# The Normalized Subspace Inclusion: Robust Clustering of Motion Subspaces

Nuno Pinho da Silva João Paulo Costeira
ISR – Instituto Superior Técnico
Torre Norte - 7 piso, Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal
(nmps,jpc)@isr.ist.utl.pt

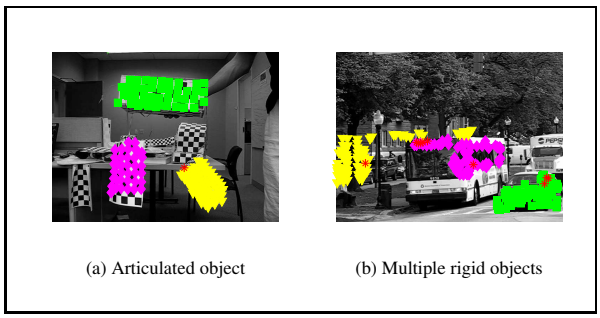
## Abstract

*Perceiving dynamic scenes of rigid bodies, through affine projections of moving $3D$ point clouds, boils down to clustering the rigid motion subspaces supported by the points' image trajectories. For a physically meaningful interpretation, clusters must be consistent with the geometry of the underlying subspaces. Most of the existing measures for subspace clustering are ambiguous, or geometrically inconsistent. 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. Relying on this similarity, 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 (Tomasi and Kanade [10]) to multiple moving objects appearing in the field of view (Costeira and Kanade [1], Ozden *et al*. [5], Yan and Pollefeys [17]), 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. 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 [11]) (red $*$ are points classified outliers). Section 4 presents results over the entire data set.

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. In section 3 we provide simple examples proving these facts. In particular, we show that a geometric consistent criterion cannot be a distance function, because it violates the identity of the indiscernibles[1] by being consistent with the inclusion of subspaces (*i.e*. 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)).

The main contribution of this paper is the *normalized subspace inclusion* (NSI), a criterion for subspace clustering consistent with the geometry of the underlying subspaces. Our approach for (rigid) motion segmentation relies on splitting the observations into geometrically meaningful clusters, and agglomerating them, under the NSI criterion, to provide the adequate interpretation (segmentation), given the total number of motions (see Figure 3).

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.

[1] We call the identity of indiscernibles to the following metric axiom that any distance function $D$ must satisfy: $D(x, y) = 0 \Leftrightarrow x = y$.



2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

Extensive validation on the Hopkins155 database (Tron and Vidal [11]) shows that our approach can leverage the segmentation results given by state-of-the-art methods (*e.g.*, ALC from Rao *et al.* [7] and LSA from Yan and Pollefeys [17]), particularly in challenging settings presenting more than two rigid bodies and outdoor scenes.

## 2. Motion Subspaces

Consider a rigid body moving relative to the camera and let $N_j$ be the number of $3D$ points $\mathbf{x}_{j_i}$ tracked over $F$ frames (the index $j$ represents the $jth$ rigid body in a multi-body scene). Assuming affine projections, the data matrix $\mathbf{W}_j \in \mathbb{R}^{2F\times N_j}$, collecting the images of the point trajectories, is written as

$$\mathbf{W}_j = [\mathbf{w}_{j_1} \dots \mathbf{w}_{j_{N_j}}],\ j = 1, \dots, M \tag{1}$$

$$\mathbf{w}_{j_i} = [u^1_{j_i}\ v^1_{j_i} \dots u^F_{j_i}\ v^F_{j_i}]^T,\ i = 1, \dots, N_j \tag{2}$$

$$[u^f_{j_i}\ v^f_{j_i}]^T = \mathbf{M}^f_j \left[\mathbf{x}^T_{j_i}\ 1\right]^T,\ f = 1, \dots, F \tag{3}$$

$$\mathbf{M}^f_j = \mathbf{K}_{2\times 4} \begin{bmatrix} \mathbf{R}^f_j & \mathbf{t}^f_j \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{4}$$

