

Multi-model fitting based on Minimum Spanning Tree

Radwa Fathalla
<http://www.aast.edu/cv.php?ser=36825>

College of Computing and Information
Technology,
Arab Academy for Science and
Technology,
Alexandria, Egypt

George Vogiatzis
<http://www.george-vogiatzis.org>

School of Engineering and Applied
Science,
Aston University,
Birmingham, UK

Abstract

This paper presents a novel approach to the computation of primitive geometrical structures, where no prior knowledge about the visual scene is available and a high level of noise is expected. We based our work on the grouping principles of proximity and similarity, of points and preliminary models. The former was realized using Minimum Spanning Trees (MST), on which we apply a *stable alignment* and *goodness of fit* criteria. As for the latter, we used *spectral clustering* of preliminary models. The algorithm can be generalized to various model fitting settings, without tuning of run parameters. Experiments demonstrate the significant improvement in the localization accuracy of models in plane, homography and motion segmentation examples. The efficiency of the algorithm is not dependent on fine tuning of run parameters like most others in the field.

1 Introduction

Simultaneous parametric estimation of multiple primitive geometric models plays a key role in the interpretation of complex 3d scenes. This is characterized in the literature as a LP3 [1] problem (Irregular sites with discrete labels), on which techniques of unsupervised classification and optimization can be applied. Formally stated, let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ be a set of n data points. It is required to find $L = \{L_i\}_{i=1}^M$, such that L is a set of models that best describe \mathbf{X} . L_i is the parameter vector of model i which, together with the variable M are unknown a priori. In addition, the data points are contaminated by varying levels of outliers.

The literature of model fitting can be broadly categorized into the following subdivisions.

Energy-based formulation. Early attempts include the work in the RANSAC-adaptation to the multi-model case [2] and MLESAC [3]. These algorithms start with randomly populating a set of models. These models compete according to some poor greedy heuristics which results in enhancing each model locally. For example, the cardinality of inliers set assumption fails in case of multiple models in presence of gross outliers. Generally, the oversimplified single objective formulation overlooks cues that are inherent to the human vision

system. These include, the compactness of points in areas that belong to the same model and the intuitive merging of adequately similar models. This gave rise to the need for regularized functions as in PEARL [10]. The starting point for that work was the energy function of the incapacitated facility location problem (FLP) [11], which incorporates total transportation cost and the establishing of a new shop cost. To this function a smoothness prior was added, ensuring the spatial coherence in the search. The random initial set of hypotheses are verified using the α -expansion graph cut optimization. Recently, Yu *et al.* [12] presented a novel energy function. Our criticism of their approach is the over emphasis on the smoothness assumption, evident in the inlier similarity and embedded in the model fidelity terms. In particular, inliers of a model are not solely determined based on residuals but rather on the presence of similar points in the consensus set of a model. Theoretically, this could be amended by the proper assignment of multipliers for these terms. A more serious problem is the inlier similarity. This is not based on spatial proximity but on their residuals towards various models. This can be very misleading in case the models are randomly generated. However, their redundancy eliminating term (Regularizer) is more physically meaningful than the label cost terms in the formulations that preceded them. Furthermore, they pose their energy in the standard Quadratic Program (QP) form handled by efficient constrained optimization techniques. This step produces a ranked list of probable structures preferable over the early hard assignment of points to models. The main shortcomings of their approach are the determination of the trade-off between the various energy terms and susceptibility to local minima. In addition, it is not difficult to find a counter example, as in figure 1 (a), for relying on absolute proximity to enforce spatial coherence.

