When Is It Possible to Identify 3D Objects from Single Images Using Class Constraints?

Ronen Basri*   Yael Moses
Dept. of Applied Math
The Weizmann Inst. of Science
Rehovot, 76100, Israel

Abstract

One approach to recognizing objects seen from arbitrary viewpoint is by extracting invariant properties of the objects from single images. Such properties are found in images of 3D objects only when the objects are constrained to belong to certain classes (e.g., bilaterally symmetric objects). Existing studies that follow this approach propose how to compute invariant representations for a handful of classes of objects. A fundamental question regarding the invariance approach is whether it can be applied to a wide range of classes. To answer this question it is essential to study the set of classes for which invariance exists. This paper introduces a new method for determining the existence of invariance for classes of objects together with the set of images from which these invariance can be computed. We develop algebraic tests that, given a class of objects undergoing affine projection, determine whether the objects in the class can be identified from single images. In addition, these tests allow us to determine the set of views of the objects which are degenerate. We apply these tests to several classes of objects and determine which of them is identifiable and which of their views are degenerate.

1 Introduction

Inferring the identity of objects despite variations due to changes in viewing position is a fundamental problem in object recognition. Two common approaches to solving this problem are the model-based and class-based invariance approaches to recognition. Model-based methods (e.g., [5, 9, 10, 13, 22]) approach recognition by storing a finite library of object models in the system memory. Given an image, the identity of objects in the image is determined by comparing the image to the models in the library. To determine whether a given object can produce the image the object model must contain 3D information of the object. Therefore, model construction generally requires the acquisition and matching of two or more images of the object.

Class-based invariance offers an alternative to the model-based approach. In this method objects are recognized by extracting image properties that are invariant to changes in viewing positions. Invariance has successfully been used for certain classes of objects, e.g., planar, bilateral symmetric, or polyhedral objects [8, 15, 16, 17, 18, 19, 20, 23], see a recent review in [26]. A recognition system that uses the invariance approach typically proceeds in two stages. In the first stage the class of the object is recognized, and in the second stage the invariant properties of the class are used for identification. The advantage of using invariance for such classes is twofold. First, invariance is computed from single images; hence also object models can be constructed from single images. Second, invariance distinguishes between objects even if models for these objects are not stored in the system memory. Despite its appeal, the class-based invariance approach cannot be used directly as a method for identifying general 3D objects from single 2D images since, as has been shown in [3, 4, 15], single 2D images of general 3D objects exhibit no viewpoint invariance. The scope of the invariance approach, therefore, is limited to certain classes of objects.

It is therefore of interest to study whether the invariance approach can be extended to handle a wide range of classes of objects. Existing studies are limited to merely a handful of examples of particular classes. These studies, however, offer no general tools to extend their results to other classes of objects. In this paper we develop a method to determine the existence of invariance for classes of objects. (The problem of determining the class of the object is not addressed in this paper.) We consider only invariance that can distinguish between *all* the different objects in a given class (as opposed to invariance that distinguishes between subsets of the objects, in which case different objects may give rise to the same invariant values). We call these classes *identifiable classes*. We introduce necessary and sufficient conditions for such invariance to exist. For the identifiable classes we can also use our method to detect the *degenerate views*, views from which the invariance cannot be computed. Our method is general and does not depend on a specific choice of a class. However, specific tests must be developed for each projection model. In this paper we present tests for the affine projection model.

The rest of the paper is organized as follows. Section 2 introduces our approach and the general framework. In Section 3 we use this framework to develop algebraic tests for classes of objects that undergo affine projection, and in Section 4 we apply these tests to a number of classes of objects.

2 General Framework

Our approach is based on the following principles. A system can identify an object unequivocally from a given image if and only if that object is the only object, among the objects of interest, that can produce the given image. (Of course one can design systems that identify objects according to, say, maximal-likelihood principles. We do not deal with such a framework here.) Clearly, no recognition algorithm can discriminate between two objects, O and O', that share all their images. We call such objects *equivalent* if and only if every image of O is also an image of O' and vice versa. The choice of projection model determines the set of equivalent objects. For example, under the affine projection model every two objects that are related by a 3D affine transformation are equivalent. When two objects are not equivalent (*different*) they may still share some of their images. We call all these images *ambiguous*. In this case too, no recognition algorithm can identify an object from an ambiguous image.

If two objects share an image, they will also share other images which differ by 2D transformations from each other. Such images are called *equivalent images*. For a given object, we define a *view* to be the set of all equivalent images. It is clear that if two objects share an image they will also share the set of images that are in the same view. (Similar to the equivalent objects, the set of equivalent images is determined by the projection model.) It follows that any recognition function cannot identify an object from any of the images in its set of *ambiguous views*. We will therefore alternately consider in the rest of this paper ambiguous views and ambiguous images.

The following proposition summarizes the above observation:

Proposition 1: *An object $O \in C$ is identifiable from a view $\hat{v}$ if and only if there exists no different object $O' \in C$ that shares the view $\hat{v}$ of O.*

Our task then is to determine, given a class of objects, the set of views from which the objects can be identified. We therefore develop a method for exploring the set of ambiguous views. If we find that all the images in the class are ambiguous we conclude that the class is not identifiable from any view. If some of the images are non-ambiguous, then the class is identifiable, and the ambiguous images are considered degenerate.

Below we develop tests to determine the set of ambiguous images of a class. These tests depend on the choice of projection model. In this paper we introduce algebraic tests for objects (given as 3D point sets) undergoing affine projection (that is, 3D-to-2D affine transformation).

