Different threshold values did not give much improvement. Similarly, different tunings of J-linkage generated various artifacts. These problems might be explained by the large level of noise in the data, as in Fig.9.

3.2 Estimating multiple homographies

In this section we use PEARL to estimate multiple homographies in uncalibrated wide-base stereo image pairs. We use SIFT features [21] as points of interest. The set of all matched pairs of features P is found using exhaustive search.

One way to measure geometric error between a pair of points $p = (p_l, p_r)$ and a given model H is the *symmetric transfer error* (STE) [17]. We generate our finite set of initial model proposals H_0 by randomly sampling four matched pairs. Model parameters are com-



(a) PEARL

(b) Multi-RANSAC (image from [40])

(c) J-linkage [27]

(d) Spectral clustering (image from [6])

Fig. 16 Multi-homography fitting for stereo images from VGG (Oxford) Merton College I.

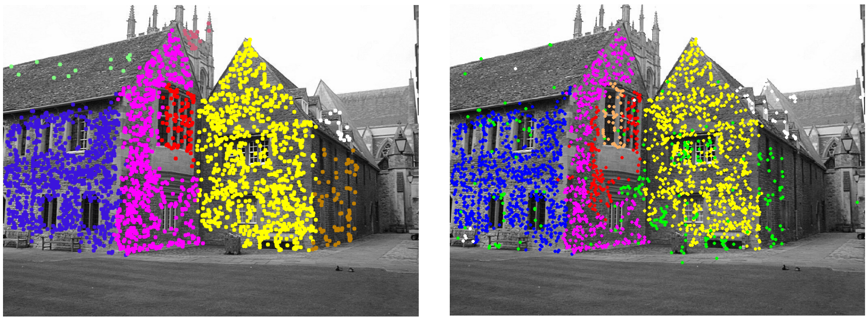
puted by minimizing the non-linear STE error using Levenberg-Marquardt, as described in [17]. The initial solution for the non-linear minimization is found using the *direct linear transform* (DLT) method. We apply PEARL to energy

$$E(\mathbf{H}) = \sum_{p} ||p - H_p|| + \lambda \sum_{(p,q)\in\mathcal{N}} \delta(H_p \neq H_q) + \beta \cdot |\mathcal{H}| \tag{9}$$

where $\mathbf{H} = \{H|p \in P\}$ is an assignment of models to data points p and the neighborhood system $\mathcal{N}$ is based on a triangulation of data points in one of the images.

In the example of Fig.16(a) PEARL identified 7 planes. The third term in energy (9) allowed to merge spatially isolated parts of yellow, green, and white planes. Unlike multi-RANSAC [40], PEARL does not require *a priori* knowledge of the number of planes and produces spatially coherent inliers. In [40] multi-RANSAC required 11604 iterations to fit 4 models to the same data, see Fig.16(b). Since each iteration sampled 4 random models, the total number of sampled homographies in Fig.16(b) is 46416. In contrast, PEARL used only 900 randomly sampled models to identify 7 planes. PEARL converged in three iterations. For qualitative comparison, Fig.16 also shows the result based on spectral clustering from [6] and the best result we could obtain using J-linkage [27].

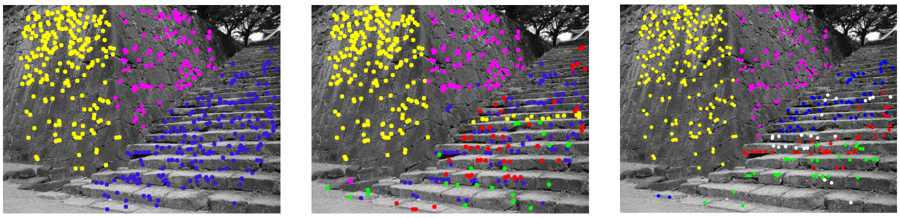
In Fig.17(a) PEARL identified 8 planes using only 3000 initial samples and converged after four iterations. PEARL finds planes with varying number of inliers. The roof on the left (light green) has only 13 inliers, while the two large walls (blue and pink) have 786 and 581 inliers, respectively. The qualitative comparison of our regularization-based method with greedy techniques like J-Linkage and sequential RANSAC in Figures 17 and 18 shows that



(a) PEARL

(b) J-linkage [27]

Fig. 17 Multi-homography fitting for stereo images from VGG (Oxford) Merton College III. Both results were obtained using standard SIFT features.