where $\mathbf{R}^f_j$ is the rotation, and $\mathbf{t}^f_j$ the translation, of the $jth$ rigid body's coordinate frame relative to the camera at time (image) $f$, and $\mathbf{K}_{2\times 4}$ collects the affine model's intrinsic parameters. Given (3), $\mathbf{W}_j$ can be decomposed into a motion matrix $\mathbf{M}_j \in \mathbb{R}^{2F\times 4}$ and structure matrix $\mathbf{S}_j \in \mathbb{R}^{4\times N_j}$, as shown by Tomasi and Kanade [10], *i.e.*

$$\mathbf{W}_j = \mathbf{M}_j\mathbf{S}_j = \begin{bmatrix} \mathbf{M}^1_j \\ \vdots \\ \mathbf{M}^F_j \end{bmatrix} \begin{bmatrix} \mathbf{x}_{j_i} & \dots & \mathbf{x}_{j_{N_j}} \\ 1 & \dots & 1 \end{bmatrix}. \tag{5}$$

Since the column subspace of $\mathbf{W}_j$ is generated by the motion matrix $\mathbf{M}_j$, it is a (rigid) *motion subspace*. From (5), the measurement matrix $\mathbf{W}_j$ is at most rank 4, and so is the dimension of the rigid motion subspace.

### 2.1. Problem Formulation

When a scene contains multiple objects (rigid or articulated), we built the data matrix $\mathbf{W} \in \mathbb{R}^{2F\times N}$ by tracking the trajectories of $M$ rigid bodies with different motions relative to the camera. Here $N = N_0 + \dots + N_M$ is the total number of point trajectories, and $N_0$ is the number of outlying trajectories, *i.e.* mismatches or nonrigid moving points. Hence, the problem of (rigid) motion segmentation is to identify the trajectories of each rigid body, as well as the outlying trajectories (Fig. 2).

Formally, given a data matrix $\mathbf{W}$, construct its canonical (segmented) form, *i.e.*

$$\mathbf{W} \quad = \quad [\mathbf{W}_0\mathbf{W}_1 \dots \mathbf{W}_M], \tag{6}$$

where $\mathbf{W}_0$ assembles an arbitrary number of outliers, $\mathbf{W}_j = \mathbf{M}_j\mathbf{S}_j$ ($j \neq 0$), and the subspaces generated by each rigid body may intersect arbitrarily.

From (5), identifying rigid bodies is tantamount to segmenting the subspaces supported by their $2D$ trajectories. However, the basis of the motion subspaces is given by stacking the affine projections of the special euclidean group (4), and segmenting, while imposing the metric (rotation and translation) constraints, leads to the cyclic dilemma: to check the constraints it is necessary to segment features and to segment it is necessary to compute the constraints.

A useful alternative is to impose the number $M$ of motions extrinsically, as shown in the works of Sugaya and Kanatani [8], Vidal *et al.* [12], Yan and Pollefeys [17], though this will not, *per se*, impose physical coherence. To attain it, the $M$ clusters must respect the geometry of the underlying motion subspaces.

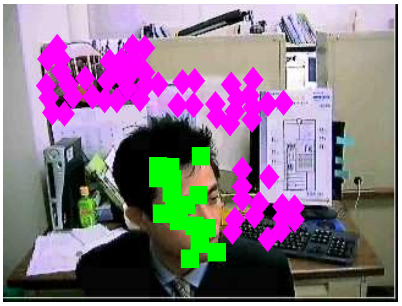
### 2.2. Related Work

Under affine projections, independent motions, and no outliers, the segmentation is obtained by the block diagonal structure of the shape interaction matrix (Costeira and Kanade [1]). Unfortunately, null cross-correlation does not hold for partially dependent motions, motivating approaches such as the Sugaya and Kanatani [8] Multi-Stage Learning (MSL), the Vidal *et al.* [12] Generalized Principal Component Analysis (GPCA) or the Yan and Pollefeys [17] Local Subspace Affinity (LSA).

