Fitting Multiple Heterogeneous Models by Multi-class Cascaded T-linkage

Luca Magri Andrea Fusiello
DPIA - Università degli Studi di Udine, Italy
name.surname@uniud.it

Abstract

This paper addresses the problem of multiple model fitting in the general context where the sought structures can be described by a mixture of heterogeneous parametric models drawn from different classes. To this end, we conceive a multi-model selection framework that extends T-linkage to cope with different nested classes of models. Our method, called MCT, compares favourably with the state-of-the-art on publicly available data-sets for various fitting problems: lines and conics, homographies and fundamental matrices, planes and cylinders.

1. Introduction

During the past decades until nowadays, the need of high-level geometric representations of visual content has incessantly inspired the research towards the design of robust multi-model fitting methods aimed at segmenting the data in meaningful structures by simultaneously estimating a collection of underlying parametric models[1]. In particular, most recent efforts focused on dealing with multiple instances of the same model in the presence of noise and outliers.

In this work we address the problem of multi-model fitting in the more general context where the sought structures can be described by a mixture of heterogeneous parametric models, i.e., of different type or class. To this end, we conceive a multi-model selection framework that extends T-linkage to cope with different nested classes of models (*e.g.* lines vs. circles, homographies vs. fundamental matrices). This problem is also referred to as multi-class/multi-model fitting.

The main motivations can be ascribed to two orders of reasons.

- Generally speaking, a wider choice between possible explanatory models enables higher level of abstraction and allows more flexibility in the interpretation of the data. In this sense, dealing with multiple model *and* classes of varying complexities can be regarded as a further step towards a better and more elaborate understanding of visual content to fill the gap between low-level vision and higher level reasoning.
- From a practical point of view, in some applications, degenerate structures may arise when the data do not provide enough constraints to determine a model uniquely. Then it is convenient to be able to switch to a simpler and more appropriate model to circumvent the pitfalls of ambiguous estimate which are extremely fragile to outliers.

The success of such endeavours involves the explicit integration of model selection criteria in the multi-structure recovery framework. In this regard, it is useful to keep in mind, as a caveat, a well-known aphorism of scientific folklore: "all models are wrong but some are useful". In many concrete situations, a qualitative dichotomy between models lags behind the quantitative nuances of continuous phenomena, and it may be puzzling to decide in favour of an interpretation of the data rather than for another.

For instance, in stereo vision, assuming a rigid scene, the borders between a planar setting – modelled by projectivities – and a full three-dimensional one – better described by the epipolar geometry– are faintly traced by the relative variation of the focal length of the camera compared to the depth of the scene. Similar difficulties arise, as instance, if a regular polygon with k sides has to be described either by several lines or by a unique circle. In these situations, the practical context of application should tip the scale of model selection towards the more appropriate class of structure for the task at hand.

Contributions On these premises, being aware that the usefulness of a model is "in the eye of the beholder", we present a multi-model recovery framework that offers a hierarchical interpretation of data in nested compatible models, which can be conveniently biased to serve specific pur-

[1] Although used somehow interchangeably, "structure" and "model" have different nuances. The term structure refers to the arrangement and relations among the data, and it is intrinsic to the data itself, whereas the model is "in the eye of the beholder", being the mathematical description of the data that an observer fits onto them.

poses exploiting user provided criteria. Specifically, we conceive a cascaded extension of T-linkage that sequentially extract simpler nested models starting form the more general ones. We leverage on geometric constraints (*e.g.* tangency between line and circle, or compatibility between homography and fundamental matrix) to guide the sampling towards simpler valid interpretation of the data and to break down a complex model selection task into a collection of easier local model selection instances.

Models belonging to different classes compete with each other only if they explain the same points and are geometrically compatible. In this way, all the intra-class model selection problems are implicitly dealt by T-linkage, whereas the inter-class model selection issues take explicitly the form of one-vs-many or one-vs-one model comparison, which can be efficiently tackled with classical tools such as GRIC [25].

2. Previous work

Even if, besides few exceptions, the most recent literature on abstract multi-model fitting has been mainly focused on the case of multiple instances of the same model, earlier work, tailored for specific applications, have faced, more or less directly, the challenges entailed by the selection of multiple models belonging to different classes.

In particular, a long and distinguished line of research on motion segmentation deals with the development of model selection strategy for the segmentation of different kinds of motion, both in pairs of images and video sequences. As an instance, the work of Torr [25, 24, 14], tackles the problem of Multi-body Structure and Motion (M-SaM) in two views and make use of fundamental matrices and homographies to robustly cope with degenerate configuration that arise in presence of planar structures and pure rotational camera displacements. Specifically, in [25] the geometric robust information criterion (GRIC), suited for geometric problems is introduced and the M-SaM problem is solved in a fashion reminiscent of sequential-Ransac: at the beginning, candidate motions are randomly instantiated using both homographies and fundamental matrices and their GRIC score is computed. Then, the models with the best GRIC are iteratively detected and their corresponding inliers are removed from the matches until all the data are exhausted. At the end, a segmentation is obtained minimising a cost function that balances between data fidelity and the complexity of the retrieved models. The well known limitations typical of sequential-Ransac affects as well this multi-class approach. In addition, the method is meant to handle only direct comparison between pairs of models, as, in one-vs-many model comparisons, the greedy selection of the model with the lowest GRIC systematically favours the simpler ones.

