Multiple structure recovery with T-linkage

Multiple Structure Recovery with T-Linkage

Luca Magri[a], Andrea Fusiello[b]

[a]*Dip. di Informatica - University of Verona, Strada Le Grazie, 15 - Verona (Italy)*
[b]*DPIA - University of Udine, Via delle Scienze, 208 - Udine (Italy)*

Abstract

This work addresses the problem of robust fitting of geometric structures to noisy data corrupted by outliers. An extension of J-Linkage (called T-Linkage) is presented and elaborated. T-Linkage improves the preference analysis implemented by J-Linkage in term of performances and robustness, considering both the representation and the segmentation steps. A strategy to reject outliers and to estimate the inlier threshold is proposed, resulting in a versatile tool, suitable for multi-model fitting "in the wild". Experiments demonstrate that our methods perform better than J-linkage on simulated data, and compare favorably with state-of-the-art methods on public domain real datasets.

Keywords: Multi-model fitting, grouping, segmentation

1. Introduction: the challenges of multi-model fitting

Finding multiple models (or structures) that fit data corrupted by noise and outliers is an ubiquitous problem in the empirical sciences, including Computer Vision, where organizing and aggregating unstructured visual content in higher level geometric structures is a necessary and basic step to derive better descriptions and understanding of a scene.

A typical example of this problem can be found in 3D reconstruction, where multi-model fitting is employed either to estimate multiple rigid moving objects and hence to initialize multi-body Structure from Motion [1, 2], or to produce intermediate geometric interpretations of reconstructed 3D point cloud by fitting geometric primitives [3, 4, 5]. Other scenarios in which the estimation of multiple geometric structure plays a primary role include face clustering, body-pose estimation, augmented reality, image stitching and video motion segmentation, to name just a few.

In all these contexts the information of interest can be extracted from the observed data and organized in semantic significant structures by estimating some underlying geometric parametric models, e.g. planar patches, homographic transformations, linear subspaces or fundamental matrices, as shown in Figure 1 where a set of two view correspondences is segmented according to the moving objects they belong to.

Multi-model fitting is a challenging and demanding task, as many issues are involved. If multiple instances of the same structure are present in the data, the problem becomes a typical example of a chicken-&-egg dilemma: in order to estimate models one needs to first segment the data, and in order to segment the data it is necessary to know the models associated with each data point. Moreover, the presence of multiple structures strains robust estimation, which has to cope with both *gross* outliers and *pseudo*-outliers (i.e. "outliers to the structure of interest but inliers to a different structure" [6]).

Figure 1: Example of multi-model fitting application: multiple rigid motions are estimated by fitting fundamental matrices to a set of 2D correspondences across two images.

The main challenge is therefore the simultaneous robust estimation of both segmentation and models without knowing in advance the correct number of models κ. As a matter of fact, the problem of multi-model fitting, is inherently ill-posed, since many different interpretations of the same dataset are possible. Making the problem tractable requires a regularization strategy that constrains the solution using prior information, usually in the form of one or more parameters, such as the number of sought structures or the noise magnitude. Unfortunately estimating these quantities turns out to be a thorny problem, as in general there is no a canonical way to judge the appropriateness of a solution. Following the spirit of Occam's razor – that one shall not presume more things than the required minimum – κ should be kept as low as possible, but finding a correct tradeoff between data fidelity and model complexity (a.k.a. bias-variance tradeoff) is an intricate model selection task.

Contributions. In this work we present T-Linkage, a framework in which the multi-model fitting problem is tackled from the segmentation horn of the chicken-&-egg dilemma, leveraging on preference analysis and hierarchical clustering in a con-

tinuous conceptual space[1]. This scheme has the merit of automatically discovering the number of structures hidden in the data, furthermore treating rogue points as micro-clusters that, in turn, can be pruned in a probabilistic framework where the reliability of a structure is measured in term of its randomness. The resulting algorithm enjoys a straightforward implementation. In addition only a global scale is required, therefore, if consensus clustering is integrated in this approach, it is possible to estimate this parameter given a proper interval search. Thanks to these features, T-Linkage is a suitable tool for multiple structure recovery "*in the wild*" when minimal to no prior knowledge of the data is available.

In this paper we wrap up and review all the material that appeared in [7, 8, 9], including more thorough descriptions, additional insights and new experiments. In particular: the background material and the description of T-Linkage have been enhanced; a new discussion on the scale estimation problem is presented; new experiments have been added, including tests on the complete Adelaide dataset and a 3D plane fitting experiment.

Outline. The article is organized as follows: after a presentation of the relevant literature in Section 2, the main steps of T-linkage algorithm are presented and explored. At first in Section 3 we concentrate on the conceptual representation: we investigate the "*preference trick*" and we propose a continuous relaxation of the binary, winner-take-all approach followed by J-linkage [10]. In this way we provide a more general framework in which we are able to integrate the use of soft functions in order to robustly depict data preferences. Density-based techniques are also employed to analyse the geometry of our conceptual space in Section 4, showing that points belonging to the same model are clustered in a high density region, whereas outliers can be characterized as the most separated points. This observation provides insights into the effectiveness of the clustering techniques described in Section 5. The problem of dealing with outliers is addressed in Section 6, whereas Section 7 concentrates on the selection of the correct inlier threshold avoiding the classical model selection trade-off of two terms (data fidelity versus model complexity).

2. Related work: a consensus and preference perspectives

To set the general context and notation, μ denotes a model and $X = \{x_1, \ldots, x_n\}$ a set of n data, possibly corrupted by noise and outliers. Multi model fitting consists in estimating κ instances of models $\theta_1, \ldots, \theta_\kappa$, also termed structures, together with $C_1, \ldots, C_\kappa$ subsets of the data such that all points described by θ_i are aggregated in C_i. We assume the structures to be parametric – i.e. they can be represented as vectors in a proper parameter space Θ – and the existence of an error function

$$\mathrm{err}\colon X \times \Theta \to \mathbb{R}^+ \tag{1}$$

that associates to every point-model pair $(x, \theta) \in X \times \Theta$ the corresponding residual error $\mathrm{err}(x, \theta)$. For the sake of illustration, in a line fitting problem such as the one reported in Figure 2, $\Theta = \mathbb{P}^{1*}$ is the projective space of all the lines of the plane, and err is simply the geometric point-line distance. In the two view motion segmentation problem encountered in Figure 1, Θ can be identified with the fundamental manifold and err can be either the sampson distance or the reprojection error.

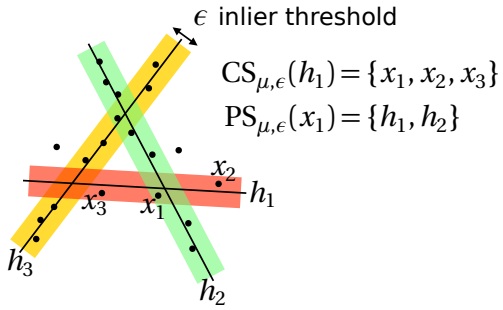


Figure 2: Consensus and preference sets: a simple line-fitting example illustrates the consensus set of the line h_1 and the preferred models of x_1.

In order to deal with the scale of the noise, a threshold $\epsilon \in \mathbb{R}^+$, commonly known as the *inlier threshold*, is introduced. A point x is said to belong to a given structure θ if

$$\mathrm{err}(x, \theta) \leq \epsilon. \tag{2}$$

Hence, the *consensus set* of a structure is simply defined as the set of points that fits the model within a certain inlier threshold ϵ:

$$\mathrm{CS}_\epsilon(\theta) = \{x \in X\colon\ \mathrm{err}(x, \theta) \leq \epsilon\}. \tag{3}$$

Dually, the *preference set* of a point is the set of models having that point as an inlier:

$$\mathrm{PS}_\epsilon(x) = \{\theta \in \Theta\colon\ \mathrm{err}(x, \theta) \leq \epsilon\}. \tag{4}$$

Most of the multi-model fitting algorithms proposed in the literature can be dichotomized according to which part of the chicken-egg-dilemma is addressed first. *Consensus* methods put the emphasis on the estimation and aim for the structures that describe as many points as possible. On the other hand, *preference* approaches concentrate on the segmentation side of the problem: point preferences are clustered in order to obtain a partition of the data that is used to estimate the structures of interest.

Hereinafter we attempt to trace the path that has been followed in the literature to address these challenging issues, starting from consensus and continuing to preference analysis.

2.1. Consensus analysis

The peculiar nature of visual data – typically affected by arbitrarily large gross measurement errors – breaks down classical least-square estimators that are fragile and sensitive to outliers. Consensus analysis stands out as one of the first attempts that is not unduly affected by rogue data. The methods belonging to this category follow a common two step paradigm. At first

[1]This notion will be defined at pg. 12, for now it suffices to say that the conceptual representation of an object is the vector of its posterior probabilities given certain classes

the parametric space Θ of all the feasible structures is approximated as a suitable finite hypothesis space H. Then a voting procedure elects the structures in H that best explain the data in terms of a consensus set.

This idea of exploiting consensus is at the core of the celebrated Ransac (Random Sample Consensus) [11], aimed at estimating the parameters of a single model in the presence of large numbers of outliers. The goal of minimizing squared residuals, typical of the Least Square method, is replaced in Ransac with the objective of maximizing the size of the consensus set of a structure, provided the inlier threshold as input. Ransac approximately maximizes this criterion by searching through a pool of putative structures H determined by random sampling. In particular at each iteration a *Minimum Sample Set* (MSS) – composed of the minimum number of points necessary to instantiate the free parameters of a structure – is drawn. In this way the estimation problem in the continuous domain Θ is converted into a selection problem in the finite discrete subset $H \subset \Theta$.

For each estimated model the corresponding consensus set is computed by counting the residuals below the inlier threshold. This procedure is repeated until a structure having enough supporting inliers is discovered. A number of efforts have been made to improve the RANSAC paradigm. For example, MSAC (M-estimator Sample Consensus) and MLESAC (Maximum Likelihood Estimation Sample Consensus) [12] propose to increasing the robustness of the Ransac paradigm incorporating the use of M-estimator techniques. A lot of other refinements [13] in terms of both accuracy and efficiency have been made; for example different sampling strategies have been proposed in the literature to reduce the number of iterations necessary to recovery an inlier structure. A nice survey on all these advancements can be found in [14] or in the more comprehensive overview of recent researches presented in [15] where Usac (Universal Framework for Random Sample Consensus) is also derived.

The Ransac strategy has been adapted to estimate multiple structures. Its most straightforward generalization is embodied by Sequential Ransac, an iterative algorithm that executes Ransac many times and removes the found inliers from the data as each structure is detected. Zuliani et al. [16] noticed some drawbacks of this greedy *estimate-and-remove* approach, which in fact may happen to be sub-optimal since the quality of the attained solution can be affected by inaccurate estimation of the initial structures.

In order to correct this behavior Zuliani et al. introduced Multi-Ransac. Remaining tied to the idea of maximizing the consensus set, Multi-Ransac replaces the sequential scheme with a parallel approach. Rather than looking for a single structure having the largest consensus, κ models having maximal support are searched for simultaneously at each iteration. This is done by updating iteratively a collection of κ models with κ new sampled structures using a fusion procedure that explicitly enforces the disjointness of the obtained consensus sets. However as demonstrated experimentally in [10], this method may yield poor results in the presence of intersecting structures.

The popular Hough transform and its randomized version (Randomize Hough Transform [17]) can be considered as well as consensus-oriented algorithms. In these approaches the parameter space Θ is approximated as a quotient space $H = \Theta/_{\sim}$ in which models are represented as equivalence classes of similar structures. The space H is hence employed to build an accumulator collecting data votes: every point adds a vote to the bins representing the structures it belongs to. After voting is complete, the accumulator is analysed to locate the maxima that individuate the desired structures. Differently from Ransac, where H is a discrete sampled version of Θ, in the Hough transform the elements of the hypothesis space provide an exhaustive representation of the parameter space, and tentative models are all considered simultaneously. This, however, comes at the cost of defining a proper quantization of the space, which rapidly becomes intractable as the degrees of freedom of the models increase. A randomized Hough Transform instead of considering the votes of all the points, exploits random sampling to approximate the accumulator for votes, reducing the computational load.

This strategy can be considered as an instance of a more general approach, that consists in finding modes directly in Θ [18]. In this way, the difficulties of the quantization step are alleviated by mapping the data into the parameter space through random sampling and then by seeking the modes of the distribution with, e.g., mean-shift [19].