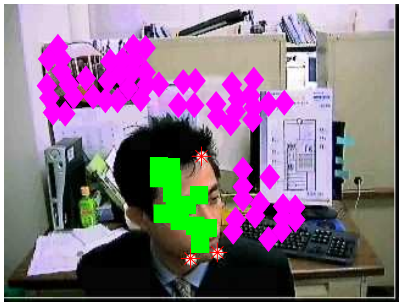
The results on the Hopkins155 database (Tron and Vidal [11]) showed that the LSA was the most accurate of these methods. However, as we show in section 3, the sum of the squared sines of the principal angles, *i.e.* the criterion underlying the affinity in the LSA method, may provide no insight into the geometric relation between the segments. Also, all previous methods rely on the global $rank$ detection of a data matrix, which is prone to errors in the presence of noise and outliers. Therefore, some additional procedure, *e.g.* RANSAC must be considered to deal with outliers. Furthermore, recent results (Rao *et al.* [7]) suggested that robustness improves the segmentation.

Recent approaches such as the MDPO (Jian and Chen [3]), the GMC (Silva and Costeira [6]) and the ALC (Rao *et al.* [7]) are intrinsically robust. The MDPO is a two-view approach, combining mixtures of Dirichlet process with RANSAC, and has the advantage of being independent of the number of clusters. However, its application is limited to independent motions in a calibrated setup. The GMC is a divisive method, using recursion over the space of the dimensions to comply with degeneracies, and exploring the riemannian structure of the Grassmann manifold to find the maximum consensus subspace. The ALC follows an agglomerative strategy, minimizing the segmentation's

(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 ina non-rigid way. (b) The segmentation given by the MSL (Sugaya and Kanatani [8]) and ALC (Rao *et al*. [7]) fails to detect them. Our approach (c) recognizes the outlying trajectories, though admitting one false negative (the forehead point).

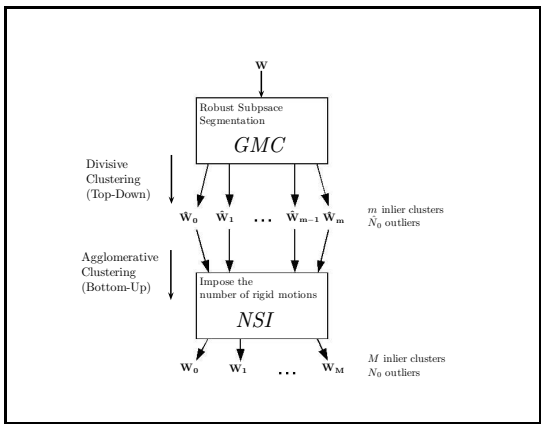



Figure 3. Two step approach: imposing the number of subspaces provides semantical meaning for the geometric segmentation, since the latter is a relaxed solution of the motion segmentation problem.

coding length over a range of distortions. It is often accurate, but very time consuming and may lose robustness during the voting process (Fig. 2).

Fig. 3 summarizes our approach. The GMC provides the initial grouping and the corresponding dimensions of the underlying subspaces. These clusters are pair-wise hierarchically agglomerated under the NSI criterion, thus finding the interpretation according to the number of motions imposed.

## 3. Normalized Subspace Inclusion (NSI)

Physically meaningful subspace clustering must be geometrically correct. We will use examples to show that most of the subspace clustering criteria are ambiguous and result into inconsistent clusters, *i.e*. clusters disrespecting the geometric relation between the underlying subspaces. The reason is that they are invariant only to the orthogonal or to the inclusion relationship between subspaces, but not to both.

Following this observation, we propose the *normalized subspace inclusion* (NSI) criterion. 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

$$\text{NSI}(\mathcal{L}_1, \mathcal{L}_2) = \frac{tr\{\mathbf{U_1^T U_2 U_2^T U_1}\}}{min(d_1, d_2)}, \qquad (7)$$

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.

The NSI measures the similarity between two subspaces by quantifying "how much a subspace belongs to another". Formally, it generalizes the $\cos^2$ between vectors by considering all principal directions between two subspaces.

