Multiple Models Fitting as a Set Coverage Problem

Luca Magri
Computer Science Dept. - University of Verona
Strada Le Grazie, 15 - 37134 Verona, IT
magri.luca.l@gmail.com

Andrea Fusiello
DPIA - University of Udine
Via delle Scienze, 208, - 33100 Udine, IT
name.surname@uniud.it

Abstract

This paper deals with the extraction of multiple models from noisy or outlier-contaminated data. We cast the multi-model fitting problem in terms of set coverage, deriving a simple and effective method that generalizes Ransac to multiple models and deals with intersecting structures and outliers in a straightforward and principled manner, while avoiding the typical shortcomings of sequential approaches and those of clustering. The method compares favorably against the state-of-the-art on simulated and publicly available real data-sets.

1. Introduction

Finding a model (or structure) that fits data corrupted by noise and outliers is an omnipresent problem in empirical sciences, including Computer Vision. When multiple instances of the same structure are present in the data, the problem has a chicken-and-egg pattern: in order to estimate models one needs to first segment the data, and in order to segment the data it is necessary to know which model points belong to. Moreover, the presence of multiple structures strains robust estimation, because, in addition to rogue points, the outliers to a structure of interest are all the inliers to the other structures.

Among the wide range of methods proposed in Computer Vision to address the challenge of multiple models geometric fitting, the analysis of *consensus* together with its counterpart, the analysis of *preferences*, can be recognized as leitmotifs recurring throughout the extensive literature on the subject. The *consensus set* of a model is simply defined as the set of points that are inliers to that model. Dually, the *preference set* of a point is the set of models to which that point is inlier. Most of the multi-model fitting techniques proposed in the literature can be ascribed to one of these two concepts, according to which horn of the chicken-egg-dilemma is addressed first.

Consensus-based algorithms put the emphasis on the estimation part and focus on models that describe as many points as possible. On the other hand, preference approaches concentrate on the segmentation side of the problem, and aim at finding a proper partition of the data, from which model estimation follows.

Both approaches conceptually work on the consensus/preference matrix P defined as

$$P(i,j) = \begin{cases} 1 & \text{if } \mathrm{err}(x_i, \theta_j) < \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $x_i \in X$ are data points, $\theta_j \in H$ tentative structures, err a suitable error function and ϵ the inlier threshold. The binary matrix P can be interpreted in several ways. It can be regarded as the incidence matrix of an hyper-graph where rows correspond to vertices and columns represent hyper-edges; alternatively its rows, identified with preference sets, can be interpreted as representations of data in high dimensional spaces. In both cases multi-model fitting boils down to cluster analysis. Changing the perspective, columns of P can be interpreted as consensus sets, whose cardinality is to be maximized.

In the remaining of this section we shall track down the path that, starting from consensus throughout preference analysis, have been followed in the literature to address the challenges of multiple structures recovery. For a review of multi-model fitting from the perspective of the optimization of a global energy functional the reader is referred to [9].

1.1. Consensus analysis

Consensus analysis stands out as one of the first efforts to address robust model estimation. The methods belonging to this category follow a common paradigm. At first the space Θ of all the feasible structures is approximated as a suitable finite hypothesis space H in different ways. Then a voting procedure elects the structures in H that best explain the data in terms of consensus.

The idea of exploiting consensus is at the core of the celebrated Ransac (Random Sample Consensus) and its variants (see [20] and references therein). A straightforward generalization to multiple models is Sequential Ransac

[29, 33], an iterative, greedy algorithm that executes Ransac many times and removes the found inliers from the data as each structure is detected. As a consequence, inaccurate detections at early stages of the algorithm can heavily deteriorate the results; in addition, points in the intersections do not contribute to the sampling of subsequent structures. As such, this strategy is inherently prone to achieve suboptimal segmentation. A parallel scheme, dubbed Multi-Ransac, has been proposed in [39] in the endeavor to mitigate its greediness. This method, however, falls short of dealing with intersecting models.

The popular Hough transform and its randomized version [35] can be regarded as consensus-oriented algorithms too. A more general approach consists in finding modes directly in Θ (*e.g.* [25]). In this way the difficulties of the quantization step are alleviated by mapping the data into the parameter space through random sampling and then by seeking the modes (*e.g.* with with mean-shift [3]).

In all these consensus based methods, alongside the voting phase, the approximation of Θ is a recurring and tricky issue. The crucial point is that, when multiple structures are hidden in the data, consensus oriented algorithms have to disambiguate between genuine structures and redundant ones, *i.e.* instances of the same model with slightly different parameter. This issue is addressed by enforcing several disjointedness criteria, either explicitly or implicitly by different approximations of the solution space.

For instance, Hough transform handles redundancy by capturing similar structures in the same equivalence class via the quantization of Θ. Along the same line, the bandwidth used in mean shift can be thought as a softer way to localize and aggregate redundant models. Also Sequential Ransac and Multi-Ransac enforce disjointedness by avoiding to sample similar models [7]. As regards Sequential Ransac, this idea can be identified in the iterative removal of the discovered inliers and in the subsequent sampling of the hypotheses on the remaining data. In Multi-Ransac this is more evident, since this algorithm explicitly search for the best collection of k disjoint models. In practice, however, using consensus as the only criterion seems short-sighted, for true models can have mutual intersections greater than redundant ones, hence the algorithm would fail in discerning authentic structures.