Before we present these tests, let us first consider an extreme case in which the class of objects contains all the objects described by n-tuples of 3D points. It is easy to show that, for this class of objects, every image is ambiguous (there are at least two different objects that can project to the image). It follows that it is impossible to identify uniquely these objects from any of their views by using the invariance approach (unless further constraints are imposed on the class). This is a weaker version of the result that for this class there exists no view invariance [3, 4, 15]. It follows that to define a recognition function it is necessary to restrict the domain of the recognition function (the set of images for which the recognition function must identify correctly the set of objects). In our work we restrict this domain by imposing two types of constraints. First we constrain the set of objects considered by the system (e.g., the class of bilaterally symmetric objects as opposed to the set of general 3D objects). In addition, we may constrain the set of images considered for each object in the class (e.g., excluding the side viewing direction of planar objects). Note that when we consider all 3D point sets it is not sufficient to restrict only the set of views since for this class of objects every view is ambiguous.

Another extreme case of class constraints is provided by the model-based approach to recognition. In this paradigm only finite sets of objects are considered. For a finite set of objects the set of ambiguous views is finite (at most quadratic in the number of objects in the set), and so by excluding the ambiguous views we obtain identifiable classes. Thus, every finite class of objects is identifiable. Below we concentrate on infinite sets of classes and we consider the case of objects that undergo affine projection.

Note that ambiguous views, which we also refer to as *degenerate* views, are related but not identical to *accidental* views. The term "accidental view" often is used to describe a view of an object from which a non-stable image is obtained (see, e.g., [24]). That is, a view is accidental if a small change in the viewing direction will cause a large change in the appearance of the object. Such instability often is caused as a result of a change of aspect of the object due to self occlusion. Degenerate views, in contrast, are views from which an object appears identical to another object of the same class. Degeneracy, therefore, is a property which depends on the class, while accidentalness is a property of the object irrespective of the class. In many cases accidental views are also degenerate. For instance, consider the side views of a planar objects. These views are accidental, since the object appears as a line. Likewise, within the class of planar objects these views are degenerate since many different planar objects can be confused from images that were taken from these viewing directions.

3 Identification under affine projection

In this section we develop tests for identification of 3D objects undergoing affine projection. An image is an affine projection of a given object if it is obtained by applying a 3D affine transformation to the object followed by an orthographic projection. Affine projection is the model of choice in several recognition studies [10, 11, 12, 21, 22] (see [10, 1] for more on this

transformation). The tests are applied to objects given as ordered 3D point sets. We assume here that all the points are visible in every considered view. It is possible, however, to use our results in cases of partial (or self) occlusion by considering subsets of the points of the objects.

3.1 Notation

Given an object O consisting of n ordered points $\mathbf{p}_1 = (x_1, y_1, z_1)^T, ..., \mathbf{p}_n = (x_n, y_n, z_n)^T$, we write O in a matrix form as a $3 \times n$ matrix $O = [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n]$. Further, we denote the rows of O by $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{R}^n$ and by row(O) the row space of O (which is the vector space spanned by $\mathbf{x}, \mathbf{y}, \mathbf{z}$). Throughout the paper we shall assume that the object points, $\mathbf{p}_1, ..., \mathbf{p}_n$ are non-coplanar. This implies that rank(row(O) $\cup$ {$\mathbf{1}$}) = 4. We shall also assume that n is sufficiently large to induce invariance whenever such invariance exits. Note that we consider two objects that contain the same set of points but ordered differently as different objects. This assumption can be dropped for some of our results (see Section 5).

We denote an image by a $2 \times n$ matrix I, that consists of the location of the object points in the image. Let I be an image obtained as a result of an affine projection of O. I is produced by applying a 3D affine transformation to O followed by an orthographic projection. Specifically, the image position $\mathbf{q}_i \in \mathcal{R}^2$ of an object point $\mathbf{p}_i \in \mathcal{R}^3$ of O taken from a view $\hat{\mathbf{v}}$ is given by $\mathbf{q}_i = A_{\hat{\mathbf{v}}} \mathbf{p}_i + \mathbf{t}$, where $A_{\hat{\mathbf{v}}}$ is a 2×3 linear matrix of rank 2 and $\mathbf{t} \in \mathcal{R}^2$. The subscript $\hat{\mathbf{v}}$ represents the viewing direction from which the image is observed. This will be defined below. The result of applying this affine transformation to O can be written in a matrix form as $I = A_{\hat{\mathbf{v}}} O + \mathbf{t} \mathbf{1}^T$. Two objects, O and O', share an image I if there exist two 2×3 matrices of rank 2, $A_{\hat{\mathbf{v}}}$ and $A_{\hat{\mathbf{u}}}$, and two vectors, $\mathbf{t}, \mathbf{s} \in \mathcal{R}^2$ such that

$$I = A_{\hat{\mathbf{v}}_1} O + \mathbf{t} \mathbf{1}^T = A_{\hat{\mathbf{v}}_2} O' + \mathbf{s} \mathbf{1}^T$$

A view under the affine projection model is the set of images that are 2D affine equivalent (namely, related by a 2D affine transformation). That is, both I and $I' = AI + \mathbf{t} \mathbf{1}^T$ (where A is 2×2) are affine equivalent and therefore belong to the same view of O. Clearly, whenever two objects share an image from a given view, they will also share all images in that view. An affine view $\hat{\mathbf{v}}$ is defined by the set of matrices $AA_{\hat{\mathbf{v}}}$, where A is an arbitrary 2×2 non-singular matrix. It is not difficult to see that under affine projection a view is determined by a unit vector $\hat{\mathbf{v}} \in \mathcal{R}^3$ such that $A_{\hat{\mathbf{v}}} \hat{\mathbf{v}} = 0$ (or, in other words, by a point on a unit sphere). Since $AA_{\hat{\mathbf{v}}} \hat{\mathbf{v}} = 0$ for every 2D affine matrix A, the vector $\hat{\mathbf{v}} \in \mathcal{R}^3$ uniquely determines the set of affine-equivalent images that were taken from viewing direction $\hat{\mathbf{v}}$. Below we shall use the notation $A_{\hat{\mathbf{v}}}$ to denote a 2×3 matrix of a given view, $\hat{\mathbf{v}}$.