For these reasons, in [14], the greedy selection strategy is replaced by a quadratic binary integer programming. Along the same line, Schindler, Suter and Wang [15] extracts motions across multiple frames. At the beginning, the pool of mixed-motion hypotheses is pruned in order to avoid redundant model by reasoning on their consensus sets. Then a segmentation of the data is derived from the refined models exploiting integer linear programming coupled with model selection criteria, remarkably special care is reserved for the treatment of intersecting models.

Motion segmentation is also considered in [18, 19] where, under the assumption of affine camera model, the feature trajectories are organised in linear spaces rather than with fundamental matrices. In order to handle degenerate motions – that give rise to subspace of lower dimension – a multi-stage approach is presented: at first degenerate 2D subspace are detected and used to segment the data exploiting an expectation maximisation scheme, then the attained model are upgraded to general motion models. The main idea is that, since a degenerate motion is a special case of a general ones, an optimal solution for a degenerate structure is unchanged when optimised by assuming a more general model. This approach produces accurate results, but depends on good initialisation and is computational expensive. Moreover, in principle, if outliers contaminate the data, the promotion of a degenerate motion to a general one can easily deteriorate the result.

The simultaneous fitting of multi-class geometric primitives to 3D data is addressed in [17], where the selection of models instantiated by random sampling is performed through a tabu-search optimisation implementing the Minimum Description Length principle.

A trend common to all the previous approaches can be singled out: after building a set of candidate models, all the methods derive a segmentation of the data by optimising a model selection criteria, namely minimising a global energy function balancing geometric errors and model complexity expressed in diverse forms, each grounded either on statistics or information theory.

This idea has been systematised in the well established framework of PEARL [6]: a general multi-model fitting method that relies on an efficient iterative optimisation technique where the regularity of inlier clusters is also exploited. Its success is confirmed by the several variations that have sprouted from its paradigm, such as [12] able to enforce geometric constraints, or [1] with improved efficiency, just to name a few.

Although, in principle, PEARL can be extended to deal with heterogeneous structures by specifying model penalties conveniently, in practice, it has been customarily employed for the estimation of multiple model instances of the same class. A step towards the multi-class case has been recently made by Multi-X [2], that generalises and improves PEARL by combining it with a Mean-Shift step carried out separately for each distinct model class. Notably, an auto-

matic parameter setting strategy is proposed to alleviate the need of user intervention.

The extraction of multiple models is faced from a different perspective in [27], where the Generalised Principal Component Analysis is introduced. One of the main merit of this work is to elegantly formulate the multi-model fitting problem in the language of algebraic geometry to unify different specific applications as the recovery of a union of linear subspace, with possibly varying dimensions, by fitting, differentiating and dividing a set of homogeneous polynomials to data. Unfortunately this approach turns to be fragile in real noisy contexts.

Inspired by the Gestalt principles and by the studies on human perceptions of patterns, the *a-contrario* framework integrates probabilistic reasoning with robust estimation for single model [13] and multi-model fitting [10]. The idea behind it, echoing Minpran [16], is to exploit the unlikeliness of a structure, measured in term of the number of false alarm (NFA), to assess its validity. This approach is very flexible and can be used as a model selection criterion for the multi-class models scenario. As an instance, [11] proposes a probabilistic model selection method to jointly fit ellipses or lines to edge maps.

The recent work of [29], which deals with the problem of co-segmentation of motion trajectories, raises a relevant objection. Data are often typified as 'almost-but-not-quite' degenerate structure, and it is nearly unfeasible to select a model with respect to another in such situations. For this reason, the authors propose to use a spectral clustering fusion scheme to combine together multiple partitions based on "complementary geometric models": homographies and fundamental matrices. Unfortunately, in this way, only the segmentation horn of the multi-model fitting problem is addressed, and this solution can not be applied to recover the model parameters underlying the attained clustering, as the type of the models are not maintained during the clustering fusion.

3. A multi-model selection framework

In this work, we also exploit the idea of combining clustering deduced from different models, but rather than extracting a segmentation from a multi-class soup of heterogeneous models, we make models of different classes to compete with each other only when they insist on the same inliers sets and satisfy geometric constraints, breaking the model selection in many easy-to-solve tasks.

For the convenience of the reader we briefly recall the T-linkage algorithm in Section 3.1 , while Section 3.2 describes how we propose to extend this approach to the case of multi-class models.

3.1. T-linkage

At high-level, T-linkage [8] performs a two steps *first-represent-then-clusterise* scheme: at first, the input data are represented by the "preferences" they grant to a pool of provisional model hypotheses, then a greedy bottom-up agglomerative clustering is performed to yield a partition of the data merging points with similar preferences according to the Tanimoto distance.

