Robust Multiple Structures Estimation with J-Linkage

Roberto Toldo and Andrea Fusiello

Dipartimento di Informatica, Università di Verona
Strada Le Grazie 15, 37134 Verona, Italy
andrea.fusiello@univr.it

Abstract. This paper tackles the problem of fitting multiple instances of a model to data corrupted by noise and outliers. The proposed solution is based on random sampling and conceptual data representation. Each point is represented with the characteristic function of the set of random models that fit the point. A tailored agglomerative clustering, called J-linkage, is used to group points belonging to the same model. The method does not require prior specification of the number of models, nor it necessitate parameters tuning. Experimental results demonstrate the superior performances of the algorithm.

1 Introduction

A widespread problem in Computer Vision is fitting a model to noisy data: The RANSAC algorithm is the common practice for that task. It works reliably when data contains measurements from a single structure corrupted by gross outliers.

When multiple instances of the same structure are present in the data, the problem becomes tough, as the robust estimator must tolerate both gross outliers and *pseudo-outliers*. The latter are defined in [1] as "outliers to the structure of interest but inliers to a different structure". The difficulty arise because robust estimators, including RANSAC, are designed to extract a single model. Sequential RANSAC – sequentially apply RANSAC and remove the inliers from the data set as each model instance is detected – has been proposed as a solution, but [2] argued that the approach is non optimal, and introduced the multi-RANSAC algorithm. The method is effective (provided that the models do not intersect each other), but the number of models is user specified, and this is not acceptable in some applications.

Another popular method based on random sampling and voting is the Randomized Hough Transform (RHT) [3]. It builds an histogram over the parameter space. A minimal sample set is randomly selected and the parameters of the unique model that it defines are recorded in the histogram. Peaks in the histogram correspond to the sought model. Its extension to multiple models is straightforward: they are revealed as multiple peaks in the parameter space. Hence, RHT does not need to know the number of models beforehand. However,

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part I, LNCS 5302, pp. 537–547, 2008.

RHT suffers from the typical shortcomings of Hough Transform methods, such as limited accuracy and low computational efficiency. Ultimately, the choice and the discretization of the parameter space turn out to be crucial.

RHT can be seen as an instance of a more general approach consisting of finding modes in parameter space (see e.g. [4]). Instead of quantize the parameter space and accumulate votes, one can map data into the parameter space through random sampling and then seek the modes of the distribution with mean-shift [5]. This, however, is not an intrinsically robust technique, even if it can be robustified with outliers rejection heuristics. Moreover, the choice of the parametrization is critical, as in RHT.

In summary, RANSAC is very robust, but it is not suited to deal with multiple structures. Mode finding in parameter space (and RHT), on the contrary, copes naturally with multiple structures, but cannot deal with high percentage of gross outliers, especially as the number of models grows and the distribution of inliers per model is uneven. Also the algebraic technique presented in [6] is effective in estimating multiple models, but it is not robust to gross outliers.

Recently [7] proposed a novel method for estimating multiple structures based on the analysis of the distribution of residuals of individual data points with respect to the hypotheses, generated by a RANSAC-like sampling process. It turns out that the modes of the residuals distribution reflects the model instances. With respect to RANSAC and other robust methods (such as LMeds, for example) this entails a change of perspective: "studying the distribution of the residuals for each data point instead of studying the distribution of residuals per each hypothesis" [7].

Residuals for each data point have peaks corresponding to the true models because hypotheses generated with random sampling tend to cluster around the true model, a fact that is also at the basis of RHT. The method, in principle, can discover the number of models automatically as in RHT and is effective as RANSAC. However, finding modes ends up to be cumbersome, as proved in our experiments. One reason is that the peak corresponding to a given model becomes less localized as the point-model distance increases. As a result, the rightmost modes in the histogram are usually drowned in the noise.

In the same spirit of RHT and [7] we exploit clustering of the hypotheses. However we do not work in the parameter 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 [7]. We adopt instead a *conceptual representation*: each data point is represented with the characteristic function of the set of models preferred by that point[1]. Multiple models are revealed as clusters in the conceptual space. Experimental comparison with sequential RANSAC, multiRANSAC, residual histogram analysis [7] and mean-shift is favorable to our algorithm.