Similarity-based formulation. This category exploits the fact that a structure can be detected by the presence of several entities sharing a certain property, defined upon a parameter, residual or conceptual space. The entities can be the given points, as the system proposed in [13]. An agglomerative algorithm clusters points and the final models are the best fits of these clusters. The points are expressed by their set of preferred models based on residuals. This approach can be problematic in case of random generation of models that may result in cross structures (figure 1 (b)). Their assumption that residuals for each data point have peaks corresponding to the true model is unrealistic in case of initial random sampling in high levels of outliers, because points can be equidistant to completely different models. Hough transform is commonly used to group models based on their parameters. In [14], parameters are binned and modes are found directly from the histogram. Whereas, in [15] they are mapped directly to the parameter space, on which they applied the mean shift algorithm. There are well-known difficulties associated with the accuracy of mode finding. More importantly, one might argue that it is not straightforward to establish a perceptually uniform space based on the parameters of the Hough transform. Other work in this category include [16].

Our proposed algorithm 1 belongs to the category of model-based similarity formulation of the model fitting problem. It provides a solution to the problems presented in figure 1 by relying on analysis of models layout in space, focusing on point arrangements rather than optimizing on residual values. This is achieved by a novel guided sampling algorithm utilizing Minimum Spanning Trees (MST) at each point. We employ *stable alignment* criterion to select the subtrees from which we hypothesize models for the initial set. The subtree selection is further enhanced by the incorporation of the margin of error criterion. Finally we find the models, after performing spectral clustering on the initial set. Each cluster promotes one of its models to the final set. Our algorithm constructs the models based on both local fitting information from the points and global contextual information from other plausible models.

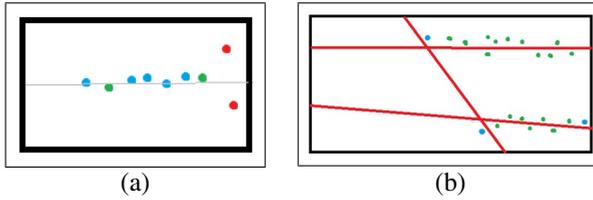


Figure 1: Snapshots of point arrangements, (a) showing 2 inliers in green with an in-between distance larger than the distances between one of them and the gross outliers in red; (b) showing 3 randomly formed models (lines). The points in blue share very similar preference based on the cross structure despite the discrepancy in their true model assignment.

This paper is outlined as follows, in section 2, we present our proposed algorithm. We report the results of the experiments in section 3. Section 4 concludes the paper.

2 Proposed Algorithm

A *good set* of initial hypotheses is vital when applying clustering techniques in multi-model geometric fitting. By a good set, we refer to the repeated presence of optimal/sub-optimal models, in order to form agglomerated dense regions in some space. Unfortunately, random sampling lacks this sense of a “good set”. In its best case scenario, it may provide the optimal/sub-optimal models in the initial set but not in high frequencies, as required. In the random case, the number of putative minimal subsets m_i required to guarantee at least one all-inlier hypothesis, given by $m_i \geq \frac{\log(\rho')}{\log(1-\varepsilon_i^p)}$ [14], where, p is the cardinality of the minimal sample set necessary to establish a model, ε_i is the ratio of inliers of the model i , to the total number of points, and ρ' is the probability of failing to form at least one all-inlier sample in withdrawals. In the general case of having k models, an estimate of the cardinality of the smallest set of inliers is typically used to provide the highest lower bound on m . Adding up the number required for all models to estimate the total, will incur much redundancy because a sample that fails to be all-inlier to one model may fulfill this condition for another. The number required in the initial set scales as a consequence of a decline in ε . This is because the added models introduce pseudo outliers over and above the gross outliers. Pseudo outliers are inliers to true structures other than the examined one. This leads to shrinking the inliers to total number of points ratio thus increasing m . Bearing in mind that this only ensures one all-inlier sample per model, we can imagine the upsurge if many all-inlier samples are needed. For this reason and because our algorithm is dependent on the existence of multiple variants of the optimal model, we have resolved to the guided sampling paradigm.