### 3.1. Subspace Clustering

A key issue in any clustering algorithm is to determine the groups' (dis)similarity. The following criteria are often employed when clustering subspaces[2]:

- *LPA*: The Least Principal Angle is a dissimilarity criterion, used by Yan and Pollefeys [17] to build kinematic chains; it is defined as

$$LPA(\mathcal{L}_1, \mathcal{L}_2) = \cos^{-1}\left(\sqrt{\sigma_{max}\left(\mathbf{U_1^T U_2}\right)}\right) \quad (8)$$

- *SD*: The Subspace Distance

$$SD(\mathcal{L}_1, \mathcal{L}_2) = \sqrt{\max(d_1, d_2) - tr\{\mathbf{U_1^T U_2 U_2^T U_1}\}} \quad (9)$$

verifies all distance axioms, and is applied in face recognition (Sun *et al*. [9] and Wang *et al*. [13]).

- *SCPA*: Considering the Sum of the $\cos^2$ of the Principal Angles

$$SCPA(\mathcal{L}_1, \mathcal{L}_2) = tr\{\mathbf{U_1^T U_2 U_2^T U_1}\}. \quad (10)$$

will also show the importance of the normalization in the NSI criterion (7).

- *SSPA*: The Sum of the $\sin^2$ of the Principal Angles is the dissimilarity employed by Yan and Pollefeys [17] in their affinity $e^{-SSPA(\mathcal{L}_1,\mathcal{L}_2)}$; it can be written as

$$SSPA(\mathcal{L}_1, \mathcal{L}_2) = \min(d_1, d_2) - SCPA(\mathcal{L}_1, \mathcal{L}_2). \quad (11)$$

The following examples motivate the usefulness of the NSI criterion[3]. Practical situations include degenerated structures clustered into more general rigid bodies, such as lines in boxes (Fig. 1 (a)), or interpreting scenes containing both independent and partially dependent motions.

**Example 1** (*LPA*) *Let* $\mathcal{L}_1 = span\{[0\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}]^T\}$ *and*

$$\mathcal{XZ} = span\left\{\begin{bmatrix}1&0\\0&0\\0&1\end{bmatrix}\right\},\ \mathcal{YZ} = span\left\{\begin{bmatrix}0&0\\1&0\\0&1\end{bmatrix}\right\},$$

*as in Fig. 4. Since* $\mathcal{L}_1 \subset \mathcal{YZ}$ *and* $\mathcal{YZ} \cap \mathcal{XZ} \neq \emptyset$*, we have*

$$LPA(\mathcal{L}_1, \mathcal{YZ}) = 0, \quad LPA(\mathcal{XZ}, \mathcal{YZ}) = 0, \quad (12)$$

[2]We do not consider the largest principal angle, defined as the matrix 2-norm of the difference between the orthogonal projectors, because it is related to the notion of distance between *equidimensional* subspaces [2], thus being of restricted application.

[3]In all cases we are given $m = 3$ subspaces and have to cluster them into $M = 2$ linear manifolds.

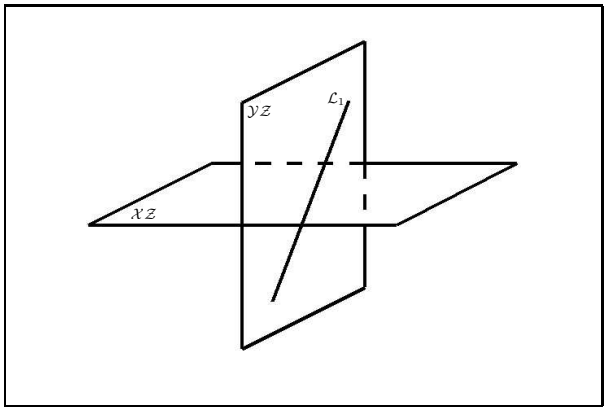



Figure 4. The subspace union of Example 1.