In all the consensus based methods, alongside the voting phase, the approximation of Θ is a recurring theme and a very critical step. The key point is that, when multiple structures are hidden in the data, consensus oriented algorithms have to disambiguate between genuine structures and redundant ones, i.e. multiple instances of the same model with slightly different parameters.

This crucial difficulty is hence addressed by enforcing several disjointedness criteria implicitly implemented in the different approximations of the solution space. For instance, the Hough transform attempts to handle redundancy by capturing similar structures in the same equivalence class via the problematic quantization of Θ. Along the same lines, the bandwidth used in mean shift can be thought of as a way to localize and aggregate redundant models. As suggested in [20] also both Sequential Ransac and Multi-Ransac enforce disjointedness by avoiding the sampling of similar models. As regards Sequential-Ransac, this idea can be individuated in the iterative removal of the discovered inliers and in the subsequent sampling of the hypotheses on the remaining data. In Multi-Ransac this is more evident, since this algorithm explicitly includes in its parallel approach a disjointedness constraint by directly searching for the best collection of κ disjoint models. In practice, however, using consensus as the only criterion seems short-sighted, as, in many cases, ground truth models can have mutual intersection greater than redundant ones, and, consequently, the plain consensus fails in discerning authentic structures.

2.2. Preference analysis

In order to overcome the difficulties inherent in consensus methods, it has been proposed to tackle the problem from a different point of view. Instead of exploiting the consensus of structures, the role of data and models are reverted: rather than

representing models and inspecting which points match them, the preference sets of individual data points are examined.

This idea can be traced back to Residual Histogram Analysis [21] where the residuals distributions of points, with respect to a set of putative structures randomly sampled, is taken into consideration. In particular, an histogram analysis of the residuals is used to reveal the most significant structures as peaks in the histograms. In addition, the number of models is automatically determined by the median number of modes found over all data points. Even if, in practice, the mode-finding step of this strategy suffers from low accuracy and depends critically on the bin size adopted, this method has the merit of reformulating the model-estimation task in a *conceptual space*, where points are described by their residuals.

The change of perspective entailed by preference analysis results in a different approach to the chicken-&-egg dilemma. Structures are recognized as groups of neighbouring points in the conceptual space so the emphasis is shifted from the estimation to the segmentation part of the problem.

J-Linkage [10] embodies the spirit of preference analysis exploiting a preference based representation of data. This algorithm has been demonstrated to be very effective in practice, and has been extensively exploited in the literature in many multi-model fitting problems[2].

At a high level, a two steps *first-represent-then-clusterize* scheme is implemented. At first, data are represented as preference sets with respect to a pool of tentative models instantiated via random sampling, then a greedy agglomerative clustering is performed to obtain a partition of the data. The key idea is that points belonging to the same structure will have similar preference sets, and the clustering algorithm proceeds in a bottom-up manner to merge clusters until all their points share a common preference.

More in detail, a cluster $U \subseteq X$, is portrayed as the preference set of all the common preferences among all the data belonging to it:

$$\mathrm{PS}_\epsilon(U) = \bigcap_{x \in U} \mathrm{PS}_\epsilon(x). \tag{5}$$

and the distance between clusters is computed as the Jaccard distance [22] between the respective preference representations. The Jaccard distance between two sets A, B is defined as

$$J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

and measures the degree of agreement between the votes of two clusters; it ranges from 0 (identical votes) to 1 (disjoint preference sets). Starting from singletons, each sweep of the algorithm merges the two clusters with the smallest Jaccard distance. The cut off value is 1. It is worth noting that, if outliers are not present in the data, the number of clusters is automatically detected by this algorithm and the only required input is the inlier threshold, used in the computation of the preference sets.

[2]For a list of applications of J-linkage see `http://www.diegm.uniud.it/fusiello/demo/jlk/j-parade.html`

Several trends in common with previous methods can be recognized: an inlier threshold ϵ needs to be provided in advance as in Ransac, and the idea of points casting votes on the models echoes the Randomize Hough Transform. Nevertheless J-Linkage does not work in a quantize space, which is at the root of the shortcoming of Hough Transform, nor in the residual space, which leads to the difficulties of modes estimation, but explicitly introduces conceptual space where points are portrayed by the *preferences* they have accorded to random provisional models.

Along the same line of J-Linkage, Kernel Fitting [23] exploits preferences to derive a kernel matrix that encapsulates the order in which models are preferred, (i.e., the order of their residuals). The rationale is that points belonging to the same ground-truth structure should have similar orders of preferred models. Exploiting this information, a transformation is applied to the data points into a space which permits the detection of outliers. The removal of outliers yields a reduced kernel matrix that, in turn, is used to over-segment the remaining inliers. Finally a merging scheme is used to reassemble these models into the final structures.

RCMSA (Random Cluster Model Simulated Annealing) [24] also takes advantage of the same idea representing data points as permutations on a set of tentative models constructed iteratively, using subsets larger than minimal. Point preferences are organized in a weighted graph and the multi-model fitting task is stated as a graph cut problem which is solved efficiently in an annealing framework. Alternatively in [25] the idea of representing points as permutations of models is also exploited by QP-MF a quadratic programming aimed to maximize the mutual preferences of inliers. Residual information is also exploited in [26], a single-model estimation technique based on random sampling, where the inlier threshold is not required.

Another stream of investigation focused on higher order clustering [27, 28, 29, 30, 31] implicitly adopts a preference based approach. In these works higher order similarity tensors are defined between n-tuple of points as the probability that these points are clustered together exploiting the residual error of the n points with respect to provisional models. In this way preferences give rise to a hypergraph whose hyperedges encode the existence of a structure able to explain the incident vertices that represent data. The problem of multi-model fitting is hence reduced to find highly connected components in this preference hypergraph. In practice, the similarity tensor is properly reduced to pairwise similarity and fed to spectral segmentation algorithms. Hypergraph Fitting [31], on the contrary, works directly with the hypergraph. At first, the hypergraph is built by expanding as much as possible its hyperedges exploiting robust statistics, then an hypergaph partition algorithm is used on the most significant hyperedges in order to detect sub-hypergraph. Hence the best representative of hyperedges in each sub-hypergraph is selected and duplicate structures are removed leveraging on mutual information theory. Also[32] works with higher order graphs, but multi-model fitting is translated into a mode-seeking problem on the data/model hypergraph.

In summary, different perspectives on the multi-model fit-

ting problem have been adopted. Consensus oriented methods look at the problem considering some kind of accumulation space – either consisting of individual models, as in Ransac, or in equivalence classes of structures as in the Hough transform – in which votes of points are collected. Structures are hence estimated from maximal consensus. This paradigm has been demonstrated to be successful in single model estimation, but it is less effective if multiple structures are present in the data, because consensus does not allow us to distinguish clearly between genuine models and redundant ones. When multiple structure recovery is viewed through the lens of preference analysis the attention is shifted to the segmentation part of the problem. Data are represented as points in a high dimensional space or as vertices of an hypergraph and clustered together using ad hoc techniques.

It goes without saying that the state-of-the art on multi-model fitting can also be described along other dimensions. For example multiple structures recovery can be seen from an optimization perspective as the minimization of a global energy functional composed of two terms: a modeling error which can be interpreted as a likelihood term, and a penalty term encoding model complexity mimicking classical MAP-MRF objectives. A survey of multi-model fitting methods from this point of view can be found in [33].

A model selection approach is taken in [34, 35, 36, 24, 33], where the cost function to be minimized is composed of a data term that measures goodness of fit and a penalty term which weighs model complexity (see e.g. [37]). Sophisticated and effective minimization techniques such as SA-RCM [24], AR-JMC [36] and PEaRL [33] have been proposed. The latter, for example, optimizes a global energy function that balances geometric errors and regularity of inlier clusters, also exploiting spatial coherence. However the relative magnitude of the penalty term with reference with the data term is not obvious to define.

In order to complete the picture, it is worth mentioning that optimization routines have also been tailored to specific instances of multi-model fitting, e.g., subspace segmentation [38, 39, 40]. In this field the use of low-rank and sparse analysis has produce a solid literature, accurately illustrated for example in [41]. Local subspace affinity (LSA) [38] is an algebraic method that uses local information around points in order to fit local subspaces and to cluster points using spectral clustering with pairwise similarities computed using angles between the local subspaces. Agglomerative Lossy Compression (ALC) [39] is a bottom up clustering algorithm that aims at segmenting the data minimizing a coding length needed to fit the points with a mixture of degenerate Gaussians up to a given distortion. SSC [40] exploits sparse representation for segmenting data. Since high dimensional data can be expressed as linear combination of few other data points, SSC use the sparse vectors of coefficients of these linear combinations as a convenient conceptual representation of points. Then spectral clustering is performed for segmenting data in this conceptual space.

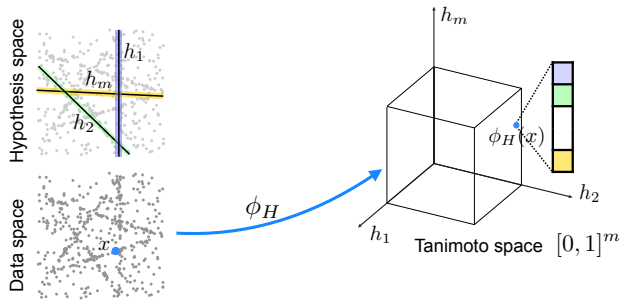


Figure 3: The preference trick in a nutshell: data X are represented via ϕ_H (defined in Eq. 12) in the Tanimoto space as preference function collecting the votes they grant to a pool of tentative hypotheses H.

3. The preference trick: a lift to Tanimoto space

The binary preference analysis implemented by J-Linkage suffers from the same drawbacks as Ransac with respect to Msac, therefore we propose enhancing it by relaxing the notion of a preference set. To this end we borrow from robust statistics the weighting functions adopted by M-estimators and use them as a voting function to express robust point preferences. As a result, we alleviate the influence of outliers and mitigate the truncating effect of the inlier threshold. We hence conceive a continuous conceptual space in which the Jaccard distance is generalized by the Tanimoto distance in order to handle the continuous representations of points.

In pattern recognition a theoretical framework for conceptual representation was settled by Pekalska and Duin in [42]:

Definition 3.1 (Conceptual representation)**.** *Given two arbitrary sets A and B, let ϕ be a non negative function, expected to capture the notion of closeness between a pair of points in $A \times B$, e.g. a similarity or a dissimilarity measure. A conceptual representation of a point $a \in A$ is a set of similarities/dissimilarities between a and the elements of B expressed as a vector*

$$a \mapsto [\phi(a, b_1), \phi(a, b_2), \ldots, \phi(a, b_m)] \in \mathbb{R}^m \tag{7}$$

B is called a representation set.

The function ϕ might be non-metric. This definition is very flexible. In the case $A = B$ the conceptual representation is a standard similarity or dissimilarity measure between a pair of objects. Allowing B to be an arbitrary set of prototypes [43], several generalizations, recently applied for classification purposes, can be derived. For example [44] exploits hidden markov model to construct a conceptual space for clustering sequential data, whereas [45] relies on one class support vector machine to represent and aggregate semantically similar images.

The representation step adopted by J-Linkage can be mapped in this framework setting $A = X$, $B = H \subseteq \Theta$, the pool of sampled structure is regarded as the representation set, and choosing as ϕ the similarity measure defined as

$$\phi(x_i, h_j) = \begin{cases} 1 & \text{if } \mathrm{err}(x_i, h_j) \leq \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In practice ϕ assesses the fitness to x_i with respect to the structure h_j. As noted in [42], this construction can be interpreted in a statistical sense as the posterior probabilities of the point x with respect to the m classes determined by the consensus set of the putative structures:

$$[\text{Prob}\,(x|\,\text{CS}_\epsilon(h_1))\,,\ldots,\text{Prob}\,(x|\,\text{CS}_\epsilon(h_m))] \in \mathbb{R}^m \tag{9}$$

Seen in this way, this conceptual representation is linked to the stream of research on higher-order clustering where probability is used to define higher-order affinity between points.

With respect to J-linkage, we introduce a continuous relaxation of the binary preference set: the *preference function* of a point. Each data point x is described by a function ϕ taking values in the whole closed interval $[0,1]$ and not only in $\{0,1\}$.