(a) PEARL

(b) sequential RANSAC

(c) J-linkage [27]

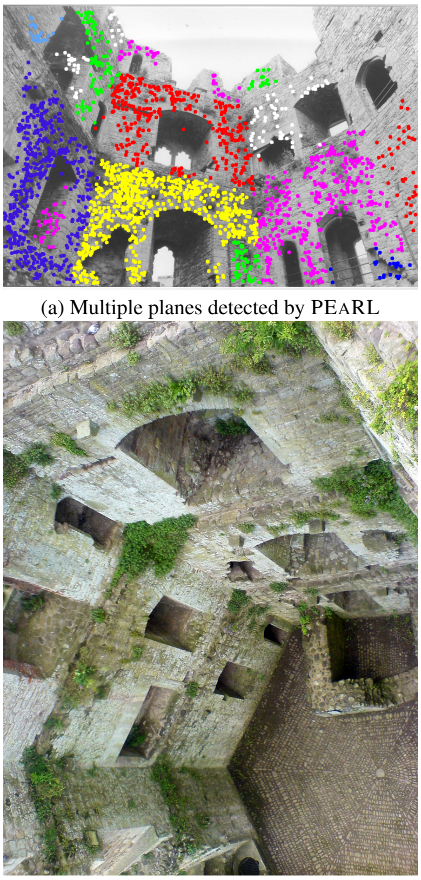
Fig. 18 Multi-homography fitting for "stairs" stereo image from VGG (Oxford). The results in (b) and (c) represent our best efforts with tuning thresholds in RANSAC and J-linkage. For example, two "walls" start to leak for larger value of thresholds, while smaller thresholds over-segment the "stairs" even more than above.

they are less robust to high noise. Similar general limitation of greedy inlier maximization approaches was previously demonstrated on synthetic data in Figures 1 and 9.

Figure 19 shows Raglan Castle Tower result for PEARL using 6000 initial labels and only four iterations to convergence. PEARL identified 13 planes. The use of relatively large number of initial labels allowed PEARL to identify very small planes. Another picture of Raglan Castle Tower from flicker confirms that the walls on the second and the third floors represent different parallel planes.

3.3 Motion Segmentation

In this section we aim to solve the multibody motion segmentation problem using multiple-views. This problem is also referred to in literature as the multibody structure from motion problem [4, 28, 10]. The goal of this problem is to cluster the scene trackable features among distinct motions, then to estimate the motions' parameters and to recover the 3D structure of the points. We are only interested in estimating the multiple motions and clustering of image features.



(a) Multiple planes detected by PEARL

(b) A different point of view (by anonymous Flicker photographer)

Fig. 19 In this Raglan Castle example from VGG (Oxford) we used the same color more than once to represent different planes (a). Only spatially connected planes are shown in different colors. An image of the same scene from a different view (b) confirms that each floor of the building corresponds to different planes.

3.3.1 Using two-views

Assume that the multiple bodies are rigid and each body undergoes a different motion. Each distinct rigid-body motion (R,t) could be described by a fundamental matrix $F = [K't]_x K'RK^{-1}$ corresponding to two views. This fundamental matrix satisfies the epipolar constraint $p_r^T F p_l = 0$ where p_r and p_l are two matching features corresponding to a 3D point on some rigid body [22,20].

We apply PEARL to estimate multiple fundamental matrices for uncalibrated image pairs. Matching pairs are found using the same procedure mentioned in section 3.2. One way to measure geometric error between a matching pair of points p and a given model F is

	PEARL	ER1	ER2	REF	GPCA	LSA *4n*	RANSAC
# frames used	2	2	2	N	N	N	N
# models fixed	-	3	3	3	3	3	3
average	21.2%	6.93%	4.2%	6.28%	31.95%	5.80%	25.78%
median	11.9%	4.15%	1.2%	5.06%	32.93%	1.77%	26.01%

