Multi-Class Model Fitting by Energy Minimization and Mode-Seeking

Daniel Barath
Machine Perception Research Laboratory
MTA SZTAKI, Budapest, Hungary
barath.daniel@sztaki.mta.hu

Jiri Matas
Centre for Machine Perception, Department of Cybernetics
Czech Technical University, Prague, Czech Republic
matas@cmp.felk.cvut.cz

Abstract

We propose a novel method, called Multi-X, for general multi-class multi-instance model fitting – the problem of interpreting input data as a mixture of noisy observations originating from multiple instances of multiple types. The proposed approach combines global energy optimization and mode-seeking in the parameter domain. Robustness is achieved by using an outlier class. Key optimization parameters like the outlier threshold are set automatically within the algorithm. Considering that a group of outliers may form spatially coherent structures in the data, we propose a cross-validation-based technique removing statistically insignificant instances.

Multi-X outperforms significantly the state-of-the-art on the standard AdelaideRMF (multiple plane segmentation, multiple rigid motion detection) and Hopkins datasets (motion segmentation) and in experiments on 3D LIDAR data (simultaneous plane and cylinder fitting) and on 2D edge interpretation (circle and line fitting). Multi-X runs in time approximately linear in the number of data points at around 0.1 second per 100 points, an order of magnitude faster than available implementations of commonly used methods.

1. Introduction

In multi-class fitting, the input data is interpreted as a mixture of noisy observations originating from multiple instances of multiple model types, e.g. as \( k \) lines and \( l \) circles in 2D edge maps, \( k \) planes and \( l \) cylinders in 3D data, multiple homographies or fundamental matrices in correspondences from a non-rigid scene (see Fig. 1). Robustness is achieved by considering assignment to an outlier class.

Multi-model fitting has been studied since the early sixties, the Hough-transform [12, 13] being the first popular method for extracting multiple instances of a single class [11, 23, 29, 40]. A widely used approach for finding a single instance is RANSAC [9] which alternates two main steps: the generation of instance hypotheses and their validation. However, extending RANSAC to the multi-instance case has had limited success. Sequential RANSAC detects instance one after another in a greedy manner, removing their inliers [36, 16]. In this approach, data points are assigned to the first instance, typically the one with the largest support, for which they cannot be deemed outliers, rather than to the best instance. MultiRANSAC [42] forms compound hypothesis about \( n \) instances. Besides requiring the number \( n \) of the instances to be known a priori, the approach increases the size of the minimum sample and thus the num-

![Figure 1: Multi-class multi-instance fitting examples. Results on simultaneous plane and cylinder (top left), line and circle fitting (top right), motion (bottom left) and plane segmentation (bottom right).](image)
ber of hypotheses that have to be validated.

Most recent approaches [15, 20, 21, 22, 33] focus on the single class case: finding multiple instances of the same model type. A popular group of methods [7, 15, 25, 27] adopts a two step process: initialization by RANSAC-like instance generation followed by a point-to-instance assignment optimization by energy minimization using graph labeling techniques [2]. Recently, an approach was proposed combining Mean-Shift and energy minimization to solve the two-view homography fitting problem in rigid scenes [1]. Another group of methods uses preference analysis, introduced by RHA [41], which is based on the distribution of residuals of individual data points with respect to the instances [20, 21, 33].

The multiple instance multiple class case considers fitting of instances that are not necessarily of the same type. This generalization has received much less attention than the single-class case. To the best of our knowledge, the last significant contribution is that of Stricker and Leonardis [31] who search for multiple parametric models simultaneously by minimizing description length using Tabu-search.

The proposed Multi-X method finds multiple instances of multiple model types drawing on progress in energy minimization and in mode seeking in the parameter domain. Mode seeking significantly reduces the label space, thus speeding up the energy minimization step and overcomes the problem of multiple instances with similar parameters, a weakness of state-of-the-art single-class approaches (see Fig. 2). The assignment of data to instances of different model types is handled by the introduction of class-specific distance functions. Multi-X can also be seen as an extension or generalization of the Hough transform: (1) it finds modes of the parameter space density without creating an accumulator and locating local maxima there, which is prohibitive in high dimensional spaces, (2) it handles multiple classes - running Hough transform for each model type in parallel or sequentially cannot easily handle competition for data points, and (3) the ability to model spatial coherence of inliers is added.

Most recent papers [20, 22, 37] report results tuned for each test case separately. The results are impressive, but input-specific tuning, i.e. semi-automatic operation with multiple passes, severely restricts possible applications. We propose an adaptive parameter setting strategy within the algorithm, allowing the user to run Multi-X as a black box on a range of problems with no need to set any parameters. Considering that outliers may form structures in the input, as a post-processing step, a cross-validation-based technique removes insignificant instances.

The contributions of the paper are: (1) energy minimization extended by median-based mode-seeking which leads to results more accurate than the state-of-the-art and order of magnitude faster speed, (2) introduction of $L_1$ based instance parameter re-estimation which considers an outlier class, gaining robustness over the classic $L_2$ approach and (3) automatic parameter setting of optimization parameters.