Given two objects, O and O', we define their $7 \times n$ joint matrix $\bar{J}(O, O')$ to contain the rows of O and O' and the vector $\mathbf{1}$ stacked on top of each other in the

form:

$$\bar{J}(O, O') = \begin{bmatrix} O \\ O' \\ \mathbf{1}^T \end{bmatrix} \qquad (1)$$

In Section 3.2 below we show that the rank of $\bar{J}(O, O')$ can be used to determine whether O and O' are affine equivalent, share a view, or are completely disjoint.

3.2 Tests for identification

Two objects, O and O', are called *affine-equivalent* if there exist an affine transformation, a 3×3 non-singular matrix A, and a vector $\mathbf{t} \in \mathcal{R}^3$, such that $O = AO' + \mathbf{t} \mathbf{1}^T$. In this case every image of O is also an image of O' and vice versa (see Proposition 2 below). Clearly, under the affine projection, it is impossible to distinguish between objects that are affine-equivalent. Otherwise, if O' cannot be obtained from O by applying an affine transformation to O then the two objects are called *affine-different*. Following is a list of necessary and sufficient conditions for affine equivalence.

Proposition 2: *The following conditions are equivalent:*

(a) $O = AO' + \mathbf{t} \mathbf{1}^T$ for some 3×3 non-singular matrix A and a vector $\mathbf{t} \in \mathcal{R}^3$.

(b) rank($\bar{J}(O, O')$)=4.

(c) Every image of O is also an image of O' and vice versa.

(d) O and O' share more than a single view.

To determine the set of ambiguous views of a class we need to develop necessary and sufficient conditions for two affine different objects to share a view. These conditions are specified in Proposition 3 below.

Proposition 3: *The following conditions are equivalent:*

(a) There exist a single direction $\hat{\mathbf{v}} \in \mathcal{R}^3$ and translation $\mathbf{t} \in \mathcal{R}^2$ such that $A_{\hat{\mathbf{v}}} O = A_{\hat{\mathbf{u}}} O' + \mathbf{t} \mathbf{1}^T$, where $A_{\hat{\mathbf{v}}}$ and $A_{\hat{\mathbf{u}}}$ are 2×3 matrices of rank 2.

(b) rank($\bar{J}(O, O')$)=5.

(c) $O = AO' + \mathbf{t} \mathbf{1}^T + \hat{\mathbf{v}} \eta^T$, where A is a 3×3 matrix and rank(A) ≥ 2, $\mathbf{t} \in \mathcal{R}^3$, $\hat{\mathbf{v}} \in \mathcal{R}^3$ is a non-zero vector, and $\eta \in \mathcal{R}^n$ is orthogonal to row(O')$\cup$\{$\mathbf{1}$\}. (O' share the view $\hat{\mathbf{v}}$ of O.)

(d) $O = AO' + \mathbf{t} \mathbf{1}^T + \hat{\mathbf{v}} \eta^T$, where A is a 3×3 matrix and rank(A) = 3, $\hat{\mathbf{v}} \in \mathcal{R}^3$ is a non-zero vector, and $\eta \in \mathcal{R}^n$ where $\eta \notin$ row(O') $\cup$ {$\mathbf{1}$}. (O' share the view $\hat{\mathbf{v}}$ of O.)

The proof of Propositions 2–3 is given in [2]. To determine whether a class is identifiable we use the above propositions in conjunction with Proposition 1.
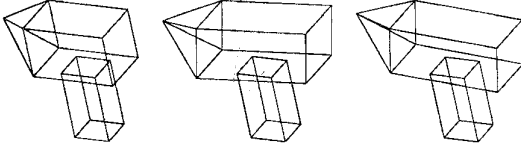
543

Figure 1: Three hammer-like shapes composed of identical parts but differ in the shapes of their heads by generic affine transformations.
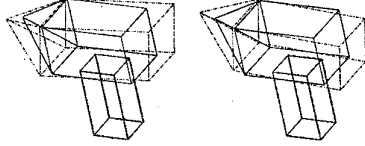

Figure 2: An overlay of the leftmost hammer (solid) with the other two hammers (dashed). Notice that the handles of the hammers are identical whereas their heads differ.

4 Examples

In this section we apply the tests developed in Section 3 to several classes of objects. For each class we determine the set of views from which the objects of the class can be identified. Some of these classes are shown to be identifiable from most or all views. Other classes are shown to be non-identifiable from any view. We begin with classes of objects which are composed of the same set of parts, but which the relative position and orientation of their parts may vary across objects (Section 4.1). We then proceed to discuss classes of objects which have two identical parts, except that one part may appear at different position and orientation with respect to the second part. In this case the shape of the parts may vary across objects (Section 4.2). Next, we consider classes that contain combinations of sub-structures, that is one part of the object is a function of the other parts (Section 4.3).