2.1 MST guided sampling

We propose an algorithm that is generic for Euclidean image space and which belongs to the group of work that focuses on increasing the probability of hitting an all-inlier sample [14]. One approach to categorizing points into possible consensus sets of different structures is relying on spatial proximity, as in [9]. The principle implies that possible inliers are closer to each other than to outliers. But, as we have shown in figure 1 (a), this produces errors in the presence of outliers. Also, it is a well known geometric fact that when building structures out

Input: \mathbf{X} : set of data points

Output: L : set of found models

for each $\mathbf{x}_i \in \mathbf{X}$ **do**

$T_i \leftarrow \emptyset$ /* T_i : subtree of the MST originating at \mathbf{x}_i */

while $\text{size}(T_i) \leq z$ **do**

 expand T_i by adding a point from the MST originating at \mathbf{x}_i

 find the current best fit model of T_i

 calculate model deviation (equation 3) based on current and preceding best fit models

end

 localize valley_of_interest in smoothed model deviation curve as in figure 3 (b)

 find T_i^{elected} which corresponds to the minimum margin of error as in figure 3 (b)

 construct model of best fit to T_i^{elected} and add to initial set

end

construct the similarity matrix between models according to equation 5

perform repeated-2 spectral clustering of models based on the similarity matrix

find centroid models of the clusters and add to the final set of models L

Algorithm 1: Proposed algorithm for model fitting

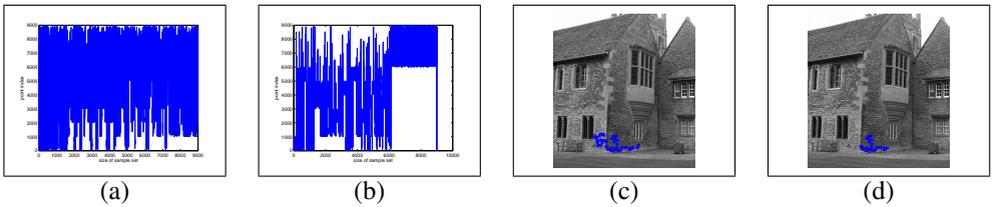


Figure 2: (a) Graphs showing the index of the point added at each iteration of the expansion of the sample set based on Front propagation; (b) based on Minimum spanning tree. In this example, points with indices > 6000 are noise points. It is evident that the bulk of noise points are accessed at late iterations after all the models points have been visited. (c) MST of size z initiated at some point. (d) Subset of the MST that satisfies our criteria.

of a consensus set, it is better to sample far apart points to provide a better fit to the whole set of inliers. In addition, methods that sample locally based on proximity tend to produce isolated patches of models in case of gaps due to partial occlusion. While, sampling based on residual as in [10], is capable of joining disconnected patches of points of a model.

We go for a compromise between proximity and spread. We begin by deterministically forming the tentative models. At each point we initiate a sample set. Gradually, this set is expanded by incorporating more points. With the addition of each point, we find the best fitting model of the formed set, the absolute residuals \mathbf{r} of the points in its sample set, and the number and absolute residuals \mathbf{s} of the points in the consensus set of this model. The consensus set is the points in \mathbf{X} , which exhibit an acceptable residual to the model. The acceptable values lie in the interval $[0, \text{mean}(\mathbf{r}) + \text{std}(\mathbf{r})]$. For the expansion process, seed fill algorithms are a common choice. However, we resort to the Prim's algorithm [11] for finding the MST, because in our applications, point clouds and feature points of sampled images, are represented as Euclidean Complete Graphs (ECG). Front propagation algorithms are ill-defined over ECGs, because there is a direct path between all-pairs. So, when the

