

Progressive-X: Efficient, Anytime, Multi-Model Fitting Algorithm

Daniel Barath^{1,2}, and Jiri Matas¹

¹ VRG, Department of Cybernetics, Czech Technical University in Prague, Czech Republic

² Machine Perception Research Laboratory, MTA SZTAKI, Budapest, Hungary

barath.daniel@sztaki.mta.hu

Abstract

The *Progressive-X* algorithm, *Prog-X* in short, is proposed for geometric multi-model fitting. The method interleaves sampling and consolidation of the current data interpretation via repetitive hypothesis proposal, fast rejection, and integration of the new hypothesis into the kept instance set by labeling energy minimization. Due to exploring the data progressively, the method has several beneficial properties compared with the state-of-the-art. First, a clear criterion, adopted from RANSAC, controls the termination and stops the algorithm when the probability of finding a new model with a reasonable number of inliers falls below a threshold. Second, *Prog-X* is an anytime algorithm. Thus, whenever is interrupted, e.g. due to a time limit, the returned instances cover real and, likely, the most dominant ones. The method is superior to the state-of-the-art in terms of accuracy in both synthetic experiments and on publicly available real-world datasets for homography, two-view motion, and motion segmentation.

1. Introduction

The multi-class multi-model fitting problem is to interpret a set of input points as the mixture of noisy observations originating from multiple model instances which are not necessarily of the same class. Examples of this problem are the estimation of k planes and l spheres in a 3D point cloud; multiple line segments and circles in 2D edge maps; a number of homographies or fundamental matrices from point correspondences; or multiple motions in videos. Robustness is achieved by considering the assignment to one or multiple outlier classes.

Multi-model fitting has been studied since the early sixties. The Hough-transform [13, 14] perhaps is the first popular method for finding multiple instances of a single class [10, 21, 24, 33]. The RANSAC [9] algorithm was extended as well to deal with multiple instances. Sequential RANSAC [29, 17] detects instances one after another by searching in the point set from which the inliers of the de-

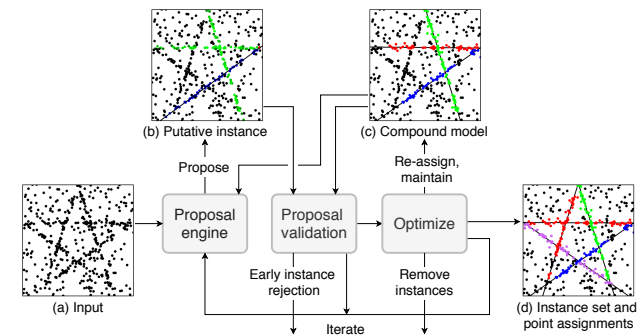


Figure 1: The Prog-X multi-model fitting pipeline gets a point set as input (a). It then repeats three steps. *First*, proposal of putative hypotheses (b) considering the active instance set (c), i.e. the compound model. In (b), the blue points are inliers of the putative instance, and the green ones are that of the compound one. *Second*, fast rejection of redundant hypotheses. *Third*, optimization by integrating new hypotheses, re-assigning points, maintaining model parameters, and removing unnecessary instances. In (c) and (d), the color codes the assignment to instances.

tected instances have been removed. The greedy approach makes RANSAC a powerful tool for finding a single instance, is a drawback in multi-instance setting. Points are assigned not to the best but to the first instance, typically the one with the largest support, for which they cannot be deemed outliers. MultiRANSAC [35] forms compound hypotheses about n instances. Besides requiring the number n of the instances to be known a priori, the processing time is high, since samples of size n times m are drawn in each iteration, where m is the number of points required for estimating a model instance. Nevertheless, RANSAC-based approaches have a desirable property of interleaving the hypothesis generation and verification steps. Moreover, they have a justifiable termination criterion based on the inlier-outlier ratio in the data which provides a probabilistic guarantee of finding the best instance.

Recent approaches for multi-model fitting [32, 15, 23,

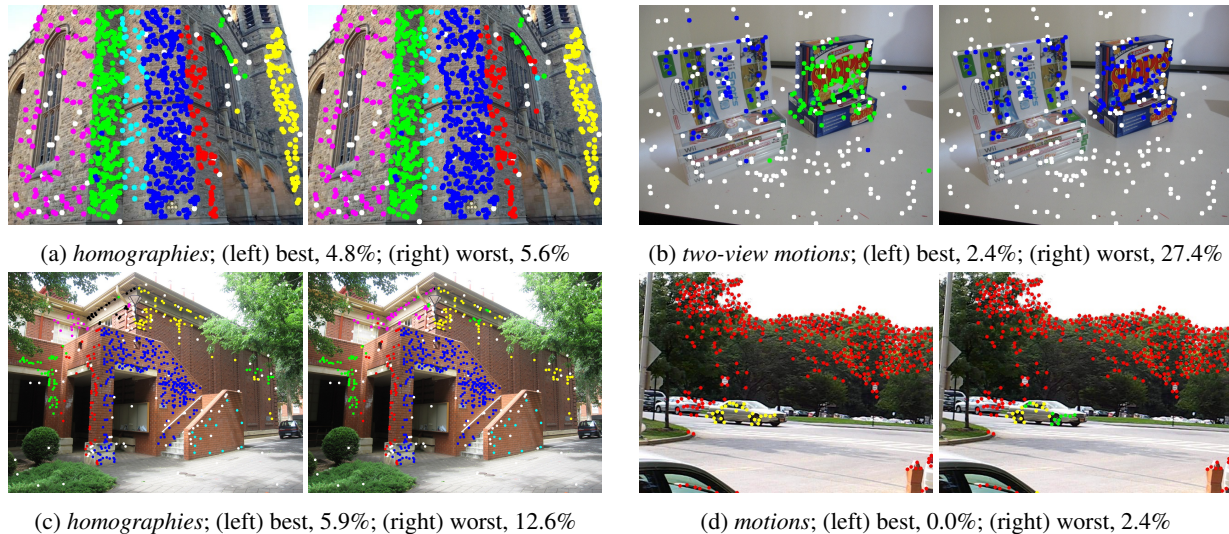


Figure 2: Prog-X – quality and stability of the results on selected problems: homography estimation (a,c), multiple epipolar geometries (b) and rigid motion (d). The best (left) and worst (right) results for five runs with fixed parameters, overlaid on one of the images. White points have been labeled as outliers, color codes the assignment to an instance. Misclassification errors, i.e. the fraction of points assigned to the wrong cluster, are shown as percentages. Results in Table 3 show that competing algorithms are much less stable.