[1] According to [8] the posterior probabilities of an object x given C classes form a *similarity conceptual representation*: $[P(x|\text{ class } 1)\cdots P(x|\text{ class } C)]$.

2 Method

The method starts with random sampling: M model hypothesis are generated by drawing M minimal sets of data points necessary to estimate the model, called minimal sample sets (MSS). Then the consensus set (CS) of each model is computed, as in RANSAC. The CS of a model is the set of points such that their distance from the model is less than a threshold ε.

Imagine to build a $N \times M$ matrix where entry (i, j) is 1 if point i belongs to the CS of model j, 0 otherwise. Each column of that matrix is the characteristic function of the CS of a model hypothesis. Each row indicates which models a points has given consensus to, i.e., which models it prefers. We call this the *preference set* (PS) of a point. Figure 1 shows an example of such a matrix in a concrete case.

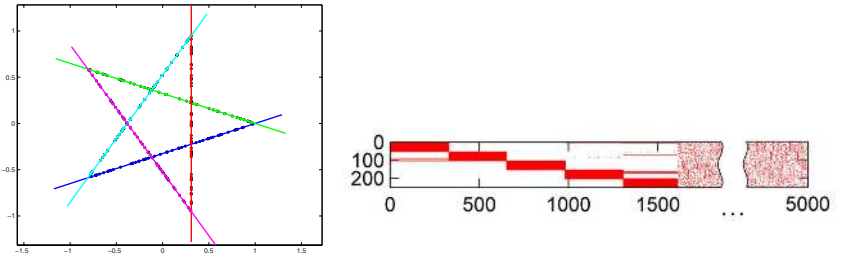


Fig. 1. Right: the data consist of 250 points on five segments forming a star. Left: Preference matrix. The rows are points (ordered by cluster), the columns are models (ordered by cluster size).

The characteristic function of the preference set of a point can be regarded as a conceptual representation of that point. Points belonging to the same structure will have similar conceptual representations, in other words, they will cluster in the *conceptual space* $\{0, 1\}^M$. This is, again, a consequence of the fact that models generated with random sampling cluster in the hypothesis space around the true models.

2.1 Random Sampling

Minimal sample sets are constructed in a way that neighbouring points are selected with higher probability, as suggested in [9,2]. Namely, if a point $\mathbf{x}_i$ has already been selected, then $\mathbf{x}_j$ has the following probability of being drawn:

$$P(\mathbf{x}_j|\mathbf{x}_i) = \begin{cases} \frac{1}{Z} \exp -\frac{||\mathbf{x}_j - \mathbf{x}_i||^2}{\sigma^2} & \text{if } \mathbf{x}_j \neq \mathbf{x}_i \\ 0 & \text{if } \mathbf{x}_j = \mathbf{x}_i \end{cases} \tag{1}$$

where Z is a normalization constant and σ is chosen heuristically.

Then for each points its preference set is computed, as the set of models such that the distance from the point is less than the inlier threshold ε (same as RANSAC).

The number M of MSS to be drawn is related to the percentage of outlier and must be large enough so that a certain number (at least) of outlier-free MSS are obtained with a given probability for all the models. Please note that if this condition is verified for the model with less inliers, it is automatically verified for all the other models.

Let S be the number of inliers for a given model and N be the total number of points. The probability of drawing a MSS of cardinality d composed only of inliers is given by:

$$p = P(E_1)P(E_2|E_1)\dots P(E_d|E_1, E_2\dots E_{d-i}) \tag{2}$$

where E_i is the event "extract an inlier at the i-th drawing". In the case of uniform sampling $P(E_i|E_1, E_2\dots E_{i-1}) = \frac{S-i+1}{N-i+1}$. In our case, the first point is sampled with uniform probability, hence $P(E_1) = S/N$, while the others are sampled with the probability function (1), therefore, after expanding the normalization constant Z, the conditional probability can be approximated as

$$P(E_i|E_1, E_2\dots E_{i-1}) = \frac{(S-i+1)\exp -\frac{\alpha^2}{\sigma^2}}{(N-S-i+1)\exp -\frac{\omega^2}{\sigma^2} + (S-i+1)\exp -\frac{\alpha^2}{\sigma^2}} \quad i = 2\dots d \tag{3}$$

where α is the average inlier-inlier distance, and ω is the average inlier-outlier distance. If $S \gg d$ then

$$p \simeq \delta \left(\frac{\delta \exp -\frac{\alpha^2}{\sigma^2}}{(1-\delta)\exp -\frac{\omega^2}{\sigma^2} + \delta \exp -\frac{\alpha^2}{\sigma^2}} \right)^{d-1}. \tag{4}$$

where $\delta = S/N$ is the inlier fraction for a given model. Therefore, assuming that ω is larger than α, the sampling strategy increases the probability of extracting an outlier-free MSS, as the intuition would also suggests.