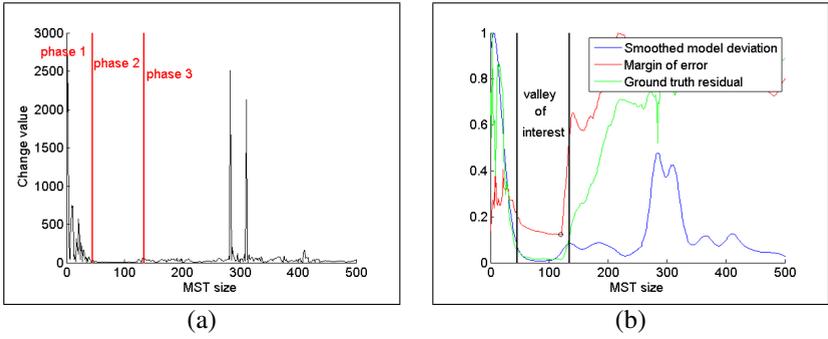


Figure 3: MS sub-T selection criteria (a) A model deviation signal showing a typical behavior of models constructed from subsets of the sample set; (b) A graph showing the smoothed model deviation, margin of error, ground truth residual. Two vertical lines marking the valley of interest. The circle shape marks the chosen subtree size for this point.

target is the shortest path, then it will be only one edge, which is the direct edge. Choosing a neighbourhood graph as a means of representation using thresholding of the distances or voxelization is a possible solution but it has the undesirable effect of inducing multiple Connected Components (CC) and binding the formed models to the localities of these CCs. In our view, the terminating criteria for the spread of a model should be based solely on the decline of its fitness, to be able to account for signal interruptions of partial occlusions. More importantly, we argue, the MST is more robust to noise. Bearing in mind, the geodesic path between 2 inliers of a model is on its surface, the MST will start by spreading over this surface (figure 2 (a)). When most of the surface points are visited, the algorithm backtracks to visit other points in the vicinity of the surface. These have a high possibility of being outliers. Hopefully, at this point the established model will be more fitting to its inliers, and will be less affected by the addition of outliers. Seed fill algorithms, on the other hand, blindly flood the nearest neighbours in a breadth first search manner. This is more prone to adding noise points as in figure 1 (a). Even, if a certain criterion was used to direct the traversal of the graph, such as the closest neighbour or a more elaborate one as a gain in the model likelihood, its lack of a backtracking mechanism causes good candidate points to be lost at high levels of the traversal tree. At each iteration of Prim's algorithm, the edge with the least value connected to the growing spanning tree is selected, provided that it increases the number of nodes in the tree.

The fundamental question is when to stop the growth of the MST. As we mentioned before, each point generates a set of plausible models. Each model corresponds to a subtree of the MST initiated at this point. The growth continues until a maximum size of z is reached and we select the optimal subtree by combining 2 criteria namely, *stable alignment* followed by *margin of error*. We start examining these generated models at each point, by recording the *model deviation*, indicating the degree of change that happened to the best fitted model over the previous single expansion step. The model deviation is basically a dissimilarity measure $d_{i(i-1)}$ between consecutive models. Typically, the graph of model deviation (figure 3 (a)) can be divided in 3 phases:

- *Phase 1*: It is characterized by sharp ripples. This shows the models undergo substantial changes at the start of the MST growth process. Because, the addition of a single

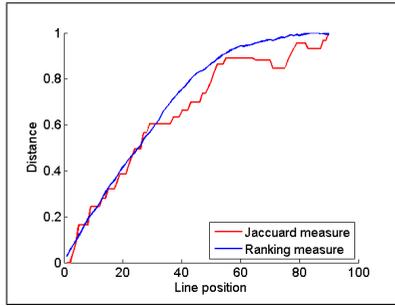


Figure 4: Normalized d_{0i} between the rotating line and the x-axis at each angle position.

point (inlier or outlier), makes a profound effect on the alignment of models of smaller sets of points.

- *Phase 2*: It can be described roughly as a plateau region. This indicates a stability in the model formation. The reason is, the growing size of the subtree enhances the spread of inlier points and the adherence of the generated model to the underlying structure. Thus, the inclusion of gross outliers in the subtree does not alter the alignment of the model. As long as, the density of the inliers exceeds that of the outliers in a manner that allows the presence of an underlying structure in the first place.
- *Phase 3*: The deviation value increases again, mainly, due to the inclusion of tangent pseudo outliers.