[18, 19, 30, 20, 3, 1] follow a two-step procedure, first, generating a number of instances using RANSAC-like hypothesis generation. Second, a subset of the hypotheses is selected interpreting the input data points the most. This selection is done in various ways. For instance, a popular group of methods [15, 23, 3, 1] optimizes point-to-instance assignments by energy minimization using graph labeling techniques [4]. The energy originates from point-to-instance residuals, label costs [7], geometric priors [23], and from the spatial coherence of the points. Another group of methods uses preference analysis based on the distribution of the residuals of data points. [34, 18, 19, 20]. Also, there are techniques [30, 31] approaching the problem as a hyper-graph partitioning where the instances are vertices, and the points are hyper-edges.

The common part of these algorithms is the initialization step when model *instances are generated blindly*, having no information about the data. As a consequence, it must be decided by the user whether to consider the worst case scenario and, thus, generate an unnecessarily high number of instances; or to use some rule of thumb, e.g. to generate twice the point number hypotheses. In practice, this is what is usually done. It, however, offers no guarantees for covering all the desired instances. The first method recognizing this problem was Multi-X [3]. It added a new step to the optimization procedure to reduce the number of instances by replacing sets of labels by the corresponding density modes in the model parameter domain. Even though this step allows the generation of more initial instances than before

without high computational demand, it still does not provide guarantees, especially, in case of low inlier ratio.

The main contribution of the proposed Prog-X is its any-time nature which originates from interleaving sampling and consolidation of the current data interpretation. This is done via repeated instance proposal minding the already proposed ones; fast rejection of redundant hypotheses; and integration of the new hypothesis into the kept instance set by energy minimization (see Fig. 1 for the basic concept). Moreover, Prog-X adopts the probabilistic guaranties of RANSAC by progressively exploring the data. We use a general instance-to-instance metric based on set overlaps which can be efficiently estimated by the min-hash method [6] and modify the model quality function of RANSAC considering the existence of multiple model instances. The method is tested both in synthetic experiments and on publicly available real-world datasets. It is superior to the state-of-the-art in terms of accuracy for homography, two-view motion, motion clustering, and simultaneous plane and cylinder fitting.

2. Terminology and notation

In this section, we discuss the most important concepts used in this paper. For the sake of generality, we consider multi-class multi-model fitting, thus, aiming to find multiple model instances not necessarily of the same model class. We adopt the notation from [3].

Multi-class multi-instance model fitting. First, we show

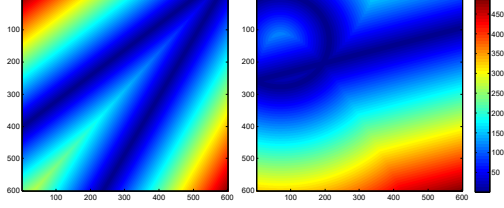


Figure 3: Examples of the compound model, represented by the union of the distance fields, generated by (left) two lines and (right) a line and circle. Best viewed in color.

the problem through examples. A simple problem is finding a pair of *line instances* $h_1, h_2 \in \mathcal{H}_l$ interpreting a set of 2D points $\mathcal{P} \subseteq \mathbb{R}^2$. Line class \mathcal{H}_l is the space of lines $\mathcal{H}_l = \{(\theta_l, \phi_l, \tau_l), \theta_l = [\alpha \ c]^T\}$ equipped with a distance function $\phi_l(\theta_l, p) = |\cos(\alpha)x + \sin(\alpha)y + c|$ ($p = [x \ y]^T \in \mathcal{P}$) and a function $\tau_l(p_1, \dots, p_{m_l}) = \theta_l$ for estimating θ_l from $m_l \in \mathbb{N}$ data points. *Multi-line fitting* is the problem of finding multiple line instances $\{h_1, h_2, \dots\} \subseteq \mathcal{H}_l$, while the *multi-class* case is extracting a subset $\mathcal{H} \subseteq \mathcal{H}_\vee$, where $\mathcal{H}_\vee = \mathcal{H}_l \cup \mathcal{H}_c \cup \mathcal{H}_\cup \cup \dots$. The set \mathcal{H}_\vee is the space of all classes including that of lines and circles. Also, the formulation includes the outlier class $\mathcal{H}_o = \{(\theta_o, \phi_o, \tau_o), \theta_o = \emptyset\}$ which behaves as if all points were equally distant, $\phi_o(\theta_o, p) = k, k \in \mathbb{R}^+$ and $\tau_o() = \emptyset$. The *objective of multi-instance multi-class model fitting* is to determine a set of instances $\{h_i\}_{i=1}^{n_i} \subseteq \mathcal{H}_\vee$ and labeling $L \in \mathcal{P} \rightarrow \mathcal{H}$ assigning each point $p \in \mathcal{P}$ to an instance from $\{h_i\}_{i=1}^{n_i}$ minimizing energy E .

3. Progressive multi-model fitting

In this section, a new pipeline is proposed for multi-model fitting. Before describing each step in depth, a few definitions are discussed.

Definition 3.1. A *putative model instance* h_p is temporary, generated by the proposal engine, not activated to take part in the optimization procedure.

Definition 3.2. An *active model instance* h_a is an instance whose parameters and support are updated in the optimization procedure.