The concept of preferences can be specified as follows: Let X be the set of data points, $\text{err}\colon X \times H \to \mathbb{R}$ an error function to measure residuals e_{ij} between data x_i and models h_j, and ϵ an inlier threshold provided as input. The method starts by generating a set of random tentative models $H = \{h_1, \ldots, h_m\}$ by drawing m subsets of data points with the minimum cardinality necessary to instantiate a model. Then an $n \times m$ matrix P is built by defining its (i, j)-th entry as

$$P(i,j) = \begin{cases} \phi(e_{ij}) & \text{if } e_{ij} = \text{err}(x_i, h_j) \leq \epsilon \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

ϕ is a decreasing function with range in $[0, 1]$ and such that $\phi(0) = 1$. In this work ϕ was set to be a Gaussian function $\phi(r) = \exp(-e^2/\sigma^2)$ where $\sigma^2 = -\epsilon^2/\log(0.05)$ in order to grant a preference of 0.05 in correspondence of residuals with magnitude ϵ. Each row P_i can be easily identified with a *preference function* $\text{pf}(x_i)$ of a given point x_i. The rationale is that points belonging to the same model will have similar preference, and therefore can be clustered in the conceptual space $[0, 1]^m$ to reveal the structures hidden in the data.

The preference function of a subset $Y \subseteq X$ is composed by the models that fit all the points in Y:

$$\text{pf}(Y) = \min_{x \in Y} \text{pf}(x). \quad (2)$$

The clustering algorithm proceeds in a bottom-up manner. At first every data point is put in its own cluster. The distance between clusters is computed as the Tanimoto distance [21] between the respective preference functions. Starting from singletons, each sweep of the algorithm merges the two clusters with the smallest Tanimoto distance, until all the preference sets of clusters are disjoint.

The parameters of the returned models are estimated by least squares fitting on each cluster of points. It is worth noting that, if outliers are not present in the data, the number of clusters is automatically detected by this algorithm. Moreover this preference approach is robust to outliers, that can be recognised as observations whose preferences deviate significantly from the rest of the data, and tend to emerge as micro-clusters, that can be pruned out a posteriori with probabilistic reasoning. The inlier threshold needs to be specified as input.

In principle, T-linkage could be used to extract a segmentation from a multi-class soup of heterogeneous models, but in this way one cannot recover the models underlying the clustering, as the class of the models are lost during the clustering fusion, in much the same way as in [29].

In the next section we shall describe our general solution based on T-linkage.

3.2. Multi-class Cascaded T-linkage

Without loss of generality, suppose we want to fit a mixture of models belonging to two *nested classes*, namely $\mathcal{A}$ and $\mathcal{B}$. In other words, we search for a partition of the data of the form $X = \bigcup_i A_i \cup \bigcup_j B_j$, where $A_i = \text{cs}(\alpha_i)$ and $B_j = \text{cs}(\beta_j)$ denote the consensus sets associated to models $\alpha_i \in \mathcal{A}$ and $\beta_j \in \mathcal{B}$ respectively. The compact notation $\text{cs}(\mu) = \{x \in X : \text{err}(x, \mu) \leq \epsilon\}$ implicitly assumes a convenient error function for each class of model and a proper inlier threshold to handle noise and to encode the desired structure resolutions. The number of structures of both classes is unknown.

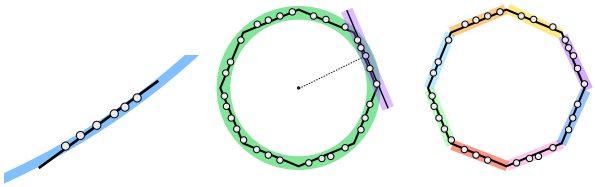
Let $\mathcal{A}$ be the more "general" class, in the sense that the structures underlying the simpler models can be also described in terms of models of type $\mathcal{A}$ without impacting data fidelity. On the other hand, there are instances in $\mathcal{A}$ that can not be explained by means of a unique model of $\mathcal{B}$ keeping the same level of accuracy. One can formalise this as follows:

1. $\forall B_j = \text{cs}(\beta_j), \beta_j \in \mathcal{B}\ \ (\exists \alpha_k \in \mathcal{A}\ \ (B_j \subseteq \text{cs}(\alpha_k)))$
2. $\exists A_i = \text{cs}(\alpha_i), \alpha_i \in \mathcal{A}\ \ (\forall \beta_k \in \mathcal{B}\ \ (A_i \neq \text{cs}(\beta_k)))$.

Classes of parametric models of varying complexity fall under this category, *e.g.* polynomials of different degree. For instance, $\mathcal{A}$ may consists of circles and $\mathcal{B}$ of lines. While a finite set of collinear points can be explained by a single circle of adequate radius (not uniquely determined) – see Fig. 1a – points in general position on a circumference can not be described by lines with a similar fitting error, and multiple instances are needed to achieve a similar accuracy. Another example consists of fundamental matrices and homographies: two images of a planar object can be related either by a homography or a fundamental matrix, even if the latter is not unique, but a single homography can not be used to describe with the same accuracy the epipolar geometry of a full 3D scene.

The pivotal observation is straightforward. Since every structure B_j can be described by models in $\mathcal{A}$, we may rewrite $B_j = \text{cs}(\alpha_k)$ and, in first instance, look for a partition $\mathcal{P}_\mathcal{A}$ of the form $X = \bigcup_i A_i \cup \bigcup_k A_k$ entirely induced by $\mathcal{A}$. This can be easily obtained with a first run of T-linkage where the pool of tentative hypotheses is instantiated by randomly sampling models of type $\mathcal{A}$.