The model deviation should capture the change in alignment when viewed globally as accurately as possible. Due to the irregular nature of the space on which the models are defined, we opt for an arbitrary dissimilarity measure. We quantify the difference between models L_i and $L_{(i-1)}$ by starting with an ordering of the n data points according to each model L_i . First, we calculate their absolute residuals to the model to form the residual vector:

$$\mathbf{r}^{(i)} = [r_1^i r_2^i \dots r_n^i] \quad (1)$$

This vector is subjected to sorting in non-descending order. The new indices given to all points are recorded in $\mathbf{a}^{(i)}$:

$$\mathbf{a}^{(i)} = [a_1^i a_2^i \dots a_n^i] \quad (2)$$

The dissimilarity d_{ij} is calculated as the total deviation in the sorting of points according to L_i and $L_{(i-1)}$, as follows:

$$d_{i(i-1)} = \sum_{o=1}^n \left| a_o^{(i)} - a_o^{(i-1)} \right| \quad (3)$$

This dissimilarity measure resembles the Jaccard distance defined on preference sets in [15], but with a more global outlook and does not suffer from the deficits of residual thresholding. To illustrate how the measure operates, we sampled a number of 2d points around a line, and then the line, was incrementally rotated by an angle of 1° for 90 times. At each angle position, the d_{0i} was recorded based on both the Jaccard and our ranking dissimilarity, (figure 4). Our measure was shown to be more linear and sensitive to small perturbations. In figure 3 (b), it is evident that phase 2 of the model deviation roughly coincides with the lowest

values of the average residuals of the ground truth model points, suggesting a good fit of the generated models in this phase. We localize this region in the graph by first convolving the model deviation signal with a smoothing 1d Gaussian filter. We perform a 1d watershed transform and find the segmented part of the first basin whose local minimum does not coincide with the first index of the model deviation vector. We then perform thresholding to trim the heights of the peaks of the valley to the shortest of them. We refer to this procedure as applying the stable alignment criterion. To elect the best fitting model, we seek the sample subset with the least margin of error t_i defined as:

$$t_i = 1.96 \times \frac{\text{std}(\mathbf{s}_i)}{\sqrt{|\mathbf{s}_i|}} \quad (4)$$

$|\mathbf{s}_i|$ is the size of the consensus set of the model i . We utilize t_i as the *goodness of fit* in our algorithm. The larger the margin of error, the less confident one could be that the data points resulted from a true structure. Geometrically, it indicates how well aligned and dense the points are in the consensus zone.

One of the challenges for this algorithm is, tangent models. One can argue that MSTs resulting from intersecting models constitute a small fraction in the first place, as explained in [8]. In addition, exploiting some domain specific knowledge to eliminate edge points will radically solve the problem. Nevertheless, we try to provide a generic algorithm that will act blindly on point clouds. Also, in our application, edge points are among the highly sampled genre of points, as they hold strong characteristics of a region. Figure 2 (b), shows that our combined criterion is capable of choosing the subtree of the MST that truly belongs to a single model. The initial set is then formed of the best fits to the MST subsets. The RANSAC recommendation provides the ceil for the size of the initial set. In practice, our algorithm managed with a fraction of this quantity. An advantage of our algorithm is, the whole set of points is available each time a new model is formed. In contrast to the multi-RANSAC, when the set of inliers of a model are completely eliminated in future iterations. In contrast to previous work on MSTs for clustering [14], we combine proximity with the stability criterion to promote it from a proximal point aggregation tool to one more capable of model detection.