Finally, the probability of drawing at least K outlier-free MSS out of M, for a given model, is given by [7]:

$$\rho = 1 - \sum_{k=0}^{K-1} \binom{M}{k} p^k (1-p)^{M-k}. \tag{5}$$

This equation is used to compute the required number of samples M for a given confidence ρ and a given K. The value of δ in (4) must be set to the smallest inliers fraction among all the models. Values of ρ vs M are shown in Fig. 2.

2.2 J-Linkage Clustering

Models are extracted by agglomerative clustering of data points in the conceptual space, where each point is represented by (the characteristic function of) its preference set.

The general agglomerative clustering algorithm proceeds in a bottom-up manner: Starting from all singletons, each sweep of the algorithm merges the two

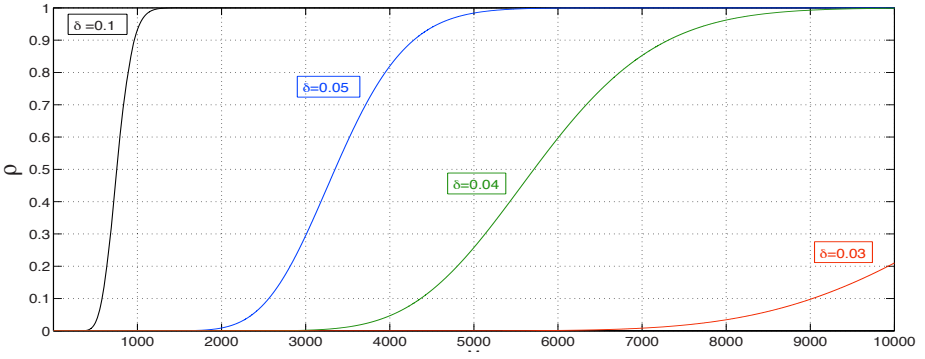


Fig. 2. Plot of ρ vs M for different values of δ with $K = 25, \alpha^2 = 0.5\sigma^2 \beta^2 = 3.0\sigma^2$

clusters with the smallest distance. The way the distance between clusters is computed produces different flavours of the algorithm, namely the simple linkage, complete linkage and average linkage [10].

We propose a variation that fits very well to our problem, called *J-linkage* (see Algorithm 1). First the preference set of a cluster is computed as the *intersection* of the preference sets of its points. Then the distance between two elements (point or cluster) is computed as the *Jaccard distance* between the respective preference sets.

Definition 1 (Jaccard distance). *Given two sets A and B, the Jaccard distance is*

$$d_{\mathrm{J}}(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}.$$

The Jaccard distance measures the degree of overlap of the two sets and ranges from 0 (identical sets) to 1 (disjoint sets).

The cut-off value is set to 1, which means that the algorithm will only link together elements whose preference sets overlap. Please note that the cut-off distance is not data dependent, but defines a qualitative behaviour of the J-linkage algorithm. Indeed, as a result, clusters of points have the following properties:

- for each cluster there exist at least one models that is in the PS of all the points (i.e., a model that fits all the points of the cluster)
- one model cannot be in the PS of *all* the points of two distinct clusters (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. The final model for each cluster of points is estimated by least squares fitting.

Outliers emerge as small clusters. Depending on the application, one may set different rejection thresholds. If the percentage of outliers is known or can be estimated (as it is assumed in RANSAC), one may reject all the smallest clusters up to the number of outliers.