Table 2 Misclassification errors for *checkerboard* motion estimation sequences in [32]. The results for REF, GPCA [33], LSA [36], and RANSAC are copied from Table 4 in [32]. These methods use all N-frames in each video sequence. In contrast, our results for PEARL with energy (11) use only 2 frames, the first and the last in the sequence. Other methods also "know" that the exact number of models is 3. In contrast, PEARL may obtain a different number of models. Figure 22 illustrates how this affects PEARL's classification errors. For a more balanced comparison we show 2 additional statistics. First, column ER1 evaluates 15 out of 26 examples where PEARL obtained exactly 3 models. Second, similarly to REF [32], column ER2 evaluates the "reference" solutions obtained by PEARL from the ground truth models in all 26 examples.

the squared Sampsons distance (SSD) [17]

$$||p - F|| = \frac{\left(p_r^T F p_l\right)^2}{(Fp_l)_1^2 + (Fp_l)_2^2 + \left(F^T p_r\right)_1^2 + \left(F^T p_r\right)_2^2} \tag{10}$$

where the $(Fp_l)_j^2$ represents the square of the j-th entry of the vector (Fp_l) . We generate our finite set of initial model proposals F_0 by random sampling eight matching pairs. Then compute the model parameters as descried in [17] by minimizing the non-linear SSD error using Levenberg-Marquardt. The initial solution for the non-linear minimization is found using the normalized 8-point algorithm [16]. The next step is to triangulate the features of one of the images (e.g the right image). Then we apply PEARL to energy

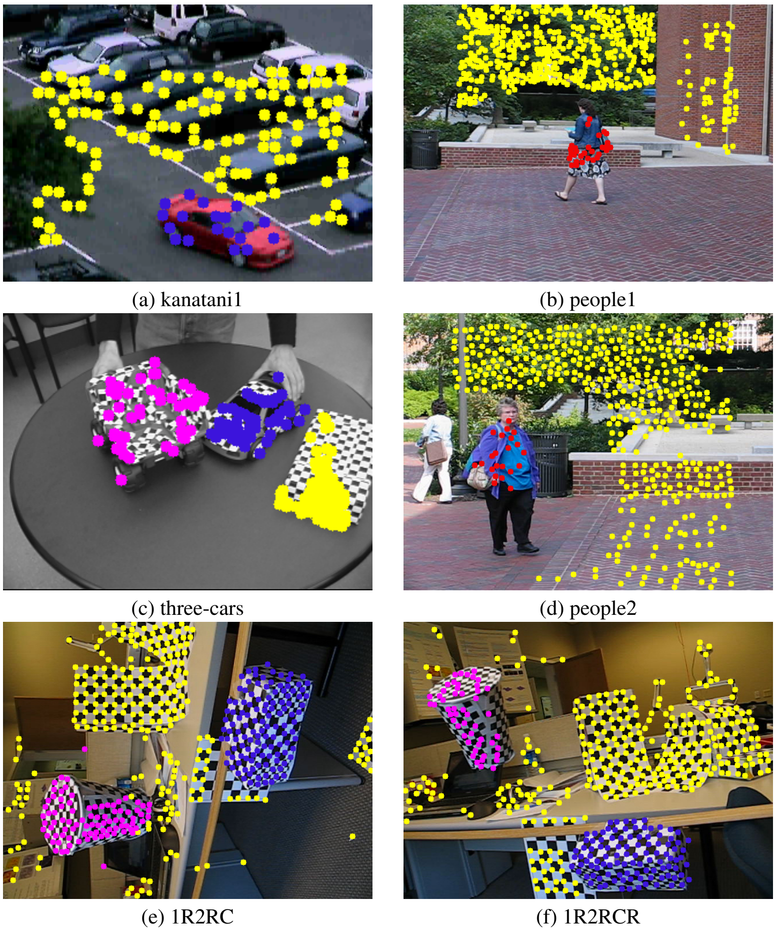
$$E(\mathbf{F}) = \sum_p ||p - F_p|| + \lambda \sum_{(p,q)\in\mathcal{N}} \delta(F_p \neq F_q) + \beta \cdot |\mathcal{F}| \tag{11}$$

where $\mathbf{F} = \{F|p \in P\}$ is an assignment of models to data points. Fig. 20 shows our representative results. In case the label cost term in (11) is dropped, it is likely that spatially isolated parts of the background could be assigned different motions, as shown in Fig.21.