2. Problem Formulation

Before presenting the general definition, let us consider a few examples of multi-instance fitting: finding a pair of line instances $h_1, h_2 \in \mathcal{H}_l$ fitting a set $\mathcal{P} \subseteq \mathbb{R}^2$ of 2D points. The line class $\mathcal{H}_l$ is the space of lines $\mathcal{H}_l = \{(\theta_l, \phi_l), \theta_l = [\alpha \ \beta]^T\}$ equipped with a distance function $\phi_l(\theta_l, p) = |x\alpha + y\beta + c|$, where $p = [x \ y]^T \in \mathcal{P}$.

Another simple example is the fitting of $n$ circle instances $h_1, h_2, \cdots, h_n \in \mathcal{H}_c$ to the same data. The circle class $\mathcal{H}_c = \{(\theta_c, \phi_c), \theta_c = [c_x \ c_y \ r]^T\}$ is the space of circles and $\phi_c(\theta_c, p) = r - \sqrt{(c_x - x)^2 + (c_y - y)^2}$ is a distance function.

**Multi-line fitting** is the problem of finding multiple line instances $\{h_1, h_2, \cdots\} \subseteq \mathcal{H}_l$, while the multi-class case is extracting a subset $\mathcal{H} \subseteq \mathcal{H}_s$, where $\mathcal{H}_s = \mathcal{H}_l \cup \mathcal{H}_c \cup \mathcal{H}_l \cup \cdots$. The set $\mathcal{H}_s$ is the space of all classes, e.g. line, circle. Our formulation also includes the outlier class $\mathcal{H}_o = \{(\theta_o, \phi_o), \theta_o = \emptyset\}$ where each instance has constant but possibly different distance to all points $\phi_o(\theta_o, p) = k, k \in \mathbb{R}^+$.  

**Definition 1 (Multi-Class Model)** The multi-class model is a space $\mathcal{H}_s = \bigcup \mathcal{H}_i$, where $\mathcal{H}_i = \{(\theta_i, \phi_i), d_i \in \mathbb{N}, \theta_i \in \mathbb{R}^{d_i}, \phi_i \in \mathcal{P} \times \mathbb{R}^{d_i} \rightarrow \mathbb{R}\}$ is a single class, $\mathcal{P}$ is the set of data points, $d_i$ is the dimension of parameter vector $\theta_i$ and $\phi_i$ is the distance function of the $i$-th class.

The objective of multi-instance multi-class model fitting is to determine a set of instances $\mathcal{H} \subseteq \mathcal{H}_s$ and labeling $L \in \mathcal{P} \rightarrow \mathcal{H}$ assigning each point $p \in \mathcal{P}$ to an instance $h \in \mathcal{H}$ minimizing energy $E$. We adopt energy

$$E(L) = E_d(L) + \lambda E_s(L) + \beta E_c(L), \quad (1)$$

to measure the quality of the fitting, where $\lambda$ and $\beta$ are weights balancing the different terms described bellow, and $E_d, E_s$ and $E_c$ are the data, spatial coherence and complexity terms, respectively.

**Data term** $E_d : (\mathcal{P} \rightarrow \mathcal{H}) \rightarrow \mathbb{R}$ is defined in most energy minimization approaches as

$$E_d(L) = \sum_{p \in \mathcal{P}} \phi_{L(p)}(\theta_{L(p)}, p),$$

penalizing inaccuracies induced by the point-to-instance assignment, where $\phi_{L(p)}$ is the distance function of model instance $h_{L(p)}$.

In our experiments, this data term worked accurately for instances having distant parameters but fails when some
of them are similar. Multi-model fitting techniques based on energy-minimization usually generate a high number of instances \( H \subseteq H_V \) randomly as a first step \([15, 25]\) \(|H| \gg |H_{real}|\), where \(|H_{real}|\) is the ground truth instance set. Therefore, the presence of many similar instances is typical. In this paper, we assume that the instances in \( H_{real} \) are supported by many points that after initialization are assigned to a number of similar instances in \( H \). The groups around the real instances in the model parameter domain can be replaced with the density modes. The proposed data term reflecting this assumption is

\[
\hat{E}_d(L) = \sum_{p \in P} \phi^\theta_{L(p)}(\theta^\theta_{L(p)}, p),
\]

where \( \Theta : H_V \rightarrow H_V \), is a mode-seeking function and pair \((\theta^\theta_{L(p)}, \phi^\theta_{L(p)}) \in \Theta(H)\) is the mode assigned to point \( p \). Data term \( \hat{E}_d(L) \) captures the point-to-mode assignment cost instead that of point-to-instance.\(^1\)

Spatial coherence term, adopted from [15], is defined as

\[
E_s(L) : (P \rightarrow H) \rightarrow \mathbb{R} = \sum_{(p,q) \in N} w_{pq} \|L(p) \neq L(q)\|,
\]

where \( N \) are the edges of a predefined neighborhood graph, the Iverson bracket \([\cdot]\) is equal to one if the condition inside holds and zero otherwise, and \( w_{pq} \) is a pairwise weighting term. In this paper, \( w_{pq} \) equals to one. \( E_s(L) \) reflects the assumption that close points are more likely to belong to the same instance. Remark: if it is required to consider higher-order geometric terms, e.g. to find three perpendicular planes, \( E_s \) can be replaced with the energy term proposed in [25].