Algorithm 1. J-LINKAGE

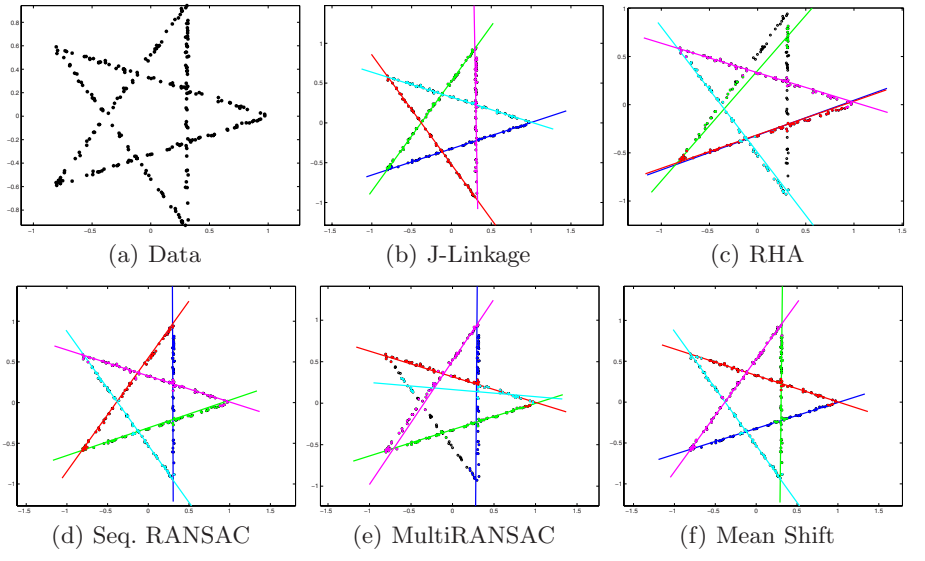
Input: the set of data points, each point represented by its preference set (PS)
Output: clusters of points belonging to the same model

1. Put each point in its own cluster.
2. Define the PS of a cluster as the *intersection* of the PSs of its points.
3. Among all current clusters, pick the two clusters with the smallest Jaccard distance between the respective PSs.
4. Replace these two clusters with the union of the two original ones.
5. Repeat from step 3 while the smallest Jaccard distance is lower than 1.

3 Experiments

We performed comparative experiments with sequential RANSAC, multi RANSAC, residual histogram analysis [7] (henceforth RHA) and mean-shift (MS). In all the experiments each model consists of 50 inliers, corrupted by variable Gaussian noise and variable outliers percentage. The data sets consist of segments in several configuration: star (*star5* and *star11*), circles (*circle5*), and horizontal (*stair4*). The latter was used also in [2].

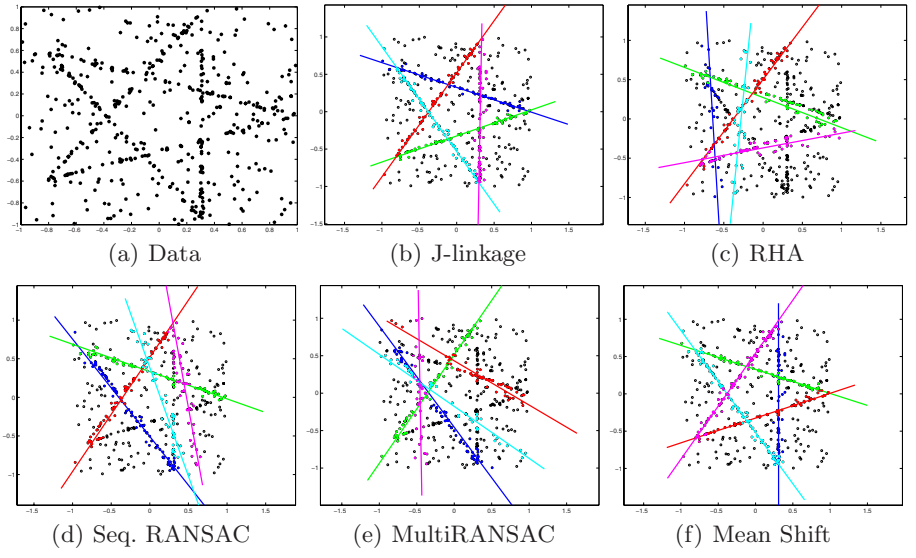
All the methods being compared are based on random sampling, hence we used the same sampling strategy (Eq. 1) and number of samples (5000) in all the experiments. The scale parameter σ in the sampling strategy is 0.2 in all the experiments but *stair4*, where it has been set to 0.05. The inlier threshold ε – variable with the noise level – was the same for sequential RANSAC, multi-RANSAC and our method. The parameters needed by MS (bandwidth) and by