To recover the less complex structures in $\mathcal{B}$, we can restrict our search to the refinement of the aforementioned



(a) (b) (c)

Figure 1: Circle $\mathcal{A}$ and lines $\mathcal{B}$ as nested models. (a) a circle can describe a set of collinear points, but it is not uniquely determined. (b) Sampling with geometric constraints from the segmentation induced by $\mathcal{A}$. A single point is sufficient to instantiate a line provided that it is tangent to the underlying circle. (c) The segmentation induced by $\mathcal{B}$.

partition $\mathcal{P}_\mathcal{A}$. Note that it may happen that some general models explain several nested structures – such as the circle in Fig. 1 whose consensus set includes the points on the sides of the octagon. The solution is therefore to perform T-linkage separately on every $\mathcal{A}$ structure, and then to adopt a model selection criterion to compare the sub-partitions induced by models in $\mathcal{B}$ internally on each A_i. As an illustration, in the case of Fig. 1 the model selection is used to decide between the interpretation of the data provided by the green circle and the one composed by the eight lines of the octagon perimeter. More generally, the model selection problem addressed here always compares a model of type $\mathcal{A}$ with one or several models of type $\mathcal{B}$.

The pool of tentative models of type $\mathcal{B}$ needed to perform T-linkage are generated enforcing geometric compatibility with the attained general models.

Our algorithm, called Multi-class Cascaded T-linkage or MCT in short, is summarised below. Points 4,7, and 8 will be expanded in the next paragraphs.

Algorithm 1 Multi-class Cascaded T-Linkage (MCT)

1. Extract models of class $\mathcal{A}$ with T-linkage
2. Reject outliers
3. For each $\mathcal{A}$ model, perform T-linkage sampling nested models $\mathcal{B}$ compatible with $\mathcal{A}$
4. Solve a model selection problem on each subset
5. Extract models of class $\mathcal{B}$ from the outliers with T-linkage
6. Reject outliers
7. Refine models

Please note that the algorithm does not have to compare all the possible combinations of type $\mathcal{B}$ models against the whole type $\mathcal{A}$ models since, inside each consensus set of the more general model, the selection of more specific structures is performed by the inner T-Linkage, which is also

able to automatically determine the correct number of instances. The first consensus set and all its nested structures (as extracted by T-Linkage) are hence compared through model selection.

Model selection. Several model selection criterion have been proposed in the literature. Let us call $L = -\frac{1}{2}\sum\left(\frac{e_i}{\sigma}\right)^2$ where e_i/σ are the normalised residuals, and let n be the number of data points, d the dimension of model manifold, k the number of model parameters and r the dimension of the measurement space (where data points belong); let $p = dn + k$ the total number of parameters to be estimated. The following are the model selection criteria we took into consideration (from [26]):

$$\text{AIC} \quad -2L + 2p \tag{3}$$

$$\text{BIC} \quad -2L + p\log(rn) \tag{4}$$

$$\text{MDL} \quad -2L + p/2\log_2(rn) \tag{5}$$

$$\text{GBIC} \quad -2L + \log(r)dn + \log(rn)k \tag{6}$$

$$\text{GRIC} \quad -2L + \lambda_1 dn + \lambda_2 k \tag{7}$$

When $\ell \geq 2$ type $\mathcal{B}$ models are compared against one type $\mathcal{A}$ model, the score is computed substituting k with ℓk to account for the actual number of parameters.

In our tests the GRIC resulted as the most versatile one, thanks to the tuning parameters λ_1 and λ_2. In our experiments we set λ_1=1 and adjusted λ_2 as reported in Tab. 1.

	k	d	r	λ_1	λ_2
Line	2	1	2	1	3
Conic	3	1			
Plane	4	2	3	1	4
Cylinder	5	2			
Fundamental	7	3	4	1	2.5
Homography	8	2			

Table 1: Model selection parameters

Outlier rejection. Following the route of the a-contrario approaches, outlying models are pruned using the statistical validation technique described in [23], the main difference is that we use it as a post-processing outlier rejection criteria on the attained models rather than as a pre-processing refinement of sampled model hypotheses. The idea is to compute the distribution of the cardinality of the consensus set of a model varying its inlier threshold and to express its unlikeliness in terms of NFA. In the most simple terms, if there are n inliers at a distance ϵ from a model, assuming a local uniform distribution of the residuals, we expect to have on average κ times more elements at a distance $\kappa\epsilon$. If this does not hold, the model has a very low probability of occurring by chance and can be retained as inliers.

Refinement. A final refinement step is carried out with the double goal of removing small models subsumed by a larger one, that may occasionally arise, and reassigning points to models in such a way that the result is not a partition of the points but a *cover*, which is a more appropriate output for multi-model fitting, as argued in [9]. With these aims, we re-assign the inlier points to the models produced by the previous step (each point is assigned to the models for which it is an inlier) and then solve a set cover problem in a greedy fashion, selecting the models in decreasing cardinality order, until all the inlier points have been covered by at least one model.