Definition 3.3. *Compound model instance.* Given a set of activate instances $\{h_{a,i}\}_{i=1}^{n_a}$, where n_a is the cardinality of the set, the compound model instance $h_\cup \in \mathcal{H}_\vee$ is defined as the union of the distance fields each generated by an individual active instance $h_{a,i}$ ($i \in [1, n_a]$). The distance of a point \mathbf{p} from h_\cup is $\phi(h_\cup, \mathbf{p}) = \min_{i=1}^{n_a} \phi(h_{a,i}, \mathbf{p})$.

Examples of compound instances are shown in Fig. 3. The color codes the distance (blue – close, red – far). The left plot shows the union of the distance fields generated by two lines. The right one shows that of a circle and a line.

Definition 3.4. *Proposal engine* $\Sigma : h_\cup \times \mathcal{P}^* \rightarrow \mathcal{H}_\vee^*$ is a function generating a putative instance from the data, using the compound model. Operator $*$ denotes the power set.

Therefore, function Σ gets the compound model and the set of points and outputs one or multiple proposals.

Definition 3.5. *Multi-instance optimization* procedure $\Theta : \{h_{a,i}\}_{i=1}^{n_a} \times \{h_p\}_{i=1}^{n_p} \times \mathcal{P}^* \rightarrow \{\hat{h}_{a,i}\}_{i=1}^{\hat{n}_a} \times L$ is a function getting the active, the putative instances and the data as input. It returns a set of active instances and a labeling $L : \mathcal{P}^* \rightarrow h_a^*$ which assigns each point to a single active instance.

Function Θ gets the active instances $\{h_{a,i}\}_{i=1}^{n_a}$, set of proposals $\{h_p\}_{i=1}^{n_p}$ and input points. It returns the optimized active instances and the labeling. Putative instances can be activated or rejected. Also, activate instances may be deactivated and removed here. The set of active instances always contain an instance of the outlier class.

3.1. Proposal engine

The *proposal engine* proposes a yet unseen instance in every iteration. For the engine, we choose a recent variant of RANSAC [9]: Graph-Cut RANSAC [2] since it is state-of-the-art and its implementation is publicly available¹. Due to assuming local structures, similarly as in [3], we choose NAPSAC [22] to be the sampler inside GC-RANSAC. The main objective of the proposal engine is to propose unseen instances, i.e. the ones which are possibly not among the active ones in $\{h_{a,i}\}_{i=1}^{n_a}$. A straightforward choice for achieving this goal is to prefer instances having a reasonable number of *points not shared with the compound instance* h_\cup . Therefore, we propose a new quality function $\hat{Q} : \mathcal{H}_\vee \times \mathcal{H}_\vee^* \times \mathcal{P}^* \times \mathbb{R} \rightarrow \mathbb{R}$, for RANSAC, measuring the score of a model instance originating from \mathcal{H}_\vee w.r.t. the compound instance (from \mathcal{H}_\vee^*), data and a manually set threshold. For the sake of easier understanding, we start from one of the simplest quality functions, i.e. the inlier counting of RANSAC. It is as follows: $Q_R(h, \mathcal{P}, \epsilon) = \sum_{\mathbf{p} \in \mathcal{P}} [\phi(h, \mathbf{p}) < \epsilon]$, where $[\cdot]$ is the Iverson bracket which is equal to one if the condition inside holds and zero otherwise. Based on Q_R , the modified quality function which takes h_\cup into account and, thus, does not count the shared points is the following:

$$\hat{Q}_R(h, \mathcal{P}, \epsilon) = \sum_{\mathbf{p} \in \mathcal{P}} [\phi(h, \mathbf{p}) < \epsilon \wedge \phi(h_\cup, \mathbf{p}) \geq \epsilon]. \quad (1)$$

The condition inside holds if and only if the distance of point \mathbf{p} from instance h is smaller than ϵ and, at the same time, its distance from the compound model h_\cup is greater than ϵ . Therefore, \hat{Q}_R counts the points which are not inliers of the compound instance but inliers of the new one.

¹<https://github.com/danini/graph-cut-ransac>

It nevertheless turned out that, in practice, the truncated quality function of MSAC [27] is superior to the inlier counting of RANSAC in terms of accuracy and sensitivity of the user-defined threshold. It is as follows:

$$Q_M(h, \mathcal{P}, \epsilon) = |\mathcal{P}| - \sum_{\mathbf{p} \in \mathcal{P}} \min \left(1, \frac{\phi(h, \mathbf{p})^2}{\gamma(\epsilon)^2} \right), \quad \gamma(\epsilon) = \frac{3}{2}\epsilon. \quad (2)$$

Considering the previously described objective, Q_M is modified as follows to reduce the score of the points which are shared with the compound model:

$$\widehat{Q}_M(h, \mathcal{P}, \epsilon) = |\mathcal{P}| - \sum_{\mathbf{p} \in \mathcal{P}} \min \left(1, \max \left(\frac{\phi(h, \mathbf{p})^2}{\gamma(\epsilon)^2}, 1 - \frac{\phi(h_{\cup}, \mathbf{p})^2}{\gamma(\epsilon)^2} \right) \right). \quad (3)$$

Consequently, for point \mathbf{p} , the implied score is zero (i) if \mathbf{p} is close to both the proposal and the compound instances, and (ii) if it is far from the proposal.

Summarizing this section, we apply GC-RANSAC as a proposal engine with NAPSAC sampling and quality function \widehat{Q}_M to propose instances one-by-one.

