MultiLink: Multi-class Structure Recovery via Agglomerative Clustering and Model Selection

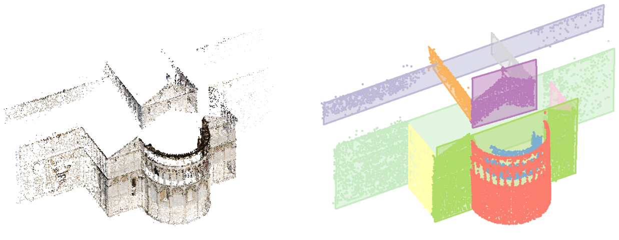
Luca Magri Filippo Leveni Giacomo Boracchi
Politecnico di Milano, DEIB
name.surname@polimi.it

Abstract

*We address the problem of recovering multiple structures of different classes in a dataset contaminated by noise and outliers. In particular, we consider geometric structures defined by a mixture of underlying parametric models (*e.g.* planes and cylinders, homographies and fundamental matrices), and we tackle the robust fitting problem by preference analysis and clustering. We present a new algorithm, termed MultiLink, that simultaneously deals with multiple classes of models. MultiLink combines on-the-fly model fitting and model selection in a novel linkage scheme that determines whether two clusters are to be merged. The resulting method features many practical advantages with respect to methods based on preference analysis, being faster, less sensitive to the inlier threshold, and able to compensate limitations deriving from hypotheses sampling. Experiments on several public datasets demonstrate that MultiLink favourably compares with state of the art alternatives, both in multi-class and single-class problems. Code is publicly made available for download*[1].

1. Introduction

Multi-structure recovery (also known as multi-model fitting) aims at organising a set of input data in multiple geometric structures described by a few underlying parametric models. This is a fundamental step in many Computer Vision and Pattern Recognition applications such as motion segmentation [40], template detection [17], primitive fitting in point clouds [12], and multi-body Structure-From-Motion [10, 22, 26]. The vast majority of fitting methods identifies multiple structures from a *single-class* of models (*e.g.* 3D planes to fit building facades) [7, 13, 18, 32], and cannot solve *multi-class* structure recovery problems, where structures have to be identified from several classes of models (*e.g.* cylinders, planes). Multi-class recovery problems have been much less investigated [3,4,21,42,43], despite they are frequently met in practical applications and their solution typically enrich the interpretation of raw data. Dealing with diverse classes of models enables a higher level of abstraction and, in a broader perspective, can be reckoned as an attempt to bridge the semantic gap separating raw visual content from reasoning. For instance, consider the 3D point cloud X in Fig. 1, where the underlying structures (groups of 3D points) can be identified by solving a 2-class multi-model fitting problem with respect to Θ_p and Θ_c, the class of planes and cylinders, respectively. Here, the proposed MultiLink successfully partitions the point cloud in s structures $X = U_1 \cup \ldots \cup U_s$, and for each structure it decides whether to fit a plane or a cylinder, providing an high-level description of the Cathedral.



(a) Input point cloud (b) Recovered structures

Figure 1: An example of multi-class recovery. MultiLink extracts planes and cylinders from the input point cloud. Structure membership is color coded.

Structure recovery algorithms have to address multiple challenges, including being robust to noise and structured-outliers, and to successfully disentangle a chicken-and-egg dilemma, since model fitting and the association of points to structures are strictly intertwined. In fact, structures are defined as set of points satisfying the same model, but models can be instantiated only once structures have been defined. Needless to say, all these problems become more difficult when coping with multiple structures *and* diverse classes of models, as we have to consider different data interpretations in order to automatically choose both the number of structures s and the right class of model for each of them. De-

[1] https://github.com/magrilu/multilink.git

spite many practical applications (*e.g.*, Scan2Bim) solved by multi-model fitting, this remains an ill-posed problem: the number/type of models are in the eye of the beholder, and either simplifying assumptions or some priors are required to make the problem tractable.

Recent solutions tackle these difficulties in two ways. On the one hand, *optimization-based* solutions make single-class algorithms able to simultaneously deal with multiple classes of explanatory models by means of sophisticated optimization techniques that minimize a suitable energy function. Multi-X [3] and Prog-X [4] are prominent examples of this approach. On the other hand, *preference-based* methods tackle the structure recovery problem from a clustering perspective, and typically result in simpler greedy algorithms. Unfortunately, preference-based algorithms are still far from solving general multi-class fitting problems in realistic conditions. There are only few solutions that either are not robust to outliers [42, 43] or that assume [21] that model classes are strictly contained within each other (*e.g.* $\Theta_p \subset \Theta_c$, since planes are also cylinders) and cannot cope with models that do not conform this "Chinese-boxes" assumption (*e.g.* cylinders and spheres).

Contributions

In this work, we present *MultiLink*, the first among preference-based algorithms that is both robust to outliers and can deal with general (*i.e.*, non-nested) classes of models, representing a viable alternative to sophisticated optimization-based solutions. Specifically, MultiLink implements an iterative agglomerative clustering scheme that successfully combines on-the-fly hypotheses sampling and a model-selection criterion to determine when structures can be conveniently merged, and, in that case, which is the most suited model among multiple classes. The transparent greedy scheme underpinning MultiLink overcomes some major limitations of other preference-based algorithms, and allow to simultaneously identify multiple, not-necessarily nested, classes of models. To the best of our knowledge, our idea of combining preference information and model selection directly into agglomerative clustering has never been explored before. MultiLink results in a multi-class multi-model fitting method that is:

General: MultiLink copes with arbitrary model classes, whereas existing robust preference-based methods, like [18, 21, 32], are limited to a single class or nested classes.

Accurate: MultiLink achieves compelling accuracy compared to recent state-of-the-art multi-class methods [3, 4, 13, 18] on both synthetic and real public datasets for structure recovery.

More stable and faster than other preference-based methods: MultiLink alleviates severe dependencies on the initial hypotheses sampling and the choice of the inlier threshold, resulting in a stable algorithm even in the single-class scenario. In addition, the proposed cluster-merging scheme doesn't need to update distances at every iteration, and this reduces the computational complexity w.r.t. [18, 21, 32].