Table 2 compares our results with some standard motion estimation methods evaluated in [32]. Note that these standard methods assume that the number of motions in the data is given. In contrast, our method automatically estimates the number of motions. The comparison in Table 2 may be not entirely meaningful since other methods benefit from *a priori* knowledge of the exact number of motions, while our method may get an incorrect number of motions contributing to gross misclassification errors, see Fig.22. One way to alleviate this problem is to report PEARL's statistics on examples where the number of models was estimated correctly, see column ER1. We also evaluated the "reference" solutions obtained when PEARL is initialized with 3 ground truth models, see column ER2.

Statistically required number of samples needed to generate good samples of fundamental matrices in the motion examples could be quite large. For example, if data points contain only 3 motions with m_i inliers (each) and m_o outliers, one can compute the probability that n independent samples have at least one good representative of each model

$$\begin{aligned}\Pr(\text{3 out of n}) \;=\;& 1 - (1-\rho_1)^n - (1-\rho_2)^n - (1-\rho_3)^n \\ & + (1-\rho_1-\rho_2)^n + (1-\rho_2-\rho_3)^n + (1-\rho_3-\rho_1)^n \\ & - (1-\rho_1-\rho_2-\rho_3)^n\end{aligned}$$


(a) kanatani1

(b) people1

(c) three-cars

(d) people2

(e) 1R2RC

(f) 1R2RCR

Fig. 20 Representative PEARL's results for motion sequences from Rene Vidal's data set [32]. Our motion estimation using mixtures of fundamental matrices (Sec.3.3.1) uses only two frames (the first and the last).

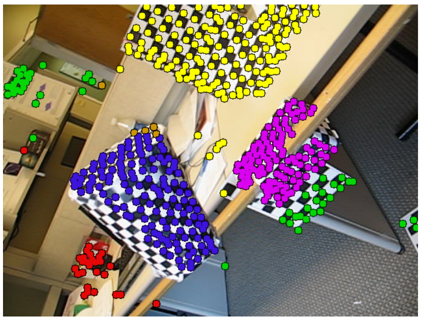


Fig. 21 Optimal solution for energy (3). In contrast to results for energy (5) in Fig.20e-h, it fails to generate one background motion from yellow, green, and red clusters. These clusters correspond to infinitesimally close motions, but they are spatially isolated. The third term in (5) addresses this issue.

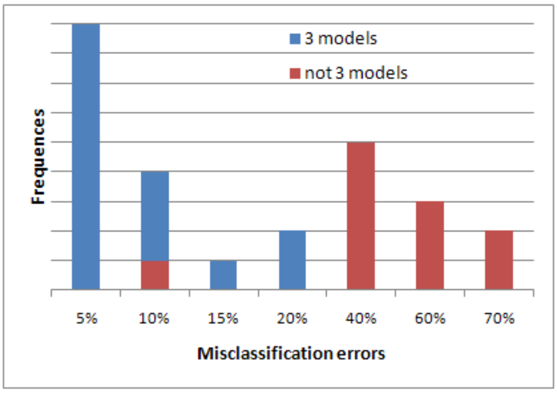


Fig. 22 Histogram of misclassification errors for PEARL with energy (11) on 26 *Chekerboard* examples [32] with 3 distinct motions. This histogram reveals two “modes”. The mode with smaller misclassification errors mainly contains examples where PEARL produced exactly 3 models. Such examples are marked in blue. Examples where PEARL produced a different number of models are marked in red. Such examples formed the “gross errors” mode. The percentage of misclassified points may not be a proper measure for comparing a method automatically computing the number of models against the methods that *a priori* know the correct number. We separately report PEARL’s error statistics for 15 (blue) examples with the correct number of obtained models, see ER1 in Table 2.

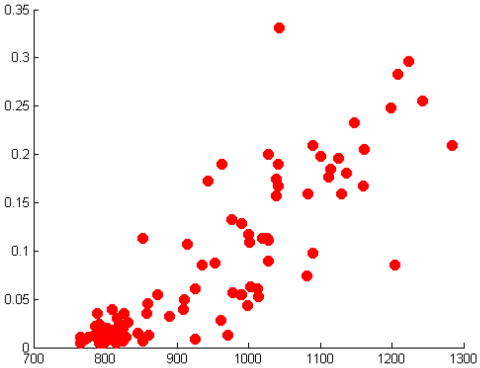


Fig. 23 Scatter plot of different solutions obtained by PEARL on one of the examples in [32] using different sets of 5000 (uniformly) sampled fundamental matrices. This number of samples is statistically insufficient for the motion estimation examples in Sec.3.3.1 (see text) and the quality of optimization may suffer. The plot also illustrates positive correlation between the values of our energy (11) and the misclassifications errors.