4. Experimental validation

This section is devoted to investigating the performances of MCT in several multi-class model fitting tasks, namely 2D and 3D primitive-fitting problems (in Sec.4.1 and 4.2 respectively) and two-views motion segmentation (Sec. 4.3). The inlier threshold ε have been tuned per-problem. The values of λ_1 and λ_2 are reported in Tab. 1. As customary the figure of merit is the misclassification error (ME), *i.e.* the percentage of misclassified point with respect to a ground truth labelling. Since the definition of ME does not envisage multiple labels, we assigned each inlier point to its closest model for the sole purpose of computing the ME.

The Matlab code of MCT is available on-line [30].

4.1. Line and conic fitting

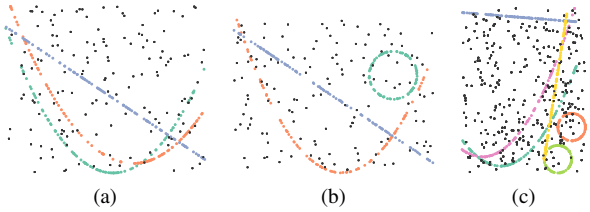
To begin with, we take into account some synthetic 2D primitive fitting problems. Specifically, in Fig. 2 we attempt to reproduce the experiment presented in [2]; the aim is to extract lines, parabolas and circles from data contaminated with 33% of outliers and a small amount of Gaussian noise.

It is interesting to note that, our method succeeds in dealing also with structures which do not conform with the nested-model assumption: we treat the linear and the parabolic model as nested ones, circles represent a third additional category that is extracted at the end from the outliers of the other two models. This is possible thanks to the refinement step, which remedies to the over-segmentation caused by the fact that the lower and the upper circular sectors of a circle can be accurately described as well by parabolas.

In second instance, we consider the problems collected in Fig. 3, where lines and circles are arranged in more complex configurations, resembling somehow blueprints of buildings. Tentative lines in circular structures are instantiated by sampling individual points and enforcing the direction of tangency with the circumference.

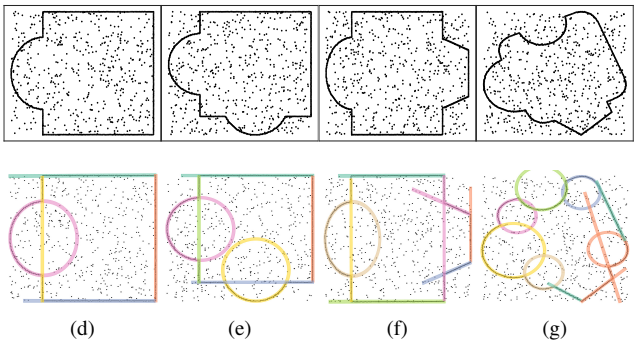
In Tab. 2 MCT is contrasted with PEARL[31] in terms of misclassication error (ME). Both methods were given the same preference matrix as input and results come from a single run; the parameters of both have been tuned to

achieve the best results and kept fixed in all instances. MCT consistently achieves the most accurate result in terms of ME for all the datasets.



(a) (b) (c)

Figure 2: **Line and conic fitting** on synthetic data generated as in [2]. Detection of multiple lines, parabolas and circles by MCT. Model assignment is colour coded.



(d) (e) (f) (g)

Figure 3: **Line and conic fitting** on "blueprint-like" synthetic data. Top row: input data. Bottom row: detection of multiple lines and circles by MCT. Model assignment is colour coded.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
PEARL [6]	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**

Table 2: ME (%) for **line and conic fitting** instances. Letters refers to Figs. 2 and 3.

To complete this set of experiments, we took in consideration circle and line detection from edge maps of real images. In Fig. 4 we collated the result of MCT with those of ELSD [11] a state-of-the-art native ellipse and line detector algorithm based on a-contrario reasoning specifically tailored for edge map classification and available online [32]. The results are promising, as our method succeeds in providing similar or even slightly better results without any further adaptations or specific assumptions.

4.2. Plane and cylinder fitting

Then we consider the problem of fitting 3D primitives to spatial data. Results on plane and cylinder fitting are

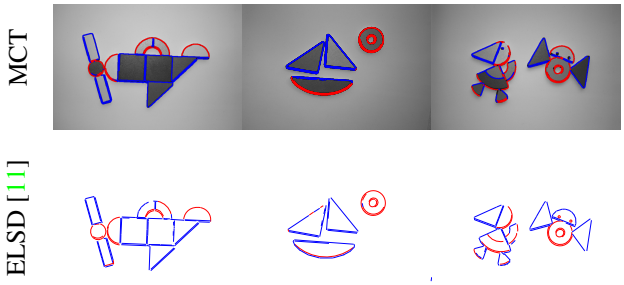


Figure 4: **Line and conic fitting**. Detection of multiple lines and circles on real images. Edgels class assignment is colour coded (circle / line) .

presented in Fig. 5. Since ground-truth is not available for these data, evaluation can be only subjective. The first two point clouds, representing a bridge and a nut, are taken from the Aim@Shape repository [33], while the third one is the Pozzoveggiani dataset [34], which has been used in several works, e.g. [12, 22, 9], but never in a multi-class framework: as a matter of fact, usually only planes are detected and the abside, which is a semi-cylinder, is left out.

The benefits of the the final refinement with set cover are shown in the bridge example, where portions of the pier have been fitted with cylinders which however ended up to be redundant because they were already covered by the plane that fitted the rest of the facade of the bridge. In the other two cases the refinement brought no visible improvement.