2. Prior work

While single-class structure recovery is a well established topic in Computer Vision that has attracted a lot of interest, Multi-class structure recovery in its full generality has been addressed only recently [3, 4, 21, 44], with a few earlier work that have addressed this problem only for specific applications [26, 27, 29, 36]. Here we survey those methods that are most relevant for the proposed solution, while in Sec. 2.1 we recall some aspects of preference-based methods that are important to understand MultiLink.

Both single and multi-class recovery methods can be broadly categorized in two major approaches: *optimization-based* and *preference-based*.

Optimization-based methods were originally conceived to deal with a single class of models [13,45]. These methods minimize an objective function composed by a data fidelity term that measures goodness of fit, and a penalty term to account for model complexity. Additional terms can be included to promote spatial coherence or further priors for the application at hand [23]. These methods typically follow a two-steps *hypothesize-and-verify* procedure and generate, during the hypothesize step, a set of models via random sampling. Hence, during the verification step, they select the models minimizing the energy function. A variety of techniques have been proposed, depending on the specific definition of the energy function: from early approaches, such as [35], to more advanced methods that rely on graph labeling [24], alpha-expansion [5, 8], convex relaxation [2] and integer linear programming [15, 20, 26]. Particular relevant to our work is [8], where model-refitting is used to escape form local minima and improve convergence of energy minimization.

Very recently, optimization-based methods addressed the challenges of fitting instances from multiple classes: Multi-X [3] combines alpha-expansion with a mean-shift step carried on in each model class, and Prog-X [4] further improves this approach by interleaving the hypothesize and the verify stages. All these methods can be considered as sophisticated implementations of a (multi-)model selection criteria, as they select the simplest models using as measure of simplicity a global energy function.

Preference-based methods represent the second mainstream approach and our work falls in this category. In contrast with optimization-based methods that concentrate on models, preference-based ones put the emphasis on structures, and cast multi-structure recovery as a clustering problem, following an *hypothesize-and-clusterize* scheme. During the first step, tentative models are randomly sampled and points are embedded in a *preference space* based on the preferences granted by the hypothesized models. A wide variety of techniques have been proposed to segment preferences: hierarchical schemes [32, 34] such as T-linkage [18], Kernel Fitting [6], robust matrix factorization [19, 31], biclustering [9, 30], higher order clustering [1, 11, 14, 46] and hypergraph partitioning [16, 25, 38, 39, 41]. In this work we build upon hierarchical clustering, that is robust to outliers and, in contrast to divisive alternatives, does not need to know the number of structures in advance.

The preference-based approach has been only lately investigated to address multi-class problems: Multi-class Cascaded T-linkage (MCT) [21] assumes that model classes are nested and it executes T-linkage in a stratified manner, from the most general to the simplest class. Then, the model selection tool [36] GRIC (Geometric Robust Information Criterion) is used to compare each cluster deduced from the general class with the corresponding nested clustering deduced from simpler structures. Unfortunately, MCT is not designed for models belonging to classes that are not strictly contained in each other. The motion segmentation algorithm presented in [42] can be seen as a multi-class preference method as well. The focus is on nearly-degenerate structures, which are difficult to characterize for real data. To overcome this limitation, rather than dealing with elusive model selection problems, authors fit models of multiple classes to data, and combine the resulting partitions through an ad-hoc multi-view spectral clustering. Regretfully, this cannot handle data contaminated by outliers.

It is also worth mentioning that structure recovery solutions based on deep-learning are now appearing. For instance, [44] tackles the multi-class multi-model fitting problem by learning, from annotation, an embedding of the points, that are subsequently segmented by k-means.

MultiLink follows a different approach and combines the strengths of optimization and preference-based methods owning both the neat formulation of model selection methods and the flexibility of clustering. Specifically, we extend preference representation to *jointly* deal with multiple, not necessarily nested, mixed classes of models.

2.1. Preference analysis

The core concept of preference analysis is the *preference embedding*, that was used for *single* class of models Θ_1. Let X be the input data and $\varepsilon > 0$ a fixed inlier threshold, the preference function of $x_i \in X$ w.r.t. a model $\vartheta_j \in \Theta_1$ is:

$$p(x_i, \vartheta_j) = \begin{cases} \phi(e_{ij}) & \text{if } e_{ij} = \text{err}(x_i, \vartheta_j) \leq \varepsilon \\ 0 & \text{otherwise} \end{cases}, \qquad (1)$$

where $\text{err}(x_i, \vartheta_j)$ measures the residual e_{ij} between a model ϑ_j and a point x_i, and ϕ is a monotonic decreasing function in $[0, 1]$ such that $\phi(0) = 1$. Intuitively, $p(x_i, \vartheta_j)$ represents the preference that a point x_i grants to a model ϑ_j: the lower $\text{err}(x_i, \vartheta_j)$, the higher the preference.

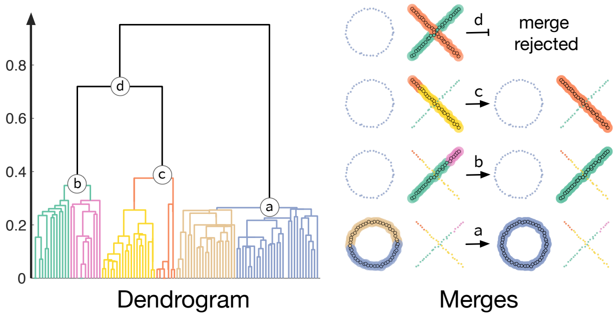


Figure 2: MultiLink combines single-linkage clustering and GRIC. Clusters are merged as long as the GRIC score improves when fitting suitable models on-the-fly. Colors indicate how cluster aggregation proceeds in the dendrogram.

