Geometrical Consistent Clustering of Linear Subspaces

Nuno Pinho da Silva João P. Costeira
Institute for Systems and Robotics – Instituto Superior Técnico
Av. Rovisco Pais, 1049–Lisboa, Portugal
{nmps,jpc}@isr.ist.utl.pt

Abstract

The perception of rigid-bodies from affine views of moving $3D$ point clouds, boils down to clustering the rigid motion subspaces supported by the image trajectories. For a physically meaningful interpretation, clusters must be consistent with the geometry of the underlying subspaces. We find that proper subspace clustering requires invariance both to the orthogonal and the inclusion relationship between subspaces. Most of the existing measures for subspace comparison do not comply with this observation. A practical consequence is that methods based on such (dis)similarities are unstable when the number of rigid bodies increase. This paper introduces the Normalized Subspace Inclusion (NSI*) criterion to resolve these issues. Combining it with a robust segmentation method, we propose a robust methodology for rigid motion segmentation, and test it, extensively, on the Hopkins155 database. The geometric consistency of the* NSI *assures the method's accuracy when the number of rigid bodies increases, while robustness proves to be suitable for dealing with challenging imaging conditions.*

1 Introduction

Extending the structure from motion framework from one rigid object to multiple moving objects appearing in the field of view (Costeira and Kanade [1],Yan and Pollefeys [8]), requires the primary task of identifying the rigid bodies in the scene, wether they are independent rigid objects, or rigid parts of articulated objects (Fig. 1).

In the finite sample scenario, rigid bodies are clouds of $3D$ points moving rigidly. Assuming affine projections and given their correspondences along the sequence, segmenting the rigid motions is framed as the robust clustering of their imaged trajectories (Fig. 2). The clustering relies on subspace comparison, because the $2D$ trajectories of rigid motion support linear subspaces.

(a) Articulated object (b) Multiple rigid objects

Figure 1. Rigid-body segmentation: our result of segmenting (a) the articulated and (b) cars10 sequences, from the Hopkins155 database (Tron and Vidal [6]) (red $*$ are points classified outliers). Section 3 presents results over the entire data set.

Extensive validation on the Hopkins155 database (Tron and Vidal [6]) shows that our approach can leverage the segmentation results given by state-of-the-art methods (*e.g.* , ALC from Rao *et al*. [3] and LSA from Yan and Pollefeys [8]), particularly in challenging settings presenting more than two rigid bodies and outdoor scenes.

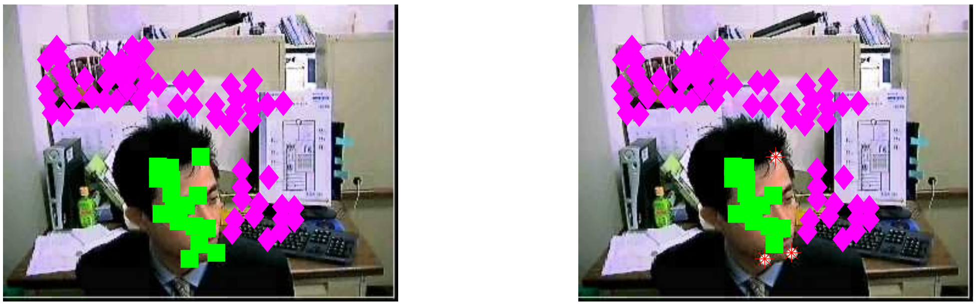
2 Robust Subspace Clustering

A key issue in any clustering algorithm is to determine the groups' (dis)similarity. We find that proper subspace clustering requires invariance both to the orthogonal and the inclusion relationship between subspaces. Most of the existing measures for subspace comparison do not comply with this observation, thus being inappropriate for a unified treatment of the problem and unstable when the number of motions increase.