In this way the conceptual space is generalized from the set of characteristic functions to $[0,1]^H = \{\phi(x,\cdot) : H \to [0,1]\}$, and we are allowed to express the preferences of a point more accurately by integrating more specific information on residuals. In practice, the idea is to mitigate the truncating effect of threshold in (8) defining the preference of a point x via the similarity:

$$\phi(x,h_j) = \begin{cases} \exp\left(-\,\text{err}(x,h_j)/\tau\right) & \text{if err}(x,h_j) < 5\tau \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The time constant τ plays the same role as the inliers threshold ϵ ; however, compared to the discrete case, the threshold defined by (10) is less critical since it replaces the abrupt truncation in (8) with an exponential decay. Please note that setting the residuals smaller than 5τ to zero is quite natural since for $u > 5\tau$ the function $e^{-u/\tau}$ can be considered almost constant, with variations that do not exceed 0.7 %.

In principle this step can also be performed by exploiting other types of functions, typical of the M-Estimator framework, e.g. one can define the similarity $\phi\colon X \times H \to [0,1]$ as

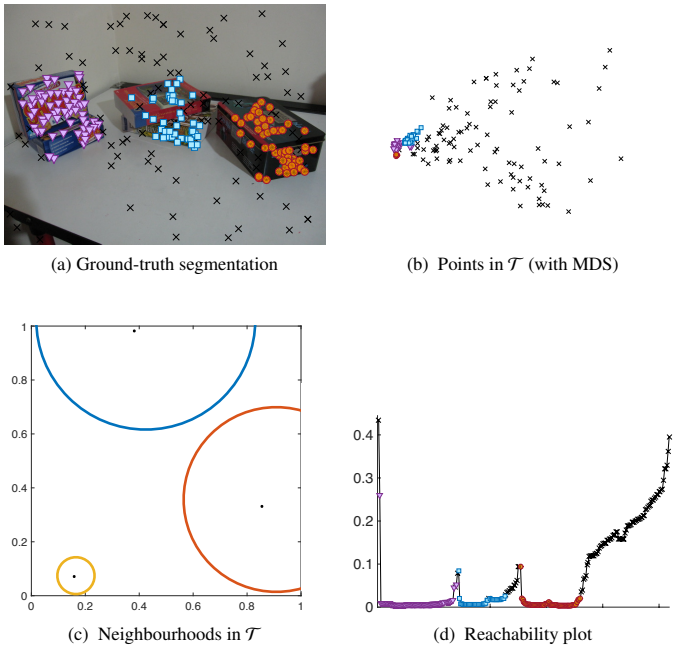
$$\phi(x_i,h_j) = w_c\left(\frac{\text{err}(x_i,h_j)}{\tau\sigma_n}\right), \tag{11}$$

where w_c can indicate any of the weighting functions whose images are contained in the interval $[0,1]$, namely the Huber, Cauchy, Geman, Welsh and Tukey weighting functions. In practice w_c plays the same role as the inlier threshold and can be tuned either using this parameter or, under the assumption of gaussian noise, as $c = k\sigma_n$ where σ_n is an estimate of the standard deviation of the residuals and k is chosen to ensure a predefined level of asymptotic efficiency on the standard normal distribution for the specific M-estimator selected.

It is straightforward to embed data points from their ambient space to the conceptual one using the vectorial mapping $\phi_H : X \to [0,1]^m$, simply defined as

$$x \mapsto [\phi(x,h_1),\ldots,\phi(x,h_m)]\,. \tag{12}$$

As depicted in Figure 3, every point x is robustly represented as a m-dimensional *preference vector* in the conceptual space whose entries are the weights, giving rise to a soft version of



(a) Ground-truth segmentation (b) Points in $\mathcal{T}$ (with MDS)

(c) Neighbourhoods in $\mathcal{T}$ (d) Reachability plot

Figure 4: Insights on the geometry of Tanimoto space. (a) one frame of the *biscuitbookbox* sequence. Model membership is color coded; black crosses (x) are outliers. (b) conceptual representation of the data in Tanimoto space are projected in the plane using Multi-Dimensional Scaling for visualization purposes. Outliers (x) are recognized as the most separated points. (c) Tanimoto neighbourhoods with the same radius in $[0,1]^2$ have a smaller Euclidean diameter if the center lies near the origin. (d) The reachability plot shows the reachability distance of ordered points (model membership is color coded according to the ground truth).

the preference set. From the preference analysis perspective, the rationale beyond this construction is that the i-th component of this vector expresses, with a soft vote in $[0,1]$, the preference granted by x to the tentative structures h_i. Please note how this parallels the difference between Ransac and Msac, if consensus sets are considered.

The next step is to introduce in the unitary cube $[0,1]^m$ a suitable metric that generalizes the Jaccard distance. This is accomplished by the Tanimoto distance [46], defined as

$$d_{\mathcal{T}}(p,q) = 1 - \frac{\langle p,q\rangle}{\|p\|^2 + \|q\|^2 - \langle p,q\rangle} \tag{13}$$

for every $p,q \in [0,1]^m$. This distance ranges in $[0,1]$ and equals 0 for preference vectors sharing the same preferences whereas reaches 1 if points have orthogonal preferences, i.e. there does not exist any model in H that can explain both the points p and q. We denote as $\mathcal{T} = ([0,1]^m, d_{\mathcal{T}})$ the metric space endowed with the Tanimoto distance [47]. Please observe that if we confine ourselves to the space $\{0,1\}^m$ the Tanimoto distance coincides with the Jaccard one. The agreement between the preferences of two points in the conceptual space reveals the multiple structures hidden in the data: points sharing the same preferences are likely to belong to the same structures as points matching the same collection of models are likely to belong to the same ground truth model.

In short, echoing the celebrated "kernel trick", which lifts a non linear problem into a higher dimension space in which it

becomes easier, this conceptual representation, shifts the data points from their ambient space to the Tanimoto one, revealing the multiple structures hidden in the data as groups of neighboring points.

Clustering can be thought of as the discrete and statistical counterpart of the continuous and geometric problem of finding connected components. With this idea as a guide, a geometric analysis of the Tanimoto space can confirm the intuition that points sharing the same preference are grouped together in the conceptual space. To illustrate qualitatively these properties we consider the multi model fitting problem reported in Figure 4a, taken from [48]. In this dataset three objects move independently, each giving rise to a set of point correspondences in two uncalibrated images: points belonging to the same object are described by a specific fundamental matrix. Outlying correspondences are also present.

Some insight into the geometrical sparseness of outliers can be reached considering a system of neighbourhoods: if we fix some $\eta \in (0, 1)$ and some $p \in \mathcal{T}$ the Tanimoto ball of radius η and center p is denoted by $N_\eta(p)$. As illustrated in Figure 4c, the Euclidean diameter of N_η changes according to the position of the center p. In particular this quantity tends to be smaller for points lying near the origin of $\mathcal{T}$, that corresponds to the region of $\mathcal{T}$ mainly occupied by rogue points. In fact outliers grant their preferences to very few sampled hypotheses, have a small Euclidean norm and consequently tend to lie near the origin. Hence the probability that two outliers live in the same ball of radius η is significantly lower than the probability that two inliers (with higher Euclidean norm) are contained in a ball with the same radius. For this reason outliers can be recognized as the most separated points in $\mathcal{T}$.

4. Density analysis

With this perspective as a guide, we can examine our conceptual representation through the lens of density based analysis in order to make these aspects of Tanimoto space more explicit.

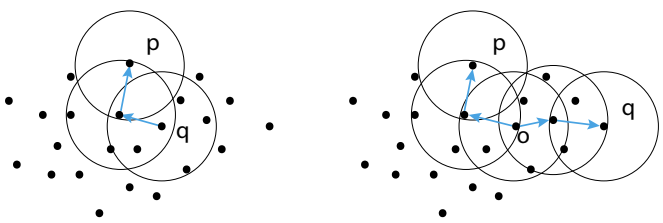


Figure 5: Illustration of reachability. Reachability is not a symmetric relation: in the example on the left p is density reachable from q, but q is not density reachable from p. On the right p and q are density-connected to each other with respect to o.

In particular we adopt the multi-scale approach offered by Optics (Ordering Points to Identify the Clustering Structure) [49]. Optics is a density-based technique which frame the geometry of the data in a reachability plot thanks to the notion of reachability distance. To start with, we tailor the definition of density-connected component proposed in [50] to Tanimoto space:

Definition 4.1. *Given $p, q \in \mathcal{T}$, the cardinality ζ of MSS and $\eta \in (0, 1)$*

- *p is said to be a* core point *if $|N_\eta(p)| > \zeta$;*
- *p is* directly density-reachable *from q with respect to η if $p \in N_\eta(q)$ and q is a core point;*
- *p is* density reachable *from q with respect to η if there is a chain of points $p_1, \ldots, p_\ell$ s.t. $p_1 = p, p_\ell = q$ and p_{i+1} is directly density reachable from p_i;*
- *p is* density-connected *to point q with respect to η if there is a point o such that both p and q are density reachable from o.*
- *a* density-connected component *is a maximal set of density-connected points.*

An illustration of these concepts is depicted in Figure 5. Density-connectivity is an equivalence relation hence all the points reachable from core points can be factorized into maximal density-connected components yielding the desired segmentation. A crucial advantage of this definition is that it deals directly with outliers which can be recognized as points not connected to any core point. In topological words, outliers can be identified as isolated points, whereas inliers are either internal or boundary points of a density-connected component. A key merit of this notion is that density-connected components may have an arbitrary shape. Note that, by definition, a density-connected component must contain at least $\zeta + 1$ points; this is consistent with the fact that at least $\zeta + 1$ points are needed to instantiate a non-trivial model (ζ points always define a model by definition of MSS).

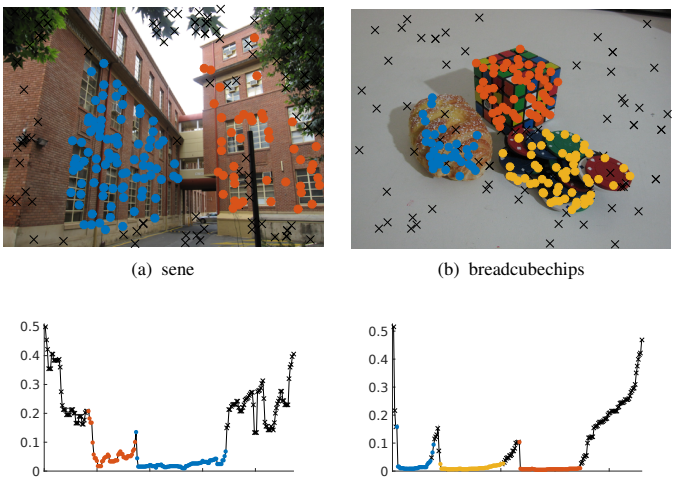
Definition 4.2. *Given the cardinality ζ of MSS,*

- *if p is a core point, the* core-distance *of p refers to the distance between p and its w-nearest neighbor.*
- *if p is a core point, the* reachability-distance *of a point p with respect to a point q is the maximum between the core distance of p and the distance $d_{\mathcal{T}}(p, q)$.*

After the data have been ordered so that consecutive points have minimum reachability distance, Optics produces a special kind of dendrogram, called a reachability plot, which consists of the reachability values on the y-axis of all the ordered points on the x-axis. The valleys of this plot represent the density-connected regions: the deeper the valley, the denser the cluster.

Figure 4d, where the *biscuitbookbox* reachability plot is shown, illustrates this. Outliers have high reachability values, while genuine clusters appear as low reachability valleys and hence are density-connected components in $\mathcal{T}$. Other examples of reachability plots are reported in Figure 6.

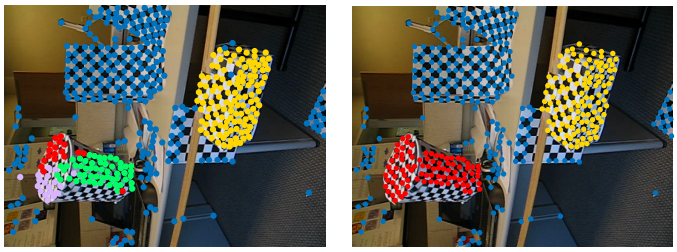
A final remark concerns the sampling strategy. Since the data we are dealing with have a geometric nature, we gain some



(a) sene (b) breadcubechips

(c) Reachability plot (d) Reachability plot

Figure 6: Examples of reachability plots (bottom row). Ground truth segmentation (top row),



(a) Uniform sampling (b) Localized sampling

Figure 7: At equal number of hypotheses, localized sampling helps in producing more accurate results. The image is a sample frame of the video segmentation experiment reported in Section 5.1.

benefits (see Figure 7) by introducing a locality bias. In particular, we adopt a mixed sampling strategy, combining uniform sampling with local sampling for selecting neighbouring points with higher probability. In this way we are able to *exploit* local information and at the same time to *explore* hypothesis space.