*disregarding that all points lying in* $\mathcal{L}_1$ *also support* $\mathcal{YX}$*, but not all points in* $\mathcal{YX}$ *belong to* $\mathcal{XZ}$*. We may use this example to show that the sum of the canonical correlations (Kim* et al. *[4]) is not inclusion consistent, because it is only invariant to the orthogonality between subspaces.*

**Example 2** (*SCPA* **and** *SD*) *Consider the following subspaces of* $\mathbb{R}^4$*:* $\mathcal{L}_1 = span\{[0\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}0]^T\}$*, and*

$$\mathcal{XYZ}{=}span\left\{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\end{bmatrix}\right\}, \mathcal{ZW}{=}span\left\{\begin{bmatrix}0&0\\0&0\\1&0\\0&1\end{bmatrix}\right\}.$$

*From (10),*

$$SCPA(\mathcal{L}_1, \mathcal{XYZ}) = 1, \quad SCPA(\mathcal{ZW}, \mathcal{XYZ}) = 1, \quad (13)$$

*showing that, without the* NSI*'s normalization (7), we have the ambiguity from the previous example. Also, applying the Subspace Distance (9),*

$$SD(\mathcal{L}_1, \mathcal{XYZ}) = \sqrt{2}, \quad SD(\mathcal{L}_1, \mathcal{ZW}) = \sqrt{\tfrac{3}{2}}, \quad (14)$$

*implies*

$$SD(\mathcal{L}_1, \mathcal{XYZ}) > SD(\mathcal{L}_1, \mathcal{ZW}), \quad (15)$$

*which is inconsistent with the fact that all points in* $\mathcal{L}_1$ *have* zero *distance to* $\mathcal{XYZ}$ *and* nonzero *distance to* $\mathcal{ZW}$*.*

**Example 3** (*SSPA*) *Consider* $\mathbb{R}^5$ *and suppose we have to cluster* $\mathcal{L}_V = span\{[0\,0\,0\,0\,1]^T\}$ *and*

$$\mathcal{XYZ}{=}span\left\{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\end{bmatrix}\right\}, \mathcal{ZW}{=}span\left\{\begin{bmatrix}0&0\\0&0\\1&0\\0&1\\0&0\end{bmatrix}\right\}.$$

*into $M = 2$ groups. Note that $\mathcal{L}_V$ is orthogonal to both $\mathcal{XYZ}$ and $\mathcal{ZW}$, whereas the latter intersect. However,*

$$SSPA(\mathcal{L}_V, \mathcal{XYZ}) = 1, \quad SSPA(\mathcal{L}_V, \mathcal{ZW}) = 1 \quad (16)$$

$$SSPA(\mathcal{ZW}, \mathcal{XYZ}) = 1, \quad (17)$$

*which is inconsistent with the orthogonality between $\mathcal{L}_V$ and the other subspaces. In this venue, the product of the squared canonical correlations (Wolf and Shashua [15]) is also orthogonally inconsistent, since a single orthogonal direction is sufficient for nullity.*

**Normalized Subspace Inclusion.** Maximizing the NSI (7) provides the correct clustering for all examples:

- **Example 1**:

$$\text{NSI}(\mathcal{L}_1, \mathcal{YZ}) = 1, \quad \text{NSI}(\mathcal{L}_1, \mathcal{XZ}) = \tfrac{1}{2}. \quad (18)$$

$$\text{NSI}(\mathcal{XZ}, \mathcal{YZ}) = \frac{1}{2}. \quad (19)$$

  So, we group $\mathcal{L}_1$ with $\mathcal{YZ}$. We will show that the NSI is unitary *iif* the subspaces verify $\mathcal{L}_1 \subseteq \mathcal{L}_2$.

- **Example 2**:Applying the NSI, we get the numerical result according to the strength of the geometric relation between the subspaces and, therefore, the correct grouping, *i.e.*

$$\text{NSI}(\mathcal{L}_1, \mathcal{XYZ}) = 1, \quad \text{NSI}(\mathcal{L}_1, \mathcal{ZW}) = \tfrac{1}{2} \quad (20)$$

$$\text{NSI}(\mathcal{ZW}, \mathcal{XYZ}) = \frac{1}{2}. \quad (21)$$

  and the clusters are $\{\mathcal{L}_1, \mathcal{XYZ}\}$ and $\mathcal{ZW}$.