In order to overcome the drawbacks inherent to consensus methods, the problem has been tackled from a different point of view, where the role of data and models are reverted: rather than representing models and inspecting which points support them, points are described by the *preference* they grant to models.

1.2. Preference analysis

The idea of describing point by their residuals can be traced back to [38] where the residuals distributions of individual points, with respect to a set of putative structures randomly sampled, is analyzed. In particular, the most significant structures are revealed as peaks in the histograms of the residuals. In addition, the number of models is automatically determined by the median number of modes found over all data points. In practice, the mode-finding step of this strategy suffers of low accuracy and depends critically on the bin size adopted.

Building on this idea, J-Linkage algorithm [28] was the first successful application of a preference-based representation of data. A two steps *first-represent-then-cluster* scheme is implemented: data are represented by the votes they grant to a set of model hypotheses, then a greedy agglomerative clustering is performed to obtain a partition of the data.

Several elements in common with previous methods can be recognized: an inlier threshold ϵ is used as in Ransac and the idea of casting points' votes echoes Randomize Hough Transform. Despite that, J-Linkage does not rely on a quantized space, which causes the shortcoming of Hough Transform, nor on the residual space, which leads to the difficulties of modes estimation, but explicitly introduces a *conceptual space* where points are portrayed by the *preferences* they have accorded to random provisional models. The changes of perspective entailed by preference analysis results in a different approach to the chicken-&-egg dilemma. Structures are recognized as groups of neighboring points in the conceptual space therefore the emphasis is shifted from the estimation to the segmentation part of the problem. T-Linkage [15] extends this idea by relaxing the notion of binary preference set allowing the use of soft votes to depict points preference more accurately.

Along the same line of J-Linkage, Kernel Fitting (KF) [2], Robust Preference Analysis [16] (RPA) and Random Cluster Model Simulated Annealing (RCMSA) [19] exploits points preferences.

KF and RPA first derive a kernel matrix to measure agreement between preferences, then a (different) transformation is applied in order to detect and remove outliers. Then the cleaned kernel matrix is used by KF to oversegment the remaining inliers and reassemble the structures with a merging scheme. RPA performs symmetric non negative factorization on the cleaned kernel matrix in order to extract the most representative sampled models. Robust statistic is then employed to assign the data to the recovered structures. RCMSA [19] organizes point preferences in a weighted graph and the multi-model fitting task is stated as a graph cut problem which is solved efficiently in an annealing framework.

Finally, we can ascribe to preference analysis also all the approaches based on higher order clustering [1, 8, 10, 37], where higher order similarity tensors are defined between n-tuple of points as the probability of points to be clustered

together measured in terms of residual errors with respect to provisional models.

1.3. Shortcoming of preference approaches

Indubitably, a preference-based approach has the great advantage of casting specific multi-model fitting problems in a very general clustering framework. Nevertheless it has been largely recognized by the research community that the segmentation/clustering problem is ill-posed, and that a "*no free lunch theorem*" [34] holds, which states that a given clustering method can be optimal only with respect to some specific type of data-set.

Moreover, Kleinberg [12] confirms that clustering techniques are inherently fraught with ambiguities: he conceives an axiomatic theory in which he defines three desirable properties that a clustering scheme ought to satisfy, namely scale-invariance, a "richness" condition that all partitions are achievable, and a consistency requirement on the shrinking and stretching of distances. In that setting an "*impossibility theorem*" is derived, demonstrating that there is no clustering function satisfying simultaneously all the three properties.

In addition, two other main issues are not satisfactorily handled by clustering techniques. In first instance, classical clustering approaches are designed to yield a *partition* of the data, hence they are not suitable for dealing explicitly with intersecting structures. As a result, intersections are either ignored or dealt indirectly with *ad hoc* post processing on the output.

In second place, the treatment reserved to outliers is not completely sound. For estimation purposes, gross outliers ought to fall in a special group of points, but clustering treats all the segments in the same way. This is the reason why partitional clustering schemes are not able to enforce robustness by simply throwing-in one additional model with the hope that outliers will be clustered together. Hierarchical methods in practice are more resilient to outliers, still they do not have a specific treatment during the clustering phase: for example T-Linkage relies on *a posteriori* specific heuristics to ensure robustness.

2. Multi-model fitting as a coverage problem

For all the reasons described in the previous section, in this work we sidestep the pitfalls of clustering and focus on the objective of maximizing consensus. However, in doing this, we shall counteract the shortcomings of Sequential Ransac and its relatives, namely greediness and poor handling of intersecting models. These requirements will naturally lead to a coverage formulation, which will be referred to as "Random sample Coverage", or RansaCov.

Let us start by assuming that all the points $x_i \in X$ are inliers (the case of outliers will be dealt with later on). This is equivalent to state that all the points are explained by some structures, in other words, the true structures determine, by means of their consensus sets, a *cover* of the data, *i.e*. a collection of sets whose union contains X:

$$F = \{S_j : j \in J\} \quad \text{such that} \quad X \subseteq \bigcup_{j \in J} S_j, \tag{2}$$

Note that we are not requiring that these sets are disjoint, so we are not limited to partitions and we can properly handle the case of intersecting models. By invoking the Occam's principle, a straightforward formulation is therefore to ask for a cover consisting of a minimal number of consensus sets. In this way we are implicitly discouraging redundant models. Thus we are naturally led to the following SET COVER problem.

Definition 2.1 (SET COVER)**.** Given a ground set X and $F = \{S_1, \ldots, S_m\}$ a cover of X, select the minimum number of subsets from F that covers X.