T-Linkage is modular with respect to sampling, and other strategies can be adopted as well, relying on different additional assumptions. Clearly it is desirable to sample as many *pure* structures as possible – i.e. structures whose MSS is entirely composed of inliers of the same model – as *spurious* hypotheses assume, along the dimension of models, the same role played by outliers along the dimension of data. As a matter of fact, efficient hypothesis generation is an active field of research and many techniques have been proposed to guide sampling towards promising models, reducing the computational burden of this step, for example [51, 52] just to name a few. An example of what can be gained in terms of accuracy by exploiting filtering techniques is represented by [53] where T-Linkage is combined with Fast Hypothesis Filtering, a method that rejects spurious and irrelevant hypotheses.

5. Preference analysis via T-linkage clustering

In this section we investigate how the robust preference trick, can be combined with clustering analysis in order to solve the multi-model fitting problem. In particular we tailor the agglomerative linkage clustering to handle continuous representations in the Tanimoto space, so that structures can be recovered as clusters of preferences in the conceptual space. In this setting outliers can be recognized as micro-clusters occurring by chance and are filtered out relying on a probabilistic framework. Some experiments are carried out in order to validate the benefit of this approach.

The abundant literature on clustering[3] offers many tools to organize data in many sensible taxonomies according to several criteria. Here we rely on hierarchical clustering which, rather than defining a static partitioning of the data as partitonal methods, aggregates points into a sequence of nested partitions, and exploits the attained hierarchy of subsets to infer the hidden structure of the data. This process can be performed in two directions, namely bottom-up or top-down. In the first case, starting from singletons, a cluster including all the data points is produced by successive merging. Vice versa in the latter case the data are sequentially split into several groups. For a data set with n elements, the top-down scheme would start by considering $2^{n-1} - 1$ possible splits of the data , which is computationally expensive. Therefore, in practice, bottom-up approaches are usually preferred.

The hierarchy of nested groups is encapsulated in a dendrogram, which depicts the formation of a cluster together with the (dis)-similarity levels that have been created by merge or split moves. The final segmentation of the data is obtained by cutting the dendrogram at the desired similarity level.

Several ways to compute the distance or the similarity measure between clusters – called linkage functions – have been proposed in the literature; the most common and popular being:

- Single linkage: where the distance between a pair of clusters is determined by the two closest elements to the different clusters. This procedure tends to generate elongated clusters, which causes the so called chaining effect.
- Complete linkage: In contrast to single linkage, the farthest distance of a pair of objects is used to define inter-cluster distance.
- Average linkage: The distance between two clusters is defined as the average of the distances between all pairs of data points, each of which comes from a different group.

Our method can be thought of as a variation of average linkage in the Tanimoto space. Each cluster is represented by a suitable prototype, but instead of averaging the preferences of the points belonging to a cluster, we take the minimum of all their votes. More precisely, we extend the preference trick to a

[3] For a short survey on the subject the interest reader is referred to [54]

subset of the data $S \subseteq X$ as

$$\phi_H(S) = \min_{x \in S} \phi_H(x). \tag{14}$$

In this way a subset S of X is represented as a vector in $[0,1]^m$ whose j-th component expresses the minimum votes granted to h_j among all the points in S. Component-wise we have:

$$[\phi_H(S)]_j = \min_{x \in S} \phi(x, h_j). \tag{15}$$

We prefer to use Equation (15), instead of averaging preferences for two reasons. First if we confine ourselves to the binary space $\{0,1\}^m$ we obtain exactly the same linkage scheme proposed in J-Linkage. Second the minimum is associative ($\min(\min(a,b),c) = \min(a, \min(b,c))$). Therefore the representation of a cluster is independent of the order in which it has been formed, and once a cluster is formed we do not need to keep track of its point preferences.

Starting from all singletons, each sweep of the algorithm merges the two clusters with the higher similarity. Tanimoto distances are hence updated and clusters are aggregated until all the distances equals 1. This means that the algorithm will only link together elements whose preference representations are not orthogonal, i.e. only as long as there exists in H a structure that received a positive vote from two clusters, will they be merged. This fact motivates the use of a voting function with a finite rejection point, if soft descenders with an infinite cutoff were adopted, small preferences, accorded to outlying structures, will cause the union of all the points in a unique cluster. On the contrary having set the votes of outliers to zero allows the use of the natural predetermined clustering-cutoff . Moreover, as a byproduct,

- for each cluster there exists at least one model for which all the points have expressed a positive preference (i.e., a model that fits all the points of the cluster)
- it is not possible that two distinct clusters grant a positive preference to the same model (otherwise they would have been linked).

Each cluster of points defines (at least) one model. If more models fit all the points of a cluster they must be very similar. As a consequence, in principle, it is sufficient to sample every genuine structure once.

From the explanatory data analysis perspective, the dendrogram produced by T-Linkage can be used to examine the attained cluster. An example is shown in Figure 8, where it can be appreciated that outliers tend to emerge as small clusters at the end of the agglomerative process, as can be inferred by the height of their subtrees. On the other hand inliers determine wider subtrees with smaller height. This means that T-Linkage, although being greedy, promotes the aggregation of points belonging to genuine structures at the very beginning of the merging process increasing the overall robustness of the method. This is also confirmed by the analysis of the linkages illustrated in Figure 9: the majority of linkages (59%) involve the fusion between clusters collecting inliers of the same structures, only few outliers (3%) are aggregated to genuine clusters, and the remaining of the linkages happens at the end of the process aggregating small clusters entirely made up of rogue points. Clearly the separation of the outliers in the Tanimoto space discussed in Section 4 is the reason for such behavior.

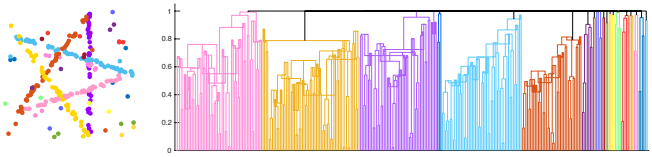
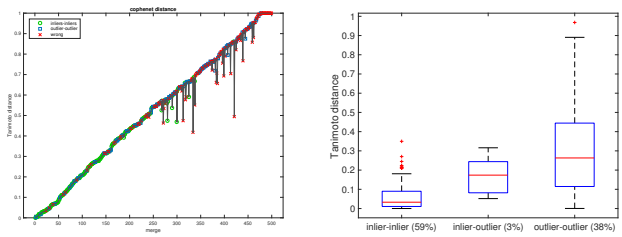


Figure 8: Dendrogram produced by T-Linkage on the *Star5* datasets. The five subtrees corresponding to inliers can be easily recognized. Best viewed in color.



(a) distance of merging moves (b) Box plots

Figure 9: Analysis of the agglomerative process for the *Star5* dataset. Left: cophenet distances registered through the agglomerative process, labeled as "inlier-inlier" when clusters associated to the same structures are merged, "outlier-outlier" when clusters of outlying points are merged and "wrong" otherwise. Right: Box plots of the cophenet distance.

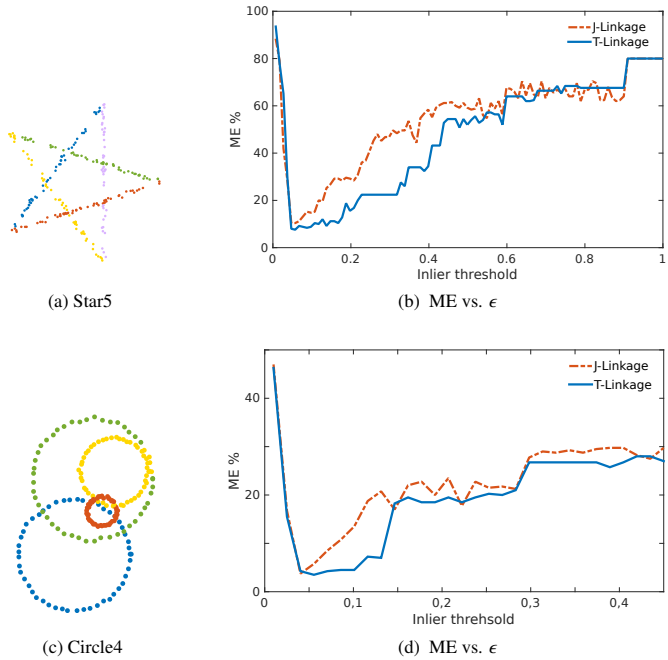
We can observe that the separation of the outliers, in the Tanimoto space, results in the fact that typically rogue points are merged later with respect to inliers. This can be appreciated from the fact that, on the contrary, inliers are linked together at the very first stage

T-Linkage, as any agglomerative clustering algorithms, fits all the data: bad models must be filtered out *a posteriori* (this aspect will be discussed in Section 6). Finally, the model for each cluster of points is estimated by least squares fitting.

As noted in Section 1, the problem of multiple fitting can be considered from two alternative points of view usually coexisting: we want to faithfully segment the data and at the same time to obtain an accurate estimate of the underlying models. Each of these two tasks can not be undertaken without the other. T-Linkage is a pure preference based method and concentrates on the first task segmenting the data in the conceptual space and extracting a model only at the end via least-squares fitting. However once models have been obtained, optionally it is possible to perform an additional *refinement* step: points are reassigned to their nearest model – if it has a distance smaller than ϵ – and finally structures are re-estimated according to this new segmentation. In this way not only can the segmentation and the model estimation step take advantages from each other, but we also gain the benefit of mitigating the greedy behavior of

T-Linkage since the final clustering depends less critically on the order in which points were merged together. Under the assumption of gaussian noise, this step can also be viewed as a maximum likelihood estimation, since minimizing the distance of points from the fitted model is equivalent to maximizing their likelihood.

Remark:. Both Optics and T-linkage can be profitably used to produce a dendrogram which captures the nature of the data. While the diagram created by Optics can be computed more efficiently with respect to the tree obtained by T-linkage, extracting automatically a clustering from the reachability plot turns out to be less convenient. In [8], an heuristic strategy based on Watershed segmentation has been proposed: the main idea is to locate the valleys of the reachability plot by detecting its local minima, requiring the user to define a sensitivity parameter. Unfortunately, this approach does not have the two guarantees (i.e. all points are explained by at least one structure, and there is no a structure explaining the points of two clusters) enforced by T-Linkage thanks to its preference updating mechanism. In addition, in some cases, the reachability plot shows great variation in steepness, and the sensitivity parameter to reject spurious local minima, is more difficult to set with respect to the inlier threshold that, by contrast, has a clear geometric meaning. For these reasons, in this work we have decided to concentrate our focus on T-linkage and its parameter, while we are leaving how the strengths of T-Linkage can be combined into the more computationally convenient framework of Optics, for future study.



(a) Star5 (b) ME vs. ϵ

(c) Circle4 (d) ME vs. ϵ

Figure 10: T-Linkage attenuates the sensitivity of ϵ. Left column: segmentations attained by T-Linkage, point membership is color coded. Right column: the ME committed by J-Linkage and T-Linkage on *Star5* (top) and *Circle4* (bottom) datasets is reported as a function of their corresponding inlier threshold parameters. T-Linkage depends less critically on the choice of the inlier threshold.

5.1. Validation

A simple experiment on simulated data with intersecting structures is here conducted in order to characterize the performances of T-linkage with respect to J-Linkage [10] and confirms the benefits of working with continuous values rather than operating with binary preferences.

We compare the performances of J-Linkage and T-linkage on fitting lines to the *Star5* data (Figure 10a) using the *misclassification error* (ME), defined as follows:

$$\text{ME} = \frac{\#\text{ misclassified points}}{\#\text{ points}}. \tag{16}$$

where a point is *misclassified* when it is assigned to the wrong model, according to the ground-truth.

The results can be appreciated in Figure 10 where the corresponding ME is reported as a function of threshold parameters for both J-Linkage and T-linkage on synthetic datasets. The advantages of T-linkage over J-linkage are twofold. On the one hand T-linkage reaches a lower ME, thereby obtaining a more refined clustering. On the other hand, the threshold parameter integrated in the weighting function is less critical compared to J-Linkage: the error function for T-linkage presents a larger plateau, i.e. a large interval of ϵ the algorithm obtains values near the optimum.