2.2 Multiplicity guided model detection

True structures in our algorithm manifest themselves with multiple structures that are relatively close to them in the initial set. This multiplicity is a quality that reinforces the existence of an underlying structure or stimuli. To exploit this principle computationally, we have utilized the spectral clustering technique. It is specifically suitable for this application, because most other methods need a definition of the absolute location of the elements in some space. In contrast, spectral clustering relies on the similarities between the elements. Thus, we construct a difference matrix of size $m \times m$ in which each cell (i, j) indicates the degree of dissimilarity d_{ij} . It is then converted to a similarity matrix as follows:

$$w_{ij} = 1 - \frac{d_{ij}}{\max_{i,j} d_{ij}} \quad (5)$$

Again, for the calculation of d_{ij} , we use equation 3 to assess a conceptual dissimilarity based on the models rankings of points. The similarity matrix is then passed to spectral clustering

Method	Random sampling	Multi-GS	MST-GS
Example 1	0.3875	0.3418	0.3328
Example 2	154.4047	93.7360	12.3048
Example 3	41.5376	10.3648	1.8447
Example 4	0.0551	0.0340	0.0244

Table 1: Average per-model residual of closest group of ground truth model points to their best fitting models in the initial set.

[8] to produce subsets. We handled the issue of the unknown number of models and subsequently clusters with the *Repeated 2-clustering method*. Clustering of similar models is done hierarchically in a top-down approach. Each node is hypothetically split into 2 subsets. The regularization function is the *Daives Bouldin* (DB) measure [9], used for internal evaluation of clustering and here it is calculated using all the current clusters. If this break down introduces an improvement i.e. decreases DB index, then it is carried out. Otherwise, the node is left as a leaf. At the end of examining all the nodes, the clusters are the leaves.

For selecting the final set of models from the clusters. We find the centroids $L_e^{(i)}$ i.e. model that is least dissimilar to the rest of models with it in cluster c_i , (equation 6). In contrast to RANSAC, this approach is unbiased to models with large consensus sets.

$$L_e^{(i)} = \arg_{L_k^{(i)} \in c_i} \min \left(\sum_{j=1}^{|c_i|} d_{kj} \right) \quad (6)$$

3 Experimental Evaluation

We validate our proposed algorithm by testing it in the applications of plane, homography fitting, and motion segmentation in the case of multiple structures. In our experiments on estimating homographies and motion segmentation, we assume the correspondence problem is solved by matching of SIFT descriptors [10] and preconditioned point matches are readily available. We follow [9] in the use of the Direct Linear Transformation (DLT) algorithm for fitting the homographies and the symmetric transfer error for calculating residuals. In case of motion segmentation, we use the normalized 8 point algorithm for estimating the fundamental matrix for each motion and the squared Sampson distance for the geometric errors. As for the planes application, Principle Component Analysis (PCA) is used to establish the fits and the residuals are the perpendicular distances from the points to the planes. We hereby present 4 examples [11], whose ground truth information is available in the form of membership of points. *Example 1*- Synthetic cube of 6 planes. Each cube face consists of 1000 points, outliers count=3000 and Gaussian noise= 0.015. *Example 2*- Wadham is a homography example of two planar regions. Models consist of 52 and 65 points respectively, outliers count=176. *Example 3*- Merton III is a homography example of three models. Models consist of 498, 394 and 570 points, outliers count= 90. *Example 4*- Dino books is a motion segmentation example that has three motions. Models consist of 78, 86 and 41 points, outliers count =155.

	Recall			Precision			M		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Ex.1	0.9143	0.8552	0.9672	0.6094	0.7004	0.6448	9	9	6
Ex.2	0.7115	0.7692	0.8316	0.7253	0.4657	0.6014	4	4	2
Ex.3	0.9164	0.9443	0.9897	0.9298	0.9011	0.9404	5	5	3
Ex.4	0.9950	0.9077	0.9915	0.7803	0.6229	0.5271	5	5	3

Table 2: Average recall, precision values and the count of detected models for results of (a) J-linkage, (b) PEARL, (c) Our proposed algorithm .