4.1 Objects with same set of parts

Many classes of interest contain objects all of which are composed of the same set of parts. The identity of an object within the class is defined by the exact position and orientation of one part with respect to another, and by the relative size or stretch of one part with respect to another. An example is given in Figure 1, where three hammer-like shapes with identical handles and heads which differ by affine transformations are shown. Formally, a class C_P is defined by the set of parts $\{Q_1, ..., Q_m\}$ where Q_i are $3 \times n_i$ matrices for some arbitrary $n_i > 3$. An object $O \in C_P$ is given by $O = [P_1, ..., P_m]$, where the $3 \times n_i$ matrices $P_i = f_i(Q_i)$ describe the parts of O. The functions f_i determine the identity of the objects in the class.

Below we limit our discussion to $\{f_i\}$ that represent affine transformations. In this case, an object $O \in C_P$ can be written in a matrix notation as $O = [P_1, ..., P_m]$ where $P_i = B_i Q_i + s_i 1^T$ for $1 \leq i \leq m$. B_i are 3×3 non-singular matrixes and $s_i \in \mathcal{R}^3$ which vary between the different objects in
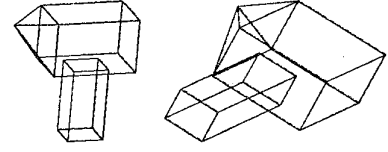

Figure 3: An overlay of the leftmost hammer (solid) and each of the other two hammers (dashed) taken from their common views. The common views are (0.29,-0.41,0.87) and (0.32,-0.73,0.60). Notice that for every view of the leftmost hammer we can produce infinitely many different hammer-like shapes that look identical to this hammer.

the class. Below we assume that every part Q_i is non-planar, that is, $rank(Q_i \cup \{1\}) = 4$.

The next proposition establishes that if the parts of the objects in the class may undergo general affine transformations between objects parts then the class is non-identifiable from any view. If however we restrict the transformation applied to the parts to scaling and stretching along the primary axes we obtain classes which are identifiable from almost all views. This implies, in particular, that it is possible in principle to extract the relative scale and stretch of objects' parts from single images.

Proposition 4: *Given* $3 \times n$ *non-planar matrices* Q_i $(1 \leq i \leq m)$ *denoting an object part, the class* $C_P = \{[P_1, ..., P_m] \mid P_i = B_i Q_i + s_i 1^T]\}$ *is:*

1. *Non-identifiable for general affine transformation, that is, for arbitrary* 3×3 *non-singular matrices* B_i *and constant* $s_i \in \mathcal{R}^3$. *(Consequently, the class is non-identifiable also if we let* s_i *be arbitrary.)*

2. *Non-identifiable for pure translation, that is, for* $B_i = I$ *and arbitrary* $s_i \in \mathcal{R}^3$.

3. *Identifiable from almost all views for scaling and stretching along the primary axes, that is, for diagonal* B_i *and constant* $s \in \mathcal{R}$. *In this case degenerate views are obtained only for* $\hat{v} = (v_1, v_2, v_3)^T$ *such that* $v_i = 0$ *for some* $1 \leq i \leq 3$.

Proof: (1) To show this we need to show that for every choice of object $O \in C_P$ and every choice of view $\hat{v}$ there exists an affine-different object $O' \in C_P$ that shares the view $\hat{v}$ of O. Let $O = [P_1, ..., P_m]$ where $P_i = B_i Q_i + s_i 1^T$ for $1 \leq i \leq m$, and let $O' = [P_1', ..., P_m']$ where $P_i' = B_i' Q_i + s_i 1^T$ for another set of B_i' and for $1 \leq i \leq m$. Since the matrices B_i determine the identity of the object it follows that given $\{B_i\}$ and $\hat{v}$ we need to find matrices $\{B_i'\}$ and a view $\hat{u}$ such that O and O' will share an image. We set $B_1' = B_1$ and $B_i' = B_i + \hat{v} r_i^T$ for some non-zero vector $r_i \in \mathcal{R}^3$ such that $rank(B_i') = 3$ and $B_i' \neq I$. It is readily verified that O and O' are affine-different (since the first parts of O and O' are related by the identity matrix, while the other parts are not). Furthermore, the two objects

O and O' share an image when both are taken from viewing direction $\hat{\mathbf{v}}$. That is, $A_{\hat{\mathbf{v}}}O = A_{\hat{\mathbf{v}}}O'$. To prove it, it is sufficient to show that $A_{\hat{\mathbf{v}}}P_i' - A_{\hat{\mathbf{v}}}P_i = 0$. This can be easily verified as follows.

$$A_{\hat{\mathbf{v}}}(P_i' - P_i) = A_{\hat{\mathbf{v}}}(B_i' - B_i)Q_i = A_{\hat{\mathbf{v}}}\hat{\mathbf{v}}r^T Q_i,$$

and since $A_{\hat{\mathbf{v}}}\hat{\mathbf{v}} = 0$ we obtain that $A_{\hat{\mathbf{v}}}(P_i' - P_i) = 0$.
(2) To show this we need to show that for every choice of object O (determined by s_i) and view $\hat{\mathbf{v}}$ there exists an affine different object O' (determined by $s_i' \neq s_i$) that shares the view $\hat{\mathbf{v}}$ of O. Let $O = [P_1, P_2, .., P_m]$ such that $P_i = Q_i + s_i\mathbf{1}$, and let $O' = [P_1', P_2', .., P_m']$ such that $P_i' = Q_i + s_i'\mathbf{1}$. We set $s_1' = s_1$ and $s_i' = s_i + \lambda_i\hat{\mathbf{v}}$ for $2 \leq i \leq n$ for some arbitrary $\lambda_i \neq 0$. In this case O and O' are affine-different since their first parts are related by a different translation than their second parts, and they share the view $\hat{\mathbf{v}}$ of O since

$$A_{\hat{\mathbf{v}}}(P_i' - P_i) = A_{\hat{\mathbf{v}}}(s_i' - s_i) = \lambda_i A_{\hat{\mathbf{v}}}\hat{\mathbf{v}} = 0.$$