where $\rho_i \approx (\frac{m_i}{m_1+m_2+m_3+m_o})^k$ is the probability that k randomly selected points (for estimating a model) come exclusively from m_i inliers of the ith model. In *checkerboard* sequences [32] there are 3 models with $m_1 = 20$, $m_2 = 24$, and $m_3 = 56$ inliers and the number of outliers is $m_o = 0$. In this case $n = 1000000$ gives only 0.92 confidence, but even this number of samples n is too large to be practical. At the same time, under-sampling could lead to suboptimal results, see Fig.23. Our current MATLAB implementation can not work with very large number of samples due to memory restrictions. To explore larger set of labels, we used a simple extension of PEARL that samples only 5000 models at a time, computes the corresponding optimal solution, and mixes the optimal models with a new batch of 5000 random samples in the next iteration. To report statistics for PEARL and ER1 in Table 2 we ran 100 of such iterations and selected the solution with the minimum energy. On average, our current MATLAB implementation finished these 100 runs in around one hour. Running only 30 iterations reduces the average running time to 20 minutes. In this case the average/median errors increase, e.g. to 11.26/4.8 for ER1. The typical run-time for PEARL in ER2 was less than a minute. While there are many opportunities for improving the running times of our MATLAB implementation, we leave them for future work.

The results in this section were obtained by sampling the data points uniformly. Instead, we can use any standard non-uniform local sampling scheme (see [7] for a recent review). The main effect of using such sampling schemes in the *propose* step of PEARL could be lowering the number of samples required to find a good (low-energy) solution. A detailed study of this effect is outside the scope of this paper.

While PEARL's results in Table 2 are obtained assuming the same level of noise for all models, the practical effect of adding noise level as a parameter for each model, see Fig.11, could be further studied on real applications such as motion estimation.

3.3.2 Using multiple views/frames

Our general energy-based model fitting approach can also be applied to the N-frame point-trajectory data used by the motion estimation methods evaluated in [32]. In fact, the standard N-frame methodology introduced by Tomasi and Kanade [28] may lead to a more accurate and faster technique. For example, instead of fundamental matrices with 7 degrees of freedom, each rigid motion is represented as a lower dimensional (4D or less) linear manifold in the space of motion trajectories. This representation assumes affine projection.

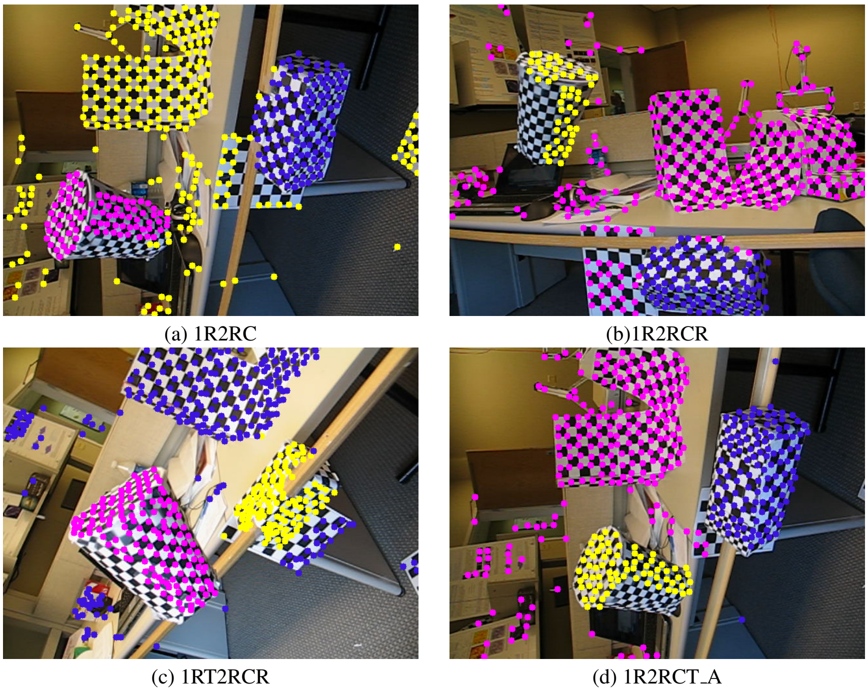
There are many ways in which our general optimization-based model fitting framework can be applied in this context. For example, it maybe used as a trajectory grouping technique instead of spectral clustering in GPCA [33] or LSA [36]. However, given the general scope of this paper, we present only the most straightforward set-up for fitting 4D hyperplanes following the basic formulation used in [32] for sequential RANSAC.