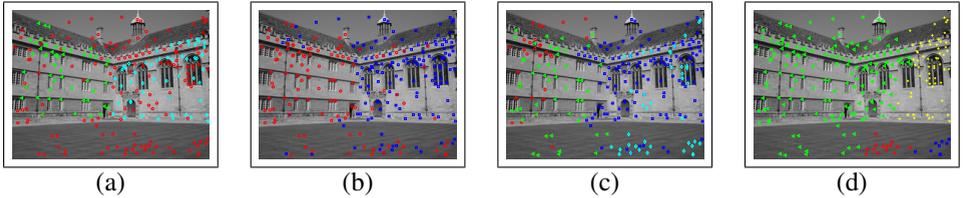


Figure 5: Wadham. (a) Ground truth (outlier points added shown as red circles). Results of (b) proposed algorithm; (c) J-linkage; (d) PEARL (results are different from those reported in their paper because utilized parameters of energy function were not given).

3.1 MST guided sampling

In this section, we compare our results against random sampling and the multi-GS [10]. It directs the selection process of points for a sample towards inliers that exhibit the same residual, as the seed point to the models in a pilot set. In the guided sampling paradigm, the prevailing performance measure is the percentage of all-inlier samples. However, this is not a guarantee for a good putative model. The sample must consist of inliers that span the manifold. In some cases, a model generated from a sample of mixed points is closer to the ground truth model than from an all-inlier counterpart, provided that the signal to noise ratio in the sample set is high. Instead, we assign a putative model to the closest group of ground truth points. Then, we calculate their average residual. The value of the z parameter ranged from 50 to 500 in our experiments based on the sparsity of the test data, which was found to be adequate for the model deviation to stabilize. As shown in table 1, MST guided sampling consistently outperformed the rest. We point out that when applying MST guided sampling, there occurred some repetitions in the sampling sets. This redundancy favorably enhances the performance of the subsequent clustering algorithm.

3.2 Multiplicity guided model detection

Figures 5, 6, and 7, show the results of our proposed algorithm, J-linkage [15] and PEARL [2] on the tested examples. We consider the classification power with respect to the data points of true models. For this purpose, we use the precision and recall values in table 2. In most cases of our approach, the precision values are lower than the other methods [2, 15], because, spectral clustering works by minimizing the cut in the similarity graph, thus maximizing the intra-cluster distance. In effect, this distributes the noise points in different clusters rather than aggregating them in a single outlier cluster. This increases the count

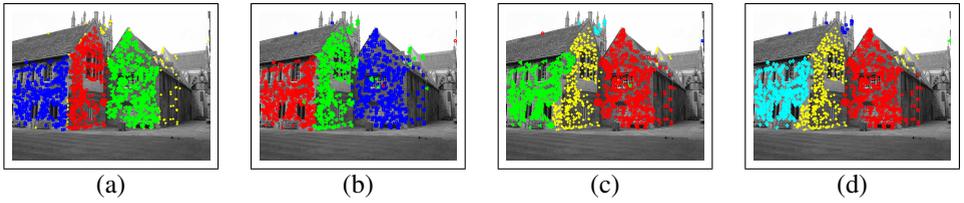


Figure 6: Merton college III. (a) Ground truth (outlier points added shown as yellow crosses). Results of (b) proposed algorithm; (c) J-linkage; (d) PEARL.

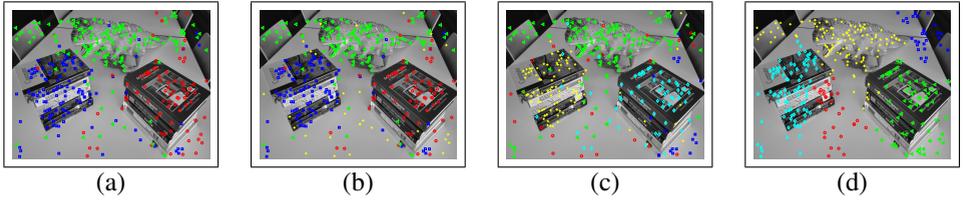


Figure 7: Dino books. Result of (a) proposed algorithm; (b) post processing of outlier residual filtering increases precision to 0.7001; (c) J-linkage; (d) PEARL.

of false positives per detected model. The precision figure is not alarming as long as the inclusion of outliers does not corrupt the models. This is proved by the enhancement in the recall figure, which shows its power in aggregating correct points to the models. This is due to the fact that our centroid finding technique is robust even in high levels of outliers, because it does not re-fit the model to the found cluster but rather elects its most probable model. We found that our algorithm always resulted in the correct minimal number of models. Such a goal has been advocated in PEARL [7] by the inclusion of a label penalty. In practice, however, the optimal choice of the label penalty is difficult.