4.3. Homography and fundamental matrix fitting

To conclude, in this section we consider homographies and fundamental matrices. We used the "`cube*`" subset of the AdelaideRMF motion dataset [28], consisting of 8 image pairs depicting a cube and other objects, with ground truth point correspondences assigned to each object. In addition we manually labelled correspondences according to the face of the cube they belong to. These examples provide a true multi-class problem, where homographies (cube faces) and fundamental matrices (other objects) coexist, and the homographies are contained in a fundamental model. As the constrained sampling is concerned, we exploit the following compatibility constraint [7, 3]: H is said to be compatible with F if and only if the product $H^\top F$ is skew-symmetric, in formula:

$$H^\top F + F^\top H = 0. \tag{8}$$

In practice, following [5, 4] once a fundamental matrix F is recovered, three correspondence $\{m_i \leftrightarrow m_i'\}_{i=1}^3$ are sampled to determine a compatible planar projectivity as

$$H = A - e'(M^{-1}b)^\top, \tag{9}$$

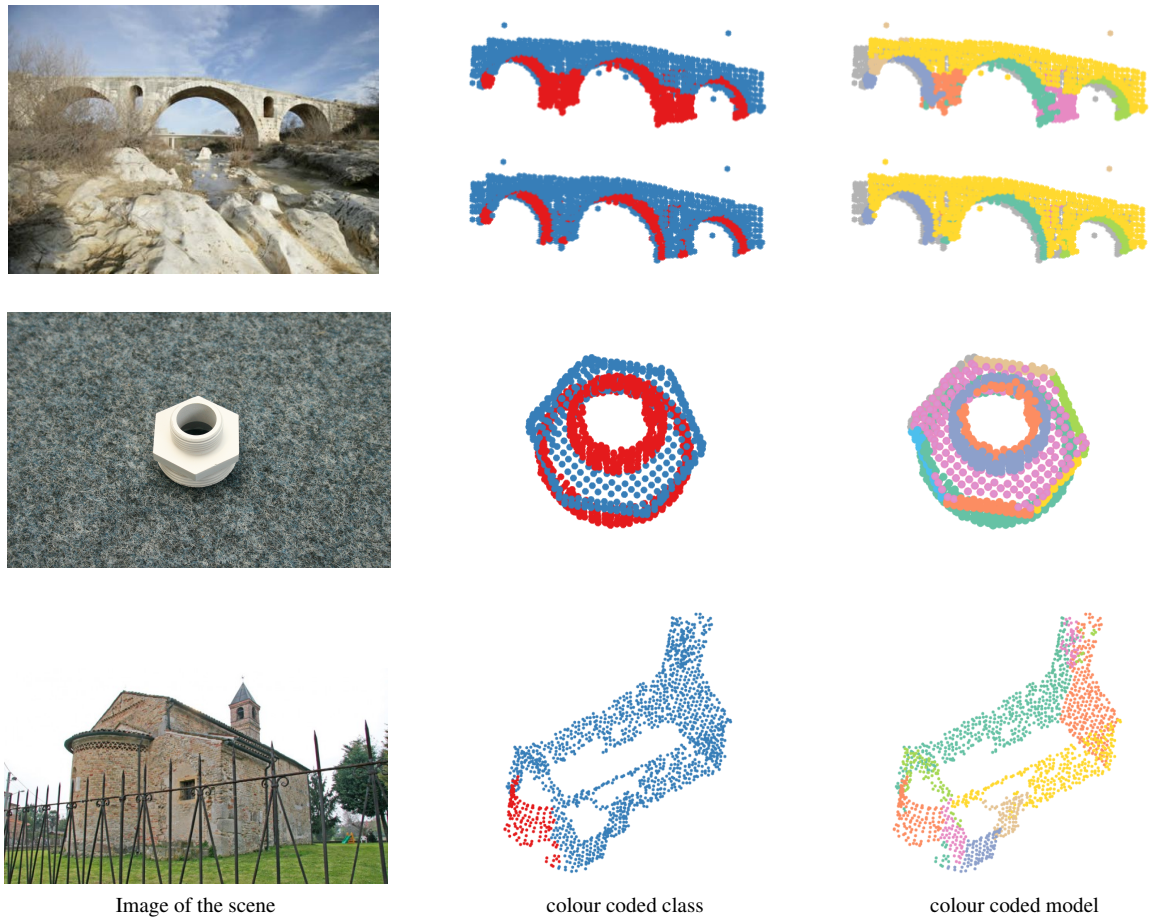


Image of the scene | colour coded class | colour coded model

Figure 5: **Plane and cylinder fitting** by MCT on 3D point clouds from Aim@Shape. The central column shows the class (point/cylinder) each point belongs to, whereas the rightmost column depicts the model each point belongs to. For the bridge the class/model assignment is show before (top) and after (bottom) the final refinement with set cover.

where $A = [e']_\times F$, $[e']_\times$ is the cross-product matrix of the epipole in the second view, b denotes the 3-dimensional vector defined component-wise as

$$b_i = (m'_i \times (Am_i))^\top (m'_i \times e') \|m'_i \times e'\|^{-2}, \qquad (10)$$

and M is a matrix collecting $m_i^\top$ by rows.