- **Example 3**: The NSI's results are

$$\text{NSI}(\mathcal{L}_V, \mathcal{XYZ}) = 0, \quad \text{NSI}(\mathcal{L}_V, \mathcal{ZW}) = 0 \quad (22)$$

$$\text{NSI}(\mathcal{ZW}, \mathcal{XYZ}) = \frac{1}{2}, \quad (23)$$

  providing the correct clusters by taking the maximum. We will show that the NSI has the property of being null *iif* the subspaces are orthogonal.

### 3.2. Properties of the NSI

Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be subspaces of $\mathbb{R}^n$, with arbitrary dimensions $d_1$ and $d_2$, and unitary bases $\mathbf{U_1}$ and $\mathbf{U_2}$, respectively. Note that

$$\begin{aligned} tr\{\mathbf{U_1^T U_2 U_2^T U_1}\} &= tr\{\mathbf{U_2 U_2^T U_1 U_1^T}\} && (24)\\ &= tr\{\mathbf{U_2^T U_1 U_1^T U_2}\} && (25)\\ &= \sum_{k=1}^{\min(d_1,d_2)} \sigma_k^2(\mathbf{U_1^T U_2}) && (26)\\ &= \sum_{k=1}^{\min(d_1,d_2)} \cos^2(\theta_k(\mathcal{L}_1, \mathcal{L}_2)) && (27)\\ &\leq \min(d_1, d_2) && (28) \end{aligned}$$

where $\sigma_k$ is the $kth$ singular value of the matrix argument, and $\theta_k$ is the $kth$ principal angle between $\mathcal{L}_1$ and $\mathcal{L}_2$.

Recalling (7), the normalized subspace inclusion (NSI) criterion has the following properties:

- **Basis Independent**: The orthogonal projectors $\mathbf{U_j U_j^T}$ in (24) are unique and independent of the chosen basis.
- **Symmetric**: From (25) we see that $\text{NSI}(\mathcal{L}_1, \mathcal{L}_2) = \text{NSI}(\mathcal{L}_2, \mathcal{L}_1)$. So, the NSI can be used to construct affinity matrices (as in Weiss [14]). However, if $d_1 \neq d_2$, the NSI should be interpreted as "how much" the lower dimensional subspace belongs to the greater dimensional one.
- **normalized**: From (26–28) we have $0 \leq \text{NSI} \leq 1$. This property sets two invariant values for the orthogonal and inclusion configuration, respectively. Also, it allows using the NSI in fuzzy clustering algorithms [16].
- **orthogonality consistency**: Note that

$$\mathcal{L}_1 \perp \mathcal{L}_2 \quad \Leftrightarrow \quad \cos(\theta_k(\mathcal{L}_1, \mathcal{L}_2)) = 0, \quad (29)$$

  $\forall\, k = 1, \ldots, \min(d_1, d_2)$, and, from (27) and (7),

$$\text{NSI}(\mathcal{L}_1, \mathcal{L}_2) = 0 \Leftrightarrow \mathcal{L}_1 \perp \mathcal{L}_2. \quad (30)$$

- **inclusion consistency**: Without loss of generality, let $d_1 \leq d_2$. Then

$$\mathcal{L}_1 \subseteq \mathcal{L}_2 \quad \Leftrightarrow \quad \cos(\theta_k(\mathcal{L}_1, \mathcal{L}_2)) = 1, \quad (31)$$

  $\forall\, k = 1, \ldots, d_1$. Recalling (27) and (7)