(a) Data (b) J-Linkage (c) RHA

(d) Seq. RANSAC (e) MultiRANSAC (f) Mean Shift

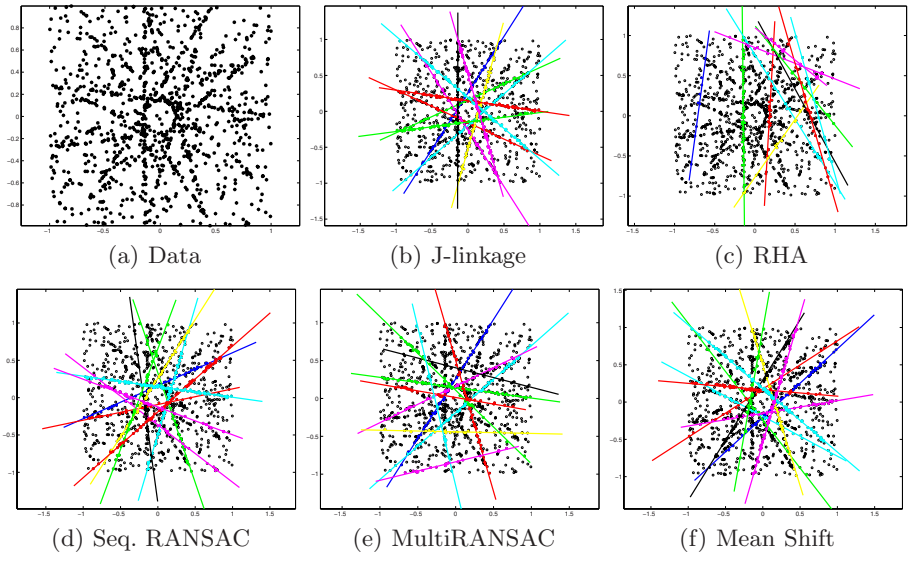
Fig. 3. *star5* set perturbed with Gaussian noise ($\sigma_n = 0.0075$) and no outliers



(a) Data (b) J-linkage (c) RHA

(d) Seq. RANSAC (e) MultiRANSAC (f) Mean Shift

Fig. 4. *star5* set perturbed with Gaussian noise ($\sigma_n = 0.0075$) and 50% outliers



(a) Data (b) J-linkage (c) RHA

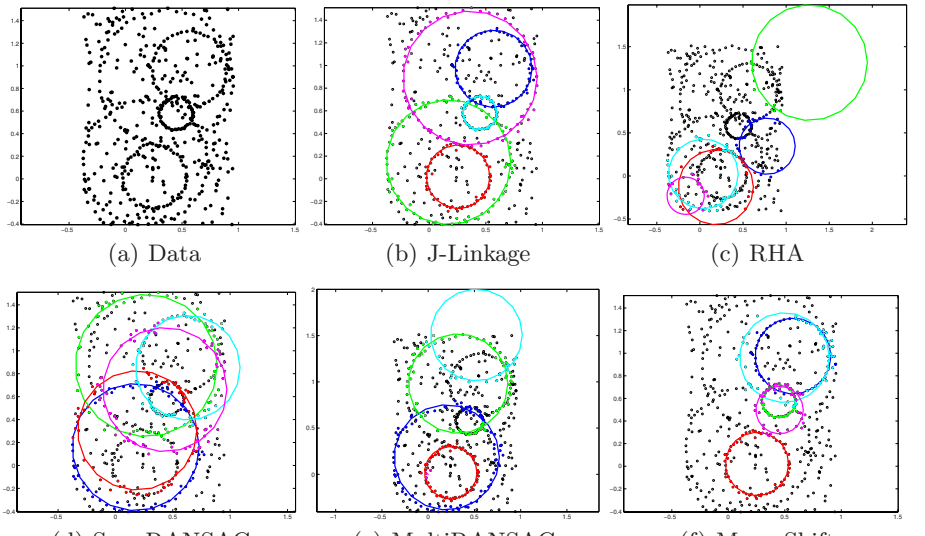
(d) Seq. RANSAC (e) MultiRANSAC (f) Mean Shift