In practice, a finite pool $\mathcal{H} = H_1 \subset \Theta_1$ of m model hypotheses is randomly sampled from the parameter space Θ_1 and used to compute the preferences as in Eq. (1), defining an embedding of data points as vectors[2] $p(x, \mathcal{H})$ in the unitary cube $[0, 1]^m$. The rationale behind this embedding is that points belonging to the same structure share similar preferences, thus are nearby. Several metrics have been proposed to measure distance/similarity in the preference space, *e.g.* Ordered Residual Kernel [6], Jaccard [32] and Tanimoto [18] distance. In this work we rely on the Tanimoto distance that, given two points $u, v \in X$, is defined as $d(u, v) = 1 - \tau(p(u, \mathcal{H}), p(v, \mathcal{H}))$, where

$$\tau(a, b) = \frac{\langle a, b \rangle}{\|a\|^2 + \|b\|^2 - \langle a, b \rangle}. \qquad (2)$$

3. Proposed method

Here we present the key principles of MultiLink by an illustrative structure recovery problem (Fig. 2). At a high level, MultiLink follows a *hypothesize-and-clusterize* framework with two major differences w.r.t. existing solutions: first, the preference embedding is computed by sampling hypotheses from a "*multi-class*" preference space, in our example $\mathcal{H} \subset \Theta_l \cup \Theta_c$, the space of lines and circles. Second, the clustering is performed in the preference space using single-linkage (see dendrogram in Fig. 2). The major novelty of MultiLink is to determine *whether* each pair of closest clusters can be conveniently merged by selecting *which class* of models describes its union at best. This problem is solved by *fitting* new models *on-the-fly*, in order to better describe points belonging to the two clusters, and by deciding whether to merge them through a *model selection* criterion. Specifically, we use GRIC to determine the best interpretation of the data in terms of both data fidelity

[2] Given $\mathcal{H} = \{h_1, \ldots, h_m\}$, $p(x, \mathcal{H})$ is a succinct notation for $[p(x, h_1), \ldots, p(x, h_m)]$.

and model complexity, hence we merge clusters when the model fitted on their union yields a GRIC score lower than the sum of individual GRIC scores. With reference to Fig. 2, clusters are first merged in a in a circle, then in b and c in a line, while in d the merge is inhibited since no model can describe the two segments better than two separate lines.

3.1. MultiLink

MultiLink is summarized in Algorithm 1 and starts by randomly sampling a finite pool of tentative models $\mathcal{H} = H_1 \cup \dots \cup H_K$ where each H_k comes from its corresponding model class Θ_k (line 1). Hence, for each $x \in X$, preferences are computed using Eq. (1) yielding vectors $p(x, \mathcal{H})$ living in the preference space $[0, 1]^{|\mathcal{H}|}$ (line 2). The agglomerative step then consists in an iterative block (lines 3-15), where the inter-cluster distance between two clusters U and V is defined using the single-linkage rule

$$d(U, V) = \min_{u \in U, v \in V} d(u, v)\,. \tag{3}$$

With a slight abuse of notation, let $\vartheta_k(U)$ denote a model in Θ_k fitted to the cluster $U \subseteq X$. Then, for each model class Θ_k, the models $\vartheta_k(U), \vartheta_k(V), \vartheta_k(U \cup V)$ are fitted *on-the-fly* to U, V and $U \cup V$, respectively. The GRIC is then computed to assess the cost of these $3K$ models (lines 7-10). The GRIC cost of a cluster $U \subseteq X$, w.r.t. a model class Θ_k, is defined as [36]:

$$g_k(U) = \sum_{x_i \in U} \rho \left(\frac{\mathrm{err}(x_i, \vartheta_k(U))}{\sigma} \right)^2 + \lambda_1 d|U| + \lambda_2 \mu, \tag{4}$$

where $\mathrm{err}(x_i, \vartheta_k(U))$ is, as in Eq. (1), a data fidelity term that measures the residual between $x_i \in U$ and the fitted model $\vartheta_k(U) \in \Theta_k$. Here σ is an estimate of the residuals standard deviation, and ρ is a robust function that bounds the loss at outliers. The other two terms in Eq. (4) account for model complexity: d is the dimension of the manifold Θ_k, μ the number of model parameters, and $|U|$ the cardinality of U.

We use GRIC to determine in a principled manner whether U and V are conveniently aggregated and, in that case, which class of models describes $U \cup V$ at best (line 11-13). To this purpose, we compare the GRIC scores of the union $g_k(U \cup V)$ with the sum of the costs for two separate fits $g_k(U) + g_k(V)$. There are two alternatives:

1. There exists $\hat{k}$ yielding the minimum cost at $g_{\hat{k}}(U \cup V)$:

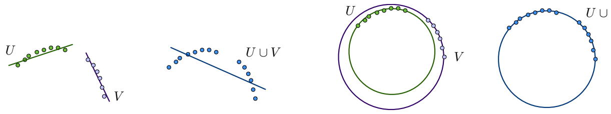
$$\exists\, \hat{k}\colon g_{\hat{k}}(U \cup V) \leq g_k(U) + g_k(V) \quad \forall k \tag{5}$$

2. There exists $\hat{k}$ yielding the minimum cost at $g_{\hat{k}}(U) + g_{\hat{k}}(V)$:

$$\exists\, \hat{k}\colon g_{\hat{k}}(U) + g_{\hat{k}}(V) < g_k(U \cup V) \quad \forall k \tag{6}$$

Algorithm 1: MultiLink

```
Input: $X$ data, $\{\Theta_k\}_{k=1,\dots,K}$ model classes, $\varepsilon$
       inlier threshold, $\lambda_1, \lambda_2$ GRIC parameters.
Output: A partition of the data in structures
        $X = U_1 \cup \dots \cup U_s$.
/* Preference embedding over $\mathcal{H}$ */
1  Sample hypotheses $\mathcal{H} = H_1 \cup \dots \cup H_K$;
2  Compute preferences $p(x_i, \mathcal{H})$ $\forall x_i \in X$ as in (1);
/* Clustering starts */
3  Put each point $x_i$ in its own cluster $\{x_i\}$;
4  Compute inter-cluster distances $d$ as in (3);
5  while $\min d < +\infty$ do
6      Find clusters $(U, V)$ with the min distance;
       // Fit models, compute GRIC (4)
7      for $k = 1 \dots K$ do
8          Fit a model $\vartheta_k(U)$ to $U$ and compute $g_k(U)$;
9          Fit a model $\vartheta_k(V)$ to $V$ and compute $g_k(V)$;
10         Fit a model $\vartheta_k(U \cup V)$ to $U \cup V$ and
            compute $g_k(U \cup V)$;
       // Test merge condition
11     if $\exists \hat{k}\colon g_{\hat{k}}(U \cup V) \leq g_k(U) + g_k(V)$ $\forall k$ then
12         merge $U$ and $V$, the structure is $\vartheta_{\hat{k}}(U \cup V)$;
13         update inter-cluster distances $d$ as in (3);
14     else
15         $d(U, V) = +\infty$;
```



(a) Rejected with lines (b) Accepted with circles

Figure 3: Illustration of merge with GRIC.