4 Conclusion

We have presented a system for multi-model detection. We resorted to the guided sampling paradigm to hypothesize models and unprecedentedly used the MST, in this respect. Other novelties in our work include, a different perspective to defining goodness-of-fit, that depends on compactness of points and stability in the layout of models. Also, we proposed a model-to-model distance measure based on ranking that was shown to be very effective in describing model variation. The algorithm has the advantages of having no model-specific assumptions, being insensitive to model size variability and tolerating high levels of outliers. Finally, we showed that our algorithm outperforms the state-of-the-art in multi-model fitting.

References

- [1] Tat-Jun Chin, Jin Yu, and David Suter. Accelerated Hypothesis Generation for Multi-structure Data via Preference Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(4): 625–638, April 2012. ISSN 0162-8828.

- [2] Dorin Comaniciu and Peter Meer. Mean Shift : A Robust Approach Toward Feature Space Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.
- [3] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition, 2009. ISBN 0262033844, 9780262033848.
- [4] David L Davies and Donald W Bouldin. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1(2):224–227, February 1979. ISSN 0162-8828.
- [5] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [6] Reza Hoseinnezhad, Ba-Ngu Vo, and David Suter. Fast segmentation of multiple motions. *Cognitive systems*, 2009.
- [7] Hossam Isack and Yuri Boykov. Energy-Based Geometric Multi-model Fitting. *Int. J. Comput. Vision*, 97(2):123–147, April 2012. ISSN 0920-5691.
- [8] Francis R Bach Michael I Jordan. Learning spectral clustering. *Advances in Neural Information Processing Systems*, 16:305, 2004.
- [9] Yasushi Kanazawa and Hiroshi Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *British Machine Vision Conference*, volume 1, pages 247–256. Citeseer, 2004.
- [10] Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. In *Proceedings of the 38th international conference on Automata, languages and programming*, ICALP' 11, pages 77–88, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-22011-1.
- [11] S.Z. Li. Markov random field models in computer vision. In *Third European Conference on Computer Vision , Part II.*, pages 361–370, 1994.
- [12] David G Lowe. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the International Conference on Computer Vision-Volume 2*, ICCV '99, pages 1150–1156, Washington, DC, USA, 1999. IEEE Computer Society. ISBN 0-7695-0164-8.
- [13] C.V. Stewart. Bias in Robust Estimation caused by discontinuities and multiple structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(8):331–338, 1997.
- [14] Mariano Tepper, Pablo Musé, and Andrés Almansa. Meaningful Clustered Forest: an Automatic and Robust Clustering Algorithm. *CoRR*, abs/1104.0, 2011.
- [15] Roberto Toldo and Andrea Fusiello. Robust Multiple Structures Estimation with J-Linkage. In *Proceedings of the 10th European Conference on Computer Vision: Part I*, ECCV '08, pages 537–547, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-88681-5.
- [16] P H S Torr and A Zisserman. MLESAC : A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding - Special issue on robusst statistical techniques in image understanding*, 78(1):138–156, 2000.

- [17] Lei Xu, Erkki Oja, and Pekka Kultanen. A new curve detection method: Randomized Hough transform (RHT). *Pattern Recognition Letters*, 11(5):331–338, 1990. ISSN 0167-8655.
- [18] Jin Yu, Tat-Jun Chin, and David Suter. A global optimization approach to robust multi-model fitting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2041–2048. IEEE, 2011.
- [19] Wei Zhang and Jana Kosecka. Nonparametric estimation of multiple structures with outliers. In *Proceedings of the international conference on Dynamical vision*, pages 60–74, 2006.