3.2. Proposal validation

The *validation* is an intermediate step between the proposal and the optimization to decide if an instance should be involved in the optimization. To do so, an instance-to-instance distance has to be defined measuring the similarity of the proposed and compound instances. If the distance is low, the proposal is likely to be an already seen one and, thus, is unnecessary to optimize. In [3], the instances are represented by problem-specific sequences of points and this approach, leads to the question of how to represent models by points. In general, the answer is not trivial and the representation affects the outcome significantly. There, is a straightforward solution for representing instances by point sets. In [26], the models are described by their preference sets, and the similarity of two instances is defined via their Jaccard score [16]. The preference set of instance h is $\mathbf{P}_h \in \{0, 1\}^{|\mathcal{P}|}$, where $\mathbf{P}_{h,j}$ is one if the j th point is an inlier of h , otherwise zero ($j \in [1, |\mathcal{P}|]$). The proposed criterion of accepting a putative instance is $J(h, h_{\cup}) = (|\mathbf{P}_h \cap \mathbf{P}_{h_{\cup}}| / |\mathbf{P}_h \cup \mathbf{P}_{h_{\cup}}|) > \epsilon_S$, where J holds if the Jaccard similarity is higher than a manually set threshold $\epsilon_S \in [0, 1]$ and J is false otherwise. We choose Jaccard similarity instead of Tanimoto distance [25, 18] since representing the instances by sets offers a straightforward way of speeding up the procedure.

Computing the Jaccard similarity in case of having thousands of points is a computationally demanding operation. Luckily, we are mostly interested in recognizing if the overlap of the instances is significant. The min-hash algorithm [6] is a straightforward choice for making the processing time of the similarity calculation fast and indepen-

dent on the number of points. *Therefore, the validation step runs in constant time.*

3.3. Multi-instance optimization

The objective of this step is to optimize the set of active model instances whenever a new putative one comes and to decide if this new instance shall be activated or should be rejected. Due to aiming at the most general case, i.e. having multiple classes, there are just a few algorithms [15, 7, 23, 3] which can approach the problem without requiring non-trivial modifications. These algorithms are based on labeling energy-minimization. In general, the major issue of labeling algorithms is their computational complexity, especially, in the case of large label space. In our case, due to proposing instances one-by-one, the label space is always kept small and, therefore, the time spent on the labeling is not significant.

Multi-X [3] could be a justifiable choice. However, in our case, it is simplified to the PEARL algorithm [15, 8] since its major contribution is a move in the label space replacing a label set with the corresponding density mode. When having just a few labels, this move is not needed. Thus, we choose PEARL as the optimization procedure.²

3.4. Termination criterion

In this section, we propose a criterion to determine when to stop with proposing new instances. We generalize the adaptive termination criterion of RANSAC which is based on formula

$$1 - \mu \leq (1 - (|\mathcal{I}|/|\mathcal{P}|)^m)^k, \quad (4)$$

where μ is the required confidence in the results typically set to 0.95 or 0.99; k is the number of iterations; m is the size of the minimal sample; $|\mathcal{I}|$ and $|\mathcal{P}|$ are the number of inliers and points, respectively. In RANSAC, Eq. 4 is formulated to determine k , i.e. the number of iterations required, using the current inlier ratio. We, instead, formulate Eq. 4 to have an estimate of the maximum number of inliers independent on the compound model as follows: $(|\mathcal{P}| - |h_{\cup}|)^m \sqrt[m]{1 - \frac{k}{1 - \mu}} \geq |\mathcal{I}|$ where $|h_{\cup}|$ is the inlier number of the compound model and $|\mathcal{P}| - |h_{\cup}|$ is the number of points which are not covered by the compound instance. Therefore, $\bar{I}(\mathcal{P}, h_{\cup}, m, k, \mu) = (|\mathcal{P}| - |h_{\cup}|)^m \sqrt[m]{1 - \frac{k}{1 - \mu}}$ is an upper limit for the inlier number of a not yet found instance with confidence μ in the k th iteration.

To find an instance, it has to fit (i.e., the residual is approximately zero) to more points than what initialized it. Thus, at least $m+1$ points have to support it, where m is the size of a minimal sample. Therefore, the algorithm updates \bar{I} in the end of every iteration and terminates if $\bar{I} < m+1$. *This constraint guaranties* that when the algorithm terminates, the probability of having an unseen model with at least $m+1$ inliers is smaller than $1 - \mu$.

²<https://github.com/nsubtil/gco-v3.0>

Note that in practice, it is more convenient to set this limit on the basis of the optimization. For instance, if the optimization does not accept instances having fewer than 20 inliers, it does not make sense to propose ones with fewer.

4. Experimental results

In this section, we evaluate the proposed Prog-X method on various computer vision problems that are: 2D line, homography, two-view motion, motion estimation, and simultaneous plane and cylinder fitting. The reported results are the averages over five runs and obtained by using fixed parameters. See Table 2 for the parameters of Prog-X.

4.1. Synthesized tests

In this section Prog-X is tested in a synthetic environment. We chose line fitting to 2D points and downloaded the dataset used in [26]. It consists of three scenes: `star4`, `star5` and `star11` (see Fig. 4a). The outlier ratio in all test cases is 0.5 thus having equal number of inliers and outliers.

Comparison of multi-model fitting algorithms. We compare Prog-X, Multi-X³, [3] RansaCov [20], RPA [19], T-Linkage⁴ [18], and PEARL [5] on line fitting in this sections. All methods were applied five times to all scenes with fixed parameters. For T-Linkage, RPA and RansaCov, we used the parameters which the authors proposed. We tuned Prog-X, Multi-X, and PEARL to achieve the most accurate average results.

The *worst results* (i.e., the ones with the highest “# fn + # fp” error) in five runs are visualized in Fig. 4. Plot (a) shows the ground truth lines and point-to-line assignments. The points of each cluster are drawn by color. The number of false negative, i.e. a line which is not found, and false positive, i.e. a found line which is not in the ground truth set, instances are reported in Table 1. It can be seen, that the proposed method leads to the most accurate results. It finds all but one lines and does not return false positives.

Any-time property. To demonstrate the any-time property of Prog-X, we applied the methods which minimize an energy function iteration-by-iteration, i.e. Prog-X, PEARL [5] and Multi-X [3], to the `star11` scene and reported their states in each iteration. All methods were applied five times and, for each, the run with the worst outcome was selected.