In this formulation, $X = \{x_1, \ldots, x_n\}$ contains the data points and the collection $F = \{S_1, \ldots, S_m\}$ is composed by the consensus sets of the sampled models $\theta_1, \ldots, \theta_m \in H$: *i.e*. $S_j = \{x \in X : \mathrm{err}(x, \theta_j) < \epsilon\}$ instantiated on minimal sample sets as in Ransac. The property that F is a cover of X can be easily enforced by requiring that every points of X is sampled at least once. SET COVER can be rephrased rigorously using the matrix P in the constraints formulation and introducing m binary variables $z_j \in \{0, 1\}$ for each subset S_j. If S_j is selected in the solution then $z_j = 1$, otherwise $z_j = 0$. In this way SET COVER can be shown to be equivalent to an Integer Linear Programming (ILP) problem:

$$\text{minimize} \quad \sum_{j=1}^{m} z_j \quad \text{subject to} \quad Pz \geq \mathbf{1}. \tag{3}$$

The constraint can be expanded as

$$\sum_{j : S_j \ni x_i} z_j \geq 1 \;\; \forall x_i \in X \tag{4}$$

where it becomes clear that it is meant to ensures that the solution $\{S_j\}_{j:z_j=1}$ is a cover of X.

If X is corrupted by rogue points we can integrate them in the formulation of the problem at the cost of introducing an additional parameter k equal to the desired number of structures. Requiring some extra information to deal with outliers seems to be unavoidable. In this respect, k is a more guessable parameter than others.

Instead of trying to find the smallest number of sets that cover all elements, we search for the largest number of points that can be covered by k sets, possibly leaving some points (the outliers) uncovered. This leads to the so called MAXIMUM COVERAGE problem.

Definition 2.2 (MAXIMUM COVERAGE)**.** Given a ground set X, $F = \{S_1, \dots, S_m\}$ a collection of subsets of X and an integer k, select from F at most k subsets that cover the maximum number of points in X.

This problem is translated in an ILP one thanks to a collection of n auxiliary variables y_i, such that $y_i = 1$ if x_i belongs to the returned subsets, 0 otherwise:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} y_i \\ \text{subject to} \quad & \sum_{j=1}^{m} z_j \leq k \\ & \sum_{j:S_j \ni x_i} z_j \geq y_i \;\; \forall x_i \in X \\ & 0 \leq y_i \leq 1, \quad z_j \in \{0,1\}. \end{aligned} \tag{5}$$

The first condition enforces that no more than k sets are picked and the second constraint ensures that if $y_i \geq 0$ then at least one set $S_j \ni x_i$ is selected.

The following preprocessing is applied to the input collection of sets. First of all, keeping in mind that our aim is to maximize consensus, we refit a structure to each consensus set via least squares, and, if the consensus has increased, we update the structure and its supporting points. The remaining sets are hence ordered by cardinality $|S_1| \geq |S_2| \geq \dots \geq |S_k|$ and a set S_j is discarded if

$$S_j \subseteq \bigcup_{i=1}^{j-1} S_i. \tag{6}$$

The rationale of this choice is to keep only those structures that cover at least a point that otherwise would be uncovered by the union of larger ones. Please note that in particular we are deleting subsets that are contained in one larger set.

SET COVER and MAXIMUM COVERAGE are long known to be NP-hard [11]: not surprisingly, since the inherent complexity of multi-model fitting does not disappear by simply rephrasing it in different terms. Nevertheless, these optimization problems are among the oldest, most studied and widespread ones in the mathematical programming literature. Therefore we can reap the outcomes of the efforts made by the scientific community in addressing this issues, and enjoy the fruits of several studies focused on approximating the solutions of these problems.

For example, the greedy strategy – hereinafter **Greedy-RansaCov** – which keeps choosing the set that covers most uncovered points until they all are covered, embodies the spirit of Sequential Ransac with the only differences that the hypothesis space is not sampled iteratively[1] and, instead of returning a partition, intersecting segments are allowed. It has been demonstrated by Feige [6] that this greedy strategy is the best possible in terms of approximation ratio. More precisely an approximation of $H(n)$ holds in the case of SET COVER problem (where $H(n)$ denotes the n-th harmonic number), and $1-1/e$ for the MAXIMUM COVERAGE problem. This result applies effortless to Greedy-RansaCov giving a provable quality measure of the solution.

Another straightforward strategy consist in solving a relaxed Linear Programming (LP) problem and converting the solution by rounding up all non-zero variables to 1. In this case [32] shows that the solution achieves an approximation guarantee of a factor equal to the the frequency of the most frequent point, where the frequency of a point is the number of sets that cover that point. Our preprocessing step, besides refining the models, improves the approximation factor of the relaxed LP solution, for it actually reduces the maximal frequency of the points.

In practice, more sophisticated strategies are used by ILP solvers, but the relaxed LP solution is a good starting point. Our algorithm – dubbed **ILP-RansaCov** – solves (5) using the `intlinprog` function of MATLAB, which attempts to tighten the LP relaxation with several heuristics and falls back to branch and bound in case of failure.