A regularization of the number of instances is proposed by Li [18] as a label count penalty

\[
E_c(L) : (P \rightarrow H) \rightarrow \mathbb{R} = \|L(P)\|,
\]

where \( L(P) \) is the set of distinct labels of labeling function \( L \). To handle multi-class models which might have different costs on the basis of the model class, we thus propose the following definition:

**Definition 2 (Weighted Multi-Class Model)** The weighted multi-class model is a space \( H_W = \bigcup \hat{H}_i \), where \( \hat{H}_i = \{(\theta_i, \phi_i, \psi_i) | d_i \in \mathbb{N}, \theta_i \in \mathbb{R}^{d_i}, \phi_i \in \mathcal{P} \times \mathbb{R}^{d_i} \rightarrow \mathbb{R}, \psi_i \in \mathbb{R}\} \) is a weighted class, \( \mathcal{P} \) is the set of data points, \( d_i \) is the dimension of parameter vector \( \theta_i \), \( \phi_i \) is the distance function, and \( \psi_i \) is the weight of the \( i \)-th class.

The term controlling the number of instances is

\[
\hat{E}_c(L) = \sum_{i \in L(P)} \psi_i, \tag{4}
\]

instead of \( E_c \), where \( \psi_i \) is the weight of the weighted multi-class model referred by label \( i \).

Combining terms Eqs. 2, 3, 4 leads to overall energy \( \hat{E}(L) = \hat{E}_d(L) + \lambda E_s(L) + \beta \hat{E}_c(L) \).

### 3. Multi-X

#### 3.1. Alternating Optimization

Given a set of model instances generated by e.g. uniform or guided sampling [33] similarly to RANSAC [9]. Energy \( \hat{E} \) uses the modes provided by function \( \Theta \) instead of the initially generated instances. It is easy to see that the optimization of \( \hat{E} \) can be separated into two steps without loss of generality:

1. Application of \( \Theta \) to \( H \), where \( H \) is the set of instances.
2. Minimization of the overall energy \( \hat{E} \) with labeling \( L \).

In order to deal with outliers a third step is introduced:

3. Detection of points not belonging to any instance.

We add a fourth step similarly to [15] as:

(4) Re-estimating the model instances \( H \) w.r.t. labeling \( L \).

In Multi-X, these steps are alternated until convergence. Next, each step is described in depth (see Alg. 1).

---

**Algorithm 1 Multi-X**

**Input**: \( P \) – data points

**Output**: \( H^* \) – model instances, \( L^* \) – labeling

1: \( H_0 := \text{InstanceGeneration}(P); i := 1; \)
2: \textbf{repeat}
3: \( H_i := \text{ModeSeeking}(H_{i-1}); \quad \triangleright \text{by Median-Shift} \)
4: \( L_i := \text{Labeling}(H_i, P); \quad \triangleright \text{by } \alpha\text{-expansion} \)
5: \( L_i := \text{OutlierRemoval}(H_i, L_i, \gamma); \quad \triangleright \text{by Weiszfeld} \)
6: \( H_i := \text{ModelFitting}(H_i, L_i, P); \quad \triangleright \text{by Weiszfeld} \)
7: \( i := i + 1; \)
8: \textbf{until} \text{Convergence}(H_i, L_i)
9: \( H^* := H_{i-1}, L^* := L_{i-1}; \)
10: \( H^*, L^* := \text{ModelValidation}(H^*, L^*); \quad \triangleright \text{Alg. 2} \)

---

**1. Mode-Seeking** is applied in the model parameter domain. Suppose that a set of instances \( H \) is given. Since the number of instances in the solution – the modes in the parameter domain – is unknown, a suitable choice for mode-seeking is the Mean-Shift algorithm [5] or one of its variants. In preliminary experiments, the most robust choice was the Median-Shift [30] using Weiszfeld- [38] or Tukey-medians [35]. There was no significant difference, but Tukey-median was slightly faster to compute.

Reflecting the fact that a general *instance-to-instance* distance is needed, we represent instances by point sets,
e.g. a line by two points and a homography by four correspondences, and define the instance-to-instance distance as the Hausdorff-distance [28] of the point sets. Even though it yields slightly more parameters than the minimal representation, thus making Median-Shift a bit slower, it is always available as it is used to define spatial neighborhood of points.

There are many point sets defining an instance and a canonical point set representation is needed. For lines, the nearest point to the origin is used and a point on the line at a fixed distance from it. For a homography \( H \), the four points are \( H[0, 0, 1]^T \), \( H[1, 0, 1]^T \), \( H[0, 1, 1]^T \), and \( H[1, 1, 1]^T \), etc. The matching step can be excluded from the Hausdorff-distance, thus speeding up the distance calculation significantly.\(^2\)

The application of Median-Shift \( \Theta_{\text{Med}} \) never increases the number of instances \(|H_i|\):

\[
|\Theta_{\text{Med}}(H_i)| \leq |H_i|.
\]

The equality is achieved if and only if the distance between every instance pair is greater than the bandwidth. Note that for each distinct model classes, Median-Shift has to be applied separately. After this step, the outlier threshold is re-estimated as discussed in Section 3.3.