(3) To show this we need to show that given an object O and a view $\hat{\mathbf{v}}$ any object O' that shares the view $\hat{\mathbf{v}}$ of O must be affine equivalent to O. Let $O = [P_1, ..., P_m]$ where $P_i = B_iQ_i + s_i\mathbf{1}^T$, and let $O' = [P_1', ..., P_m']$ where $P_i' = B_i'Q_i + s_i\mathbf{1}^T$ for another set of B_i'. Since O and O' share the view $\hat{\mathbf{v}}$ of O then there exists a view $\hat{\mathbf{u}}$ such that

$$A_{\hat{\mathbf{v}}}O = A_{\hat{\mathbf{u}}}O' + t\mathbf{1}^T$$

Without loss of generality, we can assume that $B_1 = B_1' = I$ and $s_1 = s_1' = 0$ (since we can choose to consider any of the objects that are affine equivalent to O'). This implies that

$$A_{\hat{\mathbf{v}}}P_1 = A_{\hat{\mathbf{u}}}P_1' + t\mathbf{1}^T$$
$$A_{\hat{\mathbf{v}}}(B_iQ_i + s_i\mathbf{1}^T) = A_{\hat{\mathbf{u}}}(B_i'Q_i + s_i\mathbf{1}^T + t\mathbf{1}^T).$$

The first equation implies that $\hat{\mathbf{u}} = \hat{\mathbf{v}}$ and that $t = 0$ (since $P_1 = P_1'$ and P_1 is non-planar), and so the second equation becomes

$$A_{\hat{\mathbf{v}}}(B_i - B_i')Q_i = 0.$$

Since Q_i is non-planar we obtain that $A_{\hat{\mathbf{v}}}(B_i - B_i') = 0$. If $B_i = B_i'$ for $1 \leq i \leq m$ then O and O' are identical. It follows that $B_i - B_i' = \hat{\mathbf{v}}r^T$ for some $r \neq 0$ (since the only vector $\mathbf{w}$ for which $A_v\mathbf{w} = 0$ is of the form $\mathbf{w} = \lambda\hat{\mathbf{v}}$). However, since both B_i and B_i' are diagonal also $\hat{\mathbf{v}}r^T$ must be diagonal. But since $\hat{\mathbf{v}}$ contains no zero components, the only way to obtain zeros in all non-diagonal components is by setting $r = 0$. This, however, will imply again that $B_i = B_i'$ for $1 \leq i \leq m$, and so O and O' are identical. $\square$

An example to these results is shown in Figures 1-3. In these figures we show three hammer-like objects with identical handles, but their heads differ by an arbitrary linear transformation. We constructed these hammers as follows. We first constructed the leftmost hammer. Then, we arbitrarily selected two views and modified the head of the hammer according to the proof of Proposition 4(1) so as to obtain two different hammers that will share those views with our original hammer. Note that according to the proposition we could do this to any desired view, and that at every view we could find infinitely many different hammers (determined by the choice of r) which share this view with our original hammer. It follows therefore that the class of objects with the same parts which may differ by an arbitrary affine transformation is non-identifiable. Notice that according to Proposition 4(3) if we allow the parts to only scale and stretch in the directions of the primary axes we would not be able to find two hammers with a common view except for a small set of degenerate views.

A special case of this proposition is the case that the class composed of objects of two identical parts ($Q_1 = Q_2$). Notice that the actual shape of the parts was not used in the proof, so even the negative results extend to classes such that all the parts of a given object are identical. In Section 4.2 below we further explore the case of objects such that the parts of a given object are identical, but the parts may differ across the objects in the class.

We can conclude that a set of objects that consist of identical set of parts cannot be identified if the objects differ by arbitrary affine transformation or translation of their parts. However, if the objects in the class differ by scaling or stretching of the parts, then they can be identified from almost all of their images.

4.2 Repeated structures

The next set of classes that we consider consist of objects each of which contains two identical, non-planar parts except that these parts are related by an affine transformation. An object from the class has the following form:

$$O = [P, BP + s\mathbf{1}^T].$$

In contrast to the classes discussed in Section 4.1 now the shape of the parts (denoted by P) may also vary across objects. We consider below two cases. First we consider the case that the affine transformation relating the two parts may vary across objects. Then we consider the case that this affine transformation is the same for all the objects in the class. A special case for this latter class is the class of bilaterally symmetric objects, in which the identical parts are related by a reflection.

Invariants for objects with repeated structures undergoing 3D-to-2D projective transformations were introduced in [16]. In addition, invariants for bilaterally symmetric objects and objects composed of planar repeated structure under both affine and projective transformations were presented in [7, 14, 15, 17, 20]. These studies showed that the class of repeated structures induces invariance on the set of images. The basic intuition is that, when an image of an object with repeated structures is given, this image will in general contain two copies of the same structure. Thus, under the appropriate conditions, the shape of the repeated part can be recovered up to an affine (or pro-

jective) transformation by simply using a stereo algorithm [6, 11]. Nonetheless, these studies do not determine if in general the identity of the objects can be determined from single images. Two objects with repeated structures may have exactly the same parts, but these parts may differ in their relative location or size across the objects. These studies determine the shape of the parts, but leave the relative position and size of the parts unknown. Below we show that if we permit the parts to be related by an arbitrary transformation then the class will not be identifiable (and the objects which have the same parts will be confused). This implies that the relationships between the parts cannot be uniquely recovered from single images, which is equivalent to saying that the problem of calibrating an affine camera from two images is inherently ambiguous. If however we consider a class of objects with repeated structures in which the transformation relating the parts is the same across objects (such as in the case of bilaterally symmetric objects) then for most choices of transformation the class is identifiable from almost all views.