In contrast to greedy maximization of inliers by RANSAC, we seek optimal labeling of data points. Similarly to [30,26], our general energy optimization framework can assign differentiate label costs for fitting motions of various complexities, e.g. degenerate and non-degenerate. We can also replace piece-wise constant spatial regularization by piece-wise smooth in order to better address non-independent or articulated motions. For simplicity, however, we stick to the most basic formulation. It is detailed below.

Assuming that the cameras are affine then it could be proved that the motion of a rigid body i.e. the trajectory of its features will live in a 4D subspace [4,28,33]. Let P be a set of 3D points that belong to a rigid body which undergoes a motion over F frames. The image feature of a point i in the frame f is defined by p_{fi}. Stacking all p_{fi} measurements $\forall f \in F$ and $\forall i \in P$ will form the measurement matrix

$$W = \begin{bmatrix} p_{11} & \cdots & p_{1P} \\ \vdots & & \vdots \\ p_{F1} & \cdots & p_{FP} \end{bmatrix}_{2F \times P} \tag{12}$$

which has rank 4 [4,28]. First, we project these feature trajectories from $\mathcal{R}^{2F}$ to $\mathcal{R}^5$. The extra dimension is needed to discriminate between different motions [33]. Then we use



(a) 1R2RC

(b)1R2RCR

(c) 1RT2RCR

(d) 1R2RCT_A

Fig. 24 Representative PEARL's results using energy (13) for motion sequences from Vidal's data set [32]. Unlike the fundamental matrix fitting method in Sec.3.3.1, this formulation uses all frames in the sequence.

PEARL to fit multiple 4D hyperplanes (motions). In contrast to [33,36], PEARL does not require the prior knowledge of the number of motions.

We generate our finite set of initial proposals M_0 by randomly sampling four points, i.e. four projected trajectories. Then we find the best fitting 4D hyperplane by minimizing the orthogonal distance $||p - M||$. The next step is to triangulate the 2D image features on one of the frames (e.g the last frame). Finally, we apply PEARL to the following energy

$$E(\mathbf{M}) = \sum_p ||p - M_p|| + \lambda \sum_{(p,q)\in\mathcal{N}} \delta(M_p \neq M_q) + \beta \cdot |\mathcal{M}| \qquad (13)$$

where $\mathbf{M} = \{M|p \in P\}$ is an assignment of models to data points. Figure 24 shows some representative optimal results obtained by PEARL for energy (13) using the same set of parameters. Figure 25 indicates how the classification errors may depend on the main parameters in (13): weight of the smoothness term λ and weight of the label cost term β.

The detailed empirical comparison of our approach versus standard motion estimation methods is beyond the scope of this paper. Table 3 presents only results for *checkerboard* motion estimation sequences in [32]. As discussed in Sec.3.3.1, it is not clear how to compare our general technique with methods that assume a known number of models, or with methods that assume no outliers. Thus, table 3 reports several additional performance measures for our approach (ER1 and ER2) described in Sec.3.3.1.

Comparing tables 2 and 3 shows that our approach can produce more accurate results by fitting fundamental matrices to matches between the first and the last frames in each sequence. We do not make any conclusions from this fact as we tested only one and probably

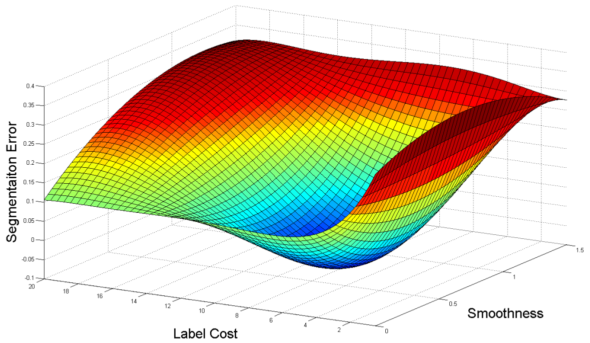


Fig. 25 Multi-motion estimation accuracy in one of the *checkerboard* sequences from Sec.3.3.2. Misclassification errors are shown for optimal solutions corresponding to different values of λ and β in energy (13).