2. Labeling assigns points to model instances obtained in the previous step. A suitable choice for such task is \( \alpha \)-expansion [3], since it handles an arbitrary number of labels. Given \( H_i \) and an initial labeling \( L_{i-1} \) in the \( i \)-th iteration, labeling \( L_i \) is estimated using \( \alpha \)-expansion minimizing energy \( \hat{E} \). Note that \( L_0 \) is determined by \( \alpha \)-expansion in the first step. The number of the model instances \(|H_i|\) is fixed during this step and the energy must decreases:

\[
\hat{E}(L_i, H_i) \leq \hat{E}(L_{i-1}, H_i).
\]

3. The Outlier Removal part of the algorithm determines points fitting no instance. A common technique to remove outliers in energy minimization tasks is to add an outlier label – several approaches are discussed in [14]. We experienced that our separation of \( \alpha \)-expansion and \( \alpha \)-expansion removal steps does not affect the accuracy (it is described in Section 3.2 in depth). However, it facilitates the adaptive parameter setting. All points are marked outlier for which \( \phi_p(\theta_p, p) > \gamma_p \), where \( \gamma_p \) is a threshold corresponding to the instance for which point \( p \in P \) is assigned to.\(^3\)

4. Model Fitting instance parameters w.r.t. the assigned points. The obtained instance set \( H_i \) is re-fitted using the labeling provided by \( \alpha \)-expansion. The number of the model instances \(|H_i|\) is constant. We use the Weiszfeld-algorithm [38], an iterative re-weighted least squares approach, to \( L_1 \) model fitting. Note that if the processing time is crucial, \( L_2 \) fitting is an appropriate choice, however, \( L_1 \) minimization is more robust.

The overall energy \( \hat{E} \) can only decrease or stay constant during this step since it consists of three terms: (1) \( \hat{E}_d \) – the sum of the assignment costs minimized, (2) \( \hat{E}_s \) – a function of the labeling \( L_i \), fixed in this step and (3) \( \hat{E}_c \) – which depends on \(|H_i|\) and \(|H_i| = |H_{i+1}| \) so \( \hat{E}_c \) remains the same. Thus

\[
\hat{E}(L_i, H_{i+1}) \leq \hat{E}(L_i, H_i).
\]

3.2. Convergence

The convergence of the algorithm depends on the number of instances in \( H_i \) and energy \( \hat{E}(L_i) \). Note that in Multi-X, energy \( \hat{E} \) can increase if the number of instances decreases but the overall state of the algorithm converges.

Definition 3 (State) The state \( n \in \mathbb{N} \times \mathbb{R} \) of Multi-X is a pair of numbers, where \( n \) and \( \mathbb{R} \) represent the number of instances \(|H_i|\) and the value of the energy \( E(L_i) \), respectively.

Convergence is reached when

\[
(|H_i|, \hat{E}(L_i)) = (|H_{i+1}|, \hat{E}(L_{i+1})).
\]

The convergence of the instance number is ensured since the first step, Median-Shift, does not introduce new instances. It can only reduce the instance number by merging two or more into a cluster corresponding to a mode. Outlier removal and \( \alpha \)-expansion steps do not affect the instances. The last, model re-estimation part changes the instance parameters but does not remove or add instances. Therefore,

\[
|H_i| \geq |H_{i+1}| \geq 0.
\]

The convergence of the energy is ensured for every number of instances \(|H_i|\), since the labeling minimizes the energy and the number of possible labelings is finite; the outlier removal step considers points as outliers if their energy is higher than the constant associated with an outlier. The outlier number is not penalized in the energy, therefore, \( \hat{E} \) cannot increase. The instance re-estimation part changes the instance parameters but does not remove or add instances. Thus

\[
\hat{E}(L_{i-1}, H_i) \geq \hat{E}(L_{i}, H_i).
\]

Overall convergence is ensured due to the previously described properties. Suppose that operator \( \geq \colon \mathbb{N} \times \mathbb{N} \times \mathbb{R} \rightarrow \{0, 1\} \) is the lexicographic ordering of states \((n_1, e_1), (n_2, e_2) \in \mathbb{N} \times \mathbb{R}\) is defined as

\[
(n_1, e_1) \geq (n_2, e_2) \iff n_1 \geq n_2 \lor (n_1 = n_2 \land e_1 \geq e_2).
\]
Due to conditions Eq. 6 and Eq. 7, \((n_i, e_i) \geq (n_{i+1}, e_{i+1})\), where \((n_i, e_i)\) and \((n_{i+1}, e_{i+1})\) are the states of the convergence in the \(i\)-th and the \((i + 1)\)-th iterations, respectively. Since the number of possible labelings is finite and the lower limit for instance number is zero, the convergence is ensured and reached as soon as \((n_i, e_i) = (n_{i+1}, e_{i+1})\).