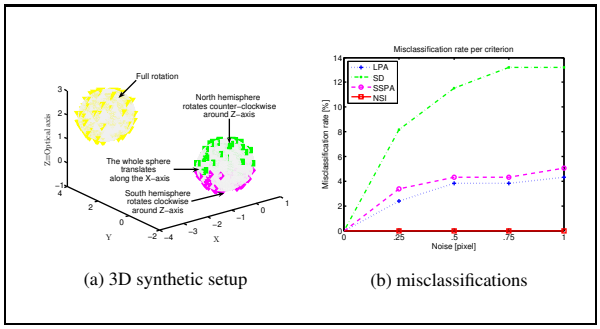
$$\text{NSI}(\mathcal{L}_1, \mathcal{L}_2) = 1 \Leftrightarrow \mathcal{L}_1 \subseteq \mathcal{L}_2. \quad (32)$$

The inclusion consistency property implies that a criterion for subspace clustering, interpreted as a dissimilarity, must not be a distance function, because it will violate the identity of the indiscernibles (one of the metric axioms) that any distance must verify. Unlike other criterion, the NSI is invariant to both the orthogonal and inclusion relationship between subspaces, thus leveraging the subspace clustering results, as shown in the next section.

## 4. Experiments

Our approach (Fig. 3) relies on a first robust segmentation of the union of the subspaces. To implement the GMC , we project the data matrix onto its $4M$-dimensional principal space ($M$ is the number of rigid motions)[4], each resulting point onto the sphere, and let the optimization threshold define the inliers' band for each group. Hence, the whole approach relies solely on the given number of (rigid) motions $M$.

[4] $4M$ is the maximum possible rank of the observations without outliers, *i.e.* it is the true rank of the data matrix only if its range is the union of $M$ independent full motion subspaces.




(a) 3D synthetic setup (b) misclassifications

Figure 5. Synthetic scene: (a) two independently moving spheres: the hemispheres of the right one rotate, in opposite directions, around the world's $z-axis$, while the whole sphere translates; (b) average misclassification rate *per* criterion (1000 trials *per* noise level, considering $512 \times 512$ images).

### 4.1. Synthetic scene

Observing the scene in Fig. 5(a) generates a sequence[5] that generalizes Examples 1–3. It combines almost all possible degeneracies: the features from right sphere support partially dependent, degenerated motion subspaces, because the motion from each hemisphere is perceived has planar (*i.e.* there is no "depth" information). Also, there is transparent motion from the south hemisphere.

The results in Fig. 5(b) were obtained using the same subspace clustering algorithm, only differing the criterion: as long as the initialization is geometrically correct, the NSI guarantees better scene interpretations than the other criteria.

### 4.2. Real data

The Hopkins155 database consists of 155 sequences, with 2 and 3 rigid bodies, divided into three classes (Fig. 6), and the points were automatically tracked, with the tracking errors manually corrected [11]. In order to be consistent with this ground-truth, we need to consider every point as an inlier, thus outliers were assigned to the group supporting its nearest subspace.

Tables 1–3 present the results of our method, together with the LSA (as provided with the Hopkins155 database) and the (sparse projection) ALC the most accurate methods tested on the entire data set (Rao *et al.* [7]).

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 [11], pointing out its robustness to real-world imaging conditions. 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 agglomerative 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).

In table 3[6], 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.

| 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 |

(a) 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 |

(a) 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]).