A second experiment is conducted on a real dataset to validate the benefits of T-linkage over its binary counterpart. For this purpose we consider a motion segmentation problem on the Hopkins 155 motion dataset [55], which is available online at `http://www.vision.jhu.edu/data/hopkins155` (see Figure 7 for a sample). The dataset consists of 155 sequences of two and three motions, divided into three categories: checkerboard, traffic, and other (articulated/non-rigid) sequences. In motion segmentation the input data consist of a set of feature trajectories corrupted by noise (no outliers are present), across a video taken by a moving camera, and the problem consists in recovering the different rigid-body motions contained in the dynamic scene. Under the assumption of affine cameras, this task boils down to clustering data-trajectories in a union of subspaces.

The more accurate results of T-linkage with respect to J-Linkage can be appreciated from tables 1 and 2.

The better result for T-linkage is due to the more expressive representation provided by the continuous conceptual space in proximity to model intersections, since residual information allows us to disambiguate more accurately between disputed points. J-Linkage on the contrary has no information to decide to which structure a point in the intersection of two inlier band has to be assigned. In this perspective we have made a little step toward the solution of intersecting models which caused the poor performances of Multi-Ransac.

Indirectly we are also able to compare T-linkage with sequential Ransac, and with algorithms tailored to subspace clustering: SSC [40] LSA [38] and ALC [39] using the figure reported in the site mentioned above. In the two-motion sequences the results of T-linkage are mixed, although it always achieves

		Ransac	LSA 4n	ALC 5	ALC sp	SSC	J-Lnkg	T-Lnkg
Checkerboard	mean	6.52	2.57	2.56	1.49	**1.12**	1.73	1.20
	median	1.75	0.72	**0.00**	0.27	**0.00**	**0.00**	**0.00**
Traffic	mean	2.55	5.43	2.83	1.75	**0.02**	0.70	**0.02**
	median	0.21	1.48	0.30	1.51	**0.00**	**0.00**	**0.00**
Others	mean	7.25	4.10	6.90	10.70	**0.62**	3.49	0.82
	median	2.64	1.22	0.89	0.95	**0.00**	**0.00**	**0.00**
All	mean	5.56	3.45	3.03	2.40	**0.82**	1.62	0.86
	median	1.18	0.59	**0.00**	0.43	**0.00**	**0.00**	**0.00**

Table 1: **Motion segmentation**: misclassification error (%) for video sequences with two motions

		Ransac	LSA 4n	ALC 5	ALC sp	SSC	J-Lnkg	T-Lnkg
Checkerboard	mean	25.78	5.80	6.78	5.00	**2.97**	8.55	7.05
	median	26.02	1.77	0.92	0.66	**0.27**	4.38	2.46
Traffic	mean	12.83	25.07	4.01	8.86	0.58	0.97	**0.48**
	median	11.45	23.79	1.35	0.51	**0.00**	**0.00**	**0.00**
Others	mean	21.38	7.25	7.25	21.08	**1.42**	9.04	7.97
	median	21.38	7.25	7.25	21.08	**0.00**	9.04	7.97
All	mean	22.94	9.73	6.26	6.69	**2.45**	7.06	5.78
	median	22.03	2.33	1.02	0.67	**0.20**	0.73	0.58

Table 2: **Motion segmentation**: misclassification error (%) for video sequences with three motions

a zero median error (as SSC does) and in one case (*Traffic*) also the best average error. The overall average is the second best after SSC, and fairly close to it.

On the three-motion sequences, the results of T-linkage are worse than in the other sequences, and are also somehow odd: on the traffic sequence it achieves the lowest ME, but on *Checkerboard* and *Others* it comes only third (it is second, however, in the mean and median ME).

6. Dealing with outliers

Despite countless efforts spent by the scientific community, there is no universally accepted definition able to capture the elusive nature of outliers. Nevertheless a multitude of approaches have been suggested to characterize outliers; among them we can single out some of the most common assumptions [56]:

- Probability-based : Outliers are a set of small-probability samples with respect to a reference probability distribution.
- Influence-based: Outliers are data that have a relatively large influence on the estimated model parameters. The influence of a sample is normally the difference between the model estimated with and without the sample.
- Consensus-based: Outliers are points that are not consistent with the structure inferred from the remainder of the data.

T-linkage is agnostic about the outliers rejection strategy that comes after; depending on the application, different rejection criteria can be adopted. Since the output of T-Linkage is a partition of data points in consensus sets of the estimated structures, a viable solution is to integrate together the approaches based on probability and consensus by analyzing the cardinality of the attained clusters in a probabilistic framework in order to distinguish between good fits from random ones. This solution can be traced back to Minpran [57] and Plunder [58] and recently appeared also in [59]. More generally this idea is supported by a stream of research rooted in gestalt theory [60, 61] that provides a formal probabilistic method for testing if a model is likely to happen at random or not. The rationale is the Helmholtz principle [62] which asserts that a strong deviation from a background model is valuable information. In our case the background model is determined by outliers, whereas structures of inliers are regarded as an unlikely structure of interest.

First of all we can safely start rejecting all those clusters that have less than $\zeta + 1$ elements since they can be deemed spurious.

Under the mild assumption that outliers are independently distributed [57], it is possible to easily estimate the probability that a cluster is entirely composed of outliers according to its cardinality and the model it defines. Consequently we retain only the groups with a high confidence of being inliers and discard those structures that "happen by chance" and do not reflect an authentic structure in the data.

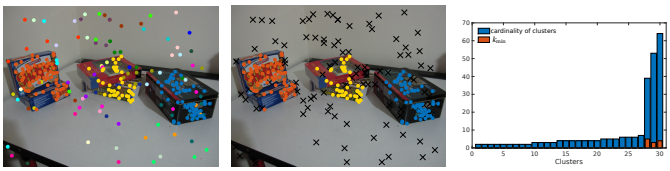
In practice, following Minpran, at first, the probability p that an outlier belongs to the consensus set of an estimated structure is computed by a Monte-Carlo simulation. The value of p can be estimated either in advance for a generic structure, or for every specific model attained by T-Linkage at the end of the clustering. The latter option takes into account the fact that in general models are not all equiprobable and avoids to considering a fixed minimum cardinality. Then the probability that k points belong to the same given model is computed as

$$\alpha(k) = 1 - F(k, n, p), \tag{17}$$

where n is the total number of data points, and F is the binomial cumulative distribution function:

$$F(k, n, p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}. \tag{18}$$

For each structure we compute $k_{\min} = \alpha^{-1}(0.01)$ the minimum cardinality necessary to be not considered mere coincidence. If the considered model is supported by less than $k_{\min}$ points, it is rejected as outlier. [10] Alternatively, based on the observation that large clusters of outliers are very unlikely, if the number κ of structures is known beforehand, it is sufficient to keep the largest κ clusters as inliers.



(a) T-Linkage output (b) Outlier rejection (c) Cluster cardinality

Figure 11: Results obtained by the proposed outlier rejection on the biscuitbookbox dataset. (a) the output produced by T-linkage, (b) outliers are detected and marked with a black ×. (c) histogram of cluster cardinality. Accuracy: 99.22% False number rate: 0%

7. Choosing the scale: a model selection problem

T-linkage does not have any scale selection strategy and the inlier threshold ϵ has to be manually specified by the user, as in MSAC. If prior knowledge about the noise in the data is available, ϵ can be easily tuned, otherwise the scale turns out to be a sensible free parameter even if the use of a soft weighting function mitigates its criticality.

In this section we develop a method for estimating the scale which results in a novel model selection technique avoiding the classical model selection trade-off of two terms in favour of a single term criterion. In particular, we borrow, from the *Consensus Clustering* technique [63], the idea, originally outlined in the context of micro-array data, that the stability of the clustering suffices in disambiguating the correct estimate of models. The rationale behind this method is that the "best" partition of the data is the one most stable with respect to input randomization. We translate this principle in the context of geometric fitting, tailoring the Consensus Clustering strategy to T-Linkage.

It is important to observe that ϵ plays a crucial role in both of these steps of T-Linkage. At first, in the conceptual representation step, the inlier threshold ϵ explicitly defines which points belong to which model (a point belongs to a model if its distance is less than ϵ). If the scale is underestimated the models do not fit all their inliers; on the contrary, if the scale is overestimated, the models are affected by outliers or pseudo outliers. With respect to the clustering step, points are linked together by T-Linkage until their vectorial representations are orthogonal. Here again, as ϵ controls the orthogonality between these vectors, also the final number of models depends on this parameter.

Given a genuine model, if the true noise variance is known, it is always possible to compute a region containing a certain fraction of the inliers. For example, under the typical assumption that the noise for inliers is Gaussian, with zero mean and variance σ^2, the squared point-model errors between an inlier and the uncontaminated model can be represented as chi-square distribution with d degrees of freedom since it is a sum of d squared Gaussian variables, where d is the codimension of the model. For this reason, in order to recover a fraction ρ of inliers, an appropriate threshold ϵ_ρ can be computed as

$$\epsilon_\rho^2 = \chi_d^{-1}(\rho)\,\sigma^2, \tag{19}$$

where χ_d^{-1} is the inverse cumulative chi-square distribution. Hence it is possible to derive the value of the inlier threshold with a certain level of confidence ρ. Many robust estimators of the noise variance have been proposed, one of the most popular ones is the so called MAD (Median Absolute Deviation) which is defined as

$$\text{MAD} = \text{median}_j |\,\text{err}(x_i, \theta) - \text{err}(x_j, \theta)|. \tag{20}$$

Even if this estimator has a 50% breakdown point, it is biased for multiple-mode cases even when the data contains less than 50% outliers.

As a matter of fact, in many real applications selecting the correct scale is a hard problem. In practice many factors hinder scale estimation: the uncertainty of the estimated models has to be taken into account, the presence of high level of contamination due to outliers and multiple structures strains robust estimation and the fact that noise does not always follow gaussian assumption complicates statistical computations. Nevertheless several solutions for automatic scale selection have been proposed. For example this problem is addressed in [26, 64] in regard to the case of single model estimation, whereas [65, 66, 67] treat the case of inlier noise estimation for multiple models exploiting elaborated robust statistics. These techniques rely on the idea of simultaneously estimating a structure together with its inlier threshold. Unfortunately, this captivating strategy turns out to be impossible to integrate in T-linkage. As a matter of fact, T-Linkage merges together clusters as long they have a common structure in their preferences. Therefore a single structure for which ϵ has been erroneously over-estimated is sufficient to cause an incorrect aggregation of clusters and to bias the result towards under segmentation. By way of illustration, one can think of the extreme case where all the sampled structures are computed with the correct scale value, but a single inlier threshold is inaccurately over-estimated in a way that the corresponding consensus set includes all the data points, in this case T-Linkage will return a single cluster.

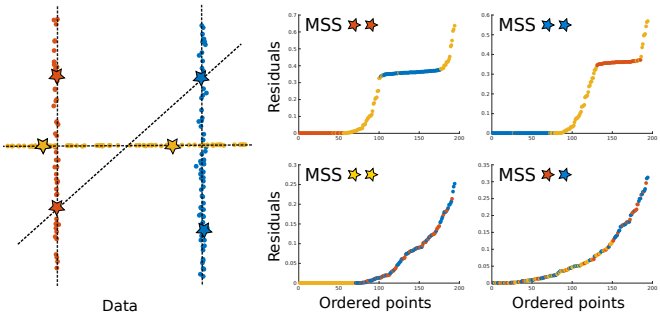


Figure 12: The difficulties inherent in scale estimation for spurious structures. On the left a contrived multi-line fitting example is presented. Data points are sampled, with different level of noise from three ground truth lines (membership of these lines is color coded). Four MSS are drawn, three MSS are pure, the fourth is mixed. Analysing the residuals of the corresponding instantiated model (on the left) clearly shows that as regards the pure MSS ordered residuals clearly exhibit the presence of multi-modal population that can be separate by suitable statistical test. On the contrary in the case of the spurious model (bottom-right) residuals do not present any regularity since there are not enough inlier points. As a consequence scale estimation can not produce a reliable result.

All the thresholds ought to be estimated accurately in a data dependent fashion. However reasoning about the distribution of inlier residuals is not a viable solution as suggested by Figure 12.