### 3.3. Implementation Details

**Adaptive Median-Shift.** For mode-seeking, the Median-Shift proposed in [30] was chosen. In contrast to Mean-Shift, it does not generate new elements in the vector space since it always return an element of the input set. With the Tukey-medians as modes, it is more robust than Mean-Shift [30]. However, we replaced Locality Sensitive Hashing [6] with Fast Approximated Nearest neighbors [24] to achieve higher speed.

To avoid manual bandwidth selection, we adopted the automatic procedure proposed in [10] which sets the bandwidth \(e_i\) of the \(i\)-th instance to the distance of the instance and its \(k\)-th neighbor. Thus each instance has its own bandwidth set automatically on the basis of the data.

**Adaptive Outlier Threshold.** Recognition of data points not belonging to any instance is critical for robust performance. The approach common in the literature is an outlier label balanced by a weight [14, 15]. This solution works well, but the balancing parameter does not have any geometric interpretation and it is hard to set automatically. We therefore do not use outlier label in the \(\alpha\)-expansion. Instead, all points which are farther from the assigned instances than a threshold are marked outlier after the labeling step. In practice, we have not experienced any significant difference in accuracy between the two approaches, but the threshold is easier to set.

Reflecting the fact that the inliers of each instance may have different distribution, a specific outlier threshold \(\gamma_i\) is assigned to each. Each \(\gamma_i\) is estimated as follows: first, distances \(\delta_{i,k} := \phi_k(p_i, \theta_i)\) of the \(i\)-th model instance and the \(k\)-th point are computed and sorted for every distinct instances \((\delta_{i,k} \leq \delta_{i,k+1}, \ i \in \{1, 2, ..., |H|\}, \ k \in \{1, 2, ..., |P|\}\). Threshold \(\gamma_i\) is set to distance \(\delta_{i,k}\) as

\[
k^* := \arg \max_{k=1, ..., |P| - 1} \delta_{i, k+1} - \delta_{i, k}, \ \gamma_i := \delta_{i, k^*}.
\]

Note that we prefer to limit each \(\gamma_i\) to a threshold estimated from the input, e.g. as the mean distance of the points.

**Model Validation.** Considering that a group of outliers may form spatially coherent structures in the data, we propose a post-processing step to remove those, statistically insignificant, models using cross-validation. The algorithm, summarized in Alg. 2, selects a minimal subset from the set of inlier points \(I/t\) times. In each iteration, an instance is estimated using the selected points and the distance to each point is calculated. The original instance is considered as a stable one if the mean of these distances is lower than a threshold \(\gamma\). Note that \(\gamma\) is the outlier threshold used in the previous sections and it is set adaptively.

**Algorithm 2 Model Validation.**

| Input: \(I\) – inlier points, \(t\) – trial number, \(\gamma\) – outlier threshold | Output: \(R \in \{\text{true}, \text{false}\}\) – response |
| 1: \(\hat{D} := 0\) |
| 2: for \(i := 1\) to \(t\) do |
| 3: \(\text{MSS} := \text{SelectMinimalSubset}(I)\) |
| 4: \(H := \text{ModelEstimation(MSS)}\) |
| 5: \(D := \text{MeanDistanceFromPoints}(H, I)\) |
| 6: \(\hat{D} := \hat{D} + D/t\) |
| 7: \(R := \hat{D} < \gamma\) |

### 4. Experimental Results

We report the performance of Multi-X applied to the following Computer Vision problems: line and circle fitting, 3D plane and cylinder fitting to LIDAR point clouds, multiple homography fitting, two-view and video motion segmentation.

**Simultaneous Line and Circle Fitting** is evaluated on 2D edges of banknotes and coins. Edges are detected by Canny edge detector and assigned to circles and lines manually to create a ground truth segmentation. Each method started with the same number of initial model instances: twice the data point (e.g. edge) number. The evaluated methods are PEARL [7, 15], T-Linkage [20] and RPA [21] since they can be considered as the state-of-the-art and their implementations are available. PEARL and Multi-X fits circles and lines simultaneously, while T-Linkage and RPA sequentially.

Fig. 2 shows three test scenes (rows) and the results of the methods (columns). Detected models are drawn onto the images: green ones are the true positives, false negatives and false positives are denoted by red and blue, respectively.

Table 1 reports the number of false negative and false positive models. Multi-X achieved the lowest error for all test cases.

**Multiple Homography Fitting** is evaluated on the AdelaideRMF homography dataset [39] used in most recent publications. AdelaideRMF consists of 19 image pairs of different resolutions with ground truth point correspondences assigned to planes (homographies). To generate initial model instances the technique proposed by Barath et al. [1] is applied: a single homography is estimated for each correspondence using the point locations together with the re-

---

http://www.diegm.uniud.it/fusiello/demo/jlk/  
http://www.diegm.uniud.it/fusiello/demo/rpa/
Figure 2: Line and circle fitting on Canny edge maps (b) extracted from three images (left). Results of T-Linkage (c), PEARL (d) and Multi-X (e): true positive (green), false negative (red), false positive (blue) instances. The number of fully overlapping instances (second row, PEARL) is indicated by blue numerals 2 and 3 in the center of the image in row (2), column (d).