Comparison with FACILITY LOCATION. The closest methods to ours in the literature are those casting multi-model fitting as a FACILITY LOCATION (FL) problem: provided a set of potential facilities (which corresponds to the pool of tentative structures), FL selects an optimal subset of facilities and assigns customers (*i.e.* data points) to one facility each, so as to minimize the sum of facility opening costs and the distances between customers to their assigned facilities. This leads to the optimization of a cost function composed by two terms: a modeling error – *i.e.* customers-facility distances – which can be interpreted as a likelihood term, and a penalty term to encode model complexity – the cost to open the facilities – mimicking classical MAP-MRF objectives. Some authors solves it with ILP [30, 14, 22, 13] while others propose different combinatorial optimization techniques [9, 36, 4, 18]. Although SET COVER and FL are related (the first can be rephrased as a special case of the second) and ILP has been used to solve both, ILP-RansaCov differs from previous work based on FL in many respects.

In first instance, FL needs to guess a correct trade-off between data fidelity and model complexity, in order to strike the proper balance between over and under fitting. For example [4, 9] regularizes the modeling-fitting error, expressed in terms of residual, by introducing a label costs that penalizes the number of different structures, whereas [13], aimed at fitting subspace of different dimensions to outlier-free data, exploits a penalty term encoding subspace dimension. In contrast, our formulation elude this thorny trade-off: in the outlier-free scenario SET COVER regular-

[1] In Sequential Ransac columns of P are generated sequentially: once a structure of inlier is detected, its supporting points are removed and successive hypotheses are sampled from the remaining of the data.

izes the solution invoking the minimality of cover, while, if outliers are present, MAXIMUM COVERAGE requires the maximum number of models as a clear, intelligible parameter, instead of balancing two incommensurable quantity in the cost function.

Second, FL minimizes the fitting error on the continuum of residuals, in the same spirit of MLE estimators, while ILP-RansaCov gains resiliency to outliers by maximizing the consensus *à la* Ransac. This, however, comes at the price of assuming that all the structures have the same error scale, while MLE-like estimators can compute the scale along the parameters of each model.

In our formulation the rogue points will be simply left uncovered by MAXIMUM COVERAGE, whereas FL copes with outliers either by introducing a special additional model for which a constant fidelity measure has to be manually tuned [9], or by requiring an upper bound to the total number of outliers [14].

Finally, FL approaches enforce hard-membership constraints, producing a partition of the data, whereas ILP-RansaCov inherently caters for intersecting solutions.

3. Experiments on simulated data

In this section we investigate the performance of ILP-RansaCov with respect to Greedy-RansaCov (which emulates Sequential Ransac), J-Linkage [28] and T-Linkage [15] on synthetic data, using *the same sampling* and the same inlier threshold for all the methods (or, equivalently, the same P matrix). We obtained the implementations of J-Linkage and T-Linkage from [27]. The MATLAB code of ILP-RansaCov is available on the web[2].

The data sets (Fig. 1) consist of segments in several configurations and circles, as in [28]. Each structure consists of 50 inliers points, contaminated by Gaussian noise and outlying points in different percentages (reported in Tab. 1). All the methods have been provided with the correct number of structures k; in the case of J-Linkage and T-Linkage, the largest k structures produced by the algorithms are considered.

The results are collected in Fig. 1 while Tab. 1 reports the misclassification errors (ME), computed as follows: first the map between ground-truth labels and estimated ones that minimize the overall number of misclassified points is found (as in [23]), then a point is deemed as correct if one of its labels corresponds to the ground-truth. The ME is the percentage of misclassified points .

First of all we can notice that in the *Stair4* experiment (firstly used in [39] to criticize Sequential Ransac), Greedy-RansaCov performs poorly: the shortcomings of this greedy strategy are here afoot: the incorrect selection of the first structure compromises the subsequent interpretation of the data. A greedy approach to the MAXIMUM COVERAGE problem yields a sub-optimal segmentation also on the *Circle4* data-set, where one of the four structures is over-segmented by Greedy-RansaCov at the expense of the smaller circle in the center.

[2] www.diegm.uniud.it/fusiello/demo/cov/

	outliers	J-Lnkg	T-Lnkg	Grdy-RansaCov	ILP-RansaCov
Stair4	50%	10.20	**10.00**	39.20	12.00
Star5	60%	15.20	14.40	10.40	**3.80**
Star11	50%	35.00	33.09	32.36	**25.18**
Circle4	50%	26.50	23.00	30.25	**11.25**
mean		20.12	20.12	28.05	**13.06**

Table 1: Misclassification error (ME %) on **simulated data.**

On *Star11* J-linkage misses a ground truth segment. During the merging process some inliers are incorrectly aggregated to spurious models, hence the recovered segment that actually corresponds to a ground truth structure collects fewer inliers, to the point that it falls outside the first k largest models and is deemed as outlier. In general the tendency of loosing inliers during the segmentation step affects J-Linkage (and T-Linkage) also in the other data-sets, *e.g.* it is particularly evident on *Circle4*,

Even when the discovered inliers are enough to recover the corresponding structures, this behavior has a detrimental effect on the model estimate, for it increases the variance.

ILP-RansaCov yields reliable segmentations in all the experiments, and it achieves the best average ME. The reason can be ascribed to the non-greedy minimization strategy (w.r.t. Greedy-RansaCov) and to the departure from the partitioning paradigm (w.r.t. J-Linkage and T-Linkage). As a matter of fact, when models do not intersect, as in *Stair4*, the performance of J-Linkage and T-Linkage are in the same range of ILP-RansaCov.