In the first case, we merge U and $V$and consider $U \cup V$ as a structure of class $\Theta_{\hat{k}}$ (line 12). A new cluster is created and the inter-cluster distances are updated accordingly to the single-linkage rule (line 13). Otherwise, U and V are considered as separate structures, and the merge is inhibited by setting $d(U, V) = +\infty$ (line 15).

Our greedy strategy guides the clustering towards the simplest explanations of the data, since a merge is accepted only when the union of two clusters leads to a decrease in the GRIC cost. Fig. 3 shows that, when the U and V do not belong to the same structure, the advantage of fitting a single model over $U \cup V$ is cancelled out by the large residuals, and the higher GRIC cost prevents the merge. On the contrary, when the points of U and V belong to the same structure of a class $\Theta_{\hat{k}}$, the cost $g_{\hat{k}}(U \cup V)$ is lower because the two regularization terms increase the cost of fitting two separate models. Note that since models are fitted on-the-fly (line 8-10) during clustering, MultiLink explores new models over $\Theta = \Theta_1 \cup \cdots \cup \Theta_K$ while being driven by the linking procedure, therefore these models turn to be more relevant than those randomly sampled in $\mathcal{H}$.

Rationale During the *hypothesize* step, MultiLink samples models from all the classes Θ and embeds all the points in the same space, disregarding the class they refer to. Clusters are initially formed by aggregating points that share similar preferences, thus there is not yet a model associated to each cluster as the latter might be too small. Nevertheless, clusters typically gather inliers of the same model, since these points have strong preferences in common. Models are then associated to clusters during the *clusterize* step, as soon as clusters contain a sufficient number of points to allow fitting models from different classes. Therefore, in the early stages, MultiLink exploits reliable preference information associated to single points, and only later exploits reliable information from clusters. During cluster merging, MultiLink performs on-the-fly model fitting to recover model instances missed in the initial sampling and, at the same, to time identify different class of models.

Implementation details The inlier threshold ε should be tuned on the level of noise that is however typically difficult to estimate in presence of multiple models. Therefore, we perform our experiments using both a manually tuned fixed inlier threshold, and combining the meta-heuristic based on Silhouette index [33]. This latter is a way to automatically estimate the best ε inside a search interval. Despite this typically yields a very rough estimate of the best ε, it was enough for MultiLink to achieve good performance confirming the good stability of the method to ε values. Finally, let us remark that, in principle, an ad-hoc inlier threshold ε_k could be defined for each class Θ_k, even though we used the same ε for all the classes in our experiments.

Preference embedding was implemented as described in Eq. (1). As in [18], we set ϕ as a Gaussian function $\phi(x) = \exp(-x^2/\sigma^2)$ where $\sigma^2 = -\varepsilon^2/\log(0.05)$. The robust function ρ in Eq. (4) is $\rho(x) = \min(x, r - d)$, where r is the dimension of the ambient space of the data and d is the dimension of the model manifold [36].

While testing the merging conditions in Eq.(5) and (6) on U and V, if any of the two clusters is too small to instantiate a model in Θ_k, we resort to the merging criterion in T-linkage [18] which merges two clusters when there exists in $\mathcal{H}$ at least one sampled model explaining all the points of the two clusters. We also find beneficial to remove, from the initial pool of models $\mathcal{H}$, those hypotheses that occurred by chance. To this purpose, we validate each hypothesis $h \in \mathcal{H}$ through the preprocessing stage presented in [30], which implements a Gestalt principle to determine whether h is significant or not. A model is not significant when its supporting points are uniformly distributed in space. This condition is verified by checking wheter the number of inliers at distance $k\varepsilon$ is almost k times inliers at ε.

	T-linkage	MCT	MultiLink
Class	single $\mathcal{H} = H_1$	nested $H_1 \subset H_2$	multi $\mathcal{H} = H_1 \cup \ldots \cup H_K$
Models	sampled	sampled	sampled and on-the-fly
Linkage	centroid-link.	centroid-link.	single-linkage & GRIC
Model selection	no	a posteriori	inside clustering

Table 1: Differences among T-linkage, MCT and MultiLink

3.2. Features and benefits

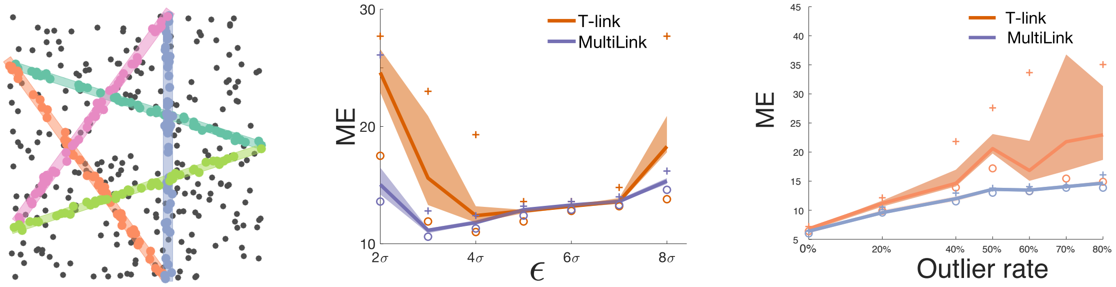
MultiLink shares a few peculiar features with preference methods based on hierarchical clustering, such as T-linkage and MCT, but, thanks to some crucial differences, it overcomes the main limitations of these two algorithms, even in the single-class scenario. Tab. 1 summarizes the key improvements of MultiLink w.r.t. T-linkage and MCT.