The number of false positive (FP) and false negative (FN) instances for simultaneous line and circle fitting is shown in Table 1.

Table 1: The number of false positive (FP) and false negative (FN) instances for simultaneous line and circle fitting.

| Image num. (see Fig. 2) | (1) | (2) | (3) |
|-------------------------|-----|-----|-----|
| FP | FN | FP | FN | FP | FN |
| PEARL [15] | 1 | 0 | 3 | 0 | 5 | 3 |
| T-Linkage [20] | 0 | 1 | 1 | 3 | 0 | 6 |
| RPA [21] | 0 | 1 | 0 | 2 | 0 | 5 |
| Multi-X | 0 | 0 | 0 | 0 | 0 | 1 |

Two-view Motion Segmentation is evaluated on the AdelaideRMF motion dataset consisting of 21 image pairs of different sizes and the ground truth – correspondences assigned to their motion clusters.

Simultaneous Plane and Cylinder Fitting is evaluated on the LIDAR point cloud data (see Fig. 5). The annotated motion clusters are denoted by color. Table 4 shows comparison with state-of-the-art methods when all methods are tuned separately for each test case. Results are the average and minimum misclassification errors (in percentage) of ten runs. All results except that of Multi-X are copied from [37]. For Table 5, all methods use fixed parameters. For both test types, Multi-X achieved higher accuracy than the reference methods.
Table 2: Misclassification error (%) for the two-view plane segmentation on AdelaideRMF test pairs: (1) johnsonna, (2) johnsonnb, (3) ladysymon, (4) neem, (5) oldclassicswing, (6) sene.

| # of planes | T-Lnk [20] | RCMSA [27] | RPA [21] | Multi-H [1] | Multi-X |
|-------------|------------|------------|----------|-------------|---------|
| (1)         | 4          | 4.02       | 4.16     | 4.02        | 6.48    | 5.90    | 5.07    | **3.75** |
| (2)         | 6          | 18.18      | 18.18    | 18.17       | 21.49   | 17.95   | 18.33   | **4.46** |
| (3)         | 2          | 5.49       | 5.91     | 5.06        | 5.91    | 7.17    | 9.25    | **0.00** |
| (4)         | 3          | 5.39       | 5.39     | 3.73        | 8.81    | 5.81    | 3.73    | **0.00** |
| (5)         | 2          | 1.58       | 1.85     | 0.26        | 1.85    | 2.11    | 0.27    | **0.00** |
| (6)         | 2          | 0.80       | 0.80     | 0.40        | 0.80    | 0.80    | 0.84    | **0.00** |
| Avg.        |             | 5.91       | 6.05     | 5.30        | 7.56    | 6.62    | 6.25    | **1.37** |
| Med.        |             | 4.71       | 4.78     | 3.87        | 6.20    | 5.86    | 4.40    | **0.00** |

Table 3: Misclassification errors (%), average and median) for two-view plane segmentation on all the 19 pairs from AdelaideRMF test pairs using fixed parameters.

|        | T-Lnk [20] | RCMSA [27] | RPA [21] | Multi-H [1] | Multi-X |
|--------|------------|------------|----------|-------------|---------|
| Avg.   | 44.68      | 23.17      | 15.71    | 14.35       | **9.72** |
| Med.   | 44.49      | 24.53      | 15.89    | 9.56        | **2.49** |

Figure 4: AdelaideRMF (top) and Hopkins (bot.) examples. Color indicates the motion Multi-X assigned a point to.

database consists of traffic signs, balusters and the neighboring point clouds truncated by a 3-meter-radius cylinder parallel to the vertical axis. The ground truth, points were manually assigned to signs (planes) and balusters (cylinders).

Multi-X is compared with the same methods as in the line and circle fitting section. PEARL and Multi-X fit cylinders and planes simultaneously while T-Linkage and RPA sequentially. Table 6 reports that Multi-X is the most accurate in all test cases except one.

Video Motion Segmentation is evaluated on 51 videos of the Hopkins dataset [34]. Motion segmentation in video sequences is the retrieval of sets of points undergoing rigid motions contained in a dynamic scene captured by a moving camera. It can be seen 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 [34].

Table 7 shows that the proposed method outperforms the state-of-the-art: SSC [8], T-Linkage [20], RPA [21], Grdy-RansaCov [22], ILP-RansaCov [22], and J-Linkage [33]. Results, except for Multi-X, are copied from [22]. Fig. 4 shows two frames of the tested videos.