4. Experiments on real data

In this section, we demonstrate the performance of ILP-RansaCov on three classical Computer Vision applications, namely: i) vanishing point detection, ii) video motion segmentation, and iii) two-views segmentation. In all these scenarios we compare ILP-RansaCov with J-Linkage [28], T-linkage [15] and RPA [16], whose implementation is taken from [21]. In addition, one reference method has been added to the comparison for each specific scenario, namely: MFIPG [18] in the vanishing point experiments, SSC [23] for video motion segmentation and RCMSA [19] for two-views segmentation. These methods have been selected because i) they are among the best performers, ii) the original code from the authors is available (MFIPG [17], RCMSA [17], SSC [24]), and iii) they have been tested on the same respective data-sets.

MFIPG and RCMSA are considered only in one scenario

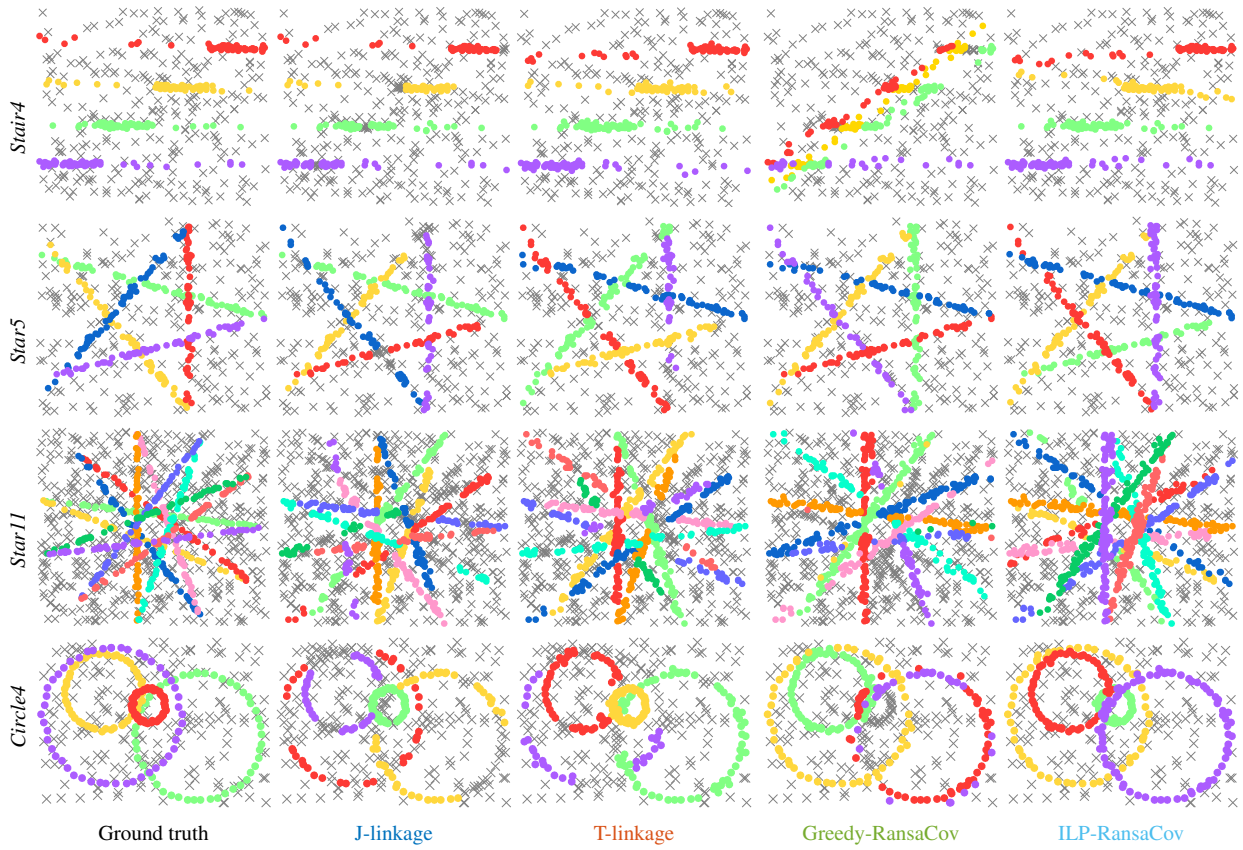


Figure 1: Comparison on **simulated data** (outliers marked as x).

out of three because the authors provided the tuning parameters only for that particular application (vanishing point detection and two-views segmentation, respectively). SSC instead is tailored specifically to subspace segmentation, hence it cannot be applied in the other two cases (where models are not linear or affine subspaces).

All the algorithms but SSC and RCMSA were provided with the same pool of putative models, sampled as in [16].

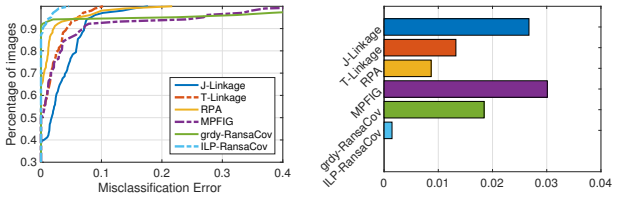
Vanishing point detection. In this experiment we compare the performances of ILP-RansaCov with MFIPG on vanishing point detection using the York Urban Line Segment Database [5], or York Urban DB in short, a collection of 102 images of architectural Manhattan-like environments (*i.e.* scenes dominated by two or three mutually orthogonal vanishing directions). Annotated line-segments that match with the 3-d orthogonal frame of the urban scene are provided with the ground-truth, no outliers are present in the data. The aim is to group the supplied segments in order to recover two or three orthogonal vanishing points.