For example, let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two motion subspaces: (i) the least principal angle criterion [8] cannot distinguish their intersection $\mathcal{L}_1 \cap \mathcal{L}_2 \neq \emptyset$ from their inclusion $\mathcal{L}_1 \subseteq \mathcal{L}_2$; (ii) the subspace distance, developed by Sun *et al*. [5]

This work was partially founded by FCT (ISR – IST plurianual funding) through the POS_Conhecimento Program that includes FEDER funds and grant PTDC/EEA-ACR/72201/2006, "MODI - 3D Models from 2D Images". Nuno Pinho da Silva is supported by FCT, under the grant SFRH/BD/24879/2005.

(a) The first, second and last frame from the Kanatani3 sequence (Hopkins155 database) with the points classified as outliers superimposed.



(b) MSL and ALC

(c) our result

Figure 2. Outlier detection: (a) Feature points in the chin are outliers: their image positions slide along the sequence in a non-rigid way. (b) The segmentation given by the MSL (Sugaya and Kanatani [4]) and ALC (Rao *et al*. [3]) fails to detect them. Our approach (c) recognizes the outlying trajectories, though admitting one false negative (the forehead point).

and Wang *et al*. [7], is inconsistent with their inclusion, *i.e.* it does not reflect that all features supporting $\mathcal{L}_1$ also support $\mathcal{L}_2$; or (iii) with features from a third independent object supporting $\mathcal{L}_3$, if $\mathcal{L}_1$ and $\mathcal{L}_2$ intersect, each is orthogonal to $\mathcal{L}_3$, and the sum of the squared sines of the principal angles criterion [8] is unable to recognize it. Simple examples may be constructed proving these facts.

A geometric consistent criterion cannot be a distance function, because it violates the identity of the indiscernibles (*i.e.* , any distance function D must satisfy $D(x, y) = 0 \Leftrightarrow x = y$) by being consistent with the inclusion of subspaces: if $\mathcal{L}_1 \subseteq \mathcal{L}_2$, where $\mathcal{L}_1$ and $\mathcal{L}_2$ are motion subspaces, the criterion must reflect that all features supporting $\mathcal{L}_1$ also support $\mathcal{L}_2$, as the trajectories of the points on the $\mathcal{L}_1$ lines also lie on the $\mathcal{L}_2$ boxes in Fig. 1(a).

Our main contribution is the *normalized subspace inclusion* (NSI), a criterion for subspace clustering consistent with the geometry of the underlying subspaces. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be linear subspaces of $\mathbb{R}^n$, such that $dim(\mathcal{L}_1) = d_1$ and $dim(\mathcal{L}_2) = d_2$. Define the NSI as

$$\mathrm{NSI}(\mathcal{L}_1, \mathcal{L}_2) = \frac{tr\{\mathbf{U_1^T U_2 U_2^T U_1}\}}{min(d_1, d_2)}, \quad (1)$$

where $tr\{\cdot\}$ is the trace function and, $\mathbf{U_1}$ and $\mathbf{U_2}$ are orthonormal bases for $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively.

3 Robust Rigid Motion Segmentation

We perceive rigid bodies by robustly segmenting the observations into geometrically meaningful clusters (using the Grassmannian Maximum Consensus [2]), and agglomerate them, under the NSI criterion, thus providing the adequate interpretation (segmentation), given the total number of motions.

Tables 1–3 compares the results of the state-of-the-art algorithms LSA (Yan and Pollefeys [8]) and ALC (Rao *et al*. [3]) with our approach.

Table 1 presents the results by class. Our approach achieves the highest accuracy in the traffic class, which contains sequences taken by handheld camera, often with degenerate motions [6], pointing out its robustness to real-world imaging conditions (and the GMC correct dimension estimation, since the NSI criterion depends on this information). Also, the misclassification rate decreases as the number of motions increases, being our approach the most accurate for all classes with $M = 3$ motions, thus showing the NSI's immunity to the higher complexity of the agglom-

Method	avg [%]	std [%]
LSA	4.86	10.29
ALC	3.37	7.97
our result	3.44	7.34

(a) 155 (all) sequences

Method	avg [%]	std [%]
LSA	9.71	14.71
ALC	6.69	11.48
our result	2.87	5.28

(b) 35 sequences with $M = 3$ rigid motions

Method	avg [%]	std [%]
LSA	3.45	8.14
ALC	2.40	6.35
our result	3.61	7.84

(c) 120 sequences with $M = 2$ rigid motions

Table 2. Average (avg) and standard deviation (std) of misclassification rates for all classes of sequences.