4.1. Indicator of Robustness

In our experience, the spatial coherence term $E_s$ is a good indicator of the model quality. In 68% of the cases, selecting the solution with the lowest $E_s$ value out of three runs lead to the most accurate choice. In 78%, there was at most a 3% difference in misclassification error between the best and the selected solution. As Multi-X is fast, it is
Multi-X is a randomized algorithm - a property it inherits from the initialization stage.

|     | KF [4] | RCG [19] | T-Lnkg [20] | AKSWH [32] | MSH [37] | Multi-X |
|-----|--------|----------|-------------|------------|----------|---------|
| Avg. | Min.  | Avg. | Min.   | Avg. | Min. | Avg. | Min. | Avg. | Min. |
| (1) | 8.42 | 4.23 | 13.43 | 9.52 | 5.63 | 2.46 | 4.72 | 2.11 | 3.80 | 2.11 | **3.45** | **1.41** |
| (2) | 12.53 | 2.81 | 13.35 | 10.92 | 5.62 | 4.82 | 7.23 | 4.02 | 3.21 | 1.61 | **2.27** | **0.40** |
| (3) | 14.83 | 4.13 | 12.60 | 8.07 | 4.96 | 1.32 | 5.45 | 1.42 | 2.69 | 0.83 | **1.45** | **0.41** |
| (4) | 13.78 | 5.10 | 9.94 | 3.96 | 7.32 | 3.54 | 7.01 | 5.18 | 3.72 | 1.22 | **0.61** | **0.30** |
| (5) | 16.87 | 14.55 | 26.51 | 19.54 | 4.42 | 4.00 | 9.04 | 8.43 | 6.63 | 4.55 | 5.24 | **1.80** |
| (6) | 16.06 | 14.29 | 16.87 | 14.36 | 1.93 | 1.16 | 8.54 | 4.99 | 1.54 | 1.16 | **0.62** | **0.00** |
| (7) | 33.43 | 21.30 | 26.39 | 20.43 | **1.06** | 0.86 | 7.39 | 3.41 | 1.74 | 0.43 | 5.32 | **0.00** |
| (8) | 31.07 | 22.94 | 37.95 | 20.80 | 3.11 | 3.00 | 14.95 | 13.15 | 4.28 | 3.57 | **2.63** | **1.52** |

Table 4: Misclassification errors (%) for two-view motion segmentation on the AdelaideRMF dataset. All the methods were tuned separately for each video by the authors. Tested image pairs: (1) cubechips, (2) cubetoy, (3) breadcube, (4) gamebiscuit, (5) breadtoycar, (6) biscuitbookbox, (7) breadcubechips, (8) cubebreadtoychips.

| RPA | RCMSA | T-Lnkg | AKSWH | Multi-X |
|-----|-------|--------|-------|---------|
| Avg. | 5.62 | 9.71 | 46.83 | **2.97** |
| Med. | 4.58 | 8.48 | 39.42 | **0.00** |

Table 5: Misclassification errors (%, average and median) for two-view motion segmentation on all the 21 pairs from the AdelaideRMF dataset using fixed parameters.

| PEARL [15] | T-Lnkg [20] | RPA [21] | Multi-X |
|------------|-------------|----------|---------|
| Avg. | 10.63 | 57.46 | 46.83 | **8.89** |
| (1) | 10.88 | 41.79 | 53.39 | **4.72** |
| (2) | 37.34 | 52.97 | 61.64 | **2.84** |
| (3) | 38.13 | 38.91 | 41.41 | **19.38** |
| (4) | 17.20 | 51.83 | 53.34 | **16.83** |
| (5) | **17.35** | 61.77 | 51.21 | 21.72 |
| (6) | 6.12 | 12.49 | 80.45 | **5.72** |

Table 6: Misclassification error (%) of simultaneous plane and cylinder fitting to LIDAR data. See Fig. 5 for examples.

 feasible to apply it repeatedly and select the solution with the lowest $E_s$ value.

### 4.2. Processing Time

Multi-X is orders of magnitude faster than currently available Matlab implementations of J-Linkage, T-Linkage and RPA. Attacking the fitting problem with a technique similar to PEARL and SA-RCM, it is significantly faster since it benefits from high reduction of the number of instances in the Median-Shift step (see Table 8).

| # | (1) M | T | (2) M | T | (3) M | T |
|----|-------|---|-------|---|-------|---|
| 100 | 0.1 | 0.4 | 0.1 | 0.3 | 0.1 | 0.2 | 0.1 | 0.4 |
| 500 | 2.0 | 14.0 | 3.2 | 8.4 | 2.1 | 8.4 | 0.8 | 7.0 |
| 1000 | 5.1 | 102.8 | - | - | - | - | 7.5 | 120.9 |

Table 8: Processing times (sec) of Multi-X (M) and T-Linkage (T) for the problem of fitting (1) lines and circles, (2) homographies, (3) two-view motions, (4) video motions, and (5) planes and cylinders. The number of data points is shown in the first column.

### 5. Conclusions

A novel multi-class multi-instance model fitting method was proposed. It combines global energy minimization and median-based mode-seeking in the parameter domain and its key parameters are set adaptively.