MFIPG (Model-Fitting- with-Interacting-Geometric-Priors) is a recently proposed method that improves on PeARL [4] adding high-level geometric priors. In particular, in this application, an additional term expressing interaction between vanishing points is included into the FL formulation, to promote the extraction of orthogonal vanishing points. The global input parameters recommended in the original paper have been optimized for each single image to enhance the results.

Figure 2 shows three images where ILP-RansaCov achieved the worst ME, which are nevertheless qualitatively correct. Figure 3(a) reports the cumulative distribution of the ME per sequence, *i.e.* the value on the ordinate corresponds to the percentage number of sequences where the algorithm achieved a ME lower than the abscissa. The differences among the methods can be better appreciated by plotting the area above the cumulative distribution of ME (Fig. 3(b)) or by analyzing the average and median ME, collated in Tab. 2. These quantitative results confirm that ILP-RansaCov is the most accurate, followed by RPA. As MFIGP enhances PeARL, figures in Tab. 2 indirectly corroborate the advantage of ILP-RansaCov over PeARL. It is worth noting that Greedy-RansaCov, a proxy of the vilified Sequential Ransac, performs better than other sophisticated methods, in this task.

Video motion segmentation. In this experiments we considered Sparse Subspace Clustering [23] a state-of-the-art

Figure 2: A sample of the worst ILP-RansaCov results on YorkUrbanDB (**vanishing point detection**). Line membership is color-coded.



(a) Cumulative ME (b) Area above the cumulative ME

Figure 3: Results on YorkUrbanDB. (a) is the cumulative distributions of the errors per sequence; (b) shows the area above the curve (the smaller the better).

	J-Lnkg	T-Lnkg	RPA	MFIGP	Grdy-RansaCov	ILP-RansaCov
Mean	2.85	1.44	1.08	3.51	2.38	**0.19**
Med	1.80	**0.00**	**0.00**	0.16	0.00	**0.00**

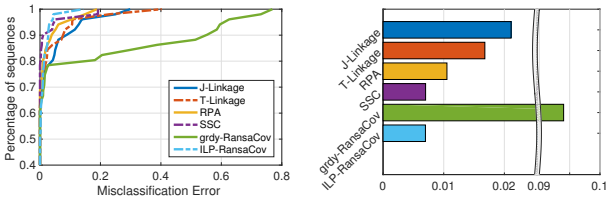
Table 2: Misclassification error (ME %) on YorkUrbanDB.

technique that exploits a sparse representation to build an affinity matrix, which in turns is segmented by spectral clustering. The input data is a set of features trajectories across a video taken by a moving camera, and the aim is to recover the different rigid-bodies. We use the 51 *real* video sequences from the Hopkins 155 data-set [31], each containing two or three moving objects, with no outliers. Following [26], in order to deal with degenerate motions, we project the data onto an affine 4-d space where the rigid-body segmentation is translated in a 3-d plane fitting problem.

Figure 4 reports some sample results, in particular three sequences belonging to *Traffic 2* and *Others 3* subsets, respectively, where ILP-RansaCov achieves sub-optimal segmentations. Figure 5 and Tab. 3 provide a comparison of the performances in terms of ME: ILP-RansaCov places in the same range of SSC and achieves the best overall results. In this case the advantage of solving the MAXIMUM COVERAGE problem with a global approach is afoot, since the greedy strategy of Greedy-RansaCov, sampling being equal, fails. Please note that, via [13], this experiment provides an indirect comparison with FLoSS.

Two-views segmentation. In this experiment we additionally compare ILP-RansaCov against RCMSA [19] on the Adelaide Robust Model Fitting Data Set, or AdelaideRMF in short, which consists of 38 image pairs, 19 related by multiple homographies (H) and 19 by multiple fundamental matrices (F), with outliers. The task involves segmenting different planes/moving objects by fitting homographies/fundamental matrices to subsets of corresponding points.

Figure 4: A sample of the worst ILP-RansaCov results on Hopkins155 (**video motion segmentation**). Point membership is color-coded.



(a) Cumulative ME (b) Area above the cumulative ME

Figure 5: Results on Hopkins155. (a) is the cumulative distributions of the errors per sequence; (b) shows the area above the curve (the smaller the better).

		J-Lnkg	T-Lnkg	RPA	SSC	Grdy-RansaCov	ILP-RansaCov
Traffic 3	Mean	1.58	0.48	**0.19**	0.76	28.65	0.35
	Med	0.34	0.19	**0.00**	**0.00**	1.53	0.19
Traffic 2	Mean	1.75	1.31	0.14	**0.06**	7.48	0.54
	Med	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Others 3	Mean	6.91	5.32	9.11	**2.13**	14.89	**2.13**
	Med	6.91	5.32	9.11	**2.13**	14.89	**2.13**
others 2	Mean	5.32	6.47	4.41	3.95	8.57	**2.40**
	Med	1.30	2.38	2.44	**0.00**	0.20	1.30
All	Mean	2.70	2.47	1.42	1.08	10.91	**0.98**
	Med	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0,00**

Table 3: Misclassification error (ME %) on Hopkins155.

All the methods are given the inlier threshold computed from the available ground truth.

Some failure examples are reported in Fig. 6. The left image is an example of under-segmentation, where a unique fundamental matrix explains both the cube and the toy (red points). In the middle image ILP-RansaCov fails in detecting one planar structure (second wall of the building from the left). In the right image the campanile (on the very right) is over-segmented, and this consumes one of the available k models, thereby preventing the nearby wall to be detected.