Preference embedding: MultiLink implements a "multi-class" preference embedding that, in contrast to MCT, allows to deal with multiple classes of models not necessarily nested. However, multi-class embedding alone would not be enough to recover multi-class structures. In fact, this in T-linkage would result in the more general models always prevailing over the simpler ones, being model complexity ignored.

On-the-fly sampling: A second key feature of MultiLink is that, during iterations, it fits additional models to those derived in the original sampling $\mathcal{H}$. Models fitted on-the-fly are more reliable than those in $\mathcal{H}$ as they are instantiated on emerging clusters of inliers rather than on minimal sample sets. This makes MultiLink more robust than T-linkage and MCT w.r.t. sampling imbalance. For this reason its performance are more stable, as demonstrated in the experiments. Note, that models are typically fitted on-the-fly in fast close-form, while more demanding non-linear refinement can be used at the end, once clusters are formed.

Agglomerative scheme: T-linkage and MCT exploit a variant of centroid-linkage specifically designed for the preference space. In practice, two clusters U and V are merged when there exists at least a sampled model in $\mathcal{H}$ that passes through all the points of $U \cup V$ within a maximum distance of ε, or equivalently when the centroids of the two clusters have distance lower than 1. Therefore, even a single outlier in a cluster can heavily affect the cluster centroid resulting in over-segmentation. The single-linkage mechanism of MultiLink not only sidesteps this problem, but also reduces the computational burden, since its distance update mechanism is more efficient for single-linkage as it does not involve the computation of a cluster representative.

Model selection: Note that GRIC was also adopted in



(a) *Worst* result (3σ)

(b) ME vs. inlier threshold ε

(c) ME vs. outlier rate

Figure 4: Line fitting: MultiLink *vs* T-linkage. Fig. 4a: the *worst* result of MultiLink for $\varepsilon = 3\sigma$ over 50 trials. Figs. 4b and 4c: the median ME (solid line), IQR (shaded area), maxima (+) and minima (○) as a function of ε and outlier ratio.

MCT to determine which model(s) fit to each structure at the end of a stratified clustering. In contrast, MultiLink uses GRIC as a key ingredient *during* clustering. Moreover, tuning λ_1, λ_2 in MCT is rather difficult, since GRIC compares a varying number of model instances. MultiLink instead always compares one-vs-a-pair of models, and we safely set $\lambda_1 = 1, \lambda_2 = 2$ in all our experiments, as in Tab. 6a.

4. Experimental validation

We test MultiLink on both single-class and multi-class structure recovery problems. We first address 2D primitive fitting problems (Sec. 4.1), which represent a standard benchmark for structure recovery algorithms, and demonstrate that MultiLink outperforms MCT. Then, we test MultiLink on real-world datasets (see Fig. 7) for the estimate of two view relations from correspondences (Sec. 4.2) and for video motion segmentation (Sec. 4.3). In all these experiments, we show that MultiLink favorably compares or performs on par with recent multi-class structure recovery alternatives [4]. The results of MultiLink on 3D primitive fitting in a sparse input point cloud [47] are reported in Fig. 1a.

Performance is measured, as customarily, in terms of misclassification error (ME), *i.e.* the fraction of misclassified points w.r.t. the ground-truth labelling. If not stated otherwise, we always report ME averaged over 5 runs. Parameters used to configure MultiLink in each dataset are reported in Tab. 6a. Matlab code of MultiLink is available on-line at [48].

4.1. 2D fitting problems

We first consider a single-class structure recovery problem (line fitting) and show that MultiLink outperforms T-linkage in terms of accuracy, robustness to outliers and runtime. Then, we address a multi-class structure recovery problem (conic fitting) and show that MultiLink outperforms MCT and PEARL.

Line fitting We consider T-linkage as the closest alternative to MultiLink on the single-class problem illustrated in Fig. 4a. The dataset, containing multiple lines corrupted by noise and outliers, and the MATLAB implementation of T-linkage are from [49].

Fig. 4b reports the median ME over 50 runs as a function of the inlier thresholds $\varepsilon = n\sigma$, where σ is the noise level and $n = 2, \dots, 8$. This plot displays the inter-quantile range (IQR) of the ME (shadowed regions), together with the minimum (○) and maximum (+) errors. Both these methods were provided with the *same* initial hypotheses $\mathcal{H}$, leading to the *same* preference representation of points. This plot indicates that MultiLink outperforms T-linkage, achieving the best performance both in terms of median, maximum and minimum ME. Remarkably, except for $\varepsilon = 2\sigma$ where both methods over-segment the data, MultiLink provides very stable outputs, as indicated by the small IQR. This confirms that, fitting new models on-the-fly during clustering, improves the stability of MultiLink w.r.t. both ε and the randomly sampled hypotheses $\mathcal{H}$. On the contrary, T-linkage, which rely exclusively on the fixed pool of models $\mathcal{H}$, suffers of higher instability across multiple runs as demonstrated by its large IQR and maximum error. We also calculated the ME on *star5* dataset at increasing outlier rates (Fig. 4c), and MultiLink always outperforms T-linkage, demonstrating to be more robust.

Despite, due to merge rejections, MultiLink features in principle a worst-case complexity that T-linkage, in practice it exhibit no computational overheads, as the single-linkage scheme makes the clustering phase of MultiLink (0.26 s) faster than that of T-linkage (0.76 s), on average. This experiment confirms that MultiLink outperforms its closest, single-class, alternative both in terms of effectiveness and efficiency, being more stable thus more practical.

Line and conic fitting Fig. 5 illustrates 2D simulated datasets used in [21] to recover lines, circles and parabolas, where we report the *worst* results attained by MultiLink. All the datasets comprise instances of lines and circles, while (a), (b) and (c) include also parabolas. Here, MultiLink recovers all the geometric structures even in the *worst* runs.

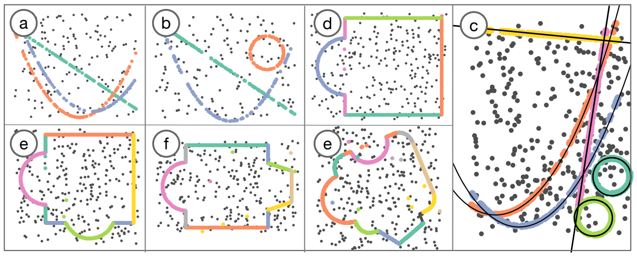
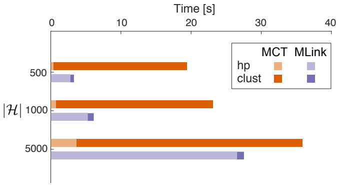


Figure 5: *Worst* results by MultiLink on conic fitting.