It is straightforward at this point to show that the class of objects with repeated structures whose parts are related by an arbitrary affine transformation is not identifiable from any view. This can be seen by applying Proposition 4(1) with $Q_1 = Q_2$, which tells us that a subset of this class, the objects that have the same parts, cannot be identified from all their views. Consequently, no invariants can distinguish between all the objects in this class.

Next we consider the case that the same affine transformation relates the two parts in every object. We show that for most choices of an affine transformation the class determined by this transformation is identifiable from almost all views. The degenerate views in this case lie along at most three great circles on the viewing sphere. We list these views in Proposition 5.

Proposition 5: *Given a 3×3 non-singular matrix B and a vector $s \in \mathcal{R}^3$, the class $\mathcal{C}_{B,s}$ is identifiable from all views $\hat{v} \in \mathcal{R}^3$ unless $\hat{v}$ is an eigenvectors of B or $\hat{v}$ is located on at most three great circles on the viewing sphere which depend on B. (The additional degenerate views are listed in Table 1.)*

The proof is given in [2]. The additional degenerate views listed in Table 1 depend on the matrix B, and, in particular, on the number of eigenvectors and eigenvalues of B. The degenerate views include all the eigenvectors of B. For convenience, we list the degenerate views according to the Jordan form of the matrix B, $\bar{B} = A^{-1}BA$. The views are given in terms of v', where $v' = A^{-1}v$. Note that given a vector v' it is straightforward to recover the corresponding view $\hat{v} = Av'$. Furthermore, $\hat{v}$ is an eigenvector of B if and only if v' is an eigenvector of the Jordan form $\bar{B}$. Note also that in certain cases the actual list of degenerate views is smaller than what is specified in Table 1 since the list of vectors v' includes vectors that correspond to complex views $\hat{v}$. Evidently, only real view vectors are geometrically feasible.

	No. of Eigen-vectors	No. of Eigen-values	Constraints on v'
a	3	3	$v'_i = 0$ for some $1 \leq i \leq 3$.
b	3	2	$v'_1 = 0$.
c	3	1	no identification from any view.
d	2	2	$v'_3 = 0$ or $v'_2 = 0$.
e	2	1	only eigenvectors.
f	1	1	$v'_3 = 0$ or $v'_1 = v'_2 = 0$.

Table 1: List of additional degenerate views for the class $\mathcal{C}_{B,s}$ listed according to the number of eigenvalues and eigenvectors of B.

As an example consider the class of bilaterally symmetric objects. In this case the matrix B representing reflection about a plane is given by $B = \text{diag}(1, 1, -1)$ and $s = 0$. B has three eigenvectors and two different eigenvalues, corresponding to the second row of the table. Since B is already in its Jordan form then $v' = \hat{v}$. Accordingly, bilaterally symmetric objects are identifiable from all views except for those which coincide with the symmetry plan, $\hat{v} = \alpha[1, 0, 0] + \beta[0, 1, 0]$, and the direction perpendicular to the symmetry plan, $\hat{v} = [0, 0, 1]$.

4.3 Combinations of sub-structures

Next we consider classes of objects in which one part can be expressed as a linear combination of the other parts. Below we consider only classes of objects with three or more sub-structures. When the number of parts is two we obtain the case of repeated structures, which was discussed in Section 4.2. One reason we are interested in such classes is that it is possible to build such classes such that their objects will have the same degrees of freedom as planar objects do ($2n$, where n is the number of points on the object). Yet, as is shown below, unlike the class of planar objects these classes are not identifiable from any view.

In a class that contains a combination of sub-structures an object is divided into $m \geq 2$ parts. The location of every point on the m'th parts can be expressed as a linear function of the $m - 1$ corresponding points in the other parts, where the same linear function is applied to all the points. The shape of an object of this class, therefore, is given by $O = [P_1, P_2, ..., P_m]$, where P_i are $3 \times n$ matrices defining the shape of the i'th part ($1 \leq i < m$), B_i are 3×3 non-singular matrices, $s \in \mathcal{R}^3$, and

$$P_m = B_1 P_1 + B_2 P_2 + ... + B_{m-1} P_{m-1} + s \mathbf{1}^T$$

In Proposition 6 below we show that for $m > 2$, identification is impossible from all views even when all the matrices B_i and s are the same for all objects in the class.

Proposition 6: *Given 3×3 non-singular matrices B_i where $1 \leq i \leq m-1$, $m \geq 3$, and given $s \in \mathcal{R}^3$, the class $\mathcal{C}_{B_1,...,B_{m-1},s}$ is not identifiable from any view.*

Proof: Here we prove the proposition for $m > 3$. The proof for $m = 3$ is given in [2]. To show this we need to show that for every choice of object O and view $\hat{\mathbf{v}}$ there exists an affine different object O' and a view $\hat{\mathbf{u}}$ so that O and O' share the view $\hat{\mathbf{v}}$ and $\hat{\mathbf{u}}$ respectively. According to Proposition 3(**d**) it is sufficient to show that there exists an object O' in the class such that $O = O' + \mathbf{v}\eta^T$ for $\eta \perp O$. (We take $A = I$ and $\mathbf{t} = 0$.)