method	checkboard avg [%] | std [%]	articulated avg [%] | std [%]	traffic avg [%] | std [%]
LSA	3.35|8.06	4.58|6.59	9.09|14.86
ALC	2.37|6.18	12.30|16.50	3.06|6.21
our result	3.54|7.39	7.79|8.22	1.69|6.33

(a) 155 (all) sequences

method	checkboard avg [%] | std [%]	articulated avg [%] | std [%]	traffic avg [%] | std [%]
LSA	5.70|10.89	7.25|9.30	25.30|19.05
ALC	5.00|9.14	21.08|28.87	8.86|13.16
our result	2.92|5.73	6.38|9.03	1.67|1.51

(b) 35 sequences with $M = 3$ rigid motions

method	checkboard avg [%] | std [%]	articulated avg [%] | std [%]	traffic avg [%] | std [%]
LSA	2.57|6.79	4.10|6.47	5.43|11.17
ALC	1.49|4.58	10.70|15.00	1.75|1.83
our result	3.75|7.89	8.05|8.51	1.69|7.00

(c) 120 sequences with $M = 2$ rigid motions

Table 1. Average (avg) and standard deviation (std) of misclassification rates by class of sequences.

Method	checkboard	articulated	traffic	all
LSA	5.13	1.93	3.96	4.58
ALC	1213.55 (∼20m)	558.36 (∼9m)	962.52 (∼16m)	1097.06 (∼18m)
our result	14.75	4.16	11.02	12.85

Table 3. Computational burden (average cpu time [s]). These times are essentially indicative (order of magnitude), because they depend on particular implementations. All code was written in $\mathrm{Matlab}^{\circledR}$.

erative task.

This is confirmed by table 2, where it can be seen that our approach achieves an average misclassification rate 1.42% better than the LSA, but 0.07% higher than the ALC, while being slightly more stable. Also, the difference between all methods is much higher for sequences with 3 groups than for sequences with 2 groups (the vast majority of the sequences in the database). Note that for sequences with $M = 3$ rigid motions, lower error rates are expected if the number of misclassified points remains fairly the same, because the total of features is often higher in the $M = 3$ group sequences than in sequences with $M = 2$ groups, since most of the latter were constructed by splitting each 3 motion sequences into its respective clusters.

In table 3, note that our segmentation is, on average, only 8 seconds (3 times) slower than the LSA and 85 times faster than the ALC, balancing accuracy with computational burden.

References

[1] J. P. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision*, 29(3):159–179, 1998.

[2] N. Pinho da Silva and J. P. Costeira. Subspace segmentation with outliers: a grassmannian approach to the maximum consensus subspace. *in IEEE CVPR*, 2008.

[3] S. R. Rao, R. Tron, R. Vidal, and Y. Ma. Motion segmentation via robust subspace separation in the presence of outlying, imcomplete, or corrupted trajectories. *in IEEE CVPR*, 2008.

[4] Y. Sugaya and K. Kanatani. Geometric structure of degeneracy for multi-body motion segmentation. *SMVP*, LNCS 3247:13–25, 2004. D. Comaniciu *et al*. (Eds.).

[5] X. Sun, L. Wang, and J. Feng. Further results on the subspace distance. *Pattern Recognition*, 40(1):328–329, 2007.

[6] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. *in IEEE CVPR*, 2007.

[7] L. Wang, X. Wang, and J. Feng. Subspace distance analysis with application to adaptive bayesian algorithm for face recognition. *Pattern Recognition*, 39(3):456–464, 2006.

[8] J. Yan and M. Pollefeys. A factorization-based approach for articulated nonrigid shape, motion, and kinematic chain recovery from video. *IEEE Trans. on PAMI*, 30(5):865–877, 2008.