	ε	λ_1	λ_2
conic (a-c)	0.180	1	2
conic (d-g)	0.900	1	2
plane seg.	0.070	1	2
2-view seg.	0.058	1	2
video seg.	[0.01, 0.3]	1	2

(a) Parameters used



(b) Execution times

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
PEARL	6.00	16.22	14.44	8.05	8.33	17.38	19.21
MCT	**0.67**	**2.00**	2.33	5.23	7.12	5.38	6.23
MLink	2.17	2.13	**1.55**	**1.83**	**0.87**	**2.46**	**4.28**

(c) ME

Figure 6: Quantitative results on conic fitting: ME (bottom) and execution times on the problem of Fig. 5.a (right).

Tab. 6c shows the performance of PEARL [13] and MCT as reported in [21], and indicates that MultiLink favorably compares with both methods achieving a lower ME on 5 cases out of 7. In the two cases where MCT scores the best, the ME of MultiLink is rather small. Fig. 6b compares the running times of MCT and MultiLink on the dataset of Fig. 5.a w.r.t. $|\mathcal{H}|$, the number of initial hypotheses. Both algorithms are implemented in MATLAB, and the code of MCT is from [50]. As expected, on-the-fly fitting makes MultiLink more efficient than the cascaded approach of MCT. In fact, the clustering step of T-linkage, which we show on *star5* experiment to be slower than that in MultiLink, is repeated several times in MCT, resulting in longer executions. MultiLink spends most of the time in generating the hypotheses (light blue bars), whereas the actual clustering step takes far less (dark blue bars). However, we experienced that this is due to the optional pre-processing step [30] used to remove irrelevant models, whose computational burden can be drastically reduced in an optimized and parallel implementation.

4.2. Two-views relations

We test MultiLink on two-views segmentation over the popular Adelaide RMF dataset [40], which consists of 36 sequences of stereo images with correspondences corrupted by noise and outliers, and annotated ground-truth matches. Specifically, we first detect planar structures by fitting homographies, and then we perform motion segmentation. This latter was cast as a multi-class recovery problem as we fit both fundamental matrices, affine fundamental matrices, and homographies.

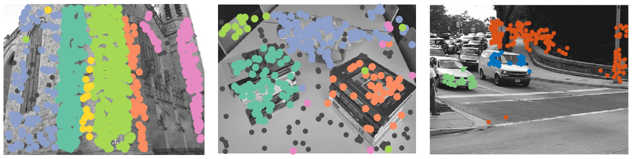


Figure 7: Sample results attained by MultiLink on plane (left) two-view (middle) and motion segmentation (right).

Plane segmentation (single-class) Results in Tab. 2a demonstrates that Prog-X and MultiLink achieve comparable best mean performance, albeit the latter yields more stable result. For fair comparison against MCT, which was used to fit a fundamental matrix and then to recover nested-compatible homographies using per-sequence tuned inlier threshold, we execute MultiLink by optimizing ε in the same way. Tab. 2b indicates that MultiLink is still the best performing algorithm. Furthermore, the difference in terms of ME between fixed and sequence-wised tuned ε for MultiLink is much smaller than for T-linkage, confirming that MultiLink is rather robust w.r.t. the choice of ε.

Two-views segmentation (multi-class) We carry out a two-view motion segmentation experiment on the 19 stereo images depicting moving objects. This dataset has been extensively used to estimate ego motions by fitting fundamental matrices, thus has become a benchmark for single-class multi-structure fitting [3, 4, 18, 40]. However, our preliminary tests suggested that some movements can also be reliably described by affine fundamental matrices, or even by homographies. Probably, these ground-truth motions can be deemed as quasi-degenerate. Therefore, we run MultiLink with three different classes of models: (1) *Fundamentals*: Θ_f the manifold of fundamental matrices; (2) *Affine fundamentals*: Θ_a the manifold of affine fundamental matrices; (3) *Mixed models*: where we consider Θ_f, Θ_a and Θ_h, the space of homographies.

Tab. 2c reports the mean ME averaged over the whole dataset, together with its standard deviation. We tested both MultiLink and T-linkage in the above three configurations using *fixed* parameters. Prog-X, Multi-X and PEARL were also tested on this dataset to fit fundamental matrices, and we report results from [4] accordingly. To test these methods on the affine fundamental and mixed models configurations, we modified the codes provided by authors of Multi-X and PEARL in [51] and [52]. This operation was not possible for Prog-X code which was not flexible enough to be used in other settings. Two relevant comments arise: first,

	Pearl	Multi-X	Prog-X	RPA	T-link	MLink
Mean	15.14	8.71	6.86	23.54	22.38	**6.46**
Std.	6.75	8.13	5.91	13.42	7.27	1.75

(a) Plane seg. *fixed parameters*

	T-link	MCT	MLink
Mean	6.60	6.13	**4.10**
Median	4.68	4.93	**2.70**

(b) Methods with ε *tuned* per-sequence

	Fundamental		Affine fund.		Mixed	
	Mean	Std.	Mean	Std.	Mean	Std.
MLink	8.59	4.67	9.84	4.09	**7.75**	4.54
T-link	32.20	50.33	41.90	7.95	38.78	8.21
Prog-X	10.73	8.73	-	-	-	-
Multi-X	17.13	12.23	10.5	2.90	9.53	1.43
Pearl	29.54	14.80	41.81	15.25	48.89	8.16

(c) Two-view seg. *fixed parameters*

	Mean	Std.
Multi-X	12.96	19.60
Prog-X	8.41	10.29
T-link +s (dim 3)	8.68	12.23
MCT +s	10.87	12.68
MLink +s (dim. 3)	**8.34**	11.93
MLink +s (mixed)	9.83	13.05

(d) Video seg. s= Silhouette index.