	PEARL	ER1	ER2	REF	GPCA	LSA *4n*	RANSAC
# frames used	N	N	N	N	N	N	N
# models fixed	-	3	3	3	3	3	3
average	19.57 %	8.97 %	8.15%	6.28%	31.95%	5.80%	25.78%
median	19.28 %	3.53 %	3.74%	5.06%	32.93%	1.77%	26.01%

Table 3 Misclassification errors for *checkerboard* motion estimation sequences in [32]. The results for REF, GPCA [33], LSA [36], and RANSAC are copied from Table 4 in [32]. All methods in this table use all frames in each video sequence. PEARL may obtain a different number of models while the other methods "know" that the exact number of models is 3. For a more balanced comparison we show 2 additional statistics. First, column ER1 evaluates 12 out of 26 examples where PEARL obtained exactly 3 models. Second, similarly to REF in [32], column ER2 evaluates the "reference" solutions obtained by PEARL from the ground truth models in all 26 examples.

the most straightforward formulation of energy (5) for multi-frame motion detection. Energies (13) and (11) are defined in very different settings and it is not obvious why one should work better than the other. The general geometric framework for (13) is to detect independent motions as hyperplanes, which is comparable to those for GPCA and RANSAC in [32]. Other multi-frame motion detection methods, e.g. LSA [36], may obtain better results in a different geometric framework: they detect motions by fitting hyperspheres to normalized data points. A detailed analysis of different formulations is beyond the scope of our paper. The result for LSA is presented in tables 2 and 3 mainly as a reference to alternative geometric frameworks.

4 Conclusions

We proposed a new general approach to geometric multi-model fitting based on global optimization. The problem is formulated as discrete labeling of data points using MRF and MDL style regularization functionals widely used in other computer vision problems. The goal is to find models "explaining" all data points based on spatial regularity and sparsity priors. The continuous space of model parameters is explored via PEARL algorithm that combines data sampling and energy minimization iterating assignment and re-estimation steps. The method automatically obtains a small number of models that "explain" data well. Many empirical tests on synthetic and real imagery demonstrate a strong potential of our general approach applicable to a wide spectrum of model fitting problems. The main conclusion are:

- The proposed general energy functionals (3) and (5) are potent criteria for solving a wide class of geometric multi-model fitting problems with *a priori* unknown number of models corrupted by noise and outliers.
- The proposed algorithmic approach for minimizing our model fitting energies (PEARL) works well in many practical applications.

The main two parameters in the proposed energies (3) and (5) are the weights λ and β for the spacial regularity and label cost terms. We did not analyze any specific technique for selecting these parameters, but we demonstrate that the method is fairly stable with respect to them. Many useful ideas for parameter selection can be borrowed from information theoretic interpretation of these energies, e.g. see [30, 11].

One can use any geometric error function $||p - L||$ in the data term to evaluate the distance between models and data points. We also showed that fitting one additional uniform error model $\emptyset$ may work well for classifying the outliers. Our general energy formulation does not require any hard thresholds, event though one can use truncated or step error measures $||p - L||$, if necessary.

Acknowledgements We would like to thank Olga Veksler, Andrew Delong, Fredrik Kahl, and our anonymous reviewers for a number of discussions and suggestions that considerably improved the quality of this paper. This work was fully supported by funding from NSECR and CFI (Canada).