From the data reported in Fig. 7 and Tab. 4, the reader can appreciate that the ME of ILP-RansaCov is consistently lower than RCMSA and in the same range of RPA.

In order to evaluate the relative importance of multi-

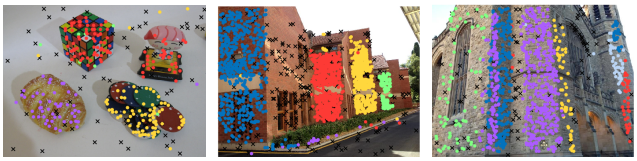
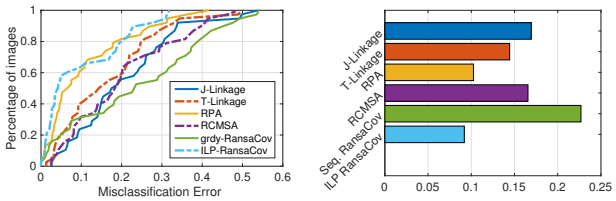


Figure 6: A sample of the worst ILP-RansaCov results on AdelaideRMF (**two-views segmentation**). Point membership is color-coded, black crosses are points outliers.



(a) Cumulative ME (b) Area above the cumulative ME

Figure 7: Results on AdelaideRMF. (a) is the cumulative distributions of the errors per sequence; (b) shows the area above the curve (the smaller the better).

		J-Lnkg	T-Lnkg	RPA	RCMSA	Grdy-RansaCov	ILP-RansaCov
F	Mean	16.43	9.37	**5.49**	12.37	17.08	6.04
	Med	14.29	7.80	4.57	9.87	21.65	**4.27**
H	Mean	25.50	24.66	17.20	28.30	26.85	**12.91**
	Med	24.48	24.53	17.78	29.40	28.77	**12.34**

Table 4: Misclassification error (ME %) for motion segmentation (F) and plane segmentation (H) on AdelaideRMF.

ple membership w.r.t. the optimization method, we have rephrased Multi-Ransac in the framework of maximal coverage: the strategy is similar to Greedy-RansaCov, the difference being that, after a set is picked, the subsequent ones are searched among those having maximal Jaccard distance with the currently covered elements, thereby maximizing disjointedness. Even if a point can be assigned to multiple model, experiments demonstrated that the performances are consistently inferior to ILP-RansaCov (ME is: 2.97 for VP, 4.58 for video sequences, 17.01 for F and 26.85 for H), confirming the crucial role of the optimization technique.

Finally, we run an experiment to probe of how the execution time scales with the input dimension and where the time is spent. To this end, we run ILP-RansaCov on a line fitting problem extracted from *Star11* with variable number of sampled models and number of points. The execution times, broken down for each step, are reported in Fig. 8. The instantiation of the consensus/preference matrix dominates the complexity for moderate point number, whereas ILP takes over when the number of points increases. Also, while the dependence from the number of sampled models appears to be polynomial, the execution time grows exponentially with the number of points, in accordance with the-

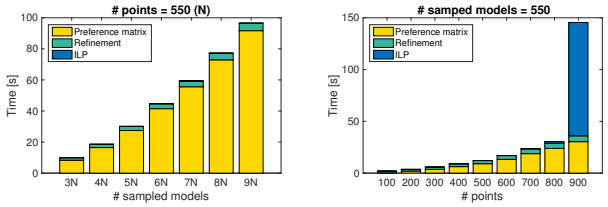


Figure 8: **Execution time** of ILP-RansaCov on simulated data w.r.t. the dimensions of the problem.

	YorkUrbanDB	Hopkins155	Adelaide (F)	Adelaide (H)
mean	8.09	41.19	52.24	146.34
median	1.14	11.56	48.79	51.12

Table 5: Execution time [s] of ILP-RansaCov on real data.

oretical prediction.

The impact of the preprocessing step, related to Eq. (6), is negligible in terms of the running time, but it improves the quality of the solution: *e.g.*, with reference to Tab. 4 (F), the mean ME of ILP-RansaCov *without this refinement* raises to 11.44.

To complete the picture on computational burden, we report in Tab. 5 the time spent by ILP-RansaCov in each experiment on real data. A comparison with other methods would have been meaningless for not all of them are coded in MATLAB as ours.

5. Conclusions

We formulated multi-model fitting in terms of SET COVER and MAXIMUM COVERAGE problems, yielding a simple and easy to implement method that generalizes Ransac to the case of multiple structures in a neat and principled manner.

As in previous work, the multi-model fitting problem is formulated in terms of optimization of a global cost function, thereby eluding the greediness of techniques such as Sequential/Multi-Ransac and J-linkage, but at the same time avoiding the difficult trade-off between data fidelity and complexity of other formulations, by resorting to consensus maximization. In both cases, we tackle the problem of intersecting models at the root, by replacing partitions with coverages.

ILP-RansaCov is modular with respect to the ILP solver and to the sampling strategy. Few intelligible parameters need to be set and tuned, namely the inlier threshold and the number of desired model.

In summary, we expect that this paper will offer practitioners a manageable tool for addressing a difficult and ubiquitous problem, and will provide the community a reference baseline for further advancements.

Acknowledgments. L. Magri gratefully acknowledge the support of 3Dflow srl.