Table 2: Mean ME (in %) on real datasets. Averages over 5 runs on each sequence.

MultiLink outperforms all the competing methods in all the three configurations. Second, MultiLink can successfully perform multi-class fitting, achieving the lowest ME when the three models are mixedly used. This result is in agreement with findings in [42], and represents an interesting application where multi-class can be successfully employed to account for nearly degenerate data. When using only affine fundamental matrices, both MultiLink and T-linkage achieve higher ME than when using fundamental matrices, suggesting that affine fundamental matrices are not flexible enough to capture the motion diversity in the whole dataset.

4.3. Video motion segmentation (multi class)

Finally we test MultiLink on the video motion segmentation tasks of the Hopkins 155 benchmark [37]. This dataset consists of 155 video sequences with 2 or 3 moving objects whose trajectories can be approximated, under the assumption of affine projection, as a union of low dimensional subspaces. The dimension of the subspaces might vary depending on the type of motions in the dynamic scenes [28]. Therefore, we run MultiLink in two configurations: *i*) single-class, where we fit affine subspaces of dimension 3, and *ii*) multi-class, where we fit both affine subspaces of dimension 2 and 3 as mixed models. Tab. 2d compares performance of MultiLink against Multi-X and Prog-X, which were reported in [4]. Both Multi-X and Prog-X were executed with fixed parameters over the whole dataset. We thus configure MultiLink with all the parameters fixed, but automatically estimate the inlier threshold $\varepsilon \in [0.01, 0.3]$ in each sequence by means of a variant of the Silhouette index as described in [33]. Estimating ε in this way represents a very practical solution that is widely applicable in real-world scenarios. We also run T-linkage and MCT coupled with Silhouette index. The results by MultiLink with subspaces of dimension 3 are in line with the ones of Prog-X. In addition, this experiment confirms that our solution is stable, as it can successfully compensate for inaccurate estimates of ε. The advantages of adopting mixed models are not apparent on the average ME over the whole dataset, but we experienced that MultiLink with mixed classes consistently improves the results as long as natural video sequences with some degenerate motions are concerned (MultiLink with mixed classes achieves a ME of 1.37% on Traffic 3 and 3.14% on Traffic 2, in contrast to the configuration with subspaces of dimension 3 that scores 7.51% and 4.18% respectively). We suspect that, in a reasonably large number of sequences, 3-d subspaces are the right model to fit, and the mixed configuration actually degrades performance.

5. Conclusions

We presented MultiLink, a simple and effective algorithm to recover structures from different classes in data affected by noise and outliers. In particular, MultiLink can fit models from different classes during clustering steps, and includes a novel cluster-merging scheme that is based on on-the-fly model fitting and model selection through Gric. Experiments on both simulated and real data demonstrates that MultiLink is faster, more stable and less sensitive to sampling and to the inlier threshold than greedy alternatives based on preference analysis and agglomerative clustering such as T-linkage and MCT. In addition MultiLink favorably compares with optimization-based methods. All in all, MultiLink represents a very flexible framework that can be further extended by modifying cluster-merging conditions to accommodate for specific constraints coming from an application at hand. Finally, MultiLink offers an easy-to-manage tool to practitioners, for addressing the difficult and ubiquitous problem of multi-class structure recovery.

References

[1] Sameer Agarwal, Jongwoo Lim, Lihi Zelnik-manor, Pietro Perona, David Kriegman, and Serge Belongie. Beyond pairwise clustering. In *Conference on Computer Vision and Pattern Recognition*, pages 838–845, 2005. 3

[2] Paul Amayo, Pedro Piniés, Lina M Paz, and Paul Newman. Geometric multi-model fitting with a convex relaxation algorithm. In *Conference on Computer Vision and Pattern Recognition*, pages 8138–8146, 2018. 2

[3] Daniel Barath and Jiri Matas. Multi-class model fitting by energy minimization and mode-seeking. In *European Conference on Computer Vision*, pages 221–236, 2018. 1, 2, 7

[4] Daniel Barath and Jiri Matas. Progressive-x: Efficient, anytime, multi-model fitting algorithm. In *Conference on Computer Vision and Pattern Recognition*, pages 3780–3788, 2019. 1, 2, 6, 7, 8

[5] Daniel Barath, Jiri Matas, and Levente Hajder. Multi-h: Efficient recovery of tangent planes in stereo images. In *British Machine Vision Conference*, pages 13.1–13.13. BMVA Press, September 2016. 2

[6] Tat-Jun Chin, Hanzi Wang, and D. Suter. Robust fitting of multiple structures: The statistical learning approach. In *International Conference on Computer Vision*, pages 413–420, 2009. 3

[7] Tat-Jun Chin, David Suter, and Hanzi Wang. Multi-structure model selection via kernel optimisation. In *Conference on Computer Vision and Pattern Recognition*, pages 3586–3593, 2010. 1

[8] Andrew Delong, Olga Veksler, and Yuri Boykov. Fast fusion moves for multi-model estimation. In *European Conference on Computer Vision*, pages 370–384, 2012. 2

[9] Matteo Denitto, Luca Magri, Alessandro Farinelli, Andrea Fusiello, and Manuele Bicego. Multiple structure recovery via probabilistic biclustering. In *Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition*, pages 274–284. Springer, 2016. 3

[10] Andrew W Fitzgibbon and Andrew Zisserman. Multibody structure and motion: 3-d reconstruction of independently moving objects. In *European Conference on Computer Vision*, pages 891–906. Springer, 2000. 1

[11] Venu M. Govindu. A Tensor Decomposition for Geometric Grouping and Segmentation. *Conference on Computer Vision and Pattern Recognition*, 1:1150–1157, 2005. 3

[12] Christian Häne, Christopher Zach, Bernhard Zeisl, and Marc Pollefeys. A patch prior for dense 3d reconstruction in man-made environments. In *Joint 3DIM/3DPVT Conference*, pages 563–570. IEEE, 2012. 1

[13] Hossam Isack and Yuri Boykov. Energy-based geometric multi-model fitting. *International Journal of Computer Vision*, 97(2):123–147, 2012. 1, 2, 7