Fig. 5 plots the evolution of the properties iteration-by-iteration. In (a), the numbers of stored instances (vertical axis) are plotted as the function of the iteration number (horizontal). It can be seen that PEARL and Multi-X have an unnecessarily high number of instances stored in the first iteration. Then, from the second one, it drops significantly and remains almost constant. For Prog-X, the number of instances increases by one in each iteration. The ground truth

number was 11 in this scene. In (b), it is simulated what would happen if a method is stopped before it terminates. In this case, all methods have a set of instances stored. In the plot, the ratio (vertical axis) of the number of desired instances kept and the number of all stored instances is shown as the function of the iteration number (horizontal). A desired instance is one which overlaps with a ground truth one. A ground truth instance can be covered only by one instance. The number of desired instances is, thus, at most the number of ground truth models in the data. It can be seen that PEARL and Multi-X are significantly worse than Prog-X in the first iteration since the number of stored instances is far bigger than the number of desired ones. Even if all the ground truth ones are covered, there are many false positives and, thus, the instances are not usable without further optimization. This ratio for Prog-X is one in all iterations and, as a consequence, *it can be stopped any time and still returns solely desired instances*. In (c), the processing times in milliseconds (vertical axis) are reported for each iteration (horizontal). PEARL and Multi-X spend seconds in the first iteration and, then, their processing times drop significantly. This is the expected behavior knowing that their most time-consuming operation is the α -expansion algorithm which has to manage a fairly large label space in the first iterations. For Prog-X, the processing time is almost constant since it alternates between the proposal and optimization step with slightly increasing label space. **Summarizing** (a-c), *PEARL and Multi-X are not any-time algorithms*. To (a) and (b), they can be stopped after the first iteration without significant deterioration in the accuracy. However, to (c), the first iteration requires more processing time than the rest in total. Prog-X can be interrupted at any time; all of the stored instances cover ground truth ones.

Plot (d) shows the change of the energy (vertical axis) iteration-by-iteration (horizontal). We divided the energy in each iteration by the energy in the first one. It can be seen that Prog-X leads to a significantly higher reduction in the energy than the competitor algorithms.

Comparison with increasing outlier ratio. To evaluate, how the outlier ratio affects the outcome of multi-model fitting algorithms, we first kept solely the noisy inliers from the `star11` scene. Then outliers, i.e. points uniformly distributed in the scene, were added to achieve a given outlier ratio ν . Fig. 6 reports (a) the processing time (in milliseconds; vertical axis) and (b) the difference between the ground truth and the obtained instance number (vertical) as the function of the outlier ratio (horizontal). It can be seen that the proposed Prog-X leads to similar processing time as Multi-X but the returned number of instances is significantly closer to the desired one. Consequently, it is the least sensitive to the outlier ratio.

³<https://github.com/danini/multi-x>

⁴<http://www.diegm.uniud.it/fusiello/demo/>

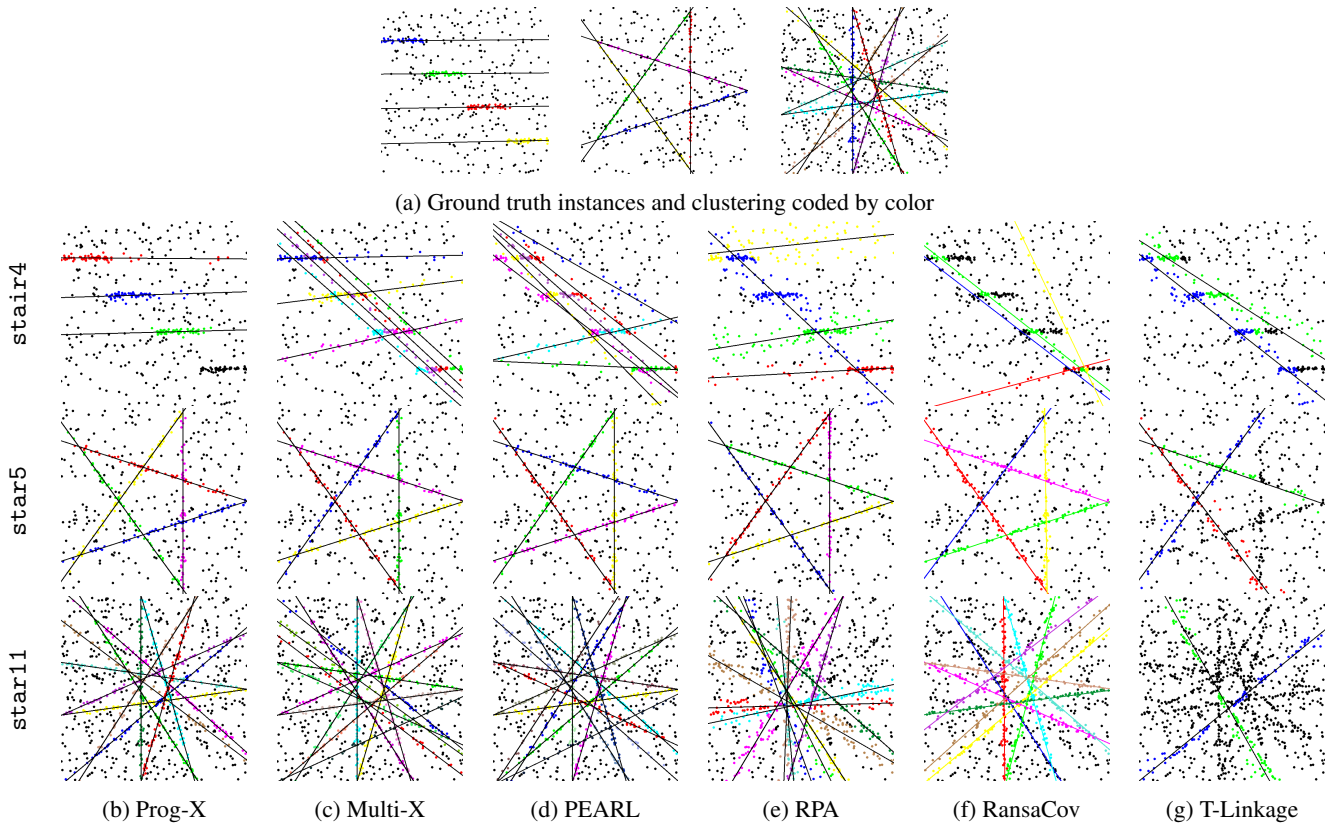


Figure 4: Fitting of multiple 2D lines to synthetic data. The *worst results* over five runs on the dataset of [26] are shown. The points are assigned to lines by color. In (a), the ground truth clusterings are shown. From (b) to (g), the results of the evaluated algorithms are put. Fixed parameters were used for all methods. See Table 1 for the details of the results.