In first instance all the scale estimators that rely on variants [68, 67] of the MAD, which has a breakdown point of 50%, can not be adopted because they are prone to over-estimation due to the large number of pseudo-outliers in a common multi-model fitting scenario. Furthermore, the presence of mixed minimal sample sets thwarts all the approaches with a higher breakdown point such as KOSE [69] and IKOSE [36], which substitute MAD with the k-th ordered absolute residual. In this

case the problem is that spurious structures do not have enough supporting points obeying the statistical assumption made by this kind of estimator. Other approaches, e.g. [70], avoid to estimating the scale using all the data points and exploit a forward search method [71]: starting from MSS, the consensus set is expanded until a statistical test on the residuals is verified. Also these methods will definitely fail, because structures arising from impure MSS, produce drifting models and, again, over-estimated scales. In short while scale estimation can work reliably for structures close to the ground truth model parameters, the automatic tuning of the inlier threshold of "random" structures is somehow unfeasible. Unfortunately the ideas presented in Section 6 can not be used in this context to recognize and discard these spurious structures, since ϵ is required as an input to measure the randomness of a model.

For these reasons we found it profitable to tackle the problem from a different perspective. The pivotal observation is that in T-linkage the tuning of ϵ turns out to be a typical model selection problem. If ϵ is too small, we are stuck in under-segmentation: multiple similar structures explain the same model in a redundant way. On the contrary, if ϵ is too large, we run into the problem of over-segmentation obtaining fewer structures than necessary that poorly describe the data. We can therefore identify our scale selection problem as a model selection one. The great advantage of this approach is that by tuning the single free parameter ϵ we are able to implicitly balance at the same time between both the complexity of the obtained structures and their fidelity to the data.

Model selection is a thorny pattern recognition problem that appears ubiquitous in the multi-model fitting literature (see e.g. [37]). As a matter of fact, following the spirit of Occam's razor, several multi-model fitting methods result in minimizing an appropriate cost function composed of two terms: a modelling error and a penalty term for model complexity. Just to name a few relevant algorithms, this approach is taken in [34, 35, 36, 24, 33] where sophisticated and effective minimization techniques such as SA-RCM [24], ARJMC [36] have been proposed. Several alternatives have been explored for encoding model complexity. PEARL [33] for example, optimizes a global energy function that balances geometric errors and regularity of inlier clusters, also exploiting spatial coherence. In [72], an iterative strategy for estimating the inlier-threshold, the score function, named J-Silhouette, is composed of a looseness term, dealing with fidelity, and a separation one, controlling complexity.

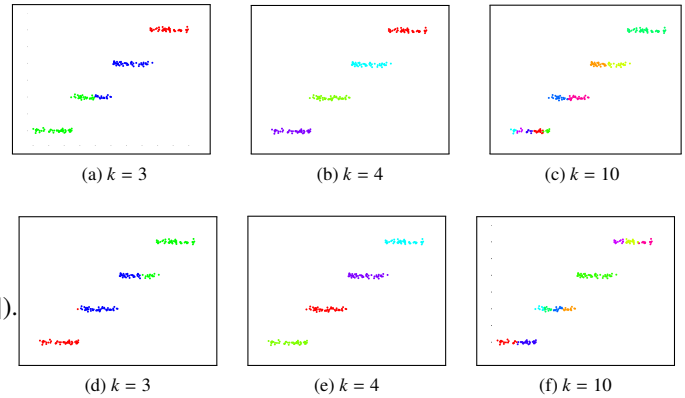
Our starting point is STARSAC [64] in which Choi and Medioni demonstrate that choosing the correct ϵ enforces the stability of the parameter of the solution in the case of a single structure. We extend this result to the multiple structures scenario, reasoning on segmentation rather than on models parameters. The idea of exploiting stability appears in the context of clustering validation. In particular in [63] the authors propose Consensus Clustering, a strategy that succeeds in estimating the number of clusters in the data with a single term model selection criterion based on stability. The next section is devoted to a presentation of the Consensus Clustering approach.

7.1. Consensus Clustering

In some cases the thorny problem of correct tradeoff data fidelity for model complexity (a.k.a. bias-variance dilemma) can be bypassed introducing a different model selection principle based exclusively on the *stability* of models.

The key idea of this approach is that good models should be found among the ones that are stable with respect to small perturbations of the data. This very general principle with the necessary specifications can be applied in many contexts, and can be exploited also in the classical segmentation problem.

For instance, consider the situation illustrated in Figure 13. In this case the models to choose are all the possible partitions of points in k disjoint subsets and model selection is employed for choosing the correct value of k. Running k-means several times on subsamples of the same data, with different values of k shows that the resulting clusterings are stable only when k expresses the nature of the data, otherwise they manifest lack of stability.



(a) $k = 3$ (b) $k = 4$ (c) $k = 10$

(d) $k = 3$ (e) $k = 4$ (f) $k = 10$

Figure 13: Clusters estimation. k-mean is run two times (rows) on subsamples of the same dataset with different values of k (columns). Only for $k = 4$ the attained segmentation is the same. This figure is best viewed in color.

This simple example sustains the intuition that the more stable models represent valid structures in the data.

In [63] the authors develop this idea and present the Consensus Clustering approach to determine the correct number of clusters by maximizing the *consensus*, i.e., the agreement of clustering after perturbation of the data.

More in detail, the Consensus Clustering approach consists in assuming a clustering algorithm, for example k-means, and a resampling scheme (e.g. bootstrapping) in order to perturb the data. Then for each possible cluster number $k = 2, 3, \ldots, k_{\max}$ the data are subsampled several times and processed by the clustering algorithm. The corresponding results are described for each k by means of a *consensus matrix* M_k which is intended to capture the mutual consensus of attained clusters. The consensus matrix M_k is defined as follows: the element $(M_k)_{ij}$ stores the number of times points i and j are assigned to the same cluster divided by the total number of times both items are selected by the resampling scheme. In other words, the consensus matrix records the proportion of clustering runs in which the two points i, j have been clustered together. For this

reason $(M_k)_{ij} \in [0, 1]$ and perfect consensus corresponds to a clean consensus matrix with all the entries equal to either 0 or 1[4], whereas a deviation from this case should be explained with lack of stability of the estimated clusters. Exploiting this observation, the k that yields the cleanest consensus matrices according to an ad hoc measure is selected as the optimal estimate of the number of models.

7.2. *T-Linkage with Consensus Clustering*

In this section we shall concentrate on tailoring Consensus Clustering to T-Linkage, in order to guide the selection of the scale ϵ, leveraging on a single term model selection criterion based on consensus stability.

As explained in Section 7, in the case of T-Linkage we do not have to select the number of clusters (that is automatically determined by T-Linkage clustering) but we shall concentrate on the scale ϵ which, is a sensitive input parameter that implicitly tunes the balance between the complexity of the obtained clusters and their fidelity to the data. If ϵ is too small, we are stuck in under-segmentation: similar multiple structures explain the same model in a redundant way. Alternatively, if ϵ is too large, we run into the problem of over-segmentation obtaining fewer structures than necessary that poorly describe the data.

The outline of our approach is sketched in Figure 14. The estimation of ϵ is iteratively laid out as follows. At first the interval search $[\epsilon_L, \epsilon_R]$ has to be defined, ensuring that the correct ϵ belongs to the interval. For this reason a sound choice of ϵ_L is a small scale value that surely over-segments the data, whereas ϵ_R has to give rise to under-segmentation (for example it can be estimated fitting a single model to all the data points and taking the maximum of their residuals). For each ϵ value belonging to the interval search, T-Linkage is run t times $t = 1, \ldots, t_{\max}$ on the data properly perturbed.

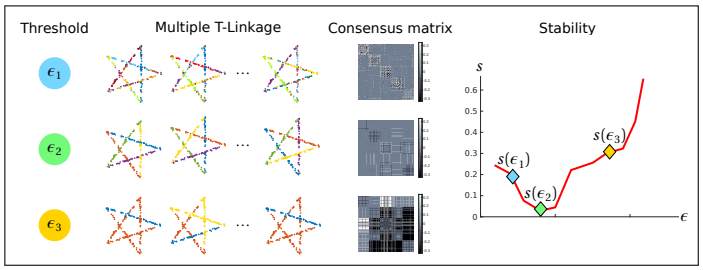


Figure 14: The proposed method in a nutshell. Different ϵ values are used for running multiple times T-Linkage on the perturbed *Star5* dataset. In correspondence to ϵ_1 (which is lower than the ground truth inlier threshold) T-Linkage over-segments the data producing unstable results, when the threshold is ϵ_2 a reliable and stable clustering composed of 5 structures is returned, and finally using the over-estimation of the threshold ϵ_3, data are under-segmented in different ways. The corresponding consensus matrices measure the mutual consensus between the attained segmentations and define the stability index. The most stable clustering, corresponding to ϵ_2, is selected. (Best viewed in color)

[4]If the data points were arranged so that points belonging to the same model are adjacent to each other, perfect consensus would translate into a block-diagonal matrix

Rather than bootstrapping the raw data in advance as in [63], we perturb their representation in the conceptual space inside T-Linkage by bootstrapping the generated hypothesis: In practice, at each iteration, instead of working with the whole hypothesis space H, we extract a subset $\hat{H}$ of tentative structures and use it to construct the conceptual representation of points. In our simulations, we build $\hat{H}$ by uniformly extracting about 90% of structures in H.

After the data have been processed we obtain $t_{\max}$ clustering outputs for each ϵ value. The intuition is that, on the correct scale, there will be consistency between the partitions produced by T-linkage. For each scale the consistency of the partitions is hence tabulated via the consensus clustering matrix M_ϵ introduced in Section 7.1.

Now we measure the consensus stability of each matrix boiling down each M_ϵ to a single consensus stability value s per scale. If we were to plot a histogram of the entries of $(M_\epsilon)_{ij}$, perfect consensus would translate into two bins centered at 0 and 1 and, in general, a histogram skewed toward 0 and 1 indicates good clustering. With this idea in mind, consider the following change of variable:

$$F(x) = \begin{cases} x & \text{if } x < 0.5 \\ x - 1 & \text{if } x \geq 0.5. \end{cases} \tag{21}$$

F redistributes the entries of M_ϵ from the $[0, 1]$ range to the interval $[-0.5, 0.5]$. The effect is to rearrange the histogram symmetrically around the origin. In this way stable entries are concentrated around 0 whereas unstable ones are accumulated at the tails of the histogram. For this reason, measuring how far the entries of $F(M_\epsilon)$ are spread out accounts for the consensus stability of a given scale ϵ. For this purpose we propose to employing the variance[5] of the vectorized upper triangular part of $F(M_\epsilon)$ and defining a *consensus stability index* as

$$s(\epsilon) = \operatorname{var}\left(\operatorname{vech}\left(F(M_\epsilon)\right)\right), \tag{22}$$

where vech returns the vectorization of the upper triangular matrix it receives in input. Then, assuming we are dealing with authentic multiple structures, the scale is selected among the ϵ values that segment the data in at least two clusters. Within these ϵ we retain as correct the smallest one obtaining the lower score of s:

$$\epsilon^* = \min\left(\arg \min_{\epsilon:\ \#\,\text{cluster}>1} s(\epsilon)\right). \tag{23}$$

The most stable solution (the one obtained with ϵ^*) is then returned.

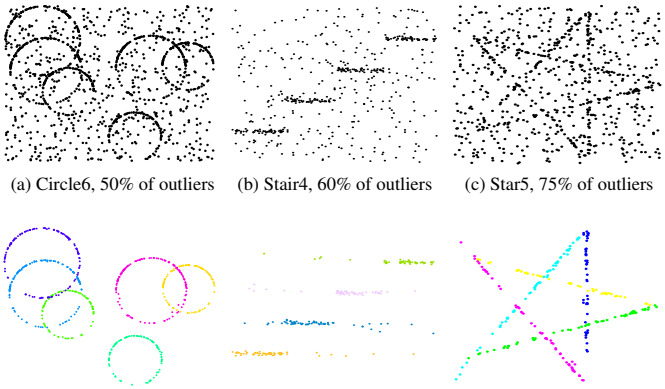
With respect to the computational complexity of this method, if c is the execution time of T-linkage, k_1 the threshold values tested and k_2 the number of bootstrapping trials, the total execution time of this method is $k_1 k_2 c$ to which the time needed for computing the consensus matrices has to be added. Even if the number of bootstrap iterations is small ($k_2 = 4$ in our experiments suffices in providing good results), there is space for improvement for example by replacing exhaustive search on the interval $[\epsilon_L, \epsilon_R]$ with a suitable (direct) minimization strategy reducing the number of scale values that are evaluated.

[5]We also tested other dispersion indices with comparable results.