Let $O = [P_1, \; P_2, \; ..., P_m]$, for $P_m = \sum_{i=1}^{m-1} B_i P_i + \mathbf{s}\mathbf{1}^T$. We set $O' = [P'_1, P'_2, ..., P'_m]$ such that

$$P'_i = P_i + \mathbf{v}\eta_i^T \quad \text{for} \quad (1 \le i \le m), \qquad (2)$$

where $\eta^T = [\eta_1^T, \eta_2^T, ..., \eta_m^T]$. The object O' belongs to the class if and only if there exists a vector η such that the m'th part of O' can be written in term of the first $m - 1$ parts. That is, there exists a vector η such that

$$P'_m = \sum_{i=1}^{m-1} B_i P'_i + \mathbf{s}\mathbf{1}^T \qquad (3)$$

From Eq. 2 (the case $i = m$) and Eq. 3 we obtain that

$$P'_m = P_m + \hat{\mathbf{v}}\eta_m^T = \sum_{i=1}^{m-1} B_i P'_i + \mathbf{s}\mathbf{1}^T \qquad (4)$$

Plugging in the first $m - 1$ equations in Eq. 2 into Eq. 4 and rearranging we obtain

$$\sum_{i=1}^{m-1} B_i \hat{\mathbf{v}}\eta_i^T - \hat{\mathbf{v}}\eta_m^T = 0. \qquad (5)$$

It is left to show that for every view $\hat{\mathbf{v}}$ there exists a non-trivial vector $\eta^T = [\eta_1^T, ..., \eta_m^T]$ that solves the above equations subject to the constraint $O\eta = 0$. This last equation and Eq. 5 contain $3n + 3$ homogeneous equations in mn unknowns. Since $m \ge 4$, $mn > 3n + 3$ for $n > 3$, and so for any choice of $\hat{\mathbf{v}}$ a non-trivial solution will exist. It follows that for every $\hat{\mathbf{v}}$ there exists an object given by $O' = O - \hat{\mathbf{v}}\eta^T$ that shares the view $\hat{\mathbf{v}}$ with O. Hence, the class is not identifiable. $\square$

We next discuss the counting argument for the case $m = 3$. In this case, an object has the form $O = [P_1, \; P_2, \; P_3]$ where $P_3 = BP_1 + CP_2 + \mathbf{s}\mathbf{1}^T$, B and C are 3×3 non-singular matrices, and $\mathbf{s} \in \mathcal{R}^3$. Below we assume in addition that $\text{rank}(\bar{J}(P_1, P_2)) = 7$.[1] If we rely only on counting arguments we may be misled to believe that the class $\mathcal{C}_{B,C,\mathbf{s}}$ is identifiable. Consider for example the counting arguments given in [25]. Given an object $O \in \mathcal{C}_{B,C,\mathbf{s}}$, suppose O includes n points. An image of O gives $2n$ measurements. Together with n class constraints (the linear relations between the first two parts and the third part) they give

---
[1] Note that dropping the assumption that $\text{rank}(\bar{J}(P_1, P_2)) = 7$ will not change our final result that $\mathcal{C}_{B,C,\mathbf{s}}$ is not identifiable since extending the class with more objects cannot reduce the set of ambiguous views.

rise to $3n$ equations. The number of variables in these equations are as follows. The shape of O is defined by $3n$ variables. The projection parameters (the parameters of a 3D-to-2D affine transformation) are eight. A 3D affine reference frame can be obtained by picking the position of four points, hence such a frame involves setting 12 parameters. Therefore, the total number of variables is $3n + 8 - 12 = 3n - 4$. According to a counting argument the number of equations obtained, $3n$, is sufficiently large to determine the values of the $3n-4$ unknowns, and so in theory using elimination we should be able to recover the shape of the object from the equations. However, the underlying assumption in counting arguments is that the counted equations are all independent and consistent. In classes which contain combinations of sub-structures this assumption is violated, and so the identity of the objects cannot be recovered from single images, as was shown above in Proposition 6.

5 Summary and Discussion

A fundamental question regarding the invariance approach is whether it can be applied to a wide range of classes. To answer this question it is essential to study the set of classes for which invariance exists. In this paper we investigated the invariant representations that discriminate between all the objects of a given class. We addressed the problem of determining, given a class of objects, the set of images from which the objects can be identified. Our approach is based on exploring the set of ambiguous images. We developed a number of algebraic tests to determine the ambiguous images under affine projection, and applied these tests to a number of classes of objects.

We now consider a number of assumptions we have made and how they should be relaxed in future work. **Projection model:** our tests were developed for objects which may undergo affine projection. We intend in the future to develop similar tests for other, more realistic projection models. We would like to note, however, that some of our results (e.g., Proposition 4) apply also to rigid objects undergoing weak-perspective projection.

Constructive tests: Although our tests can determine which classes are identifiable and from what views, the tests at their present form are not constructive. That is, the tests cannot be used to derive the invariants for the objects. Consequently, our approach may serve only as a first step in deriving invariance for new classes of objects. In particular, it can be used to avoid seeking invariance for classes for which invariance does not exist. In addition, for identifiable classes our method can be used to determine the set of views from which objects are identifiable and exclude the ambiguous views.

Dealing with noise: in our tests we did not take into account the effect of noise on identifiable classes. Obviously, if we allow for noise, more images may become ambiguous. One possible way to detect sensitivity to noise in classes of objects is by looking at the singular values of the matrix $\bar{J}(O_1, O_2)$, which was introduced in Section 3.2. This issue also is left for future research.