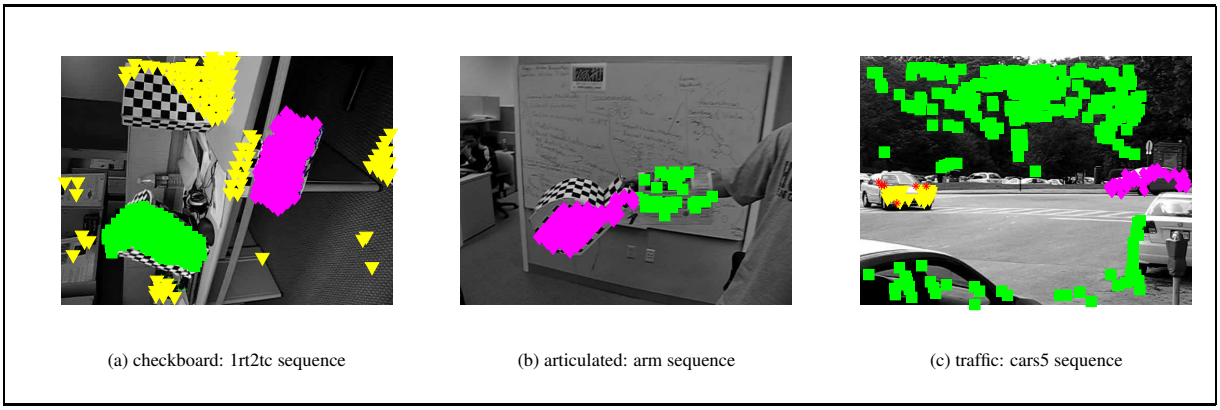
## 5. Conclusion

We introduced the Normalized Subspace Inclusion (NSI), a criterion for subspace clustering consistent with the geometry of the subspaces underlying the observations, unlike other widely used criteria, as our examples demonstrated. We combine it with the GMC [6] to obtain a motion segmentation approach that can cope with motion degeneracies and outliers, and is stable under increasing number of motions.

A natural extension is to unify motion segmentation and kinematic chain construction, two subspace cluster-

[5] 121 frames from a fixed orthographic camera (without occlusion) with the optical axis aligned with the $z - axis$ of the world coordinate system. The inter-frame rotation was 3 degrees and the translation of the right sphere was 3/121 units along the negative direction of the $x - axis$.

[6] These times are essentially indicative (order of magnitude), because they depend on particular implementations. All code was written in $Matlab^{®}$.



(a) checkboard: 1rt2tc sequence (b) articulated: arm sequence (c) traffic: cars5 sequence

Figure 6. Examples from the classes on the Hopkins155 database, with our result superimposed (the red $*$ are points classified as outliers).

| 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. For sequences with $M = 3$ rigid motions sequences, 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 sequences with 2 groups were constructed by splitting each 3 motion sequence into its respective clusters.

ing applications, under the NSI criterion, using non-parametric clustering algorithms. Further research must be conducted towards explicitly imposing the metric constraints.

## 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. 1, 2

[2] G. H. Golub, , and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, $3^{rd}$ edition, 1996. 4

[3] Y. D. Jian and C. S. Chen. Two-view motion segmentation by mixtures of dirichlet process with model selection and outlier removal. *in IEEE ICCV*, 2007. 2

[4] T. K. Kim, J. Kittler, and R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. *IEEE Trans. on PAMI*, 29(6):1005–1018, 2007. 4

[5] K. E. Ozden, K. Schindler, and L. van Gool. Simultaneous segmentation and 3d reconstruction of monocular image sequences. *in IEEE ICCV*, 2007. 1

[6] N. Pinho da Silva and J. P. Costeira. Subspace segmentation with outliers: a grassmannian approach to the maximum consensus subspace. *in IEEE CVPR*, 2008. 2, 6

[7] 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. 2, 3, 6

[8] 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.). 2, 3

[9] X. Sun, L. Wang, and J. Feng. Further results on the subspace distance. *Pattern Recognition*, 40(1):328–329, 2007. 4

[10] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992. 1, 2

[11] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. *in IEEE CVPR*, 2007. 1, 2, 6

[12] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis. *IEEE Transactions on PAMI*, 27(12):1945–1959, 2005. 2

[13] L. Wang, X. Wang, and J. Feng. Subspace distance analysis with application to adaptive bayesian algorithm for face recognition. *Pattern Recognition*, 39(3). 4

[14] Y. Weiss. Segmentation using eigenvectors:a unifying view. *in IEEE ICCV*, 1999. 5

[15] L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *J. Machine Learning Research*, 4. 5

[16] R. Xu and D. Wunsch II. Survey of clustering algorithms. *IEEE Trans. on Neural Networks*, 16(3):645–678, 2005. 5

[17] 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. 1, 2, 4