References

1. J. S. Beis and D. G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *CVPR*, pages 1000–1006, 1997. 19
2. S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *ICCV*, 1999. 3, 7, 8, 20, 21, 22
3. C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, August 2006. 3, 4
4. T. Boult and L. G. Brown. Factorization-based segmentation of motions. In *IEEE Workshop on Visual Motion*, 1991. 24, 29
5. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 2001. 6, 7, 9, 11, 12, 13
6. T.-J. Chin, H. Wang, and D. Suter. Robust fitting of multiple structures: The statistical learning approach. In *International Conference on Computer Vision (ICCV)*, 2009. 23
7. T.-J. Chin, J. Yu, and D. Suter. Accelerated hypothesis generation for multi-structure robust fitting. In *European Conference on Computer Vision (ECCV)*, 2010. 29
8. O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. *Pattern Recognition*, LNCS 2781:236–243, 2003. 13
9. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 2002. 2, 13
10. J. Costeira and T. Kanade. A multi-body factorization method for motion analysis. In *ICCV*, 1995. 24
11. A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast Approximate Energy Minization with Label Costs. *International Journal of Computer Vision (IJCV)*, 96(1):1–27, Jan. 2012. (earlier version: CVPR 2010). 3, 4, 8, 9, 13, 15, 18, 32
12. O. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images*. MIT Press, 2004. 19
13. M. A. Figueiredo and A. K. Jain. Unsupervised learning of finite mixture models. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 24(3):381–396, 2002. 3, 6
14. M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *CACM*, 1981. 2, 4, 9
15. A. Gruber and Y. Weiss. Incorporating non-motion cues into 3D motion segmentation. In *European Conference on Computer Vision (ECCV)*, 2006. 4
16. R. Hartley. In defense of the eight-point algorithm. *IEEE Transections on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1997. 26

17. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. 12, 19, 22, 23, 26
18. H. Isack. *Spatially Coherent Multi-Model Fitting*. MS Thesis, CS Dept., University of Western Ontario, London, Canada, April, 2009. 9, 11
19. Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision (IJCV)*, 3(1):73–102, May 1989. 8
20. H. Li. Two-view motion segmentation from linear programming relaxation. In *CVPR*, 2007. 3, 5, 7, 8, 25
21. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004. 18, 19, 22
22. Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3D Vision: From Images to Geometric Models*. Springer Verlag, 2003. 25
23. M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP*, 2009. 19
24. C. Olsson, O. Enqvist, and F. Kahl. A polynomial-time bound for matching and registration with outliers. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, USA, 2008. 4
25. C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (SIGGRAPH)*, August 2004. 7, 8
26. K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers through model selection. *IEEE Transections on Pattern Analysis and Machine Intelligence*, 28(6):983–995, June 2006. 8, 29
27. R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *ECCV*, 2008. 2, 13, 22, 23, 24
28. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 1992. 24, 29
29. P. Torr and A. Zisserman. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Journal of Computer Vision and Image Understanding*, 78(1):138156, 2000. 4
30. P. H. S. Torr. Geometric Motion Segmentation and Model Selection. *Philosophical Trans. of the Royal Society A*, pages 1321–1340, 1998. 2, 3, 5, 7, 8, 15, 29, 32
31. P. H. S. Torr and D. W. Murray. Stochastic motion clustering. In *European Conference on Computer Vision (ECCV)*, volume LNCS 801, pages 328–337, Stockholm, Sweden, 1994. 3, 7, 8
32. R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *CVPR*, 2007. 26, 27, 28, 29, 30, 31
33. R. Vidal, R. Tron, and R. Hartley. Multiframe motion segmentation with missing data using powerfactorization and GPCA. *IJCV*, 2008. 26, 29, 30, 31
34. E. Vincent and R. Laganiere. Detecting planar homographies in an image pair. In *ISPA*, June 2001. 2, 7, 9, 16
35. J. Wills, S. Agarwal, and S. Belongie. What went where. In *CVPR03*, pages I: 37–44, 2003. 7
36. J. Yan and M. Pollefeys. A general framework for motion segmentation: independent, articulated, rigid, non-rigid, degenerate, and non-degenerate. In *European Conference on Computer Vision (ECCV)*, 2006. 26, 29, 30, 31
37. R. Zabih and V. Kolmogorov. Spatially Coherent Clustering with Graph Cuts. In *CVPR*, June 2004. 7, 8
38. S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996. 7, 8, 15
39. R. Zrour, Y. Kenmochi, H. Talbot, L. Buzer, Y. Hamam, I. Shimizu, and A. Sugimoto. Optimal consensus set for digital line and plane fitting. *International Journal of Imaging Systems and Technology*, 21:45–57, 2011. 4
40. M. Zuliani, C. Kenney, and B. Manjunath. The multiransac algorithm and its application to detect planar homographies. In *ICIP*, 2005. 2, 9, 13, 16, 23