Fig. 5. *star11* set perturbed with Gaussian noise ($\sigma_n = 0.0075$) and 50% outliers

RHA have been tuned manually to optimize performances. The best outcome out of several trials have been recorded. As multiRANSAC requires prior specification of the number n of models, for the sake of fairness we used the same information also with the other algorithms: only the best or strongest n models among the ones produced by the algorithm were considered. For example, with

J-linkage we retained the n largest clusters, and the same for MS. In RHA we sought the n strongest modes.

The results on synthetic data are reported in Figures 3, 4, 5, 6, 7. In summary, we see that MS on the *star5* data set always produces the correct result, whereas all the other methods break when the outlier percentage grows to 50%. If the number of models increases (*star11* data set), however, only J-linkage produces the correct result, at the same level of noise and outliers. Also on the *circle5* data set, with random noise and 50% outliers, J-linkage is the only one that works correctly. On the *stair4* data set with and 60% outliers both multiRANSAC and J-linkage yield the correct result.



(a) Data (b) J-Linkage (c) RHA

(d) Seq. RANSAC (e) MultiRANSAC (f) Mean Shift

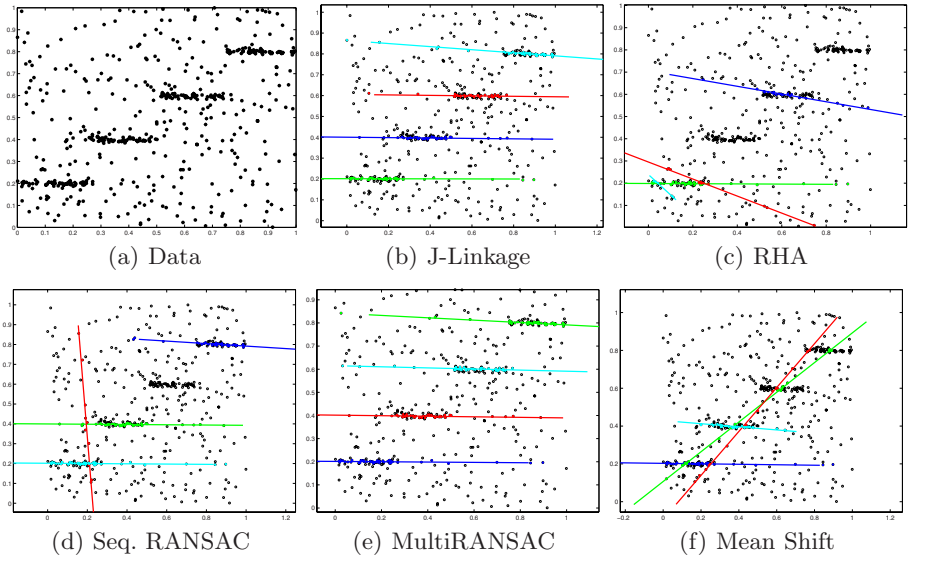
Fig. 6. *circle5* set perturbed with Gaussian noise ($\sigma_n = 0.0075$) and 50% outliers

We noted that multiRANSAC systematically fails when the models intersect each other (the problem is more evident when the models are cyclically intersecting). The reason is that the greedy approach adopted by multiRANSAC tries to maximize the number of total inliers, implicitly assuming non intersecting models.

As an example of a real case, Fig. 8 shows the results of fitting planes to a cloud of 3D points. They were produced by a Structure and Motion pipeline [11] fed with a set of images of the church of Pozzoveggiani (Italy). Despite the fact that gross outliers are absent, pseudo-outliers and the uneven distribution of points among the models (ranging from 9 to 1692) challenges any model fitting algorithm. Indeed, our method is the only one that produces the correct result. To appraise further the result of J-linkage, Fig. 8 show two images of Pozzoveggiani church with points marked according to the plane they belong to.

The MATLAB code is available for download from the web[2].