Result are shown in Fig. 6, where it can be appreciated that, in all the cases, the model selection mechanism promotes reliably the extraction of homography in correspondence of the planar facets of the cube while distinguishing, at the same time, the remaining moving object through fundamental matrices. Moreover, thanks to constrained sampling, the homographies estimated from the faces of the cube are geometrically consistent with the 3D motion of the object. The mean run time per image pair is 21s in Matlab on 2.6 GHz i7 machine.

To allow a comparison (albeit somehow contrived) with Multi-X and T-linkage [30], following the same protocol used in [2], we applied MCT to a single class problem, namely **multiple homography fitting** on the AdelaideRMF homography dataset [28], consisting of 19 image pairs with ground truth point correspondences assigned to planes (homographies). In order to bias MCT toward homographies we set $\lambda_1 = 500$ and $\lambda_2 = 0$ in this experiment. Result shown in Tab. 3 are mixed: Multi-X achieved the best median error, whereas MCT has the lowest mean error. In this case MCT works almost like regular T-linkage, except a first clustering is made with F matrices that biases the subsequent sampling of homographies. As observed by [20], initial recovery of an F matrix can profitably guide the sam-

	Multi-H	Multi-X	T-lnkg	MCT
mean	14.35	9.72	6.60	**6.13**
median	9.56	**2.49**	4.68	4.93

Table 3: ME (%) on AdelaideRMF for **multiple homography fitting**. The first two columns are copied from [2].

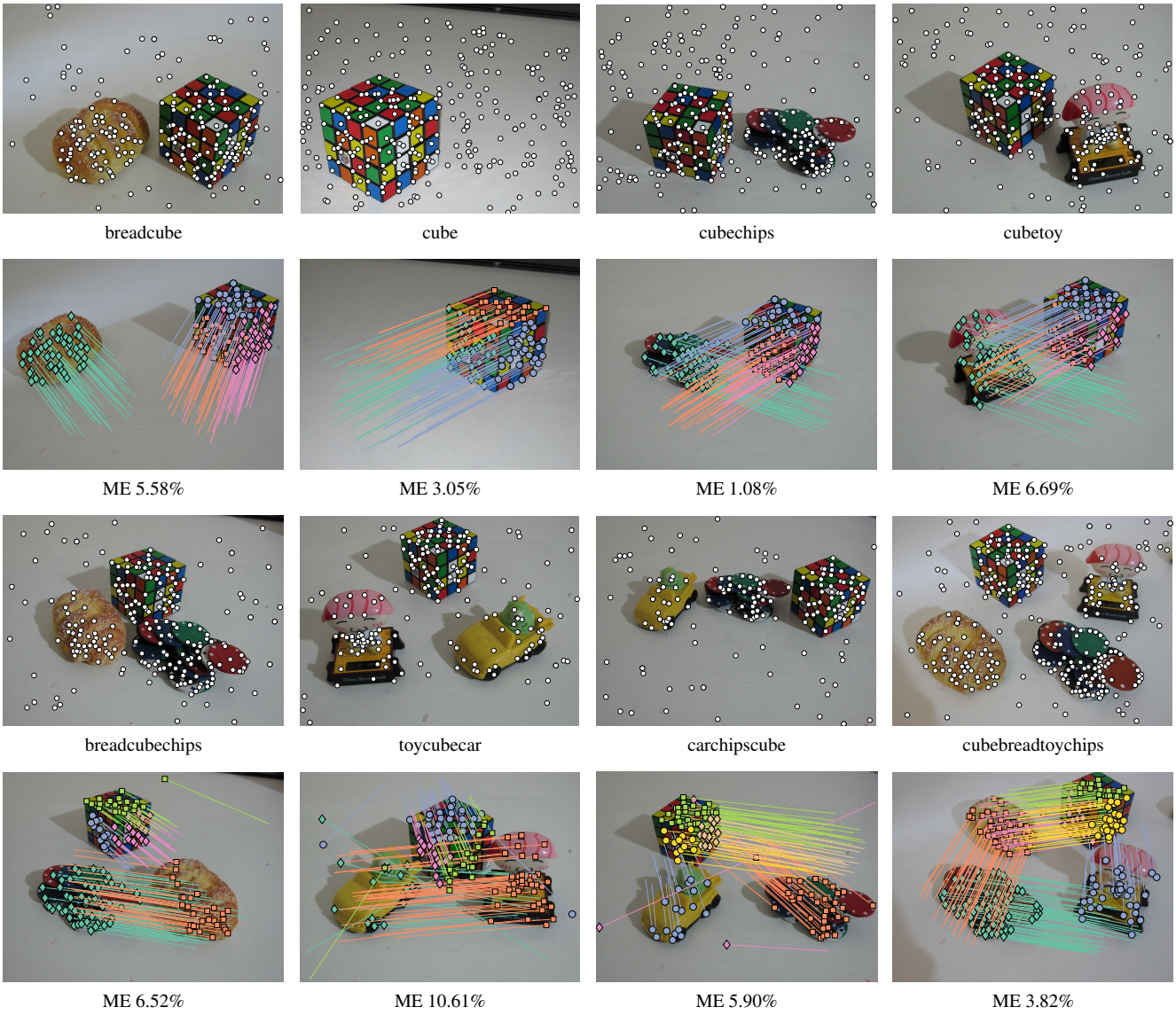


Figure 6: **Homography and fundamental matrix fitting** by MCT on `cube*` image pairs of the AdelaideRMF dataset. We sampled 5000 F matrices and 1000 H matrices per cluster. Odd rows: first input image with outlier-contaminated points overlaid. Even rows: F and H models superimposed on the second image (model assignment is colour coded).

pling towards valid homographies and also gain robustness against outliers.