[14] Suraj Jain and Venu Madhav Govindu. Efficient higher-order clustering on the grassmann manifold. In *International Conference on Computer Vision*, pages 3511–3518, 2013. 3

[15] Hongdong Li. Two-view motion segmentation from linear programming relaxation. In *Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 2

[16] Shuyuan Lin, Guobao Xiao, Yan Yan, David Suter, and Hanzi Wang. Hypergraph optimization for multi-structural geometric model fitting. In *Conference on Artificial Intelligence*, volume 33, pages 8730–8737, 2019. 3

[17] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 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*, pages 3954–3961, June 2014. 1, 2, 3, 5, 7

[19] Luca Magri and Andrea 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. 3

[20] Luca Magri and Andrea Fusiello. Multiple models fitting as a set coverage problem. In *Conference on Computer Vision and Pattern Recognition*, pages 3318–3326, June 2016. 2

[21] Luca Magri and Andrea Fusiello. Fitting multiple heterogeneous models by multi-class cascaded t-linkage. In *Conference on Computer Vision and Pattern Recognition*, pages 7460–7468, 2019. 1, 2, 3, 6, 7

[22] Kemal Egemen Ozden, Konrad Schindler, and Luc Van Gool. Multibody structure-from-motion in practice. *Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1134–1141, 2010. 1

[23] Trung Thanh Pham, Tat-Jun Chin, Kaspar Schindler, and David Suter. Interacting geometric priors for robust multimodel fitting. *IEEE Transactions on Image Processing*, 23(10):4601–4610, 2014. 2

[24] Trung-Thanh Pham, Tat-Jun Chin, Jin Yu, and David Suter. The random cluster model for robust geometric fitting. In *Conference on Computer Vision and Pattern Recognition*, 2012. 2

[25] Pulak Purkait, Tat-Jun Chin, Hanno Ackermann, and David Suter. Clustering with hypergraphs: the case for large hyperedges. In *European Conference on Computer Vision*, pages 672–687, 2014. 3

[26] Konrad Schindler, David Suter, and Hanzi Wang. A Model-Selection Framework for Multibody Structure-and-Motion of Image Sequences. *International Journal of Computer Vision*, 79(2):159–177, 2008. 1, 2

[27] Markus Stricker and Aleš Leonardis. Exsel++: A general framework to extract parametric models. In *International Conference on Computer Analysis of Images and Patterns*, pages 90–97. Springer, 1995. 2

[28] Yasuyuki Sugaya and Kenichi Kanatani. Geometric structure of degeneracy for multi-body motion segmentation. In *International Workshop on Statistical Methods in Video Processing*, pages 13–25. Springer, 2004. 8

[29] Yasuyuki Sugaya and Kenichi Kanatani. Multi-stage unsupervised learning for multi-body motion segmentation. *IEICE Transactions on Information and Systems*, 87(7):1935–1942, 2004. 2

[30] Mariano Tepper and Guillermo Sapiro. A biclustering framework for consensus problems. *SIAM Journal on Imaging Sciences*, 7(4):2488–2525, 2014. 3, 5, 7

[31] Mariano Tepper and Guillermo Sapiro. Nonnegative matrix underapproximation for robust multiple model fitting. In *Conference on Computer Vision and Pattern Recognition*, pages 2059–2067, 2017. 3

[32] R. Toldo and A. Fusiello. Robust multiple structures estimation with J-Linkage. In *European Conference on Computer Vision*, pages 537–547, 2008. 1, 2, 3

[33] Roberto Toldo and Andrea Fusiello. Automatic estimation of the inlier threshold in robust multiple structures fitting. In *International Conference on Image Analysis and Processing*, pages 123–131, 2009. 5, 8

[34] Roberto Toldo and Andrea Fusiello. Image-consistent patches from unstructured points with J-Linkage. *Image and Vision Computing*, 31(10):756–770, 2013. 3

[35] P. H. S. Torr. Geometric motion segmentation and model selection. *Philosophical Transactions of the Royal Society of London*, A 356:1321–1340, 1998. 2

[36] Philip H. S. Torr. *Model Selection for Two View Geometry:A Review*, pages 277–301. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999. 2, 3, 4, 5

[37] Roberto Tron and René Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. 8

[38] Hanzi Wang, Guobao Xiao, Yan Yan, and David Suter. Mode-seeking on hypergraphs for robust geometric model fitting. In *International Conference on Computer Vision*, December 2015. 3

[39] 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*, 41(3):697–711, 2018. 3

[40] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter. Dynamic and hierarchical multi-structure geometric model fitting. In *International Conference on Computer Vision*, 2011. 1, 7

[41] Guobao Xiao, Hanzi Wang, Taotao Lai, and David Suter. Hypergraph modelling for geometric model fitting. *Pattern Recognition*, 60:748–760, 2016. 3

[42] Xun Xu, Loong Fah Cheong, and Zhuwen Li. Motion segmentation by exploiting complementary geometric models. In *Conference on Computer Vision and Pattern Recognition*, pages 2859–2867, 2018. 1, 2, 3, 8

[43] Xun Xu, Loong Fah Cheong, and Zhuwen Li. 3d rigid motion segmentation with mixed and unknown number of models. *Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, jan 2019. 1, 2

[44] Xun Xu, Loong-Fah Cheong, and Zhuwen Li. Learning for multi-model and multi-type fitting. *arXiv preprint arXiv:1901.10254*, 2019. 2, 3

[45] Jin Yu, Tat-Jun Chin, and David Suter. A global optimization approach to robust multi-model fitting. In *Conference on Computer Vision and Pattern Recognition*, pages 2041–2048. IEEE, 2011. 2

[46] Ron Zass and Amnon Shashua. A unifying approach to hard and probabilistic clustering. In *International Conference on Computer Vision*, volume 1, pages 294–301, 2005. 3

[47] http://www.diegm.uniud.it/fusiello/demo/samantha/. 6

[48] https://github.com/magrilu/multilink.git. 6

[49] http://www.diegm.uniud.it/fusiello/demo/jlk. 6

[50] http://www.dpia.uniud.it/fusiello/demo/mct/. 7

[51] https://github.com/danini/multi-x. 7

[52] http://mouse.cs.uwaterloo.ca/code/gco-v3.0.zip. 7