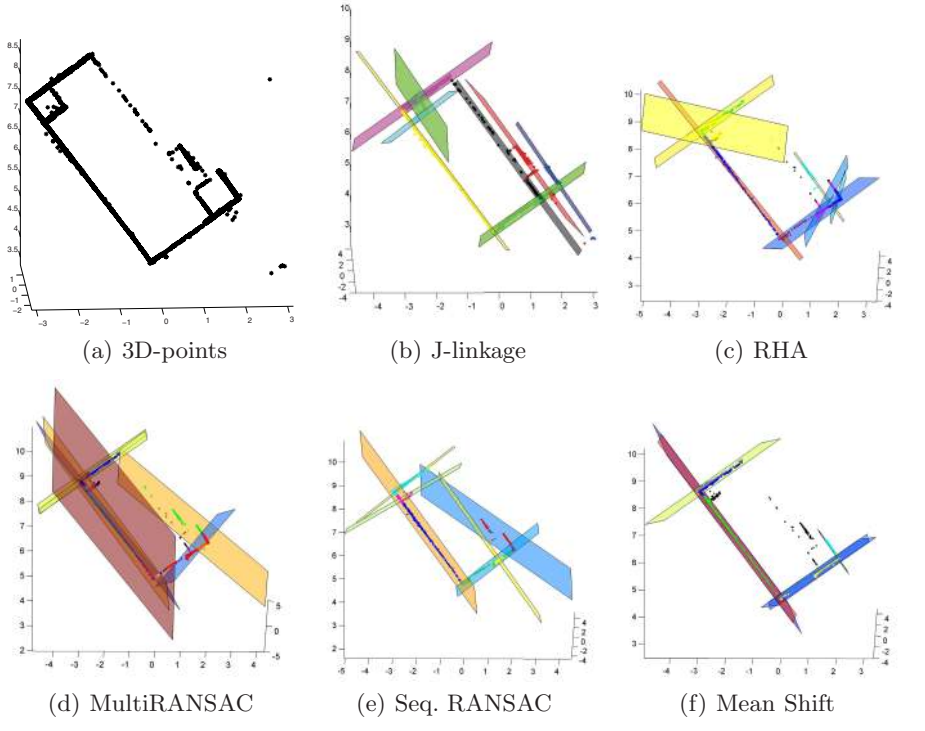
[2] http://profs.sci.univr.it/~fusiello/demo/jlk/



(a) Data (b) J-Linkage (c) RHA

(d) Seq. RANSAC (e) MultiRANSAC (f) Mean Shift

Fig. 7. *stair4* set perturbed with Gaussian noise ($\sigma_n = 0.0075$) and 60% outliers



(a) 3D-points (b) J-linkage (c) RHA

(d) MultiRANSAC (e) Seq. RANSAC (f) Mean Shift

Fig. 8. "Pozzoveggiani" dataset. 3D planes extracted by J-linkage and other algorithms viewed from the zenith.

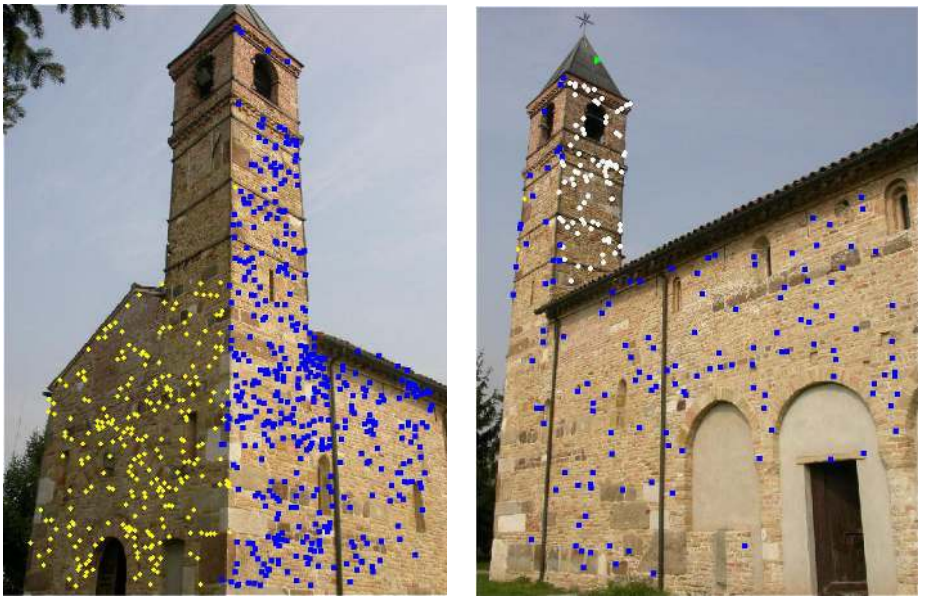


Fig. 9. Two images of the "Pozzoveggiani" dataset. Points belonging to different planes according to J-linkage are plotted with different markers.