5. Conclusions

We presented a multi-model recovery framework that offers a hierarchical interpretation of data in nested compatible models. Specifically, we conceived MCT, a cascaded extension of T-linkage that sequentially extract simpler nested models starting form the more general ones. We leverage on geometric constraints to guide the sampling towards simpler valid interpretation of the data.

In this way, all the intra-class model selection problems are implicitly solved by T-linkage, whereas the inter-class model selection issues is cast into an explicit comparison of GRIC scores *among models explaining the same data*.

Our model selection framework is simple and in principle is agnostic about the multi-model fitting technique adopted. In practice it can be generalised to other preference based multi-model fitting algorithm which can benefit from model-constrained sampling.

References

[1] P. Amayo, P. Piniés, L. M. Paz, and P. 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] D. Barath and J. Matas. Multi-class model fitting by energy minimization and mode-seeking. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XVI*, pages 229–245, 2018. 2, 5, 6, 7

[3] P. Chen and D. Suter. Simultaneously estimating the fundamental matrix and homographies. *IEEE Transactions on Robotics*, 25(6):1425–1431, 2009. 6

[4] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 772–779. IEEE, 2005. 6

[5] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 6

[6] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *International Journal of Computer Vision*, 97(2):123–147, 2012. 2, 6

[7] Q.-T. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *Computer vision and image understanding*, 64(2):193–229, 1996. 6

[8] L. Magri and A. Fusiello. T-Linkage: A continuous relaxation of J-Linkage for multi-model fitting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3954–3961, June 2014. 3

[9] L. Magri and A. Fusiello. Multiple models fitting as a set coverage problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3318–3326, June 2016. 5, 6

[10] L. Moisan, P. Moulon, and P. Monasse. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line*, 2:56–73, 2012. 3

[11] V. Patraucean, P. Gurdjos, and R. G. von Gioi. Joint a contrario ellipse and line detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(4):788–802, 2017. 3, 6

[12] T. T. Pham, T.-J. Chin, K. Schindler, and D. Suter. Interacting geometric priors for robust multimodel fitting. *IEEE Transactions on Image Processing*, 23(10):4601–4610, 2014. 2, 6

[13] J. Rabin, J. Delon, Y. Gousseau, and L. Moisan. Mac-ransac: a robust algorithm for the recognition of multiple objects. In *Fifth International Symposium on 3D Data Processing, Visualization and Transmission (3DPTV 2010)*, page 051, 2010. 3

[14] K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2:643–648 vol. 2, 2005. 2

[15] K. Schindler, D. Suter, and H. Wang. A Model-Selection Framework for Multibody Structure-and-Motion of Image Sequences. *International Journal of Computer Vision*, 79(2):159–177, 2008. 2

[16] C. V. Stewart. MINPRAN: A new robust estimator for computer vision. *Pattern Analysis and Machine Intelligence*, 17(10):925–938, 1995. 3

[17] M. Stricker and A. 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

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

[19] Y. Sugaya and K. Kanatani. Improved multistage learning for multibody motion segmentation. In *VISAPP (1)*, pages 199–206, 2010. 2

[20] Z. L. Szpak, W. Chojnacki, and A. van den Hengel. Robust multiple homography estimation: An ill-solved problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2015. 8

[21] T. Tanimoto. An elementary mathematical theory of classification and prediction. Internal technical report, IBM, 1957. 3

[22] M. Tepper and G. Sapiro. A biclustering framework for consensus problems. *SIAM Journal on Imaging Sciences*, 7(4):2488–2525, 2014. 6

[23] M. Tepper and G. Sapiro. Fast L1-NMF for multiple parametric model estimation. *arXiv preprint arXiv:1610.05712*, 2016. 5

[24] P. H. Torr, A. W. Fitzgibbon, and A. Zisserman. The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. *International Journal of Computer Vision*, 32(1):27–44, 1999. 2

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

[26] P. H. S. Torr. *Model Selection for Two View Geometry:A Review*, pages 277–301. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999. 5

[27] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca). *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(12):1945–1959, 2005. 3

[28] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter. Dynamic and hierarchical multi-structure geometric model fitting. In *Proceedings of the International Conference on Computer Vision*, 2011. 6, 7

[29] X. Xu, L. F. Cheong, and Z. Li. Motion segmentation by exploiting complementary geometric models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2859–2867, 2018. 3, 4

[30] http://www.diegm.uniud.it/fusiello/demo/jlk. 5, 7

[31] http://mouse.cs.uwaterloo.ca/code/gco-v3.0.zip. 5

[32] http://dev.ipol.im/~jirafa/ipol_demo/elsdc/. 6

[33] http://visionair.ge.imati.cnr.it/ontologies/shapes/viewmodels.jsp. 6

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