Multi-X outperforms significantly the state-of-the-art in multiple plane segmentation (homography detection), multiple rigid motion detection, motion segmentation, simulta-
neous plane and cylinder fitting, and on 2D edge interpretation (simultaneous circle and line fitting). The method runs in time approximately linear in the number of input data points which is an order of magnitude faster than available implementations of commonly used methods.\footnote{The source code and the datasets for line-circle and plane-cylinder fitting will be made available with the publication.}

References

[1] D. Barath, J. Matas, and L. Hajder. Multi-H: Efficient recovery of tangent planes in stereo images. In BMVC, 2016. 2, 5, 7

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. PAMI, 2004. 2

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. PAMI, 2001. 4, 6

[4] T.-J. Chin, H. Wang, and D. Suter. Robust fitting of multiple structures: The statistical learning approach. In ICCV, 2009. 8

[5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. PAMI, 2002. 3

[6] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In SoCG, 2004. 5

[7] A. Delong, L. Gorelick, O. Veksler, and Y. Boykov. Minimizing energies with hierarchical costs. IJCV, 2012. 2, 5

[8] E. Elhamifar and R. Vidal. Sparse subspace clustering. In CVPR, 2009. 7, 8

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 1981. 1, 3

[10] B. Georgescu, I. Shimshoni, and P. Meer. Mean shift based clustering in high dimensions: A texture classification example. In ICCV, 2003. 5

[11] N. Guil and E. L. Zapata. Lower order circle and ellipse hough transform. Pattern Recognition, 1997. 1

[12] P. V. C. Hough. Method and means for recognizing complex patterns, 1962. 1

[13] J. Illingworth and J. Kittler. A survey of the hough transform. Computer Vision, Graphics, and Image Processing, 1988. 1

[14] H. Isack. Spatially coherent multi-model fitting. CS Dept., University of Western Ontario, London, Canada, 2009. 4, 5

[15] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. IJCV, 2012. 2, 3, 5, 6, 7, 8

[16] Y. Kanazawa and H. Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In BMVC, 2004. 1

[17] N. Lazic, I. Givoni, B. Frey, and P. Aarabi. Floss: Facility location for subspace segmentation. In ICCV, 2009. 6, 7

[18] H. Li. Two-view motion segmentation from linear programming relaxation. In CVPR, 2007. 3

[19] H. Liu and S. Yan. Efficient structure detection via random consensus graph. In CVPR, 2012. 8

[20] L. Magri and A. Fusiello. T-Linkage: A continuous relaxation of J-Linkage for multi-model fitting. In CVPR, 2014. 2, 5, 6, 7, 8

[21] L. Magri and A. Fusiello. Robust multiple model fitting with preference analysis and low-rank approximation. In BMVC, 2015. 2, 5, 6, 7, 8

[22] L. Magri and A. Fusiello. Multiple model fitting as a set coverage problem. In CVPR, 2016. 2, 7, 8

[23] J. Matas, C. Galambos, and J. Kittler. Robust detection of lines using the progressive probabilistic hough transform. CVIU, 2000. 1

[24] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP, 2009. 5

[25] T. T. Pham, T.-J. Chin, K. Schindler, and D. Suter. Interacting geometric priors for robust multimodel fitting. TIP, 2014. 2, 3

[26] T. T. Pham, T.-J. Chin, J. Yu, and D. Suter. Simultaneous sampling and multi-structure fitting with adaptive reversible jump mcmc. In NIPS, 2011. 6, 7

[27] T. T. Pham, T.-J. Chin, J. Yu, and D. Suter. The random cluster model for robust geometric fitting. PAMI, 2014. 2, 6, 7, 8

[28] R. T. Rockafellar and R. J.-B. Wets. Variational analysis. Springer Science & Business Media, 2009. 4

[29] P. L. Rosin. Ellipse fitting by accumulating five-point fits. PRL, 1993. 1

[30] L. Shapiro, S. Avidan, and A. Shamir. Mode-detection via median-shift. In ICCV, 2009. 3, 5

[31] M. Stricker and A. Leonardis. ExSel++: A general framework to extract parametric models. In CAIP, 1995. 2

[32] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In ICCV, 2009. 8

[33] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In ECCV, 2008. 2, 3, 6, 7, 8

[34] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In CVPR, 2007. 7

[35] J. W. Tukey. Mathematics and the picturing of data. In ICM, 1975. 3

[36] E. Vincent and R. Lagnani`ere. Detecting planar homographies in an image pair. In ISIPA, 2001. 1

[37] H. Wang, G. Xiao, Y. Yan, and D. Suter. Mode-seeking on hypergraphs for robust geometric model fitting. In ICCV, 2015. 2, 6, 8

[38] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. Tohoku Mathematical Journal, 1937. 3, 4

[39] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter. Dynamic and hierarchical multi-structure geometric model fitting. In ICCV, 2011. 5

[40] L. Xu, E. Oja, and P. Kultanen. A new curve detection method: randomized hough transform (rht). PRL, 1990. 1

[41] W. Zhang and J. Kosecká. Nonparametric estimation of multiple structures with outliers. In Dynamical Vision, 2007. 2

[42] M. Zuliani, C. S. Kenney, and B. Manjunath. The multi-ransac algorithm and its application to detect planar homographies. In ICIP, IEEE, 2005. 1