7.3. Validation

This section is devoted to evaluating the proposed method on both simulated and real data, proving that consensus stability s can be exploited as a single term model selection criterion for automatically fit multiple structures.

Some synthetic experiments are carried on in order to qualitatively assess the proposed approach. In particular, as shown in Figure 15, we address the problem of fitting circles (Figure 15a) and lines (Figures 15b, 15c) to noisy data contaminated by gross outliers. Since the number of structures is unknown – actually it is determined by the parameter ϵ we want to estimate – we do not rely on this information for rejecting outliers. Therefore we employ the outlier rejection strategy described in 6 that discards the structures which happen by chance. It is worth noting that this criterion works properly to filter out bad models with a different percentage of outliers.



(a) Circle6, 50% of outliers (b) Stair4, 60% of outliers (c) Star5, 75% of outliers

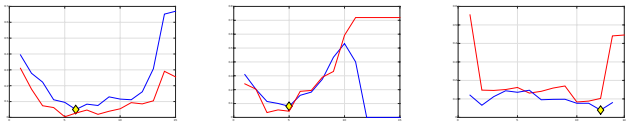
(d) Estimated models (e) Estimated models (f) Estimated models

Figure 15: *Synthetic examples*: rogue data are reported in the first row, extracted models are shown in the second one. Membership is color coded.

We validate this strategy – henceforth referred to as TLCC (T-Linkage and Consensus Clustering)– on some real datasets. We test our method on image pairs correspondences taken from the AdelaideRMF dataset[48] on both two view motion segmentation and plane experiments. The sequences in this dataset consist of matching points in two uncalibrated images with gross outliers. In the case of plane segmentation the (static) scene contains several planes, each giving rise to a set of point correspondences described by a specific homography. The aim is to segment different planes by fitting an homography matrix to subsets of corresponding points. In the second case (*motion segmentation*) the setup is similar, but the scene is not static, i.e., it contains several objects moving independently, each giving rise to a set of point correspondences described by a specific fundamental matrix. The aim is to segment the different motions by fitting fundamental matrices to subsets of corresponding points.

We compare TLCC with the results obtained by T-linkage when an optimal scale ϵ_{opt} in the interval search is given as input:

$$\epsilon_{\text{opt}} = \arg \min_{\epsilon \in [\epsilon_L, \epsilon_R]} \text{ME}(\epsilon), \tag{24}$$



(a) biscuitbookbox (b) cubebreadtoychips (c) johnsonb

Figure 16: Stability index s (blue) and ME (red) as a function of the scale ϵ parameter for some image pairs of the *motion segmentation* (16a, 16b) and *plane segmentation* experiments (16c). The estimated scale is marked with a diamond on the s curve.

in other words ϵ_{opt} is the global minimum of ME. We remind to the reader that ME, defined in Eq.16, measures the percentage of misclassified points with respect to the ground labelling. For each experiment we compare the ME(ϵ^*) achieved in correspondence with the scale ϵ^* estimated by TLCC with the ME(ϵ_{opt}) of the *optimal* scale obtained by T-Linkage^{opt}

Using the data reported in [24] we are able to compare TLCC indirectly with other state of the art algorithms inspired by the classical two term model selection approach. For fair comparison with [24], where the parameters of each sequence are individually tuned and the best outcomes out of several trials have been recorded, we adjust the localized sampling parameters per sequence separately.

In regard to fundamental matrix fitting, according to Table 3, TLCC succeeds in estimating the optimal ϵ in six cases (marked in bold) and misses the global optimum in two cases, for which we plot the ME and the stability index in Figures 16a and 16b. It can be appreciated that the profile of the ME is fairly flat near the optimum, and that the minimum of the stability index is fairly close to the optimum of ME anyway.

Our conjecture for such a behaviour is that the models have mutual intersections (or close to), and that ME does not measure properly the quality of a clustering. For instance, imagine a point P that lies in the intersection of two models, say A and B, and suppose that, according to the ground truth, it is assigned to A. A clustering that assigns P to B is penalized by ME, whereas it should not. A similar argument applies to points that lie close to two models (without belonging exactly to the intersection): the penalty for assigning a point to the wrong model should be attenuated in such close-to-ambiguous situations.

In all but three cases TLCC achieves the best result, and, if the mean ME is considered, it is the best algorithm. These cases are reported in Figures 17a, 17b 17c where it can be appreciated that the resulting segmentation is reasonable anyway.

On plane segmentation experiments, in five cases (marked in bold) the proposed method estimates an optimal scale according to ME.

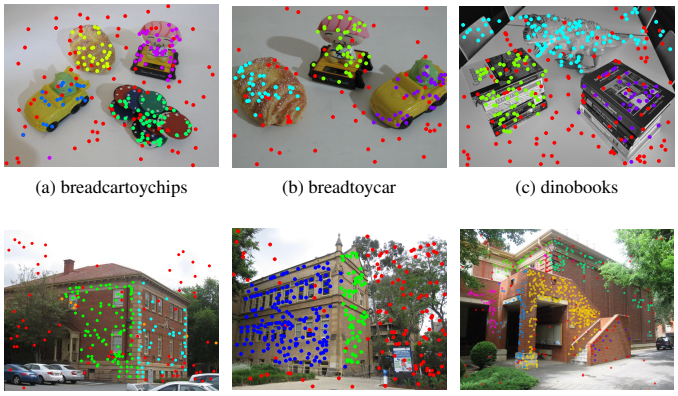
For the *johnsonb* image pairs the attained segmentation by TLCC is slightly less accurate than the optimal one, however from Fig. 16c, where the ME and the stability index are shown, it can be appreciated that the value achieved by TLCC corresponds to a plateau of ME. The segmentation produced by TLCC is presented in Figure 17f. Notice that the actual global optimum of ME can be conditioned by arbitrary tie-breaking of disputed points between models.

	Two term model selection					Stability	
	PEARL	QP-MF	FLOSS	ARJMC	SA-RCM	TLCC	T-Link^{opt}
biscuitbookbox	4.25	9.27	8.88	8.49	7.04	**2.71**	0.39
breadcartoychips	5.91	10.55	11.81	10.97	**4.81**	5.19	5.19
breadcubechips	4.78	9.13	10.00	7.83	7.85	**2.17**	2.17
breadtoycar	6.63	11.45	10.84	9.64	**3.82**	4.27	4.27
carchipscube	11.82	7.58	11.52	11.82	11.75	**1.22**	1.22
cubebreadtoychips	4.89	9.79	11.47	6.42	5.93	**4.46**	3.50
dinobooks	14.72	19.44	17.64	18.61	**8.03**	13.86	13.86
toycubecar	9.5	12.5	11.25	15.5	7.32	**3.03**	3.03
Mean	7.81	11.21	11.68	11.16	7.07	**4.62**	

Table 3: ME (%) for two-view *motion segmentation*.

	Two term model selection					Stability	
	PEARL	QP-MF	FLOSS	ARJMC	SA-RCM	TLCC	T-Link^{opt}
johnsona	4.02	18.5	4.16	6.88	5.9	**3.12**	3.12
johnsonb	18.18	24.65	18.18	21.49	17.95	**8.83**	8.81
ladysymon	**5.49**	18.14	5.91	5.91	7.17	6.17	6.17
neem	5.39	31.95	5.39	8.81	5.81	**4.78**	4.78
oldclassicswing	**1.58**	13.72	1.85	1.85	2.11	**1.65**	1.65
sene	0.80	14	0.80	0.80	0.80	**0.42**	0.42
Mean	5.91	20.16	6.05	7.62	6.62	**4.08**	

Table 4: ME (%) comparison for *plane segmentation*.



(a) breadcartoychips (b) breadtoycar (c) dinobooks

(d) ladysymon (e) oldclassicswing (f) johnsonb

Figure 17: Sample results of TLCC in two-view motion segmentation (top row) and planar segmentation (bottom row). Point membership is color coded, red dots are points rejected as outliers.

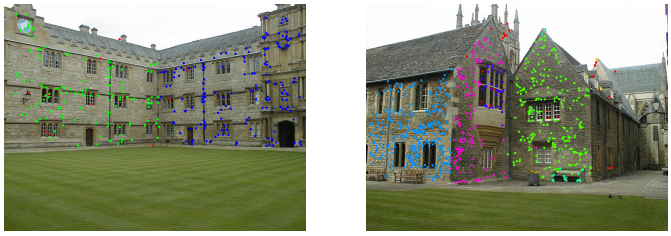
Table 4 compares TLCC with state of the art methods (results for all the methods but TLCC are taken from [24]). Our method achieves in all cases, but one, the best ME and a reasonable segmentation and it scores first on the average.

In order to simulate the performance of this method in a use case scenario where there is no any kind of prior information on the data, we also run TLCC on the whole Adelaide dataset adopting a *pure uniform random sampling* scheme without exploiting local priors. In this challenging scenario, the genuine structures are less represented in the hypothesis space and, as can be seen in Table 5, the attained results of T-Linkage^{opt} are also less accurate with respect to the mixed sampling strategy previously employed. This situation is reflected in the fact that the attained segmentations are less reliable and therefore less stable. As a consequence, TLCC is less precise and often selects a scale that is near to the best available segmentation possible. Nevertheless the achieved ME are still comparable with the optimal ones.

Table 5: Adelaide two view segmentation (uniform sampling)

motion segmentation			plane segmentation		
	TLCC	T-Link^{opt}		TLCC	T-Link^{opt}
biscuitbookbox	2.33	1.55	unionhouse	6.71	5.03
breadcartoychips	5.19	5.19	bonython	1.52	1.52
breadcubechips	3.48	3.48	physics	0.00	0.00
breadtoycar	4.88	4.32	elderhalla	4.67	4.67
carchipscube	4.27	2.44	ladysymon	11.81	10.13
cubebreadtoychips	17.83	15.92	library	7.91	7.44
dinobooks	13.86	13.86	nese	1.97	1.18
toycubecar	10.10	7.07	sene	1.69	1.27
biscuit	0.63	0.00	napiera	10.93	8.94
biscuitbook	0.29	0.29	hartley	10.63	10.00
boardgame	9.68	7.89	oldclassicswing	3.58	3.58
book	2.14	2.14	barrsmith	9.36	6.81
breadcube	0.41	0.41	neem	18.70	18.70
breadtoy	2.43	1.74	elderhallb	17.14	16.73
cube	5.30	5.30	napierb	16.88	12.24
cubetoy	2.81	1.20	johnsona	21.53	8.78
game	0.01	0.01	johnsonb	13.46	12.98
gamebiscuit	25.30	2.44	unihouse	8.72	5.70
cubechips	1.76	1.76	bonhall	14.77	14.77
mean	5.53	3.56	mean	9.58	7.92
median	2.43	1.76	median	9.36	7.44

In Figure 18 we present TLCC results on some plane segmentation experiments taken from the VGG dataset[6].



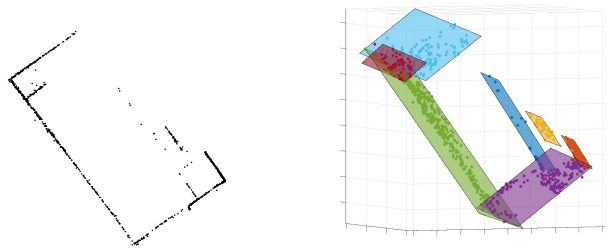
(a) Merton College II (b) Merton College III

Figure 18: Qualitative results of TLCC on stereo images from VGG, Oxford (point membership is color coded, red dots are points rejected as outliers)

We also validate our approach on the 3D plane fitting problem: given an unordered cloud of 3D points, the aim is to aggregate the visual content in significant higher-level geometric structure. Here, we describe the Pozzoveggiani dataset [10] as a combination of different planes representing the wall of the building. As can be sensed from Figure 19, TLCC succeeds in providing an accurate description of the edifice, selecting the scale that recovers all the six walls that can be clearly seen in the top view of the church.

In summary, results show that TLCC, and *a fortiori* T-Linkage, place in the same range as the state of the art competing algorithm adopting a classical two-term model selection strategy, with a free balancing parameter. Experiments show that this method succeeds in suggesting the optimal scale parameter of T-linkage and provides evidence that stability has a minimum in the "right" spot, ideally the same spot where the misclassification error (ME) achieves its minimum.