	Progressive-X		Multi-X		PEARL		RPA		RansaCov		T-Linkage	
	# fn	# fp	# fn	# fp	# fn	# fp	# fn	# fp	# fn	# fp	# fn	# fp
stair4	1	0	1	4	2	5	2	2	4	4	4	2
star5	0	0	0	0	0	0	0	0	0	0	2	0
star11	0	0	0	2	0	3	5	5	1	1	9	0
all	1	0	1	6	2	8	7	7	5	5	15	2

Table 1: Fitting of multiple 2D lines to synthetic data. The *worst results* over five runs on the dataset of [26] are shown. The reported properties are the number of false negative (# fn) and false positive instances (# fp). Fixed parameters were used for all methods. The results are visualized in Fig. 4.

4.2. Real world problems

To evaluate the proposed method on real-world problems, we downloaded a number of publicly available datasets. The error is the misclassification error (ME), i.e. the ratio of points assigned to the wrong cluster. In Fig. 2, the first images are shown of image pairs which are some of the worst (left plots) and best (right) results of Prog-X running five times on each scene with a fixed parameters (reported in Table 2). The shown percentages are the misclassification errors. The white points are labeled as outliers,

while the ones with color are assigned to an instance. It can be seen the error originates mostly from missing instances and, thus, the returned ones are usable.

Two-view motion fitting is evaluated on the AdelaideRMF motion dataset consisting of 21 image pairs of different sizes and correspondences manually assigned to motion clusters. In this case, multiple fundamental matrices were fit. In the proposal step, the 7-point algorithm [12] was applied for fitting to a minimal sample and the normalized 8-point method [11] for the polishing steps on non-minimal samples. The average (of 5 runs with fixed parameters)

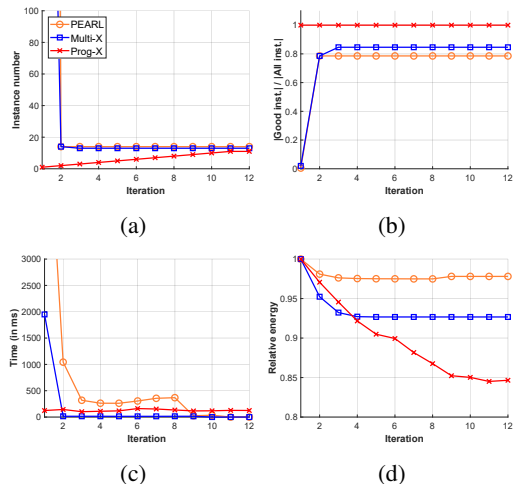


Figure 5: *The results iteration-by-iteration.* The compared methods minimizing an energy function iteratively: the proposed Prog-X (red), Multi-X [3] (blue) and PEARL [5] (orange). **(a)** The number of stored instances are shown (vertical axis) as a function of the iteration number (horizontal). The values in the 1st iteration are 2 200 for PEARL and 550 for Multi-X. **(b)** The ratio (vertical) of the number of ground truth models covered by an instance and the number of all instances stored are reported. **(c)** The processing times (vertical; in ms) are shown for each iteration (horizontal). The times in the 1st iteration are 6 904 ms for PEARL and 1 949 ms for Multi-X. In total, PEARL required 9 880 ms, Multi-X 2 065 ms and Prog-X 1 534 ms. **(d)** The energy divided by the energy in the 1st iteration (horizontal) is shown.

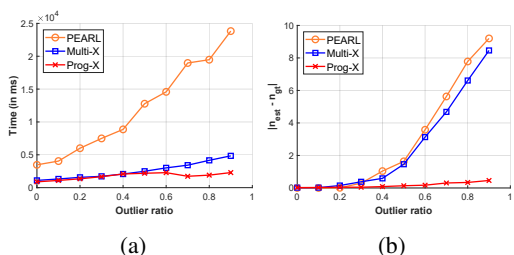


Figure 6: **(a)** The average (of 1 000 runs) processing time (in msec; vertical axis) and **(b)** the difference of ground truth and returned instance numbers (vertical) plotted as the function of the outlier ratio (horizontal) on *star11*. In each run, random, uniformly distributed, outliers are generated and the coordinates of the original noisy inliers are perturbed by zero-mean Gaussian-noise with $\sigma = 1.0$ pixels.

errors and their standard deviations are shown in the first block of Table 3 for each method (from the 2nd to the 4th columns). Prog-X is superior to the competitors in all investigated properties. Detailed results for each scene are shown

	μ	ϵ_S	w_s	ϵ	w_l
Two-view motions				1.6	43.0
Homographies	0.95	0.1	0.15	2.9	112.5
Motions				$5 * 10^{-4}$	$2.25 * 10^{-6}$
Planes, cylinders				0.5	0.035

Table 2: Parameters of Prog-X used for each problem (rows). The RANSAC confidence (μ); minimal Jaccard-distance for passing the instance validation (ϵ_T); inlier-outlier threshold (ϵ); label cost (w_l); and the weight of the spatial coherence term (w_s) are reported.

in the top-left plot of Fig. 7. It can be seen that Prog-X is always the most or the second most accurate method.

Homography fitting is evaluated on the AdelaideRMF homography dataset [32] used in most recent publications. AdelaideRMF consists of 19 image pairs of different resolutions with ground truth point correspondences assigned manually to homographies. In the proposal step, the normalized 4-point algorithm [12] was used for fitting to a minimal sample and also for the polishing steps. The results are shown in the second block of Table 3 (from the 5th to the 7th columns). Prog-X leads to the lowest average error and standard deviation. Similar results can be seen in the top-right plot of Fig. 7 as before: Prog-X always leads to the lowest or second lowest misclassification errors.