4 Discussion

The experiments showed that J-linkage compares favorably with other state-of-the-art methods. Our insight about the reason for this is that it derives the best features from the competing algorithms. Like RANSAC it is robust because only the points below the outlier threshold matters. Compared to RHA, it consider only the the first bin of the residuals histogram, while RHA takes the whole the histogram into account, which is corrupted by outliers. Like RHA, however, our conceptual representation casts the problem from the perspective of the data point, which is a strength of RHA.

The discovery of multiple models is devolved to clustering, thereby gaining a global view of the problem, whereas sequential RANSAC and multiRANSAC are forced to make local, greedy choices. Clustering, however, is not an intrinsically robust technique, in general. J-linkage, instead, is inherently robust, because it is based on the intersection of preference sets, hence it favours the growing of clusters made up of inliers only. On the contrary, modes of the residual histogram or of the parameters distribution are difficult to locate, especially when the number of gross outliers and pseudo-outliers grows, and the fraction of outlier-free sets generated by the random sampling decreases accordingly.

5 Conclusions

We described a novel method for fitting multiple instances of a model to data corrupted by noise and outliers. Each point is represented with the characteristic function of its preference set and multiple models are revealed as clusters in this conceptual space. A new agglomerative clustering algorithm, called J-linkage, have been specifically devised. The method does not require prior specification of the number of models, nor it necessitate manual parameters tuning. The only free parameter is the consensus threshold, as in RANSAC. Our method demonstrated its effectiveness in comparison with state-of-the-art competing algorithms.

Acknowledgements. This research was supported by the Italian PRIN 2006 project 3-SHIRT.

References

1. Stewart, C.V.: Bias in robust estimation caused by discontinuities and multiple structures. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(8), 818–833 (1997)
2. Zuliani, M., Kenney, C.S., Manjunath, B.S.: The multiRANSAC algorithm and its application to detect planar homographies. In: Proceedings of the IEEE International Conference on Image Processing, Genova, IT, September 11-14 (2005)
3. Xu, L., Oja, E., Kultanen, P.: A new curve detection method: randomized Hough transform (RHT). Pattern Recognition Letters 11(5), 331–338 (1990)
4. Subbarao, R., Meer, P.: Nonlinear mean shift for clustering over analytic manifolds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, York, USA, pp. 1168–1175 (2006)
5. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 603–619 (2002)
6. Vidal, R., Ma, Y., Sastry, S.: Generalized principal component analysis (gpca). IEEE Transactions on Pattern Analysis and Machine Intelligence 27(12), 1945–1959 (2005)
7. Zhang, W., Kosecká, J.: Nonparametric estimation of multiple structures with outliers. In: Vidal, R., Heyden, A., Ma, Y. (eds.) WDV 2006. LNCS, vol. 4358, pp. 60–74. Springer, Heidelberg (2006)
8. Duin, R., Pekalska, E., Paclik, P., Tax, D.: The dissimilarity representation, a basis for domain based pattern recognition? In: Goldfarb, L. (ed.) Pattern representation and the future of pattern recognition, ICPR 2004 Workshop Proceedings, Cambridge, UK, pp. 43–56 (2004)
9. Kanazawa, Y., Kawakami, H.: Detection of planar regions with uncalibrated stereo using distributions of feature points. In: British Machine Vision Conference, pp. 247–256 (2004)
10. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis, pp. 98–105. John Wiley and Sons, Chichester (1973)
11. Farenzena, M., Fusiello, A., Gherardi, R.: Efficient Visualization of Architectural Models from a Structure and Motion Pipeline. In: Eurographics 2008 - Short Papers, Crete, Greece, Eurographics Association, pp. 91–94 (2008)