The objects: in our tests we assumed that the objects are given as ordered point sets. In particular we regarded two objects that contain the same set of points ordered differently as two different objects. Several of our results are not affected by this assumption. For example, the class of objects with identical parts and the class of objects with repeated structures are closed under permutation of the objects' points. (One only has to find correspondence between the different parts of the same object.) Furthermore, classes which were shown to be non-identifiable remain non-identifiable even if we relax this requirement. We intend in the future to extend our tests to other objects that consists of non-ordered points. In addition, it is clearly of interest to develop tests for contour images, and grey-level images.

Invariance: in this paper we concentrated on analyzing whether given classes are identifiable. For classes that are not identifiable it may still be possible to extract invariants from single images. These invariants will not suffice to discriminate between all the objects of the class, but will distinguish only between subsets of the objects. Developing tests for such classes is left for future research.

Classifying the objects: the first step of applying class-based invariance to images involves classifying the object in the image. Unfortunately, it is impossible to both classify the object and recover its specific identity using invariance, since this will contradict the result that there exists no view invariance that can discriminate between all 3D objects. In practical systems it may be the case that many of the objects belong to a small number of classes, in which case one may enumerate all these classes, or find some properties which distinguish these classes from one another. This problem too is beyond the scope of this paper.

References

[1] Basri, R., 1996. "Paraperspective ≡ affine", *IJCV*, 19(2): 169–180.

[2] Basri, R., and Moses, Y., 1997. "Recognizing 3D Objects from Single Images Using Class Constraints," *The Weizmann Institute of Science, Technical Report CS97-17*.

[3] Burns, J., Weiss, R., and Riseman, E., 1992. "The non-existence of general-case view-invariants", in Mundy, J. and Zisserman, A. (Eds), *Geometric Invariance in Computer Vision*, MIT Press, Cambridge.

[4] Clemens, D. and Jacobs, D., 1991. "Space and time bounds on model indexing", *PAMI*, 13(10): 1007–1018.

[5] Fischler, M.A. and Bolles, R.C., 1981. "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography". *Com. of the A.C.M.*, 24(6): 381–395.

[6] Faugeras, O.D., 1992. "What can be seen in three dimensions with an uncalibrated stereo rig?" *ECCV*: 563-378.

[7] Fawcett, R., Zisserman, A., and Brady, J.M., 1994. "Extracting Structure from an Affine View of a 3D Point Set with One or Two Bilateral Symmetries", *Image and Vision Computing*, 12(9): 615-622.

[8] Forsyth, D., Mundy, J. L., Zisserman, A., Coelho, C., Heller, A., and Rothwell, C., 1991. "Invariant descriptors for 3-D object recognition and pose". *PAMI*, 13(10): 971–991.

[9] Huttenlocher, D.P., and Ullman, S., 1990 "Recognizing Solid Objects by Alignment with an Image", *IJCV*, 5(2): 195–212.

[10] Jacobs, D., 1992. "Space efficient 3D model indexing", *CVPR*: 439–444.

[11] Koenderink, J. and van Doorn, A., 1991. "Affine structure from motion", *J. of the Optic. Soc. of America*, 8(2): 377–385.

[12] Lamdan, Y. & H.J. Wolfson, 1988, "Geometric Hashing: A General and Efficient Model-Based Recognition Scheme," *ICCV*: 238–249.

[13] Lowe, D., 1985, *Perceptual Organization and Visual Recognition*, The Netherlands: Kluwer Academic Publishers.

[14] Mitsumoto, H., Tamura, S., Okazaki, K., Kajimi, N., and Fukui, Y., 1992. "3-D reconstruction using mirror images based on a plane symmetry recovering method". *PAMI*, 14(9): 941–946.

[15] Moses, Y. and Ullman, S., 1992. Limitations of non model-based recognition schemes, *ECCV*: 820–828.

[16] J. L. Mundy and A. Zisserman, 1994. Repeated structures: image correspondence constraints and 3D structure recovery. In J. L. Mundy, A. Zisserman, and D. A. Forsyth (Eds), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science 825, Springer-Verlag.

[17] Poggio, T. and Vetter, T., 1992. "Recognition and structure from one 2D model view: observations on prototypes, object classes and symmetries," *M.I.T., A.I. Memo No. 1347*.

[18] Rothwell, C., Zisserman, A., Forsyth, D. A., and Mundy, J. L, 1992, "Canonical Frames for Planar Object Recognition," *ECCV*: 757-772.

[19] Sparr, G., 1992. "Depth Computations from Polyhedral Images," *Image and Vision Computing*, 10:683–688.

[20] Sugimoto, A., 1994. "Geometric invariant of noncoplanar lines in a single view,". *ICPR*, Vol. 1: 190–195.

[21] Tomasi, C., and Kanade, T., 1992, "Shape and Motion from Image Streams under Orthography: a Factorization Method," *IJCV*, 9(2): 137–154.

[22] Ullman, S., and Basri, R., 1991. "Recognition by linear combinations of models". *PAMI*, 13(10): 992–1006.

[23] Weiss, I., 1988. "Projective invariants of shape," *DARPA Image Understanding Workshop*: 1125–1134.

[24] Weinshall, D., and Werman, M., 1997. "On view likelihood and stability," *PAMI*, 19(2):97–108.

[25] Werman, M., and Shashua, A., 1995. "The study of 3D-from-2D using elimination," *ICCV*: 473–479.

[26] Zisserman, A., Forsyth, D., Mundy, J., Rothwell, C., Liu, J., and Pillow, N., 1995. "3D object recognition using invariance," *Artificial Intelligence*. 78:239–288.