Motion segmentation is tested on the 155 videos of the Hopkins dataset [28]. The dataset consists of 155 sequences divided into three categories: checkerboard, traffic, and other sequences. The trajectories are inherently corrupted by noise, but no outliers are present. Motion segmentation in videos is the retrieval of sets of points undergoing rigid motions contained in a dynamic scene captured by a moving camera. It can be considered as a subspace segmentation under the assumption of affine cameras. For affine cameras, all feature trajectories associated with a single moving object lie in a 4D linear subspace in \mathbb{R}^{2F} , where F is the number of frames [28]. The results are shown in the third block of Table 3 (from the 8th to the 10th columns). Prog-X leads to the lowest average errors. Its standard deviation is the second lowest. Detailed results on the sequences of traffic are put in the bottom of Fig. 7.

Plane and cylinder fitting is evaluated on the dataset from [3]. It consists of LiDAR point clouds of traffic signs, their columns and the neighboring points. Points were manually assigned to signs (planes) and columns (cylinders). The proposal step of Prog-X alternately proposes cylinders and planes. The results are in the last three columns of Table 3. Prog-X obtains the most accurate results, but, most importantly, it is *three orders of magnitude faster* than the second fastest method. The reason is the large number of points in the scenes (from 1 260 up to 52 445). The processing time of Prog-X, due to being dominated by a number of

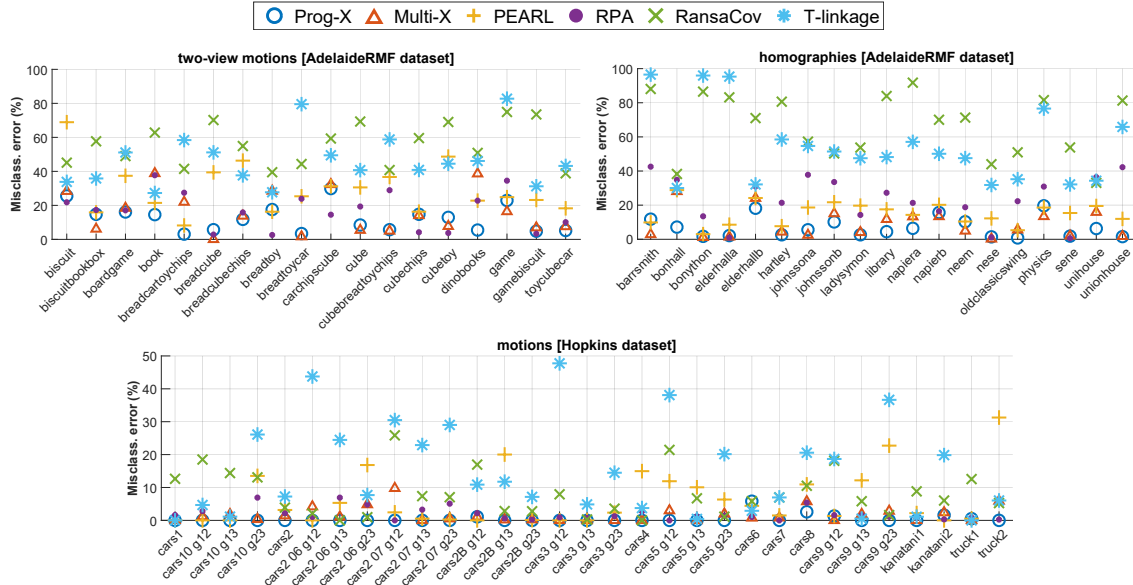


Figure 7: The average ME (vertical axis; in percentage; avg. of 5 runs) of multi-model estimators applied to problems: two-view motion (AdelaideRMF **F** dataset; top-left), multi-homography (AdelaideRMF **H** dataset; top-right), multi-motion (Hopkins dataset, traffic sequences; bottom) fitting. All methods used fixed parameters. Summarized results are in Table 3.

	Two-view motions			Homographies			Motions			Planes and cylinders		
	18 scenes			19 scenes			155 scenes			8 scenes		
	avg.	std.	time	avg.	std.	time	avg.	std.	time	avg.	std.	time
Prog-X	10.73	8.73	14.38	6.86	5.91	1.03	8.41	10.29	0.02	33.69	8.26	9.36
Multi-X [3]	17.13	12.23	1.52	8.71	8.13	0.27	12.96	19.60	0.95	35.67	11.94	1 407.39
PEARL [15]	29.54	14.80	4.94	15.14	6.75	2.61	14.25	23.23	3.30	43.89	11.99	5 142.39
RPA [19]	17.11	11.08	10.24	23.54	13.42	622.87	9.16	11.26	4.92	55.47	11.71	10 459.98
RansaCov [20]	55.61	12.42	2.33	66.88	18.44	17.69	11.13	8.00	2.04	46.65	10.31	7 914.00
T-linkage [18]	46.67	15.60	2.69	54.79	22.17	57.84	27.24	15.57	0.95	45.33	15.37	8 423.31

Table 3: Average misclassification errors (in %; 5 runs on each scene), their standard deviations and the processing times (in secs) for each problem: two-view motion fitting on the AdelaideRMF motion dataset (2nd–4th cols), homography estimation on the AdelaideRMF homography dataset (5th–7th), motion fitting on the Hopkins dataset (8th–10th), and simultaneous plane and cylinder fitting on the Multi-X dataset (11th–13th). Fixed parameters were used for all methods (see Table 2 for Prog-X). Detailed results are in Fig. 7.

RANSAC runs, depends linearly on the point number. All of the other methods depend approximately quadratically on n . In RPA, T-Linkage and RansaCov, the preference calculation has $\mathcal{O}(n^2)$ complexity. In Multi-X and PEARL, if at least n initial instances are generated, the label space is already too large for the inner α -expansion to finish early.