References

[1] S. Agarwal, J. Lim, L. Zelnik-manor, P. Perona, D. Kriegman, and S. Belongie. Beyond pairwise clustering. In *Computer Vision and Pattern Recognition*, pages 838–845, 2005. 2

[2] T.-J. Chin, H. Wang, and D. Suter. Robust fitting of multiple structures: The statistical learning approach. In *International Conf. on Computer Vision*, pages 413–420. IEEE, 2009. 2

[3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002. 2

[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. 4, 6

[5] P. Denis, J. H. Elder, and F. J. Estrada. Efficient edge-based methods for estimating manhattan frames in urban imagery. In *European Conf. on Computer Vision*, pages 197–210, 2008. 6

[6] U. Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998. 4

[7] D. F. Fouhey. *Multi-model Estimation in the Presence of Outliers*. PhD thesis, Citeseer, 2011. 2

[8] V. M. Govindu. A Tensor Decomposition for Geometric Grouping and Segmentation. *Computer Vision and Pattern Recognition*, 1:1150–1157, 2005. 2

[9] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *International Journal of Computer Vision*, 97(2):123–147, 2012. 1, 4, 5

[10] S. Jain and V. M. Govindu. Efficient higher-order clustering on the grassmann manifold. In *International Conf. on Computer Vision*, 2013. 2

[11] R. M. Karp. *Reducibility among combinatorial problems*. Springer, 1972. 4

[12] J. Kleinberg. An impossibility theorem for clustering. *Advances in neural information processing systems*, pages 463–470, 2003. 3

[13] N. Lazic, I. Givoni, B. Frey, and P. Aarabi. Floss: Facility location for subspace segmentation. In *International Conf. on Computer Vision*, pages 825–832. IEEE, 2009. 4, 7

[14] H. Li. Two-view motion segmentation from linear programming relaxation. In *Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. 4, 5

[15] L. Magri and A. Fusiello. T-linkage: A continuous relaxation of j-linkage for multi-model fitting. In *Computer Vision and Pattern Recognition*, pages 3954–3961, June 2014. 2, 5

[16] L. Magri and A. Fusiello. Robust multiple model fitting with preference analysis and low-rank approximation. In *British Machine Vision Conference*, pages 20.1–20.12. BMVA Press, September 2015. 2, 5, 6

[17] http://cs.adelaide.edu.au/~trung. 5

[18] T. T. Pham, T.-J. Chin, K. Schindler, and D. Suter. Interacting geometric priors for robust multimodel fitting. *Transactions on Image Processing*, 23(10):4601–4610, 2014. 4, 5

[19] T.-T. Pham, T.-J. Chin, J. Yu, and D. Suter. The random cluster model for robust geometric fitting. *Pattern Analysis and Machine Intelligence*, 36(8):1658–1671, 2014. 2, 5, 7

[20] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J. Frahm. Usac: a universal framework for random sample consensus. *Pattern Analysis and Machine Intelligence*, 35(8):2022–2038, 2013. 1

[21] http://www.diegm.uniud.it/fusiello/demo/rpa/. 5

[22] K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers through model selection. *Pattern Analysis and Machine Intelligence*, 28(6):983–995, 2006. 4

[23] M. Soltanolkotabi, E. Elhamifar, and E. J. Candès. Robust subspace clustering. *CoRR*, abs/1301.2603, 2013. 5, 6

[24] http://www.vision.jhu.edu/code/. 5

[25] R. Subbarao and P. Meer. Nonlinear mean shift for clustering over analytic manifolds. In *Computer Vision and Pattern Recognition*, volume 1, pages 1168–1175. IEEE, 2006. 2

[26] Y. Sugaya, Y. Matsushita, and K. Kanatani. Removing mistracking of multibody motion video database hopkins155. In *British Machine Vision Conference*, 2013. 7

[27] http://www.diegm.uniud.it/fusiello/demo/jlk/. 5

[28] R. Toldo and A. Fusiello. Robust multiple structures estimation with J-Linkage. In *European Conf. on Computer Vision*, volume 5302, pages 537–547, 2008. 2, 5

[29] 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. 2

[30] P. H. Torr and D. W. Murray. Stochastic motion clustering. In *European Conf. on Computer Vision*, pages 328–337. Springer, 1994. 4

[31] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. 7

[32] V. V. Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013. 4

[33] E. Vincent and R. Laganiére. Detecting planar homographies in an image pair. In *Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis*, pages 182–187, 2001. 2

[34] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *Transaction on evolutionary computation*, 1(1):67–82, 1997. 3

[35] L. Xu, E. Oja, and P. Kultanen. A new curve detection method: randomized Hough transform (RHT). *Pattern Recognition Letters*, 11(5):331–338, 1990. 2

[36] J. Yu, T. Chin, and D. Suter. A global optimization approach to robust multi-model fitting. In *Computer Vision and Pattern Recognition*, 2011. 4

[37] R. Zass and A. Shashua. A unifying approach to hard and probabilistic clustering. In *International Conf. on Computer Vision*, volume 1, pages 294–301, 2005. 2

[38] W. Zhang and J. Kosecká. Nonparametric estimation of multiple structures with outliers. In *European Conf. on Computer Vision*, volume 4358, pages 60–74, 2006. 2

[39] M. Zuliani, C. S. Kenney, and B. S. Manjunath. The multi-RANSAC algorithm and its application to detect planar homographies. In *International Conf. on Image Processing*, 2005. 2, 5