[6]available online at `http://www.robots.ox.ac.uk/~vgg/data/data-mview.html`



(a) 1000 input points (top view) (b) Segmentation attained by TLCC

Figure 19: Qualitative results on 3D plane segmentation on the Pozzoveggiani dataset

8. Final remarks

We have presented and examined T-Linkage, a simple and versatile tool for geometric model fitting based on preference analysis and hierarchical clustering. This method generalizes and improves J-Linkage, by relaxing the notion of a preference set in the Tanimoto space. The number of sought structures is automatically decided. Moreover, if rogue data are present, the peculiar geometry of the Tanimoto space brings robustness to the clustering step, and outliers can be easily detected and pruned. The only input required by the user is the inlier threshold, whose tuning can be guided by exploiting the proposed consensus clustering framework. In this way T-linkage can be employed to explore and analyze data, when little prior information on the data is available. At the same time, when some prior information is available, this can easily be integrated into T-Linkage, for instance, by tailoring the sampling step to the specific problem at hand.

Further enhancements are possible and are being planned for future work. For example, it would be interesting to exploit the information provided by the hierarchy of nested clusters in order to define robustly an adaptive cutoff of the dendrogram. Also cophenet distances could be profitably integrated in the outlier rejection phase.

References

[1] A. W. Fitzgibbon, A. Zisserman, Multibody structure and motion: 3-d reconstruction of independently moving objects, in: Proceedings of the European Conference on Computer Vision, Springer, 2000, pp. 891–906.

[2] K. E. Ozden, K. Schindler, L. Van Gool, Multibody structure-from-motion in practice, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (6) (2010) 1134–1141.

[3] A.-L. Chauve, P. Labatut, J.-P. Pons, Robust piecewise-planar 3d reconstruction and completion from large-scale unstructured point data, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 1261–1268.

[4] C. Häne, C. Zach, B. Zeisl, M. Pollefeys, A patch prior for dense 3d reconstruction in man-made environments, in: Proceedings of the Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission, IEEE, 2012, pp. 563–570.

[5] R. Toldo, A. Fusiello, Image-consistent patches from unstructured points with j-linkage, Image and Vision Computing 31 (10) (2013) 756–770.

[6] C. V. Stewart, Bias in robust estimation caused by discontinuities and multiple structures, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (8) (1997) 818–833.

[7] L. Magri, 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, 2014.

[8] L. Magri, A. Fusiello, Fitting multiple models via density analysis in tanimoto space, in: Proceedings of the International Conference on Image Analysis and Processing (ICIAP), Lecture Notes in Computer Science, Springer, Genova, IT, 2015, pp. 73–84.

[9] L. Magri, A. Fusiello, Scale estimation in multiple models fitting via consensus clustering, in: Proceedings of the International Conference on Computer Analysis of Images and Patterns (CAIP), Lecture Notes in Computer Science, Springer, Valletta, Malta, 2015, pp. 13 – 25.

[10] R. Toldo, A. Fusiello, Robust multiple structures estimation with j-linkage, in: Proceedings of the European Conference on Computer Vision, 2008, pp. 537–547.

[11] M. A. Fischler, R. C. Bolles, Random Sample Consensus: a paradigm model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381–395.

[12] P. H. S. Torr, A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry, Computer Vision and Image Understanding 78 (2000) 2000.

[13] K. Lebeda, J. Matas, O. Chum, Fixing the locally optimized RANSAC–full experimental evaluation, in: British Machine Vision Conference, 2012, pp. 1–11.

[14] S. Choi, T. Kim, W. Yu, Performance evaluation of RANSAC family, in: British Machine Vision Conference, 2009.

[15] R. Raguram, O. Chum, M. Pollefeys, J. Matas, J. Frahm, USAC: a universal framework for random sample consensus, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8) (2013) 2022–2038.

[16] M. Zuliani, C. S. Kenney, B. S. Manjunath, The multiRANSAC algorithm and its application to detect planar homographies, in: Proceedings of the IEEE International Conference on Image Processing, 2005.

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

[18] R. Subbarao, P. Meer, Nonlinear mean shift for clustering over analytic manifolds, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 1168–1175.

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

[20] D. F. Fouhey, Multi-model estimation in the presence of outliers, Master's thesis, Middlebury College (2011).

[21] W. Zhang, J. Kosecká, Nonparametric estimation of multiple structures with outliers, in: Workshop on Dynamic Vision, European Conference on Computer Vision 2006, Vol. 4358 of Lecture Notes in Computer Science, Springer, 2006, pp. 60–74.

[22] P. Jaccard, Étude comparative de la distribution florale dans une portion des Alpes et des Jura, Bulletin del la Société Vaudoise des Sciences Naturelles 37 (1901) 547–579.

[23] T. Chin, H. Wang, D. Suter, Robust fitting of multiple structures: The statistical learning approach, in: Proceedings of the International Conference on Computer Vision, 2009, pp. 413–420.

[24] T.-T. Pham, T.-J. Chin, J. Yu, D. Suter, The random cluster model for robust geometric fitting, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[25] J. Yu, T. Chin, D. Suter, A global optimization approach to robust multi-model fitting, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[26] R. Raguram, J.-M. Frahm, Recon: Scale-adaptive robust estimation via residual consensus., in: Proceedings of the International Conference on Computer Vision, 2011, pp. 1299–1306.

[27] S. Agarwal, J. Lim, L. Zelnik-manor, P. Perona, D. Kriegman, S. Belongie, Beyond pairwise clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 838–845.

[28] V. M. Govindu, A Tensor Decomposition for Geometric Grouping and Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1 (2005) 1150–1157.

[29] S. Jain, V. M. Govindu, Efficient higher-order clustering on the grassmann manifold, in: Proceedings of the International Conference on Computer Vision, 2013.

[30] R. Zass, A. Shashua, A unifying approach to hard and probabilistic clus-

tering, in: Proceedings of the International Conference on Computer Vision, Vol. 1, 2005, pp. 294–301.
[31] G. Xiao, H. Wang, T. Lai, D. Suter, Hypergraph modelling for geometric model fitting, Pattern Recognition 60 (2016) 748–760.
[32] H. Wang, G. Xiao, Y. Yan, D. Suter, Mode-seeking on hypergraphs for robust geometric model fitting, in: Proceedings of the International Conference on Computer Vision, 2015.
[33] H. Isack, Y. Boykov, Energy-based geometric multi-model fitting, International Journal of Computer Vision 97 (2) (2012) 123–147.
[34] K. Schindler, D. Suter, H. Wang, A model-selection framework for multibody structure-and-motion of image sequences, International Journal of Computer Vision 79 (2) (2008) 159–177.
[35] A. Delong, O. Veksler, Y. Boykov, Fast fusion moves for multi-model estimation, in: Proceedings of the European Conference on Computer Vision, 2012, pp. 370–384.
[36] T.-T. Pham, T.-J. Chin, J. Yu, D. Suter, Simultaneous sampling and multistructure fitting with adaptive reversible jump mcmc., in: Neural Information Processing Systems, 2011, pp. 540–548.
[37] P. H. S. Torr, An assessment of information criteria for motion model selection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (1997) 47–53.
[38] J. Yan, M. Pollefeys, A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and nondegenerate, in: Proceedings of the European Conference on Computer Vision, 2006, pp. 94–106.
[39] S. Rao, R. Tron, R. Vidal, Y. Ma, Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories, Pattern Analysis and Machine Intelligence 32 (10) (2010) 1832–1845.
[40] E. Elhamifar, R. Vidal, Sparse subspace clustering: Algorithm, theory, and applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (11) (2013) 2765–2781.
[41] L. Zappella, et al., Manifold clustering for motion segmentation, Ph.D. thesis, Universitat de Girona (2011).
[42] E. Pekalska, R. P. Duin, The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence), World Scientific Publishing, 2005.
[43] M. Orozco-Alzate, R. P. Duin, G. Castellanos-Domínguez, A generalization of dissimilarity representations using feature lines and feature planes, Pattern Recognition Letters 30 (3) (2009) 242–254.
[44] M. Bicego, V. Murino, M. A. Figueiredo, Similarity-based classification of sequences using hidden markov models, Pattern Recognition 37 (12) (2004) 2281–2291.
[45] C. Lai, D. M. Tax, R. P. Duin, E. Pekalska, P. Paclík, A study on combining image representations for image classification and retrieval, International Journal of Pattern Recognition and Artificial Intelligence 18 (05) (2004) 867–890.
[46] T. Tanimoto, An elementary mathematical theory of classification and prediction, Internal technical report, IBM (1957).
[47] A. H. Lipkus, A proof of the triangle inequality for the tanimoto distance, Journal of Mathematical Chemistry 26 (1-3) (1999) 263–265.
[48] H. S. Wong, T.-J. Chin, J. Yu, D. Suter, Dynamic and hierarchical multistructure geometric model fitting, in: Proceedings of the International Conference on Computer Vision, 2011.
[49] M. Ankerst, M. M. Breunig, H.-P. Kriegel, J. Sander, Optics: ordering points to identify the clustering structure, SIGMOD Record 28 (2) (1999) 49–60.
[50] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Second International Conference on Knowledge Discovery and Data Mining, 1996, pp. 226–231.
[51] T.-J. Chin, J. Yu, D. Suter, Accelerated hypothesis generation for multistructure data via preference analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence (2012) 533–546.
[52] T. Lai, H. Wang, Y. Yan, D.-H. Wang, G. Xiao, Rapid hypothesis generation by combining residual sorting with local constraints, Multimedia Tools and Applications 75 (12) (2016) 7445–7464.
[53] L. Tiwari, S. Anand, Fast hypothesis filtering for multi-structure geometric model fitting, in: Proceedings of the IEEE International Conference on Image Processing, IEEE, 2016, pp. 3728–3732.
[54] R. Xu, D. Wunsch, et al., Survey of clustering algorithms, IEEE Transactions on Neural Networks 16 (3) (2005) 645–678.
[55] R. Tron, R. Vidal, A benchmark for the comparison of 3-d motion segmentation algorithms, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.
[56] A. Y. Yang, S. R. Rao, Y. Ma, Robust statistical estimation and segmentation of multiple subspaces, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2006, pp. 99–99.
[57] C. V. Stewart, MINPRAN: A new robust estimator for computer vision., Pattern Analysis and Machine Intelligence 17 (10) (1995) 925–938.
[58] P. H. S. Torr, A. Zisserman, S. Maybank, Robust detection of degeneracy, in: Proceedings of the International Conference on Computer Vision, Springer–Verlag, 1995, pp. 1037–1044.
[59] M. Tepper, G. Sapiro, A biclustering framework for consensus problems, SIAM Journal on Imaging Sciences 7 (4) (2014) 2488–2525.
[60] A. Desolneux, L. Moisan, J.-M. Morel, From gestalt theory to image analysis: a probabilistic approach, Vol. 34, Springer Science & Business Media, 2007.
[61] L. Moisan, P. Moulon, P. Monasse, Automatic homographic registration of a pair of images, with a contrario elimination of outliers, Image Processing On Line 2 (2012) 56–73.
[62] A. Balinsky, H. Balinsky, S. Simske, On the helmholtz principle for data mining, Hewlett-Packard Development Company, LP.
[63] S. Monti, P. Tamayo, J. Mesirov, T. Golub, Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data, Machine Learning 52 (1-2) (2003) 91–118.
[64] J. Choi, G. G. Medioni, StaRSaC: Stable random sample consensus for parameter estimation., in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009.
[65] S. Mittal, S. Anand, P. Meer, Generalized projection based m-estimator: Theory and applications., in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[66] H. Wang, D. Suter, Robust adaptive-scale parametric model estimation for computer vision, IEEE Transactions on Pattern Analysis and Machine Intelligence (2004) 1459–1474.
[67] L. Fan, T. Pylvänäinen, Robust scale estimation from ensemble inlier sets for random sample consensus methods, in: Proceedings of the European Conference on Computer Vision, 2008, pp. 182–195.
[68] J. V. Miller, C. V. Stewart, Muse: Robust surface fitting using unbiased scale estimates, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 1996, pp. 300–306.
[69] K.-M. Lee, P. Meer, R.-H. Park, Robust adaptive segmentation of range images, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (2) (1998) 200–205.
[70] A. Bab-Hadiashar, D. Suter, Robust segmentation of visual data using ranked unbiased scale estimate, Robotica 17 (06) (1999) 649–660.
[71] A. Atkinson, M. Riani, et al., Exploring multivariate data with the forward search, Springer Science & Business Media, 2013.
[72] R. Toldo, A. Fusiello, Automatic estimation of the inlier threshold in robust multiple structures fitting, in: Proceedings of the International Conference on Image Analysis and Processing, 2009, pp. 123–131.