5. Conclusion

The Prog-X algorithm is proposed for geometric multi-class multi-model fitting. The method is fast and superior to the state-of-the-art in terms of accuracy in a synthetic environment and on publicly available real-world datasets for homography, two-motion, and motion segmentation. Additionally, it is an any-time algorithm. Therefore, when-

ever is interrupted, e.g. due to a strict time limit, the returned instances cover real and, likely, the most dominant ones. The termination criterion, adopted from RANSAC, makes Prog-X robust to the inlier-outlier ratio. The source code is available at <https://github.com/danini/progressive-x>.

Acknowledgement

This work was supported by the Czech Science Foundation grant GA18-05360 and OP VVV project CZ.02.1.01/0.0/0.0/16019/000076 Research Center for Informatics. D. Barath was also supported by the Hungarian Scientific Research Fund (No. NKFIH OTKA KH-126513 and K-120499).

References

- [1] Paul Amayo, Pedro Piniés, Lina M Paz, and Paul Newman. Geometric multi-model fitting with a convex relaxation algorithm. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8138–8146, 2018. 2
- [2] Daniel Barath and Jiri Matas. Graph-cut ransac. *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3
- [3] Daniel Barath and Jiri Matas. Multi-class model fitting by energy minimization and mode-seeking. In *European Conference on Computer Vision*, 2018. 2, 3, 4, 5, 7, 8
- [4] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence*, 2004. 2
- [5] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence*, 2001. 5, 7
- [6] A. Z. Broder. On the resemblance and containment of documents. In *Compression and complexity of sequences 1997. proceedings*, pages 21–29. IEEE, 1997. 2, 4
- [7] Andrew Delong, Lena Gorelick, Olga Veksler, and Yuri Boykov. Minimizing energies with hierarchical costs. *International Journal of Computer Vision*, 2012. 2, 4
- [8] Andrew Delong, Anton Osokin, Hossam Isack, and Yuri Boykov. Fast approximate energy minimization with label costs. *International journal of computer vision*, 2012. 4
- [9] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981. 1, 3
- [10] Nicolas Guil and Emilio L. Zapata. Lower order circle and ellipse hough transform. *Pattern Recognition*, 1997. 1
- [11] Richard Hartley. In defense of the eight-point algorithm. *Transactions on Pattern Analysis and Machine Intelligence*, 1997. 6
- [12] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. 2003. 6, 7
- [13] P. V. C. Hough. Method and means for recognizing complex patterns, 1962. 1
- [14] John Illingworth and Josef Kittler. A survey of the hough transform. *Computer Vision, Graphics, and Image Processing*, 1988. 1
- [15] Hossam Isack and Yuri Boykov. Energy-based geometric multi-model fitting. *International Journal on Computer Vision*, 2012. 2, 4, 8
- [16] Paul Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579, 1901. 4
- [17] Yasushi Kanazawa and Hiroshi Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *British Machine Vision Conference*, 2004. 1
- [18] Luca Magri and Andrea Fusiello. T-Linkage: A continuous relaxation of J-Linkage for multi-model fitting. In *Conference on Computer Vision and Pattern Recognition*, 2014. 2, 4, 5, 8
- [19] Luca Magri and Andrea Fusiello. Robust multiple model fitting with preference analysis and low-rank approximation. In *British Machine Vision Conference*, 2015. 2, 5, 8
- [20] Luca Magri and Andrea Fusiello. Multiple model fitting as a set coverage problem. In *Conference on Computer Vision and Pattern Recognition*, 2016. 2, 5, 8
- [21] Jiri Matas, Csaba Galambos, and Josef Kittler. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding*, 2000. 1
- [22] D. R. Myatt, Philip Torr, Slawomir J. Nasuto, Mark J. Bishop, and R. Craddock. NAPSAC: High noise, high dimensional robust estimation - its in the bag. 2002. 3
- [23] Trung Thanh Pham, Tat-Jun Chin, Konrad Schindler, and David Suter. Interacting geometric priors for robust multi-model fitting. *TIP*, 2014. 2, 4
- [24] Paul L. Rosin. Ellipse fitting by accumulating five-point fits. *Pattern Recognition Letters*, 1993. 1
- [25] T. T. Tanimoto. Elementary mathematical theory of classification and prediction. 1958. 4
- [26] Roberto Toldo and Andrea Fusiello. Robust multiple structures estimation with J-Linkage. In *European Conference on Computer Vision*, 2008. 4, 5, 6
- [27] Philip Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision*, 50(1):35–61, 2002. 4
- [28] Roberto Tron and Rene Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *Conference on Computer Vision and Pattern Recognition*, 2007. 7
- [29] E. Vincent and Robert Laganière. Detecting planar homographies in an image pair. In *International Symposium on Image and Signal Processing and Analysis*, 2001. 1
- [30] Hanzi Wang, Guobao Xiao, Yan Yan, and David Suter. Mode-seeking on hypergraphs for robust geometric model fitting. In *International Conference of Computer Vision*, 2015. 2
- [31] Hanzi Wang, Guobao Xiao, Yan Yan, and David Suter. Searching for representative modes on hypergraphs for robust geometric model fitting. *Transactions on Pattern Analysis and Machine Intelligence*, 2018. 2
- [32] Hoi Sim Wong, Tat-Jun Chin, Jin Yu, and David Suter. Dynamic and hierarchical multi-structure geometric model fitting. In *International Conference on Computer Vision*, 2011. 2, 7
- [33] Lei Xu, Erkki Oja, and Pekka Kultanen. A new curve detection method: randomized hough transform (rht). *Pattern Recognition Letters*, 1990. 1
- [34] Wei Zhang and Jana Kosecká. Nonparametric estimation of multiple structures with outliers. In *Dynamical Vision*. 2007. 2
- [35] Marco Zuliani, C. S. Kenney, and Bangalore Manjunath. The multiransac algorithm and its application to detect planar homographies. In *International Conference on Image Processing*. IEEE, 2005. 1